

## Содержание

<b>ГЕОРГИЙ ДМИТРИЕВИЧ КАРПЕЧЕНКО. К 110-ЛЕТИЮ СО ДНЯ РОЖДЕНИЯ (03.05.1899–28.07.1941)</b> <i>М.А. Вишнякова, Н.П. Гончаров</i> .....	7
<b>МОРФОГЕНЕЗ, МОРФОМЕХАНИКА И ГЕНОМ</b> <i>Л.В. Белоусов</i> .....	29
<b>ВЫЯВЛЕНИЕ НОВЫХ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ SREBP В ПРОМОТОРНЫХ РАЙОНАХ ГЕНОВ ПОЗВОНОЧНЫХ НА ОСНОВЕ КОМБИНАЦИИ БИОИНФОРМАТИЧЕСКОГО И ЭКСПЕРИМЕНТАЛЬНОГО ПОДХОДОВ</b> <i>Е.В. Игнатъева, Т.И. Меркулова, Д.Ю. Ощепков, Н.В. Климова, Г.В. Васильев, И.И. Турнаев, В.Ф. Кобзев, Н.А. Колчанов</i> .....	37
<b>ВЫЯВЛЕНИЕ НОВЫХ DRE В РЕГУЛЯТОРНОЙ ОБЛАСТИ ГЕНОВ ЧЕЛОВЕКА, КОДИРУЮЩИХ КОМПОНЕНТЫ ЦИТОЗОЛЬНОГО КОМПЛЕКСА АРИЛ-ГИДРОКАРБОНОВОГО РЕЦЕПТОРА</b> <i>Д.Ю. Ощепков, Д.П. Фурман, Е.А. Ощепкова, А.В. Катохин, М.Ю. Шаманина, В.А. Мордвинов</i> .....	46
<b>ЦИТОКИНЫ: БИОЛОГИЧЕСКИЕ СВОЙСТВА И РЕГУЛЯЦИЯ ЭКСПРЕССИИ ГЕНА ИНТЕРЛЕЙКИНА-5 ЧЕЛОВЕКА</b> <i>В.А. Мордвинов, Д.П. Фурман</i> .....	53
<b>DNA – «PROGRAMMING LANGUAGE OF LIFE»</b> <i>R. Hofestädt</i> .....	68
<b>DOMAIN – WIDE LANDSCAPE OF HUMAN GENOME</b> <i>D.A. Maximov, V.N. Babenko</i> .....	73
<b>HORIZONTAL TRANSFER OF NON-LTR RETROTRANPOSONS</b> <i>O.S. Novikova, V. Fet, A.G. Blinov</i> .....	76
<b>КОРРЕЛЯЦИИ ОПЕРОННОЙ СТРУКТУРЫ С ДЛИНОЙ ГЕНОМА У 14 ВИДОВ МИКОПЛАЗМ</b> <i>С.А. Лащин, Ю.Г. Матушкин, Т.М. Хлебодарова, В.А. Лихошвай</i> .....	84

<b>INTRON-EXON PATTERNS AS A POTENTIAL TOOL IN STUDYING GENE EVOLUTION</b> <i>A. Ruvinsky</i> .....	91
<b>GENETIC LINKAGE ANALYSIS CHALLENGES ON A DISTRIBUTED GRID ENVIRONMENT</b> <i>A. Calabria, D. Di Pasquale, A. Orro, G. Trombetti, M. Gnocchi, L. Milanesi</i> .....	96
<b>ПРИМЕНЕНИЕ НЕМЕТРИЧЕСКОГО МНОГОМЕРНОГО ШКАЛИРОВАНИЯ ДЛЯ МУЛЬТИПЛАТФОРМЕННОЙ ОБРАБОТКИ МИКРОЧИПОВЫХ ЭКСПРЕССИОННЫХ ДАННЫХ</b> <i>В.М. Ефимов, А.В. Катохин</i> .....	102
<b>A COMBINATORICS-BASED DATA-MINING APPROACH TO TIME-SERIES MICROARRAY ALIGNMENT</b> <i>N. Turenne, I. Hue</i> .....	109
<b>ФУНКЦИОНАЛЬНАЯ АННОТАЦИЯ АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ЛОКАЛЬНОГО СХОДСТВА</b> <i>К.Е. Александров, Б.Н. Соболев, Д.А. Филимонов, В.В. Поройков</i> .....	114
<b>АНАЛИЗ РАСПРЕДЕЛЕНИЯ АДЕНОЗИНФОСФАТ-СВЯЗЫВАЮЩИХ САЙТОВ БЕЛКОВ НА ЭКЗОННОЙ СТРУКТУРЕ ГЕНА</b> <i>И.В. Медведева, П.С. Деменков, В.А. Иванисенко</i> .....	122
<b>КОЭВОЛЮЦИЯ ДОМЕНОВ КЛЮЧЕВЫХ БЕЛКОВ АПОПТОЗА P53 И MDM2</b> <i>С.С. Пинтус</i> .....	128
<b>КОМПЬЮТЕРНОЕ ПРЕДСКАЗАНИЕ БИОЛОГИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ ВЕЩЕСТВ: ВИРТУАЛЬНАЯ ХЕМОГЕНОМИКА</b> <i>В.В. Поройков, Д.А. Филимонов, Т.А. Глориозова, А.А. Лагунин, Д.С. Дружиловский, А.В. Степанчикова</i> .....	137
<b>THE WITHIN-INDIVIDUAL BASIS OF BETWEEN-INDIVIDUAL DIFFERENCES</b> <i>V.G. Cherdantsev, V.A. Scobeyeva</i> .....	144
<b>МОДЕЛИРОВАНИЕ ЭВОЛЮЦИИ ТРОФИЧЕСКИ ЗАМКНУТЫХ СООБЩЕСТВ С КОМПЕНСАТОРНЫМ И НЕКОМПЕНСАТОРНЫМ МЕТАБОЛИЗМОМ</b> <i>С.А. Лашин, В.В. Суслов, Ю.Г. Матушкин</i> .....	150
<b>TOPOLOGICAL INDEX OF THE P53-MDM2 CIRCUIT</b> <i>V.P. Golubyatnikov, E. Mjolsness, Yu.A. Gaidov</i> .....	160
<b>СИСТЕМА АВТОМАТИЗИРОВАННОЙ ГЕНЕРАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ГЕННЫХ СЕТЕЙ</b> <i>Ф.В. Казанцев, И.Р. Акбердин, К.Д. Безматерных, В.А. Лихошвай</i> .....	163
<b>МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МЕТАБОЛИЗМА АУКСИНА В КЛЕТКЕ МЕ-РИСТЕМЫ ПОБЕГА РАСТЕНИЯ</b> <i>И.Р. Акбердин, Ф.В. Казанцев, Н.А. Омелянчук, В.А. Лихошвай</i> .....	170
<b>МОДЕЛИРОВАНИЕ РЕГУЛЯЦИИ АУКСИНОМ ИНИЦИАЦИИ ЛАТЕРАЛЬНЫХ ОРГАНОВ У ARABIDOPSIS THALIANA L.</b> <i>В.А. Лихошвай, Н.А. Омелянчук, В.В. Миронова, Ф.В. Казанцев, И.Р. Акбердин, В.К. Королев, С.И. Фадеев, Н.А. Колчанов</i> .....	176

<b>ГЕНЕТИЧЕСКИЙ КОНТРОЛЬ РАЗВИТИЯ МЕХАНОРЕЦЕПТОРОВ У <i>DROSOPHILA MELANOGASTER</i> – ОПИСАНИЕ В БАЗЕ ДАННЫХ «NEUROGENESIS»</b> <i>Т. Бухарина, Д.П. Фурман</i> .....	186
<b>МОДЕЛИРОВАНИЕ МЕХАНИЗМОВ РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНА <i>EVEN-SKIPPED</i> У ДРОЗОФИЛЫ</b> <i>А.Д. Матвеева, Дж. Ионидес, М.Г. Самсонова, Дж. Райнитиц</i> .....	194
<b>МОДЕЛИРОВАНИЕ РЕГУЛЯТОРНЫХ СЕТЕЙ ДЛЯ ВЫЯВЛЕНИЯ ПРОТИВООПУХОЛЕВЫХ МИШЕНЕЙ НА ПРИМЕРЕ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ</b> <i>О.Н. Коборова, Д.А. Филимонов, А.В. Захаров, А.А. Лагунин, А. Кель, Ф. Колтаков, Ю.В. Кондрахин, Р. Шарипов, В.В. Поройков</i> .....	201
<b>МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДЕЙСТВИЯ ПОТЕНЦИАЛЬНЫХ ПРОТИВОВИРУСНЫХ ПРЕПАРАТОВ НА РЕПЛИКАЦИЮ СУБГЕНОМНОГО РЕПЛИКОНА ВИРУСА ГЕПАТИТА С В КЛЕТКЕ</b> <i>Е.Л. Мищенко, К.Д. Безматерных, В.А. Иванисенко, В.А. Лихошвай, Н.А. Колчанов</i> .....	208
<b>ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО АНАЛИЗА МИКРОИЗОБРАЖЕНИЙ ЛИСТА ДЛЯ ОЦЕНКИ ХАРАКТЕРИСТИК ОПУШЕНИЯ ПШЕНИЦЫ <i>TRITICUM AESTIVUM L.</i></b> <i>А.В. Дорошков, С.И. Арсенина, Т.А. Пшеничникова, Д.А. Афонников</i> .....	218
<b>ПАМЯТИ ВАСИЛИЯ ЕФИМОВИЧА ПЕРФИЛЬЕВА (1931–2008 гг.)</b> .....	227

## Content

<b>ON THE 110th YEAR ANNIVERSARY OF GEORGYI DMITRIEVICH KARPECHENKO (03.05.1899–28.07.1941)</b>	
<i>V.A. Vishnyakova, N.P. Goncharov</i> .....	7
<b>MORPHOGENESIS AND MORPHOMECHANICS</b>	
<i>L.V. Belousov</i> .....	29
<b>PREDICTION OF NEW SREBP BINDING SITES IN THE PROMOTER REGIONS OF VERTEBRATE GENES USING A COMBINATION OF A BIOINFORMATIC AND AN EXPERIMENTAL APPROACH</b>	
<i>E.V. Ignatieva, T.I. Merkulova, D.Y. Oshchepkov, N.V. Klimova, G.V. Vasiliev, I.I. Turnaev, V.F. Kobzev, N.A. Kolchanov</i> .....	37
<b>DETECTION OF NEW DRE SITES IN REGULATORY REGION OF HUMAN GENES EN- CODING COMPONENTS OF AH RECEPTOR CYTOSOLIC COMPLEX</b>	
<i>D.Y. Oshchepkov, D.P. Furman, E.A. Oshchepkova, A.V. Katokhin, M.Y. Shamanina, V.A. Mordvinov</i> .....	46
<b>CYTOKINES: BIOLOGICAL PROPERTIES AND REGULATION OF HUMAN INTERLEU- KIN-5 GENE EXPRESSION</b>	
<i>V.A. Mordvinov, D.P. Furman</i> .....	53
<b>DNA – «PROGRAMMING LANGUAGE OF LIFE»</b>	
<i>R. Hofestädt</i> .....	68
<b>DOMAIN – WIDE LANDSCAPE OF HUMAN GENOME</b>	
<i>D.A. Maximov, V.N. Babenko</i> .....	73
<b>HORIZONTAL TRANSFER OF NON-LTR RETROTRANPOSONS</b>	
<i>O.S. Novikova, V. Fet, A.G. Blinov</i> .....	76
<b>CORRELATION BETWEEN THE OPERON STRUCTURE AND THE GENOME LENGTH IN 14 MYCOPLASM SPECIES</b>	
<i>S.A. Lashin, Yu.G. Matushkin, T.M. Khlebodarova, V.A. Likhoshvai</i> .....	84
<b>INTRON-EXON PATTERNS AS A POTENTIAL TOOL IN STUDYING GENE EVOLUTION</b>	
<i>A. Ruvinsky</i> .....	91

<b>GENETIC LINKAGE ANALYSIS CHALLENGES ON A DISTRIBUTED GRID ENVIRONMENT</b> <i>A. Calabria, D. Di Pasquale, A. Orro, G. Trombetti, M. Gnocchi, L. Milanesi</i> .....	96
<b>APPLICATION OF NONMETRIC MULTIDIMENSIONAL SCALING FOR ANALYSIS OF CROSS-PLATFORM GENE EXPRESSION MICROARRAY DATA</b> <i>V.M. Efimov, A.V. Katokhin</i> .....	102
<b>A COMBINATORICS-BASED DATA-MINING APPROACH TO TIME-SERIES MICROARRAY ALIGNMENT</b> <i>N. Turenne, I. Hue</i> .....	109
<b>FUNCTIONAL ANNOTATION OF THE AMINO ACID SEQUENCES USING LOCAL SIMILARITY</b> <i>K. Alexandrov, B. Sobolev, D. Filimonov, V. Poroikov</i> .....	114
<b>DISTRIBUTION ANALYSIS OF ADENOSINE PHOSPHATE BINDING SITES OF PROTEINS ON EXON STRUCTURE OF GENE</b> <i>I.V. Medvedeva, P.S. Demenkov, V.A. Ivanisenko</i> .....	122
<b>COEVOLUTION OF PROTEIN DOMAINS OF P53 AND MDM2 – KEY PROTEINS OF APOPTOSIS</b> <i>S.S. Pintus</i> .....	128
<b>COMPUTER-AIDED PREDICTION OF BIOLOGICAL ACTIVITY SPECTRA FOR SUBSTANCES: VIRTUAL CHEMOGENOMICS</b> <i>V.V. Poroikov, D.A. Filimonov, T.A. Glorizova, A.A. Lagunin, D.S. Druzhilovsky, A.V. Stepanchikova</i> .....	137
<b>THE WITHIN-INDIVIDUAL BASIS OF BETWEEN-INDIVIDUAL DIFFERENCES</b> <i>V.G. Cherdantsev, V.A. Scobeyeva</i> .....	144
<b>EVOLUTION MODELING OF TROPHIC WEB-LIKE COMMUNITIES WITH COMPENSATORY AND NON-COMPENSATORY METABOLISM</b> <i>S.A. Lashin, V.V. Suslov, Yu.G. Matushkin</i> .....	150
<b>TOPOLOGICAL INDEX OF THE P53-MDM2 CIRCUIT</b> <i>V.P. Golubyatnikov, E. Mjolsness, Yu.A. Gaidov</i> .....	160
<b>THE TOOL FOR AUTOMATIC GENERATION OF GENE NETWORK'S MATHEMATICAL MODELS</b> <i>F.V. Kazantsev, I.R. Akberdin, K.D. Bezmaternykh, V.A. Likhoshvai</i> .....	163
<b>MATHEMATICAL MODEL OF AUXIN METABOLISM IN MERISTEM CELLS OF PLANT SHOOTS</b> <i>I.R. Akberdin, F.V. Kazantsev, N.A. Omelyanchuk, V.A. Likhoshvai</i> .....	170
<b>MODELING OF AUXIN REGULATED LATERAL ORGAN INITIATION IN ARABIDOPSIS THALIANA L.</b> <i>V.A. Likhoshvai, N.A. Omelyanchuk, V.V. Mironova, F.V. Kazantsev, I.R. Akberdin, V.K. Korolev, S.I. Fadeev, N.A. Kolchanov</i> .....	176
<b>GENETICAL CONTROL OF MECHANORECEPTORS FORMATION IN D. MELANOGASTER – THE DISCRPTION IN DATABASE «NEUROGENESIS»</b> <i>T.A. Bukharina, D.P. Furman</i> .....	186

<b>MODELING THE EXPRESSION OF THE DROSOPHILA <i>EVEN-SKIPPED</i> (EVE) GENE</b> <i>A.D. Matveeva, J.M.C. Ionides, J. Reinitz, M.G. Samsonova</i> .....	194
<b>MODELLING OF REGULATORY NETWORKS IN IDENTIFICATION OF PROMISING DRUG TARGETS FOR BREAST CANCER THERAPY</b> <i>O.N. Koborova, D.A. Filimonov, A.V. Zakharov, A.A. Lagunin, A. Kel, F. Kolpakov, R. Sharipov, Y. Kondrachin, V.V. Poroikov</i> .....	201
<b>MATHEMATICAL MODELING OF EFFECTS OF POTENTIAL ANTIVIRALS ON SUBGENOMIC HEPATITIS C VIRUS REPLICON REPLICATION IN A CELL</b> <i>E.L. Mishchenko, K.D. Bezmaternykh, V.A. Ivanisenko, V.A. Likhoshvai, N.A. Kolchanov</i> .....	208
<b>THE USE OF COMPUTER-BASED IMAGE PROCESSING TO LEAF HAIRINESS ANALYSIS IN WHEAT <i>TRITICUM AESTIVUM</i> L.</b> <i>A.V. Doroshkov, S.I. Arsenina, T.A. Pshenichnikova, D.A. Afonnikov</i> .....	218
<b>COMMEMORATING VASILYI EFIMOVICH PERFILYEV (1931–2008)</b> .....	227

## ГЕОРГИЙ ДМИТРИЕВИЧ КАРПЕЧЕНКО К 110-ЛЕТИЮ СО ДНЯ РОЖДЕНИЯ (03.05.1899–28.07.1941)

М.А. Вишнякова<sup>1</sup>, Н.П. Гончаров<sup>2</sup>

<sup>1</sup> ГНУ ГНЦ РФ Всероссийский НИИ растениеводства им. Н.И. Вавилова, С-Петербург, Россия, e-mail: m.vishnyakova@vir.nw.ru; <sup>2</sup> Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: gonch@bionet.nsc.ru

Мировая известность пришла к Г.Д. Карпеченко в 28 лет. За неполных 20 лет научной деятельности он создал новую отрасль биологии, став, по сути, первым генным (хромосомным) инженером растений. Во все языки мира вошло слово *Raphanobrassica*, а все учебники по генетике и цитогенетике описывают его классические эксперименты по преодолению бесплодия у рукотворных отдаленных гибридов. Несмотря на всемирную известность Георгия Дмитриевича, о нем написано не слишком много. До сих пор не обработаны и не опубликованы документы, хранящиеся у его родных и в ряде архивов. Данная статья – скромная дань памяти ученого в канун его 110-летнего юбилея.

**Ключевые слова:** Г.Д. Карпеченко, история генетики, отдаленная гибридизация, амфиплоидия, *Raphanobrassica*.

Георгий Дмитриевич Карпеченко был одним из самых близких сподвижников Н.И. Вавилова, его любимцем (любимчиков у Вавилова не было). Работать по генетике растений в те годы в Советской России и не пересечься с Н.И. Вавиловым было невозможно, поскольку его Институт прикладной ботаники и новых культур (с 1930 г. Всесоюзный институт растениеводства) был средоточием всего нового в растениеводческой науке. Вот и Г.Д. Карпеченко, получив образование в Москве, оказался в Ленинграде во главе генетических исследований сначала вавиловского Института, а затем и Ленинградского университета.

Г.Д. Карпеченко родился 21 апреля (3 мая по новому стилю) 1899 г. в г. Вельске Вологодской губернии (ныне Архангельская обл.) в семье уездного землемера. Окончил вологодскую гимназию. В 1917 г. поступил на естественное отделение физико-математического факультета Пермского университета, организованного в 1916 г. на базе частично эвакуированного из прифронтовой полосы Петроградского университета. Через год перевелся на сельскохозяйственное отделение Петровской сельскохозяйственной академии (бывший Московский СХИ, с 1923 г. – Сельскохозяйственная академия им. К.А. Тими-

ряева). «Голод и отсутствие средств заставили [в] 1919 и 1920 гг. учиться в Академии только урывками; большую часть времени я работал то в Вельске – техником сельского хозяйства, преподавателем естествознания в сельской школе, то в Вологде – санинструктором по коммунальным хозяйствам, лектором ботаники Пролетарского Университета, прерывая работу раза три в год для поездок в Академию – для сдачи экзаменов и проведения практических занятий. С 1920 г., после декрета о возвращении студентов в ВУЗы, поселился в Академии безвыездно и нормально окончил ее в 1922 г.», – напишет он позднее в своей автобиографии<sup>1</sup>.

Учителями Г.Д. Карпеченко в «Петровке», определившими его научные пристрастия, стали профессор Сергей Иванович Жегалов (1881–1927), автор первого отечественного учебника для вузов «Введение в селекцию сельскохозяйственных растений» (Жегалов, 1924), в 1913–1922 гг. работавший также помощником заведующего, а с 1922 г. заведующим Селекционной станцией в Петровско-Разумовском (Полумордвинова, 2007), и Александра Гавриловна Николаева (1884–1925), занимавшаяся на этой

<sup>1</sup> Автобиография Г.Д. Карпеченко // Арх. ВНИИР им. Н.И. Вавилова. Оп.2-1. Д.77. Л. 44.





Г.Д. Карпеченко.

же станции цитологией культурных растений. Она одна из первых в мире установила наличие у видов пшениц различий по числу хромосом (Николаева, 1922/1923) и исследовала цитологию ржано-пшеничных гибридов (Николаева, 1924). А.Г. Николаева в свою очередь была ученицей С.Г. Навашина (Рубцова, 1983), школа которого интенсивно разрабатывала вопросы связи морфологии хромосом с эволюционными процессами. С.Г. Навашин ввел понятие «идиограммы» (Навашин, 1921), его ученик Л.Н. Делоне (1922) – «кариотипа». Позже оба термина стали применяться в формулировках, уточненных еще одним учеником С.Г. Навашина Г.А. Левитским (1924, 1931).

Первой научной работой Г.Д. Карпеченко было кариологическое исследование ряда видов крестоцветных, фасоли и клевера как «для выяснения генезиса их, так и для познания связи между изменениями фенотипа и кариотипа»<sup>2</sup>. У бобовых им было описано два типа эволюционного процесса: первый, связанный с преобразованием хромосомного комплекса в целом, был выявлен в роде *Trifolium* L., у видов которого было установлено наличие полиплоидных рядов с числами хромосом от 14 до 130; второй тип был обусловлен мутациями генов и

<sup>2</sup> Письмо Г.Д. Карпеченко Н.И.Вавилову от 2 января 1925 г. (здесь и далее письма цитируются по URL: <http://www.vir.nw.ru/history/v-k-letters.htm> (дата обращения: 04.12.2008)).

структурными изменениями отдельных хромосом (у всех изученных видов фасоли  $2n = 22$ ) (Карпеченко, 1924/1925). Позже работы будут продолжены его сотрудниками: А.Н. Лутков по результатам межвидовой гибридизации у двух видов гороха пришел к выводу, что «дифференциация видов в процессе эволюции шла без изменений в хромосомном числе, путем структурных изменений в самих хромосомах...» (Лутков, 1930. С. 361).

Цитологический метод, которым он овладел под руководством первой русской женщины-цитолога растений, позднее названной им «своим единственным, истинным другом»<sup>3</sup>, оказался необходимой методической базой для его дальнейших исследований по отдаленной гибридизации растений, принесших ему мировую славу. На опытных полях Селекционной станции в Петровско-Разумовском он получил от профессора С.И. Жегалова для скрещивания два вида растений из разных родов: редьку *Raphanus sativus* L. и капусту *Brassica oleracea* L. и установил, что эти виды имеют одинаковое число хромосом ( $2n = 18$ ), чрезвычайно сходных по форме (Карпеченко, 1924). Ранее А.Г. Николаева (1920), исследуя род *Avena* L., показала, что овсы распадаются на две группы, одна из которых характеризуется малыми числами хромосом ( $2n = 14-16$ ), другая – большими ( $2n = 44-48$ ). При этом различие в числе хромосом является препятствием к получению плодовитого потомства у межгрупповых гибридов.

Работа на Селекционной станции сочетала в себе возможности для проведения и полевых опытов, и фундаментальных исследований. Здесь под руководством ее основателя и первого заведующего Д.Л. Рудзинского проводились биохимические, цитологические, физиологические, фитопатологические и другие эксперименты (Елина, 2007). Поэтому, делая скрещивания и получая свои первые гибриды, Г.Д. Карпеченко (1927б) здесь же, на станции, подвергал их тщательному цитологическому анализу. Кроме А.Г. Николаевой, «всемерное содействие» проведению его работ оказывала и цитолог И.Н. Свешникова, будущий автор уникальной монографии по цитогенетике рода *Vicia* L. (Свешникова, 1979).

<sup>3</sup> Письмо Г.Д. Карпеченко Н.И.Вавилову от 7 ноября 1925 г.



После окончания Петровской сельскохозяйственной академии в 1922 г. Г.Д. Карпеченко был оставлен «для подготовки к научной деятельности» при кафедре генетики и селекции (заведующий кафедрой С.И. Жегалов). «Оставленным состоял в течение трех лет, неся в то же время в Академии обязанности преподавателя ботаники, а одно полугодие – и частного земледелия»<sup>4</sup>. Преподавал он также анатомию и систематику растений<sup>5</sup>. С.И. Жегалов писал Д.Л. Рудзинскому, что «...лабораторию удалось значительно пополнить: ... получили два хороших микроскопа..., имеем свой хороший микротом.... Удалось найти двух талантливых учеников для нее [А.Г. Николаевой – Н.Г.]: Г.Д. Карпеченко, оканчивающего срок оставления при кафедре, и И.Н. Свешникову, которую никак не могу устроить в Академию на платное место» (цит. по: Бечус, Марков, 1966. С. 92).

Вероятно, что Н.И. Вавилов и Г.Д. Карпеченко познакомились именно в «Петровке», где-то в 1920–1921 гг. К тому времени, когда Г.Д. Карпеченко стал студентом, Вавилов уже уехал из Москвы работать и преподавать в Саратов. Однако в Петровско-Разумовском он оставил много своих опытов и просил кого-нибудь присматривать за ними. В частности Софья Павловна Зыбина вела иммунологические наблюдения над разными культурами и выполняла фитопатологические учеты главнейших болезней на пшенице, овсе и ячмене (Зыбина, 2007). Наведываясь в Москву из Саратова, а потом и из Петрограда, где в 1920 г. он стал заведующим Отделом прикладной ботаники и селекции Сельскохозяйственного ученого комитета (ОПБиС СХУК), Н.И. Вавилов всегда приезжал на Селекционную станцию, чтобы проверить свои опыты, и находил время для посещения других станций, лабораторий и кафедр Академии. Он не мог, конечно же, не обратить внимание на опыты пытливого студента, затеявшего скрещивания растений из разных родов. По воспоминаниям С.П. Зыбиной (2007), в обеденные часы случалось чаевничать в лаборатории физиологии у Н.Е. Прокопенко в компании Н.И. Вавилова, А.Г. Николаевой,

Г.Д. Карпеченко и иногда присоединявшегося к ним С.И. Жегалова.

В 1923 г. между Н.И. Вавиловым и Г.Д. Карпеченко завязалась переписка. В этом году в «Трудах по прикладной ботанике и селекции», издание которых было возобновлено ОПБиС СХУК под редакцией Н.И. Вавилова, появилась первая статья студента Г. Карпеченко (1922/1923), и Николай Иванович не только выслал ему несколько томов «Трудов...», но и сообщил о готовности принять к опубликованию и его следующую работу: «...статью о гибридах охотно напечатаем в размерах печатного листа (16 страниц) и даже несколько больше, но сократите, елико возможно, число рисунков»<sup>6</sup>.

В следующем, 1924, году у Георгия Дмитриевича выходят четыре статьи и одна из них в «Journal of Genetics» (Karpechenko, 1924). Это был, безусловно, успех молодого ученого. Он отсылает оттиски Н.И. Вавилову и при этом просит о срочной поддержке его заграничной стажировки, довольно подробно ее обосновывая: «... мне очень бы хотелось повысить свои шансы рекомендацией или отзывом от Вас»<sup>7</sup>. Николай Иванович, запаздывая с рекомендацией, тем не менее, в следующий раз поддерживает стремление уже своего сотрудника «людей посмотреть»<sup>8</sup> и пишет ему подробное напутствие: «... За границей стоит побывать и повидать Винклера, Корренса, Нильсона-Элле, а особенно Герберта Нильсона. Баур плохо доступен, занят больше, как и мы грешные, администрированием, кроме того, участвует в коммерческих предприятиях. Книгу его я прочитал, меня она удовлетворила очень мало, так же как и других. Оказывается, Йогансен пишет новое издание “Элементы”. Непременно поведите Винге; вообще скандинавы народ любопытный: работают спокойно, толково и дельно. В Англии затишье. Любопытны голландцы, а в целом Лотси и Де-Фриз, но их мы знаем и по книжкам. Как я уже говорил Вам, за границей любопытно побывать, подучить языки, собрать материал, повидать больших людей, вдохновиться, но учиться особенно не-

<sup>4</sup> Автобиография Г.Д. Карпеченко // Арх. ВНИИР им. Н.И. Вавилова. Оп.2-1. Д.77. Л. 38.

<sup>5</sup> Личный листок по учету кадров Г.Д. Карпеченко // Арх. ВНИИР им. Н.И. Вавилова. Оп.2-1. Д.77. Л. 22.

<sup>6</sup> Письмо Н.И. Вавилова Г.Д. Карпеченко от 16 ноября 1923 г.

<sup>7</sup> Письмо Г.Д. Карпеченко Н.И. Вавилову от 2 января 1925 г.

<sup>8</sup> Там же.

чему. Гораздо важнее почитать побольше книг и овладеть языками, так, чтобы не лазить в словарь. Говорю откровенно, как думаю. Но при всем нашем убожестве мы еще как-то держимся на одном уровне...»<sup>9</sup>.

Срок приписки Георгия Дмитриевича к кафедре генетики и селекции СХА имени К.А. Тимирязева<sup>10</sup> заканчивался в апреле 1925 г. Узнав об этом, Николай Иванович приглашает его незамедлительно приехать работать в Ленинград, в его Всесоюзный институт прикладной ботаники и новых культур (ИПБиНК), организованный в 1924 г. на базе части ОПБиС СХУК. И, уже считая его сотрудником института, хлопочет о его зарубежной поездке: «При сем посылаю Вам бумагу для Главнауки, согласованную с Викт. Евгр. [Писаревым – М.В.]. Срок командировки прописал, как видите, полугодовой. Думаю, что и по существу было бы целесообразно пробыть не менее полугодом за границей, чтобы подучить языки и хотя бы одним языком свободно владеть. Когда поедете, поставим Вам миллион заданий, об этом напишет Вам и В. Евгр. Думаю, что возможно поддержать Вас и в финансовом отношении. Уже одно вознаграждение по должности даст Вам возможность наполовину обеспечить командировку»<sup>11</sup>.

Решение Вавилова о приеме на работу Г.Д. Карпеченко было достаточно спонтанным. Об этом свидетельствуют разночтения в архивных документах, касающиеся названия должности, даты вступления в должность и подразделения, которого на момент принятия на работу Г.Д. Карпеченко, по сути, еще не существовало. В личном деле начало его деятельности в ИПБиНК обозначено по-разному: «заведующий Генетической лабораторией ИПБ»<sup>12</sup>, «заведующий Генетическим Отделением Центральной Генетической и Селекционной Опытной Станцией в Детском Селе»<sup>13</sup>, «ученый специалист, научный руководитель по генетике

отдела селекции и генетики»<sup>14</sup>, «ученый специалист, заведующий секцией (позже отделом) общей генетики»<sup>15</sup>. Однако осознавая острую необходимость иметь в своем Институте хорошего специалиста, который мог бы возглавить генетические исследования, Н.И. Вавилов приглашает многообещающего ученого, которому в то время было только 25 лет. Экспериментальная деятельность Г.Д. Карпеченко далеко выходила за рамки практических задач, поставленных перед ним на Селекционной станции ТСХА, и он, вероятно, без больших колебаний принял приглашение Н.И. Вавилова.

Назначение Г.Д. Карпеченко на новую должность – ученого специалиста, заведующего секцией общей генетики осуществилось 1 апреля 1925 г, а чуть ранее, 28 марта, он был откомандирован на Селекционную станцию в Петровско-Разумовское – «закончить свои генетические работы»<sup>16</sup>. В августе 1925 г. Г.Д. Карпеченко выехал за рубеж, успев набрать в штат лаборатории трех человек (А.Н. Луткова, О.Н. Сорокину и С.А. Щавинскую) и обустроить их работу в одном из зданий великокняжеской усадьбы в Детском Селе, в которой Н.И. Вавилов незадолго до этого организовал Центральную генетико-селекционную станцию ИПБиНК, вскоре реорганизованную в отдел генетики и селекции Института под руководством замдиректора В.Е. Писарева (Филимонов, 1972).

За время заграничной командировки с августа 1925 по июнь 1926 гг. Георгий Дмитриевич посетил генетические лаборатории 9 европейских стран: Финляндии (Хельсинки), Швеции (Стокгольм, Лунд, Свалёф, Ландскрона), Норвегии (Осло), Дании (Копенгаген), Германии (Гамбург, Киль, Берлин, Мюнхен, Геттинген, Йена), Голландии (Гронинген, Амстердам, Вагенинген), Англии (Лондон, Мертон, Лидинг), Франции (Париж), Австрии (Вена). Он ознакомился с работами крупнейших генетиков этих стран: Э. Бауэра, У. Бэтсона, О. Винге, Г. Винклера, С. Дарлингтона, К. Корренса, Я. Лотси, К. фон Чермака и др. Общение с этими учеными, знакомство с их новейшими исследования-

<sup>9</sup> Письмо Н.И.Вавилова Г.Д. Карпеченко от февраля 1925 г.

<sup>10</sup> 10 декабря 1923 г. СНК РСФСР переименовал Петровскую сельскохозяйственную академию в Сельскохозяйственную академию им. К.А. Тимирязева.

<sup>11</sup> Письмо Н.И.Вавилова Г.Д. Карпеченко от 6 апреля 1925 г.

<sup>12</sup> Автобиография Г.Д. Карпеченко // Арх. ВНИИР им. Н.И. Вавилова. Оп. 2-1. Д. 77. Л. 38.

<sup>13</sup> Командировочное удостоверение // Арх. ВНИИР им. Н.И. Вавилова. Оп. 2-1. Д. 77. Л. 4.

<sup>14</sup> Личный листок по учету кадров // Арх. ВНИИР им. Н.И. Вавилова. Оп. 2-1. Д. 77. Л. 20.

<sup>15</sup> Характеристика на Г.Д. Карпеченко // Арх. ВНИИР им. Н.И. Вавилова. Оп. 2-1. Д. 77. Л. 30.

<sup>16</sup> Командировочное удостоверение // Арх. ВНИИР им. Н.И. Вавилова. Оп. 2-1. Д. 77. Л. 4.

ми и работа в их лабораториях были серьезным испытанием и одновременно полезной школой для Г.Д. Карпеченко: «...Работаю нервно, не могу часто заснуть до 4–5 часов, все думаю, думаю, и все об этих гибридах...»<sup>17</sup>. Между тем, и он сам, и его научный багаж были интересны принимающим его ученым. Работая в лаборатории О. Винге, Г.Д. Карпеченко (1927б) попытался экспериментально на своих редечно-капустных гибридах доказать его гипотезу гибридного происхождения полиплоидных видов (Winge, 1917). Параллельно с ним константные амфиплоиды от скрещивания двух видов табака получили R. Clausen, T. Goodspeed (1925). Однако у последних амфидиплоиды (термин С.Г. Навашина (1927)) были получены посредством межвидовой гибридизации, в то время как у Г.Д. Карпеченко – межродовой. Процессы, обнаруженные в эксперименте, могли осуществляться в природе, и наличие полиплоидных серий видов показывает, что в их образовании «играет роль умножение хромосомного комплекса» (Карпеченко, 1935б. С. 398). Кроме того, этот способ открывал широкие возможности для практической селекции на основе получения плодовых гибридов, сочетающих признаки нескольких видов (Карпеченко, 1930, 1938а; Karpechenko, 1929).

В одном из писем он признается Н.И. Вавилову: «Рекомендуясь здесь заведующим генетической лабораторией, я хронически попадаю в профессора, что ставит меня в очень неловкое положение, из которого я не всегда решаюсь выйти. Получается какое-то безобразие. Я очень бы просил Вас, дорогой Николай Иванович, писать мне (конечно, очень кратко). Меня очень интересуют дела Всесоюзного Института и Ваши личные перспективы»<sup>18</sup>.

Понимая, что лаборатория генетики ИПБиНК должна стать важнейшим звеном в создаваемом Вавиловым ряду методических подразделений института, Г.Д. Карпеченко разрабатывает методологию исследований, а именно: теоретическое и экспериментальное обоснование новых методов создания селекционного материала, его изучения и использования в селекции растений. Главными темами исследования в его

лаборатории становятся отдаленная гибридизация и цитогенетика отдаленных гибридов – продолжение уже ранее начатых им работ, экспериментальный мутагенез, методы получения и использования в селекции полиплоидных форм растений (Ригин, 2007). При этом исследования по отдаленной гибридизации – одно из наиболее результативных и интересных направлений научной работы отдела: «Овладеть отдаленной гибридизацией с целью создания новых форм, совмещающих признаки, рассеянные по разным видам и родам, представляется весьма заманчивым» (Карпеченко, 1934. С. 41). При формировании экспериментальной базы и штата лаборатории проявился незаурядный организаторский талант Георгия Дмитриевича. Он собрал коллектив блестящих генетиков: Т.В. Асеева, Е.И. Барулина, Е.А. Домбровская, А.С. Каспарян, О.Н. Сорокина, В.С. Федоров, М.И. Хаджинов, А.У. Хоменко, С.А. Щавинская, Д.Р. Щербина и др. (Ригин, 1989). Лаборатория, позднее реорганизованная в отдел генетики, стала одним из ключевых экспериментальных подразделений института. В 1934 г. в штате отдела состояло 43 человека, не считая аспирантов и практикантов. А аспирантов уже в 1931 г. было 16 человек. Впоследствии многие из них вошли в штат лаборатории.

Заместителем Г.Д. Карпеченко по лаборатории был А.Н. Лутков. Он работал в лаборатории со дня ее основания – сначала лаборантом, затем младшим и впоследствии старшим научным сотрудником. А.Н. Лутков принимал самое активное участие как в организации самой лаборатории, так и в подготовке кадров, заведя лабораторией во время командировок заведующего, порой достаточно продолжительных. Его научные интересы были связаны с цитогенетикой отдаленных гибридов, экспериментальным мутагенезом и методами получения, изучения, размножения и использования в селекции полиплоидных растений. А.Н. Лутков (1937б) получил первое экспериментальное подтверждение вавиловского закона гомологических рядов в наследственной изменчивости: им была получена безлигульная мутация у ячменя.

С 1931 г. на работу в лабораторию перешла Е.И. Барулина – жена Н.И. Вавилова, специалист по зернобобовым культурам, монограф рода *Lens L.* – Чечевица. К тому времени Елена

<sup>17</sup> Письмо Г.Д. Карпеченко Н.И. Вавилову от 15 декабря 1925 г.

<sup>18</sup> Там же.





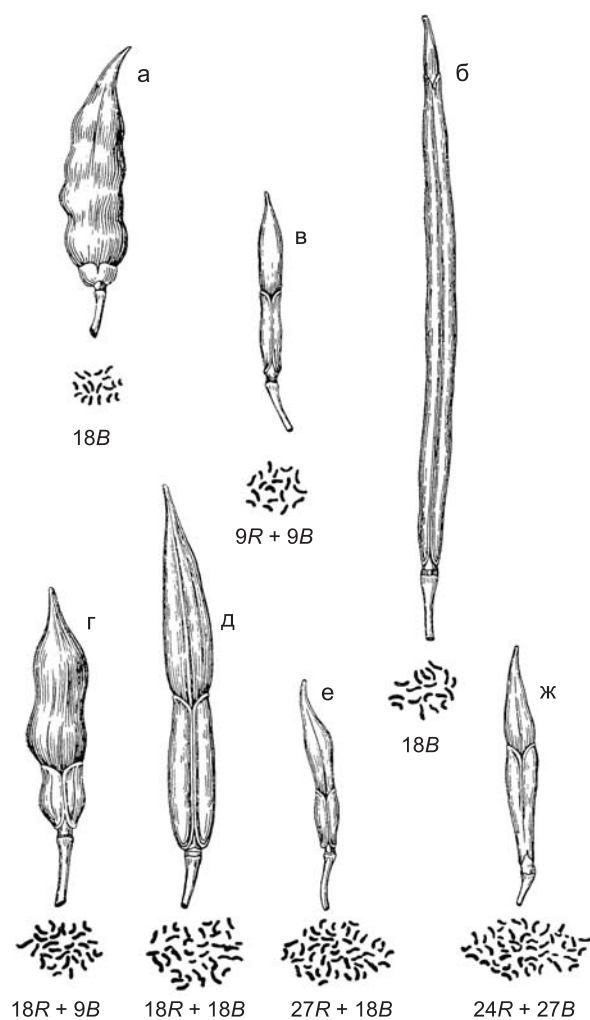
Г.Д. Карпеченко и Е.И. Барулина на экспериментальном участке отдела генетики ВИР в Детском Селе (середина 1930-х гг.).

Ивановна изучила мировое разнообразие чечевицы и опубликовала ряд сводок об этой культуре, в том числе монографию «Чечевица СССР и других стран» (Барулина, 1930). В лаборатории Г.Д. Карпеченко она занялась изучением генетики чечевицы – важной продовольственной культуры, по производству зерна которой в то время СССР был мировым лидером. В отделе генетики проработала она не долго (Вишнякова, 2007) и за это время успела изучить наследование безлигульности у тетра- и гексаплоидных пшениц (Барулина, 1937), заложив основы сравнительно-генетическому изучению признаков у видов одного рода, отличающихся уровнем плоидности.

С самого начала существования лаборатории отдаленной гибридизацией злаков (эгилопсов, пшениц, ржи) занималась О.Н. Сорокина. Идеологом этой работы был Георгий Дмитриевич (Карпеченко, 1927а), и это нашло отражение в их совместных статьях (Карпеченко, Сорокина, 1929). О.Н. Сорокиной в середине 1930-х гг. был осуществлен цикл исследований по гибридизации эгилопсов с пшеницами и эгилопсов с рожью, что позволило пролить свет на многие вопросы филогении этих родов и создать новые, несуществующие в природе виды злаков (Сорокина, 1934, 1937а, б, 1938). В отделе был получен ряд важных экспериментальных данных

о взаимоотношениях хромосом геномов мало изученных форм и видов культурных растений (Макушина, 1938; Светозарова, 1939). Это было продолжением работ А.Г. Николаенко и студенческих исследований Г.Д. Карпеченко по использованию цитологических и генетических методов в систематике полиплоидных родов растений и выяснению их генезиса, а так же для познания связи между изменениями фенотипа и кариотипа. Оценки вклада Г.Д. Карпеченко в решение данной проблемы различны (см. Розанова, 1946; Рубцова, 1973, 1975).

Главным объектом исследований самого Г.Д. Карпеченко оставался редечно-капустный гибрид (*Raphanobrassica*) (рис. 1), полученный в результате скрещивания растений разных видов семейства крестоцветных (*Raphanus sativus* L. × *Brassica oleracea* L.). При этом в качестве отцовской формы использовались образцы из четырех групп разновидностей капусты – капуста белокочанная (*B. oleracea* convar. *capitata* (L.) Alef. var. *capitata* L. f. *alba* DC), капуста савойская (*B. oleracea* convar. *capitata* (L.) Alef. var. *sabauda* L.), капуста брюссельская (*B. oleracea* convar. *oleracea* var. *gemmifera* DC) и капуста кольраби (*B. oleracea* convar. *acephala* (DC.) Alef. var. *gongylodes* L.). Гибриды между редькой и капустой впервые были получены М. Sageret (1826) и позже F. Gravatt (1914). В работе по-



**Рис. 1.** Плоды, пластинки хромосом и формула для *Raphanus* (а) и *Brassica* (б), ди- (в), три- (г), тетра- (д), пентаплоидного (е) и гипогексаплоидного (ж) гибридов (из: Карпеченко, 1927б).

следнего у гибридного растения наблюдался «сильнейший» гетерозис (оно доросло до крыши теплицы, вылезло через вентиляционное отверстие наружу и свесилось по обоим скатам крыши), но растение было стерильно. Как и многие межвидовые гибриды, растения  $F_1$  гибридов, полученных Г.Д. Карпеченко, были также стерильны. Хромосомы редьки и капусты не вступали в конъюгацию в мейозе, в результате не образовывались пыльца и семязачатки, из которых после оплодотворения могли бы развиваться семена гибридного растения (Карпеченко, Щавинская, 1930). И редька, и капуста – двулетние растения. У  $F_1$  редечно-капустных гибридов после хранения в течение зимы коче-

рыг в холодном помещении по мере отрастания стеблей произошло удвоение хромосомного набора. Успеху способствовало и то, что у обеих родительских форм было одинаковое число хромосом ( $2n = 18$ ). У новой рукотворной полиплоидной формы каждая клетка содержала диплоидный набор хромосом редьки и диплоидный набор хромосом капусты. В результате у амфидиплоида ( $2n = 36$ ) восстановилась способность к нормальному прохождению мейоза. В первом поколении было получено 19 фертильных растений, давших 821 жизнеспособное семя (Карпеченко, 1927б). Потомства плодовых растений были высажены на участках в Петровско-Разумовском (где делянки не были изолированы от редьки) и в Грибово (Станция по селекции овощных культур, также возглавляемая С.И. Жегаловым), где соблюдалась полная изоляция гибридных делянок). В первом случае было высажено 209 семян и получено 91 растение  $F_2$  гибридов. Из них 73 были изучены цитологически, при этом 61 растение оказались стерильными триплоидами ( $2n = 27$ ), что является результатом их скрещивания с редькой. В Грибово было высажено 612 семян, давших 361 растение  $F_2$  гибридов, из которых 229 были изучены цитологически – 213 оказались фертильными амфидиплоидами. Другие полиплоидные формы (пента- и гексаплоиды) также были стерильны. Всего было цитологически изучено 302 растения. Дальнейшее изучение растений  $F_2$  гибридов было проведено в 1925 г. уже в Детском Селе (Карпеченко, 1927б). Именно эти впечатляющие тщательностью цитогенетического анализа исследования межвидовых гибридов позволили Г.Д. Карпеченко уже в 1927 г. выступить с докладом о получении редечно-капустных гибридов в Берлине.

Хотя данная работа не оправдала надежд на получение гибрида редьки–капусты, сочетавшего полезные хозяйственные признаки обеих овощей (корнеплод и кочан), Г.Д. Карпеченко была показана принципиальная возможность экспериментального преодоления стерильности у растений, возникающей при отдаленной гибридизации. Тем самым были заложены теоретические и методические основы использования отдаленной гибридизации в селекционной практике и существенно расширены

представления о возможных путях эволюции цветковых растений. Работа Г.Д. Карпеченко (1927б) по редечно-капустным гибридам была опубликована в 1927 г. в «Трудах по прикладной ботанике и селекции». Впоследствии эта работа неоднократно переиздавалась, в том числе и за рубежом (Karpchenko, 1928; Классики..., 1968; Карпеченко, 1971). Позже работы с редечно-капустными гибридами велись довольно интенсивно и другими авторами (см. обзоры Yarnell, 1956; McNaughton, 1973). При этом некоторым из них удалось получить фертильные *Brassico-Raphanus* амфиплоиды (Terasawa, 1932). *Raphanobrassica* была получена заново (McNaughton, 1973; Chen, Wu, 2008). В том числе использовалась для интрогрессии генов из редьки в возделываемые виды рода *Brassica* L. (Metz et al., 1995).

Эксперименты Г.Д. Карпеченко по отдаленной гибридизации были предтечей последующих многочисленных экспериментов по получению новых форм растений, нашедших широкое практическое применение (Шумный, Першина, 1989). Отсюда, вероятно, пошли ржано-пшеничные амфиплоиды – знаменитые АД (сокращ. от амфидиплоид) В.Е. Писарева (1964). Ранее в вировской лаборатории Г.А. Левитского впервые в мире были цитологически изучены спонтанные пшенично-ржаные гибриды с Саратовской опытной станции (Левитский, Бенедская, 1931). Следует отметить удивительное сосуществование в одном учреждении (даже здании) двух подразделений, занимавшихся весьма схожими проблемами (Лебедев, Абрамова, 1994). Это было взаимопроникающее тяготение: лаборатория цитологии Г.А. Левитского (1929) тяготела к генетике, а отдел Г.Д. Карпеченко (1929) – к цитологии. (Н.И. Вавилов ругал последнего, что ему неудобно, что «Карпеченко хлеб у Левитского отнимает»). Н.И. Вавилов (1938) придавал большое значение использованию отдаленной гибридизации в селекции. Успешно получал амфидиплоиды у пшениц и табака работавший в 1930-е гг. в его же Институте генетики АН СССР болгарский генетик Д. Костов (1934, 1936а, б, 1940). По методу Костова в отделе генетики Г.Д. Карпеченко проводилась гибридизация нескольких видов пшеницы для получения так называемых тройных гибридов, сочетающих ценные качества нескольких видов (Александров, 1936).

Большое место в тематике отдела генетики ВИР занимала проблема получения тетраплоидных форм растений. В начале своих исследований сотрудники отдела применяли метод «вегетативного» получения тетраплоидов путем использования тетраплоидных клеток из соматических тканей растений, названный по имени его автора «метод Йоргенсена». У молодых растений обрезалась верхушка и удалялись все пазушные почки, на срезе возникал каллус, а образовавшиеся из него новые побеги подвергались черенкованию. До 10 % растений из таких побегов оказывались тетраплоидными (Jørgensen, 1928). С.А. Щавинской таким способом были получены тетраплоидная и октаплоидная формы капусты (Щавинская, 1937б, в), а также тетраплоидные томаты и была восстановлена плодовитость у стерильной герани (Щавинская, 1937а). Г.Д. Карпеченко (1937б) использовал данный метод для получения гибрида капусты *B. oleracea* с абиссинской горчицей *B. carinata* Al. Braun. У растений, не дающих на месте среза каллус, хромосомы удваивали методом Рандольфа посредством воздействия высоких температур в момент оплодотворения (Randolf, 1932). Таким способом Г.Д. Карпеченко (1938б) получил тетраплоидные формы ячменя. Позднее эти и другие способы «мягкого» получения полиплоидов уступили место методу колхицинирования (Blakeslee, Avey, 1937). При этом Г.Д. Карпеченко (1940а, б) впервые экспериментально было показано возникновение под влиянием колхицина не только клеток с удвоенным числом хромосом, но и ряда хромосомных aberrаций и разрывов хромосом по центромерам. Это было крайне важное для дальнейшего развития таких исследований наблюдение.

В 1927 г. Г.Д. Карпеченко вместе с Н.И. Вавиловым, С.С. Четвериковым и А.С. Серебровским участвовал в V Международном генетическом конгрессе в Берлине, где выступил с докладом (Karpetschenko, 1929). Съезд был своего рода триумфом советской генетики. Делегация генетиков из СССР была одной из самых многочисленных и состояла из 64 человек. И на следующем, VI Генетическом конгрессе в Итаке (США) в 1932 г. советские генетики могли бы быть одной из самых значимых групп и должны были сделать 5 пленарных докладов из запла-





С.С. Четвериков, А.С. Серебровский, Г.Д. Карпеченко, Н.И. Вавилов (слева направо) на V Международном генетическом конгрессе в Берлине (1927 г.).

нированных 25. Георгию Дмитриевичу Оргкомитетом было предложено сделать пленарный доклад об отдаленной гибридизации у растений. К сожалению, на этом форуме генетиков из СССР представлял лишь один Н.И. Вавилов – вице-президент конгресса. Остальные докладчики и участники «не смогли присутствовать...». В 1934 г. Г.Д. Карпеченко не разрешили выехать на VI Международный ботанический конгресс в Амстердам, где он был избран вице-президентом генетической секции. Но все это будет еще впереди. Пока же работа ученого проходила поистине счастливо. Э. Линдстром представил его к Рокфеллеровской стипендии. Отзывы дали Э. Баур, Э. Блексли, Э. Бебкок, Т.С. Морган и Н.И. Вавилов. Последний писал: «В прекрасной, тщательно продуманной работе Г.Д. Карпеченко удалось получить совершенно новый промежуточный вид или, вернее, род... Обнаруженные исследователем факты открывают широкие возможности в межвидовой гибридизации растений, и ныне пропасть, которая до недавнего времени отделяла виды и роды, становится проходимой... Мы считаем своим долгом отметить выдающееся значение этой работы нашего соотечественника» (цит. по: Шварц, 2008).

10–16 января 1929 г. в Ленинграде проходил Всесоюзный съезд по генетике, селекции, семеноводству и племенному животноводству.

В его работе участвовало более 1,5 тысяч отечественных и зарубежных ученых. Председательствовал на съезде Н.И. Вавилов, а Георгий Дмитриевич был генеральным секретарем съезда. Он проделал огромную организационную работу по подготовке съезда, сбору и редактированию его трудов и, будучи одной из ключевых фигур форума, сделал на секции «Цитология наследственности» два доклада: «О половом обособлении редечно-капустных тетраплоидов» (совместно с С.А. Щавинской) и «К синтезу константного гибрида из трех видов (гибриды редечно-капустных тетраплоидов с дикой редькой, репой, горчицей и рапсом)» (Всесоюзный съезд..., 1929). Обе работы были изданы в трудах Съезда (Карпеченко, 1930; Карпеченко, Щавинская, 1930).

Вскоре по приглашению Немецкого общества селекционеров Г.Д. Карпеченко выступил на Немецком селекционном съезде в Кенигсберге.

В этом же году наряду с профессорами Н.А. Максимовым (ИПБиНК) и А.Г. Дояренко (ТСХА) Н.И. Вавилов выдвигает Г.Д. Карпеченко кандидатом на премию им. В.И. Ленина. «Его работа “К проблеме экспериментального видообразования, полиплоидные гибриды редьки и капусты” представляет исключительный интерес в области генетики за последние годы», – пишет он в письме от 23 апреля 1929 г. академику О.Ю. Шмидту, занимавшему в то





Ф.Г. Добржанский, Н.П. Добржанская и Г.Д. Карпеченко в Пасадене, Калифорния, США, октябрь 1930 г. (Фото Н.И. Вавилова).

время ряд крупных постов: заведующего секцией естественных и точных наук Комакадемии<sup>19</sup>, члена Коллегии Наркомпроса и Президиума Государственного ученого совета (Матвеева, 1993).

1929 г. был поистине звездным для советской генетики, для Н.И. Вавилова и Г.Д. Карпеченко. Однако он оказался и последним спокойным для них годом в родном отечестве. В октябре этого же года Г.Д. Карпеченко уезжает на стажировку в США за счет средств присужденной ему Рокфеллеровской стипендии. Более года – по февраль 1931 г. – он работал вначале в Калифорнийском университете в лаборатории ботаника Э. Бабкока (E. Babcock) в Беркли, а затем во всемирно известной лаборатории Т.Х. Моргана в Калифорнийском технологическом институте в Пасадене, успев за это время посетить все самые известные генетические учреждения США.

В Калифорнии в Пасадене он много общался с русским генетиком Феодосием Григорьевичем Добржанским (Theodosius Dobzhansky), учеником заведующего кафедрой генетики Ленинградского университета Ю.А. Филипченко, находящимся на стажировке в лаборатории

Т.С. Моргана так же по стипендии Рокфеллеровского фонда (Бабков, 2007). Ко времени приезда Г.Д. Карпеченко Ф.Г. Добржанский пробыл в этой лаборатории уже два года. В письме от 1 января 1929 г. Ю.А. Филипченко, советуя Ф.Г. Добржанскому не спешить возвращаться и задержаться в США на более длительный срок, исходил из того, что «...3–4 года работы в моргановской лаборатории сделали бы из Вас, Феод. Григ., если не второго Моргана, то по крайней мере одного из самых блестящих морганоидов вроде Стертеванта» (цит. по: Коначев, 1991. С. 241). Соскучившись по родине, продлевая командировку, Ф.Г. Добржанский, намереваясь вернуться «на старое и милое место», в письме от 21 января 1929 г. к Ю.А. Филипченко отвечал на это предложение: «Что касается возможностей для меня остаться здесь на ряд лет, то таковая решительно отсутствует. ...Вы, конечно, совершенно правы в оценке тех преимуществ, какие бы отсюда проистекали для меня, и я это чувствую (именно чувствую, больше, чем что-либо иное). Но есть и другое чувство, называемое любовью к родине, которое развилось у меня особенно остро на чужбине» (У истоков..., 2002. С. 153).

Молодой энергичный Г.Д. Карпеченко появился в Калифорнии именно во время нелегких размышлений Ф.Г. Добржанского о своей дальнейшей судьбе (У истоков..., 2002). Гость

<sup>19</sup> В 1918 г. была организована Социалистическая академия общественных наук, переименованная в 1923 г. в Коммунистическую (Комакадемия), в 1936 г. она включена в АН СССР на правах отделения.

из СССР был полон восторженных впечатлений от успехов ИПБиНК, в котором работал, от Н.И. Вавилова, от своих исследований. Он без устали рассказывал об институте, об институтских делах, о генетических экспериментах на детскосельских делянках и, конечно, о своих рафанобрассиках – новом, донныне не существовавшем виде растений. Более того, Карпеченко имел посеы ячменя и своих *Raphanobrasica* на ближайшей к Пасадене ферме, и Ф.Г. Добржанский время от времени возил его туда на своей машине.

Разговаривали, как и полагается по русскому обычаю, часто до поздней ночи. Говорили не только о генетике, не только о гибридах и дрозофиле. Возвращаться или оставаться? Там, в Пасадене, гуляя в парке среди вековых деревьев, они обсуждали, спорили и никак не могли прийти к согласию, ехать ли домой или искать прибежища на чужбине. Новый, 1930-й, год они встречали вместе. «Мы очень сблизились с ним; его приезд для нас – прямо-таки счастье – все-таки родная душа», – писал Ф.Г. Добржанский на родину (У истоков..., 2002. С. 263).

Они хорошо понимали друг друга, но к согласию все-таки не пришли. Много лет спустя в одном из своих писем в Россию Ф.Г. Добржанский напишет: «Наши пути радикально расходились. Пока Карпеченко был в обстановке субтропической Калифорнии, он считал мой путь неправильным... Он вообще был большим патриотом..., – восхищался Добржанский оптимизмом Карпеченко и его верой в науку. – Но оптимизм этот не был наивным непониманием ужасов того времени. Это был высший оптимизм преодоленного пессимизма» (цит. по: Шварц, 2005).

Успел принять участие в этих дискуссиях и Н.И. Вавилов, кометой мелькнувший в Пасадене по пути в Мексику в октябре 1930 г.: «Я здесь с Карпеченко и Добржанским – это лучшие наши генетики» (цит. по: Шварц, 2005). Перед самим Николаем Ивановичем подобного выбора не было никогда. Он был одержим идеей из своего многострадального государства сделать «обновленную землю», считая, что возможности служить человечеству, существующие в СССР, настолько огромны и вдохновляющи, что во имя их следует научиться не обращать внимания на жестокости существующей реальности.

Он убеждал Ф.Г. Добржанского, что нигде в мире так высоко не ценится труд ученого, как в СССР, и при этом нисколько не сомневался, что Г.Д. Карпеченко вернется на родину.

Ф.Г. Добржанский остался в Соединенных Штатах навсегда (Конашев, 1993; Бабков, 2007). Георгий Дмитриевич вернулся в Советский Союз, написав последнее письмо Добржанскому из Парижа, где останавливался на неделю. В нем сквозила грусть и как-будто предчувствие непоправимого (Шварц, 2005). Однако черные времена для Г.Д. Карпеченко были еще впереди.

После ухода Ю.А. Филипченко с заведования кафедрой генетики и экспериментальной зоологии Ленинградского университета и его безвременной смерти (Медведев, 2006) на биологическом факультете ЛГУ по инициативе Н.И. Вавилова в 1931 г. была открыта самостоятельная кафедра генетики растений и Г.Д. Карпеченко было предложено ее возглавить. Вновь созданную кафедру генетики животных возглавил А.П. Владимирский. Г.Д. Карпеченко был самым молодым профессором ЛГУ, где он до 1941 г. читал общий курс генетики и «специальные главы генетики растений». Как и в отделе вавиловского института, Георгий Дмитриевич создал на кафедре замечательный коллектив единомышленников, пригласив туда в качестве профессоров вировцев Григория Андреевича Левитского, Марию Александровну Розанову и Леонида Ипатьевича Говорова. Цитогенетик Б.И. Васильев был заместителем заведующего и ученым секретарем кафедры. Для чтения лекций на кафедру приглашались известные зарубежные ученые: ученики Т.Х. Моргана К. Бриджес и Дж. Мёллер, которые бывали и работали у Н.И. Вавилова и в Ленинграде, и в Москве, болгарский ученый Дончо Костов, читавший курс частной генетики растений. Наряду с чтением лекций Георгий Дмитриевич руководил аспирантами и возглавлял лабораторию генетики растений Петергофского биологического института ЛГУ. В отдел генетики в ВИР и на кафедру университета к Г.Д. Карпеченко ехали отовсюду, его научные семинары пользовались исключительным успехом. Так отдел генетики ВИР и кафедра генетики и селекции растений ЛГУ стали центром подготовки генетических кадров для всей страны.



Дом сотрудников ВИР в г. Пушкине, Московское шоссе № 23. Веранда и окна в центре на первом этаже – летняя квартира Н.И. Вавилова, крайнее окно справа – квартира Г.Д. Карпеченко (Фото В. Уржумцева).

Как и Н.И. Вавилов, Г.Д. Карпеченко все время курсировал между «Селом и городом»<sup>20</sup>. Его семья имела квартиру в Детском Селе в доме № 23 (ныне дом № 27) по Московскому шоссе, по соседству с летней квартирой Вавилова. Его жена, Галина Сергеевна, молодая, красивая, хорошо образованная женщина всегда готова была помочь мужу. Они были красивой парой современных во всех отношениях людей. Галина Сергеевна была заядлой теннисисткой. Они любили общество, много читали, были в курсе новостей культурной жизни города. В семье росла дочь<sup>21</sup>. Когда готовилась в свет книга Т. Морган «Экспериментальные основы эволюции» (Морган, 1936), Галина Сергеевна помогла мужу перевести ее на русский язык. Впоследствии она многие годы была помощницей Н.П. Дубинина – работала в Институте общей генетики в Москве. В 1991 г. в числе других ученых-генетиков Галина Сергеевна была награждена орденом Трудового Красного Знамени.

2 ноября 1934 г. Президиум Всесоюзной академии сельскохозяйственной наук им. В.И. Ле-

<sup>20</sup> «... раздваиваюсь между Селом и городом...» – цитата из письма Н.И. Вавилова П.П. Подъяпольскому от 18.03.1921.

<sup>21</sup> Валентина Георгиевна Карпеченко (р. 1934) живет и работает в Москве.

нина присудил Георгию Дмитриевичу ученую степень доктора биологических наук по разделу «генетика растений». 29 августа 1935 г. квалификационная комиссия ВАСХНИЛ утвердила его в ученом звании действительного члена научно-исследовательского учреждения. К званию профессора представлен ЛГУ и утвержден в нем 28 декабря 1938 г. (Лебедев, 1980).

Ведущий генетик ВИР, заведующий кафедрой генетики растений ЛГУ, высоко ценимый в мире, прекрасно образованный, любимый преподаватель студентов биофака университета, с середины 1930-х гг., конечно же, стал одной из основных мишеней лысенковцев. Требования ликвидации лаборатории генетики ВИР, стоявшей «на позиции хромосом и генов», нападки на кафедру генетики растений ЛГУ – «оплот реакции» – озвучивались с трибун, печатались в университетской многотиражке и в стенгазетах обоих учреждений.

В 1939 г. Н.И. Вавилов впервые был вынужден публично заговорить о «практической значимости расхождений», ставших очевидными за последние годы. На заседании выездной сессии Ленинградского областного бюро Секции научных работников профсоюза вузов и научно-исследовательских учреждений в ВИР



15 марта он объясняет, что логика развития генетики как науки такова, что она естественно переживает отход от задач практической селекции, в генезисе науки необходим период накопления знания, его оформления. Этот период в развитии генетики был отмечен всюду, во всей мировой науке, но нигде это не ставилось в вину ученым. На этом же заседании Георгий Дмитриевич произносит горькие слова: «Я принадлежу к тем, кого называют всякими страшными словами, я – генетик. Положение наше отчаянное... Чрезвычайно жуткое... Я, знаете, в борьбе закалился. Но страшно тяжело морально... Такое чувство все время, что ты не научный работник, а сектант. Я говорю о том, как нам фактически приходится работать... Мы сейчас слышим, что Мендель – это лженаука. Николай Иванович сказал, что тот, кто знает материал, не может с этим согласиться, и нам остается гореть на костре за эти 3:1»<sup>22</sup>.

Г.Д. Карпеченко участвовал во всех сессиях и заседаниях ВАСХНИЛ (после ее реорганизации в 1934 г.), посвященных вопросам генетики и селекции (Карпеченко, 1937а). В конце 1930-х гг. сотрудники Отдела генетики ВИР очень много времени тратили на проверку «теорий» Т.Д. Лысенко и его школы, на дискуссии (Карпеченко, 1937а). Кроме того, в отделе продолжали интенсивно изучать мутагенез у растений под воздействием X-лучей и химических веществ (Хлоп, 1936; Лутков, 1937а), получили полиплоиды воздействием высокой температуры и с использованием колхицина (Карпеченко, 1938б, 1940б) и т. д. Н.И. Вавилов настаивал, что в план ВИР «Ваше направление надо сформулировать не как бесконечное получение полиплоидов, а преодоление бесплодия гибридов, и реально поставить конкретные задачи. Сейчас все внимание концентрировано на прививочных гибридах. Весь фронт вести не по линии получения полиплоидов, а по линии резкого изменения в сторону преодоления бесплодия гибридов и по линии радикального изменения скрещиваемости, крупных наследственных сдвигов»<sup>23</sup>.

Сам Г.Д. Карпеченко занимался гибридизацией географически отдаленных разновидно-

стей ячменя и выделил ценные безостые формы этой культуры с длинными колосьями, которые должны были вскоре пройти сортоиспытание (Лебедев, 1994). Кроме того, он рассматривает географическое и экологическое обособление как факторы, ведущие к процессам видообразования: «Расселяясь по земному шару, встречаясь с новыми условиями обитания, формы, сходные в начале, могут, очевидно, накапливать различия как вследствие того, что в неодинаковых условиях выживают различные мутации, так и потому, что сам по себе мутационный процесс в различных условиях протекает по-разному, причем и все нарастающее несходство в генотипах также действует на него» (Карпеченко, 1935а. С. 298). Им и его сотрудниками были получены интересные полиплоидные формы ячменя, томата, картофеля, льна (Карпеченко, 1938б). При этом были уже отработаны методы их массового получения (Лутков, 1939).

Для трехтомных «Теоретических основ селекции растений», вышедших под редакцией Н.И. Вавилова в 1935–1937 гг., им были написаны две главы: «Теория отдаленной гибридизации» (Карпеченко, 1935а) и «Экспериментальная полиплоидия и гаплоидия» (Карпеченко, 1935б). В первой из них он подчеркивал относительность деления скрещиваний на внутри- и межвидовые и предложил свою классификацию: конгруэнтные – скрещивания видов, имеющих совместимые геномы, приводящие к получению фертильных и жизнеспособных гибридов, и инконгруэнтные – неприводящие к подобному результату и дающие стерильные гибриды (Карпеченко, 1935а). Хотя такая классификация не прижилась, она представляет определенный интерес для теоретического решения проблемы успешной гибридизации (Зарубайло, 1975). Им было показано, что успех при отдаленной гибридизации сильно зависит и от того, какая форма берется в качестве материнской. Так, на примере рода *Triticum* L., в котором межвидовая гибридизация проводится давно и вопросы скрещиваемости, плодовитости и всхожести зерновок F<sub>1</sub> изучены довольно подробно, было показано, что скрещивание разнохромосомных видов удается легче, если в качестве материнского растения берется вид с меньшим числом хромосом, а всхожесть гибридных зерен выше в случае, когда меньшее

<sup>22</sup> Стенографический отчет ВИР (выездная сессия) от 15 марта 1939 г. на областном бюро Секции научных работников // Личный архив Ю.Н. Вавилова.

<sup>23</sup> Письмо Н.И. Вавилова Г.Д. Карпеченко от 10 октября 1938 г. (цит по: Левина, 1987. С. 2017–2018).

число хромосом имеет отцовский компонент (Карпеченко, 1935а). Дальнейшее развитие проблема несовместимости получила в работе А.Ф. Шулындина (1978), вскрытие причин которой во многих случаях служит основой для разработки методов ее успешного преодоления (Шумный, Першина, 1989).

Георгия Дмитриевича арестовали прямо на рабочем месте – в главном здании Детскосельской усадьбы ВИР 17 февраля 1941 г. Очевидцы видели, как сопровождаемый «людьми в штатском» он шел по антресолям, опоясывающим двухэтажный холл, и спускался по лестнице коттеджа, в котором проработал 16 счастливых лет... На 4-й день после ареста, 21 февраля 1941 г. Г.Д. Карпеченко обратился с письмом к наркому просвещения РСФСР В.П. Потемкину: «В день ареста я начал писать работу о действии отбора в чистой линии после удвоения у растения числа хромосом. Это важнейшее открытие и очень бы хотелось, чтобы оно осталось за советской наукой. Не сможете ли Вы попросить органы Наркомвнудела, ведущие мое дело, разрешить мне написать эту работу хотя бы в виде предварительного сообщения в «Докладах Академии наук», для чего я должен иметь таблицу, сводящую данные по фертильности ячменя с удвоенным числом хромосом за несколько лет. Эта табличка осталась в лаборатории генетики Института растениеводства или в бумагах, взятых агентами Наркомвнудела. Что это не что иное, как таблица по фертильности, могут подтвердить все товарищи по лаборатории» (цит. по: Лебедев, 1994. С. 219). Ему предъявили обвинение в «шпионско-вредительской деятельности», к которой была добавлена открытая борьба под руководством Н.И. Вавилова против «передовых методов научно-исследовательской работы и ценнейших достижений академика Лысенко по получению высоких урожаев» (Лебедев, 1994. С. 223). Приговор Военной коллегии Верховного суда СССР ему был вынесен по обвинению в участии в антисоветской вредительской организации. Г.Д. Карпеченко и Н.И. Вавилов были приговорены к расстрелу в один день (9 июля 1941 г.) (Захаров, 2000). Позднее расстрел Н.И. Вавилову был заменен на 20-летнее заключение. Георгий Дмитриевич был расстрелян 28 июля 1941 г., реабилитирован посмертно 21 апреля 1956 г. Подробности

пересмотра его дела освещены в статье Д.В. Лебедева (1994).

В 1941 г. кафедра генетики растений ЛГУ была разгромлена и прекратила свое существование (Медведев, 2006). Еще раньше, в конце 1940 г., был ликвидирован отдел генетики в ВИР (Лебедев, 1994).

Семья Георгия Дмитриевича в ту пору еще не знала о его трагической кончине. Летом 1941 г. его жена Галина Сергеевна с дочкой жили на даче у своего отца в Ильинке Московской области, куда пригласили и жену Н.И. Вавилова с сыном Юрием. Это был жест великого гражданского мужества – пригласить жену арестованного академика, от которой в ту пору отвернулись многие.

О кончине Г.Д. Карпеченко не знал никто еще более десятилетия. Академик Дмитрий Николаевич Прянишников, который с величайшим мужеством боролся с властями за реабилитацию своего ученика Н.И. Вавилова, дважды: в 1943 г. и после окончания войны 20 июля 1945 г., обращается к наркому внутренних дел Л.П. Берии с просьбой о помиловании Г.Д. Карпеченко: «Профессора Карпеченко следует отнести к выдающимся представителям работников науки, от него можно ожидать дальнейшего крупного развития и роста. Учитывая ту пользу, которую Карпеченко принес и может принести в дальнейшем нашей родине, я обращаюсь к Вам с просьбой способствовать обеспечению возможности работы профессору Карпеченко в одном из крупных исследовательских институтов нашей страны. Применение, в связи с победой над гитлеровской Германией, указа об амнистии к профессору Карпеченко сыграло бы большую роль в усилении нашей науки и ее дальнейшем росте после великих побед над ненавистным фашизмом. В случае же, если бы закон об амнистии оказался неприложимым к делу Г.Д. Карпеченко, то, по крайней мере, необходимо поставить его в условия, благоприятные для работы в самой системе Наркомвнудела, чтобы его познания и дарования могли бы послужить на пользу нашей родине» (Россиянов, 1991. С. 531).

Г.Д. Карпеченко не увидел практической реализации своих идей в селекции растений. Но его сотрудники и ученики внесли значительный вклад в развитие исследований по

использованию полиплоидии и отдаленной гибридизации в практику растениеводства (Чуксанова, 1974). Уже А.Н. Лутков со своими сотрудниками получили полиплоидные формы льна, сахарной свеклы, редиса, капусты, кукурузы, ржи, гороха, мяты и ряда других растений (Гончаров, Шумный, 2005). Академик М.И. Хаджинов создал принципиально новую методологию получения высокопродуктивных гибридов кукурузы на основе ЦМС (Лебедев, Светозарова, 1994). Ю.П. Мирюта разрабатывал генетические и цитогенетические основы гетерозиса (Захаров, Шумный, 2005а, б). Г.Д. Карпеченко и М.С. Навашину принадлежит идея закрепления гетерозиса через апомиктическое размножение (Соколов и др., 2007). О.Н. Сорокина (1958) показала возможность передачи иммунитета от эгилопсов в селекционный материал посредством получения эгилопсо-пшеничных гибридов. Р.Х. Макашева разрабатывала генетические основы селекции гороха (Вишнякова и др., 2002).

Этот перечень достижений может быть очень длинным, поскольку многие генетики, и не являвшиеся непосредственными учениками Г.Д. Карпеченко, продолжают дело, начатое им.

### Благодарности

Считаем своим приятным долгом поблагодарить проф. И.К. Захарова, к.б.н. Л.И. Лайкову, д.б.н. В.А. Соколова (ИЦиГ СО РАН, Новосибирск) и к.б.н. О.Ю. Елину (ИИЕиТ РАН, Москва) за полезное конструктивное обсуждение статьи, а также зав. архивом ВНИИР им. Н.И. Вавилова З.И. Михайлову за возможность работы с архивными материалами.

### Литература

- Александров А.Б. ВИР и его очередные задачи в борьбе за поднятие урожайности // Селекция и семеноводство. 1936. № 5. С. 11–14.
- Бабков В.В. Путевые письма и микроэволюция Ф.Г. Добржанского // Информ. вестник ВОГиС. 2007. Т. 11. № 2. С. 463–469.
- Барулина Е.И. Чечевица СССР и других стран (Ботанико-морфологическая монография). Л., 1930. 319 с. (Прилож. 40-е к Трудам по прикл. ботан., генет. и селекции).
- Барулина Е. Сравнительно-генетическое изучение видов *Triticum*. I. Генетика признака *ligula* у разнотелосомных видов пшениц: *T. vulgare* Vill., *T. compactum* Host, *T. durum* Desf. // Тр. по прикл. ботан., генет. и селекции. 1937. Сер. I. Т. 5. С. 127–166.
- Бечус К.М., Марков Х.Н. Пионер селекции. Вильнюс: Минтис, 1966. С. 92.
- Вавилов Н.И. Межвидовая и межродовая гибридизация в селекции // Изв. АН СССР. Сер. биол. 1938. № 3. С. 543–563.
- Вишнякова М.А. «Моя милая прекрасная Леночка...». Елена Барулина – жена и соратница Николая Вавилова. СПб.: Серебряный век, 2007. 152 с.
- Вишнякова М.А., Булынец С.В., Петрова М.В. Пауза Хадиевна Макашева. СПб.: «Копи-Р», 2002. 30 с. (Серия «Люди науки»).
- Всесоюзный съезд по генетике, селекции, семеноводству и племенному животноводству. Программа занятий съезда. Л.: Изд. Организационного бюро, 1929. 27 с.
- Гончаров Н.П., Шумный В.К. Методы генетики в селекции растений: к 80-летию Сибирского НИИ растениеводства и селекции // Информ. вестник ВОГиС. 2005. Т. 9. № 2. С. 395–403.
- Делоне Л.Н. Сравнительно-кариологическое исследование видов *Muscari* Mill. и *Bellevalia* Lapeyr. // Вестн. Тифлис. ботан. сада. 1922. Т. 2. Вып. 1. С. 1–32.
- Елина О.Ю. «Наш учитель» Дионисий Леопольдович Рудзинский: к истокам дисциплинарного строительства селекции растений в России // Информ. вестник ВОГиС. 2007. Т. 11. № 3/4. С. 575–590.
- Зарубайло Т.Я. Г.Д. Карпеченко и межвидовая гибридизация растений // Тр. по прикл. ботан., генет. и селекции. 1975. Т. 54. Вып. 1. С. 262–267.
- Захаров И.А. Николай Иванович Вавилов и страницы истории советской генетики. М.: ИОГен им. Н.И. Вавилова РАН, 2000. С. 124–125.
- Захаров И.К., Шумный В.К. Мирюта Юрий Петрович: к 100-летию со дня рождения (25.02.1905–22.10.1976) // Генетика. 2005а. Т. 41. № 3. С. 286–288.
- Захаров И.К., Шумный В.К. 100-летие со дня рождения генетика и селекционера Юрия Петровича Мирюты // Информ. вестник ВОГиС. 2005б. Т. 9. № 1. С. 73–78.
- Зыбина С.П. Воспоминания о Н.И. Вавилове. М.: ФГОУ ВПО РГАУ-МСХА им. К.А. Тимирязева, 2007. 123 с.
- Жегалов С.И. Введение в селекцию сельскохозяйственных растений. Л.: Гос. изд-во, 1924. 484 с.
- Карпеченко Г.Д. Число хромосом и генетические взаимоотношения у культурных *Cruciferae* // Тр. по прикл. ботан., генет. и селекции. 1922/1923. Т.

13. Вып. 2. С. 4–14.
- Карпеченко Г.Д. Межвидовые гибриды *Raphanus sativus* L. × *Brassica oleracea* L. // Науч.-агроном. журнал. 1924. Т. 1. № 3/4. С. 390–410.
- Карпеченко Г.Д. Кариологический очерк рода *Trifolium* L. // Тр. по прикл. ботан. и селекции. 1924/1925. Т. 14. Вып. 1. С. 271–279.
- Карпеченко Г.Д. О хромосомах видов фасоли // Тр. по прикл. ботан. и селекции. 1925. Т. 14. Вып. 2. С. 143–148.
- Карпеченко Г.Д. Новые данные по гибридизации *Aegilops* с пшеницами // Тр. по прикл. ботан. и селекции. 1927а. Т. 17. Вып. 4. С. 343–350.
- Карпеченко Г.Д. Полиплоидные гибриды *Raphanus sativus* L. × *Brassica oleracea* L. (К проблеме экспериментального видообразования) // Тр. по прикл. ботан. и селекции. 1927б. Т. 17. Вып. 3. С. 305–410.
- Карпеченко Г.Д. Успехи генетики в области формообразования // Достижения и перспективы в области прикладной ботаники и селекции. Л.: Изд. ИПБиНК и ГИОА, 1929. С. 71–86.
- Карпеченко Г.Д. К синтезу константных гибридов из трех видов // Тр. Всесоюз. съезда по генетике, селекции, семеноводству и племенному животноводству. Л.: Изд. редколлегии съезда, 1930. Т. 2. С. 277–294.
- Карпеченко Г.Д. Изучение отдаленной гибридизации в СССР // Соц. растениеводство. Сер. А. 1934. № 10. С. 41–50.
- Карпеченко Г.Д. Теория отдаленной гибридизации // Теоретические основы селекции растений. Т. 1. М.; Л.: Сельхозгиз, 1935а. С. 293–354.
- Карпеченко Г.Д. Экспериментальная полиплоидия и гаплоидия // Теоретические основы селекции растений. Т. 1. М.; Л.: Сельхозгиз, 1935б. С. 398–434.
- Карпеченко Г.Д. Выступление по докладам // Спорные вопросы генетики и селекции. М.: Изд-во ВАСХНИЛ, 1937а. С. 281–284.
- Карпеченко Г.Д. Экспериментальное получение тетраплоидных гибридов *Brassica oleracea* × *Brassica carinata* Al. Braun // Тр. по прикл. ботан., генет. и селекции. Сер. 2. 1937б. № 7. С. 53–68.
- Карпеченко Г.Д. Скрещиваемость видов и удвоение хромосомного набора // Изв. АН СССР. 1938а. Сер. биол. № 3. С. 693–694.
- Карпеченко Г.Д. Тетраплоидные ячмени, полученные действием высокой температуры // Биол. журнал. 1938б. Т. 7. № 2. С. 287–294.
- Карпеченко Г.Д. О поперечном делении хромосом под влиянием колхицина // Докл. АН СССР. 1940а. Т. 29. № 5/6. С. 402–404.
- Карпеченко Г.Д. Тетраплоидные шестирядные ячмени, полученные обработкой колхицином // Докл. АН СССР. 1940б. Т. 27. № 1. С. 48–51.
- Карпеченко Г.Д. Избранные труды. М.: Наука, 1971. 303 с.
- Карпеченко Г.Д., Сорокина О.Н. Гибриды *Aegilops triuncialis* L. с рожью // Тр. по прикл. ботан., генет. и селекции. 1929. Т. 20. Вып. С. 563–584.
- Карпеченко Г.Д., Щавинская С.А. О половом обособлении тетраплоидных гибридов *Raphanus* × *Brassica* // Тр. Всесоюз. съезда по генетике, селекции, семеноводству и племенному животноводству. Т. 2. Л.: Изд. Редколлегии съезда, 1930. С. 267–276.
- Классики советской генетики (1920–1940 гг.). Л.: Наука, 1968. С. 461–511.
- Конашев М.Б. Об одной научной командировке, оказавшейся бессрочной // Репрессированная наука. Л.: Наука, 1991. С. 240–263.
- Конашев М.Б. «Невозвращенец» поневоле // Российские ученые и инженеры в эмиграции. М.: ПО «Перспектива», 1993. С. 136–146.
- Костов Д. Экспериментальное получение полиплоидных гибридов в роде *Nicotiana* и их значение // Природа. 1934. № 8. С. 44–49.
- Костов Д. Изучение полиплоидных растений. XI. Амфилоид *T. timopheevii* Zhuk. × *T. monocosmum* L. // Докл. АН СССР. 1936а. Т. 1 (10). № 1. С. 32–36.
- Костов Д. Наследование естественного иммунитета у растений и получение иммунных разновидностей путем межвидовой гибридизации // Проблемы иммунитета культурных растений: Тр. майской сессии 1935 г. М.; Л.: Изд-во АН СССР, 1936б. С. 22–41.
- Костов Д. Происхождение и селекция пшениц с цитогенетической точки зрения // Изв. АН СССР. Отд. биол. наук. 1940. № 1. С. 56–94.
- Лебедев Д.В. Георгий Дмитриевич Карпеченко // Выдающиеся советские генетики. Сб. биографических очерков. М.: Наука, 1980. С. 37–48.
- Лебедев Д.В. Георгий Дмитриевич Карпеченко // Соратники Н.И. Вавилова. СПб.: ВИР, 1994. С. 210–228.
- Лебедев Д.В., Абрамова Л.И. Григорий Андреевич Левитский // Соратники Н.И. Вавилова. СПб.: ВИР, 1994. С. 307–322.
- Лебедев Д.В., Светозарова В.В. Михаил Иванович Хаджинов // Соратники Н.И. Вавилова. СПб.: ВИР, 1994. С. 556–565.
- Левина Е.С. Из истории отечественной генетики. Н.И. Вавилов и Г.Д. Карпеченко // Генетика, 1987. Т. 22. № 11. С. 2007–2019.
- Левитский Г.А. Материальные основы наследственности. Киев: ГИЗ Украины, 1924. 166 с.
- Левитский Г.А. Успехи генетической цитологии и ее применение к культурным растениям // До-



- стижения и перспективы в области прикладной ботаники и селекции. Л.: Изд-во ИПБиНК и ГИОА, 1929. С. 87–98.
- Левитский Г.А. Морфология хромосом и понятие «кариотипа» в систематике (на основе исследования подсемейства *Hellebogaee*) // Тр. по прикл. ботан., генет. и селекции. 1931. Т. 27. Вып. 1. С. 187–240.
- Левитский Г.А., Бенецкая Г.К. Цитология пшенично-ржаных амфидиплоидов // Тр. по прикл. ботан., генет. и селекции. 1931. Т. 27. Вып. 1. С. 241–264.
- Лутков А.Н. Межвидовые гибриды *Pisum humile* Boiss. × *Pisum sativum* L. // Тр. Всесоюз. съезда по генетике, селекции, семеноводству и племенному животноводству. Т. 2. Л.: Изд. Редколлегии съезда, 1930. С. 353–367.
- Лутков А.Н. Об экспериментальном получении полиплоидных гамет под влиянием низкой температуры и хлороформа // Тр. по прикл. ботан., генет. и селекции. 1937а. Сер. II. Вып. 7. С. 127–150.
- Лутков А.Н. Экспериментальное получение безлигильной формы ячменя под влиянием X-лучей // Тр. по прикл. ботан., генет. и селекции. 1937б. Сер. II. Вып. 7. С. 197–202.
- Лутков А.Н. Массовое получение тетраплоидных растений льна действием колхицина // Докл. АН СССР. 1939. Т. 29. № 4. С. 177–181.
- Макушина Е.Н. Новый вид пшеницы *Triticum armeniacum* (Jakubz.) sp. n. // Докл. АН СССР. 1938. Т. 21. № 7. С. 350–353.
- Матвеева Л.В. Отто Юльевич Шмидт. М.: Наука, 1993. 202 с.
- Медведев Н.Н. Юрий Александрович Филипченко: 1882–1930. Изд. 2-е исп. и доп. М.: Наука, 2006. 230 с.
- Морган Т.Х. Экспериментальные основы эволюции. М.; Л.: Биомедгиз, 1936. 250 с.
- Навашин С.Г. Резюме возражений на доклад Л.Н. Делоне // Журн. русск. ботан. об-ва. 1921. № 6.
- Навашин С.Г. Об изменении числа и морфологических признаков хромосом и межвидовых гибридов // Тр. по прикл. ботан. и селекции. 1927. Т. 17. Вып. 3. С. 121–150.
- Николаева А.Г. Применение цитологического метода при решении некоторых вопросов генетики // Тр. III Всерос. съезда по селекции и семеноводству. Саратов, 1920. С. 31.
- Николаева А.Г. Цитологическое исследование рода *Triticum* // Тр. по прикл. ботан. и селекции. 1922/1923. Т. 13. Вып. 1. С. 33–44.
- Николаева А.Г. Материалы по цитологии ржано-пшеничных гибридов // Научно-агрон. журнал. 1924. Т. 1. № 9/10. С. 570–576.
- Писарев В.Е. Амфиплоиды «яровая пшеница × яровая рожь» // Писарев В.Е. Селекция зерновых культур. Избр. работы. М.: Колос, 1964. С. 286–315.
- Полумордвинова И.В. «Ученый с выдающейся эрудицией... верным глазом, редкой интуицией...» (штрихи к портрету С.И. Жегалова) // ВИЕТ. 2007. № 2. С. 140–164.
- Ригин Б.В. Г.Д. Карпеченко и развитие генетики во Всесоюзном институте растениеводства // Тр. по прикл. ботан., генет. и селекции. 1989. Т. 128. С. 97–103.
- Ригин Б.В. Н.И. Вавилов и основные направления и результаты исследований в отделе генетики ВНИИ растениеводства // Информ. вестник ВОГиС. 2007. Т. 11. № 3/4. С. 525–536.
- Розанова М.А. Экспериментальные основы систематики растений. М.; Л.: Изд-во АН СССР, 1946. 256 с.
- Россиянов К.О. Из истории борьбы академика Д.Н. Прянишникова за генетику // Репрессированная наука. Л.: Наука, 1991. С. 528–533.
- Рубцова З.М. Значение работ Г.Д. Карпеченко для развития эволюционной цитогенетики // Из истории биологии. Т. 4. М.: Наука, 1973. С. 148–159.
- Рубцова З.М. Развитие эволюционной генетики растений в СССР. (1920–1940 гг.). Л.: Наука, 1975. 172 с.
- Рубцова З.М. Хромосомные перестройки и полиплоидия // Эволюционные учения в СССР (1917–1970 гг.). Л.: Наука, 1983. С. 92–116.
- Светозарова В.В. О втором геноме *T. timopheevi* Zhuk. // Докл. АН СССР. 1939. Т. 23. № 5. С. 472–476.
- Свешникова И.Н. Цитогенетика рода *Vicia*. М.: Наука. 1979. 152 с.
- Соколов В.А., Тараканова Т.К., Абдырахманова Э.А. Отдаленные скрещивания как источник расширения экологической устойчивости кукурузы // Информ. вестник ВОГиС. 2007. Т. 11. № 3/4. С. 650–656.
- Сорокина О.Н. Гибридизация эгилопсов с пшеницей // Тр. по прикл. ботан., генет. и селекции. 1934. Сер. 2. Вып. 6. С. 7–37.
- Сорокина О.Н. К синтезу видов эгилопса // Тр. по прикл. ботан., генет. и селекции. 1937а. Сер. 2. Вып. 7. С. 151–160.
- Сорокина О.Н. Новые эгилопсо-пшеничные амфиплоиды // Тр. по прикл. ботан., генет. и селекции. 1937б. Сер. 2. Вып. 7. С. 161–173.
- Сорокина О.Н. Роль амфиплоидов и других сбалансированных форм в отдаленных скрещиваниях // Докл. АН СССР. 1938. Т. 7. № 7/8. С. 597–600.
- Сорокина О.Н. Иммуный эгилопсо-пшеничный гибрид и его использование // Тез. совещ. по отд. гибридизации растений и животных. М., 1958. Т. 1.

- У истоков академической генетики в Санкт-Петербурге / Ред. Э.И. Колчинский, К.В. Манойленко, М.Б. Конашев. Сост. М.Б. Конашев. СПб.: Наука, 2002. 558 с.
- Филимонов П.Н. Хроника полувека // Бюл. ВНИИР им. Н.И. Вавилова. 1972. Вып. 24. С. 81–84.
- Хлоп М.Л. Проблема экспериментального получения мутаций у растений под влиянием X-лучей // Тр. по прикл. ботан., генет. и селекции. 1936. Сер. А. Соц. раст. № 18. С. 149–159.
- Чуксанова Н.А. Полиплоидия и видообразование у растений // Теоретические и практические проблемы полиплоидии. М.: Наука, 1974. С. 64–80.
- Шварц А.Л. Тогда, в Пасадене // «Word/Слово». 2005. № 47. (URL: <http://magazines.russ.ru/slovo/2005/47/shv25.html> (дата обращения: 04.02.2009)).
- Шварц А.Л. Этот счастливец Г.Д. Карпеченко // «Word/Слово». 2008. № 60. (URL: <http://magazines.russ.ru/slovo/2008/60/sh35.html> (дата обращения: 04.02.2009)).
- Шульдин А.Ф. Генетические закономерности расщепления отдаленных гибридов растений // Проблемы отдаленной гибридизации растений. София: Изд-во Болг. АН, 1978. С. 33–44.
- Шумный В.К., Першина Л.А. Отдаленная гибридизация растений // Вавиловское наследие в современной биологии. М.: Наука, 1989. С. 220–230.
- Щавинская С.А. Восстановление плодовитости у герани (*Pelargonium radula roseum* W.) путем удвоения хромосомного комплекса // Тр. по прикл. ботан., генет. и селекции. Сер. 2. 1937а. № 7. С. 101–106.
- Щавинская С.А. Тетраплоидная капуста, полученная путем регенерации // Тр. по прикл. ботан., генет. и селекции. Сер. 2. 1937б. № 7. С. 13–36.
- Щавинская С.А. Экспериментально полученная октоплоидная капуста // Тр. по прикл. ботанике, генетике и селекции. Сер. 2. 1937в. № 7. С. 69–77.
- Blakeslee A.F., Avey A.C. Methods of inducing doubling of chromosomes in plants // J. Hered. 1937. V. 28. № 12. P. 393–411.
- Chen H.G., Wu J.S. Characterization of fertile amphidiploid between *Raphanus sativus* and *Brassica alba glabra* and the crossability with *Brassica* species // Genet. Res. Crop Evol. 2008. V. 55. P. 143–150.
- Clausen R., Goodspeed T. Inheritance hybridization in *Nicotiana*. II. A tetraploid *glutinosa-tabacum* hybrid, an experimental verification of Winge's hypothesis // Genetics. 1925. V. 10. P. 278–284.
- Gravatt F. A radish – cabbage hybrid // J. Heredity. 1914. V. 5. P. 269–272.
- Jørgensen C.A. The experimental formation of heteroploid plants in the genus *Solanum* // J. Genetics. 1928. V. 19. P. 133–211.
- Karpechenko G.D. Hybrids of *Raphanus sativus* L. × *Brassica oleracea* L. // J. Genetics. 1924. V. 14. № 3. P. 375–394.
- Karpechenko G.D. Polyploid hybrids of *Raphanus sativus* L. × *Brassica oleracea* L. // Z. Inductive Abstammungs und Vererbungslehre. 1928. Bd. 48. H. 1. S. 1–84.
- Karpetschenko G.D. Konstantwerden von Art- und Gattungsbastarden durch Verdoppelung der Chromosomenkomplexe // Züchter. 1929. Bd. 1. H. 5. S. 133–140.
- McNaughton I.H. Synthesis and sterility of *Raphanobrassica* // Euphytica. 1973. V. 22. P. 70–88.
- Metz P.L.J., Nap J.-P., Stiekema W.J. Hybridization of radish (*Raphanus sativus* L.) and oilseed rape (*Brassica napus* L.) through a flower-culture method // Euphytica. 1995. V. 83. P. 159–168.
- Randolf L.F. Some effects of high temperature on polyploidy and other variations in maize // Proc. Natl Acad. Sci. USA. 1932. V. 18. P. 222–229.
- Sageret M. Considérations sur la production des hybrides, des variantes et des variétés en général, et sur celles de la famille des Cucurbitacées en particulier // Annals Sci. Nat. 1826. V. 8. P. 294–314.
- Terasawa Y. Konstante amphidiploide *Brassicoraphanus* Bastarde // Proc. Imp. Acad. (Tokyo). 1932. V. 8. P. 312–314.
- Winge O. Chromosomes. Their numbers and general importance // C.R. Trav. Lab. Carsberg. 1917. V. 13. P. 131–275.
- Yarnell S.H. Cytogenetics of the vegetable crops. II. Crucifers // Bot. Rev. 1956. V. 22. № 2. P. 81–166.

**COMMEMORATING THE 110th ANNIVERSARY  
OF GEORGYI DMITRIEVICH KARPECHENKO  
(03.05.1899–28.07.1941)**

**V.A. Vishnyakova<sup>1</sup>, N.P. Goncharov<sup>2</sup>**

<sup>1</sup> N.I. Vavilov Institute of Plant Industry, St.-Petersburg, Russia; e-mail: m.vishnyakova@vir.nw.ru;

<sup>2</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: gonch@bionet.nsc.ru

**Summary**

G.D. Karpechenko became world-famous at the age of 28. He became in fact the first genetic engineer who has developed a new branch of biology during less than 20 years of his scientific activity. The word ‘Raphanobrassica’ enriched the vocabulary of all the languages. His classical experiments of overcoming the infertility of man-maid hybrids are described in all the genetic and cytogenetic textbooks. However, not so much is written about G.D. Karpechenko. Till now the documents kept in archives and by his relatives are not processed and published. The present article is an attempt to pay a tribute to the scientist on the occasion of his 110th anniversary.

**ИЗБРАННЫЕ ТРУДЫ  
VI МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ  
ПО БИОИНФОРМАТИКЕ  
РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМА  
(BGRS'2008)**

**Новосибирск, Россия  
22–28 июня 2008 г.**

**Ответственные редакторы  
И. Р. Акбердин, д.б.н. Д. П. Фурман**

**Международный программный комитет конференции:**

- Н. А. Колчанов**, Институт цитологии и генетики СО РАН, Новосибирск, Россия (председатель)
- Ralf Hofstaedt**, Билефельдский Университет, Германия (сопредседатель)
- Д. П. Фурман**, Институт цитологии и генетики СО РАН, Новосибирск, Россия (ученый секретарь)
- Д. А. Афонников**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- Shandar Ahmad**, Национальный институт биомедицинских инноваций, Япония
- Samir Brahmachari**, Институт геномики и интегративной биологии, Дели, Индия
- Philip Bourne**, Калифорнийский университет Сан-Диего, США
- Е. И. Верещагин**, Сибирский центр фармакологии и биотехнологии, Новосибирск, Россия
- Е. Е. Витяев**, Институт математики им. Соболева СО РАН, Новосибирск, Россия
- В. В. Власов**, Институт химической биологии и фундаментальной медицины СО РАН, Новосибирск, Россия
- М. С. Гельфанд**, Институт проблем передачи информации РАН, Москва, Россия
- Б. М. Глинский**, Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия
- Н. П. Гончаров**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- Л. А. Животовский**, Институт общей генетики РАН, Москва, Россия
- А. С. Иванов**, Институт биомедицинской химии РАМН, Москва, Россия
- Manfred Kayser**, Университет им. Эразма Роттердамского, Роттердам, Нидерланды
- Ф. А. Кондрашов**, Калифорнийский университет в Сан-Диего, США
- А. В. Кочетов**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- Jun-ichi Kudoh**, Центр исследований Северо-Восточной Азии, Университет Тохоку, Япония
- В. Н. Кузнецов**, Отделение анализа генома и экспрессии генов, Институт биоинформатики, Сингапур
- М. М. Лаврентьев**, Институт математики им. Соболева, Институт автоматизации и электрометрии СО РАН, Новосибирск, Россия
- С. Лукашук**, Институт фундаментальной и прикладной математики, Университет Халла, Великобритания
- Kenji Mizuguchi**, Группа биоинформатики и компьютерной биологии, Национальный институт биомедицинских инноваций, Осака, Япония
- Luciano Milanesi**, Институт биомедицинской технологии, Милан, Италия
- Ming Chen**, Отделение биоинформатики, Чжецзянского университета, Ханьжоу, КНР
- М. П. Мошкин**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- В. В. Поройков**, Институт биомедицинской химии РАМН, Москва, Россия
- Jagath S. Rajapakse**, Школа компьютерной технологии, Наньянский технологический университет, Сингапур
- И. Б. Rogozin**, Национальный центр биотехнологической информации, Национальный институт здоровья, Бесезда, США
- Н. Б. Рубцов**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- М. Г. Самсонова**, Санкт-Петербургский политехнический университет, Санкт-Петербург, Россия
- Akinori Sarai**, Институт технологии Кюсю (КИТ), Иизука, Япония
- К. Г. Скрябин**, Центр «Биоинженерия» РАН, Москва, Россия
- Н. Ю. Сурнина**, Институт цитологии и генетики СО РАН, Новосибирск, Россия
- A. F. Famili**, Университет Оттавы, Национальный научно-исследовательский совет Канады, Оттава, Канада
- Charlie Hodgman**, Мультидисциплинарный центр интегративной биологии, Университет Ноттингема, Великобритания

## Организаторы



Учреждение Российской академии наук  
Институт цитологии и генетики  
Сибирского отделения РАН (ИЦиГ СО РАН)



Сибирское отделение Российской академии наук



Вавиловское общество генетиков и селекционеров



Лаборатория теоретической генетики ИЦиГ СО РАН



Новосибирский государственный университет



Кафедра информационной биологии НГУ



PBSoft Ltd.



Институт вычислительных технологий

## Спонсоры



Российский фонд фундаментальных исследований



Carl Zeiss



Bio-Rad



Scientific Professional Equipment



Сибирское отделение Российской академии наук



Федеральное агентство по науке и инновациям



Hewlett Packard



HTLabAG

## МОРФОГЕНЕЗ, МОРФОМЕХАНИКА И ГЕНОМ

Л.В. Белоусов

Биологический факультет Московского государственного университета, Россия,  
e-mail: morphogenesis@yandex.ru

Морфогенез относится к процессам самоорганизации и поэтому должен быть основан на нелинейных обратных связях между его динамическими компонентами. Предлагается модель обратных связей между пассивными и активными механическими напряжениями в эмбриональных тканях, которые могут играть роль движущих сил морфогенеза. Обсуждается соотношение генетических факторов и морфогенетических процессов в индивидуальном развитии и эволюции.

**Ключевые слова:** морфогенез, механическое напряжение, модели развития, экспрессия генов.

Морфогенез в широком понимании – это образование новых форм и структур (от внутриклеточных до макроскопических) в ходе эмбрионального развития организмов. (Внутриклеточные морфогенезы обычно называют цитодифференцировками.) С морфогенезом связан буквально каждый процесс индивидуального развития, равно как и эволюционные преобразования. Бесспорно, морфогенез представляет собой наиболее подробный и в то же время упорядоченный процесс самоусложнения из всех происходящих как в живой, так и в неживой природе.

Тем не менее общепринятой теории морфогенеза до сих пор не существует. Мы не можем ответить, почему по ходу развития одни пространственные структуры сменяются строго определенными другими или почему в некоторых случаях закономерные финальные структуры возникают из переменных и даже беспорядочных, находящихся на грани хаоса. Но самое парадоксальное состоит в том, что, как правило, вопрос так и не ставится, и что подавляющее большинство исследователей вообще не рассматривает морфогенез как самостоятельную проблему. Более того, пока считается, что такой проблемы вообще не существует.

Такой взгляд достаточно традиционен. Долгое время в эмбриологии господствовал преформизм – учение, утверждающее, что морфогенез как процесс самоусложнения вообще фиктивен,

а все структуры организма пространственно размечены «изначально», в некий загадочный начальный момент развития. И хотя открытие Г. Дришем эмбриональных регуляций более 100 лет тому назад строго доказало отсутствие изначальной «разметки», отдельные фрагменты этого воззрения существуют до сих пор. К ним относятся утверждения, что сам по себе развивающийся организм не обладает самостоятельной динамикой, а лишь пассивно воспринимает предельно детализированную информацию, записанную на какой-либо другой матрице. Таковы существующие представления о том, что морфогенез и вообще все процессы развития «запрограммированы генетически». Доведение этого утверждения до его логического предела означало бы, что если мы располагаем полной информацией о структуре генома особи данного вида, мы сможем предсказать его морфогенез.

Но вся совокупность имеющихся фактов – как классических, так и недавних – показывает несостоятельность этого утверждения (впрочем, и без того экспериментально непроверяемого). Прежде всего, морфогенез лишь в малой степени видоспецифичен – его наиболее фундаментальные процессы протекают сходным образом у генетически весьма различных организмов. Еще важнее помнить, что геном и морфогенез – сущности совершенно разного порядка. При всех изменениях взглядов на гены со времен Менделя они всегда рассматриваются



как статические дискретные факторы, тогда как морфогенез – это разворачивающийся в пространстве–времени континуальный (нерасчленимый по ходу своему на отдельные признаки) процесс. Понятно, что прямая проекция одного на другой невозможна. Даже если принять, что каждый шаг морфогенеза связан с активацией или репрессией определенных генов (на самом деле это не так), то пространственно-временное расписание активации/репрессии генов должно определяться не ими самими, а вне- (эпи-) генетическими факторами, прямо или косвенно связанными с морфогенезом. Однако исследования последних десятилетий говорят о существенно большем: оказалось, что одни и те же (точнее, гомологичные) наборы генов, белковые факторы и/или сигнальные каскады обеспечивают как у разных видов, так и на разных стадиях развития одного и того же вида совершенно разные морфогенетические процессы. И, с другой стороны, гомологичные эмбриональные структуры у близких видов могут иметь негомологичное генетическое обеспечение. Это было ясно уже Н.И. Вавилову (Чайковский, 2008. С. 165) и позже было подтверждено во множестве работ (Wray, Abouheif, 1998). Следовательно, зная, какой ген экспрессируется или какой сигнальный каскад в данный момент развития и в данной локализации работает, мы ничего не можем сказать о том, какой морфогенетический процесс состоится. Очевидно, что природа располагает весьма ограниченным набором генетических факторов и сигнальных путей, которыми зародыш распоряжается согласно плану своего развития как необходимыми, но не содержащими информацию орудиями: молоток необходим, чтобы забить гвоздь, но не он определяет, где и когда гвоздь будет забит.

Таким образом, трудно избежать парадоксального на первый взгляд утверждения, что *генетические факторы при всей своей важности вовсе не содержат в себе сколько-нибудь однозначной информации о развитии*. Последняя, очевидно, имеется (а точнее – возникает по ходу дела) в самой развивающейся системе как таковой. Это вплотную подводит нас к представлению о морфогенезе как о процессе *самоорганизации*.

Данное утверждение не является чисто словесным, как иногда думают. Современная

теория самоорганизации (ТСО) – это мощная синтетическая отрасль современной науки с математической «сердцевиной» и ответвлениями в сторону физики, химии и биологии. Для первого ознакомления с ТСО можно рекомендовать работу Капра (2002). Ничего ближе к фундаментальным проблемам морфогенеза, нежели ТСО, в науке пока не создано. Какие же свойства морфогенеза могут быть адекватно описаны на языке ТСО? Прежде всего – это «самоусложнение». Данный не вполне строгий термин корректно переводится на язык ТСО как *спонтанное* (протекающее без внешних воздействий) понижение порядка симметрии объекта (диссимметризация). Многовековая тенденция считать, будто морфогенеза не существует, была связана с (часто подспудным) убеждением, что «самоусложнение» противоречит законам природы. Однако для рассматриваемых в ТСО удаленных от термодинамического равновесия нелинейных систем это не так: в определенных условиях они теряют устойчивость, и тогда возможно понижение порядка их симметрии под влиянием ничтожно малых воздействий, т. е. практически спонтанно.

Нелинейные взаимодействия подразумевают наличие обратных связей, из которых наибольшее значение для морфогенеза имеют «+, –» обратные связи между двумя переменными (первая из них стимулирует вторую, а вторая ингибирует первую). При этом возникает широкий спектр периодических временных и пространственных структур (автоколебаний и автоволн) – главных элементарных составляющих практически всех морфогенетических процессов. Важно, что такие системы непременно являются иерархиями, состоящими из структур-процессов разных характерных размеров (Sch) и времен (Tch). Факторы, управляющие их поведением, сами имеют иерархическую структуру. Факторы с малыми Sch и Tch (узколокализованные в пространстве и времени) называются динамическими, а с большими Sch и Tch («размытыми» в пространстве–времени) – параметрическими. Если динамическое управление имеет сходство с традиционным (для линейных систем) сугубо локальным пониманием причинно-следственных связей, то нелокальное параметрическое управление – принципиально новая черта нелинейных систем, причем именно

она определяет наиболее фундаментальные их особенности (например наличие и набор неустойчивостей).

Уже эти самые общие понятия ТСО совершенно по-новому высвечивают отношения между генетическими факторами и морфогенезом. Если рассматривать геном как набор отдельных генов, то последние надо относить к категории динамических факторов, поскольку они действуют на малых отрезках времени и достаточно локально. Как уже говорилось выше, эта категория генетических факторов действует не самостоятельно, а в тесной связи с факторами эпигенетическими. С другой стороны, чтобы геном играл роль параметра, он должен состоять из крупных блоков, инвариантных во времени и пространстве развития. Современные системные подходы к геному (интерес к его «крупноблочной» и трехмерной структуре и к генетическим сетям (Huang, Ingber, 2000; Dahl *et al.*, 2008; Системная ..., 2008)) к этому близки.

В последнее время ТСО активно используется для построения имитационных моделей, воспроизводящих с той или иной степенью точности некоторые реальные морфогенезы. Обсудим цель подобных работ. Понятно, что если для имитации берутся относительно короткие отрезки морфогенеза и тем более, если они не сопровождаются понижением порядка симметрии, то при свободном подборе параметров имитация заранее заданной точности «обречена на успех». Таких «подгоночных» моделей типа «что вложили, то и получили» в литературе по морфогенезу, да и в других областях биологии довольно много. На наш взгляд, они мало дают для решения фундаментальных проблем. Модели, действительно важные для понимания морфогенеза, должны обладать следующими чертами:

1) они должны предлагать схему максимально универсальных для морфогенеза (т. е. действующих на максимально протяженных отрезках пространства–времени развития) обратных связей, приводящих, помимо прочего, к понижению порядка симметрии. По крайней мере, для простейших ситуаций наличие таких обратных связей должно быть эмпирически подтверждено;

2) число свободных параметров должно быть минимальным, а сами параметры биологически правдоподобными;

3) модели должны быть достаточно «грубыми», т. е. нечувствительными к малым «шевелениям» параметров и/или начальных/граничных условий;

4) они должны порождать (пусть в некотором приближении) реальные морфологические структуры, а не некоторые их «коды» или «символи», трансформация которых в реальные структуры потребовала бы дополнительных допущений.

К настоящему времени предложено два класса моделей, имеющих отношение к морфогенезу. Первый из них – это так называемые реакционно-диффузионные модели, основанные на допущении «+, –» обратных связей между химическими компонентами, диффундирующими с разными скоростями (Meinhardt, 1982). Разработка этих моделей впервые натолкнула на важный вывод о том, что в морфогенетических процессах близкое действие имеет преимущественно активирующий характер, а дальное действие – ингибирующий. Тем не менее они не удовлетворяют пункту 1 (постулируемые обратные связи не доказаны, господство свободной диффузии в многоклеточных системах проблематично, особенности клеточной структуры вообще не учитываются) и пункту 4 (на выходе – не реальные структуры, а концентрации «морфогенов», трансформация которых в морфологические структуры требует специальных *ad hoc* допущений).

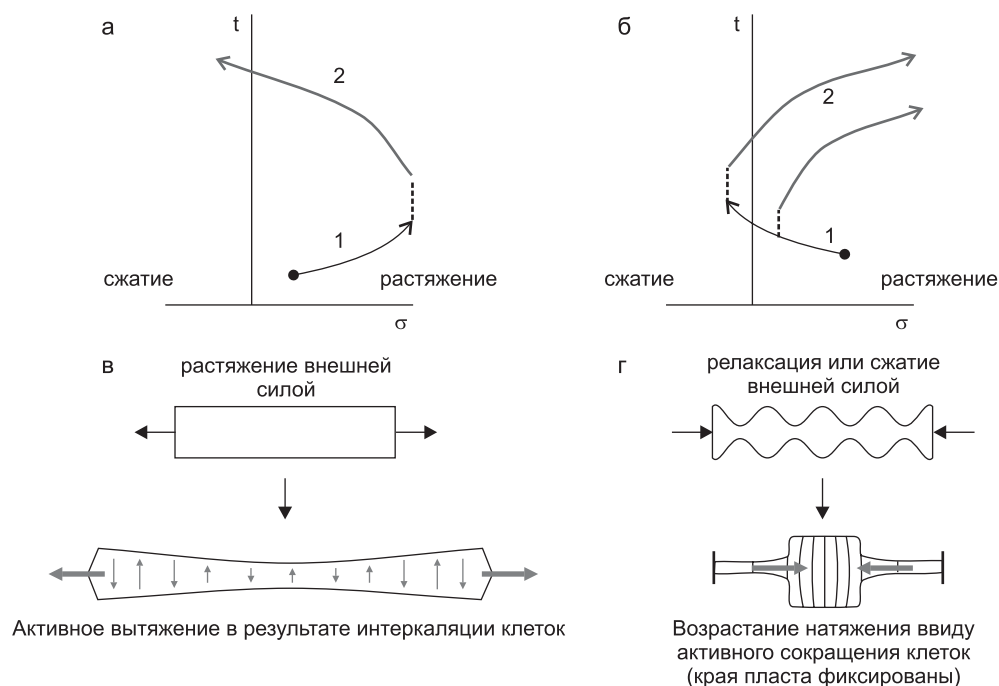
Значительно ближе к реальности, на наш взгляд, класс так называемых механо-химических моделей. Их основой является утверждение о ведущей морфогенетической роли обратных связей между так называемыми пассивными и активными механическими напряжениями (МН) в эмбриональных тканях. Пассивные МН – это те, которые вызваны механическими силами, внешними относительно рассматриваемого участка эмбриональной ткани. Активные же МН генерируются (естественно, при затрате энергии) «молекулярными машинами», локализованными внутри данного участка. Экспериментальная основа этих моделей – наличие закономерных, стадиоспецифичных рисунков МН в развивающихся зародышах (Belousov, 2008) и возможность самоорганизации в группировках клеток, посеянных на эластичные субстраты, которые они сами растягивают (Harris *et al.*, 1984).

Исторически первая механохимическая модель (Odell *et al.*, 1981) постулировала, что пассивное растяжение некоторого участка (А) клеточного пласта внешней силой порождает силу активного сокращения внутри А, которая растягивает следующий за ним участок Б, и т. д. В рамках этой модели структура обратных связей решающим образом зависит от того, закреплены или нет края пласта: если они свободны или закреплены нежестко, то активное сокращение ведет к релаксации пассивного растяжения (получаем «+, -» обратные связи), но тогда растяжение не распространяется дальше по пласту. Если же края закреплены (что соответствует большинству реальных ситуаций), то растяжение распространяется, но обратные связи будут выражаться символом «+, +», что исключает возможность формирования самоограниченных в пространстве структур, т. е. самую суть морфогенеза.

Модель Б.Н. Белинцева с соавт. (Белинцев, 1988) исходит из противоположного, но экспериментально подтвержденного допущения: не растяжение, а пассивная релаксация эпителиальной клетки стимулирует ее активное

сокращение. Тогда возникают «+, -» обратные связи, работающие при закрепленных краях пласта, что соответствует реальности. Эта модель вполне соответствует требованиям (1–4). Она не только воспроизводит широко распространенные в морфогенезе процессы сегрегации эпителиальных пластов на домены распластанных и столбчатых клеток, но и обеспечивает их масштабную инвариантность (Дришевские регуляции). Тем не менее и эта модель недостаточно универсальна для интерпретации основных процессов морфогенеза. В качестве дальнейшего обобщения (включающего в себя предыдущие модели в виде частных случаев) была предложена модель гипервосстановления МН в следующей формулировке (Белоусов, Миттенталь, 1992; Belousov, 2008).

При любом изменении величины пассивного МН, приложенного извне к образцу эмбриональной ткани, в последнем возникают активные силы, направленные на восстановление исходного значения МН, но, как правило, перекрывающие его в обратную сторону («гипервосстановление», ГВ) (рис. 1, а, б). Если навязанные извне изменения МН распределены



**Рис. 1.** Диаграммы (а, б) и схематические примеры (в, г) реакций гипервосстановления (ГВ).

а, в – ГВ-реакции на растяжение, б, г – ГВ-реакции на релаксацию/сжатие. На а, б по горизонтальной оси – механические напряжения, по вертикальной оси – время. Траектории 1 – внешние силы, вертикальный пунктир – лаг-период перед активной реакцией, траектории 2 – активные ГВ-ответы (по: Belousov, 2008).

неравномерно или анизотропно, то активные ответы будут направлены на уменьшение наибольших из навязанных МН.

Наиболее распространенные способы осуществления реакций ГВ показаны на рис. 1, в, г. Один из наиболее типичных ответов на пассивное растяжение клеточного пласта внешней силой – так называемая интеркаляция клеток, генерирующая силу внутреннего давления, и как следствие – активное вытяжение пласта (рис. 1, в). В эту же сторону направлена многократно описанная пролиферативная реакция – стимуляция клеточных делений растяжением ткани (Nelson *et al.*, 2005). Типичный ответ на релаксацию или сжатие – активное сокращение клеток (рис. 1, г) или их уход из пласта. В обоих случаях исходное натяжение восстанавливается с «перехлестом». В пассивно изогнутых клеточных пластах на растянутой выпуклой поверхности осуществляется реакция, соответствующая рис. 1, в, а на сжатой выпуклой – рис. 1, г. В результате кривизна, намеченная внешней силой, будет активно нарастать. Другим примером ГВ-реакций является увеличение площади межклеточных контактов при растяжении пласта и уменьшение при его релаксации (очевидно, что первая реакция направлена на уменьшение, а вторая – на увеличение натяжений между клетками пласта). Эти и другие ГВ-реакции многократно наблюдались (Belousov *et al.*, 2006; Belousov, 2008).

ГВ-реакции являются достаточно «грубыми» (нечувствительными к малым вариациям параметров и динамических переменных) и могут в широком диапазоне условий приводить к понижению порядка симметрии. Так, если сокращение клеток в ответ на релаксацию или интеркаляция в ответ на растяжение начнется не одновременно на всем протяжении клеточного пласта (а именно такой ответ является грубым), то в зависимости от значений параметров пласт будет расчленяться на различное число чередующихся доменов распластанных-столбчатых клеток, т. е. метамеров. При этом понизится порядок его трансляционной симметрии. Исходно сферический или цилиндрический пласт при этих же начальных условиях утратит симметрию вращения. Аналогично, если в некоторой малой области плоского или сферического клеточного пласта будет локально включен ме-

ханизм усиления кривизны, возникнет эстафета складкообразования, понижающая порядок симметрии (Belousov, Grabovsky, 2007).

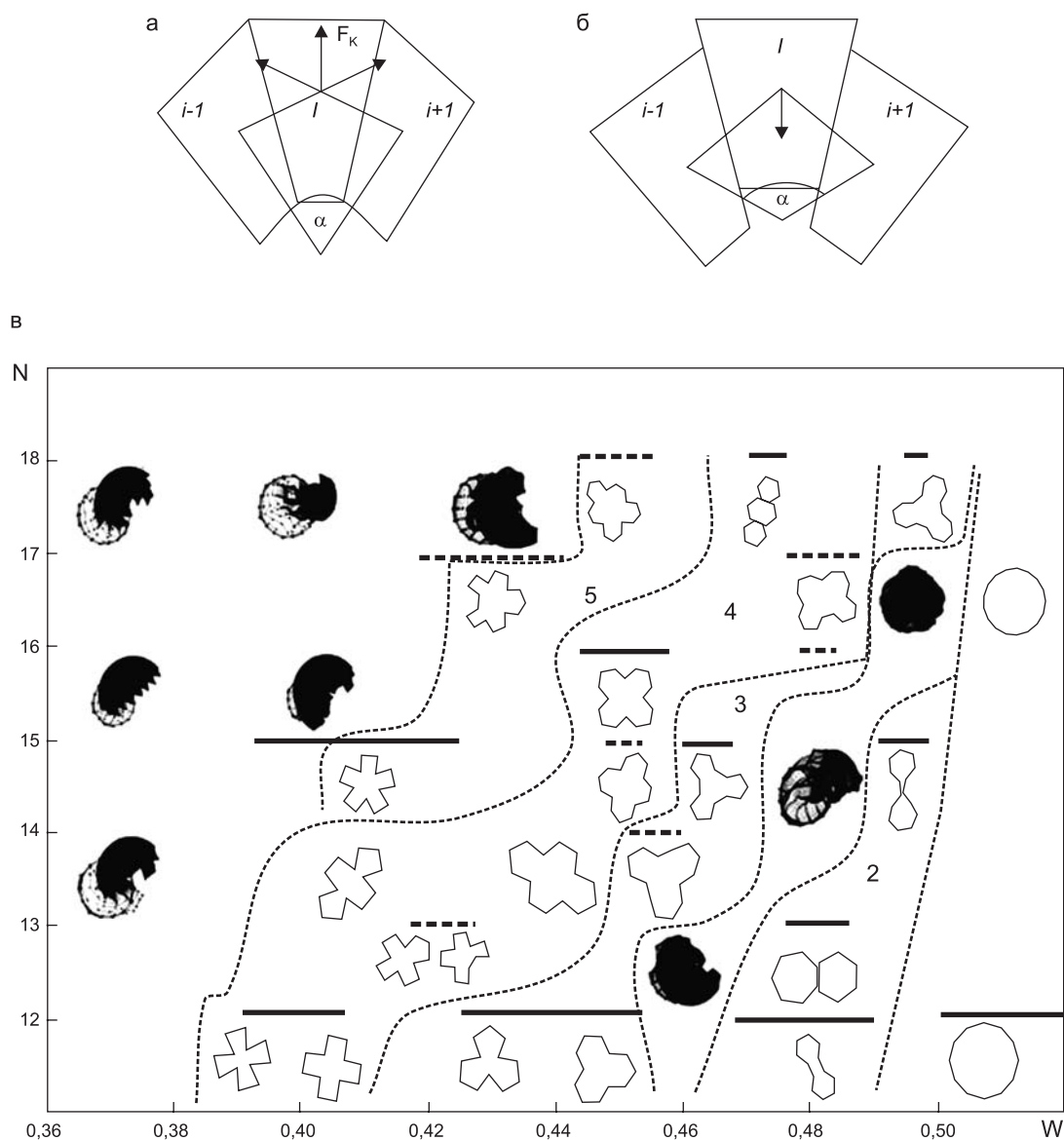
Было также показано, что разные ГВ-реакции поддерживают и усиливают друг друга. Например, в клеточном пласте с закрепленными краями активное растяжение некоторого участка вызовет сначала пассивное сжатие/релаксацию, а затем и активное сокращение смежного участка. Поэтому ГВ-реакции могут играть роль движущих факторов морфогенеза. На примере гастрюляции амфибий последовательные шаги морфогенеза удастся представить как цепи взаимообусловленных ГВ-реакций (Belousov *et al.*, 2006). Основные элементы данной схемы могут быть приложены и к другим типам гастрюляции.

Действует ли предлагаемая схема обратных связей в процессах цитодифференцировки? За последнее время были получены впечатляющие данные об активации генов механическими силами (Farge, 2003) и о решающей роли механических свойств субстрата для выбора стволовыми клетками направления их дифференцировки (Engler *et al.*, 2006). Мы получили предварительные данные о связи сжатия–растяжения эмбриональных тканей с экспрессией в них либо нейральных, либо мезодермальных генов (Белоусов и др., 2007). Все эти данные согласуются с гипотезой ГВ. Они позволяют заключить, что фундаментальные закономерности морфогенеза и цитодифференцировки едины и отражают механозависимые кооперативные процессы в цитоскелете и клеточных мембранах.

В заключение вернемся к вопросу о связях морфогенеза с эволюционными преобразованиями, имея в виду, что разные виды и систематические группы отличаются друг от друга не столько по своему генетическому составу или спектру экспрессируемых генов, сколько по «характеру использования» идентичных или, по крайней мере, гомологичных генетических факторов и сигнальных путей. Это еще раз свидетельствует о наличии собственных (негенетических) законов морфогенеза, используемых как в онтогенезе, так и в эволюции (Черданцев, 2003; Шишкин, 2006). Генетические факторы могут влиять на морфогенез лишь через посредство этих законов вырожденным

и неоднозначным образом. В качестве иллюстрации к сказанному рассмотрим следующую модель (Belousov, Grabovsky, 2007). Задано кольцо из  $C$  клеток, расположенных в один слой и оказывающих друг на друга (в результате пролиферации, тургорного давления или/и радиальных сокращений) тангенциальное давление, приводящее к их центрифугальным смещениям (рис. 2, а). Клетки сгруппированы в  $N$  кинематически независимых единиц ( $N \leq C$ ),

способных центрифугально смещаться независимо друг от друга, образуя складки.  $N$  является параметром изгибной жесткости диска, он равен числу складок и обратно пропорционален их окружной длине. Вводится также параметр упругости  $W$  (мера сопротивления радиальным смещениям клеток) (рис. 2, б). На рис. 2, в результаты моделирования представлены в пространстве параметров  $W$  и  $N$ . Видно, что это пространство подразделяется на примерно



**Рис. 2.** Моделирование форм согласно работе (Belousov, Grabovsky, 2007).

а – схема радиального сдвига ( $F_k$ ) клетки  $i$  в результате тангенциального давления на нее клеток  $i-1$  и  $i+1$ ; б – противодействующее действие упругой силы  $W$ ; в – диаграмма форм в пространстве параметров  $W$  и  $N$ . Цифры обозначают число лучей. Сплошные линии накрывают формы, устойчивые при неограниченном числе итераций, пунктир – устойчивые при ограниченном числе итераций. Зачерненные фигуры неустойчивы (непрерывно вращаются).



диагональные зоны (разделенные волнистыми линиями), занятые либо формами, выходящими после некоторого числа итераций на стационар (светлые контуры), либо не имеющими стационаров (зачерненные контуры). Разные стационарные зоны различаются между собой по порядку радиальной симметрии форм (числу лучей).

Поскольку параметры  $N$  и  $W$  принимаются постоянными для всего объекта и всего времени моделирования, их можно считать либо косвенно, либо напрямую связанными с геномом. В таком случае они могут соответствовать обширным и в целом мало специфичным группам генов, определяющим механические свойства ткани.

Данная схема – не более чем грубый набросок. Но при этом можно прийти к следующим выводам о соотношении генетических и морфогенетических факторов эволюции. Как дискретность надвидовых таксономических единиц (без чего классическая систематика была бы вообще невозможна), так и наличие закономерных рядов эволюционных преобразований (типа гомологических рядов Н.И. Вавилова) определяются в первую очередь негенетическими законами морфогенеза. С другой стороны, выбор конкретных вариантов из задаваемого этими законами множества – функция генома, действующего, скорее всего, вырожденным образом и через посредство обширных генных блоков. Дальнейшее уяснение неоднозначных и сильно нелинейных связей между морфогенезом и генетическими факторами – одна из наиболее актуальных проблем биологии ближайшего будущего.

### Литература

- Белинцев Б.Н. Физические основы биологического формообразования. М.: Наука, 1988. 254 с.
- Белоусов Л.В., Корвин-Павловская Е.Г., Лучинская Н.Н., Корникова Е.С. Роль коллективных клеточных движений и механо-геометрических условий в разметке осевых зачатков у зародышей шпорцевой лягушки // *Онтогенез*. 2007. Т. 38. № 3. С. 192–204.
- Белоусов Л.В., Миттенталь Дж. Гипервосстановление механических напряжений как возможный движущий механизм морфогенеза // *Журн. общ. биологии*. 1992. Т. 53. С. 797–807.
- Капра Ф. Паутина жизни. Новое научное понимание живых систем. «София» ИД «Гелиос». 2002. 336 с.
- Системная компьютерная биология / Под ред. Н.А. Колчанов, С.С. Гончаров, В.А. Лихошвай, В.А. Иванисенко. Новосибирск: Изд-во СО РАН, 2008. 762 с.
- Чайковский Ю.В. Активный связанный мир. Опыт теории эволюции жизни. М.: КМК, 2008. 726 с.
- Черданцев В.Г. Морфогенез и эволюция. М.: КМК, 2003. 360 с.
- Шишкин М.А. Индивидуальное развитие и уроки эволюционизма // *Онтогенез*. 2006. Т. 37. С. 179–198.
- Belousov L.V. Mechanically based generative laws of morphogenesis // *Physical Biol.* 2008. V. 5. 015009.
- Belousov L.V., Grabovsky V.I. Information about a form // *BioSystems*. 2007. V. 87. P. 204–214.
- Belousov L.V., Louchinskaia N.N., Ermakov A.S., Glagoleva N.S. Gastrulation in amphibian embryos, regarded as a succession of biomechanical feedback events // *Int. J. Dev. Biol.* 2006. V. 50. P. 113–122.
- Dahl K.N., Ribeiro A.J.S., Lammerding J.N. Nuclear shape, mechanics and mechanotransduction // *Circ. Res.* 2008. V. 102. P. 1307–1318.
- Engler A.J., Sen S., Sweeney H.L., Discher D.E. Matrix elasticity directs stem cell lineage specification // *Cell*. 2006. V. 126. P. 677–689.
- Farge E. Mechanical induction of Twist in the *Drosophila* foregut/stomodaeal primordium // *Curr. Biol.* 2003. V. 13. P. 1365–1377.
- Harris A.K., Stopak D., Warner P. Generation of spatially periodic patterns by a mechanical instability: a mechanical alternative to the Turing model // *J. Embryol. Exptl Morphol.* 1984. V. 80. P. 1–20.
- Huang S., Ingber D.E. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks // *Exptl Cell Res.* 2000. V. 261. P. 91–103.
- Meinhardt H. Models of biological pattern formation. N.Y., L.: Academic Press, 1982.
- Nelson C.M., Jean R.P., Tan J.L. *et al.* Emergent patterns of growth controlled by multicellular form and mechanics // *PNAS*. 2005. V. 102. P. 11594–11599.
- Odell G.M., Oster G., Alberch P., Burnside B. The mechanical basis of morphogenesis. I. Epithelial folding and invagination // *Developm. Biol.* 1981. V. 85. P. 446–462.
- Wray G.A., Abouheif E. When is homology not homology? // *Curr. Opin. Genet. Devel.* 1998. V. 8. P. 675–680.

**MORPHOGENESIS AND MORPHOMECHANICS****L.V. Belousov**

Faculty of Biology, Moscow State University, Moscow, Russia,  
e-mail: morphogenesis@yandex.ru

**Summary**

So far as morphogenesis belongs to self-organizing processes, it should be based upon non-linear feedbacks between its dynamic components. We suggest a scheme of the feedbacks between the passive (imposed outside) and active (generated within given parts) mechanical stresses in embryonic tissues, which may play a role of the driving forces of development. We discuss the relations between genetic factors and morphogenetic processes both during individual development and evolution.



## ВЫЯВЛЕНИЕ НОВЫХ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ SREBP В ПРОМОТОРНЫХ РАЙОНАХ ГЕНОВ ПОЗВОНОЧНЫХ НА ОСНОВЕ КОМБИНАЦИИ БИОИНФОРМАТИЧЕСКОГО И ЭКСПЕРИМЕНТАЛЬНОГО ПОДХОДОВ

Е.В. Игнатьева<sup>1,2</sup>, Т.И. Меркулова<sup>1,2</sup>, Д.Ю. Ощепков<sup>1,2</sup>, Н.В. Климова<sup>1</sup>,  
Г.В. Васильев<sup>1</sup>, И.И. Турнаев<sup>1</sup>, В.Ф. Кобзев<sup>1</sup>, Н.А. Колчанов<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики Сибирского  
отделения РАН, Новосибирск, Россия, e-mail: eignat@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Активность транскрипционных факторов семейства SREBP (Sterol Regulatory Element Binding Protein) зависит от внутриклеточного уровня холестерина. Для формирования целостной картины влияния повышенного уровня холестерина на клеточные функции необходима идентификация возможно большего числа генов, регулируемых факторами SREBP. На основе метода SITECON нами выявлены консервативные конформационные и физико-химические особенности сайтов связывания SREBP (SRE типа) и разработан компьютерный метод распознавания SRE. Выявлены новые потенциальные SRE в регуляторных районах генов липидного метаболизма и клеточного цикла позвоночных. Проведенная затем экспериментальная проверка *in vitro* способности предсказанных сайтов специфически связываться с белком SREBP1-а, который эффективно взаимодействует с SRE, продемонстрировала высокую точность метода. Таким образом, выявлен ряд новых сайтов связывания SREBP, которые с высокой степенью вероятности окажутся функциональными.

**Ключевые слова:** SREBP, сайты связывания транскрипционных факторов, методы распознавания сайтов, экспериментальная проверка.

### Введение

Транскрипция генов эукариот осуществляется при участии транскрипционных факторов, специфически взаимодействующих с короткими (5–20 п.о.) участками ДНК в регуляторных районах генов – сайтами связывания транскрипционных факторов (ССТФ). Каждый фактор обеспечивает координированную активацию или подавление транскрипции большого числа генов, определяя адекватный уровень их экспрессии в ответ на различные сигналы внешней и внутренней среды, а также ее тканеспецифичность.

Транскрипционные факторы семейства SREBP функционируют в составе холестеринного сенсора, обеспечивая регуляцию экспрессии генов в зависимости от уровня холес-

терина в клетке (Gimpl *et al.*, 2002). Семейство SREBP включает белки SREBP-1a, SREBP-1c, SREBP-2. Они принадлежат к классу спираль–петля–спираль/лейциновая застежка (basic Helix–loop–helix/leucine zipper (bHLH-ZIP)) повсеместно экспрессирующихся факторов. N-терминальный домен белков SREBP (~ 480 аминокислотных остатков) соответствует зрелой активной форме факторов и включает ДНК-связывающий основной район (basic region) и димеризационный HLH-Zip район (Wang *et al.*, 1993). Факторы семейства SREBP синтезируются в форме неактивных предшественников preSREBP с молекулярной массой ~ 125 kd и присутствуют в клетке, будучи прикрепленными к мембранам эндоплазматического ретикула (Yokoyama *et al.*, 1993). Образование активных транскрипционных факторов SREBP

(с мол. массой ~ 68 kd) осуществляется путем расщепления предшественников preSREBP в аппарате Гольджи. Транспорт preSREBP в аппарат Гольджи усиливается при низком уровне холестерина и замедляется при его высоком содержании в клетке.

Факторы SREBP влияют на транскрипцию генов, взаимодействуя с сайтами связывания двух типов: SRE (Sterol Regulatory Element) и E-боксы. Сайты E-боксы типа (ATCACGTGA) соответствуют инвертированному повтору тетрауклеотида TCAC. SRE (ATCACCCAC) включают несовершенный прямой повтор 5'-ATCAC-3' и 5'-CCCAC-3' (Kim *et al.*, 1995). Показано, что способность к взаимодействию с сайтами SRE типа является уникальной только для SREBP. Напротив, последовательности типа E-боксы являются сайтами связывания как для SREBP, так и для ряда других, родственных им, факторов семейства bHLH, экспрессирующихся в широком круге тканей (USF, TFE3, TFEB и др.) (Kim *et al.*, 1995; Hu *et al.*, 2005). По этой причине только сайты связывания SRE типа способны эффективно опосредовать регуляцию транскрипции генов в ответ на изменение уровня холестерина (Athanikar *et al.*, 1998).

Известно, что факторы SREBP регулируют транскрипцию многих генов системы липидного метаболизма: ферментов пути биосинтеза холестерина и жирных кислот, белков, обеспечивающих транспорт липидов и гидрофобных молекул (ApoA2, CETP, STAR), транскрипционных факторов (SREBP2, PPAR $\gamma$ ). В числе SREBP регулируемых генов имеются также гены из систем клеточного цикла и апоптоза: *CDKN1A*, кодирующий циклин-зависимый ингибитор киназы человека p21 WAF1/CIP1, задерживающий клеточный цикл на стадии G1 (Inoue *et al.*, 2005), и *CASP2*, кодирующий апоптоз-зависимую цистеиновую пептидазу человека (Logette *et al.*, 2005). Последние два примера указывают на то, что регуляторная роль SREBP может выходить за рамки системы липидного метаболизма. Необходимость регуляции экспрессии генов клеточного цикла в зависимости от внутри-клеточного уровня холестерина определяется тем, что рост и деление клеток сопровождается построением клеточной мембраны, в состав которой входит холестерин. Очевидно, что при недостатке холестерина пос-

троение клеточной мембраны будет затруднено, а следовательно, и клеточный цикл должен осуществляться медленнее.

Целью настоящей работы является создание эффективного компьютерного метода распознавания сайтов связывания транскрипционных факторов SREBP и поиск новых сайтов связывания SREBP в генах позвоночных. Для ее достижения на первом этапе на основе данных базы TRRD (Kolchanov *et al.*, 2002) была сформирована выборка экспериментально подтвержденных сайтов связывания SREBP. С использованием компьютерного метода SITECON (Oshchepkov, 2004a) выявлены консервативные конформационные и физико-химические особенности сайтов SRE типа и разработан компьютерный метод распознавания SRE. На следующем этапе с использованием метода SITECON были выявлены новые потенциальные SRE в регуляторных районах генов липидного метаболизма и клеточного цикла. Проведенная затем экспериментальная проверка способности предсказанных сайтов специфически связываться с изоформой SREBP1-а, которая эффективно взаимодействует с SRE (Amemiya-Kudo, 2002), продемонстрировала высокую эффективность метода. Результаты экспериментов *in vitro*, а также наличие сайтов SREBP1 в регуляторных районах генов-ортологов указывают на то, что ряд выявленных нами сайтов с высокой долей вероятности окажутся функциональными.

## Материалы и методы

Выборка сайтов связывания SREBP SRE типа (38 последовательностей) была сформирована с использованием информации из базы TRRD (<http://www.bionet.nsc.ru/trrd/>), в рамках которой имеется тематический раздел LM-TRRD, посвященный регуляции транскрипции генов липидного метаболизма (Kolchanov *et al.*, 2002). В состав выборки были включены последовательности сайтов связывания SREBP длиной 38 нуклеотидов, для которых имелись строгие экспериментальные подтверждения связывания с рассматриваемым белком. Поиск потенциальных SRE осуществляли в 5'-фланкирующих областях 46 генов из системы липидного метаболизма из базы LM-TRRD, для которых

на момент составления выборки регуляция факторами SREBP еще не была известна. Для 38 генов в анализ были взяты участки промоторных районов  $-1000$  до  $+100$  нуклеотидов относительно старта транскрипции, а для 8 генов – районы  $-600/+100$ <sup>1</sup>. Нами также были проанализированы промоторные ( $-1000/+100$ ) районы 20 генов человека из базы данных TRRD, продукты которых участвуют в регуляции клеточного цикла.

Для распознавания потенциальных сайтов использовался метод SITECON, основанный на анализе консервативных конформационных и физико-химических свойств в позициях выборки ССТФ. Подробно метод описан ранее (Oshchepkov, 2004a, b).

Экспериментальное исследование связывающей способности потенциальных SRE осуществлялось методом задержки комплекса ДНК/белок в геле (EMSA) с использованием рекомбинантного SREBP-1а человека (Kolchanov *et al.*, 2007).

## Результаты

**Консервативные конформационные и физико-химические особенности сайтов SRE.** В результате анализа выборки сайтов связывания SRE типа был обнаружен ряд особенностей конформационных и физико-химических свойств ДНК, три из которых представляются наиболее интересными (рис. 1). Оказалось, что в районе сайта ширина большой и ширина малой бороздки значимо консервативны, при этом ширина большой бороздки (рис. 1, а) уменьшена, а ширина малой бороздки (рис. 1, б) увеличена по сравнению со случайными последовательностями. Кроме того, во фланкирующих районах обоих тринуклеотидов САС консенсусной последовательности наблюдаются пониженные значения жесткости к изгибу в сторону большой бороздки (рис. 1, в) по сравнению со случайными последовательностями. Выявленные особенности находятся в согласии с данными рентгеноструктурного анализа (РСА) комплекса гомодимера SREBP1-а с сайтом SRE

<sup>1</sup> Длина 5'-фланкирующих участков генов, взятых в анализ, определялась наличием последовательностей необходимой длины в карточках базы нуклеотидных последовательностей EMBL.

типа (Párraga *et al.*, 1998). Согласно этим данным, структура ДНК сайта в комплексе с белком мало отличается от В-формы, за исключением двух параметров: средняя величина ширины большой бороздки в районе сайта увеличена, а ширина малой бороздки уменьшена относительно значений этих свойств для В-формы ДНК. Кроме того, в центральных 4 парах ДНК (соответствующих нуклеотидам СССС консенсусной последовательности) не наблюдалось изгиба, в то время как по обоим концам сайта наблюдался изгиб в сторону большой бороздки. Как видно, эти данные хорошо совпадают с особенностями свойств ДНК для сайта SRE, предсказанными при использовании технологии SITECON.

## Характеристика точности метода распознавания SRE на основе технологии SITECON.

Нами было проведено сравнение точности предсказания SRE для метода SITECON и метода весовых матриц. Метод весовых матриц был запрограммирован на основе работы Gershenson *et al.* (2005). Расчет ошибок производился идентично для обоих подходов на основе выборки SRE, описанной выше. Ошибка I рода рассчитывалась с помощью подхода Jack-knife (Efron, Gong, 1983). Эта процедура включала в себя распознавание одного из сайтов из выборки SRE на основе обучения по неполной выборке, не включающей распознаваемый сайт, такая процедура проводилась последовательно для всех сайтов выборки. Для расчета ошибки II рода обоих методов использовалась одна и та же случайная последовательность длиной  $1 \times 10^6$  п.о., сгенерированная путем многократного случайного перемешивания нуклеотидов обучающей выборки. Из графика рис. 2 видно, что метод SITECON и метод весовых матриц показывают сходное качество распознавания. При этом для значений ошибки II рода, пригодных для геномных исследований (менее чем  $10^{-3}$ ), до критических значений ошибки I рода (более 90 %) качество распознавания методом SITECON даже незначительно превосходит качество метода весовых матриц. Для выбранного нами порога уровня конформационного сходства ошибка I рода метода SITECON составляет 0,53. При заданном уровне ошибки I рода (0,53) метод весовых матриц дает ошибку II рода  $9,90 \times 10^{-4}$ . В то же время метод SITECON позволяет получить меньшую ошибку:  $7,55 \times 10^{-4}$ .

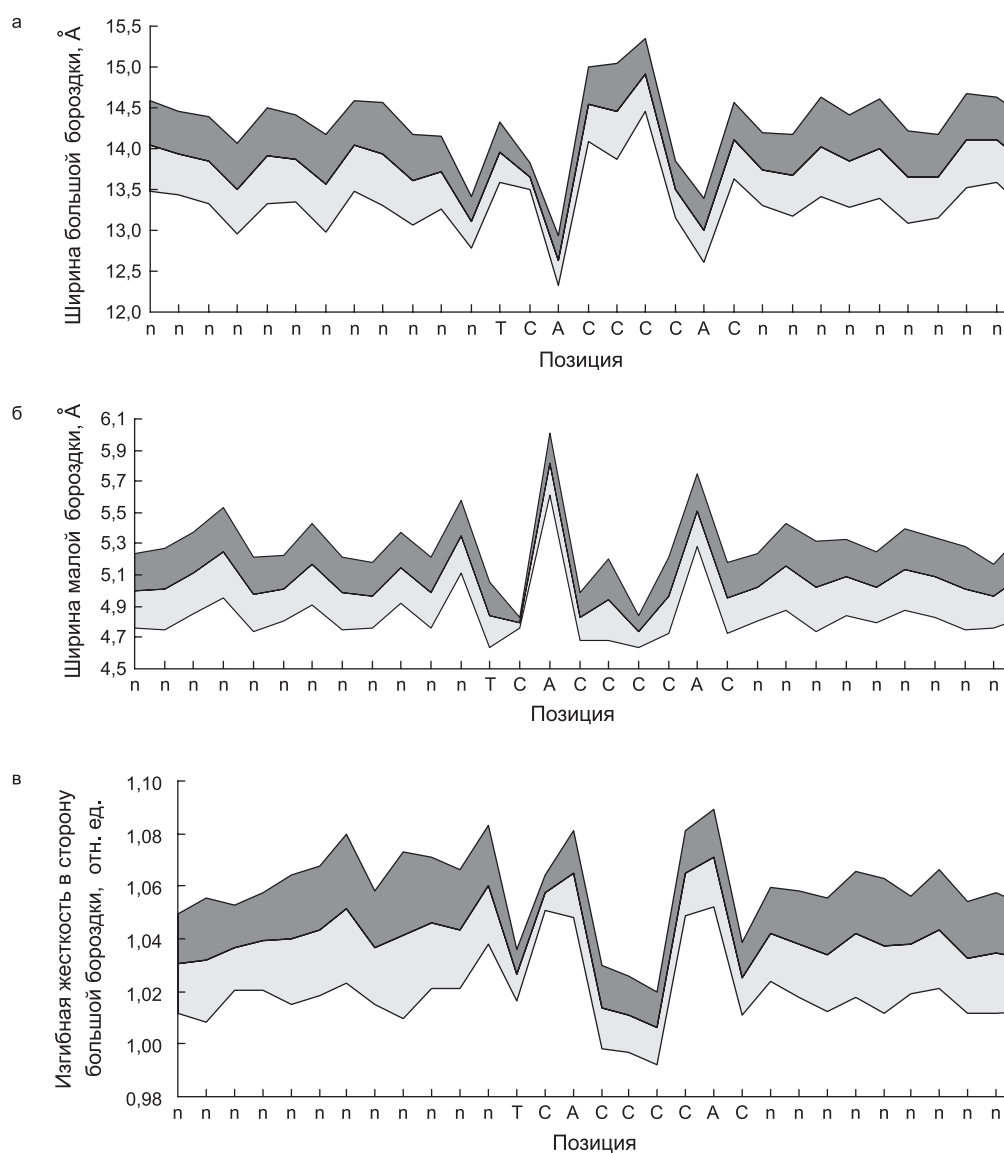


Рис. 1. Значения конформационных и физико-химических свойств ДНК и их консервативность, рассчитанные для выборки выровненных сайтов SRE типа.

а – ширина большой бороздки; б – ширина малой бороздки; в – изгибная жесткость ДНК в сторону большой бороздки. Центральная кривая соответствует среднему значению свойства для каждой из позиций выборки; верхняя кривая соответствует среднему значению плюс значение стандартного отклонения для позиции; нижняя кривая соответствует среднему значению минус значение стандартного отклонения для позиции. Участок, соответствующий сайту SRE, обозначен по оси OX его консенсусной последовательностью TCACCCAC.

**Выявление новых SRE в регуляторных районах генов позвоночных и экспериментальное исследование связывающей способности новых SRE.** При анализе 5'-фланкирующих областей генов позвоночных методом SITECON выявлены потенциальные SRE. Сайты выявлены как на прямой, так и на обратной цепи ДНК и располагаются на разном удалении

от старта транскрипции. Часть потенциальных SRE, выявленных в генах липидного метаболизма, представлена в табл. 1. Выявлены также SRE в 15 генах из системы регуляции клеточного цикла человека: *CDKN2A*, *UNG*, *RBL2*, *GADD45G*, *CCNA1*, *TGFB1*, *BIRC5*, *CDKN1A*, *CCND1*, *MYB*, *JUN*, *RB1*, *CCNE1*, *E2F1*, *GADD45A*. Гены, содержащие потенциальные

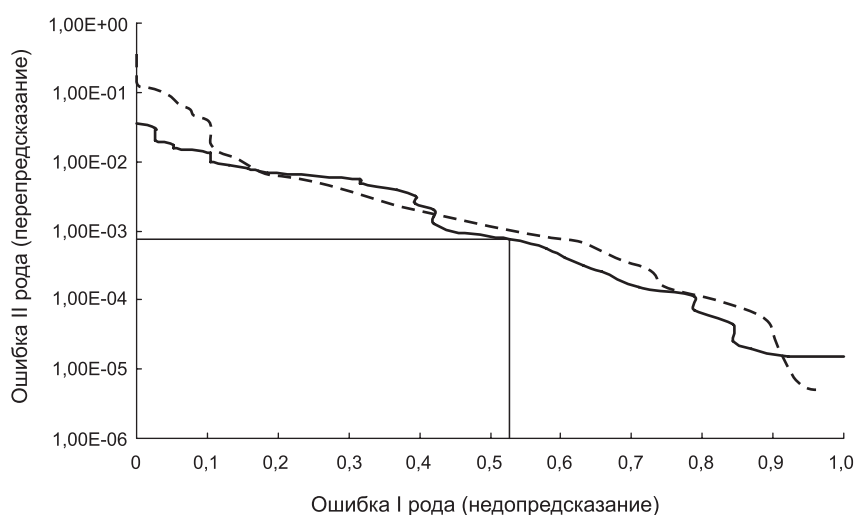


Рис. 2. Сравнение ошибок I и II рода для методов SITECON (сплошная линия) и метода весовых матриц (пунктир).

Линиями обозначены значения ошибок I и II рода ( $0,53$  и  $7,55 \times 10^{-4}$  соответственно) для порога, на котором проводилось предсказание методом SITECON.

SRE, кодируют белки с различными функциями: транскрипционные факторы (LXRA, E2F, JUN), ферменты (FAS, CPT2, LPL), транспортные белки (MSR, IBABP) и др.

Выборочной экспериментальной проверке *in vitro* подверглись 30 сайтов, распознанных в генах липидного метаболизма (табл. 1). Типичные результаты, полученные методом задержки

в геле, представлены на рис. 3. На дорожке 2, соответствующей ДНК-белковому комплексу, включавшему консенсусную последовательность SRE, выявлено 2 полосы задержки, что согласуется с данными других авторов (Moop, 2000). ДНК-белковые комплексы с такой же подвижностью выявлены и на дорожках 4–14 и 17–26.

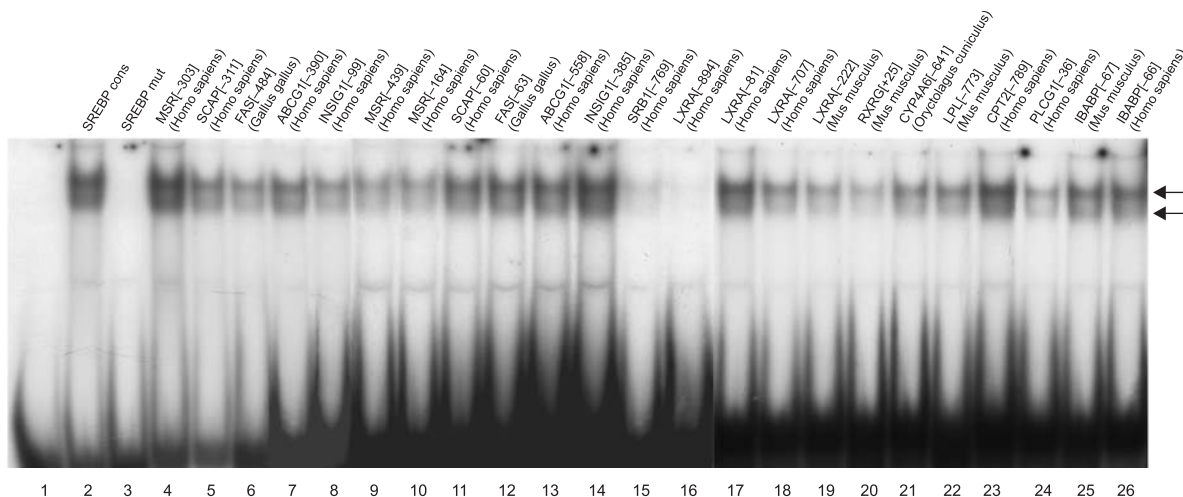


Рис. 3. Экспериментальное подтверждение SRE, предсказанных методом SITECON.

Инкубация олигонуклеотидной пробы (1 нг), меченной  $^{32}P$ , в отсутствие белка (1-я дорожка) либо с белком SREBP-1a (0,5 мкг) (дорожки 2–26). SREBP cons – олигонуклеотид, включавший консенсусную последовательность, SREBPmut – олигонуклеотид, включавший консенсусную последовательность с мутацией. Сверху указаны названия генов, виды организмов и позиции сайтов в генах относительно старта транскрипции. Стрелками показаны комплексы ДНК с белком SREBP-1a, доказывающие связывающую способность предсказанных сайтов.



Таблица 1

Потенциальные SRE, предсказанные методом SITECON в промоторах генов системы липидного метаболизма, и результаты их экспериментальной проверки

№	Название гена (вид)	Позиция*	Ориентация	R**	Последовательность ССТФ	Экспериментальное подтверждение
а) потенциальные SRE, имеющие уровень конформационного сходства более 73 %						
1	<i>MSR</i> (человек)	-439	+	0,750	GGACCATGAGGTCAGGGGATCAAGACCATC	+
2	<i>MSR</i> (человек)	-303	+	0,778	GAGATCACACCACTGCACTCCACCCTGGTG	+
3	<i>MSR</i> (человек)	-164	-	0,796	AGATAACTCACTCTTACCCCCATTTAGGCC	+
4	<i>SCAP</i> (человек)	-311	+	0,777	TCTCCCCCGTCAACACCAGGGGCCCGCGGA	+
5	<i>FAS</i> (цыпленок)	-484	+	0,756	GCCTAGCAACACCCTCACGGGCTCCGCATT	+
6	<i>FAS</i> (цыпленок)	-63	+	0,749	GGAGTAAACCGTCAGCCCATGTGGTGGCCG	+
7	<i>ABCG1</i> (человек)	-558	-	0,732	CCCTTCCCGCCCCACCCACCCCGACGGTGT	+
8	<i>INSIG1</i> (человек)	-385	+	0,736	GGCCCGAAAAAATCACCCAAGGATGCAAGC	+
9	<i>INSIG1</i> (человек)	-99	+	0,769	GCGGGGCAAGCTCAGGCCACGCCCTGGGC	+
10	<i>LXRA</i> (человек)	-81	-	0,780	CTGCCGAGGGGTCAGGCCACCCCTCCTTCC	+
11	<i>CYP4A6</i> (кролик)	-641	-	0,783	ACAAGCATTTCACCCCACTGCCCTCAACT	+
12	<i>LPL</i> (мышь)	-773	+	0,805	CCACACTCTAGCACCCCATATCCCCTCGC	+
13	<i>CPT2</i> (человек)	-789	+	0,731	CAATCCGAGGTCACCTAAGATAACAAA	+
14	<i>IBABP</i> (мышь)	-67	-	0,863	AGCCCCAAAGTCACCCCACTTCTCAGTCCT	+
15	<i>IBABP</i> (человек)	-66	-	0,881	CAGCCCCAAGTCACCCCACTTCTTCTCCC	+
б) потенциальные SRE, имеющие уровень конформационного сходства менее 73 %						
16	<i>LXRA</i> (мышь)	-222	+	0,729	TCTCTGTCCCCTCAGGGCATCCACACACAA	+
17	<i>PLCG1</i> (человек)	-36	+	0,728	GGGGTGGGGCCTCAGGGCAGCCCCGCCCG	+
18	<i>RXRG</i> (мышь)	+25	+	0,727	GGAAGAGTCGCTCACAGCAGCCCAACCCAG	+
19	<i>SRBI</i> (человек)	-769	+	0,725	GGGGGCCCGCCTCACGGGTTGGCATCCCAG	-
20	<i>MSR</i> (человек)	-323	-	0,704	AGTGGTGTGATCTCAGCTCACTGCAACCTC	+
21	<i>LXRA</i> (человек)	-707	-	0,717	CAACTCAGGTGGCAGCCCAGGATGGACCTG	+
22	<i>LXRA</i> (человек)	-894	-	0,717	CCAGCCCTGTTTCAGGGGTGTTGGGGGAGA	-
23	<i>FAS</i> (цыпленок)	-429	-	0,716	AGTGCCGGGCGGCACCCAACGCGGGGGCG	-
24	<i>CAVI</i> (мышь)	-239	-	0,715	GACCCACAACCTCACCTGCTCCGTGGTCC	-
25	<i>SIP</i> (человек)	-101	+	0,713	GAACACGCCTGGGCACTCCATTCGGGGCTG	-
26	<i>ABCG1</i> (человек)	-390	+	0,712	CCCCGCAGCCTCACCCGTGCTGGCCCAGCC	+
27	<i>INSIG1</i> (человек)	-47	-	0,712	GCGGGGCCCGCATCACGTGAGCCTGCCCGG	-
28	<i>FAS</i> (цыпленок)	-129	+	0,705	CAGGGGCTGGCATCACCCGGGGCCGGCTC	+
29	<i>CAVI</i> (мышь)	-501	-	0,699	TGCCAAGACCTGTCACCCCCACCCCATGC	-
30	<i>SCAP</i> (человек)	-60	+	0,694	CAGGCGCCCTGAGCACCCGCGTGACGGTAC	+

\* Позиция указана относительно старта транскрипции; \*\* уровень конформационного сходства с известными сайтами связывания SREBP, оцененный методом SITECON.

Уровень конформационного сходства с ССТФ из обучающей выборки варьировал от 69 до 88,1 %. Способность взаимодействовать с белком SREBP-1a подтверждена для 23 из 30 сайтов. При этом все 15 сайтов, обладающих уровнем конформационного сходства более чем 73 %, оказались способными к связыванию с белком SREBP-1a *in vitro* (табл. 1, а). Только 8 из 15 потенциальных сайтов с уровнем конформационного сходства менее чем 73 % оказались способными к связыванию с белком SREBP-1a *in vitro* (табл. 1, б). Следует отметить тот факт, что 4 из 5 сайтов с уровнем конформационного сходства от 72 до 73 % оказались способными к связыванию с белком SREBP-1a *in vitro*. Этот результат подтверждает тот факт, что показатель «уровень конформационного сходства» адекватно отражает потенциальную способность предсказанных сайтов к взаимодействию с факторами SREBPs.

### Обсуждение

Во всех известных к настоящему времени работах выявление потенциальных сайтов связывания SREBP осуществляется на основе консенсуса или методом весовых матриц с использованием данных из базы TRANSFAC. Однако достоверных данных об эффективности этих подходов до настоящего времени нет, поскольку для их получения необходима массовая экспериментальная проверка связывающей способности потенциальных сайтов. Настоящая работа и была направлена на решение подобной задачи.

На первом этапе сформирована выборка геномных последовательностей SRE, превышающая по объему представленные в литературе компиляции. Для распознавания SRE был разработан метод SITECON, в основе которого лежат анализ и выявление консервативных контекстно-зависимых конформационных и физико-химических свойств спирали ДНК в районе сайтов связывания (Oshcherkov *et al.*, 2004a). Данный подход с успехом применялся ранее для анализа и распознавания сайтов связывания E2F и SF1 (Oshcherkov *et al.*, 2004b; Игнатъева и др., 2007). Настоящее исследование вновь подтвердило эффективность технологии SITECON. С помощью метода SITECON для сайтов SRE был выявлен ряд свойств ДНК, которые со-

гласуются с данными рентгеноструктурного анализа ДНК-белкового комплекса этого сайта с фактором SREBP-1a (Párraga *et al.*, 1998). Экспериментальная проверка потенциальных сайтов, предсказанных методом SITECON, продемонстрировала высокую эффективность метода. Все 15 сайтов с уровнем конформационного сходства выше 73 % обладали высокой связывающей активностью по отношению к белку SREBP-1a в условиях *in vitro*.

В исследовании связывающей способности потенциальных сайтов с уровнем конформационного сходства ниже выбранного нами порога 73 % (от 69 до 72,7 %) выявлено (табл. 1, б), что только 53 % потенциальных сайтов этой группы оказались способными к связыванию с белком SREBP-1a *in vitro*. Таким образом, преимуществом метода SITECON является не только новое знание о биологической модели взаимодействия ДНК–белок, но и то, что показатель «уровень конформационного сходства» адекватно отражает потенциальную способность предсказанных сайтов к взаимодействию с факторами SREBP. Наличие связывающей активности всех 15 протестированных сайтов с уровнем конформационного сходства более 73 % подтверждает правильность выбранного порога распознавания для выявления высокодостоверных SRE.

Для подтверждения функциональности выявленных нами новых сайтов связывания SREBP безусловно необходимы дальнейшие эксперименты. Однако уже имеющиеся данные свидетельствуют о высокой вероятности того, что они являются функциональными. Например, роль SREBP в регуляции транскрипции генов *ABCG1* человека и *FAS* цыпленка была подтверждена экспериментально другими авторами (Gondret *et al.*, 2001; Ecker *et al.*, 2007). Кроме того, поскольку показано, что SREBP участвует в регуляции транскрипции *LPL* человека (Schoonjans *et al.*, 2000), выявленный нами сайт (в позиции –773) в одноименном гене мыши, по всей вероятности, также является функциональным. Для остальных генов, в которых предсказанные SRE продемонстрировали способность взаимодействовать с SREBP-1a *in vitro*, регуляция факторами SREBP также достаточно вероятна вследствие того, что кодируемые этими генами белки являются важными элементами системы липидного метаболизма.

Нами также выявлены сайты связывания SREBP в 15 генах клеточного цикла человека, для ряда из которых (*E2F*, *JUN*, *GADD45A* и др.) при анализе ортологичных генов у мыши и крысы также были выявлены сайты SRE. Мы считаем выявленные SRE в генах клеточного цикла чрезвычайно перспективными для дальнейшего изучения, поскольку участие SREBP в регуляции генов клеточного цикла имеет достаточно ясный биологический смысл.

Проведенное нами исследование позволило разработать метод распознавания SRE на базе технологии SITECON, идентифицировать новые потенциальные SRE в генах позвоночных, оценить эффективность метода и адекватность используемого уровня конформационного сходства на основе выборочной проверки предсказанных сайтов.

### Благодарности

Работа поддержана грантом РФФИ № 08-04-01214, проектом № 8 программы Президиума РАН № 22 «Молекулярная и клеточная биология», проектом № 29 программы Президиума РАН № 22 «Биологическое разнообразие», подпрограммой «Эволюция гео-биологических систем» программы Президиума РАН «Происхождение и эволюция биосферы», интеграционными проектами СО РАН (№ 26, 113, 119, 115, 18.13), а также НШ-2447.2008.4. Авторы выносят благодарность Т.Ф. Осборне (University of California, Irvine, CA) за предоставление плазмиды pCMV5, И.В. Лоховой за библиографическую поддержку, А.Л. Проскура за участие в аннотировании данных в базе TRRD.

### Литература

- Игнатъева Е.В., Климова Н.В., Ощепков Д.Ю. и др. Поиск новых сайтов связывания транскрипционного фактора SF1 методом SITECON: экспериментальная проверка и анализ регуляторных районов генов-ортологов // Докл. РАН. 2007. Т. 415. № 1. С. 120–124.
- Amemiya-Kudo M., Shimano H., Nasty A.H. *et al.* Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different target promoters of lipogenic and cholesterologenic genes // J. Lipid Res. 2002. V. 43. № 8. P. 1220–1235.
- Athanikar J.N., Osborne T.F. Specificity in cholesterol regulation of gene expression by coevolution of sterol regulatory DNA element and its binding protein // Proc. Natl Acad. Sci. USA. 1998. V. 95. № 9. P. 4935–4940.
- Ecker J., Langmann T., Moehle C., Schmitz G. Isomer specific effects of conjugated linoleic acid on macrophage ABCG1 transcription by a SREBP-1c dependent mechanism // Biochem. Biophys. Res. Commun. 2007. V. 352. № 3. P. 805–811.
- Efron B., Gong G. A leisure look at the bootstrap, the jackknife and cross-validation // Am. Stat. 1983. V. 37. P. 36–48.
- Gershenzon N.I., Stormo G.D., Ioshikhes I.P. Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites // Nucl. Acids Res. 2005. V. 33. № 7. P. 2290–2301.
- Gimpl G., Burger K., Fahrenholz F. A closer look at the cholesterol sensor // Trends Biochem. Sci. 2002. V. 27. № 12. P. 596–569.
- Gondret F., Ferré P., Dugail I. ADD-1/SREBP-1 is a major determinant of tissue differential lipogenic capacity in mammalian and avian species // J. Lipid Res. 2001. V. 42. № 1. P. 106–113.
- Hu J., Banerjee A., Goss D.J. Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: importance of protein-protein and protein-DNA interactions // Biochemistry. 2005. V. 44. № 35. P. 11855–11863.
- Inoue N., Shimano H., Nakakuki M. *et al.* Lipid synthetic transcription factor SREBP-1a activates p21WAF1/CIP1, a universal cyclin-dependent kinase inhibitor // Mol. Cell. Biol. 2005. V. 25. № 20. P. 8938–8947.
- Kim J.B., Spotts G.D., Halvorsen Y.D. *et al.* Dual DNA binding specificity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain // Mol. Cell. Biol. 1995. V. 15. P. 2582–2588.
- Kolchanov N.A., Ignatieva E.V., Ananko E.A. *et al.* Transcription Regulatory Regions Database (TRRD): its status in 2002 // Nucl. Acids Res. 2002. V. 30. № 1. P. 312–317.
- Kolchanov N.A., Merkulova T.I., Ignatieva E.V. *et al.* Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes // Briefings in Bioinformatics. 2007. V. 8. № 4. P. 266–274.
- Logette E., Solary E., Corcos L. Identification of a functional DNA binding site for the SREBP-1c transcription factor in the first intron of the human caspase-2 gene // Biochim. Biophys. Acta. 2005. V. 1738. № 1/3. P. 1–5.
- Moon Y.A., Lee J.J., Park S.W. *et al.* The roles of sterol regulatory element-binding proteins in the transactivation of the rat ATP citrate-lyase promoter // J. Biol. Chem. 2000. V. 275. № 39.

- P. 30280–30286.
- Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A. *et al.* SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition // Nucl. Acids Res. 2004a. V. 32. P. W208–W212.
- Oshchepkov D. Yu., Turnaev I.I., Pozdnyakov M.A. *et al.* SITECON—a tool for analysis of DNA physicochemical and conformational properties: E2F/DP transcription factor binding site analysis and recognition // Bioinformatics of Genome Regulation and Structure / Ed. N. Kolchanov, R. Hofstaedt. Boston/Dordrecht/London: Kluwer Academic Publ., 2004b. P. 93–102.
- Párraga A., Bellolell L., Ferré-D’Amaré A.R., Burley S.K. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution // Structure. 1998. V. 6. № 5. P. 661–672.
- Schoonjans K., Gelman L., Haby C. *et al.* Induction of LPL gene expression by sterols is mediated by a sterol regulatory element and is independent of the presence of multiple E boxes // J. Mol. Biol. 2000. V. 304. № 3. P. 323–334.
- Wang X., Briggs M.R., Hua X. *et al.* Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter // J. Biol. Chem. 1993. V. 268. P. 14497–14504.
- Yokoyama C., Wang X., Briggs M.R. *et al.* SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene // Cell. 1993. V. 75. № 1. P. 187–197.

**PREDICTION OF NEW SREBP BINDING SITES  
IN THE PROMOTER REGIONS OF VERTEBRATE GENES  
USING A COMBINATION OF A BIOINFORMATIC  
AND AN EXPERIMENTAL APPROACH**

**E.V. Ignatieva<sup>1,2</sup>, T.I. Merkulova<sup>1,2</sup>, D.Y. Oshchepkov<sup>1,2</sup>, N.V. Klimova<sup>1</sup>, G.V. Vasiliev<sup>1</sup>,  
I.I. Turnaev<sup>1</sup>, V.F. Kobzev<sup>1</sup>, N.A. Kolchanov<sup>1,2</sup>**

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: eignat@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

**Summary**

The activity of transcription factors in the family of sterol regulatory element binding proteins (SREBPs) is regulated by cellular cholesterol content. To have better insight into the implications of elevated cholesterol for cellular functions, it is critical to identify as many genes that SREBPs regulate, as possible. Using SITECON method, we have determined the conserved conformational, physical and chemical properties of SREs, which are SREBP binding sites, and developed a computer method for SRE prediction. New potential SREs in the regulatory regions of the vertebrate genes involved in control of lipid metabolism and cell cycle have been predicted. An *in vitro* validation of the ability of potential SREs to specifically bind to SREBP1-a – an isoform, which binds efficiently to SREs – showed a high accuracy of SITECON-based recognition. In summary, new SREBP binding sites have been found, which will most likely be functional.

## ВЫЯВЛЕНИЕ НОВЫХ DRE В РЕГУЛЯТОРНОЙ ОБЛАСТИ ГЕНОВ ЧЕЛОВЕКА, КОДИРУЮЩИХ КОМПОНЕНТЫ ЦИТОЗОЛЬНОГО КОМПЛЕКСА АРИЛ-ГИДРОКАРБОНОВОГО РЕЦЕПТОРА

Д.Ю. Ощепков<sup>1</sup>, Д.П. Фурман<sup>1,2</sup>, Е.А. Ощепкова<sup>1</sup>, А.В. Катохин<sup>1</sup>,  
М.Ю. Шаманина<sup>1</sup>, В.А. Мордвинов<sup>1</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Арил-гидрокарбонный рецептор (AhR) – активируемый лигандом транскрипционный фактор, который участвует в ответе на ксенобиотики или эндогенные сигналы, модулируя экспрессию многих генов в различных тканях у разных видов. В отсутствие лиганда AhR находится в цитоплазме клетки в составе гетеротетрамерного корового цитоплазматического комплекса. В составе этого же комплекса он взаимодействует с лигандом и транслоцируется в ядро. В ядре связанный с лигандом AhR образует с белком Arnt транскрипционно активный комплекс, который обладает высоким сродством к DRE (dioxin response elements). В работе представлены результаты поиска DRE в генах, кодирующих белки цитозольного комплекса арил-гидрокарбонного рецептора. С использованием программы SITECON идентифицировано 27 новых потенциальных DRE в пяти последовательностях из регуляторных областей соответствующих генов человека. Вновь выявленные DRE позиционируются в промоторных районах генов, что позволяет предположить функциональную значимость выявленных DRE. Полученные данные могут использоваться для разработки моделей AhR-зависимой регуляции экспрессии генов, кодирующих белки цитозольного комплекса.

**Ключевые слова:** арил-гидрокарбонный рецептор, DRE, диоксин, коровый цитозольный комплекс, авторегуляторная петля.

### Введение

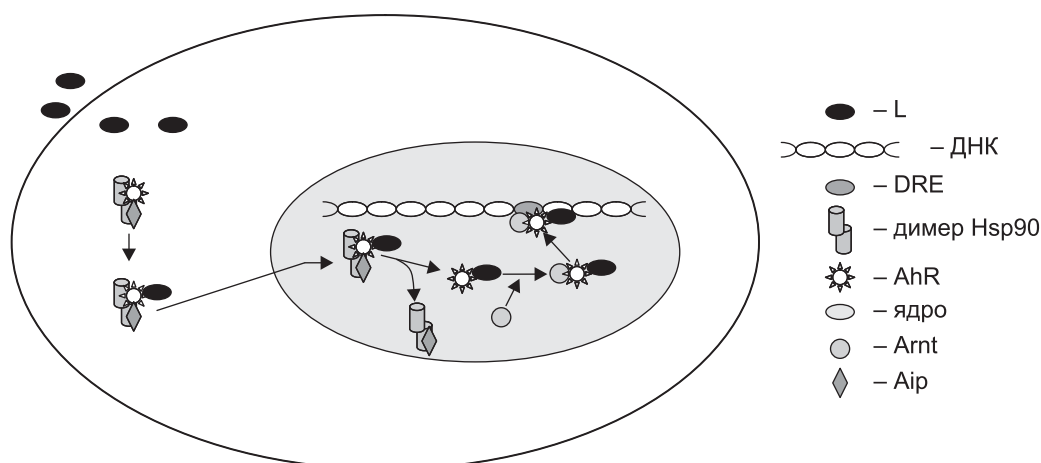
Несмотря на повышенный интерес исследователей к эффектам действия диоксина и его производных на организм млекопитающих и человека, механизмы его участия в регуляции экспрессии генов эукариот изучены недостаточно. В частности до сих пор неполны данные о наличии в составе регуляторных районов генов, кодирующих компоненты цитозольного комплекса арил-гидрокарбонного рецептора, «диоксин-чувствительных элементов» (dioxin response elements, DRE), ответственных за инициацию их транскрипции.

Арил-гидрокарбонный рецептор (AhR) – активируемый лигандом (L) транскрипционный фактор, который участвует в ряде критических клеточных событий в ответ на ксенобиотики или эндогенные сигналы (Harper *et al.*, 2006), моду-

лируя экспрессию многих генов в различных тканях у разных видов (Sogawa, Fujii-Kuriyama, 1997). В отсутствие лиганда AhR находится в составе гетеротетрамерного корового цитоплазматического комплекса (КЦК), состоящего из AhR (собственно лиганд-связывающая субъединица), иммунофилин-подобного белка Air и димера Hsp90, с которым менее тесно ассоциирован ряд других белков (Petruilis, Perdew, 2002) (рис. 1).

При связывании лиганда, примером которого может служить 2,3,7,8-тетрахлородibenзо-п-диоксин (2,3,7,8-ТХДД) – наиболее токсичный из известных до настоящего времени ксенобиотиков, AhR/L в составе комплекса транслоцируется в ядро. В ядре КЦК диссоциирует, и AhR/L образует транскрипционно-активный комплекс с Arnt (aryl hydrocarbon receptor nuclear translocator) (AhR/L-Arnt), который способен с высоким сродством связываться со специфи-





**Рис. 1.** Роль КЦК в AhR-опосредованном пути активации экспрессии генов, содержащих в своей регуляторной области DRE.

ческими последовательностями (DRE – dioxin response elements) в регуляторной области генов-мишеней, инициируя их транскрипцию (Whitlock, 1999; Ma, 2001).

Механизмы регуляции транскрипции генов, кодирующих собственно белки КЦК, в связи с воздействием диоксина до конца неясны. Существуют экспериментальные данные, демонстрирующие наличие AhR-опосредованной регуляции экспрессии гена *AhR* у крыс. Показано, в частности, что связывание лиганда с AhR в печени крыс повышало уровень транскрипции мРНК самого *AhR* зависимым от типа лиганда образом: так, при введении 2,3,7,8-ТХДД уровень мРНК AhR возрастал в 3 раза (Franc *et al.*, 2001; Brauze *et al.*, 2006). Ряд результатов указывает на возможное существование такой регуляции и у человека (Hayashi *et al.*, 1994; Jana *et al.*, 1998; Pitt *et al.*, 2001; Wu *et al.*, 2003). В связи с этим представляется интересным исследовать регуляторные районы генов человека, кодирующих компоненты КЦК, с целью получения более полного представления об их организации и данных для построения в дальнейшем моделей, отражающих механизмы регуляции экспрессии этих генов и динамику биосинтеза/распада комплекса.

### Материалы и методы

Поиск DRE в регуляторных районах генов проводился с помощью подхода SITECON

(<http://www.mgs.bionet.nsc.ru/mgs/programs/sitecon/>) (Oshchepkov *et al.*, 2004). Его эффективность была ранее подтверждена экспериментально при распознавании сайтов связывания транскрипционных факторов SF-1 и SREBP (Kolchanov *et al.*, 2007).

В качестве обучающей выборки использовалось 13 экспериментально выявленных DRE длиной 19 п.н. (Sun *et al.*, 2004). Все эти сайты содержат инвариантную последовательность 5'-GCGTG-3' и вариабельные фланговые последовательности. На основе выборки был определен набор консервативных конформационных и физико-химических свойств DRE, который использовался для создания правил распознавания и последующего определения уровня конформационного сходства тестируемой последовательности с обучающей выборкой. Ошибка 1-го рода (недопредсказание) рассчитывалась с помощью подхода «складного ножа» (jack-knife) (Efron, Gong, 1983). Эта процедура включала в себя распознавание одного из сайтов из выборки DRE на основе обучения по неполной выборке, не включающей распознаваемый сайт, и проводилась последовательно для всех сайтов из обучающей выборки. Для расчета ошибки 2-го рода (перепредсказание) использовалась случайная последовательность длиной 500 000 п.о., сгенерированная путем многократного случайного перемешивания нуклеотидов из последовательностей обучающей выборки. Критерием принадлежности

тестируемой последовательности к потенциальным DRE было значение уровня сходства, превышающее пороговое значение  $P = 0,95$ , при котором ошибка 1-го рода равна нулю (табл. 1), все сайты обучающей выборки безошибочно распознаются на контроле.

**Таблица 1**  
Ошибки распознавания DRE  
методом SITECON для разных уровней  
конформационного сходства

Пороговое значение уровня сходства	Ошибка 1-го рода (недопредсказание)	Ошибка 2-го рода (перепредсказание)
0,95	0,0000	1,11e-003 (1/901)
0,96	0,0769	9,10e-004 (1/1099)
0,97	0,2308	6,70e-004 (1/1492)
0,98	0,3077	1,50e-004 (1/6665)

Эффективность подхода SITECON для поиска DRE тестировалась на последовательностях генов *CYP1A1*, *CYP1B1* человека и гена эписрегулина крысы. При сравнении полученных результатов с экспериментальными данными (Kress *et al.*, 1998; Tsuchiya *et al.*, 2003; Patel *et al.*, 2006) оказалось, что с помощью SITECON выявляются все белок-связывающие и функционально-активные DRE. Следует отметить, что значение уровня сходства составило не менее 0,974 для всех этих элементов, включая и функционально активный сайт в гене эписрегулина крысы, который не был выявлен другими методами распознавания (Sun *et al.*, 2004).

### Результаты

Проанализировано 5 последовательностей из регуляторных областей генов человека, кодирующих компоненты цитозольного комплекса арил-гидрокарбонowego рецептора: *AhR*, *AIP*, *HSP90AA1*, *HSP90AB1*. Для генов *AIP*, *HSP90AA1* и *HSP90AB1* в анализ были взяты участки промоторных районов от -5000 до +600 нуклеотидов относительно старта транскрипции. Для *AhR* район поиска был расширен до -15000/+600. В результате обнаружено 30 потенциальных DRE, 27 из которых выявлены впервые: 1 в *AhR*, 5 в *AIP*, 5 в *HSP90AB1* и по 8

в регуляторной зоне гена *HSP90AA1* для двух альтернативных транскриптов. Все DRE распознаны программой SITECON с существенным превышением выбранного порога, равного 0,95. Полученные результаты суммированы в табл. 2.

Проведенный анализ промотора гена *AhR* человека показал наличие 3 потенциальных DRE (позиции +8, +406 и +696), лишь один из которых (в позиции +406) был уже выявлен (Sun *et al.*, 2004). Кроме того, потенциальный DRE был найден в позиции от -7398 до -7385, в районе, где ранее поиск DRE не проводился или проводился, но соответствующий сайт не был идентифицирован.

Поскольку установлен факт AhR-опосредованной регуляции экспрессии гена *AhR* крысы (Franc *et al.*, 2001; Brauze *et al.*, 2006), нами был проведен анализ его промоторной зоны в районе -15000/+600. В результате выявлено три потенциальных DRE с высоким уровнем конформационного сходства (0,97–0,98, табл. 2), причем позиция одного совпадает с позицией DRE, выявленного ранее (Sun *et al.*, 2004).

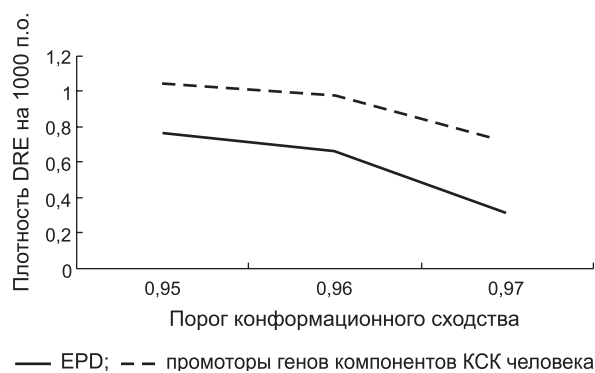
Для верификации полученных результатов было проведено сравнение плотности распознаваемых DRE (число DRE на 1000 п.о.) в выборке промоторов генов человека (1867 последовательностей) из базы данных Eukaryotic Promoter Database (Schmid *et al.*, 2006) и в промоторах перечисленных выше генов – компонентов КСК человека (5 последовательностей). Следует отметить, что анализируемая выборка из EPD включала как нецелевые для воздействия ксенобиотиков гены, так и заведомо содержащие DRE, поскольку априорную классификацию по признаку наличия/отсутствия DRE в имеющихся последовательностях провести не представляется возможным по причине недостаточной изученности степени участия соответствующих генов в ответе на действие ксенобиотиков. Распознавание в обоих случаях проводилось для участка -5000/+600 относительно старта транскрипции. Результаты представлены на рис. 2. Видно, что для порогов 0,95, 0,96 и 0,97 плотность DRE в промоторах генов компонентов КСК человека превышает их плотность в промоторах из базы данных EPD, причем для порога 0,97 более чем в 2 раза. Установленное различие позволяет с высокой степенью веро-

Таблица 2

Потенциальные DRE, выявленные с использованием метода SITECON  
в генах *AhR*, *HSP90*, *AIP* человека (Nedosekina *et al.*, 2007) и в гене *AhR* крысы.  
Сравнение с результатами Sun *et al.* (2004) и Boutros *et al.* (2004)

Гены человека	Последовательность DRE	SITECON		Sun <i>et al.</i> , 2004	Boutros <i>et al.</i> , 2004
		Позиция	Сходство	Позиция	Позиция
<i>Hs:AhR</i> (NM_001621)	CGAGAGCGTGCCCC	+8	0,975	НН	НН
	AGCCTGCGTGAGCC	+406	0,984	+467**	НН
	CACCCGCGTGCCCTG	+696	0,984	НН	НН
	GAATCGCGTGAACC	-7392	0,971	НА	НН
<i>Rn:AhR*</i> (NM_013149)	CCCACGCGTGACAGC	+112	0,974	НН	НН
	TTTTTGCGTGAAAA	-4933	0,984	-4904**	НН
	CCCATGCGTGCTAT	-9119	0,987	НН	НН
<i>Hs:HSP90AA1</i> (NM_001017963)	TTCCTGCGTGTGAT	+550	0,959	НА	НА
	ACCCCGCGTGCTGG	+256	0,973	НА	НА
	ACCTGGCGTGCTCC	-39	0,972	НА	НА
	CGCAGGCGTGCTCA	-52	0,970	НА	НА
	GGTGGGCGTGATCC	-197	0,968	НА	НА
	GCCCTGCGTGCCGA	-993	0,988	НА	НА
	CGGTGGCGTGCTCA	-2830	0,972	НА	НА
	TACAGGCGTGAGCC	-4138	0,968	НА	НА
<i>Hs:HSP90AA1</i> (NM_005348)	TGTTGCGTGCGGC	+306	0,973	НА	НН
	CGGGGGCGTGCGAG	-241	0,970	НА	-234
	CCCGGGCGTGCCCT	-364	0,971	НА	НН
	GTCGTGCGTGGACG	-465	0,955	НА	НН
	CGGGCGCGTGAGAC	-491	0,970	НА	-484
	TACAGGCGTGAGCC	-2157	0,968	НА	НН
	GAATGGCGTGAACC	-3506	0,968	НА	НН
	TAGAGGCGTGAGCC	-3687	0,968	НА	НН
<i>Hs:HSP90AB1</i> (NM_007355)	CTACTGCGTGCCCC	-22	0,985	НН	НН
	AGGCCGCGTGACGA	-1132	0,971	НН	НН
	TGGGAGCGTGATCC	-1701	0,973	НН	НН
	CAGTGGCGTGATCT	-2618	0,970	НН	НН
	TACAGGCGTGAGCC	-3175	0,968	НН	НН
<i>Hs:AIP</i> (NM_003977)	GAATGGCGTGAACC	-858	0,968	НН	НН
	TACAGGCGTGAGCC	-1037	0,968	НН	НН
	CACAGGCGTGACCC	-1293	0,971	НН	НН
	TACAGGCGTGAGCC	-2092	0,968	НН	НН
	TACAGGCGTGAGCC	-2404	0,968	НН	НН

Обозначения: НН – не найдено; НА – не анализировалось. \* Отсчет позиции DRE проводился относительно крайней 5'-позиции соответствующей кДНК. \*\* Наблюдаемое расхождение в позициях первого нуклеотида, выявленного DRE, связано с разницей при определении позиции старта транскрипции.



**Рис. 2.** Сравнение плотности выявляемых методом SITECON DRE в промоторах из базы данных EPD и в промоторах генов компонентов КСК человека (*AhR*, *AIP*, *HSP90AB1*, и два транскрипта *HSP90AA1*) в зависимости от порогового значения уровня конформационного сходства.

ятности говорить о регуляции экспрессии генов КЦК через взаимодействие комплекса *AhR/Arnt* с DRE в их промоторных районах.

### Обсуждение

Ранее функциональность КЦК и влияние *AhR* на экспрессию соответствующих генов-мишеней рассматривались в основном в связи с ролью белок-белковых взаимодействий между партнерами, формирующими комплекс (Carver *et al.*, 1998). Установлено, что внутри КЦК *AhR* взаимодействует с обоими белками в составе димера *Hsp90* (Perdew, Bradfield, 1996). В свою очередь димер *Hsp90* поддерживает *AhR* в конформации, необходимой для связывания лиганда (Whitelaw *et al.*, 1995). *Aip* образует комплекс с *Hsp90* и *AhR*, но способен также к независимому связыванию как *Hsp90*, так и *AhR* (Meyer *et al.*, 2000).

В настоящей работе внимание сосредоточено на компьютерном анализе регуляторных областей генов, кодирующих белки КЦК человека, на предмет наличия DRE, обеспечивающих возможность *AhR*-опосредованной регуляции их экспрессии.

Полученные результаты говорят в пользу существования такого типа регуляции. Действительно, в структуре всех генов КЦК обнаружены DRE – обязательные составляющие регуляторного пути с участием *AhR*. Авторегуляция

экспрессии гена *AhR* крыс, в структуре которого нами выявлены DRE, показана экспериментально (Brauze *et al.*, 2006).

При сравнении организации промоторов генов *AhR* крысы и человека оказалось, что характер расположения потенциальных DRE в них сходен: в обеих последовательностях имеются DRE ниже старта транскрипции (1 в гене крысы в позиции +112 и 3 в гене человека в позициях +8, +406 и +696). Кроме того, в промоторах обоих генов есть по одному DRE в дистальной области на значительном расстоянии от старта транскрипции (в гене крысы в позиции –4932 и в гене человека в позиции –7392). Никаких подтверждений функциональности DRE, локализованных ниже старта транскрипции, пока не найдено, однако в модельных экспериментах показана функциональная активность DRE, расположенных выше старта транскрипции на расстоянии вплоть до –7000 от него (Kawajiri *et al.*, 1986; Kress *et al.*, 1998; Lakhman *et al.*, 2007). Эти данные могут служить аргументом в пользу предположения о функциональности DRE, выявленного в дистальной области промотора гена *AhR* человека.

Известно, что цитозольный белок *HSP90*, важный компонент КЦК, продуцируется двумя паралогичными генами – *HSP90AA1* и *HSP90AB1*. Считается, что *HSP90AA1* частично индуцибелен, тогда как *HSP90AB1* экспрессируется по большей части конститутивно (Chen *et al.*, 2005). В промоторах двух известных изоформ *HSP90AA1*, отличающихся по положению старта транскрипции, нами обнаружено по 8 потенциальных DRE. Пять потенциальных DRE обнаружены и в промоторе важнейшего компонента КЦК – белка *AIP*. Известно, что *AhR*-опосредованная экспрессия гена-мишени при воздействии 2,3,7,8-ТХДД или диоксиноподобных соединений зависит от числа DRE, расположенных в энхансерной области (Kim *et al.*, 2006; Tsyglov, 2006). Поэтому гены, имеющие 5 или более потенциальных DRE в своей регуляторной зоне, – *HSP90AA1* (8 DRE), *AIP* (5 DRE) и *HSP90AB1* (5 DRE) – могут являться первоочередными генами-мишенями для воздействия ксенобиотиков.

Для подтверждения функциональности выявленных потенциальных DRE *in vivo* безусловно необходимы дальнейшие эксперименты,



однако уже сейчас можно сделать несколько умозаключений. Детектирование потенциальных DRE в регуляторной области гена *AhR* и других генов, кодирующих белки КЦК человека, позволяет рассматривать их как структурную основу для AhR-опосредованных путей регуляции транскрипционной активности и говорить о наличии авторегуляторной петли для гена *AhR* в процессах ответа на ксенобиотик. Данные, полученные ранее для ряда генов человека и крысы (Hayashi *et al.*, 1994; Jana *et al.*, 1998; Pitt *et al.*, 2001; Wu *et al.*, 2003; Brauze *et al.*, 2006), свидетельствуют в пользу существования такого типа регуляции в поддержании собственной экспрессии AhR. В свою очередь, если принять во внимание множественные (5 и более) потенциальные DRE, обнаруженные в регуляторной области генов, кодирующих остальные белки КЦК, увеличение экспрессии белка AhR при воздействии ксенобиотика, вероятнее всего, модулирует и экспрессию генов, кодирующих белки корового цитозольного комплекса.

Таким образом, представляется вероятным, что при наличии авторегуляторной петли для *AhR* экспрессия всех генов, кодирующих белки КЦК человека, может регулироваться AhR-зависимым образом и интегрально влиять на реакцию клетки на ксенобиотики.

### Литература

- Boutros P.C., Moffat I.D., Franc M.A. *et al.* Dioxin-responsive AHRE-II gene battery: identification by phylogenetic footprinting // *Biochem. Biophys. Res. Commun.* 2004. V. 321. № 3. P. 707–715.
- Brauze D., Widerak M., Cwykiel J. *et al.* The effect of aryl hydrocarbon receptor ligands on the expression of AhR, AhRR, ARNT, Hif1alpha, CYP1A1 and NQO1 genes in rat liver // *Toxicol. Lett.* 2006. V. 167. № 3. P. 212–220.
- Carver L.A., LaPres J.J., Jain S. *et al.* Characterization of the Ah receptor-associated protein, ARA9 // *J. Biol. Chem.* 1998. V. 273. № 50. P. 33580–33587.
- Chen B., Piel W.H., Gui L. *et al.* The HSP90 family of genes in the human genome: Insights into their divergence and evolution // *Genomics.* 2005. V. 86. № 6. P. 627–637.
- Efron B., Gong G. A leisure look at the bootstrap, the jackknife and cross-validation // *Am. Stat.* 1983. V. 37. P. 36–48.
- Franc M.A., Pohjanvirta R., Tuomisto J., Okey A.B. *In vivo* up-regulation of aryl hydrocarbon receptor expression by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in a dioxin-resistant rat model // *Biochem. Pharmacol.* 2001. V. 62. P. 1565–1578.
- Harper P.A., Riddick D.S., Okey A.B. Regulating the regulator: Factors that control levels and activity of the aryl hydrocarbon receptor // *Biochem. Pharmacol.* 2006. V. 72. № 3. P. 267–279.
- Hayashi S.-I., Watanabe J., Nakachi K. *et al.* Interindividual difference in expression of human Ah receptor and related P450 genes // *Carcinogenesis.* 1994. V. 15. 5. P. 801–806.
- Jana N.R., Sarkar S., Yonemoto J. *et al.* Strain differences in cytochrome P4501A1 gene expression caused by 2,3,7,8-tetrachlorodibenzo-p-dioxin in the rat liver: role of the aryl hydrocarbon receptor and its nuclear translocator // *Biochem. Biophys. Res. Commun.* 1998. V. 248. № 3. P. 554–558.
- Kawajiri K., Watanabe J., Gotoh O. *et al.* Structure and drug inducibility of the human cytochrome P-450c gene // *Eur. J. Biochem.* 1986. 159(2). P. 219–225.
- Kim W.K., In Y.-J., Kim J.-H. *et al.* Quantitative relationship of dioxin-responsive gene expression to dioxin response element in Hep3B and HepG2 human hepatocarcinoma cell lines // *Toxicol. Lett.* 2006. V. 165. 2. P. 174–181.
- Kolchanov N.A., Merkulova T.I., Ignatieva E.V. *et al.* Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes // *Brief Bioinform.* 2007. 8(4). P. 266–274.
- Kress S., Reichert J., Schwarz M. Functional analysis of the human cytochrome P4501A1 (CYP1A1) gene enhancer // *Eur. J. Biochem.* 1998. V. 258. P. 803–812.
- Lakhman S.S., Chen X., Gonzalez-Covarrubias V. *et al.* Functional characterization of the promoter of human carbonyl reductase 1 (CBR1). Role of XRE elements in mediating induction of CBR1 by ligands of the aryl hydrocarbon receptor // *Mol. Pharmacol.* 2007. V. 72. № 3. P. 734–743.
- Ma Q. Induction of *CYP1A1*. The AhR/DRE paradigm: transcription, receptor regulation, and expanding biological roles // *Curr. Drug Metab.* 2001. V. 377. P. 205–213.
- Meyer B.K., Petrusis J.R., Perdew G.H. Aryl hydrocarbon (Ah) receptor levels are selectively modulated by hsp90-associated immunophilin homolog XAP2 // *Cell Stress Chaperones.* 2000. V. 5. N 3. P. 243–254.
- Nedosekina E.A., Oshchepkov D.Y., Katokhin A.V. *et al.* Detection of new potentially active DRE sites in regulatory region of human genes encoding components of Ah receptor cytosolic complex // *Organohalogen Compounds.* 2007. V. 69. P. 1889–1892.
- Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A. *et al.* SITECON: a tool for detecting conservative con-



- formational and physicochemical properties in transcription factor binding site alignments and for site recognition // *Nucl. Acids Res.* 2004. V. 32. W208–212.
- Patel R.D., Kim D.J., Peters J.M., Perdew G.H. The Aryl Hydrocarbon Receptor Directly Regulates Expression of the Potent Mitogen Epiregulin // *Toxicological Sci.* 2006. V. 89(1). P. 75–82.
- Perdew G.H., Bradfield C.A. Mapping the 90 kDa heat shock protein binding region of the Ah receptor // *Biochem. Mol. Biol. Int.* 1996. V. 39. № 3. P. 589–593.
- Petrulis J.R., Perdew G.H. The role of chaperone proteins in the aryl hydrocarbon receptor core complex // *Chem. Biol. Interact.* 2002. V. 141. № 1/2. P. 25–40.
- Pitt J.A., Feng L., Abbott B.D. *et al.* Expression of AhR and ARNT mRNA in cultured human endometrial explants exposed to TCDD // *Toxicol. Sci.* 2001. V. 62. P. 289–298.
- Schmid C.D., Perier R., Praz V., Bucher P. EPD in its twentieth year: towards complete promoter coverage of selected model organisms // *Nucl. Acids Res.* 2006. V. 34. P. 82–85.
- Sogawa K., Fujii-Kuriyama Y. Ah receptor, a novel ligand-activated transcription factor // *J. Biochem. (Tokyo)*. 1997. V. 122. № 6. P. 1075–1079.
- Sun Y.V., Boverhof D.R., Burgoon L.D. *et al.* Comparative analysis of dioxin response elements in human, mouse and rat genomic sequences // *Nucl. Acids Res.* 2004. V. 32. № 15. P. 4512–4523.
- Tsuchiya Y., Nakajima M., Yokoi T. Critical enhancer region to which AhR/ARNT and Sp1 bind in the human CYP1B1 gene // *J. Biochem. (Tokyo)*. 2003. V. 133. P. 583–592.
- Tsyrllov I.B. The role of body burden 2,3,7,8-TCDD in triggering malignancy-associated human viruses: from early data to mechanistic concept // *Organohalogen Comp.* 2006. V. 68. P. 552.
- Whitelaw M., McGuire J., Picard D. *et al.* Heat shock protein hsp90 regulates dioxin receptor function *in vivo* // *Proc. Natl Acad. Sci. USA*. 1995. V. 92. № 10. P. 4437–4441.
- Whitlock J.P. Induction of cytochrome P4501A1 // *Ann. Rev. Pharmacol. Toxicol.* 1999. V. 39. P. 103–125.
- Wu J., Ramesh A., Nayyar T., Hood D.B. Assessment of metabolites and AhR and CYP1A1 mRNA expression subsequent to prenatal exposure to inhaled benzo(a)pyrene // *Int. J. of Dev. Neuroscience*. 2003. V. 21. 6. P. 333–346.

## DETECTION OF NEW DRE SITES IN REGULATORY REGION OF HUMAN GENES ENCODING COMPONENTS OF AH RECEPTOR CYTOSOLIC COMPLEX

**D.Y. Oshchepkov<sup>1</sup>, D.P. Furman<sup>1,2</sup>, E.A. Oshchepkova<sup>1</sup>, A.V. Katokhin<sup>1</sup>,  
M.Y. Shamanina<sup>1</sup>, V.A. Mordvinov<sup>1</sup>**

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

### Summary

Aryl hydrocarbon receptor (AhR) is a ligand-activated transcription factor involved in response to xenobiotics and endogenous signals. AhR modulates expression of many target genes and acts in various tissues of different species. Without a ligand AhR exists as heterotetrameric cytoplasmic core complex. AhR binds a ligand and then the AhR/ligand/core complex is translocated to the nucleus. After that the complex forms heterodimer with nuclear protein Arnt, binds to DRE (dioxin response element) and acts as a transcription factor. The results of DRE search in the genes encoding proteins of heterotetrameric cytoplasmic complex are presented in this paper. Twenty seven new potential DREs located in regulatory regions of five target human genes were found using the computer program SITECON. These DREs are promoter elements and may have functional activity. Obtained data can be used for modeling of AhR-dependent regulation of genes encoding cytoplasmic complex proteins.

## ЦИТОКИНЫ: БИОЛОГИЧЕСКИЕ СВОЙСТВА И РЕГУЛЯЦИЯ ЭКСПРЕССИИ ГЕНА ИНТЕРЛЕЙКИНА-5 ЧЕЛОВЕКА

В.А. Мордвинов<sup>1</sup>, Д.П. Фурман<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: mordvin@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Интерлейкин-5 (ИЛ-5) принадлежит к обширному семейству цитокинов – информационных молекул, секретируемых клетками иммунной системы. Основным источником ИЛ-5 являются зрелые Т-лимфоциты. ИЛ-5 играет центральную роль в развитии и регуляции патологических состояний, связанных с эозинофилией, в том числе при аллергическом воспалении и паразитарных инфекциях. В обзоре приводятся общие данные о цитокинах, систематизированы результаты исследований биологической роли ИЛ-5, освещены вопросы структурно-функциональной организации и регуляции экспрессии гена ИЛ-5, а также механизмы влияния глюкокортикоидов на эозинофилию и экспрессию гена ИЛ-5.

**Ключевые слова:** Т-лимфоциты, цитокины, интерлейкин-5, экспрессия гена.

### Введение

Иммунной системе принадлежит ведущая роль в обеспечении и поддержании гомеостаза организма, а также формировании согласованных реакций его отдельных систем в ответ на внешние воздействия. Множество фактов, полученных при проведении исследований в области молекулярной биологии, иммунологии и физиологии, указывают на существование коммуникационной сети, обеспечивающей межклеточное взаимодействие на тканевом, органном и системном уровнях как основу для функционирования иммунной системы. Межклеточные связи реализуются посредством медиаторов, секретируемых специализированными клетками лимфоидного и миелоидного ряда.

Роль медиаторов выполняют цитокины – большая группа регуляторных белков, осуществляющих передачу сигналов через контакт со специфическими рецепторами на поверхностях клеток. В совокупности цитокины формируют разветвленную и многоуровневую цитокиновую сеть. Для ее правильного функционирования необходимо строгое соблюдение баланса как самих цитокинов, так и их рецепторов, содержание которых подвергается существенным

изменениям в зависимости от состояния участвующих во взаимодействиях клеток.

Нормальная работа цитокиновой сети во многом базируется на механизмах, лежащих в основе регуляции экспрессии генов цитокинов. Однако эти механизмы изучены далеко не достаточно, что существенно затрудняет понимание закономерностей взаимодействия клеток различных типов, отдельных систем организма и формирования его реакций. Иными словами, многие общие и частные проблемы физиологии, иммунологии и других биологических дисциплин требуют исследования механизмов регуляции экспрессии генов цитокинов. Следует добавить, что изучение контроля экспрессии как отдельных генов цитокинов, так и генных кластеров, объединяющих гены цитокинов определенного типа, способствует решению одной из важнейших проблем молекулярной биологии – исследования механизмов регуляции экспрессии генов эукариот.

### Семейство цитокинов: общие сведения

Интерлейкин-5 (ИЛ-5) принадлежит к обширному семейству цитокинов – информационных молекул, секретируемых клетками иммун-

ной системы (Stark *et al.*, 2007), биологические эффекты которых распространяются, однако, далеко за ее пределы. Цитокины выступают посредниками межклеточных и межсистемных взаимодействий, регулируя выживаемость клеток, стимуляцию или подавление их роста, дифференциацию, функциональную активность и апоптоз и обеспечивая согласованность действия иммунной, эндокринной и нервной систем как в нормальных условиях, так и в ответ на разнообразие патологические воздействия.

Продуцентами цитокинов являются макрофаги, лимфоциты, нейтрофилы, тромбоциты, фибробласты, эндотелиоциты, кератиноциты, стромальные клетки, а также ряд других типов клеток. Один и тот же цитокин может синтезироваться различными по гистогенетическому происхождению типами клеток в разных органах и тканях организма (Fainboim *et al.*, 2007). В то же время продукция многих цитокинов является тканеспецифичной, и некоторые из них производятся практически только одним типом клеток. В частности, основным источником ИЛ-5 являются зрелые Т-лимфоциты, и вклад других клеток в общую продукцию этого фактора крайне незначителен (Santana, Rosenstein, 2003).

История изучения цитокинов началась во второй половине XX в. (Vilcek, Feldmann, 2004). Термин «цитокины», предложенный С. Коухеном (Cohen *et al.*, 1974), отражает основное назначение этих молекул, состоящее в передаче сигналов от клетки к клетке, и не специфицирует клетки-источники информации.

Долгое время цитокины получали названия по вызываемым биологическим эффектам. Феноменологический подход к обозначению цитокинов, практиковавшийся фактически до 1990-х гг., нашел отражение в их номенклатуре. В частности, ИЛ-5 имеет 15 альтернативных названий (<http://www.copewithcytokines.de/copy.cgi?key=Cytokines>).

Термин «интерлейкины», т. е. медиаторы, осуществляющие связь между лейкоцитами, был предложен в 1978 г. на II Международном рабочем совещании, посвященном лимфокинам, для обозначения и систематизации регуляторных факторов, секретируемых клетками иммунной системы. Однако очень скоро выяснилось, что биологические эффекты этих

факторов распространяются далеко за ее пределы, и собирательным названием всей группы стал термин «цитокины». К цитокинам относят несколько подгрупп медиаторов:

- интерлейкины (ИЛ) с исторически сложившимися порядковыми номерами;
- интерфероны (ИФН) – факторы, обладающие противовирусными свойствами;
- колониестимулирующие факторы (КСФ), активирующие продукцию и дифференцировку клеток-предшественников различных ростков гемопоэза на разных этапах их созревания;
- трансформирующие ростовые факторы;
- факторы некроза опухолей (ФНО);
- хемокины, или хемотактические цитокины, обеспечивающие активацию процессов миграции различных типов лейкоцитов и некоторых других клеток.

Существующее разделение во многом условно и скорее отражает историю изучения цитокинов. На сегодняшний день достоверно установлено, что один и тот же цитокин способен выполнять различные функции (Gutcher, Becher, 2007). Плейотропизм действия, вероятно, объясняется множественностью и специализацией клеток-мишеней соответствующих регуляторных белков.

В то же время разные цитокины могут приводить к одному и тому же биологическому результату, сохраняя, однако, уникальность в отношении точек или времени приложения своего действия. Например ИЛ-2, ИЛ-7 и ИЛ-15 могут быть охарактеризованы как ростовые факторы Т-лимфоцитов. Каждый из этих медиаторов обладает наибольшим эффектом на определенной фазе клеточного развития, и его эквивалентная замена другим фактором возможна лишь вне этой фазы (Porter, Malek, 2000; Schimpl *et al.*, 2002).

Феномен взаимозаменяемости биологического действия цитокинов можно объяснить тем, что в большинстве случаев их действие реализуется по сетевому принципу с участием нескольких комбинирующихся медиаторов. При этом цитокины действуют в отношениях как синергизма, так и антагонизма. Они способны каскадно индуцировать взаимную выработку, а также инициировать трансмодуляцию поверхностных рецепторов Т-лимфоцитов к другим медиаторам (Sporn, Roberts, 1988; Galve-de

Rochemonteix *et al.*, 1993; Gounni *et al.*, 2000; Kastelein *et al.*, 2007).

Среди клеток иммунной системы – продуцентов цитокинов – значительное место занимают Т-лимфоциты (тимус-зависимые), являющиеся важнейшим компонентом системы клеточного иммунитета. Субпопуляция Т-лимфоцитов, названная Т-хелперами (от англ. helper – помощник), выполняет ведущую роль в запуске и регуляции иммунного ответа. Т-хелперы получили свое название за функцию, которую они выполняют, – индуцировать размножение и (или) дифференцировку других клеток иммунной системы. В зависимости от вырабатываемых цитокинов Т-хелперы разделяют на несколько типов (Mossman *et al.*, 1986). «Наивные», или «нулевые», Т-хелперы, выходящие из тимуса (Тх0), синтезируют практически весь спектр Т-клеточных цитокинов и представляют собой клетки-предшественники двух типов зрелых хелперных клеток – Т-хелперов 1-го или 2-го типов (Тх1 и Тх2).

Т-хелперы типа 1 (Тх1) в основном секретируют интерферон-гамма (ИФН- $\gamma$ ) и интерлейкин-2 (ИЛ-2), стимулируя клеточный иммунитет и активность макрофагов. Тх1 участвуют в иммунных реакциях гиперчувствительности замедленного типа и в защите от вирусных инфекций и от инфекций, вызываемых внутриклеточными микроорганизмами.

Т-хелперы типа 2 (Тх2) обеспечивают регуляцию гуморального иммунитета и реакции антипаразитарной защиты с участием эозинофилов. Их основными цитокинами являются ИЛ-4 и ИЛ-5, важные в патогенезе аллергических заболеваний: ИЛ-4 необходим для синтеза IgE, а ИЛ-5 запускает пролиферацию и дифференцировку эозинофилов (Gelfand, Finkel, 1996; Clement, 1996).

В соответствии с типом продуцируемых их клеток на две полярные группы (Тх1 и Тх2) разделяются и цитокины (Kutlu *et al.*, 2007). Динамическое равновесие цитокинов Тх1 и Тх2 обеспечивает гибкость и пластичность и гуморального, и клеточного иммунного ответа.

Разнообразие индивидуальных свойств цитокинов не мешает, однако, выделить ряд общих характеристик этих молекул. Как уже отмечалось, цитокины представляют собой группу регуляторных факторов, чаще всего представленных

небольшими гликозилированными белками с молекулярной массой до 50 кДа. Они могут как медиаторы секретироваться в межклеточное пространство или ассоциироваться с мембранами синтезирующих их клеток. В обеих формах цитокины обладают биологической активностью.

Важно отметить, что экспрессия генов цитокинов является индуцибельным процессом, т. е. большинство цитокинов не синтезируется клетками вне воспалительной реакции и иммунного ответа. При физиологическом состоянии, принятом за норму, спектр детектируемых мРНК цитокинов узок и уровень экспрессии соответствующих генов невысок. При повреждении тканей, воспалении, опухолеобразовании и во многих других физиологических и патологических ситуациях спектр экспрессирующихся генов цитокинов, обладающих как местной, так и дистантной активностью, значительно расширяется, а уровень экспрессии генов, обладающих базальной активностью, многократно возрастает (Rose-John, Schooltink, 2007).

Цитокины синтезируются в ответ на стимуляцию продуцирующих их клеток через короткий промежуток времени. Синтез прекращается в результате действия отрицательных обратных связей, опосредуемых простагландинами, кортикостероидными гормонами и другими факторами, а также за счет механизмов саморегуляции. В основе «выключения» синтеза лежат, как правило, события, ведущие к блокаде транскрипции или/и сокращению времени жизни мРНК.

Действие цитокинов инициируется в результате их взаимодействия со специфическими клеточными рецепторными комплексами на поверхностях клеток-мишеней. Количество рецепторов для разных медиаторов значительно варьирует (от 100 до 100 000) (Fischereder, 2007). Рецепторные комплексы связывают свои лиганды-цитокины с очень высокой аффинностью, причем отдельные цитокины способны использовать общие субъединицы рецепторов (Allen *et al.*, 2007; Alves *et al.*, 2007).

Сигнал, генерируемый комплексом «цитокин–рецептор», может регулировать как экспрессию данного цитокина и его рецептора, так и синтез целого спектра медиаторов и поверхностных рецепторов. Цитокины индуцируют либо подавляют синтез самих себя, других цитокинов и их рецепторов, участвуя в фор-



мировании цитокиновой сети. Они действуют как аутокринные регуляторы, воздействуя на клетку, синтезирующую и секретирующую данный цитокин; паракринные, активируя клетки, расположенные вблизи клетки-продуцента, например, в очаге воспаления или в лимфоидном органе; эндокринные, воздействуя дистантно на клетки любых органов и тканей после попадания цитокина в кровеносную систему. В последнем случае действие этих медиаторов напоминает действие гормонов.

В рамках иммунной системы цитокины осуществляют взаимосвязь между неспецифическими защитными реакциями и специфическим иммунитетом, действуя в обоих направлениях. Они не имеют антигенной специфичности биологического действия и влияют на функциональную активность клеток, принимающих участие в реакциях врожденного и приобретенного иммунитета. Тем не менее, воздействуя на Т- и В-лимфоциты, цитокины способны стимулировать антиген-зависимые процессы (Corgu, Kheradmand, 1999).

Двумя важными характеристиками экспрессии генов цитокинов являются ее тканеспецифичность и зависимость от активации клеточных сигнальных путей.

Значительная часть цитокинов производится практически только одним типом клеток. Так, у человека и животных основным источником ИЛ-2 являются зрелые Т-лимфоциты, тогда как вклад эозинофилов в общую продукцию этого белка крайне незначителен (Santana, Rosenstein, 2003). Хотя есть цитокины, гены которых эффективно экспрессируются в клетках нескольких типов, список таких клеток весьма ограничен (Yarilin, Belyakov, 2004).

Известно, что тканеспецифический характер экспрессии генов формируется в процессе клеточной дифференцировки (Li *et al.*, 2007) и часто определяется присутствием в соответствующих клетках специфического набора транскрипционных факторов (Rudensky *et al.*, 2006). Кроме того, вклад в тканеспецифичность экспрессии того или иного гена вносит и его «доступность» для транскрипции в контексте организации хроматина (Felsenfeld *et al.*, 1996; Wu, 1997; Narlikar *et al.*, 2002). «Доступность» генов цитокинов, в частности ИЛ-2, для транскрипции достигается в результате дифферен-

цировки клеток иммунной системы. Показано, что в ходе дифференцировки Т-клеток в тимусе хроматин в районе расположения гена ИЛ-2 подвергается изменениям, обуславливающим экспрессию ИЛ-2 только в зрелых Т-лимфоцитах (Bunting *et al.*, 2006).

Вторая важная характеристика экспрессии генов цитокинов – ее регуляция с участием сигнальных путей. Каскад внутриклеточных сигнальных реакций, следующий за взаимодействием определенных лигандов с их рецепторами на поверхности лимфоцита, завершается формированием комплексов регуляторных районов генов цитокинов с конститутивными и/или индуцибельными транскрипционными факторами, что в свою очередь приводит к инициации или ингибированию экспрессии генов (Holloway *et al.*, 2002).

Таким образом, цитокины представляют собой группу белковых медиаторов, участвующих в формировании и регуляции защитных реакций организма. К общим главным свойствам цитокинов относятся плеiotропизм и взаимозаменяемость биологического действия, индуцибельный характер синтеза, отсутствие антигенной специфичности действия, саморегуляция продукции. Цитокиновая сеть обеспечивает кооперацию иммунокомпетентных клеток и может рассматриваться как самостоятельная система регуляции основных функций организма, существующая наряду с нервной и эндокринной системами регуляции и связанная, в первую очередь, с поддержанием гомеостаза при внедрении патогенных агентов и нарушении целостности тканей.

### Биологические свойства ИЛ-5

Около 30 лет назад несколько независимых исследовательских групп начали изучение свойств цитокина, обладающего *in vitro* свойством активатора дифференцировки В-лимфоцитов мыши. Этот белок был известен как фактор, заменяющий Т-клетки (T cell-replacing factor, или TRF) (Takatsu *et al.*, 1988), фактор роста В-лимфоцитов II (B cell growth factor II, или BCGFII) (Swain *et al.*, 1988) и фактор дифференцировки В-лимфоцитов  $\mu$  (B cell differentiation factor- $\mu$ , или BCDF $\mu$ ) (Vitetta *et al.*, 1984). В 1985 г. группой Takatsu было показано, что



названные факторы идентичны (Harada *et al.*, 1985). Приблизительно в это же время был обнаружен фактор дифференцировки эозинофилов мыши (eosinophil differentiation factor, или EDF) (Sanderson *et al.*, 1985; Warren, Sanderson, 1985). В 1986 г. в результате анализа аминокислотной последовательности EDF было установлено, что этот белок идентичен BCGFII (Sanderson *et al.*, 1986). В настоящее время цитокин, носивший ранее все вышеперечисленные названия, известен как ИЛ-5 (Mordvinov, Sanderson, 2001).

ИЛ-5 контролирует количество и активность зрелых эозинофилов в организмах животных и человека. Эозинофилы – это гранулярные лейкоциты, которые продуцируются в костном мозге и после недолгого пребывания в периферической крови мигрируют в ткани. Принято считать, что эти клетки представляют в системе иммунитета основное звено противопаразитарной защиты (Sanderson, 1992). В норме содержание эозинофилов в организме человека не превышает 6 % от всех лейкоцитов, однако при определенных патологиях их количество может увеличиваться в несколько раз. Так, наблюдаемый при паразитарных заболеваниях рост числа эозинофилов, как правило, способствует элиминации паразитов и является признаком адекватной реакции организма (Voehringer, 2007). В противоположность этому, при патологиях аллергической природы, в частности астме, рост числа эозинофилов ассоциируется с обострением заболевания (Kiley *et al.*, 2007). Наличие этих клеток в легких астматиков связано с повреждением ткани, а их количество коррелирует со степенью поздней астматической реакции (Bousquet *et al.*, 1990; Jacobsen *et al.*, 2007). Выявлена также негативная роль эозинофилов при развитии ряда аутоиммунных и злокачественных заболеваний (Di Stefano, Amoroso, 2005). Таким образом, знание механизмов, регулирующих продукцию ИЛ-5, а следовательно, и продукцию зрелых эозинофилов в организме человека является ключом к пониманию патогенеза ряда серьезнейших заболеваний и основой для создания фармакологических препаратов для их лечения.

Основными продуцентами ИЛ-5 являются активированные Тх2 клетки, которые синтезируют также и другие Тх2 цитокины, прежде всего ИЛ-4. В большинстве случаев активиро-

ванные Тх2 одновременно секретируют ИЛ-5 и ИЛ-4 (Zhu *et al.*, 2006), однако коэкспрессия генов ИЛ-5 и ИЛ-4 не является абсолютно обязательной, и существуют механизмы индивидуального запуска экспрессии отдельных генов Тх2 цитокинов (Warren, Sanderson, 1985; Bohjanen *et al.*, 1990; Naora *et al.*, 1994).

Помимо Т-хелперов, ИЛ-5 при определенных условиях могут продуцировать и другие Т-лимфоциты – Т-киллеры (Erard *et al.*, 1993; Coyle *et al.*, 1995; Erb, Le Gros, 1996). Этот факт указывает на то, что в обострении астматического состояния, связанного с эозинофилией, заметную роль, наряду с Т-хелперами, могут играть и Т-киллеры (Akdisa, 2006; Umetsua, DeKruyffa, 2006).

Т-клетки являются основным, но не единственным источником ИЛ-5. Так, было показано, что у мышей SCID с дефицитом Т-лимфоцитов наработка мРНК ИЛ-5 происходит в клетках селезенки (Castro *et al.*, 1995). Дополнительным источником ИЛ-5 могут служить также К-лимфоциты (естественные клетки-киллеры) (Warren *et al.*, 1995), тучные клетки (Okayama *et al.*, 1995; Ryan *et al.*, 2007), В-клетки (Paul *et al.*, 1990; Kis *et al.*, 2006) и сами эозинофилы (Bao *et al.*, 1996; Trivedi, Lloyd, 2007).

### Активация экспрессии гена ИЛ-5

ИЛ-5 принадлежит к семейству индукторов кроветворения и иммунного ответа, членами которого являются также ИЛ-3, ИЛ-4, ИЛ-9, ИЛ-13 и ГМ-КСФ (Lampinen *et al.*, 2004). Гены, кодирующие эти белки, расположены кластером на участке q31 хромосомы 5 (McKenzie *et al.*, 1993; Willman *et al.*, 1993). В субпопуляциях Т-лимфоцитов Тх0 и Тх2 эти гены могут экспрессироваться одновременно, и в литературе часто встречается выражение «активация экспрессии локуса Тх2 цитокинов» (Mosmann, Coffman, 1989; Abbas *et al.*, 1996; Stranick *et al.*, 1997; Lavender *et al.*, 2000; Ansel *et al.*, 2006; Dean, 2006). Однако есть факты, указывающие на то, что, помимо системы инициации экспрессии всех названных генов, существуют регуляторные механизмы, позволяющие адресно индуцировать экспрессию отдельных генов. Так, показано, что в отличие от механизма активации экспрессии большинства генов

локуса, да и генов цитокинов в целом, запуск механизма инициации транскрипции гена ИЛ-5 зависит от синтеза белка *de novo*. Ингибиторы синтеза белка циклогексимид и анизомицин полностью блокируют появление мРНК ИЛ-5 в активированных первичных Т-лимфоцитах, не препятствуя, однако, активному синтезу мРНК ИЛ-2, ИЛ-3, ИЛ-4, ИЛ-10 и ГМ-КСФ (Van Straaten *et al.*, 1994; Naora, Young, 1995; Stranick *et al.*, 1997). Таким образом, в активированных Т-лимфоцитах нарабатывается по крайней мере один белок, необходимый для инициации транскрипции гена ИЛ-5, но несущественный для индукции экспрессии других генов цитокинов (Schwenger *et al.*, 2002).

Активация Т-лимфоцитов *in vivo* происходит в результате их взаимодействия с антиген-представляющими клетками (Davis, 2002; Voes, Ploegh, 2004). Для полной активации Т-клетка должна получить два сигнала, иницирующих работу различных сигнальных путей (Rincon, Flavell, 1994; Lavender *et al.*, 2000; Wetzel, Parker, 2006).

Основной сигнальный путь реализуется с участием рецептора Т-лимфоцита (TCR – англ. T-cell Receptor). Индукция TCR-зависимого сигнального пути происходит в результате взаимодействия TCR с антигеном, связанным с молекулами главного комплекса гистосовместимости антиген-представляющих клеток. Возбуждение второго, дополнительного стимулирующего сигнала происходит при взаимодействии CD28 – поверхностного рецептора Т-лимфоцитов – с его лигандом, поверхностным белком антиген-представляющих клеток (Jenkins, Johnson, 1993; Ward, Kaufman, 2007).

Т-лимфоциты, стимулированные антиген-представляющими клетками, продуцируют широкий диапазон регуляторных факторов, и ИЛ-5 часто, но не всегда, экспрессируется одновременно с цитокинами ИЛ-4 и ИЛ-13 (Sanderson *et al.*, 1999). Неясно, почему одни антигены стимулируют продукцию ИЛ-5, а другие нет. Однако этот факт еще раз указывает на то, что экспрессия гена ИЛ-5 контролируется индивидуальной системой регуляции, и запуск механизмов этой системы зависит от антиген-специфической активации Т-клеток.

В экспериментах *in vitro* показано, что в регуляции экспрессии гена ИЛ-5 участвует еще

один дополнительный сигнальный путь, связанный с сАМР, причем он, по-видимому, является специфичным для ИЛ-5 (Hallet *et al.*, 1992; Lee *et al.*, 1993; McCaffrey *et al.*, 1993; Abbas *et al.*, 1996; Karlen *et al.*, 1996; Kaminuma *et al.*, 1999; Kiefer *et al.*, 2002; Klein-Hessling *et al.*, 2003; Quintana *et al.*, 2005; Wu *et al.*, 2005).

Состав транскрипционных факторов, принимающих участие в регуляции экспрессии гена ИЛ-5, зависит от условий клеточной активации или, иными словами, от участия того или иного дополнительного сигнального пути в индукции транскрипции гена (Mordvinov *et al.*, 1999a). В частности, установлено, что в клетках лимфоцитарной культуры клеток человека PER-117 (Kees *et al.*, 1987) циклоспорин А, ингибитор экспрессии многих генов цитокинов, существенно понижает РМА/ $\alpha$ CD28- и РМА/СаI-индуцированную продукцию ИЛ-5 и не влияет на РМА/сАМР-активированный синтез этого белка. В основе механизма подавления экспрессии генов цитокинов циклоспорином А лежит его способность блокировать активность  $Ca^{2+}$ /кальмодулин-зависимой фосфатазы кальцинериум (Cho *et al.*, 2003), необходимой для перевода цитоплазматической формы фактора транскрипции NF-AT в активное состояние и последующей ядерной транслокации этого белка (McCaffrey *et al.*, 1993; Kelly, 2006). NF-AT является ключевым фактором для запуска транскрипции значительного числа генов цитокинов. Устойчивость или чувствительность экспрессии ИЛ-5 к действию циклоспорина А указывает на то, что в зависимости от условий клеточной стимуляции доминирующую роль в активации экспрессии гена ИЛ-5 могут играть различные регуляторные факторы. Так, при стимуляции культур с помощью РМА в комбинации с  $\alpha$ CD28 или СаI вклад NF-AT в активацию экспрессии гена ИЛ-5 весьма заметен. Однако при использовании РМА в комбинации с сАМР NF-AT не играет существенной роли в экспрессии гена (Mordvinov, Sanderson, 2001).

### **Структура гена ИЛ-5 и регуляция его экспрессии**

Ген ИЛ-5 человека обладает стандартной структурой генов эукариот. Кодировочная область гена представлена 4 экзонами, размеры и

первичная последовательность которых консервативны. Размер зрелой мРНК ИЛ-5 составляет 900 п.о. (Sanderson, 1990).

Регуляция экспрессии гена ИЛ-5, как и ряда других генов цитокинов, осуществляется на уровне транскрипции, инициация и активность которой контролируются кластером *cis*-регуляторных элементов, локализованных в пределах проксимального промотора (Karlen *et al.*, 1998; Schwenger, Sanderson, 1998; Sanderson, Urwin, 2000; Mordvinov, Sanderson, 2001; Ngoc *et al.*, 2005).

Важнейшей составляющей этого кластера является CLE0 (консервативный лимфокиновый элемент 0). Он расположен непосредственно за ТАТА-боксом и представляет собой композиционный элемент, активируемый комплексом регуляторных белков (Thomas *et al.*, 1999; Salerno *et al.*, 2001). Показано, что этот элемент играет главную роль в реализации механизмов, контролирующих экспрессию гена ИЛ-5, поскольку независимо от условий клеточной активации, любые нарушения последовательности CLE0 полностью блокируют инициацию транскрипции (Sanderson *et al.*, 1999; Thomas *et al.*, 1999).

CLE0 – консервативный регуляторный элемент, присутствующий в проксимальных отделах промоторов большинства генов лимфоцитарных цитокинов (The Cytokine ..., 2003). Его вклад в регуляцию транскрипции различных генов значительно варьирует, но только в контексте промотора ИЛ-5 CLE0 является критическим для активации экспрессии (Mordvinov, Sanderson, 2001). Этот факт ещё раз указывает на существование индивидуального механизма регуляции экспрессии гена ИЛ-5.

Последовательность CLE0 гена ИЛ-5 (CLE0 ИЛ-5) обладает значительным сходством с CLE0 ИЛ-3, ИЛ-4 и ГМ-КСФ. Так, гомология CLE0 ИЛ-5 и ГМ-КСФ составляет 94 %, однако эти элементы взаимодействуют с различными регуляторными комплексами. Установлено, что в отличие от CLE0 ГМ-КСФ, связывающего факторы транскрипции AP-1, NF-AT и Ets (Masuda *et al.*, 1993; Jenkins *et al.*, 1995), CLE0 ИЛ-5 образует комплекс с белками AP-1 и Octamer (Takatsu *et al.*, 1999; Thomas *et al.*, 1999; Salerno *et al.*, 2001).

Фактор транскрипции AP-1 принимает участие в активации транскрипции многих генов

цитокинов и довольно часто входит в состав регуляторного комплекса CLE0 (Jenkins, Johnson, 1993; Karlen *et al.*, 1996, 1998). Этот фактор представляет собой гетеро- или гомодимер, субъединицами которого являются ядерные белки, принадлежащие к семействам Jun и Fos. Оказалось, что субъединицами AP-1, входящего в состав регуляторного комплекса CLE0 ИЛ-5 человека, являются JunD и Fra-2 (Thomas *et al.*, 1999), синтез которых может быть индуцирован в результате активации как протеинкиназы C, так и протеинкиназы A (Mechta *et al.*, 1989; den Hertog *et al.*, 1992; Gruda *et al.*, 1994; Miller *et al.*, 1998). Иными словами, стимуляция и TCR-зависимого сигнального пути, общего для активации экспрессии генов Т-клеточных цитокинов, и cAMP-зависимого сигнального каскада, селективно активирующего экспрессию гена ИЛ-5, приводит к формированию CLE0 ИЛ-5-связывающего AP-1. Эти данные указывают на чрезвычайную важность AP-1 как компонента регуляторного комплекса, обеспечивающего активацию экспрессии гена ИЛ-5. Исследование механизма формирования комплекса CLE0/AP-1, а также роли субъединиц фактора AP-1 в активации и регуляции активности транскрипции показало, что активация экспрессии гена ИЛ-5 зависит от синтеза *de novo* белка Fra-2 (Schwenger *et al.*, 2002).

Как уже упоминалось, в состав регуляторного комплекса CLE0 ИЛ-5 входят также белки, принадлежащие к семейству факторов транскрипции Octamer. Экспериментально установлено, что в первичной последовательности CLE0 мотивы связывания факторов транскрипции Octamer и AP-1 частично перекрываются (Thomas *et al.*, 1999). Подобные примеры взаимодействия трех факторов транскрипции с одним регуляторным элементом были ранее описаны при исследовании структурно-функциональной организации промоторов генов ИЛ-2 и ИЛ-4 (Pfeuffer *et al.*, 1994). Вполне вероятно, что тесная ассоциация ядерных белков в пределах одного регуляторного элемента является обязательным условием их взаимодействия с транскрипционным комплексом элемента ТАТА и фактически лимитирует инициацию транскрипции. В этом случае не только сайты связывания с регуляторными белками, входящие в CLE0 ИЛ-5, но и вся последовательность этого

элемента должны определять эффективность кооперативного взаимодействия регуляторных элементов. Действительно, изменения в последовательности CLE0 как разрушающие белок-связывающие мотивы, так и не затрагивающие их, фактически блокируют работу промотора. Эти данные еще раз указывают на то, что CLE0 является одним из ключевых регуляторных элементов промотора гена ИЛ-5 человека, участвующих в инициации транскрипции.

Интересно, что Oct-1 в комбинации с фактором YY1 входит в состав ДНК-белковых комплексов репрессорного элемента, находящегося на расстоянии 20 нуклеотидов от CLE0 (Mordvinov *et al.*, 1999b). Таким образом, Oct-1 принимает участие в активации двух регуляторных элементов, выполняющих противоположные функции в регуляции транскрипции ИЛ-5.

Кроме Octamer-связывающего фрагмента, в 5'-фланкирующей области CLE0 в районе позиции -70 промотора гена ИЛ-5 расположен мотив связывания факторов транскрипции семейства GATA (Sanderson *et al.*, 1999). Известно, что GATA-3 играет главную роль в активации локуса T<sub>H</sub>2 цитокинов и теоретически выполняет функции регулятора координированной экспрессии генов ИЛ-4 и ИЛ-5 (Zheng *et al.*, 1997; Ouyang *et al.*, 1998; Zhang *et al.*, 1999; Nakamura, Hoshino, 2005; Romagnani, 2006; Zhu *et al.*, 2006). При исследовании функциональной значимости и белок-связывающей активности сайта GATA, расположенного в районе позиции -70 промотора гена ИЛ-5, было показано, что этот мотив взаимодействует с фактором транскрипции GATA-3 и, очевидно, совместно с CLE0 контролирует активацию экспрессии гена (Schwenger *et al.*, 2001).

При функциональном картировании промотора гена ИЛ-5 был обнаружен еще один район, взаимодействующий с белками GATA. Установлено, что этот район, охватывающий участок между позициями -110 и -170, участвует в активации транскрипции гена (Prieschl *et al.*, 1995). Присутствие в составе промотора второго сайта связывания белков GATA может указывать на высокую значимость этих факторов транскрипции в регуляции экспрессии гена ИЛ-5. Действительно, в результате развития работ по выявлению и функциональному тестированию элементов GATA, входящих в состав

промоторов генов ИЛ-5 и ИЛ-4, были получены данные, подтверждающие это предположение (Schwenger *et al.*, 2001; Швенгер и др., 2005).

Предметом активного обсуждения являлась роль фактора NF-AT в регуляции транскрипции гена ИЛ-5 (De Boer *et al.*, 1999). Дискуссия подогревалась обилием противоречивых результатов экспериментальных исследований структурно-функциональной организации промотора гена ИЛ-5, обусловленных разнообразием культур клеток и условий клеточной активации. Как упоминалось выше, при определенных условиях клеточной активации экспрессия гена ИЛ-5 может инициироваться без участия фактора транскрипции NF-AT (Mordvinov *et al.*, 1999a). Этот феномен долгое время считался артефактом, что, естественно, не способствовало формированию общепринятого взгляда на роль NF-AT. Кроме того, разные исследовательские группы использовали в качестве клеточной модели либо перевиваемые клетки мыши, либо культуры клеток человека, не продуцирующие ИЛ-5 (Sanderson *et al.*, 1999; Mordvinov, Sanderson, 2001). Полученные с помощью таких моделей данные характеризовали, очевидно, индивидуальные особенности клеточных линий, а не общие механизмы регуляции экспрессии гена ИЛ-5 в первичных Т-клетках человека.

В одной из последних работ, выполненной с использованием культуры клеток HSB-2 человека, результаты которой были верифицированы на первичных Т-лимфоцитах человека, показано, что, помимо факторов GATA-3 и AP-1, в активации транскрипции гена ИЛ-5 участвует фактор Ets1, причем действие этих факторов синергично (Wang, Young, 2007).

Таким образом, на сегодняшний день установлено, что в состав транскрипционного комплекса ИЛ-5, как правило, входят транскрипционные факторы AP-1, GATA-3, Oct1/2, YY1 и Ets1, а при определенных условиях клеточной активации еще и NF-AT.

#### **Влияние глюкокортикоидов на эозинофилию и экспрессию гена ИЛ-5**

Как упоминалось выше, у человека и животных следствием повышенной продукции ИЛ-5 является эозинофилия. Накопление эозинофилов в периферической крови и тканях может



приводить к развитию целого ряда патологий. К ним относятся, прежде всего, астма и аллергический ринит – воспалительные заболевания дыхательных путей, от которых страдает значительная часть населения индустриально развитых стран (Foley, Hamid, 2007; Taylor *et al.*, 2007; Trivedi, Lloyd, 2007). На сегодняшний день одним из наиболее эффективных средств лечения заболеваний, сопровождающихся эозинофилией, являются препараты на основе глюкокортикоидных гормонов (Hall, 2006; Abdullah, Khan, 2007; Spahn, Szefer, 2007).

В результате применения этих препаратов сокращается общее число эозинофилов и секретируемых ими продуктов, снижается содержание наиболее активной фракции этих клеток, обладающих фенотипом *hypodense* (Giembycz, Lindsay, 1999). Кроме того, установлено, что глюкокортикоиды подавляют миграцию эозинофилов к очагам воспаления, вызванного аллергеном (Elwood *et al.*, 1992; Underwood *et al.*, 1997; Birrell *et al.*, 2003).

Практическое применение в качестве лекарственных средств, помимо кортизона и гидрокортизона, нашел целый ряд их синтетических заменителей: преднизон, преднизолон, метилпреднизолон, дексаметазон, триамцинолон, флуметазон, бетаметазон и др. Эти соединения, как правило, более активны, чем природные глюкокортикоиды, и действуют в меньших дозах (Rogatsky, Ivashkiv, 2006). При проведении исследовательских работ по изучению механизмов действия глюкокортикоидов на различные клеточные процессы и экспрессию генов часто используется дексаметазон (Smith, Herschman, 2004). Следует отметить, что этот синтетический глюкокортикоид является также одним из самых распространенных препаратов для лечения заболеваний, сопровождающихся эозинофилией (Mygind, Andersson, 2006; Jacobsen *et al.*, 2007).

Характерной особенностью глюкокортикоидов является иммунодепрессивная активность, не связанная, однако, с митостатическим действием. В ранних работах по исследованию действия глюкокортикоидов на созревание различных клеток гемопоэтического ряда было обнаружено, что в культуре костного мозга эти вещества ингибируют формирование колоний, обогащённых эозинофилами (Butterfield *et al.*,

1986). Оказалось, что этот эффект возникает в результате подавления секреции ИЛ-5 клетками культуры и не обусловлен прямым действием глюкокортикоидов на клетки-предшественники эозинофилов (Buckingham, 2006). Точно так же подавление накопления эозинофилов в очаге воспаления происходит не в результате прямого действия глюкокортикоидов на адгезивные свойства этих клеток, а опять-таки вследствие подавления секреции ИЛ-5 Т-клетками (Giembycz, Lindsay, 1999). Таким образом, основной терапевтический эффект глюкокортикоидов в контексте патологий, сопровождающихся эозинофилией, обусловлен их ингибирующим действием на продукцию ИЛ-5.

Установлено, что глюкокортикоиды ингибируют экспрессию гена ИЛ-5 на транскрипционном уровне (Sanderson *et al.*, 1999; Mordvinov, Sanderson, 2001). Действительно, эти соединения часто выступают в роли модуляторов транскрипции генов цитокинов (Adcock, Caramori, 2001; Arthaningtyas *et al.*, 2005; Мордвинов и др., 2005). Подавление транскрипции глюкокортикоидами может осуществляться как с использованием механизма прямого воздействия на промоторную активность гена-мишени, так и опосредованно, через активацию экспрессии белка, обладающего свойствами транскрипционного репрессора гена-мишени. Происходит это следующим образом: после проникновения в клетку глюкокортикоиды образуют комплекс с цитоплазматическим глюкокортикоидным рецептором (ГР). Затем комплекс, который чаще называют «активированный ГР», перемещается в ядро, где действует уже как фактор транскрипции. Промоторы ряда генов содержат регуляторные элементы, при взаимодействии с которыми активированный ГР способен инициировать, усиливать или подавлять их транскрипцию (Hayashi *et al.*, 2004; Barnes, 2006).

Активированный ГР может регулировать транскрипцию и без прямого контакта с ДНК. В этом случае он влияет на активность факторов транскрипции посредством белок-белковых взаимодействий, что, как правило, приводит к существенному снижению эффективности транскрипции соответствующих генов-мишеней. Такой механизм репрессии реализуется чаще всего для ди- и мультимерных транскрипционных факторов, состоящих из отдельных



субъединиц, и экспериментально показан для NF-κB и AP-1 (Jonat *et al.*, 1990; Yang-Yen *et al.*, 1990; Konig *et al.*, 1992; Ponta *et al.*, 1992; Ray, Prefontaine, 1994).

Изучение механизмов действия глюкокортикоидов на экспрессию генов цитокинов и в частности на экспрессию гена ИЛ-5, остается «горячей» темой молекулярной биологии и иммунологии (Walter, Holtzman, 2005; Gibson, 2006; Roufousse *et al.*, 2006). Недавно было показано, что дексаметазон подавляет экспрессию гена ИЛ-5 в клетках PER-117 на транскрипционном уровне, и его мишенью в контексте регуляторных элементов гена ИЛ-5 является CLE0 (Мордвинов и др., 2005; Arthaningtyas *et al.*, 2005). Однако многие вопросы, связанные с реконструкцией механизма подавления транскрипции гена ИЛ-5 этими соединениями, не выяснены до сих пор. В частности, не идентифицированы факторы, обеспечивающие в определенных ситуациях нечувствительность экспрессии гена ИЛ-5 к их действию (Inagaki *et al.*, 2006; Kosak *et al.*, 2006). Результаты, полученные при исследовании этих вопросов, могут быть полезны как для решения задач академического характера, так и при выполнении прикладных работ по созданию новых фармакологических препаратов для лечения заболеваний, сопровождающихся эозинофилией.

### Литература

- Мордвинов В.А., Кок Ч.Ч., Артанингтас Э. и др. Дексаметазон подавляет активность промотора гена интерлейкина-5 человека // Бюл. эксперим. биол. медицины. 2005. Т. 140. С. 91–95.
- Швенгер Г.Т.Ф., Мордвинов В.А., Сандерсон К.Дж. Фактор транскрипции GATA-3 участвует в подавлении активности промотора гена интерлейкина-4 человека // Биохимия. 2005. Т. 70. С. 1289–1294.
- Abbas A.K., Murphy K.M., Sher A. Functional diversity of helper T lymphocytes // Nature. 1996. V. 383. P. 787–793.
- Adcock I.M., Caramori G. Cross-talk between pro-inflammatory transcription factors and glucocorticoids // Immunol. Cell. Biol. 2001. V. 79. P. 376–384.
- Abdullah A.K., Khan S. Evidence-based selection of inhaled corticosteroid for treatment of chronic asthma // J. Asthma. 2007. V. 44. P. 1–12.
- Akdisa A. Allergy and hypersensitivity: Mechanisms of allergic disease // Curr. Opin. Immunol. 2006. V. 18. P. 718–726.
- Allen S.J., Crown S.E., Handel T.M. Chemokine: receptor structure, interactions, and antagonism // Ann. Rev. Immunol. 2007. V. 25. P. 787–820.
- Alves N.L., Arosa F.A., van Lier R.A. Common gamma chain cytokines: dissidence in the details // Immunol. Lett. 2007. V. 108. P. 113–120.
- Ansel K.M., Djuretic I., Tanasa B., Rao A. Regulation of Th2 differentiation and I4 locus accessibility // Annu. Rev. Immunol. 2006. V. 24. P. 607–656.
- Arthaningtyas E., Kok C.C., Mordvinov V.A., Sander-son C.J. The conserved lymphokine element 0 is a powerful activator and target for corticosteroid inhibition in human interleukin-5 transcription // Growth Factors. 2005. V. 23. P. 211–221.
- Bao S., McClure S.J., Emery D.L., Husband A.J. Interleukin-5 mRNA expressed by eosinophils and gamma/delta T cells in parasite-immune sheep // Eur. J. Immunol. 1996. V. 26. P. 552–556.
- Barnes P.J. Corticosteroid effects on cell signaling // Eur. Respir. J. 2006. V. 27. P. 413–426.
- Birrell M.A., Battram C.H., Woodman P. *et al.* Dissociation by steroids of eosinophilic inflammation from airway hyperresponsiveness in murine airways // Respir. Res. 2003. V. 4. P. 3–20.
- Boes M., Ploegh H.L. Translating cell biology *in vitro* to immunity *in vivo* // Nature. 2004. V. 430. P. 264–271.
- Bohjanen P.R., Okajima M., Hodes R.J. Differential regulation of interleukin 4 and interleukin 5 gene expression: a comparison of T-cell gene induction by anti-CD3 antibody or by exogenous lymphokines // Proc. Natl Acad. Sci. USA. 1990. V. 87. P. 5283–5287.
- Bousquet J., Chanaz P., Lacoste J.Y. *et al.* Eosinophilic inflammation in asthma // N. Engl. J. Med. 1990. V. 323. P. 1033–1039.
- Buckingham J.C. Glucocorticoids: examples of multi-tasking // Br. J. Pharmacol. 2006. V. 147. P. S258–S268.
- Bunting K., Wang J., Shannon M.F. Control of interleukin-2 gene transcription: a paradigm for inducible, tissue-specific gene expression // Vitam. Horm. 2006. V. 74. P. 105–145.
- Butterfield J.H., Ackerman S.J., Weiler D. *et al.* Effects of glucocorticoids on eosinophil colony growth // J. Allergy Clin. Immunol. 1986. V. 78. P. 450–457.
- Castro A.G., Silva R.A., Minoprio P., Appelberg R. *In vivo* evidence for a non-T cell origin of interleukin-5 // Scand. J. Immunol. 1995. V. 41. P. 288–292.
- Cho C.S., Elkahwaji J., Chang Z. *et al.* Modulation of the electrophoretic mobility of the linker for activation of T cells (LAT) by the calcineurin inhibitors CsA and FK506: LAT is a potential substrate for PKC and calcineurin signaling pathways // Cell Signal. 2003. V. 15. P. 85–93.
- Clement L.T. Cellular interactions in the human immune

- response // *Immunologic Disorders of Infants and Children* / Ed. E.R. Stiehm. 4th ed. PA: W.B. Saunders. Philadelphia, 1996. P. 75–93.
- Cohen S., Fisher B., Yoshida T., Bettigole R.E. Serum migration-inhibitory activity in patients with lymphoproliferative diseases // *N. Engl. J. Med.* 1974. V. 290. P. 882–886.
- Corry D.B., Kheradmand F. Induction and regulation of the IgE response // *Nature.* 1999. V. 402. P. 18–23.
- Coyle A.J., Erard F., Bertrand C. *et al.* Virus-specific CD8+ cells can switch to interleukin 5 production and induce airway eosinophilia // *J. Exptl. Med.* 1995. V. 181. P. 1229–1233.
- Davis M.M. A new trigger for T cells // *Cell.* 2002. V. 110. P. 285–287.
- Dean A. On a chromosome far, far away: LCRs and gene expression // *Trends Genet.* 2006. V. 22. P. 38–45.
- De Boer M.L., Mordvinov V.A., Thomas M.A., Sanderson C.J. Role of nuclear factor of activated T cells (NFAT) in the expression of interleukin-5 and other cytokines involved in the regulation of hemopoietic cells // *Int. J. Biochem. Cell. Biol.* 1999. V. 31. P. 1221–1236.
- den Hertog J., de Groot R.P., de Laat S.W., Kruijer W. EGF induced jun B expression in transfected P19 embryonal carcinoma cells expressing EGF receptors is dependent on Jun D // *Nucl. Acids Res.* 1992. V. 20. P. 125–130.
- Di Stefano F., Amoroso A. Clinical approach to the patient with blood eosinophilia // *Allerg. Immunol. (Paris).* 2005. V. 37. P. 380–386.
- Elwood W., Lotvall J.O., Barnes P.J., Chung K.F. Effect of dexamethasone and cyclosporin A on allergen-induced airway hyperresponsiveness and inflammatory cell responses in sensitized Brown-Norway rats // *Am. Rev. Respir. Dis.* 1992. V. 145. P. 1289–1294.
- Erard F., Wild M.T., Garcia-Sanz J.A., Le Gros G. Switch of CD8 T cells to noncytolytic CD8-CD4- cells that make TH2 cytokines and help B cells // *Science.* 1993. V. 260. P. 1802–1805.
- Erb K.J., Le Gros G. The role of Th2 type CD4+ T cells and Th2 type CD8+ T cells in asthma // *Immunol. Cell. Biol.* 1996. V. 74. P. 206–208.
- Fainboim L., Chernavsky A., Paladino N. *et al.* Cytokines and chronic liver disease // *Cytokine Growth Factor Rev.* 2007. V. 18. P. 143–157.
- Felsenfeld G., Boyes J., Chung J. *et al.* Chromatin structure and gene expression // *Proc. Natl Acad. Sci. USA.* 1996. V. 93. P. 9384–9388.
- Fischereder M. Chemokines and chemokine receptors in renal transplantation – from bench to bedside // *Acta Physiol. Hung.* 2007. V. 94. P. 67–81.
- Foley S., Hamid Q. Role of inflammatory T cells and eosinophils in chronic rhinosinusitis // *Clin. Allergy Immunol.* 2007. V. 20. P. 79–91.
- Galve-de Rochemonteix B., Wiktorowicz K., Kushner I., Dayer J.M. C-reactive protein increases production of IL-1 alpha, IL-1 beta, and TNF-alpha, and expression of mRNA by human alveolar macrophages // *J. Leukoc. Biol.* 1993. V. 53. P. 439–445.
- Gelfand E.W., Finkel T.H. The T-lymphocyte system // *Immunologic Disorders of Infants and Children* / Ed. E.R. Stiehm. 4th ed. PA: W.B. Saunders. Philadelphia, 1996. P. 14–34.
- Gibson P.G. Allergic bronchopulmonary aspergillosis // *Semin. Respir. Crit. Care Med.* 2006. V. 27. P. 185–191.
- Giembycz M.A., Lindsay M.A. Pharmacology of the eosinophil // *Pharmacol. Rev.* 1999. V. 51. P. 213–340.
- Gounni A.S., Gregory B., Nutku E. *et al.* Interleukin-9 enhances interleukin-5 receptor expression, differentiation, and survival of human eosinophils // *Blood.* 2000. V. 96. P. 2163–2171.
- Gruda M.C., Kovary K., Metz R., Bravo R. Regulation of Fra 1 and Fra 2 phosphorylation differs during the cell cycle of fibroblasts and phosphorylation *in vitro* by MAP kinase affects DNA binding activity // *Oncogene.* 1994. V. 9. P. 2537–2547.
- Gutcher I., Becher B. APC-derived cytokines and T cell polarization in autoimmune inflammation // *J. Clin. Invest.* 2007. V. 117. P. 1119–1127.
- Hall I.P. Pharmacogenetics of asthma // *Chest.* 2006. V. 130. P. 1873–1878.
- Hallet M.M., Peyrat M.A., Soullillou J.P., Moreau J.F. Simultaneous transcription of eleven cytokines in human alloreactive T lymphocyte clones after stimulation by phorbol ester and A23187 // *Eur. Cytokine Netw.* 1992. V. 3. P. 477–483.
- Harada N., Kikuchi Y., Tominaga A. *et al.* BCGFII activity on activated B cells of a purified murine T cell-replacing factor (TRF) from a T cell hybridoma (B151K12) // *J. Immunol.* 1985. V. 134. P. 3944–3951.
- Holloway A.F., Rao S., Shannon M.F. Regulation of cytokine gene transcription in the immune system // *Mol. Immunol.* 2002. V. 38. P. 567–580.
- Hayashi R., Wada H., Ito K., Adcock I.M. Effects of glucocorticoids on gene transcription // *Eur. J. Pharmacol.* 2004. V. 500. P. 51–62.
- Inagaki N., Shiraishi N., Igeta K. *et al.* Inhibition of scratching behavior associated with allergic dermatitis in mice by tacrolimus, but not by dexamethasone // *Eur. J. Pharmacol.* 2006. V. 546. P. 189–196.
- Jacobsen E.A., Ochkur S.I., Lee N.A., Lee J.J. Eosinophils and asthma // *Curr. Allergy Asthma Rep.* 2007. V. 7. P. 18–26.
- Jenkins F., Cockerill P.N., Bohmann D., Shannon M.F. Multiple signals are required for function of the

- human granulocyte macrophage colony stimulating factor gene promoter in T cells // *J. Immunol.* 1995. V. 155. P. 1240–1251.
- Jenkins M.K., Johnson J.G. Molecules involved in T cell costimulation // *Current Opin. Immunol.* 1993. V. 5. P. 361–367.
- Jonat C., Rahmsdorf H.J., Park K.K. *et al.* Antitumor promotion and antiinflammation: down-modulation of AP-1 (Fos/Jun) activity by glucocorticoid hormone // *Cell.* 1990. V. 62. P. 1189–1204.
- Kaminuma O., Mori A., Ogawa K. *et al.* Cyclic AMP suppresses interleukin-5 synthesis by human helper T cells via the downregulation of the calcium mobilization pathway // *Brit. J. Pharmacol.* 1999. V. 127. P. 521–529.
- Karlen S., De Boer M.L., Lipscombe R.J. *et al.* Biological and molecular characteristics of interleukin 5 and its receptor // *Int. Rev. Immunol.* 1998. V. 16. P. 227–247.
- Karlen S., Mordvinov V.A., Sanderson C.J. How is expression of the interleukin-5 gene regulated? // *Immunol. Cell. Biol.* 1996. V. 74. P. 218–223.
- Kastelein R.A., Hunter C.A., Cua D.J. Discovery and biology of IL-23 and IL-27: related but functionally distinct regulators of inflammation // *Annu. Rev. Immunol.* 2007. V. 25. P. 221–242.
- Kees U.R., Ford J., Price P.J. *et al.* PER 117: a new human ALL cell line with an immature thymic phenotype // *Leukemia Res.* 1987. V. 11. P. 489–498.
- Kelly D.A. Current issues in pediatric transplantation // *Pediatr. Transplant.* 2006. V. 10. P. 712–720.
- Kiefer F., Vogel W.F., Arnold R. Signal transduction and co-stimulatory pathways // *Transpl. Immunol.* 2002. V. 9. P. 69–82.
- Kiley J., Smith R., Noel P. Asthma phenotypes // *Curr. Opin. Pulm. Med.* 2007. V. 13. P. 19–23.
- Kis L.L., Takahara M., Nagy N. *et al.* IL-10 can induce the expression of EBV-encoded latent membrane protein-1 (LMP-1) in the absence of EBNA-2 in B lymphocytes and in Burkitt lymphoma- and NK lymphoma-derived cell lines // *Blood.* 2006. V. 107. P. 2928–2935.
- Klein-Hessling S., Jha M.K., Santner-Nanan B. *et al.* Protein kinase A regulates GATA-3-dependent activation of IL-5 gene expression in Th2 cells // *J. Immunol.* 2003. V. 170. P. 2956–2961.
- Kocak A.K., Bor O., Yildiz B. *et al.* T-lymphocyte activation and the levels of eosinophilic cationic protein and interleukin-5 in asthmatic children with acute exacerbation and effect of glucocorticoid treatment // *Allergy Asthma Proc.* 2006. V. 27. P. 371–377.
- Konig H., Ponta H., Rahmsdorf H.J., Herrlich P. Interference between pathway-specific transcription factors: glucocorticoids antagonize phorbol ester-induced AP-1 activity without altering AP-1 site occupation *in vivo* // *Embo J.* 1992. V. 11. P. 2241–2246.
- Kutlu A., Bozkurt B., Ciftci F., Bozkanat E. Th1-Th2 interaction: is more complex than a see-saw? // *Scand. J. Immunol.* 2007. V. 65. P. 393–395.
- Lampinen M., Carlson M., Hakansson L.D., Venge P. Cytokine-regulated accumulation of eosinophils in inflammatory disease // *Allergy.* 2004. V. 59. P. 793–805.
- Lavender P., Cousins D., Lee T. Regulation of Th2 cytokine gene transcription // *Chem. Immunol.* 2000. V. 78. P. 16–29.
- Lee H.J., Koyano-Nakagawa N., Naito Y. *et al.* cAMP activates the IL5 promoter synergistically with phorbol ester through the signaling pathway involving protein kinase A in mouse thymoma line EL4 // *J. Immunol.* 1993. V. 151. P. 6135–6142.
- Li B., Carey M., Workman J.L. The role of chromatin during transcription // *Cell.* 2007. V. 128. P. 707–719.
- Masuda E.S., Tokumitsu H., Tsuboi A. *et al.* The granulocyte macrophage colony stimulating factor promoter cis acting element CLE0 mediates induction signals in T cells and is recognised by factors related to AP1 and NF-AT // *Mol. Cell. Biol.* 1993. V. 13. P. 7399–7407.
- McCaffrey P.G., Luo C., Kerppola T.K. *et al.* Isolation of the cyclosporin sensitive T cell transcription factor NFATp // *Science.* 1993. V. 262. P. 750–754.
- McKenzie A.N., Li X., Largaespada D.A. *et al.* Structural comparison and chromosomal localization of the human and mouse IL 13 genes // *J. Immunol.* 1993. V. 150. P. 5436–5444.
- Mechta F., Piette J., Hirai S.I. *et al.* Stimulation of protein kinase C or protein kinase A mediated signal transduction pathways shows three modes of response among serum inducible genes // *New Biol.* 1989. V. 1. P. 297–304.
- Miller C., Zhang M., He Y. *et al.* Transcriptional induction of cyclooxygenase 2 gene by okadaic acid inhibition of phosphatase activity in human chondrocytes: co stimulation of AP 1 and CRE nuclear binding proteins // *J. Cell. Biochem.* 1998. V. 69. P. 392–413.
- Mordvinov V.A., Peroni S.E., De Boer M.L. *et al.* A human T cell line with inducible production of interleukins 5 and 4. A model for studies of gene expression // *J. Immunol. Methods.* 1999a. V. 228. P. 163–168.
- Mordvinov V.A., Schwenger G.T., Fournier R. *et al.* Binding of YY1 and Oct1 to a novel element that downregulates expression of IL-5 in human T cells // *J. Allergy Clin. Immunol.* 1999b. V. 103. P. 1125–1135.
- Mordvinov V.A., Sanderson C.J. Regulation of IL-5 expression // *Arch. Immunol. Ther. Exp. (Warsz).*

2001. V. 49. P. 345–351.
- Mosmann T.R., Cherwinski H.M., Bond M.W. *et al.* Two types of murine helper T cell clones. I. Definition according to profiles of lymphokine activity and secreted proteins // *J. Immunol.* 1986. V. 136. P. 2348–2357.
- Mosmann T.R., Coffman R.L. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties // *Annu. Rev. Immunol.* 1989. V. 7. P. 145–173.
- Mygind N., Andersson M. Topical glucocorticosteroids in rhinitis: clinical aspects // *Acta Otolaryngol.* 2006. V. 126. P. 1022–1029.
- Nakamura Y., Hoshino M. TH2 cytokines and associated transcription factors as therapeutic targets in asthma // *Curr. Drug Targets Inflamm. Allergy.* 2005. V. 4. P. 267–270.
- Naora H., Altin J.G., Young I.G. TCR-dependent and -independent signaling mechanisms differentially regulate lymphokine gene expression in the murine T helper clone D10.G4.1 // *J. Immunol.* 1994. V. 152. P. 5691–5702.
- Naora H., Young I.G. Comparison of the mechanisms regulating IL-5, IL-4, and three other lymphokine genes in the Th2 clone D10.G4.1 // *Exp. Hematol.* 1995. V. 23. P. 597–602.
- Narlikar G.J., Fan H.Y., Kingston R.E. Cooperation between complexes that regulate chromatin structure and transcription // *Cell.* 2002. V. 108. P. 475–487.
- Ngoc P.L., Gold D.R., Tzianabos A.O. *et al.* Cytokines, allergy, and asthma // *Curr. Opin. Allergy Clin. Immunol.* 2005. V. 5. P. 161–166.
- Okayama Y., Petit-Frère C., Kassel O. *et al.* IgE-dependent expression of mRNA for IL-4 and IL-5 in human lung mast cells // *J. Immunol.* 1995. V. 155. P. 1796–1808.
- Ouyang W., Ranganath S.H., Weindel K. *et al.* Inhibition of Th1 development mediated by GATA-3 through an IL-4 independent mechanism // *Immunity.* 1998. V. 9. P. 745–755.
- Paul C.C., Keller J.R., Armpriester J.M., Baumann M.A. Epstein-Barr virus transformed B lymphocytes produce interleukin-5 // *Blood.* 1990. V. 75. P. 1400–1403.
- Pfeuffer I., Klein-Hessling S., Heinfling A. *et al.* Octamer factors exert a dual effect on the IL-2 and IL-4 promoters // *J. Immunol.* 1994. V. 153. P. 5572–5585.
- Ponta H., Cato A.C., Herrlich P. Interference of pathway specific transcription factors // *Biochim. Biophys. Acta.* 1992. V. 1129. P. 255–261.
- Porter B.O., Malek T.R. Thymic and intestinal intraepithelial T-lymphocyte development are each regulated by the gammac-dependent cytokines IL-2, IL-7, and IL-15 // *Semin. Immunol.* 2000. V. 12. P. 465–574.
- Prieschl E.E., Gouilleux-Gruart V., Walker C. *et al.* A nuclear factor of activated T cell-like transcription factor in mast cells is involved in IL-5 gene regulation after IgE plus antigen stimulation // *J. Immunol.* 1995. V. 154. P. 6112–6119.
- Quintana A., Griesemer D., Schwarz E.C., Hoth M. Calcium-dependent activation of T-lymphocytes // *Pflugers Arch.* 2005. V. 450. P. 1–12.
- Ray A., Prefontaine K.E. Physical association and functional antagonism between the p65 subunit of transcription factor NF-kappa B and the glucocorticoid receptor // *Proc. Natl Acad. Sci. USA.* 1994. V. 91. P. 752–756.
- Rincon M., Flavell R.A. AP 1 transcriptional activity requires both T cell receptor mediated and co stimulatory signals in primary T lymphocytes // *EMBO J.* 1994. V. 13. P. 4370–4381.
- Rogatsky I., Ivashkiv L.B. Glucocorticoid modulation of cytokine signaling // *Tissue Antigens.* 2006. V. 68. P. 1–12.
- Romagnani S. Regulation of the T cell response // *Clin. Exp. Allergy.* 2006. V. 36. P. 1357–1366.
- Rose-John S., Schooltink H. Cytokines are a therapeutic target for the prevention of inflammation-induced cancers // *Recent Results Cancer Res.* 2007. V. 174. P. 57–66.
- Roufosse F., Goldman M., Cogan E. Hypereosinophilic syndrome: lymphoproliferative and myeloproliferative variants // *Semin. Respir. Crit. Care Med.* 2006. V. 27. P. 158–170.
- Rudensky A.Y., Gavin M., Zheng Y. FOXP3 and NFAT: partners in tolerance // *Cell.* 2006. V. 126. P. 253–256.
- Ryan J.J., Kashyap M., Bailey D. *et al.* Mast cell homeostasis: a fundamental aspect of allergic disease // *Crit. Rev. Immunol.* 2007. V. 27. P. 15–32.
- Salerno M.S., Schwenger G.T., Sanderson C.J., Mordvinov V.A. Binding of octamer factors to the murine IL-5 CLE0 in primary T-cells and a T-cell line // *Cytokine.* 2001. V. 15. P. 4–9.
- Sanderson C.J. Eosinophil differentiation factor (interleukin-5) // *Immunol. Ser.* 1990. V. 49. P. 231–256.
- Sanderson C.J. IL-5, eosinophils and disease // *Blood.* 1992. V. 79. P. 3101–3109.
- Sanderson C.J., De Boer M., Senna Salerno M. *et al.* Regulation of interleukin 5 gene expression // *Interleukin-5: From Molecule to Drug Target for Asthma* / Ed. C.J. Sanderson. N.Y.: Marcel Dekker, Inc., 1999. V. 125. P. 267–287.
- Sanderson C.J., O'Garra A., Warren D.J., Klaus G.G. Eosinophil differentiation factor also has B-cell growth factor activity: proposed name interleukin 4 // *Proc. Natl Acad. Sci. USA.* 1986. V. 83. P. 437–440.
- Sanderson C.J., Urwin D. Interleukin-5: a drug target



- for allergic diseases // *Curr. Opin. Investig. Drugs*. 2000. V. 1. P. 435–441.
- Sanderson C.J., Warren D.J., Strath M. Identification of a lymphokine that stimulates eosinophil differentiation *in vitro*. Its relationship to interleukin 3, and functional properties of eosinophils produced in cultures // *J. Exptl. Med.* 1985. V. 162. P. 60–74.
- Santana M.A., Rosenstein Y. What it takes to become an effector T cell: the process, the cells involved, and the mechanisms // *J. Cell Physiol.* 2003. V. 195. P. 392–401.
- Schimpl A., Berberich I., Kneitz B. *et al.* IL-2 and autoimmune disease // *Cytokine Growth Factor Rev.* 2002. V. 13. P. 369–378.
- Schwenger G.T., Fournier R., Kok C.C. *et al.* GATA-3 has dual regulatory functions in human interleukin-5 transcription // *J. Biol. Chem.* 2001. V. 276. P. 48502–48509.
- Schwenger G.T., Kok C.C., Arthaningtyas E. *et al.* Specific activation of human interleukin-5 depends on *de novo* synthesis of an AP-1 complex // *J. Biol. Chem.* 2002. V. 277. P. 47022–47027.
- Schwenger G.T., Sanderson C.J. New directions in understanding interleukin-5 gene expression // *Leuk. Lymphoma*. 1998. V. 28. P. 443–450.
- Smith J.B., Herschman H.R. Targeted identification of glucocorticoid-attenuated response genes: *in vitro* and *in vivo* models // *Proc. Am. Thorac. Soc.* 2004. V. 1. P. 275–281.
- Spahn J.D., Szefer S.J. Steroid therapy for asthma in children // *Curr Opin Pediatr.* 2007. V. 19. P. 300–305.
- Sporn M.B., Roberts A.B. Peptide growth factors are multifunctional // *Nature*. 1988. V. 332. P. 217–219.
- Stark J., Chan C., George A.J. Oscillations in the immune system // *Immunol. Rev.* 2007. V. 216. P. 213–231.
- Stranick K.S., Zambas D.N., Uss A.S. *et al.* Identification of Transcription Factor Binding Sites Important in the Regulation of the Human Interleukin-5 Gene // *J. Biol. Chem.* 1997. V. 272. P. 16453–16465.
- Swain S.L., McKenzie D.T., Dutton R.W. *et al.* The role of IL4 and IL5: characterization of a distinct helper T cell subset that makes IL4 and IL5 (Th2) and requires priming before induction of lymphokine secretion // *Immunol. Rev.* 1988. V. 102. P. 77–105.
- Takatsu K., Tominaga A., Harada N. *et al.* T cell-replacing factor (TRF)/interleukin 5 (IL-5): molecular and functional properties // *Immunol. Rev.* 1988. V. 102. P. 107–135.
- Takatsu K., Tominaga A., Harada N. *et al.* The activity of the human interleukin-5 conserved lymphokine element 0 is regulated by octamer factors in human cells // *Eur. J. Biochem.* 1999. V. 265. P. 300–307.
- Taylor P.E., Jacobson K.W., House J.M., Glovsky M.M. Links between pollen, atopy and the asthma epidemic // *Int. Arch. Allergy Immunol.* 2007. V. 144. P. 162–170.
- The Cytokine Handbook / Eds A.W. Thomson, M.T. Lotze. 4th ed. London. UK: Academic Press, 2003.
- Thomas M.A., Mordvinov V.A., Sanderson C.J. The activity of the human interleukin-5 conserved lymphokine element 0 is regulated by octamer factors in human cells // *Eur. J. Biochem.* 1999. V. 265. P. 300–307.
- Trivedi S.G., Lloyd C.M. Eosinophils in the pathogenesis of allergic airways disease // *Cell. Mol. Life Sci.* 2007. V. 64. P. 1269–1289.
- Umetsua D.T., DeKruyffa R.H. Immune dysregulation in asthma // *Curr. Opin. Immunol.* 2006. V. 18. P. 727–732.
- Underwood S.L., Raeburn D., Lawrence C. *et al.* RPR 106541, a novel, airways-selective glucocorticoid: effects against antigen-induced CD4+ T lymphocyte accumulation and cytokine gene expression in the Brown Norway rat lung // *Br. J. Pharmacol.* 1997. V. 122. P. 439–446.
- Van Straaten J.F., Dokter W.H., Stulp B.K., Vellenga E. The regulation of interleukin 5 and interleukin 3 gene expression in human T cells // *Cytokine*. 1994. V. 6. P. 229–234.
- Vilcek J., Feldmann M. Historical review: Cytokines as therapeutics and targets of therapeutics // *Trends Pharmacol. Sci.* 2004. V. 25. P. 201–209.
- Vitetta E.S., Brooks K., Chen Y.W. *et al.* T-cell-derived lymphokines that induce IgM and IgG secretion in activated murine B cells // *Immunol. Rev.* 1984. V. 78. P. 137–157.
- Voehringer D. Allergies and parasites // *Dtsch. Med. Wochenschr.* 2007. V. 132. P. 991–994.
- Walter M. J., Holtzman M.J. A centennial history of research on asthma pathogenesis // *Amer. J. Respir. Cell. Mol. Biol.* 2005. V. 32. P. 483–489.
- Wang J., Young I.G. Eosinophilic inflammation: mechanisms regulating IL-5 transcription in human T lymphocytes // *Allergy*. 2007. V. 62. P. 1131–1138.
- Ward R.C., Kaufman H.L. Targeting costimulatory pathways for tumor immunotherapy // *Int. Rev. Immunol.* 2007. V. 26. P. 161–196.
- Warren H.S., Kinnear B.F., Phillips J.H., Lanier L.L. Production of IL-5 by human NK cells and regulation of IL-5 secretion by IL-4, IL-10, and IL-12 // *J. Immunol.* 1995. V. 154. P. 5144–5152.
- Warren D.J., Sanderson C.J. Production of a T-cell hybrid producing a lymphokine stimulating eosinophil differentiation // *Immunology*. 1985. V. 54. P. 615–623.
- Wetzel S.A., Parker D.C. MHC transfer from APC to T cells following antigen recognition // *Crit. Rev. Immunol.* 2006. V. 26. P. 1–21.
- Willman C.L., Sever C.E., Pallavicini M.G. *et al.* Deletion of IRF-1, mapping to chromosome 5q31.1,



- in human leukemia and preleukemic myelodysplasia // *Science*. 1993. V. 259. P. 968-971.
- Wu C. Chromatin remodeling and the control of gene expression // *J. Biol. Chem.* 1997. V. 272. P. 28171–28174.
- Wu C., Liu F., Zhou X. *et al.* Effect of protein kinase C on proliferation and apoptosis of T lymphocytes in idiopathic thrombocytopenic purpura children // *Cell. Mol. Immunol.* 2005. V. 2. P. 197–203.
- Yamaguchi N. T cell-replacing factor (TRF)/interleukin 5 (IL-5): molecular and functional properties // *Immunol. Rev.* 1988. V. 102. P. 107–135.
- Yang-Yen H.F., Chambard J.C., Sun Y.L. *et al.* Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein–protein interaction // *Cell*. 1990. V. 62. P. 1205–1215.
- Yarilin A.A., Belyakov I.M. Cytokines in the thymus: production and biological effects // *Curr. Med. Chem.* 2004. V. 11. P. 447–464.
- Zhang D.H., Yang L., Cohn L. *et al.* Inhibition of allergic inflammation in a murine model of asthma by expression of a dominant negative mutant of GATA 3 // *Immunity*. 1999. V. 11. P. 473–482.
- Zheng W., Flavell R.A. The transcription factor GATA 3 is necessary and sufficient for Th2 cytokine gene expression in CD4 T cells // *Cell*. 1997. V. 89. P. 587–596.
- Zhu J., Yamane H., Cote-Sierra J. *et al.* GATA-3 promotes Th2 responses through three different mechanisms: induction of Th2 cytokine production, selective growth of Th2 cells and inhibition of Th1 cell-specific factors // *Cell Res.* 2006. V. 16. P. 3–10.

## CYTOKINES: BIOLOGICAL PROPERTIES AND REGULATION OF HUMAN INTERLEUKIN-5 GENE EXPRESSION

V.A. Mordvinov, D.P. Furman

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: mordvin@bionet.nsc.ru

### Summary

Interleukin-5 (IL-5) is a hormone in a large family of cytokines, signaling molecules produced by immune system cells, basically mature T-lymphocytes. IL-5 has the central role in the development and regulation of pathological conditions associated with eosinophilia, including allergic inflammation and parasitic infection. This effort provides an input on cytokines in general, a review of research data on the biological role of IL-5, an overview of various aspects of the structural and functional organization of the IL-5 gene and the regulation of its expression, and also a discussion of the mechanisms by which glucocorticoids reduce eosinophilia and IL-5 gene expression.

## DNA – «PROGRAMMING LANGUAGE OF LIFE»

R. Hofestädt

Bielefeld University, e-mail: hofestae@techfak.uni-bielefeld.de

During the last decades methods of molecular biology are able to identify and sequence different gene functional units (DNA-units). Most of these functional units are analysed in syntax (sequence) and semantic (metabolic function). This kind of information is represented by different molecular databases and information systems, which are distributed across the world (internet). Based on this knowledge it is now possible to discuss the old and still open question if DNA can be interpreted as a programming language. In this paper we will show that the DNA can be interpreted as a programming language in the sense of computer science.

**Key words:** DNA functional units, formal language, programming language, data, control instruction, basic instruction, operon, gene regulation, language of life and synthetic biology.

### Introduction

Regarding cellular processes two different and fundamental levels of genetic languages can be identified. The polypeptide level is coded by 20 different amino acids, which represent the alphabet of this formal language, and the polynucleotide level represented by different nucleic acids (A,T,G,C {U}). The fundamental level is the polynucleotide level with it's DNA, which represents the functional units of the metabolism. Therefore, the alphabet of the fundamental language is given by  $X = \{A,T,G,C (U)\}$  and the DNA is a word over this alphabet. Furthermore, specific "programming language of life" is  $L \subseteq X^*$  (where \* is the star operator and defines all words over X). During the 60ties Ratner (Ratner, 1977) presented the idea of a genetic language that represents different levels: codon, cistron, scripton, replicon, segregon and genome. This paper will focus on the level of functional units of the DNA, which is defined by Ratners cistron level. The idea of our interpretation will be based on two assumptions. The first is that for bacteria and virus most of the DNA/RNA-units are known today. The second is that most of the known DNA-units are universal, which means that besides the universality of the genetic code the universality of DNA-units, which includes the universality of the DNA language, can be suggested. Additionally, this paper will show that analyzed

DNA-units can be interpreted as a programming language. We will present this interpretation in two steps: 1) the specification of the DNA-units in syntax (nucleotide sequence) and semantic (function) and 2) the proof that the DNA-units represent fundamental mechanisms of a computer programming language.

### DNA-units

At the beginning we have to discuss the question if most of the relevant DNA-units are already analyzed. Regarding the sequenced and analyzed genomes including latest molecular knowledge we can assume:

1. Most of the DNA-units are known (Knippers, 2006).
2. For most of these units the function is known and seems to be universal.

The syntax of the DNA-units can be specified using specific Chomsky-type-2 grammars as shown in (Hofestädt, 2007). Similar to the computer programming languages the semantic is described informal using the natural language. A list of the most relevant DNA-units is shown in Table 1.

### Basic features – programming language

Regarding the v. Neumann computer system architecture (Burks *et al.*, 1946), which is still the

Table 1

## Selection of relevant DNA-units (Knippers, 2006)

DNA-unit	Comment
Intron	sub-sequence of the structure gene
Exon	sub-sequence of the structure gene
Leader	sub-sequence of specific structure gene
Structure gene	input sequence for the protein synthesis process
Spacer	separator sequence of genes
Palindrome	antidromic DNA sequence
Terminator	end of the transcriptional unit
Pribnow-Box	subsequence of the promoter to specify promoter affinity
Promoter	start point of the transcriptional unit
Operator	sequence of the gene regulation process
Regulator	specific structure gene
Operon	unit of the protein synthesis process
Telomer	specific sequence of the end of the chromosome
Origin	startpoint of the DNA-Polymerase
Segregon	heredity unit
IS-Element	dynamical structure of the genome
Transposon	dynamical structure of the genome
Virus-DNA-RNA	dynamical structure of the genome
Enhancer	controls the promoter affinity
Overlapping gene	specific structure gene
Homeotic gene	operon, which shows the modularity of the genome

kernel of our computer, we are able to detect the following fundamental features of a programming language:

F1. **Data type** (at least one is sufficient): Computer instructions can modify data so that at least one data type (e. g. integer) must be available.

F2. **Instruction**: Each computer/computer language is offering a set of instructions, which can modify data (e. g. add, multiply, and etc.).

F3. **Control instruction**: Specific instruction which controls the order of the next executable instruction of the program.

F4. **Punctuation mark**: Begin and end symbol (word) is defined.

Besides the numeric/logical instructions, which will modify the program data, the control instructions are fundamental. Regarding all control instructions we can differentiate between three fundamental classes:

**C1: Composition**  $S_1; S_2; \dots; S_n$

The semicolon denotes the following operator. The semantic of this operator is that instruction ( $S_{i+1}$ ) will be executed after execution of instruction ( $S_i$ ).

**C2: If-Instruction (If B then S)**

Let S be an instruction and B a condition, which can be true/false. Instruction S will be executed if and only if condition B is true – otherwise S will not be executed.

**C3: While-instruction (While B do S)**

Let S be an instruction and B a condition, which can represent the value true/false. The meaning is that the instruction S will be executed as long as B is true.

The theoretical model of the v. Neumann architecture is the so-called Turing machine (Hopcroft, Ulman, 1979), which belongs to the class of the universal computational concepts. That means each problem, which is intuitive computable, can be solved using the adequate Turing machine

(v. Neumann computer). Regarding the DNA-units, which are shown in Table 1, we can define the basic instruction of the DNA language and the activation of this instruction. A DNA-unit is called basic instruction, if at least one promoter and one terminator are included. If a **basic instruction** will pass the biosynthetic process, this is called **activation** of the basic instruction.

An operon can be interpreted as a basic instruction, which includes structure genes, operator genes, one or more promoter genes and a terminator gene. The activation of a basic instruction includes fundamental metabolic processes like transcription and translation (Knippers, 2006). It is similar to the instruction execution process of a computer system.

### Interpretation

Assume that the DNA is the genetic program of the cell. In that case the cytoplasm can be interpreted as the data type which represents metabolites. Metabolites can be modified by different biochemical reactions. Therefore, we can assume that the data type (metabolite) is available (see F1). Enzymes can catalyze biochemical substances so that a substrate will be modified into a product catalysed by a specific enzyme. Therefore, instructions are chemical reactions

caused by enzymes, which are presented by structure genes (see F2). Furthermore, structure genes are controlling the metabolism indirectly. Regarding specific cells we can see specific genes, which are active during specific time periods. This behaviour shows that specific DNA-units control the activity of genes, which can be interpreted as the control instruction (see F3). Finally, the DNA-unit, which is called telomer, can be interpreted as the punctuation mark of this system (see F4). Defining and regarding specific operons it is possible to show that the control instructions C1 – C3 can be simulated by gene controlled regulatory networks. Composition of basic instructions can be interpreted as a sequence of basic instructions represented by structure genes or operons separated by spacer units (Fig. 1).

Furthermore, we can show that an operon can be interpreted as an If-instruction (see C2). Therefore, we focus on the operon L14 of *E. coli*, which regulates its own synthesis. The mechanism of this regulation process can be illustrated (Fig. 2).

Under this interpretation the boolean value of condition B will be specified by the state of the operator, which can be true (*Operator\_X* gene is free) or false (*Operator\_X* is blocked by the repressor). Under this interpretation the instruction **If B then S** is simulated, because the operon will

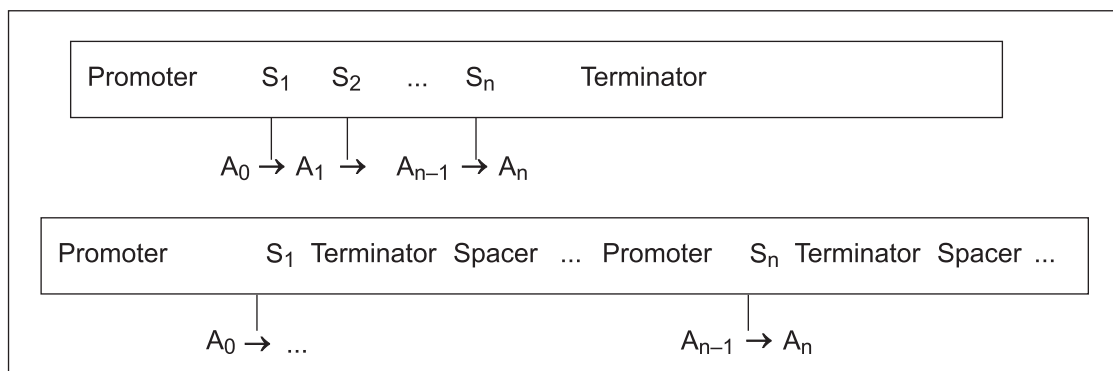


Fig. 1. Composition of basic instructions of the DNA language.

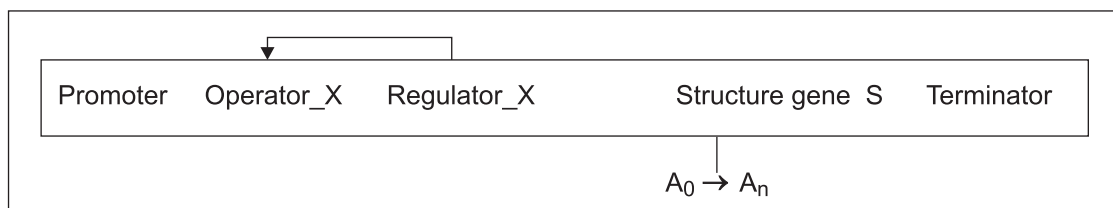


Fig. 2. Abstract representation of the operon L14 (Knippers, 2006), which can be interpreted as the If-instruction.

be blocked by itself after activation of this basic instruction (operon). The synthesis of *Structure gene S* will realise the instruction S and the synthesis of *Regulator\_X* will block the synthesis of operon L14. Regarding the example of the illustrated operon L14 the theoretical extension to the **While instruction** can be realised. Deleting the *Regulator\_X* gene, which is inside of the operon L14, will produce this effect. Therefore, we will discuss the structure and function of the *Tryptophan* operon, which shows this effect (Fig. 3).

Regarding the *Tryptophan* operon, we can see a composition of structure genes (S) and the boolean value of condition B can be true (Operator is free) or false (Operator is blocked).

As long as the operon represents the state true the basic instruction is activated. It will be activated until the operator gene will be blocked. This interpretation shows the «While operator» (see C3). Furthermore, the telomere sequence (chromosome) and the terminator/promoter sequence (cistron level) can be interpreted as the begin- and end-mark (F4).

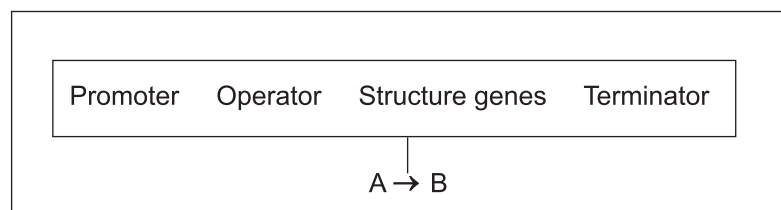
### Features of DNA-languages

Regarding the semantic of the analyzed DNA-units we can identify the following features of this language. First of all the genome is modular organized like most of the programming languages. The modular organization can be shown regarding the semantic of the homeotic genes (Watson *et al.*, 1985). These genes are controlling a battery of genes and can be interpreted as a function, procedure or module in the sense of programming languages. Regarding the semantic of the promoter we can see that the basic element will be activated based on a probability value, which is defined by the specific sequence of the promoter sequence. These sequences specify the so-called promoter affinity of the operon (Knippers, 2006). One more interesting and complex

feature is the dynamic behavior of DNA-units, which is shown for example by transposons and IS-elements (Watson *et al.*, 1985). The semantic of transposons and IS-elements is that DNA sequences can change their localization inside the chromosome (program). Furthermore, changing the localization other DNA-units can be destroyed or modified. Until now the semantic of these dynamic DNA-units is not known exactly. Parallelization is one more feature of molecular processing. The opposite to von Neumann computer basic molecular instructions can be activated in parallel. This concept is called dataflow concept and the semantic is that each basic instruction will be executed if all biochemical conditions are satisfied (operator genes are free and RNA-polymerase is available etc.). Regarding these features we can say the programming language of life is much more complex than a v. Neumann computer language. Regarding theoretical models of computation discussed in theoretical computer science we can see parallel machines (Fortune, Wyllie, 1978), probabilistic Turing machines (Gill, 1977), and hardware modification machines (Cook, 1980). However, there is no discussion about a dynamic-parallel-probabilistic-modular Turing machine model.

### Summary

The key idea of this paper is to show that the DNA can be interpreted as a programming language based on the level of analyzed functional DNA-units. Therefore, we presented a subset of well-known DNA-units and extracted the main features of a von Neumann programming language. Our paper shows that DNA-units can be interpreted as a programming language based on the level of DNA-units, which was called cistron level by the fundamental definition of Ratner (Ratner, 1977). Therefore, metabolites can be interpreted



**Fig. 3.** The *Tryptophan* operon (Knippers, 2006) represents: Promoter, terminator, operator and a composition of structure genes: *trpE*, *trpD*, *trpC*, *trpB*, *trpA*.



as the fundamental data type and operons which synthesize enzymes that can be interpreted as data modifying instructions. Specific operons can also be interpreted to simulate the fundamental computer instructions as: composition, If- and While-statements. Overall section 4 showed that the fundamental structures of a programming language are represented by basic DNA-units. Therefore, the DNA-units represent the fundamental language of life. Further discussions showed that this language is representing complex language structures as features of parallel, probabilistic, dynamic and modular computing. Regarding computational models of theoretical computer science we can see no model, which represents such a complex computational mechanism today.

Thinking about the post-genome era it is the main idea of this work to study and understand this kind of complex languages (computational methods), which we could identify in this paper. Based on this work and the understanding of such methods and concepts it will be possible to specify the language of life in detail, which will represent the framework for the realization of the ideas of synthetic biology.

## References

- Burks A., Goldstine H., Neumann von J. Preliminary Discussion of the Logical Design of an Electronical Computing Instrument // Research Report. Institute of Advanced Study, Princeton, 1946.
- Cook S. Towards a complexity theory of synchronous parallel computation // L'Enseignement Mathematique. 1980. 27. P. 99–124.
- Fortune S., Wyllie J. Parallelism in Random Access Machines // Proc. 10th ACM Symp. on Theory and Computing, 1978. P. 114–118.
- Gill J. Computational complexity of probabilistic Turing Machines // SIAM J. of Computing. 1977. 6. P. 675–695.
- Hofestädt R. Extended Backus-System for the representation and specification of the genome // J. of Bioinformatics and Computational Biol. 2007. 5-2(b). P. 457–466.
- Hopcroft J.E., Ullman J.D. Automate Theory, Languages, and Computation, Addison-Wesley Publ. Co., London – Amsterdam Don Mills, Ontario – Sydney 1979.
- Knippers R. Molekulare Genetik, Stuttgart; N.Y.: Georg Thieme Verlag, 2006.
- Ratner V. Molekulargenetische Steuerungssysteme. Stuttgart: Gustav Fischer Verlag, 1977.
- Watson J.D., Tooze J., Kurtz D. Rekombinierte DNA. Spektrum der Wissenschaft, Heidelberg 1985.

## DOMAIN – WIDE LANDSCAPE OF HUMAN GENOME

**D.A. Maximov, V.N. Babenko**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: bob@bionet.nsc.ru

Two facts are currently inferred in the course of eukaryotic genome investigations. The first finding is that the highly expressed genes in eukaryotes maintain short introns (Petrov *et al.*, 1998; Castillo-Davis *et al.*, 2002). The second observation is that there is domain-wide regulation of gene expression in human, which comprises regions of ~80–90 genes per domain on average, exhibiting a particular level of integral expression (Gierman *et al.*, 2007; Huvet *et al.*, 2007). In this work we analyzed the features of genes in regard to the domains identified in the papers mentioned above.

We have found that there are 2 distinct groups of genes in low expressed domains. One contains extremely long genes. We observed expression preference in the brain tissue for them. Another group comprises short genes featured as cluster-structured gene loci with various activities, including testis specific and liver-specific genes.

**Key words:** human, chromatin, replication start, gene structure, gene length, testis-specific genes.

### Introduction

It's long appreciated that there is a higher order transcriptome regulation on the level of chromatin state and its exploration is going on (Schübeler, 2007). The range of authors currently incline that the transcriptional domains are regulated to a large extent by histone modification. There is an extensive volume of papers devoted to the subject (Schübeler, 2007), but the locations and roles of histone modifications elsewhere in the genome remain unclear (Heintzman *et al.*, 2007).

Analysis of recent experimental timing data (Woodfine *et al.*, 2005) confirmed that, in a number of cases, domain borders coincide with replication initiation zones active in the early S phase (Yurov, Liapunova, 1977), whereas the center regions replicate in the late S phase (Huvet *et al.*, 2007). Around the putative replication origins, genes are abundant and broadly expressed. These features weaken progressively with the distance from putative replication origins. At the center of domains, genes are rare and expressed in few tissues.

682 successive N-shaped nucleotide compositional skew domains were identified in Huvet *et al.*, 2007. We explored the features of the genes located within the N-shaped regions (N-domains) and

gained some insights on the specific structural traits for them. The genes located within domain center maintain significantly longer intergenic length. There is an abundance of very short and extra long genes which usually implies their intron length. We checked GNF atlas expression resource (Su *et al.*, 2002) to define gene expression in various parts of the domains and found that tissue-specific genes, placenta genes and liver-specific genes are presented in significant abundance in the center of the domains. Flanking genes corresponded to housekeeping genes which is consistent with the previous observations (Huvet *et al.*, 2007). It is worth noting that overall the tissue-specific genes in the center area are considered low expressed (Castillo-Davis *et al.*, 2002) compared with their flank counterparts, but they are quite intense in the terms of timing, e.g. they are highly expressed within short period of time (Huang, Niu, 2008). Genes containing long introns were specifically expressed in the nervous tissue.

We extensively explored testis-specific genes since they exhibit vivid positive selection evolution mode (Kouprina *et al.*, 2004, 2007). The intron length features of the center domain genes sought to be under neutral/disruption selection mode what makes them either extremely long or quite short (Hughes *et al.*, 2008).

## Materials and Methods

Oligonucleotide microarray data were extracted from the Gene Expression Atlas [http://expression.gnf.org, (Su *et al.*, 2002)] that contains 25 human and 45 mouse non-tumoral tissues. The sample replicates corresponding to the same tissue were averaged. The signals of probes corresponding to the same gene were averaged. In total, 7735 different human mRNAs and 5297 mouse mRNAs are represented in the resulting data set. As recommended by the authors (Su *et al.*, 2002), genes whose expression level exceeded 200 arbitrary units were noted as expressed. Microarray data were available for 1276 genes in 22 normal tissues belonging to N-domains. Testis specific, liver specific and placenta specific genes were defined by 3-times larger than average log normalized expression in the corresponding tissue with at most one alternate tissue expression instance.

Gene and intron length profiles were built along the human chromosomes with overlapping window of 1 Mb in size and shift of 10 kb. The number of gene (intron) starts per window was calculated for the gene (intron) profile plot. Each gene was represented by the longest transcript unless mentioned otherwise.

## Results and Discussion

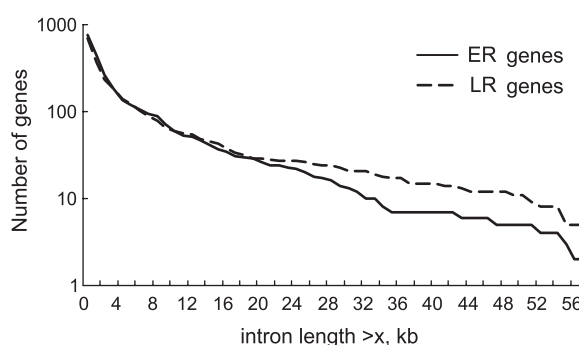
We binned the genes of the 682 N-domain regions into «Late replication (LR)» gene sample as those located within 50 % of domain length in the center of the regions, and the genes comprised in 50 % of the domain territory flanking the center region («Early replication (ER)» gene sample).

Overall gene abundance (gene density) in LR set was significantly lower than in the ER gene set (2184 vs 3274 genes,  $p < 1e-31$  with binomial test).

Next we compared the distribution of genes length within and between the bins (Fig. 1). It was revealed that the genes less than 5 kb in length were abundant in LR sample in comparison with ER sample (Fig. 1,  $p < 0.0001$  with  $\chi^2$  2x2 table test). While overall genes' deficiency in LR region over the lengths of 7–24 kb had been observed, the extra long genes were also abundant in LR set (Fig. 1). Nearly half of the short genes were single exon genes, the abundance was also very high (Table 1). The number of merged genes was preferred in the ER genes.

Testis-specific genes are fast evolving genes possessing several tissue-specific features, namely: a) the vast majority of them are short (Su *et al.*, 2008), e.g. less than 10kb in length; b) they are organized in small clusters containing 1–5 genes which are dispersed along chromosomes (Su *et al.*, 2008). They evolved rapidly (Kouprina *et al.*, 2004, 2007; Thurman *et al.*, 2008), and usually possess rather strong tissue-specific expression level (Su *et al.*, 2008).

It has been shown that chromatin remodeling is a specific epigenetic feature of spermatogenesis (Pradeepa, Rao, 2007; Tachiwana *et al.*, 2008) as well as other tissue specific expression (Thurman *et al.*, 2008). They are present in eukaryotic genomes



**Fig. 1.** Comparison of Late replication and Early replication gene sets. The maximal transcript was taken for each gene. Each point represents number of transcripts possessing the length greater than the abscissa value.

**Table 1**  
Comparison of specific gene numbers  
for a range of features

Replication timing	Number of merged genes comprised within gene loci	Number of single-exon genes	Number of genes total
ER genes	199	128	3274
LR genes	82	211	2184

Given the total number of ER genes and LR genes we assess the  $\chi^2$  2x2 tables test value equal to 12,8,  $p < 0,0002$  to reject equal distribution null hypothesis for merged genes against total gene number. The single exon genes are represented skewed towards abundance in the LR regions as well with  $\chi^2 = 65,0$ ,  $p < 1e-7$ .

from mammals down to insects (Spellman, Rubin, 2002). While relatively constant gene number, the gene content evolves rapidly in *Drosophila* group (Spellman, Rubin *et al.*, 2002).

We calculated the number of testis-specific genes in the LR regions and found it significantly abundant compared to their number in the flanking regions with the p-value  $1e-8$  (785 vs 324 testis-specific genes). Thus we can say that testis-specific genes are attributable to the LR region.

### Conclusion

We sought that the late replication origins are the key regulation units that are responsible for tissue-specific expression. We found the key structural end expression gene features are similar within ER and LR domains. Therefore we may say the domain wide chromatin structural features we observe are quite widespread phenomenon in humans and probably, in all eukaryotic species that modulate stage and tissue expression. We believe that the future studies will gain much more insight on the stage-specific mechanics of such structures.

### Acknowledges

We are grateful to the BGRS 2008 chair committee for providing the opportunity to report and publish this study. We thank the referees for quite useful comments. We also thank Institute of Computational Mathematics and Mathematical Geophysics SB RAS ([www.sccc.ru](http://www.sccc.ru)) for providing computing facilities. This work was partly supported by Integration Grant from Russian Academy of Science «Biodiversity and Genome Dynamics».

### References

- Castillo-Davis C.I., Mekhedov S.L., Hartl D.L. *et al.* Selection for short introns in highly expressed genes // *Nat. Genet.* 2002. V. 31. № 4. P. 415–418.
- Heintzman N.D., Stuart R.K., Hon G. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome // *Nat. Genet.* 2007. V. 3. P. 311–318.
- Huang Y.F., Niu D.K. Evidence against the energetic cost hypothesis for the short introns in highly expressed genes // *BMC Evol. Biol.* 2008. V. 8. P. 154.
- Hughes S.S., Buckley C.O., Neafsey D.E. Complex selection on intron size in *Cryptococcus neoformans* // *Mol. Biol. Evol.* 2008. 25(2). P. 247–253.
- Huvet M., Nicolay S., Touchon M. *et al.* Human gene organization driven by the coordination of replication and transcription // *Genome Res.* 2007. V. 17. № 9. P. 1278–1285.
- Gierman H.J., Indemans M.H., Koster J. *et al.* Domain-wide regulation of gene expression in the human genome // *Genome Res.* 2007. V. 17. № 9. P. 1286–1289.
- Kouprina N., Mullokandov M., Rogozin I.B. *et al.* The SPANX gene family of cancer/testis-specific antigens: rapid evolution and amplification in African great apes and hominids // *Proc. Natl Acad. Sci. USA.* 2004. V. 101. № 9. P. 3077–3082.
- Kouprina N., Noskov V.N., Pavlicek A. *et al.* Evolutionary diversification of SPANX-N sperm protein gene structure and expression // *PLoS ONE.* 2007. V. 2. № 4. P. e359.
- Petrov D.A., Lozovskaya E.R., Hartl D.L. High intrinsic rate of DNA loss in *Drosophila* // *Nature.* 1998. V. 384. P. 346–349.
- Pradeepa M.M., Rao M.R. Chromatin remodeling during mammalian spermatogenesis: role of testis specific histone variants and transition proteins // *Soc. Reprod. Fertil. Suppl.* 2007. V. 63. P. 1-10.
- Schübeler D. Enhancing genome annotation with chromatin // *Nat Genet.* 2007. V. 39. № 3. P. 284–285.
- Spellman P.T., Rubin G.M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome // *J. Biol.* 2002. V. 1(1). P. 5.
- Su A.I., Cooke M.P., Ching K.A. *et al.* Large-scale analysis of the human and mouse transcriptomes // *Proc. Natl Acad. Sci. USA.* 2002. № 99. P. 4465–4470.
- Su W.L., Modrek B., GuhaThakurta D. *et al.* Exon and junction microarrays detect widespread mouse strain- and sex-bias expression differences // *BMC Genomics.* 2008. V. 9. P. 273.
- Tachiwana H., Osakabe A., Kimura H., Kurumizaka H. Nucleosome formation with the testis-specific histone H3 variant, H3t, by human nucleosome assembly proteins in vitro // *Nucl. Acids Res.* 2008. V. 36. № 7. P. 2208–2218.
- Thurman R.E., Noble W.S., Stamatoyannopoulos J.M. Multi-scale correlations in continuous genomic data // *Pac. Symp. Biocomput.* 2008. P. 201–215.
- Woodfine K., Beare D.M., Ichimura K. *et al.* Replication timing of human chromosome 6 // *Cell Cycle.* 2005. V. 4. P. 172–176.
- Yurov Y.B., Liapunova N.A. The units of DNA replication in the mammalian chromosomes: evidence for a large size of replication units // *Chromosoma.* 1977. V. 60. P. 253–267.

## HORIZONTAL TRANSFER OF NON-LTR RETROTRANSPOSONS

O.S. Novikova<sup>1</sup>, V. Fet<sup>2</sup>, A.G. Blinov<sup>1</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: novikova@bionet.nsc.ru;

<sup>2</sup> Marshall University, Huntington, West Virginia, USA

Horizontal transfer is a complex phenomenon usually used as explanation of phylogenetic inconsistency which can not be interpreted in terms of vertical evolution. Most examples of horizontal transfer of eukaryotic genes involve transposable elements. An intriguing feature of horizontal transfer is that its frequency differs among transposable elements' classes. Although horizontal transfer is well known for DNA transposons and LTR retrotransposons, majority of previously described cases of horizontal transfer for non-LTR retrotransposons were dismissed based on the analysis of non-LTR retrotransposon sequence change rates. Nevertheless, recent findings suggest that non-LTR retrotransposons can be horizontally transmitted. Thus, one cannot completely exclude the possibility of horizontal transfer for this group of transposable elements. We review the occurrence of horizontal transfer of non-LTR retrotransposons and examine the criteria used to infer such transfers.

**Key words:** non-LTR retrotransposons, horizontal transfer, evolution, Fungi, Maculinea

### Introduction

Eukaryotic transposable elements (TEs) have the ability to replicate and spread in genome. They have been found virtually in all investigated eukaryotes and represent the ubiquitous components of eukaryotic genomes. For example, at least 46 % of the human genome is represented by TE-derived sequences (Bannert, Kurth, 2004) and more than 90 % of the plants' genomes can be composed by TEs (Vitte, Panaud, 2005). The relative abundance and diversity of TEs have contributed to the differences in the structure and size of eukaryotic genomes, especially in plant kingdom (Vitte, Panaud, 2005). Recent evidence suggests that TEs may provide the genome with potent agents to generate genetic and genomic plasticity (Kidwell, Lisch, 1997). TEs may have reshaped the human genome by ectopic rearrangements, by creating new genes, and by modifying and shuffling existing genes (Lander *et al.*, 2001). In some cases, TEs perform critical biological functions in their host (Kidwell, Lisch, 1997). For example, the preferential insertion of some retrotransposons in *Drosophila* at telomeric locations has removed the need for a telomerase function (Pardue *et al.*, 2005).

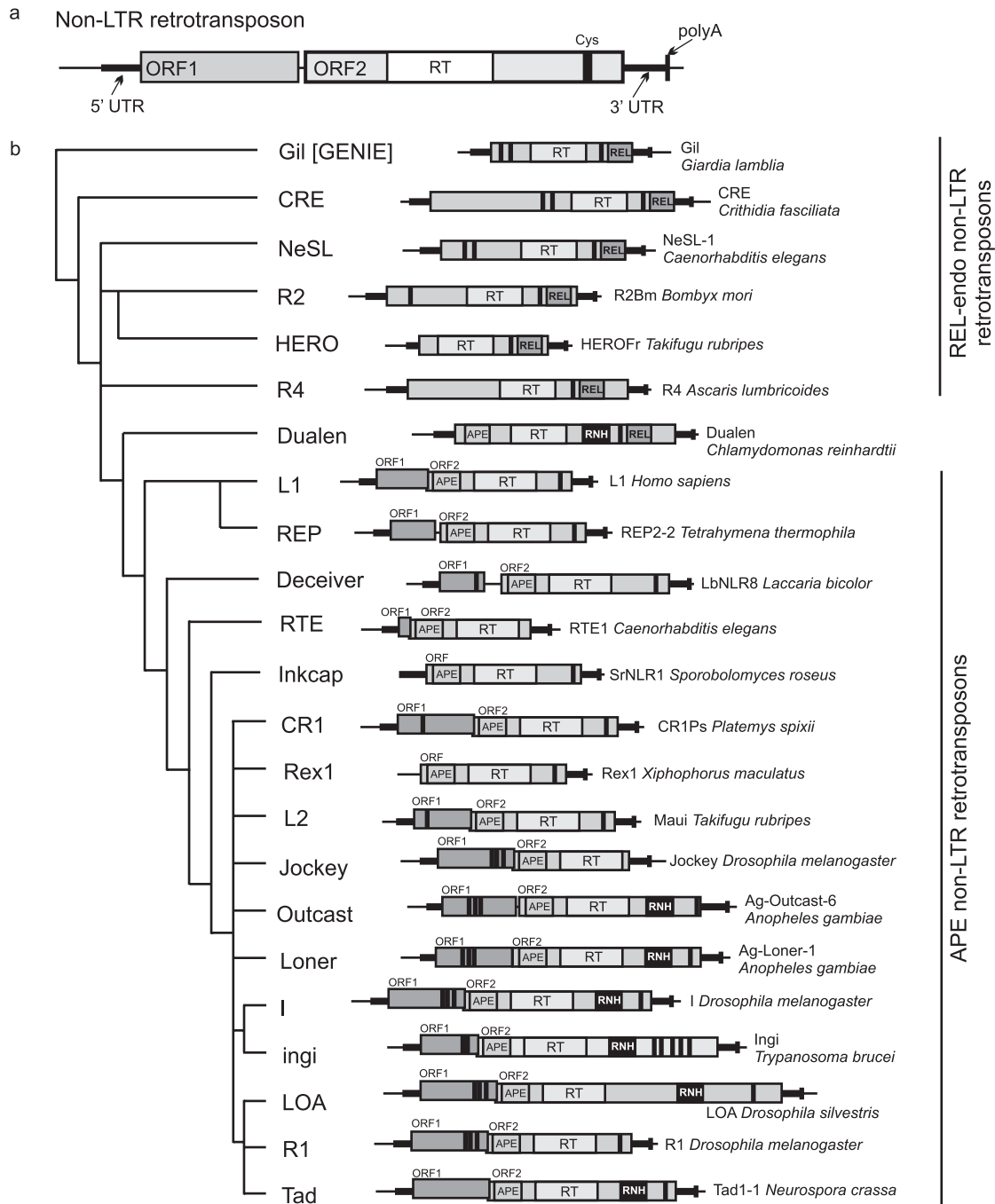
Two major classes of TEs are recognized (Wicker *et al.*, 2007). Class I elements use RNA-mediated mechanisms for their transposition and are called retrotransposons. Class II elements transpose through DNA-mediated mechanisms and are called DNA transposons (Feschotte, Pritham, 2007). Eukaryotic DNA transposons can be divided into three major subclasses. Elements from the first, the best-known subclass, are the classic "cut-and-paste" transposons that excise as double-stranded DNA and reinsert elsewhere in the genome. Elements from the second subclass, *Helitrons*, utilize a mechanism probably related to rolling-circle replication (Kapitonov, Jurka, 2007). For the last subclass, *Mavericks*, mechanism of transposition is not yet well understood, but that likely replicate using a self-encoded DNA polymerase (Pritham *et al.*, 2007). Both *Helitrons* and *Mavericks* most likely rely on distinct transposition mechanisms involving displacement and replication of a single-stranded DNA intermediate, respectively.

Retrotransposons transpose by the process involving transcription, reverse transcription, and integration of cDNA. Five orders of retrotransposons are recognized: LTR retrotransposons, which have



long terminal repeats (LTRs); non-LTR retrotransposons, which lack LTRs; *DIRS* retro-transposons; *Penelope*-like retrotransposons; and short interspersed nuclear elements (SINEs) (Wicker *et al.*, 2007). Non-LTR retrotransposons contain 5' and

3'-untranslated regions (UTR), two or one open reading frame (ORF), preceded by the promoter for RNA polymerase II, and are usually terminated by poly(A) tail (Fig. 1, a). All intact non-LTR retrotransposons encode a reverse transcriptase (RT),



**Fig. 1.** Structure (a) and evolution of non-LTR retrotransposons (b).

Abbreviations: 5'UTR and 3'UTR – 5' and 3' untranslated regions; ORF – open reading frame; RT – reverse transcriptase; REL – restriction-enzyme-like endonuclease; APE – apurinic/apyrimidinic endonuclease; RNH – ribonuclease H.

which provides for their reverse transcription. Additionally, their ORFs encode a restriction-enzyme-like apurinic/apyrimidinic endonuclease, as well as a ribonuclease H (Malik *et al.*, 1999; Wicker *et al.*, 2007; Novikova *et al.*, 2008).

Malik *et al.* (1999) first used the term «clade» for groups of non-LTR retrotransposons, which have high phylogenetic support, share the same structural features, and are estimated to date back to the Precambrian time (older than ~570 Myr). Based on the phylogenetic analysis of RT domains, 23 diverse clades of non-LTR retrotransposons have been described (e.g. Malik *et al.*, 1999; Volff *et al.*, 2000; Biedler, Tu, 2003; Novikova *et al.*, 2008) (Fig. 1, b). The most ancestral clades of non-LTR retrotransposons (GENIE, CRE, R2, NeSL-1, and R4) contain only one ORF and show site-specific distribution in the genomes (Malik *et al.*, 1999; Malik, Eickbush, 2000). They have a restriction-enzyme-like endonuclease (REL-endo) domain. During further evolution of mobile elements, the REL-endo domain is suggested to have been substituted with an apurinic/apyrimidinic (APE) endonuclease acquired from host cells. All younger clades (L1, RTE, Tad, R1, LOA, I, Jockey, CR1, Rex1, and L2) possess the APE endonuclease domain and are called APE retrotransposons (Zingler *et al.*, 2005). The acquisition of the APE endonuclease resulted in a loss of target site specificity for all the elements (except R1 clade and some elements from L1 clade), and coincided with the origin of a second ORF in front of the RT-encoding ORF. Our knowledge of non-LTR retrotransposon diversity is likely to increase further: almost all studies of the entire genomic sequences detect new phylogenetic groups, especially those which have a limited distribution and/or are represented by a few copies per genome (e.g. Biedler, Tu, 2003; Novikova *et al.*, 2008).

### Horizontal transfer of TEs

Horizontal transfer (HT) can be defined as the process by which genes can move between reproductively isolated species. It is not surprising that most examples of horizontal transfer of eukaryotic genes involve TEs (Kidwell, 1992; Hartl *et al.*, 1997). Three criteria could be used for HT event recognition. The first criterion is the inconsistency between the phylogenies of TEs

and host species. Traditionally, horizontal transfer was implied when highly similar transposable elements had been found in distantly related taxa accompanied by their discontinuous distribution, and such phenomenon could not be explained in terms of vertical inheritance (Kidwell, 1992; Hartl *et al.*, 1997). There are potential problems with application of this criterion for HT detection. Multiple TEs lineages can be present within genomes. Moreover, TEs are multicopy components of genomes. Comparisons of paralogous TE copies instead of orthologs along with varying rates of their sequence evolution are the main sources for incongruence in phylogenetic analysis, which could be misidentified as HT.

The second criterion, which seems to offer the strongest evidence, is a higher degree of observed sequence similarity for TEs than for functional genes, so-called 'slowdown effect on evolutionary rates'. Once inserted, a new copy of transposable element is presumed to evolve without functional constraints. Thus, all types of mutations should have an equal chance to be fixed. This means that the rates of synonymous and nonsynonymous nucleotide substitutions are identical, and coding regions in the sequence are disrupted by stop codons and indels (Volff *et al.*, 2000; Bensasson *et al.*, 2001). The lower than expected sequence divergence of TEs in comparison with non-mobile nuclear genes of the host species can be explained either by strong selective constraints in TE sequence coupled with a strict vertical transmission, or by horizontal transfer (Kordiš, Gubenšek, 1998; Novikova *et al.*, 2007; Novikova *et al.*, 2008). It is known that the insertion of a transposable element can alter gene expression and be selectively advantageous. However, only some of such transposable elements evolve under selective pressure (Ono *et al.*, 2001).

The third criterion of inferring HT is the discontinuous distribution of TEs among closely related taxa, i.e., presence of a TE in one lineage and its absence in a sister lineage. Such discontinuous distribution could be due to random loss of TEs, ancestral polymorphism, and independent sorting of copies into descendant species. By itself, this kind of evidence provides only weak support for HT since TE can be lost through population dynamics or ecological forces that are difficult to reconstruct (Kaplan *et al.*, 1985; Lohe *et al.*, 1995).

### Horizontal transfer of non-LTR retrotransposons

While HT is well known for DNA transposons and LTR retrotransposons (Robertson, 1993; Silva, Kidwell, 2004), non-LTR retrotransposons rarely undergo HT, and their phylogenies are largely congruent to those of their hosts (Malik *et al.*, 1999). Moreover, the analysis of TEs sequence change rates led Malik *et al.* (1999) to dismiss majority of previously described HT cases. It has been suggested that non-LTR retrotransposons are rarely, if at all, horizontally transferred and their mode of transposition might be sufficient to explain this rarity since the RNA intermediate of non-LTR retrotransposons is reverse-transcribed directly into the chromosomal target site (Malik *et al.*, 1999). Nevertheless, there are a few cases of putative HT for non-LTR retrotransposons (Table 1).

One of the most reliable cases of HT for non-LTR retrotransposons was described for *CR1B* elements from CR1 clade (Novikova *et al.*, 2007). Initially, TE sequences with extremely high similarity were identified in genomes of silkworm *Bombyx mori* (family Bombycidae) and *Maculinea* large blue butterflies (family Lycaenidae). Further comparative and phylogenetic analysis provided more evidence for possible HT. Analysis of closely related species showed that *CR1B* element presented only in *Maculinea* butterflies and in both studied Bombycidae moths (*Bombyx mori* and *Oberthueria*

*caeca*) (Fig. 2, a). All three criteria were satisfied including unexpectedly high similarity between elements, phylogenetic inconsistency, and discontinuous distribution.

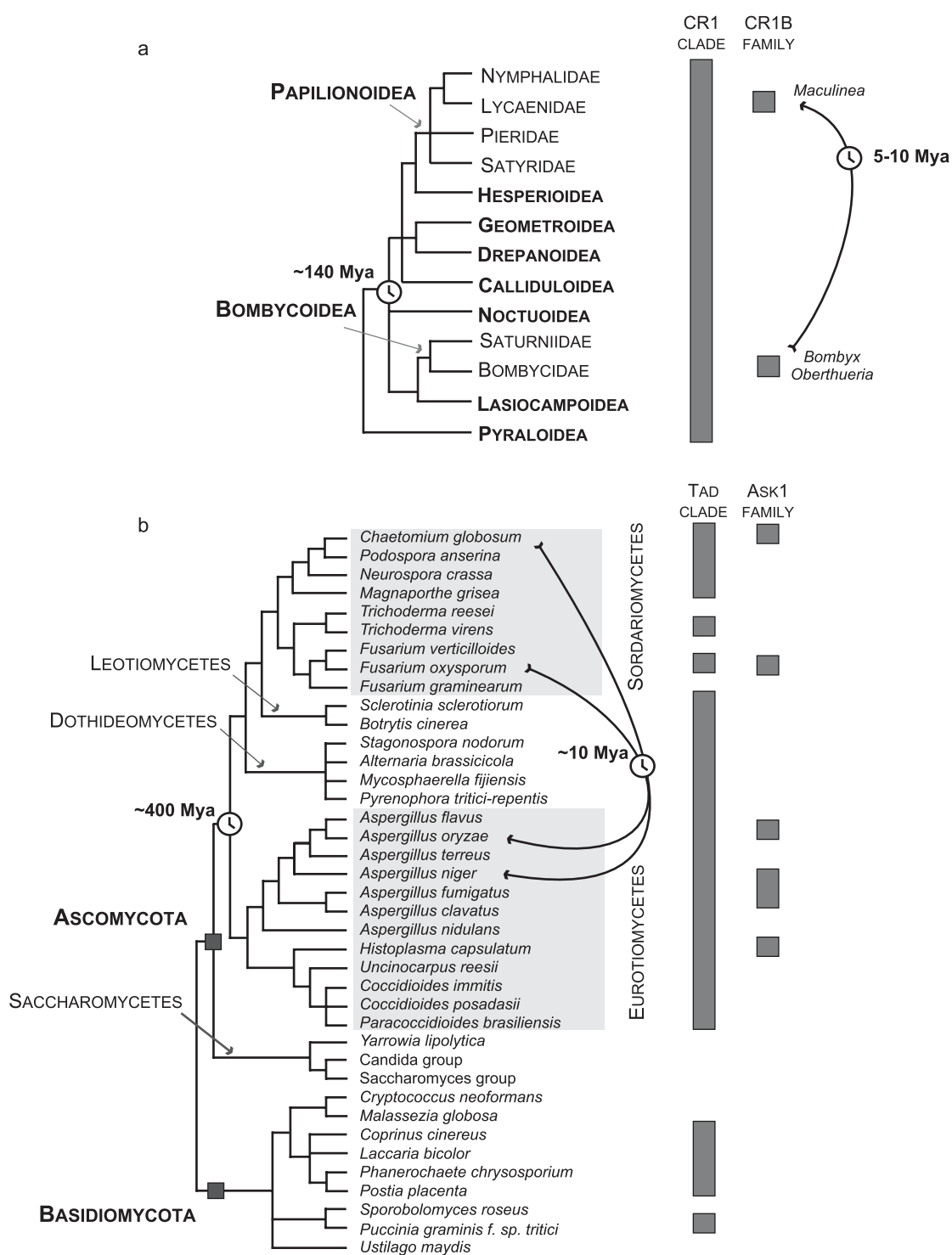
Along with three listed criteria, the 'divergence-versus-age' analysis can be used for HT inference (Kordiš, Gubenšek, 1998; Malik *et al.*, 1999). It includes the comparison of divergence rates between the enzymatic domains of the TEs with the host divergence time estimates. This method is widely used for non-LTR retrotransposon analysis (Kordiš, Gubenšek, 1998; Malik *et al.*, 1999; Novikova *et al.*, 2007; Novikova *et al.*, 2008). Amino acid sequence distances between the reverse transcriptase (RT) domains of various non-LTR retrotransposon clades (R1, R2, Jockey, I, CR1, RTE, Tad and L1) are plotted against estimates of host divergence time (Fig. 3). The last common ancestor of lepidopteran superfamilies Papilionoidea (*Maculinea* spp.) and Bombycoidea (*Bombyx mori*) has been estimated to exist 140 million years ago (Mya) (Gaunt, Miles, 2002). The point for *BmCR1B* (from *B. mori*) versus *MteCR1B* (from *Maculinea teleius*) fell markedly below all other comparisons. It means that the rate of evolution is much lower than expected for *BmCR1B* and *MteCR1B* elements. The slowdown effect on evolutionary rates can be explained by HT event (Kordiš, Gubenšek, 1998; Malik *et al.*, 1999; Župunski *et al.*, 2001; Novikova *et al.*, 2007; Novikova *et al.*, 2008). The HT of *CR1B* elements between lepidopteran taxa is an example of the

**Table 1**

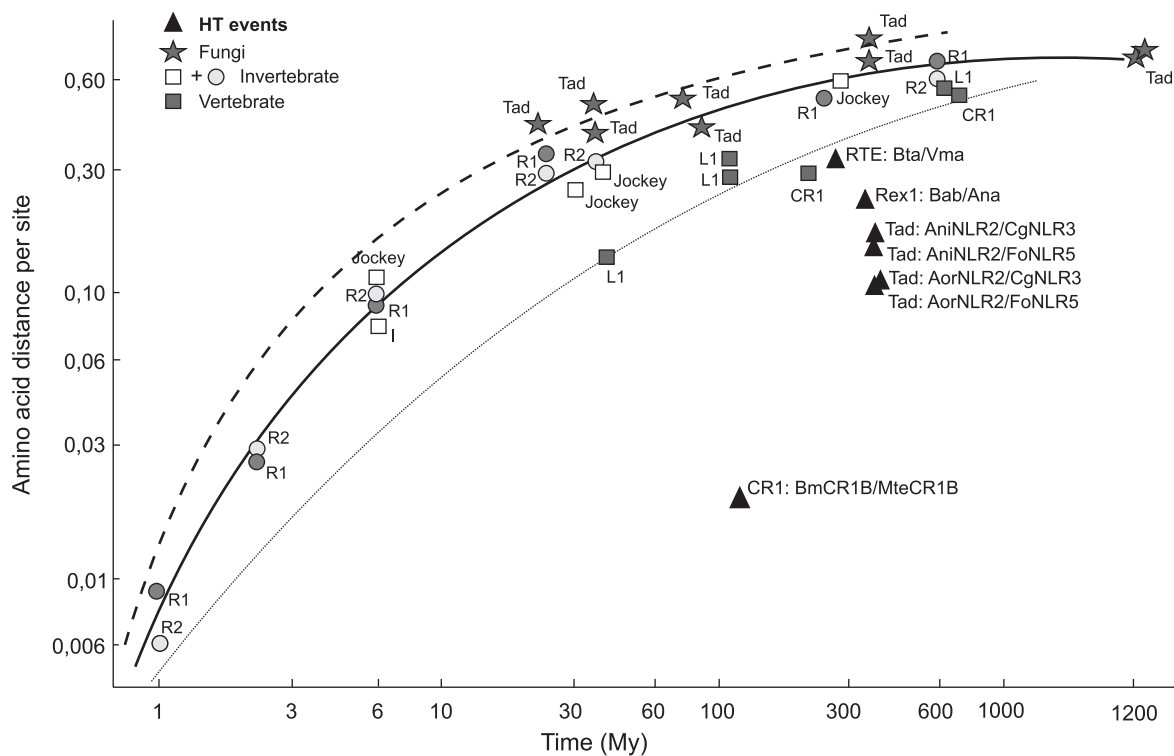
Putative cases of horizontal transfer of non-LTR retrotransposons

Clade: elements	Direction of transfer	Time estimate (Mya*)	References
Jockey: diverse elements	Insects: between Drosophilidae	~ 5–12	Sánchez-Gracia <i>et al.</i> , 2005
RTE: <i>Bta/Vma</i>	Vertebrates: from an ancestral snake lineage (Boidae) to the ancestor of ruminant mammals	~ 40–50	Kordiš, Gubenšek, 1998 Župunski <i>et al.</i> , 2001
Rex1: <i>Bab/Ana &amp; Bab/Anj</i>	Vertebrates: from Percomorpha fishes to eels ( <i>Anguilla</i> )	less than 90	Volff <i>et al.</i> , 2000
CR1: <i>BmCR1B/MteCR1B</i>	Insects: from Bombycidae moths to <i>Maculinea</i> butterflies (Lycaenidae)	~ 5–10	Novikova <i>et al.</i> , 2007
Tad: <i>AorNLR2 &amp; AniNLR2/FoNLR5 &amp; CgNLR3</i>	Fungi: from Sordaryomycetes ( <i>Chetomium globosum</i> and <i>Fusarium oxysporum</i> ) to Eurotiomycetes ( <i>Aspergillus oryzae</i> and <i>A. niger</i> )	less than 10	Novikova <i>et al.</i> , 2008

\* Mya – million years ago.



**Fig. 2.** Horizontal transfers of non-LTR retrotransposons HT of CR1 elements between *Maculinea* butterflies and *Bombyx* moths, dated 5–10 Mya (a), and HT of Tad elements between Sordaryomycetes (*Chaetomium globosum* and *Fusarium oxysporum*) to Eurotiomycetes (*Aspergillus oryzae* and *A. niger*) probably occurred 10 Mya (b). Phylogeny of superfamilies and families of butterfly suborder Ditrysia based on Tree of Life (<http://tolweb.org/>). Evolutionary tree sequenced fungal genomes is represented according to Fitzpatrick *et al.* (2006) with minor modifications.



**Fig. 3.** Divergence-versus-age analysis of non-LTR retrotransposons.

Amino acid sequence distances were calculated from the sequences of the complete reverse transcriptase (RT) domain. The curves for arthropods and vertebrates are reproduced from Malik *et al.* (1999); the curve for fungi is reproduced based on Novikova *et al.* (2008). Putative HT events are shown.

most recently occurred transmission of non-LTR retrotransposons between reproductively isolated species that belong to the taxa separated more than 100 million years ago.

Horizontal transfer is a complex phenomenon which has only indirect evidence. Moreover, it is impossible to investigate HT directly in a laboratory; all known HT cases between eukaryotes were discovered accidentally during phylogenetic studies (e.g. Kordiš, Gubenšek, 1998; Volff *et al.*, 2000; Novikova *et al.*, 2007; Novikova *et al.*, 2008). For example, HT of *CR1B* elements was identified along with the analysis of non-LTR retrotransposons diversity in *Maculinea* genomes. Bioinformatic approaches can greatly expand opportunities to observe the distribution and evolutionary history of non-LTR retrotransposons in various eukaryotic taxa and to identify novel cases of HT. The most suitable eukaryotic group for such investigations proved to be Fungi. More than 100 fungal genomes are available for analysis in various genomic databases and the number of projects is

constantly increasing (e.g. Galagan *et al.*, 2005). Fungi have small genomes, usually with limited amounts of repetitive DNA (including non-LTR retrotransposons) that makes the genomic analysis faster and easier.

Indeed, the genomic analysis of large set of fungal genomes revealed a possible case of HT of Tad-like non-LTR retroelements between Eurotiomycetes and Sordariomycetes, which is estimated to have taken place less than 10 Mya (Fig. 2, b). The time since divergence of Eurotiomycetes and Sordariomycetes is estimated between 310 Myr and 670 Myr (Berbee, Taylor, 1993; Heckman *et al.*, 2001). The oldest well-documented ascomycete fossils are found in the 400-Myr-old Rhynie chert (Taylor *et al.*, 1999). Based on this finding, it was suggested that 400 Myr for the Eurotiomycetes and Sordariomycetes divergence would seem to provide a conservative date estimate; even earlier dates could be expected (Kasuga *et al.*, 2002). Elements from Eurotiomycetes *Aspergillus niger* and *A. oryzae* (*AniNLR2* and *AorNLR2*) shared higher



similarity with elements from Sordariomycetes *Chaetomium globosum* and *Fusarium oxysporum* (CgNLR3 and FoNLR5) than with other non-LTR retrotransposons from *Aspergillus*; this cannot be explained solely by vertical evolution (Novikova *et al.*, 2008). Divergence-versus-age analysis as well as comparative analysis of evolutionary rates for functional genes and described elements confirmed the presence of slowdown effect on evolutionary rates, which always accompanies possible HTs (Fig. 3).

The criterion of discontinuous distribution cannot be applied in this case since all investigated fungal species differ in number and diversity of non-LTR retrotransposon clades presented in their genome (Novikova *et al.*, 2008). There are several main processes which could affect the copy number and diversity of non-LTR retrotransposons in fungal genomes: stochastic loss of non-LTR retrotransposons; burst of retrotransposition; the limitation of copy number increase by natural selection, which removes deleterious insertions; horizontal transfer; passive and active inactivation of repetitive sequences; self-regulation of transposition (decrease of transposition rate when the copy number increases) (e.g. Hua-Van *et al.*, 2005; Le Rouzic, Capy, 2005). Complex interactions between these processes lead to the formation of a unique repertoire of non-LTR retrotransposons in each fungal species. Although the criterion of “patchy” distribution does not work for fungi, the two strongest criteria were satisfied; therefore, the demonstrated inconsistency can be due to a possible HT.

### Conclusions

Horizontal transfer (HT) of non-LTR retrotransposons appears to be not just an artifact of analysis as suggested earlier (Malik *et al.*, 1999). Non-LTR retrotransposons, in fact, can be horizontally transmitted, although it is a very rare event in comparison to other types of transposable elements. The actual mechanisms of horizontal transfer are still unknown for eukaryotic TEs since it is not possible to show experimentally how the HT can occur. Parasites, symbionts, bacteria, or viruses all could be suggested as potential vectors for horizontal transfer.

### References

- Bannert N., Kurth R. Retroelements and the human genome: new perspectives on an old relation // Proc. Natl Acad. Sci. USA. 2004. V. 101. (Suppl. 2). P. 14572–14579.
- Bensasson D., Petrov D.A., Zhang D.X. *et al.* Genomic gigantism: DNA loss is slow in mountain grasshoppers // Mol. Biol. Evol. 2001. V. 18. P. 246–253.
- Berbee M.L., Taylor J.W. Dating the evolutionary radiations of the true fungi // Can. J. Bot. 1993. V. 71. P. 1114–1127.
- Biedler J., Tu Z. Non-LTR retrotransposons in the African malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity // Mol. Biol. Evol. 2003. V. 20. P. 1811–1825.
- Feschotte C., Pritham E.J. DNA transposons and the evolution of eukaryotic genomes // Annu. Rev. Genet. 2007. V. 41. P. 331–368.
- Fitzpatrick D.A., Logue M.E., Stajich J.E., Butler G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis // BMC Evol. Biol. 2006. V. 6. P. 99.
- Galagan J.E., Henn M.R., Ma L.J. *et al.* Genomics of the fungal kingdom: insights into eukaryotic biology // Genome Res. 2005. V. 15. P. 1620–1631.
- Gaunt M.W., Miles M.A. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks // Mol. Biol. Evol. 2002. V. 19. P. 748–761.
- Hartl D.L., Lohe A.R., Lozovskaya E.R. Modern thoughts on an ancient marinere: function, evolution, regulation // Annu. Rev. Genet. 1997. V. 31. P. 337–358.
- Heckman D.S., Geiser D.M., Eidell B.R. *et al.* Molecular evidence for the early colonization of land by fungi and plants // Science. 2001. V. 293. P. 1129–1133.
- Hua-Van A., Le Rouzic A., Maisonhaute C., Capy P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences // Cytogenet. Genome Res. 2005. V. 110. P. 426–440.
- Kapitonov V.V., Jurka J. Helitrons on a roll: eukaryotic rolling-circle transposons // Trends Genet. 2007. V. 23. P. 521–529.
- Kaplan N., Darden T., Langley C.H. Evolution and extinction of transposable elements in Mendelian populations // Genetics. 1985. V. 109. P. 459–480.
- Kasuga T., White T.J., Taylor J.W. Estimation of nucleotide substitution rates in Eurotiomycete fungi // Mol. Biol. Evol. 2002. V. 19. P. 2318–2324.
- Kidwell M.G. Horizontal transfer // Curr. Opin. Genet. Dev. 1992. V. 2. P. 868–873.
- Kidwell M.G., Lisch D. Transposable elements as sources of variation in animals and plants // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 7704–7711.

- Kordiš D., Gubenšek F. Unusual horizontal transfer of a long interspersed nuclear element between distant vertebrate classes // *Proc. Natl Acad. Sci. USA*. 1998. V. 95. P. 10704–10709.
- Lander E.S., Linton L.M., Birren B. *et al.* Initial sequencing and analysis of the human genome // *Nature*. 2001. V. 409. P. 860–921.
- Le Rouzic A., Capy P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift // *Genetics*. 2005. V. 169. P. 1033–1043.
- Lohe A.R., Moriyama E.N., Lidholm D.A., Hartl D.L. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements // *Mol. Biol. Evol.* 1995. V. 12. P. 62–72.
- Malik H.S., Burke W.D., Eickbush T.H. The age and evolution of non-LTR retrotransposable elements // *Mol. Biol. Evol.* 1999. V. 16. P. 793–805.
- Malik H.S., Eickbush T.H. NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans* // *Genetics*. 2000. V. 154. P. 193–203.
- Novikova O., Fet V., Blinov A. Non-LTR retrotransposons in fungi // *Funct. Integr. Genomics*. 2008. doi:10.1007/s10142-008-0093-8.
- Novikova O., Sliwińska E., Fet V. *et al.* CR1 clade of non-LTR retrotransposons from *Maculinea* butterflies (Lepidoptera: Lycaenidae): evidence for recent horizontal transmission // *BMC Evol. Biol.* 2007. V. 7. P. 93.
- Ono R., Kobayashi S., Wagatsuma H. *et al.* A retro-transposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21 // *Genomics*. 2001. V. 73. P. 232–237.
- Pardue M.L., Rashkova S., Casacuberta E. *et al.* Two retrotransposons maintain telomeres in *Drosophila* // *Chromosome Res.* 2005. V. 13. P. 443–453.
- Pritham E.J., Putliwala T., Feschotte C. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses // *Gene*. 2007. V. 390. P. 3–17.
- Robertson H.M. The mariner transposable element is widespread in insects // *Nature*. 1993. V. 362. P. 241–245.
- Sánchez-Gracia A., Maside X., Charlesworth B. High rate of horizontal transfer of transposable elements in *Drosophila* // *Trends Genet.* 2005. V. 21. P. 200–203.
- Silva J.C., Kidwell M.G. Evolution of P elements in natural populations of *Drosophila willistoni* and *D. sturtevantii* // *Genetics*. 2004. V. 168. P. 1323–1335.
- Taylor T.N., Hass H., Kerp H. The oldest fossil ascomycetes // *Nature*. 1999. V. 399. P. 648.
- Vitte C., Panaud O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model // *Cytogenet. Genome Res.* 2005. V. 110. P. 91–107.
- Volff J.N., Körting C., Schartl M. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes // *Mol. Biol. Evol.* 2000. V. 17. P. 1673–1684.
- Wicker T., Sabot F., Hua-Van A. *et al.* A unified classification system for eukaryotic transposable elements // *Nat. Rev. Genet.* 2007. V. 8. P. 973–982.
- Zingler N., Weichenrieder O., Schumann G.G. APE-type non-LTR retrotransposons: determinants involved in target site recognition // *Cytogenet. Genome Res.* 2005. V. 110. P. 250–268.
- Župunski V., Gubenšek F., Kordiš D. Evolutionary dynamics and evolutionary history in the RTE clade of non-LTR retrotransposons // *Mol. Biol. Evol.* 2001. V. 18. P. 1849–1863.

## КОРРЕЛЯЦИИ ОПЕРОННОЙ СТРУКТУРЫ С ДЛИНОЙ ГЕНОМА У 14 ВИДОВ МИКОПЛАЗМ

С.А. Лашин<sup>1,2</sup>, Ю.Г. Матушкин<sup>1</sup>, Т.М. Хлебодарова<sup>1</sup>, В.А. Лихошвай<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: lashin@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

В работе анализируются геномы 14 видов *Mycoplasma* существенно различающихся размерами своих геномов. Показано, что при уменьшении длины генома у *Mycoplasma* уменьшаются плотность предсказанных сайтов терминации транскрипции (количество сайтов, нормированное на число генов), плотность предсказанных единиц транскрипции и возрастает среднее количество генов, входящих в предсказанные единицы транскрипции. Для предсказания сайтов терминации транскрипции (термотивы) разработан новый метод. В работе формулируется и обосновывается гипотеза о существовании эволюционной тенденции к укрупнению оперонных структур в процессе уменьшения размеров генома в ходе дегенеративной эволюции у *Mycoplasma*.

**Ключевые слова:** *Mycoplasma*, терминация транскрипции, опероны, эволюция.

### Введение

Бактерии рода *Mycoplasma* – чрезвычайно полиморфные микроорганизмы, которые обладают примитивной организацией и самыми малыми геномами среди прокариот. В процессе дегенеративной эволюции они потеряли клеточную стенку и превратились в настоящих паразитов, которые 99 % времени проводят внутри клеток хозяина. Размеры их генома варьируют от 580 т.п.н. для *Mycoplasma genitalium* до 1380 т.п.н. для *M. mycoides* subsp. *mycoides* LC, т. е. различаются более чем в 2 раза (Razin *et al.*, 1998). Сравнительный анализ геномов двух тесно связанных видов *M. genitalium* и *M. pneumoniae* (размер генома 816 т.п.н.) показал, что структура этих геномов существенно различается. 209 генов (точнее, открытых рамок считывания) не выявлено у *M. genitalium* в сравнении с геномом *M. pneumoniae* (Himmelreich *et al.*, 1997). Анализ полностью секвенированных геномов микоплазм показал, что эти организмы не только полностью или частично утратили гены, контролирующие метаболические пути синтеза аминокислот и структур клеточной стенки, но и претерпели значительную потерю

генов, участвующих в таких процессах, как репарация и рекомбинация ДНК и клеточное деление (Razin *et al.*, 1998).

*Mycoplasma* относятся к Firmicutes, для которых характерен р-независимый механизм терминации транскрипции. Поиск сайтов терминации транскрипции и анализ их распределения в геномах *Mycoplasma* позволяет прояснить некоторые аспекты формирования транскриптонов у этих организмов, что особенно интересно в свете столь значительных изменений структуры их геномов, приобретенных в ходе эволюции. Под термином «предсказанные транскриптоны» (ПТ) или «предсказанные транскрипционные единицы» (ПТЕ) мы понимаем как опероны, состоящие из цистронов, т. е. отдельных кодирующих полипептиды последовательностей, так и комплексы генов тРНК или рРНК, которые не транслируются, но транскрибируются единой РНК, а затем процессируются. В контексте работы последовательности ДНК, кодирующие полипептид, тРНК и рРНК, будем называть термином «ген».

Данная статья посвящена изучению особенностей формирования транскрипционных единиц в пределах рода *Mycoplasma*. Проана-

лизированы полногеномные нуклеотидные последовательности 14 видов *Mycoplasma*. В работе формулируется и обосновывается гипотеза о наличии эволюционной тенденции к увеличению средней длины транскрипционных единиц микоплазм при уменьшении размеров их геномов. Для подтверждения гипотезы в работе проведен анализ распределения потенциальных сайтов терминации транскрипции (тер-мотивов) с использованием нового оригинального метода их определения.

### Материалы и методы

В работе использованы нуклеотидные последовательности геномов 14 видов *Mycoplasma*, характеристика которых приведена в табл. 1. Последовательности взяты из базы данных GenBank (<ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria>).

Для предсказания сайтов терминации транскрипции использованы программа FindTerm, параметры которой были выбраны по умолчанию (<http://linux1.softberry.com/berry.phtml>), и программа TermPred. В программе TermPred реализован оригинальный метод поиска тер-мотивов, описанный в данной статье. Параметры метода

TermPred подобраны таким образом, чтобы минимизировать отношение тер-мотивов ложных по положению к тер-мотивам истинным по положению. Истинными по положению считаются тер-мотивы, расположенные и ориентированные относительно генов таким образом, что они не вступают в смысловой конфликт с первичной оперонной разметкой генома. Обычно истинными по положению тер-мотивами являются те, которые лежат в межгенных участках, а ложными – расположенные внутри генов. Первичная оперонная разметка проведена с использованием программ FGenesB (<http://linux1.softberry.com/berry.phtml>) и OperonSet (Matushkin *et al.*, 2007). Для поиска тер-мотивов использовались последовательности длиной 660 нуклеотидов, лежащие в 3'-областях генов, с координатами (-60,600) относительно последнего нуклеотида гена. Статистическая обработка результатов проводилась с использованием критерия Стьюдента для коэффициента корреляции Пирсона.

### Результаты и обсуждение

Микоплазмы представляют род одноклеточных организмов, в который входят как сво-

Таблица 1

Генетическая статистика для 14 видов рода *Mycoplasma*

Название вида	Количество генов*	Длина генома, п.н.	Средняя длина гена	Количество генов на 1000 п.н.
<i>M. genitalium</i>	519	580086	1117,70	0,89
<i>M. arthritidis_158L3_1</i>	666	820453	1231,91	0,81
<i>M. mobile_163K</i>	667	777079	1165,04	0,86
<i>M. hyopneumoniae_7448</i>	696	920079	1321,95	0,76
<i>M. hyopneumoniae_J</i>	698	897405	1285,68	0,78
<i>M. synoviae_53</i>	713	799476	1121,28	0,89
<i>M. hyopneumoniae_232</i>	727	892758	1228,00	0,81
<i>M. pneumoniae</i>	732	816394	1115,29	0,90
<i>M. gallisepticum</i>	765	996422	1302,51	0,77
<i>M. agalactiae_PG2</i>	792	877438	1107,88	0,90
<i>M. pulmonis</i>	814	963879	1184,13	0,84
<i>M. capricolum_ATCC_27343</i>	854	1010023	1182,70	0,85
<i>M. mycoides_subsp_mycoides_SC_str_PG1</i>	1052	1211703	1151,81	0,87
<i>M. penetrans</i>	1069	1358633	1270,94	0,79

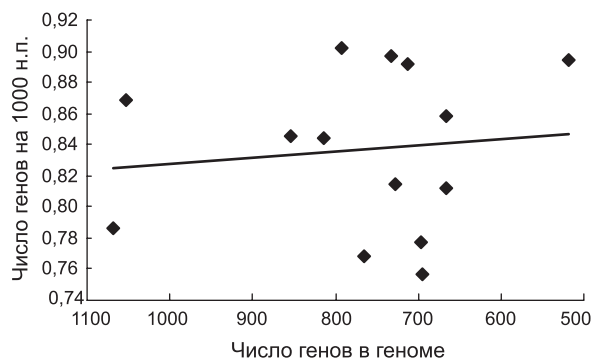
\* В контексте изложения под геном понимается участок ДНК, кодирующий белок, рРНК или тРНК.

бодноживущие, так и паразитирующие виды. В настоящее время секвенированы геномы 14 видов микоплазм. Их геномы содержат от 0,5 до 1,4 мегабаз и от 500 до 1100 генов (табл. 1). Считается, что уменьшение генома микоплазм связано с переходом к паразитическому образу жизни. Этот процесс сопровождается утерей генов и упрощением организации метаболизма и морфологии микоплазм-паразитов. В связи с тем, что накопилось достаточное количество секвенированных геномов, появилась возможность проанализировать плотность упаковки информации в геномах микоплазм в зависимости от их длины. Естественно предположить, что по мере уменьшения длины генома будет происходить увеличение плотности генов на геном (количества генов на единицу длины, например на 1000 н.п.). Как видно из рис. 1, подобная тенденция, действительно, выявляется, однако она незначительная. По линейному тренду плотность генов возрастает при уменьшении длины генома примерно на 3 %, но корреляция этих величин является недостоверной:  $r = -0,095$ ,  $p = 0,373$ .

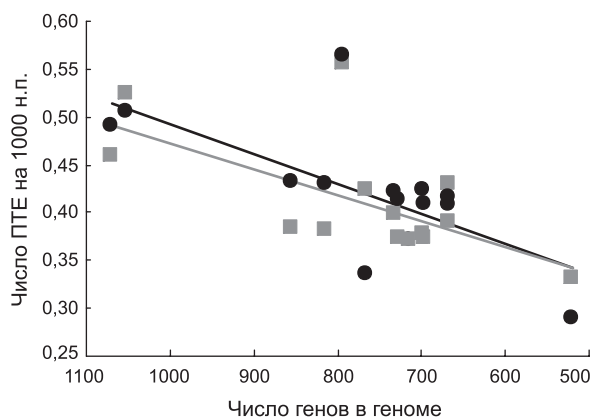
В то же время из рис. 2 видно, что плотность предсказанных транскрипционных единиц падает при уменьшении длины генома почти на 50 %. Коэффициент корреляции между количеством генов в геномах и плотностью предсказанных транскрипционных единиц, найденных программой OperonSet, равен  $r = 0,68$  и достоверен,  $p = 0,004$ . Для ПТЕ, выявленных программой FGenesB, коэффициент корреляции  $r = 0,63$  и также достоверен,  $p = 0,007$ .

На рис. 3 приведены результаты анализа среднего количества генов в предсказанной транскрипционной единице в зависимости от длины генома *Mycoplasma*. Видно, что среднее количество генов в ПТЕ увеличивается практически в 2 раза при уменьшении количества генов в геномах в 2 раза. Это свидетельствует о том, что в среднем плотность ПТЕ в геноме уменьшается с уменьшением длины генома (количества генов).

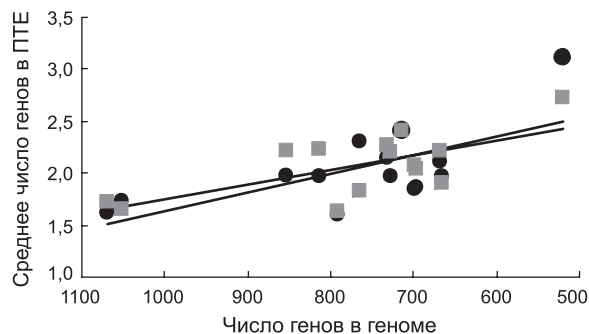
Таким образом, на основе представленных результатов мы формулируем гипотезу о существовании эволюционной тенденции к укрупнению оперонных структур в процессе уменьшения размеров генома в ходе дегенеративной эволюции у микоплазм.



**Рис. 1.** Динамика изменения плотности генов (число генов/1000 н.п.) в геномах *Mycoplasma* в зависимости от их размеров.



**Рис. 2.** Динамика изменения плотности предсказанных транскрипционных единиц (ПТЕ/1000 н.п.) в геномах *Mycoplasma* в зависимости от их размера. Черные круги и линия тренда соответствуют расчетам программы OperonSet, серые – FGenesB. Зависимости полностью сохраняются при разных параметрах соответствующих программ.



**Рис. 3.** Зависимость среднего числа генов в предсказанной OperonSet транскрипционной единице (ПТЕ) от размера генома *Mycoplasma*.



Для обоснования гипотезы мы предприняли дополнительный анализ распределения термотивов в геномах бактерий рода *Mycoplasma*. Для анализа мы использовали два метода поиска термотивов. Первый метод реализован в интернет-доступной версии программы FindTerm. Второй, TermPred, разработан в настоящей статье и описывается ниже.

### Метод распознавания термотивов TermPred

На рис. 4 приведена схема типичного мотива, который определяет терминацию транскрипции у *Mycoplasma* и других Firmicutes. Особенностью данного мотива является наличие GC-богатого инвертированного повтора, который в процессе синтеза РНК формирует вторичную структуру в виде шпильки, и следующего за ним Т-богатого участка (рис. 4). Появление такой структуры на 3'-конце растущей РНК приводит к запуску р-независимого механизма терминации транскрипции (Farnham, Platt, 1981; Lynn *et al.*, 1988; Wilson, von Hippel, 1995; Wang *et al.*, 1997; Yarnell, Roberts, 1999). Далее в статье такую последовательность будем называть термотивом.

На основании имеющихся данных о структуре р-независимого терминатора транскрипции нами был разработан оригинальный компьютерный метод TermPred предсказания термотивов в геномах прокариот. Для этого определяются четыре характеристики фиксированной в окне нуклеотидной последовательности: 1) энергия Гиббса формирования шпильки ( $G_{score}$ ); 2) насы-

щенность тимином Т-богатого участка ( $T_{score}$ ); 3) величина отклонения от оптимальной длины шпильки ( $C_{score}$ ) и (4) длина интервала между шпилькой и Т-богатым участком ( $P_{score}$ ). На основе этих характеристик вычисляется индекс потенциала терминации ( $TPI$ ) анализируемой последовательности:

$$TPI = G_{score} + T_{score} + C_{score} + P_{score}. \quad (1)$$

Превышение значения  $TPI$  некоторого фиксированного порогового значения служит основанием для идентификации анализируемой последовательности как термотива.

Значение  $G_{score}$  (энергия Гиббса шпильки) рассчитывалось программой UNAFold (Markham, Zuker, 2008).

Для расчета величины  $T_{score}$  Т-богатого участка длиной 15 нуклеотидов используется формула, предложенная в работах (d'Aubenton Carafa *et al.*, 1990; de Hoon *et al.*, 2005):

$$T_{score} = T_0 \cdot \sum_{i=0}^{14} \delta_{nuc(i)} \cdot \exp(-\lambda i), \quad (2)$$

где член  $\exp(-\lambda i)$  описывает вклад позиции нуклеотида в значение  $T_{score}$ , параметр  $\delta_i$  – вклад типа нуклеотида, расположенного в  $i$ -й позиции по формуле

$$\delta_i = \begin{cases} 1 & nuc(i) = t \\ -1 & nuc(i) = a \\ -2 & nuc(i) = g \text{ или } c \end{cases}$$

Значения параметров  $\delta_p$ ,  $\lambda = 0,144$  и  $T_0 = 2$  подобраны таким образом, чтобы максимизировать и минимизировать предсказание термотивов, истинных и ложных по положению соответственно.

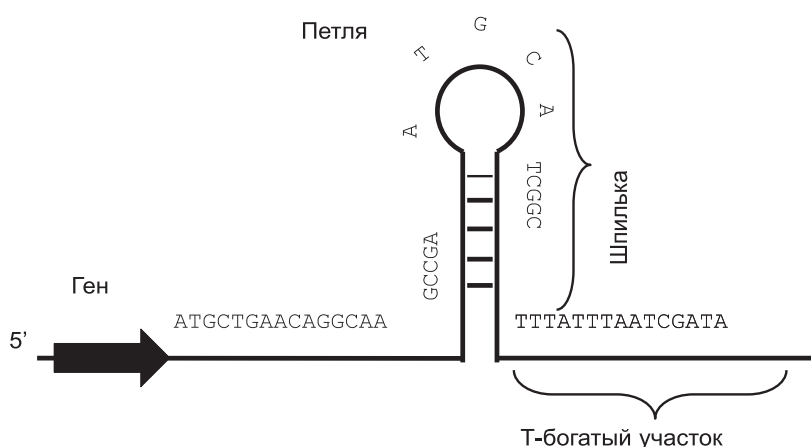


Рис. 4. Структура р-независимого терминатора транскрипции: шпилька и Т-богатый участок (Kingsford *et al.*, 2007).

Для расчета величины  $C_{score}$  используется формула:

$$C_{score} = C_0 \cdot \left( \frac{1 + \left[ \frac{n}{k_{11}} \right]^{h_{11}}}{1 + \left[ \frac{n}{k_{12}} \right]^{h_{12}}} + \frac{\left[ \frac{n}{k_{12}} \right]^{h_{21}}}{1 + \left[ \frac{n}{k_{22}} \right]^{h_{22}}} \right), \quad (3)$$

где  $n$  – длина исследуемой шпильки, а параметры  $C_0 = -5$ ,  $k_{11} = 20,5$ ,  $k_{12} = 10$ ,  $k_{21} = 43$ ,  $k_{22} = 43$ ,  $h_{11} = 4$ ,  $h_{12} = 6$ ,  $h_{21} = h_{22} = 10$  описывают значимость отклонения ее длины от некоторой оптимальной для терминации транскрипции. Значения параметров подобраны с учетом минимизации  $C_{score}$  для шпилек оптимальной длины.

Для расчета величины  $P_{score}$  используется формула:

$$P_{score} = P_0 \cdot (\exp(\eta l) - 1), \quad (4)$$

где  $l$  – длина интервала между шпилькой и Т-богатым участком, а  $\eta = 2,5$  и  $P_0 = -1$  – параметры скорости роста  $P_{score}$  при увеличении этого разрыва. Значения параметров формулы (4) были подобраны таким образом, чтобы в качестве тер-мотива принимались последовательности, у которых расстояние между шпилькой и Т-поворотом не превышает двух нуклеотидов. При этом последовательности с разрывом в два нуклеотида принимаются в качестве тер-мотивов только при очень больших значениях  $G_{score}$  и  $T_{score}$ .

Анализируемая последовательность принимается в качестве тер-мотива, если энергия Гиббса меньше 1 ккал/моль и величина  $TPI$  больше нуля.

### Предсказание тер-мотивов в геномах бактерий рода *Mycoplasma*

На рис. 5 представлена зависимость плотности тер-мотивов от длины геномов. Отчетливо выявляется более быстрое падение плотности тер-мотивов по сравнению со скоростью уменьшения количества генов в геноме. Корреляция между этими параметрами была достоверной при использовании обеих программ. Для программы TermPred коэффициент корреляции равняется  $r = 0,46$ ,  $p = 0,049$ , а для FindTerm –  $r = 0,53$ ,  $p = 0,03$ .

Таким образом, полученные расчеты подтверждают высказанную гипотезу об укруп-

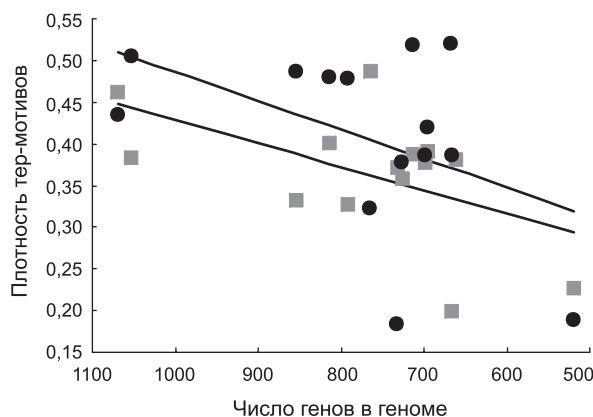


Рис. 5. Зависимость плотности тер-мотивов (число тер-мотивов/число генов) от длины генома (в генах).

Данные, полученные с помощью программы TermPred, обозначены черными кружками, а FindTerm – серыми квадратами.

нению структуры транскрипционных единиц генома при уменьшении его длины. Как следует из рис. 5, наблюдаемая тенденция к укрупнению достаточно хорошо выражена: плотность тер-мотивов относительно числа генов падает вдоль тренда примерно на 36 % при расчетах обеими программами.

Результаты качественно не меняются и в случае оценки плотности тер-мотивов в пересчете на длину генома в парах нуклеотидов (данные не приводятся).

### Заключение

В работе формулируется гипотеза о существовании эволюционной тенденции к укрупнению оперонных структур в процессе уменьшения размеров генома в ходе дегенеративной эволюции у микоплазм. Для обоснования гипотезы в работе разработан новый метод предсказания сайтов терминации транскрипции (тер-мотивов) и с его помощью проанализированы геномы 14 видов микоплазм, существенно различающихся размерами своих геномов. Альтернативный поиск сайтов терминации транскрипции осуществлялся также программой FindTerm. Первичная оперонная разметка геномов была проведена с использованием двух программ FGenesB и OperonSet. Показано, что с уменьшением размера генома у микоплазм

уменьшается и плотность предсказанных сайтов терминации транскрипции, и плотность предсказанных транскрипционных единиц. Следовательно, среднее количество генов, входящих в предсказанные транскрипционные единицы, возрастает. Эти данные свидетельствуют о том, что в процессе дегенеративной эволюции у микоплазм происходили не только уменьшение размеров генома в связи с потерей целого ряда генов, контролируемых утраченными ими функции и структуры, но также и изменение структуры их геномов в целом, выражающееся в уменьшении числа транскрипционных структур за счет потери части сайтов терминации транскрипции и слияния их в более крупные образования.

### Благодарности

Исследования частично поддержаны грантами РФФИ №06-04-49556 и № 08-04-01008, программами РАН «Молекулярная и клеточная биология» (проект № 10.7 «Компьютерное моделирование и экспериментальное конструирование генных сетей») и «Происхождение и эволюция биосферы» (проект №18.13 «Эволюция молекулярно-генетических систем: компьютерный анализ и моделирование», госконтрактом № 10104-37/П-18/110-327/180608/015 «Экосистемно-биоценоотические и генетические механизмы биологической эволюции и корреляция биологических событий») и грантом НШ-2447.2008.4. Научная школа Н.А. Колчанова «Биоинформатика и системная компьютерная биология».

### Литература

d'Aubenton Carafa Y., Brody E., Thermes C. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures // J. Mol. Biol. 1990. V. 216. P. 835–858.

- de Hoon M.J., Makita Y., Nakai K., Miyano S. Prediction of transcriptional terminators in *Bacillus subtilis* and related species // PLoS Comput. Biol. 2005. V. 1(3). P. e25.
- Farnham P.J., Platt T. Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription *in vitro* // Nucl. Acids Res. 1981. V. 9. P. 563–577.
- Himmelreich R., Plagens H., Hilbert H. *et al.* Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium* // Nucl. Acids Res. 1997. V. 25. P. 701–712.
- Kingsford C.L., Ayanbule K., Salzberg S.L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake // Genome Biol. 2007. V. 8. P. R22.
- Lynn S.P., Kasper L.M., Gardner J.F. Contributions of RNA secondary structure and length of the thymidine tract to transcription termination at the *thr* operon attenuator // J. Biol. Chem. 1988. V. 263. P. 472–479.
- Markham N.R., Zuker M. UNAFold: software for nucleic acid folding and hybridization // Methods Mol. Biol. 2008. V. 453. P. 3–31.
- Matushkin Yu.G., Vishnevskiy O.V., Volod'ko V.B. *et al.* Computer system GenomeMarker for annotation of bacterial genome structure-functional organization // Proc. of the 4th Moscow Intern. Congr. of Biotechnology: State of the Art and Prospects of Development. Moscow, 2007. V. 2. P. 399.
- Razin S., Yogev D., Naot Y. Molecular biology and pathogenicity of mycoplasmas // Microbiol. Mol. Biol. Rev. 1998. V. 62. P. 1094–1156.
- Wang D., Severinov K., Landick R. Preferential interaction of the his pause RNA hairpin with RNA polymerase beta subunit residues 904-950 correlates with strong transcriptional pausing // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 8433–8438.
- Wilson K.S., von Hippel P.H. Transcription termination at intrinsic terminators: the role of the RNA hairpin // Proc. Natl Acad. Sci. USA. 1995. V. 92. P. 8793–8797.
- Yarnell W.S., Roberts J.W. Mechanism of intrinsic transcription termination and antitermination // Science. 1999. V. 28. P. 611–615.

## **CORRELATION BETWEEN THE OPERON STRUCTURE AND THE GENOME LENGTH IN 14 MYCOPLASM SPECIES**

**S.A. Lashin**<sup>1,2</sup>, **Yu.G. Matushkin**<sup>1,2</sup>, **T.M. Khlebodarova**<sup>1</sup>, **V.A. Likhoshvai**<sup>1,2</sup>

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: lashin@bionet.nsc.ru;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

### **Summary**

We have analyzed 14 genomes of various mycoplasma species with quite different size. Both density of predicted transcription termination sites (number of sites scaled with number of genes) and density of predicted transcription units have been shown to decrease in view of genome length reduction. Thereafter an average number of genes in a transcription unit increases. Transcription termination sites (ter-motifs) were predicted with the use of a novel method developed. The hypothesis is formulated and proved that there is an evolutionary tendency of operon structures to increase during degenerative evolution (genome length reduction) of mycoplasmas.

## INTRON-EXON PATTERNS AS A POTENTIAL TOOL IN STUDYING GENE EVOLUTION

A. Ruvinsky

University of New England, Armidale NSW 2350, Australia, e-mail: aruvinsk@une.edu.au

The majority of introns are ancient elements and their phases and positions in genes were preserved for a long time. A string of intron phases represents a structure which carries essential information about organization and evolution of genes, which is usually ignored. Numerous observed strings have non-random intron phase patterns caused by intragenic repeats. Correlation between the lengths of CDS and the number of introns per human gene is high. Lengths of exons often remain constant in homologous and even paralogous genes belonging to distant species. Alignment of exon-intron strings provides useful visualization and generates new knowledge about evolution of gene families. It unravels intragenic duplications, intron gains and losses as well as extensions and contractions of exons. This additional information seems to be useful for studying gene evolution.

**Key words:** intron, exon, alignment, intragenic duplications, gene families

### Introduction

Positions and phases of the majority of introns show a great deal of conservation (Rogozin *et al.*, 2003; Roy, Gilbert, 2005). There are 3 phases, in which introns can be inserted: between codons (phase 0) and after the first or second nucleotides of a codon (phases 1 & 2). Shifts of intron-exon boundaries changing intron phases are rare events and have limited effect on the overall picture (Rogozin *et al.*, 2000). Intron gains and losses are more frequent and they certainly affect exon-intron structures of genes but do not necessarily influence corresponding proteins. Intragenic duplications likely played an important role in evolution of some genes (Jacob, 1983; Li, 1983; Patthy, 1987). According to available estimates the proportion of duplicated exons in long human genes is at least 6% (Fedorov *et al.*, 1998) and duplicated sequences occur in about 14% of all proteins (Marcotte *et al.*, 1999). There are hundreds of highly redundant genes in the human genome (Ruvinsky, Watson, 2007) and frequency of internal duplications has been increasing during metazoan evolution (Chen *et al.*, 2007). Intron-exon patterns allow tracing past events and could be helpful in evolutionary

reconstructions. For example, a string of intron phases, like 011212111111211211211121112111, representing a structure of human *GTF2I* gene, coding for general transcription factor 2I, contains valuable data. Three genes were identified in this family (Makeyev *et al.*, 2004). Lengths of exons which in some cases remains stable for lengthy evolutionary periods is another useful source of information. More detailed analysis of *GTF2I* gene confirms presence of several intragenic duplications and sheds light on the evolution of the gene. Those genes, which are prone to internal duplications, eventually became lengthy and their evolutionary pathways could be affected. Duplications involving an exon and sections of surrounding introns or several exon-intron pairs, if they framed by introns in the same phase, do not affect reading frame as well as exon lengths. Alignments of exon-intron structures of several genes from different species belonging to the same gene family could provide valuable information. This approach may help discriminate orthologs and paralogs and show the differences in evolutionary pathways of genes, including losses and gains of exons and introns and other intragenic rearrangements. The challenge is to understand the reasons behind these changes.



## Materials and Methods

The data was extracted from the exon-intron database (Saxonov *et al.*, 2000), which was extensively purged. The longest of the duplicate genes were left in the database and considered the constitutive form. The total numbers of studied genes were: *Hs*-11,315, *Dm*-8,497, *Ce*-10,312 and *At*-9,914. Some information was also obtained from genome browser Ensembl (<http://www.ensembl.org/index.html>) Statistical analysis was performed using methods described in our recent publication (Ruvinsky, Watson, 2007).

## Results

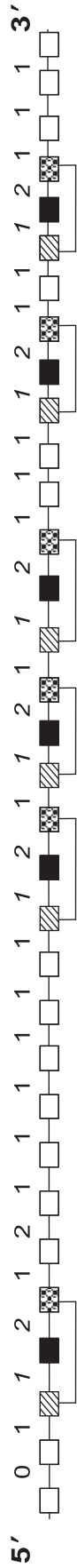
Comparisons of entropy values between observed intron strings and randomly simulated in Bernoulli schemes revealed that numerous observed strings have non-random intron phase patterns. The frequency of outliers among human genes which are beyond  $Z_{2,58}$  threshold (0,01 of the normal distribution) is 3,2 times higher than expected and is getting much higher for stricter  $Z$  thresholds (Ruvinsky, Watson, 2007). Many of such outliers have intragenic repeats. Correlation between the lengths of CDS and the number of introns per human gene is high ( $r = 0,83$ ) and getting stronger as number of introns increases. A possible interpretation of this fact is that intragenic duplications are more frequent in the genes with numerous introns and, because exons are also parts of the duplications, the length of coding sequence stronger correlates with introns number. Recently Chen *et al.* (2007) came to a comparable conclusion studying repeats in proteins. *GTF2I* is an example of a human gene with several intragenic repeats (Fig. 1).

Highly conservative exons located in the middle of these 6 repeats show significant DNA sequence similarity and hence the origin from a common ancestral sequence. All these 6 repeated exons have exactly the same length, there is no sequence gap in any of them and there are many conservative positions. The level of sequence identity varies from 66 % to ~ 40 % in 184 nucleotides. The total number of duplication events is likely to be 5. Identity of amino acid sequences coded by the conservative exons varies from 66.7 to 38,3 % and they belong to a highly conserved domain (pfam02946.12.) with DNA binding function (Vull-

horst, Buonanno, 2005). Alignment of exon-intron structures of genes from *GTF2I* family from several vertebrate species (Table 1) shows a great deal of conservation particularly between *GTF2I* orthologs from *Homo sapiens*, *Gallus gallus* and *Xenopus tropicalis*. Three other orthologs (*GTF2IRD1*) from fish species *Danio rerio* and *Oryzias latipes* and *Takifugu rubripes*, being paralogs to the tetrapod genes, show both similarities and differences in exon-intron structure. Intron insertions are likely the cause of the steadily increasing number of exons between the first and the second repeats. The fish species have only one lengthy exon following the first *GTF2I* repeat, while in frogs there are 5 exons, in birds 6 and in mammals 7, all of which are rather short. Intron loss, on the contrary, is a plausible explanation for the existence 268 nucleotides exons in fish species. The corresponding position of the gene in other compared vertebrate species contain two exons of 68 and 184 nucleotides, total of which is equal to 268. Taking into consideration that the 184-nucleotide exon is an ancient element in this gene family and surrounding introns are in the same phases, more parsimonious assumption is loss of the intron in the common ancestor of fish species. An alternative explanation based on insertion of phase 1 intron in higher vertebrates seems unlikely. Comparisons of exon-intron structures also show shifts of reading frames. For instance, shifting exon-intron boundary can be observed in *Xenopus tropicalis* 33 nucleotides exon (Table 1, underlined exon). It differs from the corresponding exons in other species by 4 extra nucleotides, such addition must change phase of the following intron from 1 to 2. This expectation is matched by the observation. The *GTF2IRD1* genes from fish species also contain modified repeat at the 3' end, which has length of 193 nucleotides (184 + 9) and thus has 3 extra codons. This is another example of exon expansion. The tetrapod species also have *GTF2IRD1* genes (not shown at Table 1), which are very similar to the fish species. However, *GTF2I* orthologs are not known for the fish species.

## Discussion

Intragenic duplications can, at least in some degree, explain creation of introns and exons. Studies of protein families revealed distinct duplication patterns and improved current understanding of



**Fig. 1.** Schematic presentation of exon-intron structure of human gene coding for general transcription factor 2I, *GTF2I*.

Numbers indicate intron phases. Rectangulars represent exons and lines introns. The brackets below the figure show the areas of repeats, which include highly conservative exons (black rectangulars), as well preceding (striped rectangulars) and the following exons (clear) between them. Conservative exons within the repeats are framed by introns in phase 1 and 2 (italic). Numeration of exons (not shown) starts from the first protein coding exon at the 5' end and finishes with the last protein coding exon at the 3' end.

**Table 1**

Alignment of exon lengths and intron phases for genes from GTF2I family in several vertebrate species

Species*	Length in aa	Number of exons	Number of GTF2 repeats	Exon lengths (above) and intron phases (below)
<i>Hs</i>	998	33	6	104,139,135,184,29,55,44,78,60,57,63,111,66,184,59,72,184,59,72,184,59,75,102,66,184,56,81,84,184,29,42,42,76 0112121111112112112111211121112111
<i>Gg</i>	984	32	6	106,139,135,184,29,49,44,78,60,63,102,66,184,59,72,184,59,84,96,66,184,53,111,84,184,29,69,15,76 0112121111112112112111211121112111
<i>Xt</i>	930	31	6	96,139,123,184,29,43,44,78,54,96,66,184,56,72,184,59,72,184,59,63,93,66,184,53,78,84,184,33,5,48,80 0112121111112112112111211121112111
<i>Dr</i>	919	26	5	126,146,140,184,225,96,84,184,26,130,23,81,37,23,51,66,184,62,39,268,29,78,193,38,116,66 021211212112112112112112110
<i>Ol</i>	958	27	5	126,142,63,174,184,269,90,84,184,26,124,98,30,36,55,43,40,184,62,78,268,29,78,193,38,110,69 011121112121112012112112112110
<i>Tr</i>	930	5	5	45,18,45,186,65,231,317,87,84,184,26,130,116,126,90,184,62,72,268,29,78,193,38,110,69 000002211212112112112

3 *GTF2I* orthologs from the tetrapods and 3 *GTF2IRD1* orthologs from fish species (tinted) represent two paralogous groups of genes from the same family. Exon lengths are shown above and intron phases below. The alignment of exon-intron strings was confirmed by DNA sequence similarity. Exons with length 268 are italicized and exons with length of 33 and 193 nucleotides are underlined. Explanations are in the text.

\* *Homo sapiens* – human; *Gallus gallus* – chicken; *Xenopus tropicalis* – frog; *Danio rerio* – zebrafish; *Oryzias latipes* – medaka; *Takifugu rubripes* – pufferfish.

the process. Tandem repeats of certain domains can be observed in many proteins (Björklund *et al.*, 2006). A model of gene formation based on essential role of introns in the duplication process was recently suggested (Street *et al.*, 2006). Similar observation relevant to MHC-linked *tenascin-X* gene has been earlier made by Hughes (1999). Our data support the view that intragenic duplications were used extensively during evolution of lengthy genes. Symmetric exons or clusters of neighbouring exons framed by introns in the same phase are preferable for duplication process (Long *et al.*, 1998). If the breaks occur in the surrounding introns, which are inserted in the same phase, this does not shift the reading frame and might not cause negative consequences. As we observed, several consecutive duplications create highly repetitive intron strings detectable by measuring their entropy.

A combined search for exons of the equal length framed by introns in the same phase suggested here is the efficient approach for finding intragenic duplications. Finally such intragenic duplications, involving a single exon-intron pair or more complex grouping, can be confirmed by the alignments of DNA and protein sequences. Long genes resulted from numerous internal duplications are not very common, but could become important if their proteins became «hubs» of proteome interactions (Dosztányi *et al.*, 2006). In some cases considered in this paper, intragenic repeats have a tandem structure, which might be a product of unequal recombination. In other situations intragenic repeats are dispersed. The basic point, however, remains unchanged, intragenic repeats regardless of their lengths or positions have to be framed by introns in the same phase. This is an essential condition for successful unequal recombination; otherwise shift of reading frame is inevitable.

Alignments of exon-intron structures from the same gene family may provide useful information, which can add to classical methods of DNA and protein sequences comparisons. Easy visualization of very lengthy alignments is the obvious advantage. It also can be helpful in distinction between orthologous and paralogous genes from the same family, because it utilises information about intron phase distribution and exon length never used by the standard methodology. Lastly, the alignments of exon-intron structures provide a wealth of new

knowledge about all kinds of intragenic rearrangements, including intron gains and losses, exon expansions and contractions as well as other changes, which should bring additional opportunities for reconstruction of gene evolution.

### Acknowledgement

The author is grateful to Dr. C. Watson for his contribution to the original paper «Intron phase patterns in genes: preservation and evolutionary changes» published in *The Open Evolution Journal* 2007, 1, 1–14.

### References

- Björklund Å.K., Ekman D., Elofsson A. Expansion of protein domain repeats // *PLoS Computational Biol.* 2006. 2. P. 0959–0970.
- Chen C.-C., Li W.-H., Sung H.-M. Patterns of internal gene duplication in the course of metazoan evolution // *Gene.* 2007. 396. P. 59–65.
- Dosztányi Z., Chen J., Dunker A.K. *et al.* Disorder and sequence repeats in hub proteins and their implications for network evolution // *J. Proteome Res.* 2006. 5. P. 2985–2995.
- Fedorov A., Fedorova L., Starchenko V. *et al.* Influence of exon duplication on intron and exon phase distribution // *J. Mol. Evol.* 1998. 46. P. 263–271.
- Hughes A.L. Concerted evolution of exons and introns in the MHC-linked *tenascin-X* gene in mammals // *Mol. Biol. Evol.* 1999. 16. P. 1558–1567.
- Jacob F. *Molecular tinkering in evolution.* Cambridge: Cambridge Univ. Press, 1983.
- Li W.-H. Evolution of duplicate genes and pseudogenes // *Evolution of Genes and Proteins / Eds M. Nei, R.K. Koehn.* Sunderland, MA.: Sinauer Associates Inc, 1983. P. 14–37.
- Long M., de Souza S., Rosenberg C., Gilbert W. Relationship between “proto-splice sites” and intron phases: evidence from dicodon analysis // *Proc. Natl Acad. Sci. USA.* 1998. 95. P. 219–223.
- Makeyev A.V., Erdenechimeg L., Mungunsukh O. *et al.* GTF2IRD2 is located in the Williams-Beuren syndrome critical region 7q11.23 and encodes a protein with two TFII-I-like helix-loop-helix repeats // *Proc. Natl Acad. Sci. USA.* 2004. 101. P. 11052–11057.
- Marcotte E.M., Pellegrini M., Yeates T.O., Eisenberg D. A census of protein repeats // *J. Mol. Biol.* 1999. 293. P. 151–160.
- Patthy L. Intron-dependent evolution: preferred types of exons and introns // *FEBS Lett.* 1987. 214. P. 1–7.
- Rogozin I.B., Lyons-Weiler J., Koonin E.V. Intron sliding in conserved gene families // *Trends in Genet.*

2000. 16. P. 430–432.
- Rogozin I.B., Wolf Y.I., Sorokin A.V. *et al.* Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution // *Curr. Biol.* 2003. V. 13. № 17. P. 1512–1517.
- Roy S.W., Gilbert W. Rates of intron loss and gain: implications for early eukaryotic evolution // *Proc. Natl Acad. Sci. USA.* 2005. 102. P. 5773–5778.
- Ruvinsky A., Watson C. Intron phase patterns in genes: preservation and evolutionary changes // *The Open Evol. J.* 2007. 1. P. 1–14.
- Saxonov S., Daizadeh I., Federov A., Gilbert W. EID: the Exon-Intron Database – an exhaustive database of protein-coding intron-containing genes // *Nucl. Acids Res.* 2000. 28. P. 185–190.
- Street T.O., Rose G.D., Barrick D. The role of introns in repeat protein gene formation // *J. Mol. Biol.* 2006. 360. P. 258–266.
- Vullhorst D., Buonanno A. Multiple GTF2I-like repeats of general transcription factor 3 exhibit DNA binding properties: evidence for a common origin as a sequence-specific DNA interaction module // *J. Biol. Chem.* 2005. 280. P. 31722–31731.

## GENETIC LINKAGE ANALYSIS CHALLENGES ON A DISTRIBUTED GRID ENVIRONMENT

A. Calabria, D. Di Pasquale, A. Orro, G. Trombetti, M. Gnocchi, L. Milanese

Institute of Biomedical Technologies ITB-CNR, Milano, Italy, e-mail: luciano.milanese@itb.cnr.it

The aim of the present work is to enable the use of high performance computing infrastructures, such as the EGEE-III Grid platform for the execution of genetic linkage analysis on very large SNPs (Single Nucleotide Polymorphism) markers data sets. The linkage analysis of SNPs has recently become a very popular approach for genetic epidemiology and population studies, aiming to discover the genetic correlation in complex diseases. It is a statistical method for detecting genetic linkage between disease loci and markers of known locations by following their inheritance in families through the generations. This is a NP-hard problem and the computational cost and memory requirements of the major algorithms proposed in literature grow exponentially with either pedigree size or number of markers. Implementations of the mentioned algorithms reflect these limits making analyses of medium/large data sets very hard on a single CPU. A web-based facility has been set up upon a high performance infrastructure, the EGEE Grid, in order to enable a tool for achieving a whole-genome linkage analysis. Test cases have been performed with 10.000 to 1 million SNPs per Chip.

**Key words:** linkage analysis, grid, distributed computing.

### Introduction

The haploid human genome is a sequence of over 3 billion DNA base pairs and contains an estimated twenty to twenty five thousand protein-coding genes, therefore finding the relationships between the expression of a particular gene or genes and the outbreak of human diseases is a hard task. A common approach to this challenge starts with the Genetic Linkage Analysis, which is a statistical method used to identify the location on a chromosome of a given gene involved in a disease, relative to a known location of chromosome markers. This is obtained by comparing genetic data with information on pedigree structure, and following the inheritance of phenotypic alterations in families through the generations, exploiting the tendency for genes and genetic markers to be inherited together due to their location near one another on the same chromosome.

Markers used in this analysis are *Microsatellites* (polymorphic loci that consist of repeating units of 1–6 base pairs in length) and, recently, also *Single Nucleotide Polymorphisms* (SNPs, DNA sequence variations involving a single nucleotide). Two major

algorithms (Lander, Green, 1987; Kruglyak *et al.*, 1996; Elston, Stewart, 1997) have been proposed for the assessment of the genetic linkage, with several implementations (Schaffer, 1996; Markianos *et al.*, 2001; Fishelson, Geiger, 2002); the problem is NP-hard and the computational cost of both algorithms grows exponentially for one of the two variables of the LOD Score equation, pedigree size and number of markers (Table 1). Implementations of the mentioned algorithms reflect these limits and the computational time for a medium size problem can require several CPU hours and huge RAM memory allocation. Therefore the role of high performance computing is getting more and more relevant in this field and more in general in the field of biological and medical scientific research, also due to the increasing quantity of data produced by the high throughput analysis techniques emerging nowadays. In this scenario, the use of distributed architecture environments can be an appropriate solution both from the computational point of view and for the data management, but submitting jobs, monitoring their status and retrieving the results can be challenging when working on a huge quantity of data.



Table 1

## Main algorithms characteristics

Algorithms	Applications	Computational bounds	Increasing computational time by		
			individuals	loci	time
Elston-Stewart	Linkage, SLink, Fastlink, Vitesse	$n^\circ$ loci: $\sim 8$	linear	exponential	$O((m2n)p)$
Lander-Green	GeneHunter, Allegro, Merlin	$n^\circ$ loci: $> 20$	exponential	linear	$O(m24p)$
Bayesian Networks	Superlink	$n^\circ$ loci: $nr$	linear	linear	$nr$

**Objectives**

Actual technologies for chips (for example Illumina) provide SNP genotyping arrays, from 10.000 SNPs to more than 1 million; pedigree files, that collect the information of the family structure, are often large, counting more than 30 individuals. Computational time and space on a single CPU is unreasonable with these preconditions, therefore stands the need for a distributed and high performance computation infrastructure and a system that enables linkage analysis with large datasets.

The aim of the present work is to enable the use of the EGEE Grid Infrastructure for the execution of linkage analysis on very large SNPs data sets, creating a pipeline for the linkage process. These large calculation challenges are launched into the Grid infrastructure, distributing the needed processes among different computing elements. In this context, a user friendly web based access to grid resources is provided. This proposed facility has been tested with challenges performed with markers data collected using the largest chips currently available (up to 1M SNPs).

**The computational problem**

The Genetic Linkage Analysis method is based on the calculation of the *LOD score*, a test parameter defined as the Logarithm Of Odds ratio between hypothesis of linkage versus the null one. Its computation belongs to the class of NP-Hard problems (Piccolboni, Gusfield, 1999), thus it has a complexity at least equal to the class NP-complete but it could also fall into the NP set. So far, approaches and algorithms proposed to solve the linkage analysis problem (quantitative and qualitative) (Table 1), are very computer intensive and the implementations of

them suffer of hardware limitations. For example the Elston-Stewart algorithm, which has been adopted by applications such as Linkage, SLink, Faslink, Vitesse, grows exponentially for markers and linearly for individuals' parameters, whereas the Lander-Green one has an opposite behavior. Since we deal with SNP chips and thus with a very large number of markers, we adopt the algorithm proposed by Lander and Green: in this way we can handle a set of about 100 markers each time, but we have to monitor the pedigree size variable.

**The computational infrastructure**

The computational infrastructure used to implement this work is the Grid operated by the European Project EGEE (Enabling Grids for E-science, Phase III) for the European scientific and research communities with more than 240 sites worldwide, providing access to more than 50.000 CPUs. It consists of a collection of computers, storages, special devices and services that are heterogeneous in every aspect, geographically distributed and dynamically linked by a wide-area network which can be accessed on-demand by a set of users with appropriate authentication and authorization. The Grid was originally invented as a practical solution to the problems of storing and processing the large quantities of data that are going to be produced by CERN's Large Hadron Collider (LHC). This computational facility can be seen as a service for sharing computer power and data storage capacity over the Internet, going beyond simple communication between computers, and aiming ultimately to turn the global network of computers into one vast computational resource. In this scenario, this infrastructure enables data sharing between thousands of scientists with multiple interests, tries to ensure that all data is accessible anywhere, anytime and

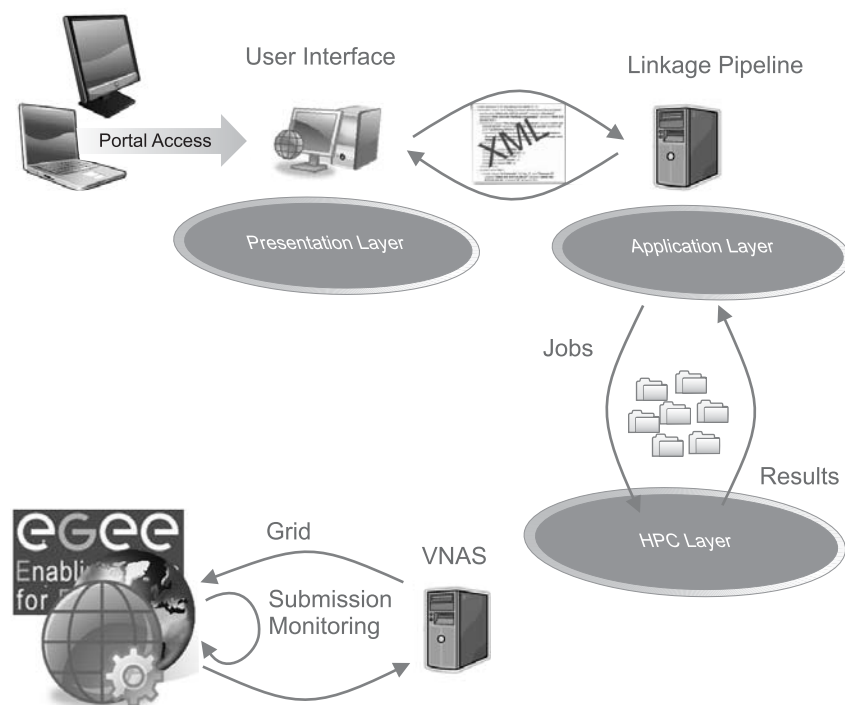
copies with different computer centers access policies, ensuring data security.

The EGEE Grid infrastructure runs upon a set of middleware services called gLite, which is integrated, certified and distributed by the project itself and world-widely deployed on the computational resources. The services available in the gLite distribution can be broadly classified into two categories: the Grid Foundation Middleware, covering the security of the infrastructure, the information, monitoring and accounting systems and the access to computing and storage resources; the other is a higher-level Grid Middleware, including services for job management, data catalogs and data replication, providing applications with end-to-end solutions. In order to use this infrastructure a personal authentication certificate is required. The certificate is released by a Certification Authority and associated to a Virtual Organization, a geographically independent group of collaborating scientists; the access to computational resources is executed through dedicated User Interfaces running the gLite middleware and its CLI commands.

The tests for the correctness of the linkage analysis algorithm and for the evaluation of its performances were performed on the computation cluster *Michelangelo* provided by the LITBIO project and is composed of 70 nodes, and 18.5TB of redundant storage. Each node is composed of two 275 dual-core AMD Opteron CPUs (total of 4 cores) and 8GB ram; all the nodes are connected via 10 Gbit/sec Infini-band for maximum scalability of the algorithms.

## Methods

The system is designed in 3 different layers, as schematized in Fig. 1. The presentation layer shows a web based user interface which has been created to aid users to setup linkage challenges and to mask the complexity of low level interactions with the Grid middleware. The web page makes use of javascript as client side scripting language to increase interactivity and user friendliness, using asynchronous communications with back end PHP server pages to manage information exchange with the application layer and input file uploads. Through the web interface users can:



**Fig. 1.** Diagram of system's design.

In the presentation layer users interact with the application which produces an XML file; the application layer parses the XML file to extract the logics of execution and linkage parameters; in the final step the HPC layer interacts with the Grid middleware submitting jobs and monitoring their execution.

– Create the data pipeline using draggable modules that represent each operative step: the arrangement of logical blocks, managed by client-side scripting, makes building the data flow, from the input files to the outputs retrieving, a visual matter.

– Customize each step: choose and upload the input files, define optional data pre-processing to adapt inputs to different algorithms requirements, select one of the available algorithms and set the proper parameters.

– Launch the analysis and monitor the jobs' state, retrieve and download the results once the challenge is complete.

The communication between the Presentation and the Application Layers is obtained through the creation of an XML file that describes the data pipeline and all its parameters: this file provides a complete description of the challenge characteristics making possible successive quick resubmissions; the choice of XML language reflects the future possibility of the system to drift to web services technology, adapting the Application Layer to this protocol with a minor effort. The communication in the opposite direction, between the Application and the Presentation Layer, is obtained with asynchronous data exchanges polled by client scripts that request information to the server backend and display the available ones without the need of page reloads.

In the application layer the system first parses the XML file created by the presentation layer, then executes pre-processing operations (error detection, etc) and finally runs the workflow on the HPC layer, monitoring its status: real time information on jobs status is returned through the interface to the user. Preprocessing operations are related to select the correct SNPs from our genotype database, then formatting files splitting SNP markers on the basis of the parameters reported in the XML file, such as distance between two consecutive markers or interval in kilo-bases (nucleotides, DNA measure), and so forth. The last activity in this phase is the creation of execution files for the HPC infrastructure according to the setting specified by user. Once files have been created the challenge execution can begin and the monitoring process, which is demanded to the VNAS component of the HPC Layer, is activated. When each single job has finished its computation, all results are merged into a single CSV file, that is easily readable by biolo-

gists or domain experts, and a final global plot for each chromosome is created.

In the HPC layer, the submission engine splits the workload into small jobs and distributes analysis tasks over the available resources. This is achieved by a software layer, called VNAS (Trombetti, 2007), built on top of the grid middleware which monitors each single grid process and ensures the success of its computation by managing the resubmission of failed jobs automatically. When all tasks are computed the results are retrieved, merged and made available for downloading through the web interface.

The VNAS framework is an advanced system for the submission and monitoring of jobs onto the Grid environment. VNAS has integrated strategies for the detection of failures and hang-ups of Grid jobs and can perform automatic resubmission for jobs detected in such conditions. VNAS hence provides an abstraction with reliability over the Grid platform which significantly eases the task of developing new applications for the Grid. VNAS also significantly simplifies the development of certain complex Grid workflows by providing a callback system that eases the creation of arbitrarily complex multi-stage computational applications. In addition, VNAS provides an abstracted virtual sandbox which bypasses certain Grid limitations such as the maximum sandbox size, while simultaneously reducing the usage of Grid bandwidth and storage resources. The latter is achieved by transparently detecting equality of virtual sandbox files based on content, across different submissions, even when performed by different users, and by performing automatic garbage collection of files after N days of no-use.

The VNAS framework has been used in various projects including the present one, to raise the reliability and reduce the development time for new Grid applications and pipelines being developed.

## Results

Tests were run to evaluate both efficiency on computational time and functionality of the proposed approach, obtaining an estimate of the Grid performances in comparison to a single 2 GHz CPU workstation and to the *Michelangelo* cluster composed by 280 CPU cores. The test, run using the Genehunter software, involved analysis of pedigrees composed by 25 subjects, including

individuals genotyped with different genotyping chips of markers from 10k up to 1 million of SNPs each. Considering the trade off between CPU load and memory requirements, a data subset of 50 SNPs was evaluated as the optimal workload for a Grid node and was assigned to each program run; the total number of runs needed to process all SNP data produced by each genotyping chip was split into jobs with an estimated duration of around 6 hours on a mid-range multicore CPU. Test results are summarized as follows: Table 2 shows the duration of the challenges in hours and Fig. 2 displays the relative plot.

Comparing the results of different computation infrastructures we can see that distributed analysis pipelines with number of data (linkage variables) close to the computational limits for a mid-range workstation achieved improvements of about 65–70 % in computational time compared to dual-core 2 GHz CPU execution and considering markers chips greater than 100 k, the advantage of the distributed architecture gets proportionally bigger, due to the difference between linear increase of computational time for the sequential run on the single CPU and the saturation trend of the parallelized data flow obtained distributing the workload on the Grid’s computing elements. It must also be highlighted that computations with a higher number of individuals in the pedigree tree, were still performed on the distributed infrastructure, but resulted infeasible on the desktop workstation

due to memory overflow. The tests also show that the average performance of the proposed system can’t compete with a cluster environment in terms of pure computing time, even if it shows a comparable trend due to the similar workload distribution technique adopted (not MPI).

**Conclusions**

This application enables the user to launch genetic linkage analysis computations for medium to large challenges over a distributed computa-

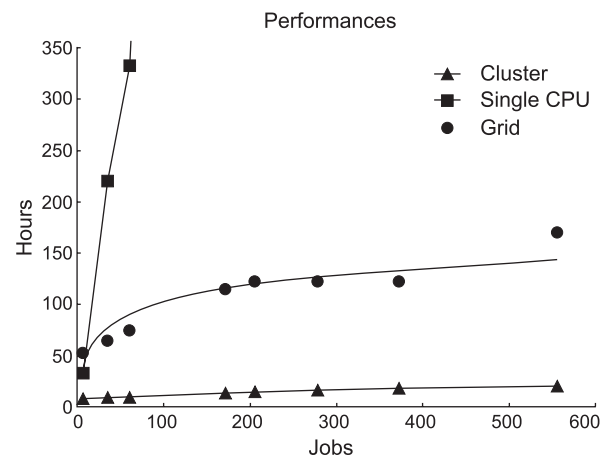


Fig. 2. Graph of the test results.

The computation efficiency of the distributed approach of our Grid-based system grows with the size of the challenge, being lower than a dedicated Cluster’s but much higher than a sequential run on a single processor unit.

**Table 2**

Duration of the challenges. Test results

Genotyping chip (# of SNPs)	Runs (50 SNP)	Jobs (6 h)	Computational cost (hours)		
			single 2GHz CPU	cluster (70 nodes 280 CPU Cores)	grid
10 k	200	6	33	8	53
66 k	1320	35	220	9.5	64
100 k	2000	60	333	10	75
317 k	6340	172	1056	13	115
370 k	7400	206	1233	15	120
500 k	10000	278	1665	16	122
670 k	13400	373	2233	18	122
1 M	20000	556	3332	20	170

Notes. Data derived from different genotyping chips were analyzed with the Genehunter software using 3 computational infrastructures: our Grid-based system, a 280 cores Cluster and a single 2GHz CPU Workstation. The left side columns show the challenges characteristics, on the right the challenges durations expressed in hours as resulted for different environments.

tional infrastructure such as the EGEE Grid. The application offers an interface to customize the data pipeline and achieve, through the submission engine, the parallel processing of the pipeline tasks on the distributed resources. The low-level interactions with the Grid environment are managed with a reliable software layer that hides any complexity for the final user, allowing monitoring and results retrieving to be easily managed with the web interface. Tests made on the proposed system showed that this approach is mostly useful in high-end challenges, where Grid overheads are affecting overall execution times less compared to single CPU performances; only very small challenges may show higher efficiency when run in a single CPU workstation, while very large analyses are made accessible even without a dedicated cluster, resulting as a good and affordable solution when considering the cost-benefit ratio of the two infrastructures.

#### Acknowledgments

This work has been also supported by the Enabling Grids for E-science (INFSO-RI-222667), CNR-BIOINFORMATICS, LITBIO and ITALBIONET and Italian-Canada FIRB-MUR Projects.

#### References

- Elston R.C., Stewart J. A general model for the analysis of pedigree data // *Hum. Hered.* 1997. № 21. P. 523–542.
- Fishelson M., Geiger D. Exact genetic linkage computations for general pedigrees // *Bioinformatics.* 2002. V. 18. № 1. S189–S198.
- Lander E.S., Green P. Construction of multilocus genetic linkage maps in humans // *Proc. Natl Acad. Sci. USA.* 1987. № 84. P. 2363–2367.
- Kruglyak L., Daly M.J., ReeveDaly M.P., Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach // *Amer. J. Hum. Genet.* 1996. № 58. P. 1347–1363.
- Markianos K., Daly M., J., Kruglyak L. Efficient multipoint linkage analysis through reduction of inheritance space // *Amer. J. Hum. Genet.* 2001. V. 68. № 4. P. 963–977.
- Schaffer A. Faster linkage analysis computations for pedigrees with loops or unused alleles // *Hum. Hered.* 1996. V. 46. № 4. P. 226–235.
- Trombetti G.A. *et al.* Data handling strategies for high throughput pyrosequencers // *BMC Bioinformatics.* 2007. № 8(1). P. S22.
- Piccolboni A., Gusfield D. On the complexity of fundamental computational problems in pedigree analysis. Technical report of University of California, Computer Science Department, 1999.



## ПРИМЕНЕНИЕ НЕМЕТРИЧЕСКОГО МНОГОМЕРНОГО ШКАЛИРОВАНИЯ ДЛЯ МУЛЬТИПЛАТФОРМЕННОЙ ОБРАБОТКИ МИКРОЧИПОВЫХ ЭКСПРЕССИОННЫХ ДАННЫХ

В.М. Ефимов, А.В. Катохин

Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: efimov@bionet.nsc.ru

Разработан многомерный метод согласованного мультиплатформенного поиска дифференциально экспрессирующихся генов-маркеров в массивах микрочиповых экспрессионных данных для одного и того же набора образцов. Метод применен для совместного анализа двух массивов микрочиповых экспрессионных данных, полученных с применением платформ CodeLink и Affymetrix (Borovecki *et al.*, 2005). Построены списки дифференциально экспрессирующихся генов-маркеров, соответствующих каждой из четырех групп образцов периферической крови (пациенты с болезнью Хантингтона, с предсимптоматикой этой болезни, субнорма, норма).

**Ключевые слова:** профили экспрессии генов, неметрическое шкалирование, болезнь Хантингтона.

### Введение

К числу многообещающих направлений современной биологии относится анализ экспрессии генов с помощью микрочипов. Микрочипы – упорядоченные комплекты фрагментов ДНК или РНК (зондов), иммобилизованные на специальных носителях (пластинках из стекла, пластика или кремния, каплях геля), называемых платформами (вместе с сопутствующими технологиями, как правило, разных производителей). Технология микрочипов позволяет одновременно анализировать экспрессию десятков тысяч генов в нескольких десятках проб (образцов), обычно представляемую в виде матрицы уровней экспрессии, в которой столбцам отвечают образцы, а строкам – гены. Столбцы и строки этой матрицы называются профилями экспрессии соответственно образцов или генов. (Иногда возникает терминологическая путаница: термин «профиль экспрессии генов» в некоторых работах может означать «профиль экспрессии образца по всем генам»).

Одной из существенных проблем, возникающих при анализе микрочиповых экспрессионных данных, является несогласованность результатов, получаемых при анализе одних и тех

же образцов на разных платформах. Основная трудность заключается в том, что вследствие различия технологий не удается напрямую сравнивать профили экспрессии генов. Во-первых, наборы генов разных платформ совпадают лишь частично. Во-вторых, даже одни и те же гены разных платформ можно лишь условно отождествить друг с другом из-за того, что на разных платформах иммобилизуются, вообще говоря, несколько разные зонды, хотя и относящиеся к одному и тому же гену (изоформы, клоны), что естественно приводит к различиям в профилях экспрессии. Имеется и ряд других межплатформенных различий, например, в интенсивности гибридизации, уровне «шума», характере случайных ошибок и т. д. (Cheadle *et al.*, 2007; Vemmo *et al.*, 2008). Тем не менее необходимость объединения результатов, получаемых на разных платформах, остается крайне актуальной.

Существует несколько подходов к решению этой проблемы. Простейший и наиболее часто применяемый способ – это ее обойти. По каждому массиву микроэкспрессионных данных отдельно решается одна и та же задача, например, поиск генов-кандидатов, которые можно использовать в качестве маркеров онкологии-

ческих или иных заболеваний. Далее списки генов-кандидатов сравниваются и выбираются те, которые присутствуют в нескольких списках. Большим недостатком такого подхода является то, что при любых уровнях значимости гены, представленные только на одной платформе, не имеют никакого шанса попасть в общий список. Например, в работе P. Pedotti *et al.* (2008) уровень экспрессии генов в 10 образцах (5 – трансгенные мыши и 5 – дикий тип) измерен с помощью 5 разных платформ, и на 10 %-м уровне значимости по критерию FDR (Benjamini, Hochberg, 1995) отобраны гены с дифференциальной экспрессией на разных группах мышей. Барьер преодолели 4, 130, 3071, 54 и 13 генов от каждой платформы соответственно. Только два гена оказались общими для всех 5 списков и 4 встретились по 4 раза.

Другой способ – ограничиться только одними и теми же зондами одних и тех же генов и состыковать их профили по всем платформам. Недостатки: приходится уделять специальное внимание отождествлению зондов по их описаниям, которые не всегда имеются или достаточны для этого (Barnes *et al.*, 2005). Однако эту подзадачу можно решать эмпирически. В работе H.K. Lee *et al.* (2004), например, проанализировано 60 массивов и выбраны пары зондов, корреляция между которыми значима в трех и более списках (максимум 31). Такие зонды считаются совпадающими. Различия в технологиях тоже остаются, что влечет необходимость различного рода предварительных трансформаций. Кроме того, так же, как и в предыдущем случае, серьезным недостатком является то, что при этом теряется значительная часть уникальной информации.

Мы предлагаем новый способ. Для каждой платформы вычисляется матрица различий между профилями экспрессии образцов по всему набору генов. Далее конструируется объединенная матрица различий, по ней методом неметрического шкалирования строится общее для всех платформ евклидово пространство профилей экспрессии образцов малой размерности, выявляются интересующие нас направления в этом пространстве (например между больными и контролем) и из всех платформ выбираются профили экспрессии генов, максимально коррелирующие с этими направлениями.

## Материалы и методы

Анализировался массив микрочиповых экспрессионных данных, полученных с применением платформы CodeLink. Массив после процедуры фильтрования (строки с отсутствующими значениями удалялись) содержал 17525 полных профилей экспрессии генов (строки) по 31 образцу периферической крови. Эти же образцы анализировались с применением платформы Affymetrix – 22282 полных профиля. Файлы с данными этих экспериментов извлечены из базы данных GEO (Barrett *et al.*, 2005; <http://www.ncbi.nlm.nih.gov/geo/>). Образцы периферической крови взяты у 12 пациентов с болезнью Хантингтона, 5 – с предсимптоматической этой болезнью (предрасположенных), 14 – здоровых. Болезнь вызывается мутацией по гену, кодирующему белок хантингтин, и проявляется с возрастом. Больные характеризуются наличием мутации и клиническими проявлениями болезни Хантингтона, предрасположенные – наличием мутации, но отсутствием клинических проявлений, здоровые – отсутствием и мутации, и клинических проявлений. Уровень клинических проявлений определялся опытным неврологом по унифицированной шкале болезни Хантингтона, наличие мутации – генетическим тестированием (Borovecki *et al.*, 2005). Массивы неоднократно анализировались другими авторами и фактически стали тестовым множеством для проверки различных методов анализа микрочиповых данных (<http://www.genesifter.net/web/webinars.php>). Оба массива логарифмировались, затем центрировались и нормировались сначала по столбцам для устранения неоднородности по образцам, затем по строкам для устранения эффектов масштаба.

После этого для каждой платформы вычислялась матрица попарных евклидовых расстояний между профилями экспрессии образцов. (Следует подчеркнуть, что метод работает для любых мер сходства–различия, а не только для расстояний, удовлетворяющих аксиомам метрики. Поэтому далее в любом случае будем именовать их различиями, чтобы отличать от евклидовых расстояний, генерируемых методом неметрического многомерного шкалирования.). Межплатформенный коэффициент корреляции между матрицами различий равен 0,61, а

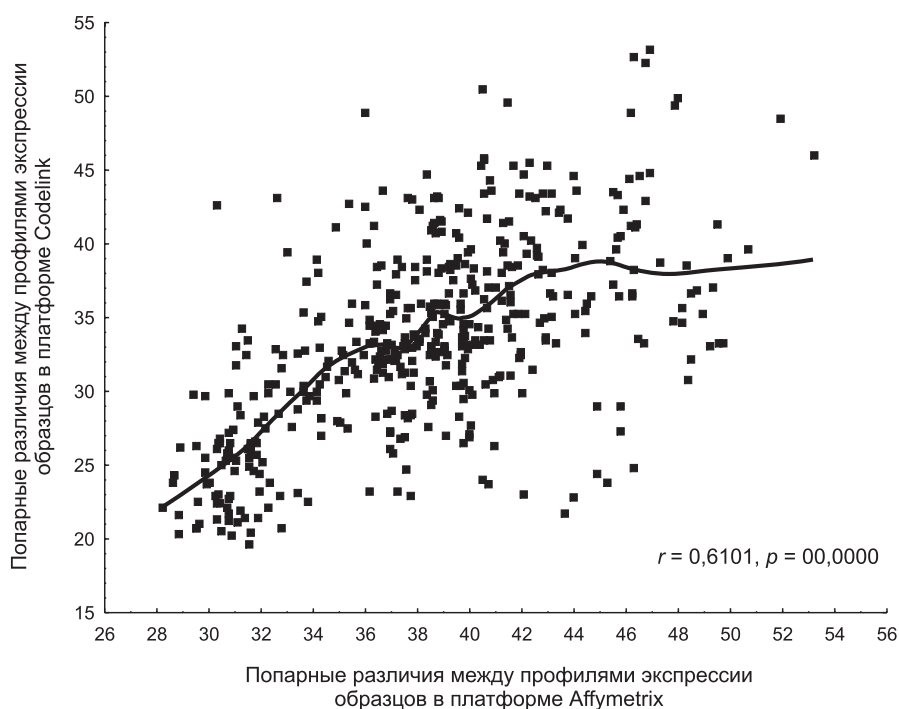
зависимость различий друг от друга оказалась явно нелинейная (рис. 1). Полученные матрицы ранжировались и усреднялись. Средний ранг различий достаточно хорошо воспроизводит исходные различия между образцами по каждой платформе (корреляция с Affymetrix – 0,89, с CodeLink – 0,88).

Матрица средних рангов различий между профилями экспрессии образцов обрабатывалась методом неметрического многомерного шкалирования (Taguchi, Oono, 2005). В этом методе каждому объекту исходного множества ставится в соответствие точка в евклидовом пространстве малой размерности, чаще всего на плоскости. Далее точки в этом пространстве передвигаются таким образом, чтобы матрица расстояний между ними как можно лучше соответствовала матрице различий между объектами исходного множества. За критерий соответствия принят коэффициент ранговой корреляции между расстояниями и различиями, что эквивалентно критерию, приведенному в работе Taguchi, Oono (2005).

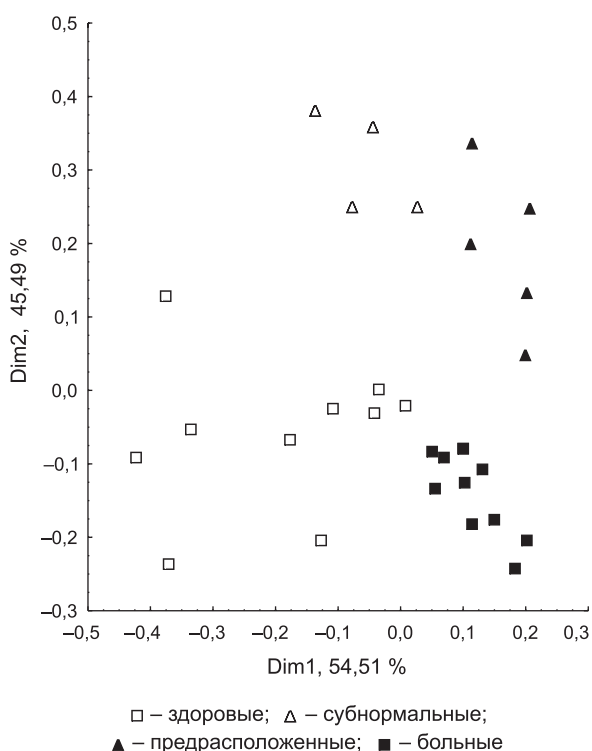
## Результаты

Все профили экспрессии образцов распались на группы, соответствующие пациентам, страдающим болезнью Хантингтона, предрасположенным к этой болезни и здоровым (рис. 2). Надо подчеркнуть, что имеющаяся информация об изначальной принадлежности пациентов к тем или иным группам при обработке методом неметрического шкалирования никак не использовалась. Тем не менее четверо пациентов из числа здоровых образовали группу, которая достаточно далеко отстоит от других здоровых и примыкает к группе предрасположенных. Это послужило основанием для выделения их в отдельную группу, которая была названа нами «субнормальной». Все четыре группы занимают отдельные области на плоскости.

Для каждой группы вычислены центроиды и направления на них от общего центра тяжести. Для всех профилей экспрессии генов обеих платформ вычислены и отложены на плоскости коэффициенты корреляции с осями неметрического



**Рис. 1.** Соответствие попарных различий между профилями экспрессии образцов для платформ Affymetrix и CodeLink (линия – lowess-регрессия).



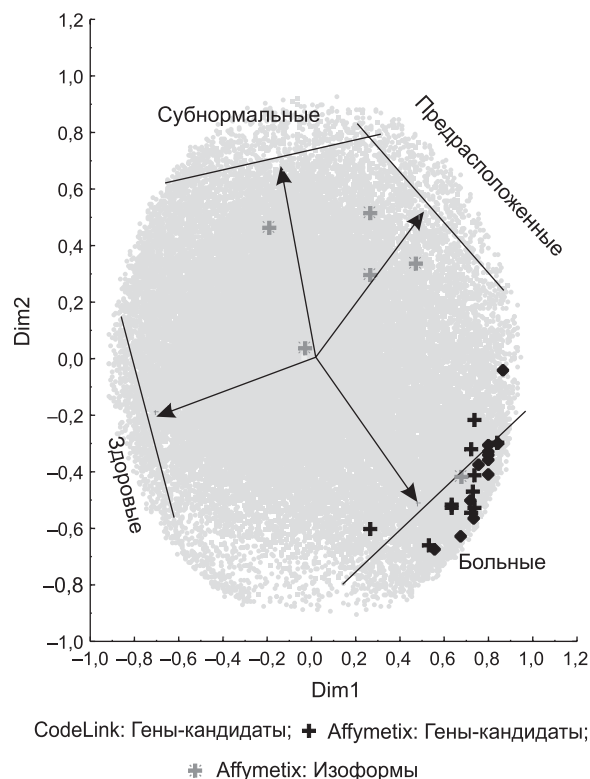
**Рис. 2.** Расположение профилей экспрессии образцов на плоскости неметрического двумерного шкалирования.

двумерного шкалирования (рис. 3). Каждому направлению на плоскости экспрессии образцов (рис. 2), в частности на центроид любой группы, взаимно однозначно соответствует направление на плоскости экспрессии генов (рис. 3). Профили экспрессии генов, наиболее отстоящие от центра вдоль этого направления, характеризуются самыми высокими значениями на объектах этой группы.

Профили, расположенные на противоположной стороне рисунка (не показано), наоборот характеризуются самыми низкими значениями. Поэтому для каждой группы образцов можно выделить соответствующую группу профилей экспрессии генов. Границу можно проводить из разных соображений. На рис. 3 граница проведена таким образом, чтобы вероятность ее превышения для максимального по модулю случайного коэффициента корреляции из 39807 (22282+17525) была не более 0,05. Следует заметить, что это достаточно жесткий барьер, практически гарантирующий отсутствие случайных профилей экспрессии генов в выделенной области. Тем не менее его преодолели

3670 генов (750 – Affymetrix, 2920 – CodeLink), что является довольно неожиданным. Общими для обеих платформ являются 8600 генов, из них совместно преодолели барьер 241. Крестиками и ромбиками обозначены профили экспрессии генов-кандидатов, выделенных в работе (Borovecki *et al.*, 2005), которые, как и следовало ожидать, почти все попадают в выделенную область. Часть генов платформы Affymetrix представлены несколькими изоформами, поэтому некоторые из них оказались разбросаны по всей плоскости.

Расположение групп позволяет высказать гипотезу о возможности последовательного перехода по круговой траектории: здоровые → субнормальные → предрасположенные → больные, причем каждой группе соответствует свой набор профилей экспрессии генов. Метод неметрического многомерного шкалирования позволяет выделить эти наборы для дальнейшего анализа средствами генной онтологии.



**Рис. 3.** Расположение профилей экспрессии генов на плоскости коэффициентов корреляции с осями неметрического двумерного шкалирования (Affymetrix + CodeLink).

### Обсуждение

В чем состоит идея предлагаемого способа обработки? Схема рассуждений довольно проста. Есть одни и те же образцы, и для них на разных платформах получены профили экспрессии по всем генам. Если эти образцы отличаются друг от друга по экспрессии генов, то эти отличия должны проявляться на каждой платформе. И наборы отличий должны быть хоть в чем-то подобны друг другу, иначе пропадает сама идея применения микрочипов. С другой стороны, поскольку каждая платформа, безусловно, имеет свою специфику, они не обязаны совпадать. Первая задача состоит в том, чтобы выявить то общее, что есть в наборах отличий между профилями образцов для разных платформ.

Метод неметрического многомерного шкалирования оказался исключительно подходящим для решения этой задачи. В этом методе входной информацией служит матрица ранжированных различий между профилями образцов. Мы любым удобным для себя образом определяем различия между профилями образцов по одной из платформ, вычисляем матрицу этих различий, упорядочиваем и заменяем каждое различие его рангом. То есть нас интересует не конкретное значение различия между двумя профилями образцов, а только его ранг среди всех других различий. Естественно предположить, что этот ранг должен быть близок к рангу различия между профилями этих же образцов, вычисленного по другой платформе, т. е. должна наблюдаться корреляция между рангами различий в разных платформах. И действительно, в рассматриваемом случае корреляция и между различиями, и между их рангами составила 0,61. Корреляция не слишком высока, что указывает на значительную разницу между обеими изучаемыми платформами. Но оба коэффициента корреляции одинаковы, и это означает, что переход к рангам не привел к заметной потере информации.

Развивая эту идею дальше, мы можем просто вычислить средний ранг различий для каждой пары образцов по обоим платформам и снова его упорядочить, взяв его ранг. Тем самым мы усилим то общее, что есть в обеих матрицах ранжированных различий, и ослабим специфику каждой отдельной платформы.

Очевидно, что число платформ совершенно непринципиально, так же, как и способы измерения различия. Ранжирование, и в этом его несомненный плюс, все приводит к единой шкале с равномерным распределением, избавляя от проблем, связанных с нелинейностью, разномасштабностью и неоднородностью распределений.

Далее к матрице новых рангов применим собственно сам метод неметрического многомерного шкалирования в варианте Шепарда-Тагучи-Ооно (Taguchi, Oono, 2005). В этом методе с каждым образцом сначала случайным образом сопоставляется точка в евклидовом пространстве малой размерности, например на плоскости. Между всеми точками вычисляется попарная матрица расстояний, которые ранжируются. Если ранг различия между какой-то парой образцов меньше, чем ранг расстояния между соответствующими точками, то эти точки несколько приближаются друг к другу, если больше – то удаляются друг от друга. Процедура повторяется до тех пор, пока не сойдется. Если коэффициент корреляции между рангами различий и расстояний не слишком высок, то размерность евклидова пространства увеличивается на единицу и весь процесс повторяется сначала. Получившиеся на предыдущем шаге координаты используются в качестве начальной конфигурации для нового шага. В конечном итоге точки располагаются так, что матрица расстояний между ними максимально соответствует матрице различий между объектами исходного множества. Мы получаем представление совокупности образцов множеством точек в многомерном евклидовом пространстве и можем анализировать их взаимное расположение относительно друг друга как визуально, так и с помощью стандартной техники многомерного анализа.

То, что в результате применения этого метода все образцы распались на группы, соответствующие различным стадиям заболевания, говорит о том, что предлагаемая процедура действительно позволила уловить существенные различия между образцами (рис. 2). Но анализ взаимного расположения соответствующих им точек на графике говорит гораздо больше. Точки, соответствующие предрасположенным и части нормальных, которых мы назвали суб-



нормальными, заметно сдвинулись вверх от точек, соответствующих здоровым и больным. Поскольку каждому направлению на плоскости профилей образцов (рис. 2) взаимно однозначно отвечает направление на плоскости профилей генов (рис. 3), это означает, что у субнормальных и предрасположенных экспрессируются совсем не те гены, что у здоровых и больных. Между здоровыми и больными больше общего, чем между здоровыми и предрасположенными или субнормальными и больными. Из рис. 2 очевидно, что задачу поиска генов-кандидатов, перспективных для выявления болезни Хантингтона, можно ставить по-разному. Можно искать гены, экспрессия которых различается у больных и здоровых, что обычно и делается. Но очевидно, что при этом в группу больных попадет и какая-то часть предрасположенных. А можно искать гены, экспрессия которых специфична именно для больных и выделяет их среди всех остальных, включая и субнормальных, и предрасположенных, что, собственно, и сделано в нашей работе. Наши результаты подтвердили выводы работы F. Borovecki *et al.* (2005), несмотря на разницу в методах обработки. Однако в работе H. Runne *et al.* (2007) на другой группе людей статистической разницы между больными и контролем по этим 12 генам найдено не было. Возможные причины расхождения с результатами H. Runne *et al.* (2007), на наш взгляд, могут заключаться или в различиях критериев отнесения пациентов к больным и контролю, или в несколько иной генетической природе анализируемых групп. Этот вопрос требует дальнейших исследований.

Природа вновь выявленной группы, названной нами субнормальной, достаточно загадочна. Во-первых, требует тщательного содержательного анализа то обстоятельство, что точки, соответствующие группе предрасположенных вместе с группой субнормальных, лежат не между точками, соответствующим группам здоровых и больных, как логично было бы ожидать, а заметно сдвинуты по отношению к ним (рис. 2). Во-вторых, по критериям отбора в группу здоровых четыре субнормальных индивидуума не должны быть носителями мутации по гену, кодирующему белок хантингтин. Именно увеличенное число повторов CAG в этом гене, переданное по наследству и продолжающее

увеличиваться в нейронах у потомков, проявляется обычно после 50 лет в виде болезни Хантингтона – сильного нейродегенеративного расстройства, прогрессирующего с возрастом (Landles, Bates, 2004). Однако недавно стало известно, что количество повторов в этом гене может увеличиваться с возрастом и в соматических тканях, в частности в нейронах головного мозга у формально здоровых людей, и превышать порог в 35 повторов, после которого оно считается мутацией (Shelbourne *et al.*, 2007). Таким образом, анализ различий в экспрессии генов между нормальными особями и индивидами из выделенной нами субнормальной группы мог бы пролить свет на природу факторов (по-видимому, общих для соматических и генеративных тканей), приводящих к мутированию в результате нарушений в процессе репликации или репарации ДНК участков генов, содержащих тринуклеотидные повторы. Поэтому проблема с подбором людей для контрольной группы становится очень существенной для адекватных сравнений профилей экспрессии генов и надлежащего выбора биомаркеров такого сложного метаболического заболевания, как болезнь Хантингтона.

Работа поддержана грантами НШ-2447. 2008.4, РФФИ 07-04-00441-а и программой РАН «Молекулярная и клеточная биология» (проект № 10.7).

## Литература

- Barnes M., Freudenberg J., Thompson S. *et al.* Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms // *Nucl. Acids Res.* 2005. V. 33. № 18. P. 5914–5923.
- Barrett T., Suzek T.O., Troup D.B. *et al.* NCBI GEO: mining millions of expression profiles – database and tools // *Nucl. Acids Res.* 2005. V. 33. P. 562–566.
- Bemmo A., Benovoy D., Kwan T. *et al.* Gene expression and isoform variation analysis using affymetrix exon arrays // *BMC Genomics.* 2008. V. 9. P. 529–543.
- Benjamini Y., Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing // *J. Roy. Stat. Soc. B.* 1995. V. 57. P. 289–300.
- Borovecki F., Lovrecic L., Zhou J. *et al.* Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease // *Proc. Natl*

- Acad. Sci. USA. 2005. V. 102. P. 11023–11028.
- Cheadle C., Becker K.G., Cho-Chung Y.S. *et al.* A rapid method for microarray cross platform comparisons using gene expression signatures // *Mol. and Cellular Probes*. 2007. V. 21. P. 35–46.
- Landles Ch., Bates G.P. Huntingtin and the molecular pathogenesis of Huntington's disease // *EMBO Rep*. 2004. V. 5. № 10. P. 958–963.
- Lee H.K., Hsu A.K., Sajdak J. *et al.* Coexpression analysis of human genes across many microarray data sets // *Genome Res*. 2004. V. 14. P. 1085–1094.
- Pedotti P., Hoen P.A., Vreugdenhil E. *et al.* Can subtle changes in gene expression be consistently detected with different microarray platforms? // *BMC Genomics*. 2008. V. 9. P. 124–136.
- Runne H., Kuhn A., Wild E.J. *et al.* Analysis of potential transcriptomic biomarkers for Huntington's disease in peripheral blood // *PNAS*. 2007. V. 104. № 36. P. 14424–14429.
- Shelbourne P.F., Keller-McGandy C., Bi W.L. *et al.* Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain // *Hum. Mol. Genet*. 2007. V. 16. № 10. P. 1133–1142.
- Taguchi Y.H., Oono Y. Relational patterns of gene expression via non-metric multidimensional scaling analysis // *Bioinformatics*. 2005. V. 21. № 6. P. 730–740.

## APPLICATION OF NONMETRIC MULTIDIMENSIONAL SCALING FOR ANALYSIS OF CROSS-PLATFORM GENE EXPRESSION MICROARRAY DATA

V.M. Efimov, A.V. Katokhin

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: efimov@bionet.nsc.ru

### Summary

The multidimensional method of the conformed multiplatform search of genes-markers with differential expression in the microarray data for the same set of samples is developed. The method is applied to the joint analysis of two different microarray platforms, CodeLink and Affymetrix (Borovecki *et al.*, 2005). Lists of genes-markers with differential expression corresponding to each of four groups of samples of peripheral blood (symptomatic Huntington disease patients, presymptomatic, subnorm, norm) are built.

## A COMBINATORICS-BASED DATA-MINING APPROACH TO TIME-SERIES MICROARRAY ALIGNMENT

N. Turenne<sup>1</sup>, I. Hue<sup>2</sup>

<sup>1</sup> INRA, Unité Mathématique Informatique et Génome UR1077, F-78350 Jouy-en-Josas, France, e-mail: turenne@jouy.inra.fr; <sup>2</sup> INRA, UMR 1198 Biologie du Développement et Reproduction, F-78350 Jouy-en-Josas, France, e-mail: isabelle.hue@jouy.inra.fr

One of the biological issues aiming at understanding bovine embryo development implies the analysis of proliferation and differentiation processes. Using published data from model species (mouse, human) we used a double-step classical clustering approach. First step runs a k-mean clustering for each chip individually. Second step runs a fuzzy consensus clustering to merge a few clusters (i.e. megaclusters) between microarrays. Hence we make temporal gene profiles using the symbolic time property of simultaneity and precedence according expression in ensemble of clusters. Finally with the help of a Jaccard coefficient between temporal gene profiles across species, we extract a list of genes revealing a similarity with a target gene of interest. Depending on the species and on the target gene, this list of genes differed in size and content, thus highlighting the interest of such cross-species comparisons to gain insights from different literature contexts.

**Key words:** microarray alignment, cross-species comparison, clustering, consensus, time series.

### Motivation and Aim

Large-scale biological experiments such as microarrays are now available and require data analysis to process huge amounts of results (Eisen *et al.*, 1998). One of the purposes when studying not well-know species is to refer to model-species. In our case we aimed at understanding bovine embryo development (Hue *et al.*, 2007) and analysed at first genes related to proliferation processes. Since embryo development and proliferation processes have been well studied in mouse and human species, we based our study on the comparison of published data sets in these three species. However, selecting temporal series for each species, it appeared that each dataset referred to a different time scale (detailed in datasets). Hence, comparing different sets of sequential time-series data could be done through alignment as largely done with DNA sequences since a few decades (Smith *et al.*, 1981; Altschul *et al.*, 1990) and even improved now (Kucherov *et al.*, 2004). For gene expression data the multidimensional property has not been handled so far since the classical way to align microarray

datasets deals with curves of univariate time-series (Aach, Church, 2001; Ernst *et al.*, 2005). Nevertheless, this cannot be adapted to our purpose since curves can be deformed if they are compressed in a short-time duration. Lots of combinations are available, among which the Dobinski formula  $B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}$  that gives the number of all partitions of a set of n objects. For  $n = 26$ , this formula gives  $1,6 \cdot 10^{21}$  combinations. With a 2Ghz clock, and approximating 1 cycle for a partition, time processing is about 800 years. To solve this, our approach consists in aligning only some parts of the microarrays, thus restricting the space of the alignments. We first used a classical clustering approach on each dataset and merged a few clusters by consensus to study gene interactions around genes of interest while using symbolic time property of simultaneity and precedence as described in Turenne and Schwer, 2008. Our project through microarray alignment was also to cross this kind of relational (temporal) information with known relational information from the literature or database softwares (IPA, PubMed, Gene Ontology for example). Two

corpora (document databases, PubMed) about mouse and human species have been designed and lists of gene names have been extracted from commercial software (IPA) and both corpora.

genes. Samples have been extracted at time 3hrs, 6hrs, 12hrs, 24hrs, 48hrs, 3days, 7days (GSE 3553), hence it covers 7 time points. Hereafter we called this chip: BiopH (Xu *et al.*, 2002).

**Datasets**

We compared pairwise Bovine microarray data with Human and Mouse microarray data. Bovine embryos on days 7, 14 and 19/20/21, extra-embryonic membranes on days 27/28 and fetuses on days 27/28 were collected to represent early embryo, elongating embryo, pre-implantation embryo, post-implantation extra-embryonic membrane and fetus, respectively (hence it covers 7 time-points). In total, the processed microarray contains 4,607 cDNA covering about 2,000 unique genes (GSE 1414) and call hereafter this chip: BiopB (Ushizawa, 2004). Dataset of murine embryonic developmental time course consists of morphologically staged samples from E6.25 to E9.0 (at approximately 0,25 day intervals, hence it covers 11 time points). In total, the processed microarray contains 43,000 cDNA covering 12,000 unique genes (GSE 9046) hereafter we call this chip: BiopM (Mitiku, Baker, 2007). Human embryonic stem cells were treated in pairs with or without BMP4. This was followed by RNA extraction and amplification and microarray analysis on DNA chips containing 43,000 cDNA clones, which represented about 25,000 unique

**Method and Algorithm**

Our method is currently developed under R tool and we used Clue library (method DWH) for clustering consensus (Hornik, 2005). The first step of the methodology relies on gene clustering for each microarray. This part could be done by *k*-means or descendant agglomerative hierarchy using an Euclidian distance for similarity. We obtained clusters of resembling genes through resembling expression profiles. At the second stage we merged clusters from each microarray in a way explained by Mirkin and Cherny (1970) or Meila (2005). A megacluster is a set of clusters (resulting from consensus clustering) assigned to a specific gene. A consensus distance was used for merging two partitions. The size of a partition is a vector of unique objects to classify (i.e. genes). For instance 12 items are clustered leading to a partition  $P_1$  and a partition  $P_2$ . In Partition  $P_1$  the 6 first items belong to the first microarray, the 6 following ones to the second microarray as illustrated in Table 1. Let us suppose that the first item belongs to cluster 1 (in  $P_1$ ) and cluster 16 (in  $P_2$ ). Because of this item the clusters 1 and 16 merged in the resulting partition.

**Table 1**

Example of consensus result

	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	Gene10	Gene11	Gene12
$P_1$	1	1	1	2	2	2	3	4	5	6	7	8
$P_2$	16	10	11	12	13	14	15	15	15	16	16	16
Consensus	1	1	1	2	2	2	3	3	3	1	1	1

Notes. Values represent indices of clusters.

From this example we see that gene 3 belongs to cluster 1 but as cluster 1 is merged with cluster 16, a megacluster for gene 3 will be (cluster 1, cluster 16). Hence for a target gene we can identify to which megacluster it belongs and

assess a time correlation matrix across both microarrays. Indeed, as we see below, if a cluster  $C_1$  belongs to  $P_1$  and  $C_2$  belongs to  $P_2$ , without intersection, three possibilities are possible and all of them are:

$$\begin{matrix}
 & P_1 & P_2 \\
 C_1 & 1 & 0 \\
 C_2 & 0 & 1
 \end{matrix}
 \quad \text{we deduce from the matrix}
 \quad \left\{ \begin{array}{l}
 \begin{matrix} C_1 \\ C_2 \end{matrix} \\
 C_1 C_2 \text{ cluster 1 precedes} \\
 C_2 C_1 \text{ cluster 2 precedes}
 \end{array} \right. \quad (1)$$

In our alignment method, through megaclusters, as shown below with  $M_1, M_2$ , a cluster  $M_2$  can be compared to  $M_1$  for a given time point  $t$  (of  $P_1$  in

this case). Let  $q_1(t)$  (resp.  $q_2$ ) be the measurement of a megacluster at time  $t$  for partition  $P_1$  (resp.  $P_2$ ) and  $S$  and a threshold of level expression.

$$\begin{matrix} & P_1 & P_2 \\ M_1 & 1 & 0 \\ M_2 & 0 & 1 \end{matrix} \quad \text{we deduce from the matrix} \quad \begin{cases} \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} (t, P_1) & \text{if } S \leq q_1(t) \wedge S \leq q_2(t) \\ M_1 M_2 (t, P_1) & \text{if } S \leq q_1(t) \leq q_2(t) \\ M_2 M_1 (t, P_1) & \text{if } S \geq q_2(t) \geq S \end{cases} \quad (2)$$

Through megaclusters we can more easily combine using now measurement of a partition (i.e.  $P_1$  as above for instance). Hence we can compute a time correlation matrix according to the preceding and use the symbol B for *before*, A for *after* and D if a current megacluster has a measurement greater

than a target megacluster. We can compute the time occurrence of a given Gene at a given time point compared to next or previous time point knowing the level expression (measurement) of either activation (state p), either inhibition (state m) of clusters in which it is supposed to belong, as:

Let  $Mod_t$  be a temporal mode  $e \in \{A, D, B\}$ ,  $t$  a time point,  $g$  a given gene,  $G$  a target gene and  $P$  a biochip:

$$\begin{cases} A_g(t, P) & \text{if } S \leq q_g(t + 1, P) \\ B_g(t, P) & \text{if } S \leq q_g(t - 1, P) \\ D_g(t, P) & \text{if } \frac{q_g(t, P)}{q_G(t, P)} \geq S \end{cases} \quad (3)$$

**Table 2**

Time correlation matrix for a target gene

(a)	Cluster	Target gene	$T1(P_1)$	$T2(P_1)$	$T3(P_1)$	$T4(P_1)$	$T1(P_2)$	$T2(P_2)$	$T3(P_2)$
	1	3	AD	ABD	BD	0	0	0	0
	16	3	0	A	D	B	A	D	0
(b)	State	Target gene	$T1(P_1)$	$T2(P_1)$	$T3(P_1)$	$T4(P_1)$	$T1(P_2)$	$T2(P_2)$	$T3(P_2)$
	P	3	AD	ABD	BD	B	A	D	0

Above, Table 2(a) shows an example of time correlation matrix for a target gene and its megacluster (1,16) and activation measurement; Table 2(b) gives the final resulting matrix for activation state (p), and should be completed by inhibition state (m). We use a Jaccard similarity index to compare the sub-matrix corresponding to a gene and the sub-matrix corresponding to the target gene and decide whether two genes are close to each other in their temporal profiles. These are the main steps of the algorithm in two stages. First stage is a preprocessing of data. Second stage applies consensus (Table 3).

**Results**

Complexity of consensus approach from CLUE library is  $O(n \times k)$  in memory and  $O(n \times k^3)$  in time but the DHW use an optimization solver is found on Hungarian algorithm and takes  $O(n^2)$  in space. We used a multiprocessor cluster of 162 nodes running under Sun Grid Engine, each node having 4 processors exploiting between 4 Gb and 8 Gb each one (processor Xeon EMT 64 3,2 Ghz / 4 Go; processor WoodCrest 2,33 Ghz / 8 Go). A job is assigned to the most available node. On the cluster a selection of 30 % of BiopM or BiopH microarray



**Table 3**

## Stage 1 (left), Stage 2 of Microarray Alignment algorithm

<p>Input: Two microarray datasets (D) (matrix with in column time-points), a measurement threshold (M)</p> <p>1 – implement a classical k-means clustering on 2 Datasets (D) adding a column of unique cluster index</p> <p>2 – delete measurement values &lt; M</p> <p>Output: D cleaned with a column pointing to a cluster index for each gene.</p>	<p>Input: Datasets (D), a Given Gene (G), a threshold of expression level (S), a threshold of temporal similarity (Js)</p> <p>1 – Compute mean expression values for clusters</p> <p>2 – Create Gene Dictionary D</p> <p>3 – Create 2 Partitions of Gene Dictionary with Clusters for D</p> <p>4 – Apply consensus to obtain a unique partition P</p> <p>5 – Create a Mapping MegaCluster ↔ clusters (MGC) using P</p> <p>6 – Generate the Temporal Matrix (TM) for all clusters</p> <p>7 – Compute a submatrix of TM for G (TMG) using MGC (as in Table 2. (b) )</p> <p>8 – For each gene g of D</p> <ul style="list-style-type: none"> <li>– compute submatrix (TMg) using MGC according to expressions in (3)</li> <li>– compute Jaccard value J between TMG and TMg</li> </ul> <p>Output: List of Genes Temporally Similar to G having <math>J &gt; J_s</math></p>
--	--

size and overall of BiopB lead to memory limit of computation: 8 Gb of computation to manage a unique gene partition out of 9000 genes and 5 hours (40 minutes for consensus). We have tested the new method presented in the chapter below, running it on data described in chapter Datasets. We thus observed that (i) each target gene has a different context in each array dataset (ii) this context varies depending on the similarity threshold or megacluster size and (iii) the use of temporal gene profiles based on a symbolic time property of simultaneity and precedence identifies target gene contexts which might be of high interest (Table 4).

By studying these contexts with the IPA software, we found similar gene networks around the *alg5* target gene with the bovine-human and bovine-mouse megaclusters whereas those surrounding the *eif2s3* target gene were rather different (Table 5). At first glance, these networks make sense with

those identified as important in embryos from cows or from other ungulates such as sheep or pig (recently reviewed in Blomberg, 2008). We thus feel confident that this approach is interesting to pursue. Whether these gene contexts make sense at certain stages more than others along the GSE time-series analysed here, in the understanding of the proliferation and differentiation processes involved in bovine embryo development, clearly awaits further studies.

### Conclusion

Using published microarray datasets and comparing two time series from two different species, we addressed a question for correlation between time-points and gene comparison across microarrays where time was neither equivalent nor linear between species. All combinations of occurrence

**Table 4**

Gene contexts identified with our combinatorial approach  
(Tb is threshold for Bovine, T is threshold for the other microarray)

Target genes	Similarity threshold	Bovine (B) & Human (H) arrays				Bovine (B) & Murine (M) arrays			
		megacluster (# cluster)	B & H genes	B genes	H genes	megacluster (# cluster)	B & M genes	B genes	M genes
alg5	Tb=0,7; T=0,9	16	14	18	0	12	25	43	37
	Tb=0,7; T=0,1	11	14	18	0	15	12	20	0
eif2s3	Tb=0,7; T=0,9	16	12	10	0	15	208	298	2265
	Tb=0,7; T=0,1	10	76	81	574	5	6	16	0

**Table 5**

Gene contexts identified with our combinatorial approach and analysed with the IPA software.  
The highest scores indicate the more significant functions (or top functions) identified in these gene contexts

Genes	Networks	Score
Alg5 bov hum = Alg5 bov mus	Connective tissue disorders, genetic disorders, cancer	22
	Cancer, cell to cell signalling and interaction, cellular assembly and organisation	14
Eif2s3 bov hum	Molecular transport, organ morphology, reproductive system development and function	24
Eif2s3 bov mus	Reproductive system disease, cardiovascular system development and function, organismal development	46
	Cancer, cell to cell signalling and interaction, cellular function and maintenance	22
	Cell cycle, cellular assembly and organisation, DNA replication, recombination and repair	22
	Cell cycle, cell morphology, connective tissue development and function	18
	Post-translational modification, protein folding, cancer	11

of genes between time points are possible, hence all partitions. We propose a methodology based on merged clusters and use a time correlation matrix to compute a time profile over two microarrays. This kind of combinatorial-mining approach could be a first step to multidimensional sequence alignment which seems close to known 1-dimension sequence alignment but is more highly combinatorial.

### References

- Aach J., Church G.M. Aligning gene expression time series with time warping algorithms // *Bioinformatics*. 2001. V. 17. № 6. P. 495–508.
- Altschul S.F., Gish W., Miller W. *et al.* Basic local alignment search tool // *J. Mol. Biol.* 1990. 215 (3). P. 403–410.
- Blomberg L., Hashizume K., Viebahn C. Blastocyst elongation, trophoblastic differentiation, and embryonic pattern formation // *Reproduction*. 2008. 135. P. 181–195.
- Eisen M.B., Spellman P., Brown P.O., Botstein D. Cluster Analysis and Display of Genome-Wide Expression Patterns // *Proc. Natl Acad. Sci. USA*. 1998. 95(25). P. 14863–14868.
- Ernst J., Nau G.J., Bar-Joseph Z. Clustering short time series gene expression data // *Bioinformatics*. 2005. V. 21. Suppl. 1. P. 159–168.
- Hornik K. A CLUE for CLUster Ensembles // *J. Stat. Software*. 2005. 14(12).
- Hue I., Degrelle S.A., Campion E., Renard J.P. Gene expression in elongating and gastrulating embryos from ruminants // *Soc. Reprod. Fertil Suppl.* 2007. 64. P. 365–377. Review.
- Kucherov G., Noй L., Ponty Y. Estimating seed sensitivity on homogeneous alignments // *Proc. of the IEEE 4th Symp. on Bioinformatics and Bioengineering (BIBE)*. May 19–21, 2004, Taichung (Taiwan). P. 387–394. IEEE Computer Society Press.
- Meila M. Comparing clusterings - an axiomatic view // *Proc. of the 22nd Intern. Conf. on Machine Learning (ICML'2005)*.
- Mirkin B., Cherny L. Deriving a distance between partitions on a finite set // *Automation and Remote Control*. 1970. № 5. P. 120–127.
- Mitiku N., Baker J.C. Genomic analysis of gastrulation and organogenesis in the mouse // *Dev. Cell*. 2007. 13(6). P. 897–907.
- Smith T.F., Waterman M.S. Identification of common molecular subsequences // *J. Mol. Biol.* 1981. 147. P. 195–197.
- Turenne N., Schwer S.R. Temporal Representation of Gene Networks // *J. of Data Mining and Bioinformatics (JDMB)*. 2008. V. 2(1).
- Ushizawa K., Herath C.B. *et al.* cDNA microarray analysis of bovine embryo gene expression profiles during the pre-implantation period // *Reprod. Biol. Endocrinol.* 2004. 24. 2:77.
- Xu R.H., Chen X., Li D.S. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast // *Nat. Biotechnol.* 2002. 20(12). P. 1261–1264.

## ФУНКЦИОНАЛЬНАЯ АННОТАЦИЯ АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ЛОКАЛЬНОГО СХОДСТВА

К.Е. Александров, Б.Н. Соболев, Д.А. Филимонов, В.В. Поройков

ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН, Москва, Россия,  
e-mail: dzimmu@yandex.ru

Разработан новый метод распознавания функциональных классов белков на основе оригинального способа описания аминокислотной последовательности. Каждая последовательность из обучающей выборки сравнивается с аннотируемой последовательностью, на основе чего вычисляются оценки локального сходства для каждой аминокислотной позиции. Эти оценки используются в качестве входных данных для оригинального классификатора. Метод тестировался на 56 классах белков «Золотого стандарта» («Gold Standard» (Brown *et al.*, 2006)), сериновых протеазах и других классах белков. Разработанная нами программа показала высокую точность предсказания – для большинства классов 100 %-я точность. При отнесении белков к классам Международной классификации ферментов (ЕС) наша программа превосходит по точности программу SVMProt (на основе метода опорных векторов) и сопоставима с программами HMMer (на основе скрытых Марковских моделей) и PROF\_PAT (на основе паттернов множественных мотивов). Описанный метод предполагается использовать как для предсказания функциональных классов белков, так и для поиска сайтов функциональной специфичности в аминокислотных последовательностях.

**Ключевые слова:** функциональная аннотация, аминокислотная последовательность, точность предсказания.

### Введение

Функциональная аннотация вновь секвенированных генов представляет собой одну из наиболее важных задач в биоинформатике. Лишь малая часть белков, кодируемых известными нуклеотидными последовательностями, охарактеризована экспериментально. Это обуславливает необходимость развития методов компьютерной аннотации аминокислотных последовательностей. Действительно, в базе данных UniProt содержится более 4 млн аминокислотных последовательностей; база данных Gene Ontology содержит около 160 000 функциональных аннотаций, из которых экспериментально охарактеризовано менее 10 000. Первоначально компьютерная аннотация аминокислотных последовательностей производилась на основе гомологии с экспериментально охарактеризованными белками (Devos, Valencia, 2000, 2001). Этот подход обеспечивает

особенно точное предсказание при использовании филогенетических методов, которые зачастую требуют ручной корректировки. Весьма полезны методы, представляющие наборы выровненных последовательностей (белковые семейства) в обобщенном виде с помощью регулярных выражений, позиционных частотных матриц (профилей) и скрытых Марковских моделей. Этот подход позволяет одновременно решать задачи классификации и функционального картирования. Методы этой группы используются в популярных информационных ресурсах: PROSITE (Hofmann *et al.*, 1999), BLOCKS (Henikoff S., Henikoff J., 1991; Henikoff *et al.*, 1999), PRINTS (Attwood *et al.*, 1999), PROF\_PAT (Bachinsky *et al.*, 1997), PFAM (Finn *et al.*, 2008). В связи с необходимостью классификации непрерывно растущего числа последовательностей в последнее время все более популярными становятся автоматизированные методы, основанные на обучении

с использованием выборки экспериментально аннотированных белков без процедуры выравнивания. В этой группе методов используются такие подходы, как наивный Байесовский классификатор, искусственные нейронные сети, метод  $k$ -ближайших соседей, дерево решений и метод опорных векторов (Han *et al.*, 2006). Методы машинного обучения показывают высокую точность распознавания функциональных классов – для ряда классов более 95 %. Однако многие функциональные группы не могут быть предсказаны с необходимой точностью (Cai *et al.*, 2003). Поэтому задача предсказания функции белка далека от окончательного решения.

В данной работе мы предлагаем новый метод машинного обучения PAAS (Проекция аминокислотных последовательностей, Projections of Amino Acid Sequences), который основан на новом способе представления аминокислотных последовательностей и оригинальном алгоритме классификации.

### Представление аминокислотных последовательностей

Мы предлагаем описывать аннотируемую последовательность  $A$  как совокупность оценок локального сходства данной последовательности с последовательностью  $B$ , взятой из

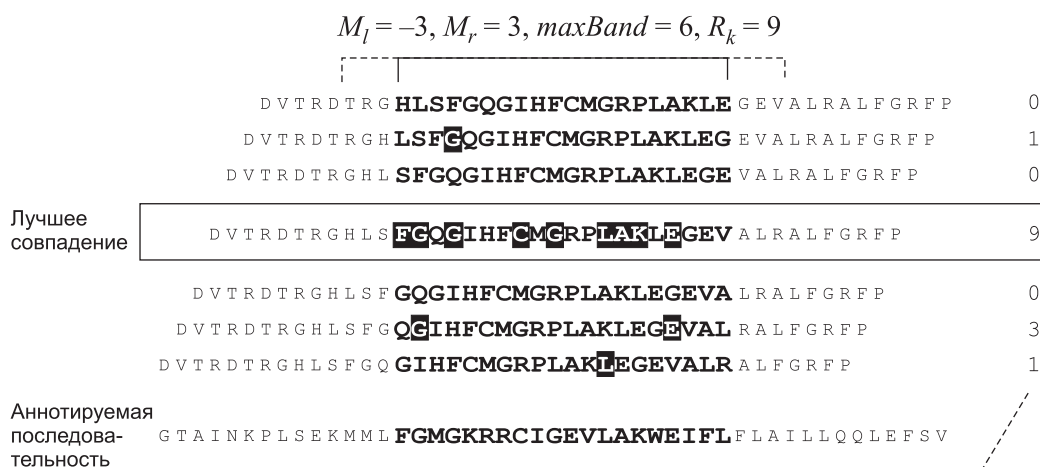
обучающей выборки. Эта выборка содержит аминокислотные последовательности белков с установленными функциональными характеристиками. Первичные оценки локального сходства рассчитываются с помощью серии сдвигов последовательности  $A$  относительно последовательности  $B$  (рис. 1). Каждый фрагмент последовательности  $A$  (длина которого определяется параметром «frame») сравнивается с каждым совмещенным фрагментом последовательности  $B$ .

Первичная оценка  $R_k$  для позиции  $k$  последовательности  $A$  равна мере сходства между фрагментом последовательности  $A$  и наиболее сходным с ним фрагментом последовательности  $B$  (в простейшем случае – числу совпадающих остатков) согласно следующему выражению:

$$R_k = \max_j (I_{k+F-1,j} - I_{kj}), \tag{1}$$

$$I_{kj} = \sum_{i=1}^k s(a_i, b_{i+j}), \quad M_l \leq j \leq M_r,$$

где  $R_k$  – первичная оценка сходства последовательности  $A$  с последовательностью  $B$  в позиции  $k$ ,  $F$  – значение параметра «frame»,  $s(a_i, b_{i+j})$  – мера сходства остатков  $a_i$  и  $b_{i+j}$  из последовательностей  $A$  и  $B$ ,  $j$  – величина сдвига. Величина сдвига  $j$  ограничена справа и слева величинами  $M_l$  и  $M_r$ , которые определяются параметром  $maxBand$ . Если последовательности равны по длине, то  $M_l = -maxBand / 2$ , а  $M_r = maxBand / 2$ .



Первичные оценки (количество совпадающих аминокислотных остатков в последовательности из обучающей выборки и аннотируемой последовательности)

Рис. 1. Расчет первичных оценок локального сходства.

Сглаженная оценка локального сходства  $S_k$  для позиции  $k$  определяется как максимальная из всех оценок  $R_k$ , рассчитанных для всех участков последовательности  $A$ , которые включали данную позицию  $k$ .

Описанная процедура поиска сходных участков напоминает построение матрицы сходства (McLachlan, 1971). Как известно, парное выравнивание может рассматриваться как отбор таких диагональных фрагментов, которые обеспечивают наилучшую оценку выравнивания. При этом отобранные фрагменты сосредоточены в достаточно узкой полосе. Наш подход позволяет учитывать такие фрагменты, которые могут соответствовать функционально важным участкам, но игнорируются при построении оптимального выравнивания. Выявление отдельных гомологичных фрагментов на матрице сходства, не обязательно совпадающих с общим выравниванием, применялось и ранее (Туманян, Поройков, 1984). Новизна нашего подхода состоит в том, что рассчитанные для всех позиций аннотируемой последовательности оценки локального сходства  $S_k$  применяются как входные данные для программы-классификатора.

Оценки локального сходства могут использоваться и для функционального картирования аминокислотных последовательностей. На рис. 2 показано распределение усредненных оценок локального сходства аннотируемой последовательности и последовательностей трех классов. Локальные максимумы на соответствующей



**Рис. 2.** Распределение усредненных (по классам ЕС) оценок локального сходства.

Утолщенная линия соответствует классу, к которому в действительности относится аннотируемая последовательность.

кривой позволяют отличить класс, к которому действительно принадлежит аннотируемый белок, от других классов.

В данной работе мы использовали простейшую меру сходства аминокислотных остатков: 1 для идентичных остатков и 0 для различающихся (было показано, что использование матрицы замен не приводит к повышению точности предсказания).

### Алгоритм классификации

Используемый алгоритм классификации основан на наивном Байесовском классификаторе (Alexandrov *et al.*, 2008). Принадлежность аннотируемой последовательности к классу  $A$  оценивается с помощью специальной статистики (В-статистики), рассчитанной по следующим формулам:

$$t_0 = \frac{\sum_{k=1}^N [W_k(A) - W_k(\neg A)]}{\sum_{k=1}^N [W_k(A) + W_k(\neg A)]}, \quad (2)$$

$$t_i = \frac{\sum_{k=1}^N S_{ik} [W_k(A) - W_k(\neg A)]}{\sum_{k=1}^N S_{ik} [W_k(A) + W_k(\neg A)]}, \quad (3)$$

$$t = \text{Sin} \left[ \frac{1}{n} \sum_{i=1}^n \text{ArcSin}(t_i) \right], \quad (4)$$

$$B = \frac{t - t_0}{1 - t t_0}, \quad (5)$$

где  $N$  – количество последовательностей в обучающей выборке,  $W_k(A)$  и  $W_k(\neg A)$  – веса последовательности  $k$  в классе  $A$  и в его дополнении  $\neg A$ ,  $S_{ik}$  – оценка локального сходства аннотируемой последовательности в позиции  $i$  с последовательностью  $k$  из обучающей выборки,  $n$  – количество аминокислотных остатков в аннотируемой последовательности.

### Оценка точности предсказания

Для оценки точности предсказания мы использовали скользящий контроль с исключением по одному. На каждом этапе этой процедуры из обучающей выборки мы удаляли одну последовательность и использовали ее в качестве



аннотируемой. Для каждого класса  $A$  на основе полученных значений В-статистики рассчитывался независимый критерий точности прогноза (Independent Accuracy of Prediction, IAP):

$$IAP = \frac{\sum_{i,j} \theta(B_{i \in A} - B_{j \in \neg A})}{N_A \cdot N_{\neg A}}, \quad (6)$$

где  $B_i$  – оценка принадлежности последовательности  $i$  классу  $A$ , если  $i$  действительно принадлежит классу  $A$ ;  $B_j$  – оценка принадлежности последовательности  $j$  классу  $A$ , если  $j$  на самом деле принадлежит его дополнению  $\neg A$ ;  $\theta(x) = 1$ , если  $x > 0$ ,  $\theta(x) = 1/2$ , если  $x = 0$ ,  $\theta(x) = 0$ , если  $x < 0$ ;  $N_A$  – количество последовательностей в классе  $A$ ,  $N_{\neg A}$  – количество последовательностей в дополнении  $A$ . При 100 %-й точности предсказания значение  $IAP = 1$ .

### Проверка точности метода

Мы провели проверку точности метода с использованием процедуры скользящего контроля с исключением по одному на двух выборках. Одна из них представляла сериновые протеазы – 566 последовательностей, разделенных на классы согласно ЕС. Другая выборка содержала 817 последовательностей из так называемого «Золотого стандарта» (Brown *et al.*, 2006), представляющих 56 семейств, различающихся по функциональным характеристикам и объединенных в 5 надсемейств. На рис. 3 показаны значения IAP, усредненные по всем группам

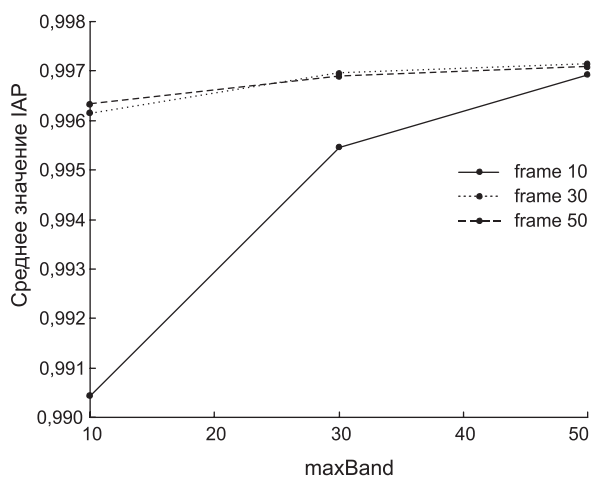


Рис. 3. Проверка точности метода на выборке сериновых протеаз.

выборки сериновых протеаз, при разных значениях  $frame$  и  $maxBand$ .

Результаты проверки точности метода на выборке «Золотого стандарта» схожи с таковыми для сериновых протеаз; 45 из 56 семейств «Золотого стандарта» предсказывались со 100 %-й точностью (рис. 4). Однако наблюдаются различия для семейств и надсемейств. В случае с надсемействами достижение высоких значений IAP происходит при значении  $frame = 50$ , в то время как для семейств точность предсказания достигает максимума при меньшем значении этого параметра (30) (рис. 5). По-видимому, это отражает большую дивергенцию после-

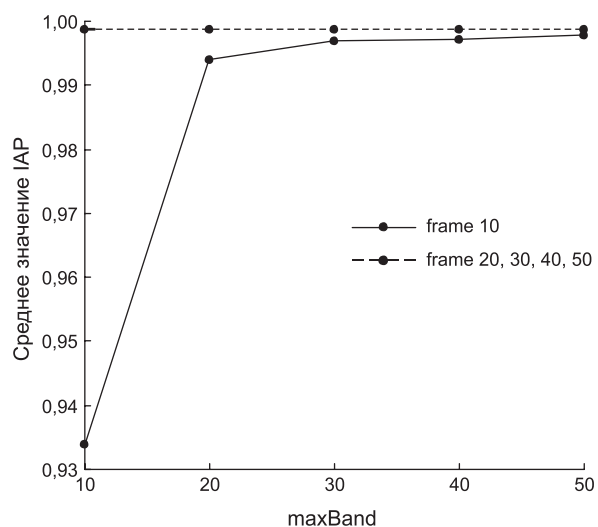


Рис. 4. Проверка точности метода на выборке семейств «Золотого стандарта».

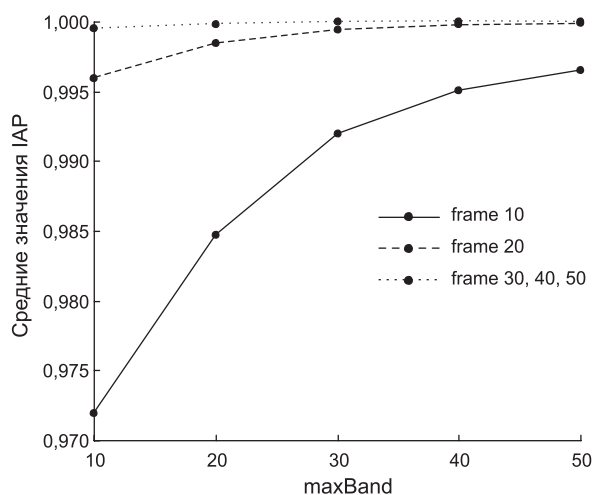


Рис. 5. Проверка точности метода на выборке надсемейств «Золотого стандарта».

довательностей, входящих в надсемейства по сравнению с семействами.

### Сравнение с другими методами

Для сравнения точности предсказания функциональных классов с помощью нашего метода мы выбрали программы, основанные на известных методах: HMMer, SVMProt и PROF\_PAT.

Программа SVMProt (Cai *et al.*, 2003) реализует один из методов машинного обучения (метод опорных векторов) и так же, как PAAS, работает с обучающей выборкой из невыровненных последовательностей. Сравнение предложенного метода с методом опорных векторов проводилось на основе данных о чувствительности и специфичности предсказания для 46 различных классов белков, приведенных авторами программы SVMProt. На основании этих данных мы рассчитали значения IAP.

$$IAP = \frac{\frac{TP}{TP + FN} + \frac{TN}{FP + TN}}{2}, \quad (7)$$

где  $TP$  – число истинно положительных,  $TN$  – число истинно отрицательных,  $FP$  – число ложноположительных и  $FN$  – число ложноотрицательных результатов.

Для расчета значений IAP при предсказании тех же 46 классов нашим методом мы применили процедуру скользящего контроля с исключением по одному. Расчеты производились при значении  $frame = 50$  и различных значениях  $maxBand$ . Источником последовательностей для обучающей выборки была база данных Brenda (<http://www.brenda.unikoeln.de>).

Средняя точность предсказания с помощью нашего метода сравнима с точностью предсказания программы SVMProt и даже превышает ее при значении  $maxBand = 5000$ . При этом значении наш метод предсказывает 32 класса с более высоким IAP и 14 классов с более низким IAP, чем SVMProt. Стоит отметить, что с помощью PAAS 4 класса белков были предсказаны со 100 %-й точностью. Однако есть классы, для которых точность предсказания с помощью PAAS относительно низка. Например для класса EC 4.4 величина IAP составляет 0,5971. В данном случае низкая точность прогноза объясняется тем, что данный класс обучающей выборки состоит из 7 белков, относящихся к 5 разным белковым

семействам (согласно классификации PFAM), так что 4 белка из 7 относятся к 4 различным классам. По-видимому, высокая точность прогноза наблюдается для тех функциональных классов, которые включают одну или несколько хорошо представленных групп гомологичных последовательностей.

Данные сравнения нашего метода с SVMProt приведены в табл. 1.

Программа HMMer (<http://hmmer.janelia.org/>) использует скрытые Марковские модели. Для оценки этого метода из «Золотого стандарта» были отобраны 5 надсемейств и 8 семейств, обладающих номерами EC. Представительность отобранных классов составляла от 7 до 215 последовательностей. Используя программу ClustalW со стандартными настройками, мы получили выравнивания для каждого класса; «ручная» коррекция выравниваний не проводилась. На основе полученных выравниваний были построены скрытые Марковские модели для соответствующих классов, а затем проводилось сопоставление всех отобранных последовательностей «Золотого стандарта» с построенными моделями. На основе выходных данных программы HMMer для каждого класса можно было определить пороговое значение: все последовательности, действительно относящиеся к данному классу, получали оценки выше порога, а все неотносящиеся – ниже порога. Можно сказать, что HMMer продемонстрировал 100 %-ую точность распознавания отобранных классов. В табл. 2 приведены оценки точности предсказания (на основе рассчитанных значений IAP) для тех же классов, полученные с помощью нашего метода ( $frame = 50$  и  $maxBand = 50$ ) при проведении скользящего контроля с исключением по одному.

Таким образом, программная реализация предложенного метода уступает программе HMMer только в 1 из 13 классов.

Программа PROF\_PAT ([http://www.mgs.bionet.nsc.ru/mgs/programs/prof\\_pat/](http://www.mgs.bionet.nsc.ru/mgs/programs/prof_pat/)) использует паттерны белковых семейств, создаваемые в автоматическом режиме на основе выровненных последовательностей из баз данных SwissProt и TrEMBL. База данных PROF\_PAT содержит паттерны более чем 13 000 белковых семейств. Для ориентировочного сопоставления точности данного метода с точностью PAAS мы исполь-

**Таблица 1**

**Сравнение PAAS и SVMProt\***

Функциональный класс	IAP для SVMProt	IAP для PAAS	ΔIAP
EC 1.1	0,9164	0,9746	<b>0,0582</b>
EC 1.2	0,9195	0,9719	<b>0,0524</b>
EC 1.3	0,8397	0,9273	<b>0,0876</b>
EC 1.4	0,881	0,9542	<b>0,0732</b>
EC 1.5	0,7469	0,8766	<b>0,1297</b>
EC 1.6	0,9469	0,7756	-0,1713
EC 1.8	0,8246	0,8755	<b>0,0509</b>
EC 1.9	0,976	0,7673	-0,2087
EC 1.10	0,8438	0,6175	-0,2263
EC 1.11	0,9161	0,9942	<b>0,0781</b>
EC 1.13	0,9087	0,8195	-0,0892
EC 1.14	0,9141	0,9682	<b>0,0541</b>
EC 1.15	0,9426	0,8597	-0,0829
EC 1.17	0,8956	0,5836	-0,312
EC 2.1	0,8581	0,8209	-0,0372
EC 2.2	0,9179	0,8652	-0,0527
EC 2.3	0,8947	0,942	<b>0,0473</b>
EC 2.4	0,8722	0,9399	<b>0,0677</b>
EC 2.5	0,8746	0,9081	<b>0,0335</b>
EC 2.6	0,8885	0,9913	<b>0,1028</b>
EC 2.7	0,8259	0,95	<b>0,1241</b>
EC 2.8	0,7869	0,984	<b>0,1971</b>
EC 3.1	0,5747	0,9505	<b>0,3758</b>
EC 3.2	0,897	0,9675	<b>0,0705</b>
EC 3.3	0,9375	0,8125	-0,125
EC 3.4	0,8717	0,9623	<b>0,0906</b>
EC 3.5	0,8355	0,8921	<b>0,0566</b>
EC 3.6	0,9268	0,9623	<b>0,0355</b>
EC 4.1	0,8959	0,8463	-0,0496
EC 4.2	0,8632	0,9129	<b>0,0497</b>
EC 4.3	0,9196	0,7773	-0,1423
EC 4.4	0,7495	0,5971	-0,1524
EC 4.6	0,7604	1	<b>0,2396</b>
EC 5.1	0,8362	0,802	-0,0342
EC 5.2	0,8255	0,969	<b>0,1435</b>
EC 5.3	0,9293	0,7567	-0,1726
EC 5.4	0,8537	0,9479	<b>0,0942</b>
EC 6.1	0,9406	0,9836	<b>0,043</b>
EC 6.2	0,8717	0,9985	<b>0,1268</b>
EC 6.3	0,8961	0,9124	<b>0,0163</b>
EC 6.4	0,9389	1	<b>0,0611</b>
EC 6.5	0,8446	1	<b>0,1554</b>

## Окончание таблицы 1

Функциональный класс	IAP для SVMProt	IAP для PAAS	$\Delta$ IAP
Рецепторы, сопряженные с G-белком	0,9616	0,9977	<b>0,0361</b>
Ядерные рецепторы	0,9352	1	<b>0,0648</b>
Тирозинкиназные рецепторы	0,8562	0,955	<b>0,0988</b>
Управляемые электрохимическим потенциалом транспортеры (симпортеры, унипортеры, антипортеры)	0,923	0,9872	<b>0,0642</b>

\*IAP для SVMProt – значения IAP, рассчитанные по данным, приведенным в статье Cai *et al.* (2003); IAP для PAAS – значения IAP для результатов предсказания, полученных с помощью метода PAAS при значениях frame = 50 и maxBand = 5000;  $\Delta$ IAP – разность между значениями IAP для SVMProt и PAAS (в тех случаях, когда точность PAAS выше точности SVMProt, т. е.  $\Delta$ IAP > 0, значения выделены).

**Таблица 2**  
Сравнение точности предсказания  
PAAS и HMMer

Функциональный класс	PAAS	HMMer
Амидогидролазы (sf)	++	++
Кротоназы (sf)	++	++
Енолазы (sf)	++	++
VOC (sf)	++	++
Галоацидные дегалогеназы (sf)	+	++
Гистоновая ацетилтрансфераза(f)	++	++
Енолаза (f)	++	++
АМФ-деаминаза (f)	++	++
D-гидантоиназа (f)	++	++
Дигидрооротаза 2 (f)	++	++
Гуаниновая деаминаза (f)	++	++
АТФаза р-типа (f)	++	++
Уреаза (f)	++	++

«++» – точность предсказания 100 %; «+» – точность предсказания > 98 %; «f» – семейства, «sf» – надсемейства.

зовали часть последовательностей «Золотого стандарта», для которых известен номер ЕС. Это условие было введено для облегчения сопоставления результатов предсказания с помощью программы PROF\_PAT. Были установлены следующие значения параметров поиска – матрица аминокислотных замен Blosum62 и 100 %-й уровень сходства (similarity level). Часть семейств «Золотого стандарта» не имеет

номеров ЕС, а часть результатов предсказания с помощью программы PROF\_PAT представлены в виде названий функциональных классов, а не в виде номеров ЕС. С помощью программы PROF\_PAT для 89,9 % последовательностей «Золотого стандарта» были предсказаны номера ЕС, которые действительно относятся к соответствующим белкам. Таким образом, эффективность программы PAAS сопоставима с эффективностью метода PROF\_PAT.

### Выводы

Предложенный подход показал высокую точность предсказания при различных уровнях функциональной классификации белков, сопоставимую с точностью трех широко используемых методов функциональной аннотации. Метод обладает рядом преимуществ: а) формализация данных о локальном сходстве последовательностей в сочетании с использованием оригинального классификатора обеспечивает высокую точность предсказания; б) предоставляется возможность как классифицировать новые аминокислотные последовательности, так и строить их функциональные карты; в) предложенный алгоритм может быть реализован в виде автоматизированного инструмента функциональной аннотации белков; г) программная реализация метода требует небольших вычислительных ресурсов, что позволяет проводить массовую аннотацию последовательностей.

## Литература

- Туманян В.Г., Поройков В.В. Установление оптимального соответствия между аминокислотными (нуклеотидными) последовательностями // Биофизика. 1984. Т. 24. № 6. С. 917–920.
- Alexandrov K., Sobolev B., Filimonov D., Poroikov V. Recognition of protein function using the local similarity // J. Bioinf. Comp. Biol. 2008. V. 6. № 4. P. 709–725.
- Attwood T.K., Flower D.R., Lewis A.P. *et al.* PRINTS prepares for the new millennium // Nucl. Acids Res. 1999. V. 27. P. 220–225.
- Bachinsky A.G., Yarigin A.A., Guseva E.H. *et al.* A bank of protein family patterns for rapid identification of possible functions of amino acid sequences // Comput. Appl. Biosci. 1997. V. 13. № 2. P. 115–122.
- Brown S.D., Gerlt J.A., Seffernick J.L., Babbitt P.C. A gold standard set of mechanistically diverse enzyme superfamilies // Genome Biol. 2006. V. 7. № 1. R8.
- Cai C.Z., Han L.Y., Ji Z.L. *et al.* SVM-Prot: Web-based support vector machine software for functional classification of a protein from its primary sequence // Nucl. Acids Res. 2003. V. 31. P. 3692–3697.
- Devos D., Valencia A. Intrinsic errors in genome annotation // Trends Genet. 2001. V. 17. № 8. P. 429–431.
- Devos D., Valencia A. Practical limits of function prediction // Proteins. 2000. V. 41. P. 98–107.
- Finn R.D., Tate J., Mistry J. *et al.* The Pfam protein families database // Nucl. Acids Res. 2008. V. 36 (Database issue):D281–8. Epub 2007. Nov 26.
- Han L., Cui J., Lin H. *et al.* Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity // Proteomics. 2006. V. 6. № 14. P. 4023–4037.
- Henikoff S., Henikoff J.G. Automated assembly of protein blocks for database searching // Nucl. Acids Res. 1991. V. 19. P. 6565–6572.
- Henikoff J.G., Henikoff S., Pietrovski S. New features of the Blocks Database servers // Nucl. Acids Res. 1999. V. 27. P. 226–228.
- Hoffmann K., Bucher P., Falquet L., Bairoch A. The PROSITE database, in status in 1999 // Nucl. Acids Res. 1999. V. 27. P. 215–219.
- McLachlan A.D. Test for comparing related amino acid sequences: Cytochrome C and cytochrome C551 // J. Mol. Biol. 1971. V. 61. P. 409–424.

## FUNCTIONAL ANNOTATION OF THE AMINO ACID SEQUENCES USING LOCAL SIMILARITY

K. Alexandrov, B. Sobolev, D. Filimonov, V. Poroikov

Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, Russia,  
e-mail: dzimmu@yandex.ru

## Summary

We have developed a new method, which enables to recognize the protein functional classes based on the original description of an amino acid sequence. Each sequence of the training set is compared with the annotated sequence and local similarity scores for all amino acid positions are calculated. These scores are used as input data for the original classifier. The method was tested on 56 classes of proteins included into the Gold Standard (Brown *et al.*, 2006), serine proteases and other protein classes. Protein classes mentioned above were predicted with high accuracy; most of them were predicted with 100 % accuracy. Our program predicted Enzyme classification classes with the accuracy superior to SVMProt program (based on the support vector machine) and comparable with HMMer (based on hidden Markov models) and PROF\_PAT (based on the multiple motif patterns). We suppose that the suggested method can be used in the prediction of the functional classes of proteins and in the revealing of functional specificity sites in the amino acid sequences.



## АНАЛИЗ РАСПРЕДЕЛЕНИЯ АДЕНОЗИНФОСФАТ-СВЯЗЫВАЮЩИХ САЙТОВ БЕЛКОВ НА ЭКЗОННОЙ СТРУКТУРЕ ГЕНА

И.В. Медведева, П.С. Деменков, В.А. Иванисенко

Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: brukaro@bionet.nsc.ru; salix@bionet.nsc.ru

Проблема сопоставления пространственной структуры белка и кодирующей его последовательности гена возникла с открытием первых методов кристаллографии. Информация о пространственных структурах белков, в частности, позволила с высокой точностью определять их доменную структуру. Многие работы посвящены исследованию взаимосвязи границ экзонов и границ доменов белков, кодируемых соответствующим геном. В данной работе было поставлено целью выявление взаимосвязи между функциональными сайтами белков и экзонной структурой кодирующих их генов на примере сайтов связывания аденозинфосфатов. Показано, что длина экзонов, кодирующих аденозинфосфат-связывающие сайты, предположительно меньше длины экзонов, не участвующих в кодировании этих сайтов. Кроме того, аминокислоты рассматриваемых сайтов связывания удалены друг от друга на первичной структуре белка, несмотря на то что большая часть этих функциональных сайтов кодируется соседними по последовательности соответствующего гена экзонами. Также оказалось, что аденозинфосфат-связывающие сайты формируются не только внутри одного домена белка, но и из других участков последовательности, включая другие домены. Распределение аминокислот функционального сайта относительно положения кодирующего экзона в последовательности показало, что сайты связывания АТФ, АДФ, АМФ в большинстве случаев кодируются первыми экзонами последовательности.

**Ключевые слова:** функциональные сайты белков, экзонная структура гена, аденозинфосфат-связывающие сайты.

### Введение

В 1960-х гг. появились экспериментальные методы определения пространственной структуры белка с помощью кристаллографического анализа и ядерного магнитного резонанса (Richards, 1963; Andreeva, 1964). Экспериментальные методы позволили идентифицировать структурные домены белков, т. е. минимальный по длине фрагмент последовательности, обладающий способностью к самосборке в глобулярную пространственную структуру, идентичную структуре этого фрагмента в составе полноразмерного белка. В настоящее время существует множество свидетельств того, что структурные домены также могут сохранять функциональные свойства, которые им были присущи в составе полноразмерных белков (Kaessmann *et al.*, 2002). С появлением данных

о пространственной структуре белков возникли вопросы о том, как же соотносятся между собой структурная организация белков и структура кодирующих их генов. Одной из первых задач было выявление соответствия между границами экзонов гена и границами кодируемых этим геном доменов белка. В частности, это могло объяснить молекулярные механизмы возникновения многодоменных белков, некоторые из доменов которых были идентичны у разных организмов, а остальные домены различались.

Гилбертом было выдвинуто предположение, согласно которому один экзон гена кодирует один домен соответствующего белка (Gilbert, 1978). Это предположение давало простое объяснение появлению таких многодоменных белков за счет возможной перетасовки экзонов, при которой происходят встройки экзонов одного гена в экзонно-интронную структуру дру-

гого гена. Явление перетасовки экзонов было открыто следом за работой Гильберта (Maki *et al.*, 1980). Позже была показана ключевая роль такой перетасовки в перестройке доменной структуры белков в ходе эволюции (Kaessmann *et al.*, 2002). В то же время были показаны случаи, когда один домен может кодироваться несколькими экзонами так же, как и один экзон может кодировать несколько доменов (Graug *et al.*, 2000). Эволюционные перестройки доменной структуры таких белков могли происходить за счет множественной перетасовки экзонов, кодирующих один домен белка (Kaessmann *et al.*, 2002).

Функция белков в значительной степени определяется их функциональными сайтами, которые могут кодироваться как одним, так и несколькими экзонами. Особенности организации функциональных сайтов белков на уровне экзонной структуры гена могут иметь большое значение для таких эволюционных событий, как формирование доменных структур белков, в том числе и в результате перетасовок экзонов (Ponting *et al.*, 2002). Однако работ по анализу особенностей распределения функциональных сайтов белков на экзонной структуре генов пока не проведено. Исследованию этого вопроса на примере сайтов связывания АТФ, АДФ, АМФ (аденозинфосфат-связывающие сайты, АФСС) посвящена данная работа. Метаболические, а также регуляторные процессы, проходящие с участием этих лигандов, распространены во всех организмах. Белки, вовлеченные в эти процессы путем связывания аденозинфосфатов, как правило, имеют длительную эволюционную историю (Oswald *et al.*, 2006). Поиск взаимосвязей между структурно-функциональной организацией белков и экзонной структурой генов может иметь большое значение для понимания процессов молекулярной эволюции.

### Материалы и методы

Для получения информации о разметке сайтов связывания АТФ, АДФ, АМФ на последовательностях белков использовалась база данных (БД) функциональных сайтов белков PDBSite (Ivanisenko *et al.*, 2005). Эта база данных содержит информацию о координатах атомов аминокислотных остатков различных

типов функциональных сайтов, а также об их положении в последовательностях белков, представленных в БД пространственных структур белков PDB (Berman *et al.*, 2000). Поскольку последовательности одного и того же белка, приведенные в БД PDB и БД SWISS-Prot (Watanabe, Narayama, 2001), могут различаться, то проводилось их выравнивание с использованием программы BLAST (Altschul *et al.*, 1997). Различия в последовательностях этих двух баз связаны с тем, что в БД PDB часто приводятся фрагменты белка, которые удалось кристаллизовать, а также мутантные формы белков. Таким образом, с помощью выравнивания нами были установлены позиции аденозинфосфат-связывающих сайтов в последовательностях SWISS-Prot. Дополнительно на этих последовательностях проводилась разметка границ экзонов, которая извлекалась из БД EMBL (Cochrane *et al.*, 2008). Для этих целей нуклеотидную последовательность транслировали в аминокислотную, которая затем выравнивалась с последовательностью соответствующего белка из базы данных SWISS-Prot. В данном случае выравнивание осуществлялось с помощью программы ClustalW (Thompson *et al.*, 2002).

Анализ проводился на наборе негомолочных белков с уровнем сходства между собой менее 30 %. Рассматривались только многоклеточные организмы и только те гены, для которых была информация об их полной экзон-интронной структуре. Таким образом, для анализа было отобрано 63 белка, содержащих аденозинфосфат-связывающие сайты следующих организмов: человек – 51 белок (81 %); мышь – 3 белка (5 %); нематода – 3 белка (5 %); бык – 2 белка (3 %); крыса – 2 белка (3 %); плодовая мушка – 1 белок (2 %); курица – 1 белок (2 %). В выборке белков были представлены все 6 классов ферментов по номенклатуре ЕС: оксидоредуктазы (4 белка), трансферазы (28 белков), изомеразы (1 белок), лигазы (4 белка), лиазы (2 белка), гидролазы (10 белков). Доля ферментов составила 71 %, доля транспортных белков 8 % (5 белков), доля белков клеточного цикла 5 % (3 белка), доля сократительных белков 3 % (2 белка), доля других белков 13 % (10 белков).

Статистический анализ выборок осуществлялся с помощью математического пакета STATISTICA.

## Результаты

Предварительный анализ выборки показал, что в генах в среднем присутствовало 12 экзонов, в то время как среднее число экзонов в генах позвоночных равно 7. При этом 90 % генов содержат 12 и менее экзонов в последовательности (Lewin, 1994). Из них четыре экзона было задействовано в кодировании аденозин-фосфат-связывающего сайта. Среднее количество аминокислотных остатков, образующих в белке АФСС сайт, составило 15. Таким образом, сайты связывания АТФ, АДФ и АМФ включают значительное число аминокислот (по сравнению, например, с каталитическими триадами) и кодируются несколькими экзонами.

Для дальнейшего анализа интерес представляла задача сравнения распределений длин экзонов, кодирующих и не кодирующих АФСС. Соответствующие распределения приведены на рис. 1. Тест  $\chi^2$  показал их статистически значимые различия ( $P < 0,00001$ ), при этом значение  $\chi^2$  было равным 962,5 при 45 степенях свободы. Средние длины экзонов, кодирующих и не кодирующих АФСС, оказались 135 и 147 п.о. соответственно.

Также была выдвинута гипотеза о том, что аминокислоты АФСС кодируются различными экзонами последовательности с равной вероятностью, не зависят от длины экзона и от его положения в последовательности. Для проверки этой гипотезы позиции аминокислот АФСС в каждой последовательности анализируемой выборки менялись случайным образом

и фиксировались их новые позиции в рамках границ экзонов, ранее отмеченных на последовательности гена. При этом позиции границ экзонов не менялись. Перемешивание позиций аминокислот проводили 1000 раз для каждой последовательности из выборки, и на основе результатов таких теоретических экспериментов было получено ожидаемое по случайным причинам распределение соответствующего числа аминокислот АФСС, закодированных одним экзоном. Сравнение двух распределений, наблюдаемого и ожидаемого по случайным причинам (рис. 2), показало по критерию  $\chi^2$  статистически значимое различие между ними с уровнем значимости, близким к нулю (статистика  $\chi^2$  была равна 681 при 25 степенях свободы).

Для того чтобы оценить, насколько удалены друг от друга аминокислоты АФСС в кодирующей экзонной структуре гена, необходимо использование специального оценочного коэффициента. Для этих целей рассчитывался коэффициент разрывности  $\alpha_1$ , характеризующий количество экзонов, кодирующих АФСС в заданной последовательности и их положение друг относительно друга в последовательности гена:

$$\alpha_1 = 1 - \frac{N}{p_N^E - p_1^E + 1}, \quad (1)$$

где,  $p_1^E$  – порядковый номер первого экзона в последовательности гена, кодирующего АФСС,  $p_N^E$  – порядковый номер последнего экзона в последовательности гена, кодирующего АФСС,

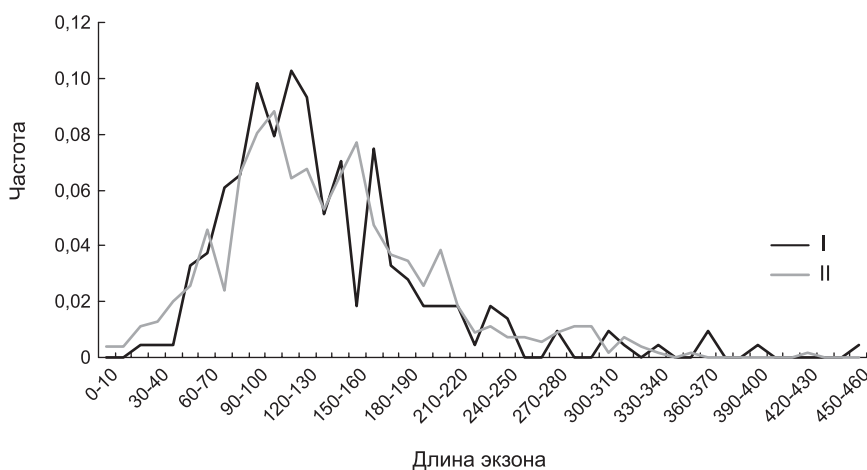


Рис. 1. Плотность распределения экзонов, кодирующих АФСС (I) и не кодирующих его (II).



**Рис. 2.** Наблюдаемое и ожидаемое по случайным причинам распределения числа аминокислот АФСС, располагающихся в границах одного экзона.

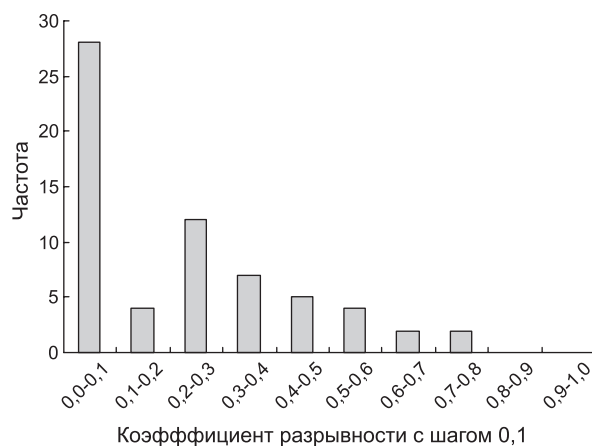
$N$  – число экзонов последовательности гена, кодирующих АФСС.

На рис. 3 показано распределение коэффициента разрывности  $\alpha_1$ . Чем больше значение коэффициента, тем дальше расположены друг относительно друга экзоны последовательности гена, кодирующие аминокислоты АФСС белка. Исходя из построенного распределения коэффициента разрывности можно заключить, что 28 % АФСС кодируются исключительно соседними экзонами в последовательности.

Коэффициент разрывности  $\alpha_2$  использовался для оценки удаленности друг от друга аминокислот АФСС в первичной структуре белка:

$$\alpha_2 = 1 - \frac{M}{p_M^A - p_1^A + 1},$$

где  $M$  – количество аминокислот АФСС в последовательности,  $p_M^A$  – позиция последней аминокислоты сайта,  $p_1^A$  – позиция первой.



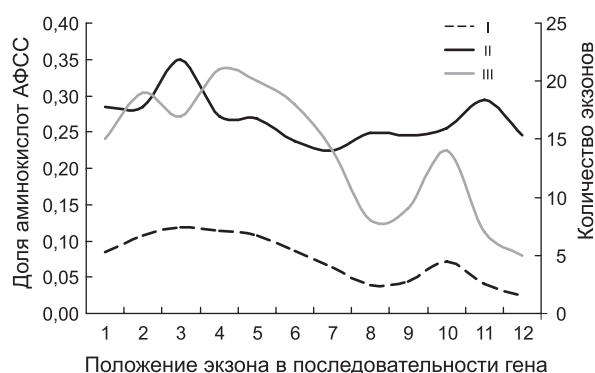
**Рис. 3.** Распределение коэффициента разрывности  $\alpha_1$ .

Распределение этого коэффициента, полученное для анализируемой выборки АФСС, показано на рис. 4. Из рисунка можно видеть, что в отличие от распределения коэффициента  $\alpha_1$  наблюдается рост числа АФСС при увеличении коэффициента разрывности по аминокислотной последовательности. Большинство АФСС характеризуются  $\alpha_2$ , близким к максимальному, в то время когда экзоны, их кодирующие, оказываются сближенными в последовательности гена. В этой связи было интересным узнать, как соотносятся позиции аминокислот АФСС в последовательности белка и границы функциональных доменов в нем. Для определения границ функциональных доменов на анализируемых последовательностях белков использовались данные из БД Pfam-A (Finn *et al.*, 2008). В результате оказалось, что лишь 60 % АФСС (38 АФСС) принадлежит одному известному функциональному домену, 5 % (3 АФСС) – двум, а 35 % (22 АФСС) принадлежат полностью или частично той части последовательности белка, в которой не обнаружен известный или гипотетический домен.

Существуют свидетельства того, что функциональные сайты белков с различной частотой присутствуют на N- и C-конце последовательности (Hodgman, 1986). В связи с этим было интересно узнать позиции экзонов в последовательности гена, которые кодируют аминокислоты АФСС. Для этого мы рассматривали только гены, содержащие более 5 экзонов. Графики, отражающие частоту позиции экзона, кодирующего аминокислоты АФСС, приведены на рис. 5. Из рисунка видно, что с наибольшей



**Рис. 4.** Распределение коэффициента  $\alpha_2$ .



**Рис. 5.** Частота, с которой аминокислоты АФСС закодированы экзонам в определенной позиции последовательности гена.

I – доля аминокислот АФСС, закодированных в одном экзоне в зависимости от его положения в последовательности гена; II – тот же показатель, что и в случае I, но рассчитанный только для экзонов, кодирующих АФСС; III – количество экзонов, кодирующих хотя бы одну аминокислоту АФСС в зависимости от их положения в гене.

частотой аминокислоты АФСС кодируются первыми 5 экзонами, а также экзонами, располагающимися на конце последовательности гена.

### Обсуждение

Аденозинфосфат-связывающие белки распространены во всех живых организмах, участвуют во многих процессах: клеточное дыхание, фотосинтез, биосинтетические реакции, деление клетки, передача сигналов и т. д.

Анализ выборки генов, кодирующих аденозинфосфат-связывающие белки, продемонстрировал несколько особенностей распределения АФСС на экзонной структуре этих генов. Одним из примечательных свойств можно назвать достоверное различие распределений длин экзонов, кодирующих и не кодирующих АФСС. Такое различие показывает существование эволюционного отбора на определенную длину экзонов, кодирующих аминокислоты АФСС. При этом на исследованном наборе генов средняя длина экзона, кодирующего аминокислоты АФСС, оказалась меньше средней длины экзонов, которые не участвуют в кодировании аминокислот этих сайтов. Является ли этот факт следствием вставки интронов в процессе эволюции, может показать дальнейшее исследование

ортологических групп генов. Свидетельством в пользу такого предположения может быть также тенденция кодирования аминокислот АФСС несколькими экзонами, близко расположенными в последовательности гена, несмотря на значительное удаление аминокислот АФСС в первичной структуре белка.

Еще одной особенностью распределения АФСС оказалась более высокая частота их кодирования на 5'-конце последовательности гена, т. е. в первых экзонах последовательности. Для понимания найденной закономерности также необходимо проведение дополнительных исследований.

Более глубокое изучение эволюционных особенностей распределения АФСС на экзонной структуре гена в дальнейшем нами планируется на основе филогенетического анализа последовательностей. Следует принять во внимание, что исследованный набор белков также содержит другие функциональные сайты, такие, как каталитические центры, сайты связывания субстратов, кофакторов и т. д. Наличие таких сайтов может влиять на особенности организации экзонной структуры гена. Поэтому нами будет проведено комплексное изучение распределения также и других типов функциональных сайтов для данного набора белков.

### Благодарности

Работа частично была поддержана проектами СО РАН № 115, 10.7, 18.13, государственным контрактом с ФАНИ № 02.514.11.4065, а также НШ-2447.2008.4. Научная школа Н.А. Колчанова «Биоинформатика и системная компьютерная биология».

### Литература

- Altschul S.F., Madden T.L., Schaffer A.A. *et al.* Gapped BLAST and PSIBLAST: a new generation of protein database search programs // *Nucl. Acids Res.* 1997. V. 25. P. 3389–3402.
- Andreeva N.S. The structure of globular proteins according to x-ray structural crystallography data // *Usp. Sovrem. Biol.* 1964. V. 58. P. 3–21.
- Berman H.M., Westbrook J., Feng Z. *et al.* The Protein Data Bank // *Nucl. Acids Res.* 2000. V. 28. P. 235–242.
- Cochrane G., Akhtar R., Aldebert P. *et al.* Priorities



- for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database // *Nucl. Acids Res.* 2008. V. 36. (Database issue). P. D5–12.
- Finn R.D., Tate J., Mistry J. *et al.* // *Nucl. Acids Res.* 2008. Database Issue 36. P. D281–D288.
- Gilbert W. Why genes in pieces // *Nature*. 1978. V. 271(5645). P. 501.
- Graur D., Li W.H. *Fundamentals of molecular evolution.* Sinauer Associates Inc., Sunderland. Second ed. 2000. P. 481.
- Hodgman T.C. The elucidation of protein function from its amino acid sequence // *Bioinformatics*. 1986. V. 2. № 3. P. 181–187.
- Ivanisenko V.A., Pintus S.S., Grigorovich D.A., Kolchanov N.A. PDBSite: a database of the 3D structure of protein functional sites // *Nucl. Acids Res.* 2005. V. 33. (Database issue). P. D183–187.
- Kaessmann H., Zollner S., Nekrutenko A., Li W.H. Signatures of domain shuffling in the human genome // *Genome Res.* 2002. V. 12(11). P. 1642–1650.
- Lewin B. *Genes* V. N.Y.: Oxford University Press. 1994. P. 1272.
- Maki R., Traunecker A., Sakano H. *et al.* Exon shuffling generates an immunoglobulin heavy chain gene // *Proc. Natl Acad. Sci. USA*. 1980. V. 77(4). P. 2138–2142.
- Oswald C., Holland I.B., Schmitt L. The motor domains of ABC-transporters. What can structures tell us? // *Naunyn Schmiedebergs Arch. Pharmacol.* 2006. V. 372(6). P. 385–399.
- Ponting C.P., Russel R.R. The natural history of protein domains // *Annu. Rev. Biophys. Biomol. Struct.* 2002. V. 31. P. 45–71.
- Richards F.M. Structure of proteins // *Annu. Rev. Biochem.* 1963. V. 32. P. 269–300.
- Thompson J.D., Gibson T.J., Higgins D.G. Multiple sequence alignment using ClustalW and ClustalX // *Curr. Protoc. Bioinformatics*. 2002. Chapter 2. Unit 2.3.
- Watanabe K., Harayama S. SWISS-PROT: The curated protein sequence database on Internet // *Protein, Nucl. Acid and Enzyme*. 2001. V. 46. P. 80–86.

## DISTRIBUTION ANALYSIS OF ADENOSINE PHOSPHATE BINDING SITES OF PROTEINS ON EXON STRUCTURE OF GENE

I.V. Medvedeva, P.S. Demenkov, V.A. Ivanisenko

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia,  
e-mail: brukaro@bionet.nsc.ru, salix@bionet.nsc.ru

### Summary

The problem of comparison of tertiary structure of protein and the structure of the encoding gene appeared after the development of the first methods of x-ray crystallography. A lot of papers were devoted to the investigation of the relationships between the exon borders and the domain borders that belong to the protein encoded by the respective gene. The aim of this work was the investigation of the relationships between protein functional sites and exon structure of the encoding genes for the adenosine phosphate binding sites because the conservative processes that adenosine phosphate ligands are involved. It is shown that the length of the exons encoding adenosine phosphate binding sites differs from the other in nonrandom manner. Besides this, the binding sites discovered to be discontinued on the protein primary structure although the most part of them is encoded by the adjacent exons. Also it has been shown that these functional sites' possess discontinuity not only on primary structure but tertiary structure of protein too. The functional sites aminoacids distribution through the encoding exons positions has shown that the most part of binding sites' of ATP, ADP, AMP are encoded by the first few exons.

## КОЭВОЛЮЦИЯ ДОМЕНОВ КЛЮЧЕВЫХ БЕЛКОВ АПОПТОЗА p53 И MDM2

С.С. Пинтус

Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: pintus@bionet.nsc.ru

Исследованы коэволюционные отношения между доменами белков p53 и mdm2, связанными между собой регуляторными отношениями. Найдены свидетельства коэволюции между отдельными участками взаимодействующих доменов белков. Также оценены скорости аминокислотных замен, предположительно характерные для коэволюционирующих, а также для нейтрально эволюционирующих районов этих белков.

**Ключевые слова:** p53, mdm2, белок, эволюционное расстояние, коэволюция, скорость замен, пространственная структура.

### Введение

Белок p53 относится к супрессорам опухолей. Он способен как останавливать деление клеток, так и вызывать их апоптоз по митохондриальному пути. Белок p53 участвует в передаче клеточного сигнала посредством как активации транскрипции генов-мишеней, так и белок-белковых взаимодействий. Для выполнения функции фактора транскрипции белку p53 необходим трансактивационный домен (TAD), но связывание p53 с регуляторной областью гена-мишени осуществляется посредством домена связывания с ДНК (DBD) (Levine, 1997).

Транскрипционная активность белка p53 в клетке подавляется посредством связывания домена MDM2 убиквитин-лигазы mdm2 с трансактивационным доменом p53. Также это приводит к уменьшению концентрации p53 в клетке за счет убиквитинилирования. В свою очередь, ген *mdm2* является мишенью белка p53, который активирует транскрипцию собственного ингибитора. Таким образом, осуществляется механизм обратной связи, который позволяет поддерживать крайне низкую концентрацию p53 в нормальной клетке (Prives, 1998).

Известно, что эволюционная история белка p53 была связана с его специализацией на функции «стража генома» (Yang *et al.*, 2002), что

подразумевает приобретение им новых генов мишеней и соответственно адаптивные изменения в аминокислотной последовательности трансактивационного домена (TAD). Учитывая важность белок-белкового взаимодействия между TAD-доменом белка p53 и MDM2-доменом белка mdm2, можно предположить, что в истории этих белков имела место коэволюция между трансактивационным доменом p53 и MDM2-доменом mdm2. Таким образом, целью настоящей работы было исследовать коэволюционные отношения между доменами белков p53 и mdm2.

Исследование коэволюции на молекулярном уровне интенсивно ведется в последние годы. Обзор данной области доступен в работе Galtier, Dutheil (2007).

### Материалы и методы

Современные методы исследования коэволюции белков основаны на анализе геномных последовательностей и представлены двумя группами. Первая группа подходов основывается на методе анализа корреляции между заменами. Данный метод был впервые предложен в работе Goh *et al.* (2000) и развит в работе Pazos, Valencia (2001). Другая группа методов основана на сходстве топологий филогенетических деревьев (Jothi *et al.*, 2005).

В настоящей работе был использован подход корреляции матриц аминокислотных расстояний для оценки коэволюции белков, аналогичный подходу, предложенному группой Пазоса в их методе поиска взаимодействий между белками на основании их коэволюции (Pazos, Valencia, 2001). Однако в этом подходе не были учтены различия между аминокислотными сайтами с различными скоростями замен.

Подход, основанный на сходстве топологий филогенетических деревьев, представляется менее предпочтительным, поскольку он требует надежной реконструкции филогении. Надежность реконструкции зависит, в частности, от числа сайтов выравнивания, использованного для реконструкции. В настоящей работе были использованы выборки из сравнительно небольшого числа сайтов аминокислотного выравнивания. При исследовании коэволюции между этими выборками снижение надежности филогении оказало бы свое влияние, и на этом основании был выбран подход корреляции матриц аминокислотных расстояний.

Ортологичные аминокислотные последовательности белков семейств p53 и mdm2 были взяты из базы данных Ensembl (Curwen *et al.*, 2004) ([www.ensembl.org](http://www.ensembl.org)) для различных организмов, за исключением курицы (*G. gallus*), для которой последовательности были взяты из базы данных Swiss-Prot (Bairoch *et al.*, 2005) ([www.uniprot.org/uniprot](http://www.uniprot.org/uniprot)). Последовательности были отобраны таким образом, чтобы набор представленных видов организмов был одним и тем же для белков p53 и mdm2 (табл. 1).

Из базы данных PFAM ([pfam.sanger.ac.uk](http://pfam.sanger.ac.uk)) были взяты границы доменов TAD, DBD и MDM2 для видов, у которых последовательности белков p53 и mdm2 представлены в базе данных Swiss-Prot. С помощью программы ClustalW 2 были получены выравнивания аминокислотных последовательностей белков p53 и mdm2. При выравнивании были использованы параметры по умолчанию: весовая матрица Gonnet (Gonnet, 2000), штраф на открытие разрыва – 10, штраф на продолжение разрыва – 0,20. Для mdm2 показатель идентичности последо-

Таблица 1

Список видов организмов и идентификаторы белковых последовательностей белков p53 и mdm2 в базе данных Ensembl

Вид организма	Идентификатор аминокислотной последовательности p53	Идентификатор аминокислотной последовательности mdm2
<i>Canis familiaris</i>	ENSCAFP00000024579	ENSCAFP00000000608
<i>Danio rerio</i>	ENSDARP00000051548	ENSDARP00000091133
<i>Equus caballus</i>	ENSECAP00000006210	ENSECAP00000011666
<i>Gallus gallus</i>	P53_CHICK	MDM2_CHICK
<i>Gasterosteus aculeatus</i>	ENSGACP00000025614	ENSGACP00000001415
<i>Homo sapiens</i>	ENSP00000269305	ENSP00000258148
<i>Macaca mulatta</i>	ENSMMUP00000011334	ENSMMUP00000018642
<i>Mus musculus</i>	ENSMUSP00000005371	ENSMUSP00000020408
<i>Ochotona princeps</i>	ENSOPRP00000005906	ENSOPRP00000008864
<i>Oryzias latipes</i>	ENSORLP00000008029	ENSORLP00000021473
<i>Pan troglodytes</i>	ENSPTRP00000014836	ENSPTRP00000050082
<i>Pongo pygmaeus</i>	ENSPPYP00000008923	ENSPPYP00000005415
<i>Rattus norvegicus</i>	ENSRNOP00000047840	ENSRNOP00000008388
<i>Takifugu rubripes</i>	ENSTRUP00000030646	ENSTRUP00000020873
<i>Xenopus tropicalis</i>	ENSXETP00000053759	ENSXETP00000003026

Примечание. Для курицы (*G. gallus*) приведены идентификаторы из базы Swiss-Prot.

вательностей (identity) оказался равным 69/558 (12 %), показатель сходства (similarity) – 100/558 (18 %). Для p53 показатели идентичности и сходства составили 81/424 (19 %) и 116/424 (27 %) (согласно ClustalW 2).

Из полученных выравниваний с учетом известных границ доменов были извлечены столбцы, соответствующие доменам TAD и DBD для p53 и домену MDM2 для mdm2. В случае домена MDM2 области выравниваний границ доменов были отредактированы вручную с учетом аминокислотных последовательностей границ доменов, аннотированных в базах Swiss-Prot и Pfam. В частности, известно, что домен MDM2 довольно консервативен (Bennett-Lovsey *et al.*, 2002), тогда как выравнивание, полученное с помощью ClustalW 2, содержало длинную

последовательность разрывов, соответствующую обширной делеции у человека и макаки, в результате чего консервативная С-концевая последовательность домена у этих организмов смещается (рис. 1, а). Разрывы были удалены вручную таким образом, чтобы С-концевые последовательности совпадали (рис. 1, б).

Из полученных выравниваний были извлечены и выравнены заново фрагменты, соответствующие доменам TAD, DBD и MDM2 (ClustalW 2, параметры по умолчанию). Число сайтов в каждом из выравниваний составило 53, 193, и 81 соответственно (рис. 2).

Три полученных выравнивания были поданы на вход программы PhyML, использовалась дискретная модель гамма-распределения замен, число категорий – 4 (согласно рекомендации ав-

а	
Chicken	VKEHRR IYS <u>MI SRNL</u> IAINQQDSTLAVPPEDDAKFRLEKENVLKESMQELEEKQTSSN--ATSQPTTSRR
Chimpanzee	----KIYT <u>MI YRNL</u> VVVNQEQESSDGSVSENRCHEGGSDQKDLVQELQEEKPSSSHLVS RPSTSSRR
Dog	VKEHRKIYT <u>MI YRNL</u> VVVNQHEPSDGSVSENSCHREGGSDQKDPVQELQEEKPSSDLISRPSTSSRR
Fugu	VKEPRVLFAM <u>ITMKNL</u> VAVKNPESSSQSDSNTDDQR--EKQEEPATPEEEELQEQRSS----LSPTHWRSS
Horse	VKEHRKIYT <u>MI YRNL</u> VVVNQPEPSDGSVSENSCPLGGSDQKAAVQELQEEKPSSSALVSRPPTSSRR
Human	-----VLFYLGQYIMTKRLYDEKQQHIVYCSNDLLG--DLFGVPSFVSVEHRKIYT <u>MI YR</u> ---- <u>NL</u>
Macaque	-----VLFYLGQYIMTKRLYDEKQQHIVYCSNDLLG--DLFGVPSFVSVEHRKIYT <u>MI YR</u> ---- <u>NL</u>
Medaka	VKEPRVLFAM <u>ITKKNL</u> VAVNKAAVNSNKEVKKNKYELEEMLKTL-----L IKKKSS--VKLQMEHQLVA
Mouse	VKEHRKIYA <u>MI YRNL</u> VAVSQQ--DSGTSLSERRQPEGGSDLKDP LQAPPEEKPSDDLISRLSTSSRR
Orangutan	VKEHRKIYT <u>MI YRNL</u> VVVNQEQESSDGSVSENRCHEGGSDQKDLVQELQEEKPSSSHLVS RPSTSSRR
Pika	VKEHRKIYT <u>MI YRNL</u> VVVNQEQEPSDGSMSI SEDRYHPEGGNDKEKPEGQELQEEKPSSSLVSRPSTSSRR
Rat	VKEHRKIYA <u>MI YRNL</u> VVVSQQ--DSGTSPESESRQPEGGSDLKDPVQASQEEKPSSSDVSRPSTSSRR
Stickleback	VKEPRVLFAM <u>ITKKNL</u> VAVKSQESPASSSEAGASRTEEAESSESHSPERRGRKRRRRSRFRSSSEPSGSS
X.tropicalis	VKEPRRLYA <u>MI SRNL</u> VSAAVKESSEVYV--NHVCSFPDKQKSQKELLQELP-EKVIAPAYDSKPCNSSQR
Zebrafish	VKEPRALFAL <u>INRNL</u> VTVKNPESQSTFSEPRSQSEPD RGP GD TDS-----DSRSS-----TSQQQRRR
б	
Chicken	MKEVIFYLGQYIMSKQLYDEKEQHIVHCANDLLGDLFGVTSFVSVEHRR IYS <u>MI SRNL</u>
Chimpanzee	MKE-----KIYT <u>MI YRNL</u>
Dog	MKEVIFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Fugu	MKEVMFYLGQYI IQKQLYDQKQQHIVHCSHDELGRV LGVDSFVSVEPRVLFAM <u>ITMKNL</u>
Horse	MKEVIFYLGQYIMTKRLYDERQQHVMVYCSNDLLGDLFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Human	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Macaque	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Medaka	MKEVMFYLGQYIMQKQLYDQKQQHVMVHCAQDALGRV LGVDSFVSVEPRVLFAM <u>ITKKNL</u>
Mouse	MKEIIFYIGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYA <u>MI YRNL</u>
Orangutan	MKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Pika	MKEVIFYLGQYIMTKRLYDEK--HIVYCSND--LGD LFGVPSFVSVEHRKIYT <u>MI YRNL</u>
Rat	MKEIIFYIGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFVSVEHRKIYA <u>MI YRNL</u>
Stickleback	MKEVMFYLGQYIILKQLYDQKQQHIVHCSQDALGRV LGVDSFVSVEPRVLFAM <u>ITKKNL</u>
X.tropicalis	MKEVIFYLGQYIMAKQLYDEKQQHIVHCSNDPLGELFGVQEFVSVEPRRLYA <u>MI SRNL</u>
Zebrafish	MKEVMFYLGQYIMSKELYDKQQHIVHCGEDPLGAVLVGKVSFVSVEPRALFAL <u>INRNL</u>

Рис. 1. Фрагмент выравнивания последовательностей белка mdm2 до и после ручной правки.

а – выравнивание до ручной правки. Остатки С-концевой последовательности домена MDM2 выделены. Видно смещение положения С-концевой последовательности у человека (Human) и макаки (Macaque). б – фрагмент выравнивания последовательностей mdm2 после ручной правки.

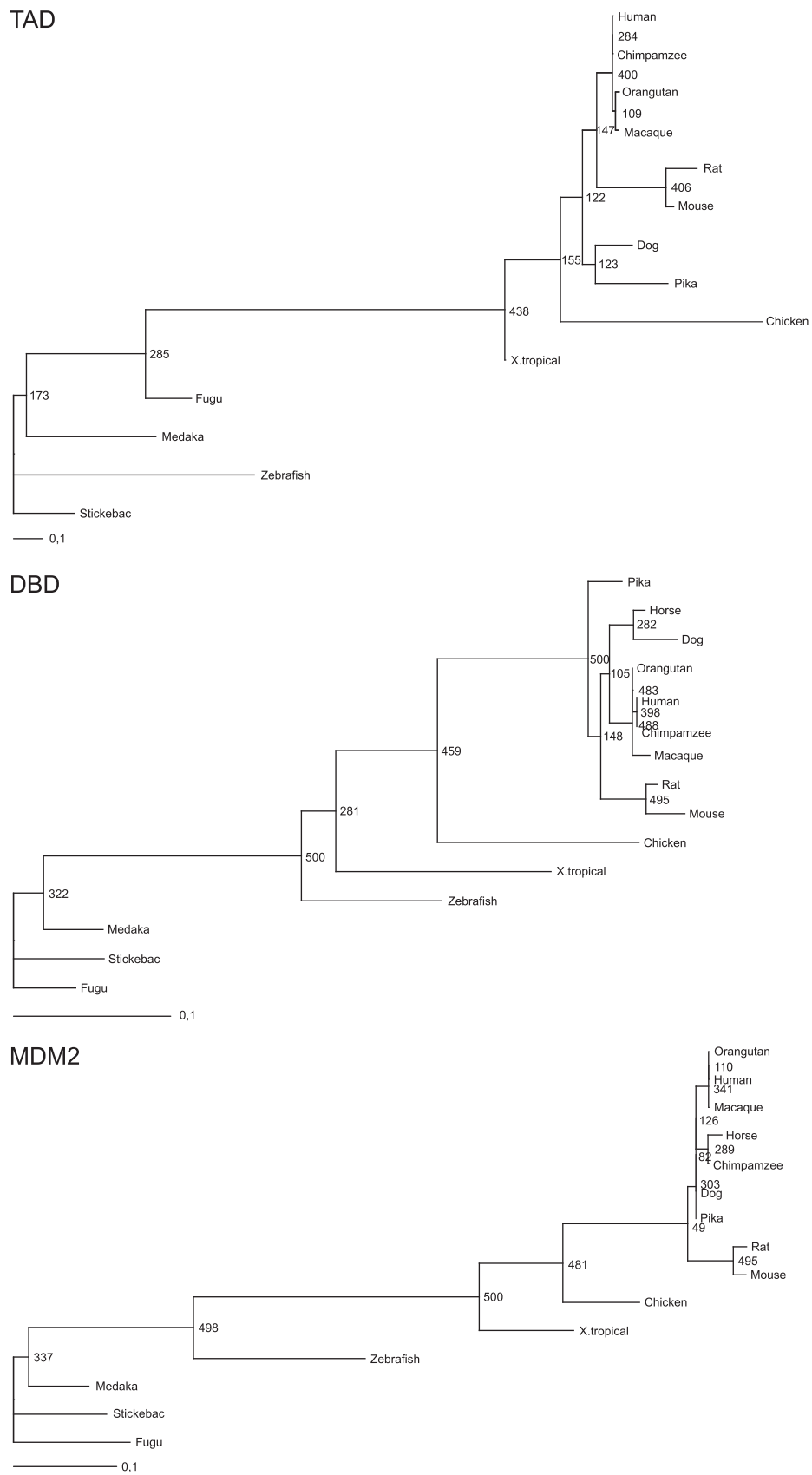
<b>TAD</b>	
14	53
Chicken	-----MAEEMEP-LLEP---TEVFMDLWSMLP-----
Chimpanzee	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Dog	-----MQEPQSELNIDPPLSQETFSELNLLPENNVLSSE---
Fugu	MED---EGFS-----LPLSQDTFQDLWENVML-----
Human	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Macaque	-----MEEPQSDPSIEPPLSQETFSDLWKLLPENNVLSPL---
Medaka	-----WESMDPVP--DLPESQGSFQELWETVY-----
Mouse	-----MEESQSDISLELPLSQETFSGLWKLLPPEDILP-----
Orangutan	-----MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPL---
Pika	-----MEEQPSELSIEPPLSQETFPELWNLPL---LLSTN---
Rat	-----MEDSQSDMSIELPLSQETFSCWLKLLPPDDILPTTATG
Sticklebac	MEGTGFEGFLALGQ--DLPDSQDSFAELWGNVVS-----
X.tropical	-----ME-PSSETGMEPPLSQETFEDLWSLLP-----
Zebrafish	-----MAQ-----NDSQE-FAELWEKNLI-----
<b>DBD</b>	
15	193
Chicken	DYGGDFDFRV GFVEAGTAKS VTCTYSPVLN KVYCRLAKPC PVQVRVGVAP
Chimpanzee	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Dog	TYPGTYGFRL GFLHSGTAKS VTWTYSPLLN KLFQQLAKTC PVQLWVSSPP
Fugu	DYPGEYGFEL RFQKSGTAKS VTSTYSEILN KLYCQLAKTS LVEVLLIKKP
Horse	-----YSPTLN KLFQQLAKTC PVQLLVSSPP
Human	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Macaque	TYHGSYGFRL GFLHSGTAKS VTCTYSPDLN KMFCQLAKTC PVQLWVDSTP
Medaka	DYPGSYELEL RFQKSGTAKS VTSTYSETLN KLYCQLAKTS PIEVRVSKEP
Mouse	TYQGNYGFRH GFLQSGTAKS VMCTYSPPLN KLFQQLAKTC PVQLWVSATP
Orangutan	TYQGSYGFRL GFLHSGTAKS VTCTYSPALN KMFCQLAKTC PVQLWVDSTP
Pika	TYPGTYGFRL GFLQSGTAKS VTCTYSPSLN KLFQQLAKTC PVQLKVDSTP
Rat	TYQGNYGFRH GFLQSGTAKS VMCTYISLN KLFQQLAKTC PVQLWVTSTP
Sticklebac	DHPGEFGFQL RFQKSGTAKS VTSTYSEILN KLYCQLAKTS LIEVLVTKEL
X.tropical	DYAGSYGLKL EFQQNGTAKS VTCTYSTDLN KLFQQLAKTC PLLVVRVERP
Zebrafish	DYPGDHGFRL RFPQSGTAKS VTCTYSPDLN KLFQQLAKTC PVQMVDVAP
<b>MDM2</b>	
15	81
Chimpanzee	LVRPKPLLLK LLKSVGAQKD TYTMKE----
Horse	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDERQQHMVY
Human	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Macaque	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Orangutan	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Dog	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEKQQHIVY
Pika	LVRPKPLLLK LLKSVGAQKD TYTMKEVIFY LGQYIMTKRL YDEK--HIVY
Mouse	LVRPKPLLLK LLKSVGAQND TYTMKEIIFY IGQYIMTKRL YDEKQQHIVY
Rat	LVRPKPLLLK LLKSVGAQKD IYTMKEIIFY IGQYIMTKRL YDEKQQHIVY
Chicken	LVRPKPLLLK LLKLAGAEKD TFTMKEVIFY LGQYIMSKQL YDEKEQHIVH
X_tropical	LVRPKPLLLS LLKSAGAQKE TFTMKEVIYH LGQYIMAKQL YDEKQQHIVH
Medaka	LVRPKVEFCS LLQHAGATKD VFTMKEVMFY LGRYIMQKQL YDQKQQHMVH
Sticklebac	LVQPKLEFRT LLHQAGATKE VFTMKEVMFY LGQYIILKQL YDQKQQHIVH
Fugu	QVRPNVEFHR LLQDAGAIKD VFTMKEVMFY LGQYI IQKQL YDQKQQHIVH
Zebrafish	-VRPKVQLKS LLEDAGADKD VFTMKEVMFY LGKYIMSKEL YDKQQQHIVH

Рис. 2. Фрагменты выравниваний доменов TAD, DBD и MDM2.

торов метода), параметр гамма-распределения ( $\alpha$ ) и число инвариантных сайтов подбирались программой автоматически, использовалась матрица аминокислотных замен WAG (Whelan, Goldman, 2001), значение бутстреп – 500. В результате были получены три филогенетических дерева (рис. 3).

Построенные деревья, а также полученные выравнивания доменов были использованы для измерения скоростей замен в аминокислотных сайтах выравниваний с помощью программы CODEML пакета PAML 3.14. Аминокислотные сайты каждого домена были разбиты программой на 5 категорий в зависимости от скорости





**Рис. 3.** Филогенетические деревья доменов TAD, DBD и MDM2, построенные программой PhyML.

Числа обозначают значения бутстрепа.

замен. Средние скорости замен в категориях составили 0,02, 0,16, 0,47, 1,11 и 3,25 млн лет<sup>-1</sup>. При этом использовалась матрица аминокислотных замен WAG (Whelan, Goldman, 2001), параметр гамма-распределения ( $\alpha$ ) определялся программой, итерации подбора параметра начинались с 0,5.

Выбор количества категорий определялся представленностью различных значений скоростей замен, с одной стороны, и количеством сайтов в категории, достаточным для статистического исследования (построения матрицы корреляции), с другой стороны. Так, при выборе 4 категорий разбиение сайтов программой CODEML таково, что не представлена отдельная группа сайтов со средней скоростью от 1–2 млн лет<sup>-1</sup>, имеющая довольно большое значение для интерпретации результатов настоящей работы, как поясняется ниже. Однако при выборе 6 категорий в категориях со средними скоростями 0,10 и 0,31 млн лет<sup>-1</sup> не представлено ни одного аминокислотного сайта домена MDM2.

Аминокислотные сайты выравнивания каждого из трех доменов были разбиты на 5 групп в зависимости от категории средней скорости замен, к которой они были отнесены программой CODEML. Каждая такая группа составила отдельную выборку (всего 15 выборок). Для каждой выборки была построена матрица эволюционных расстояний с помощью программы LAPD, разработанной Л. Арвстадом (Arvestad, 2006). Между матрицами расстояний, соответствующими одинаковым группам скоростей замен, вычислялся коэффициент попарной корреляции Пирсона, который использовался в качестве меры коэволюции соответствующих участков белков.

### Результаты

Частоты встречаемости сайтов (в процентах), относящихся к различным категориям скорости замен, приведены в табл. 2. Можно видеть, что трансактивационный домен p53 отличался от двух других исследованных доменов большей вариабельностью, а также тем, что лишь у него встречались слабовариабельные аминокислотные сайты (0,16 млн лет<sup>-1</sup>). Интересно, что доля наиболее вариабельных сайтов была примерно одинакова у всех трех доменов, тогда как доли

Таблица 2

Частоты встречаемости аминокислотных сайтов (в процентах от общего числа сайтов домена), относящихся к различным категориям скорости замен

Категория сайтов	Скорость замен, млн лет <sup>-1</sup>	mdm2 MDM2	p53 DBD	p53 TAD
1	0,02	0,43	0,47	0,23
2	0,16	0,00	0,00	0,15
3	0,47	0,09	0,11	0,25
4	1,11	0,27	0,22	0,17
5	3,25	0,21	0,20	0,21

полностью консервативных сайтов (0,02 млн лет<sup>-1</sup>) существенно различались. Также стоит отметить, что распределения сайтов по скоростям замен схожи у доменов mdm2 MDM2 и p53 DBD, тогда как распределение для домена p53 TAD существенно отличается от первых двух большей равномерностью (рис. 4).

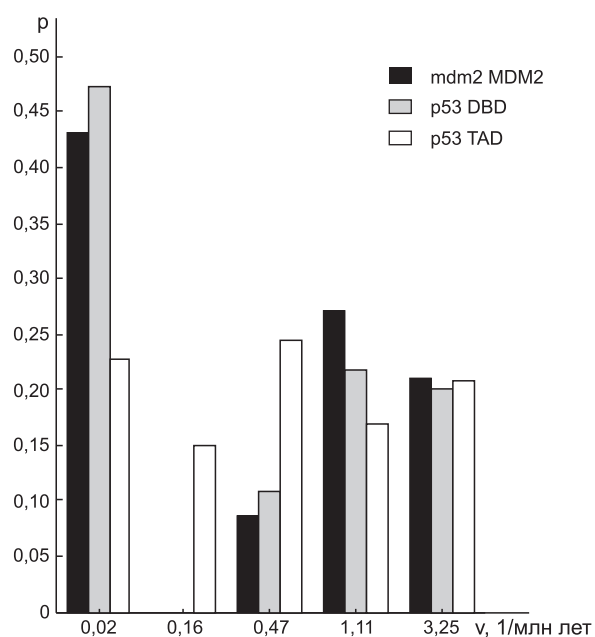


Рис. 4. Распределение частот встречаемости аминокислотных сайтов (в процентах от общего числа сайтов домена), относящихся к различным категориям скорости замен.

По оси абсцисс отложены средние скорости аминокислотных замен для каждой категории (млн лет<sup>-1</sup>), по оси ординат – частота встречаемости аминокислотных сайтов данной категории.

Значения коэффициента корреляции Пирсона между матрицами эволюционных расстояний в парах доменов белков представлены в табл. 3.

Данные для категории 1 отсутствуют, так как в нее попадали полностью консервативные сайты, и все три матрицы эволюционных расстояний для этой категории состояли из нулей. Ввиду отсутствия сайтов, относящихся к категории 2 у доменов mdm2 MDM2 и p53 DBD, не было возможности оценить коэффициент корреляции и для них.

Как видно из табл. 3, распределение коэффициентов корреляции между матрицами эволюционных расстояний для пары невзаимодействующих доменов mdm2 MDM2 – p53 DBD существенно отличается от таковых для пар взаимодействующих доменов mdm2 MDM2 – p53 TAD и p53 DBD – p53 TAD. Корреляция между высоковариабельными сайтами из категории 5 довольно высока для пары mdm2 MDM2 – p53 DBD и низка для остальных двух пар. Корреляция же между относительно консервативными сайтами из категории 3 наоборот особенно низка для пары mdm2 MDM2 – p53 DBD. Следует также отметить, что для всех трех пар доменов высока корреляция между сайтами из категории 4.

### Обсуждение

Отличие распределения сайтов по категориям скоростей замен для транскрипционного домена p53 от распределения для двух других исследованных доменов может отражать события движущего отбора в ходе эволюционной истории белка p53, связанные с приобретением новых генов-мишеней в процессе специализации p53 на функции «стража генома».

Ранее было неоднократно показано, что уровень филогенетического сигнала тем ниже,

чем больше сила естественного отбора (Hansen, Martins, 1996; Diniz-Filho, 2001; Martins *et al.*, 2002; Jose *et al.*, 2008). Для сайтов, несущих самый низкий уровень филогенетического сигнала, можно предположить самые низкие значения корреляции между невзаимодействующими доменами MDM2 и DBD, самые высокие – между постоянно взаимодействующими доменами одного белка p53 и помезуточные – для эпизодически взаимодействующих доменов MDM2 и TAD. Такое распределение значений корреляции наблюдается только для сайтов из категории 3 (0,47 млн лет<sup>-1</sup>, табл. 3). Отсюда можно сделать вывод, что сайты со скоростью замен порядка 0,47 млн лет<sup>-1</sup> несут наименьший уровень филогенетического сигнала в белках p53 и mdm2. Можно предположить, что фиксация замен именно в этих сайтах регулировалась отбором, связанным с коэволюцией этих белков.

В пространственной структуре домена DBD человека сайтам категории 3 соответствуют 20 остатков, преимущественно расположенных на границах между неструктурным и структурированным элементами (8 на границах с  $\beta$ -листами, 3 – на границе с петлями и еще 1 на границе с  $\alpha$ -спиралью). Внутри  $\beta$ -листов располагаются 3 наиболее консервативных остатка из 3-й категории. Еще 2 остатка располагаются внутри  $\alpha$ -спиралей. Они менее консервативны. Остальные остатки, относящиеся к категории 3, располагаются в неструктурных элементах.

Преобладание в 3-й категории остатков, располагающихся на границах структурных элементов, а значит, важных для пространственной структуры домена, дает дополнительный аргумент в пользу отбора по остаткам из категории 3.

В пространственной структуре  $\alpha$ -спирали домена TAD, взаимодействующей с MDM2 (Kussie *et al.*, 1996), остатки категории 3 (S20, D21, K24, L25, L26) прилегают к консерватив-

Таблица 3

Значения коэффициента корреляции Пирсона между матрицами эволюционных расстояний в парах доменов белков mdm2 и p53

Категория сайтов	Скорость замен, млн лет <sup>-1</sup>	MDM2 – TAD	DBD – TAD	MDM2 – DBD
3	0,47	0,390 ± 0,004	0,719 ± 0,002	0,207 ± 0,004
4	1,11	0,526 ± 0,004	0,710 ± 0,002	0,730 ± 0,002
5	3,25	0,187 ± 0,005	0,220 ± 0,005	0,825 ± 0,001

ным остаткам F19, T23, P27, взаимодействующим с MDM2. Остаток P27 также взаимодействует с MDM2, но относится к 3-й категории.

В пространственной структуре домена MDM2 остатки, соответствующие сайтам категории 3 (V52, Q71, I74, S90, M102), располагаются вблизи остатков L54, Q72, H73, V75, F91, T100, взаимодействующих с TAD. Более того, остатки I74, S90, а также K31 располагаются на границах  $\beta$ -листов аналогично домену DBD.

Такое распределение сайтов категории 3 на пространственной структуре взаимодействующих доменов TAD и MDM2 позволяет сделать вывод, что остатки, непосредственно участвующие во взаимодействии белков p53 и mdm2, консервативны, тогда как по аминокислотным сайтам белков, оказывающим непосредственное влияние на положение взаимодействующих остатков и относящимся к 3-й категории по значению скорости замен, возможен заметный естественный отбор.

Одинаково высокие значения корреляции между сайтами из категории 4 свидетельствуют о том, что сайты из данной категории несут наибольший уровень филогенетического сигнала. Это позволяет предположить, что в данную категорию попали сайты, эволюционирующие нейтрально. Это подтверждается довольно высокой средней скоростью замен в сайтах данной категории – 1,11 млн лет<sup>-1</sup>. С другой стороны, для сайтов из категории 4 значение корреляции между доменами MDM2 и TAD не так высоко, как в парах доменов MDM2-DBD и DBD-TAD.

Это можно объяснить относительно невысоким уровнем филогенетического сигнала, который несут аминокислотные сайты домена TAD. В пользу этого предположения говорит то, что именно в домене TAD категория 3 наиболее представлена (рис. 4, табл. 2). Другим аргументом является разброс длин ветвей филогенетического древа домена TAD, гораздо больший, чем для доменов MDM2 и DBD. Разброс длин ветвей свидетельствует о том, что различия скоростей эволюции в разных таксонах у домена TAD были наибольшими из всех трех доменов. Следовательно, для домена TAD характерен наименьший филогенетический сигнал.

Особый интерес представляют низкие значения корреляции сайтов категории 5 для пар взаимодействующих доменов mdm2 MDM2 –

p53 TAD и p53 DBD – p53 TAD и высокое значение для невзаимодействующих доменов MDM2. Из 39 аминокислотных сайтов DBD человека, относящихся к 5-й категории, 35 располагаются в неструктурных элементах, соединяющих структурные элементы. Два других – L206 и V217 – располагаются в центре  $\beta$ -листов «бочонка», являющегося структурным ядром домена. Если сайт V217 инвариантен у всех таксонов, за исключением трех видов рыб, у которых в этом сайте наблюдаются остатки треонина, цистеина и фенилаланина, то сайт L206 весьма вариабелен. Учитывая важность структуры «бочонка», образованного  $\beta$ -листами, для функции всего домена, можно предположить, что наблюдаемые высокие скорости замен в некоторых сайтах 5-й категории связаны не с нейтральной эволюцией, а с отбором. Скорость замен в части сайтов данной категории, следовательно, можно считать переоцененной, что может быть связано с недостатком последовательностей p53 промежуточных таксонов и вытекающей из этого переоценкой вариабельности остатков в сайтах этой категории. Этот вывод подтверждается тем, что 15 из 17 сайтов 5-й категории на структуре домена MDM2 человека располагаются в структурных элементах, однако преимущественно удалены от интерфейса взаимодействия доменов.

Из распределения сайтов 5-й категории на пространственной структуре доменов MDM2 и P53 можно предположить, что высокое значение корреляции для пары невзаимодействующих доменов MDM2 и DBD обусловлено большой долей нейтрально эволюционирующих сайтов в белке p53, попавших в 5-ю категорию, но несущих филогенетический сигнал. Благодаря невысокой доле неструктурных элементов в доменах MDM2 и TAD, сайты 5-й категории из этих доменов оказались свободны от филогенетического сигнала, благодаря чему не наблюдается высокой корреляции для этой пары доменов. Низкие значения корреляции для этих сайтов в парах доменов MDM2-TAD и DBD-TAD можно объяснить отбором, связанным не с взаимодействием доменов, а с поддержанием структурной целостности каждого домена. Это косвенно подтверждается удаленностью сайтов 5-й категории в MDM2 от интерфейса взаимодействия.

Работа поддержана междисциплинарным интеграционным проектом фундаментальных исследований СО РАН № 26 «Математические модели, численные методы и параллельные алгоритмы для решения больших задач СО РАН и их реализация на многопроцессорных суперЭВМ», а также междисциплинарным интеграционным проектом фундаментальных исследований СО РАН № 113 «Разработка вычислительных методов, алгоритмов и аппаратурно-программного инструментария параллельного моделирования природных процессов».

### Литература

- Arvestad L. Efficient methods for estimating amino acid replacement rates // *J. Mol. Evol.* 2006. V. 62. P. 663–673.
- Bairoch A., Apweiler R., Wu C.H. *et al.* The Universal Protein Resource (UniProt) // *Nucl. Acids Res.* 2005. V. 33. P. D154–D159.
- Bennett-Lovsey R., Hart S.E., Shirai H., Mizuguchi K. The SWIB and the MDM2 domains are homologous and share a common fold // *Bioinformatics.* 2002. V. 18. P. 626–630.
- Curwen V., Eyraas E., Andrews T.D. *et al.* The Ensembl automatic gene annotation system // *Genome Res.* 2004. V. 14. P. 942–950.
- Diniz-Filho J.A. Phylogenetic autocorrelation under distinct evolutionary processes // *Evolution.* 2001. V. 55. P. 1104–1109.
- Galtier N., Duthel J. Coevolution within and between genes // *Genome Dyn.* 2007. V. 3. P. 1–12.
- Goh C.S., Bogan A.A., Joachimiak M. *et al.* Co-evolution of proteins with their interaction partners // *J. Mol. Biol.* 2000. V. 299. P. 283–293.
- Gonnet G.H., Korostensky C., Benner S. Evaluation measures of multiple sequence alignments // *J. Comput. Biol.* 2000. V. 7. P. 261–276.
- Hansen T.F., Martins E. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data // *Evolution.* 1996. V. 50. P. 1404–1417.
- José J., Puma-Villanueva W.J., Von Zuben F.J., Diniz-Filho J.A. Phylogenetic inheritance of genetic variability produced by neutral models of evolution // *Genet. Mol. Res.* 2008. V. 7. P. 1327–1343.
- Jothi R., Kann M.G., Przytycka T.M. Predicting protein-protein interaction by searching evolutionary tree automorphism space // *Bioinformatics.* 2005. V. 21. P. i241–i250.
- Kussie P.H., Gorina S., Marechal V. *et al.* Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain // *Science.* 1996. V. 274. P. 948–953.
- Levine A.J. p53, the cellular gatekeeper for growth and division // *Cell.* 1997. V. 88. P. 323–331.
- Martins E.P., Diniz-Filho J.A., Housworth E.A. Adaptive constraints and the phylogenetic comparative method: a computer simulation test // *Evolution.* 2002. V. 56. P. 1–13.
- Pazos F., Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction // *Protein Eng.* 2001. V. 14. P. 609–614.
- Prives C. Signaling to p53: breaking the MDM2-p53 circuit // *Cell.* 1998. V. 95. P. 5–8.
- Whelan S., Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach // *Mol. Biol. Evol.* 2001. V. 18. P. 691–699.
- Yang A., Kaghad M., Caput D., McKeon F. On the shoulders of giants: p63, p73 and the rise of p53 // *Trends Genet.* 2002. V. 18. P. 90–95.

## COEVOLUTION OF PROTEIN DOMAINS OF P53 AND MDM2 KEY PROTEINS OF APOPTOSIS

S.S. Pintus

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: pintus@bionet.nsc.ru

### Summary

Coevolution of domains of p53 and mdm2 proteins, that are involved in mutual regulation, is studied. Evidences for coevolution of particular regions of interacting domains of these two proteins are found. Amino acid substitution rates are estimated for regions of the proteins that are supposed to undergo coevolution as well as for regions of the proteins that are supposed to undergo neutral selection.



## КОМПЬЮТЕРНОЕ ПРЕДСКАЗАНИЕ БИОЛОГИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ ВЕЩЕСТВ: ВИРТУАЛЬНАЯ ХЕМОГЕНОМИКА

В.В. Поройков, Д.А. Филимонов, Т.А. Глоризова, А.А. Лагунин,  
Д.С. Дружиловский, А.В. Степанчикова

ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН,  
e-mail: vladimir.poroikov@ibmc.msk.ru

На основе прогнозируемых с помощью компьютерной программы PASS спектров биологической активности можно выявлять наиболее вероятные молекулярные мишени, с которыми химические вещества взаимодействуют в биологических объектах, и обусловленные этими взаимодействиями биологические эффекты. Средняя точность прогноза составляет около 94 %, а разработанный метод устойчив по отношению к неполноте информации в обучающей выборке. Широкомасштабное использование данного подхода в качестве «инструмента» для виртуальной хемогеномики (<http://www.ibmc.msk.ru/PASS>) позволило в ряде случаев подтвердить компьютерные предсказания независимыми экспериментами.

**Ключевые слова:** спектр биологической активности, прогнозирование, PASS, хемогеномика, мишени, лиганды.

### Введение

Хемогеномика определяется различными авторами (Caron *et al.*, 2001; Kubinyi, 2006; Rognan, 2007) как новая мультидисциплинарная область науки, конечной целью которой является установление полного соответствия между всеми лигандами и всеми мишенями в биологических объектах. Хемогеномика использует методы установления закономерностей на основе компьютерного анализа информации, содержащейся в химических и биологических базах данных (Rognan, 2007).

Целью создания и развития компьютерной программы PASS (Prediction of Activity Spectra for Substances) является прогнозирование всех видов биологической активности, которые химическое вещество может проявить при взаимодействии с биологическими объектами (Филимонов и др., 1995; Поройков, 1999; Poroikov, Filimonov, 2002, 2005; Филимонов, Поройков, 2006). Прогноз осуществляется на основе анализа обширной обучающей выборки, содержащей информацию о структуре извест-

ных биологически активных веществ и их взаимодействиях с биологическими объектами.

Оригинальность разработанного нами подхода к прогнозированию большого числа видов биологической активности по структурной формуле химического соединения, реализованного в компьютерной программе PASS, была отмечена в ряде публикаций (Muegge, 2003; Jonsdottir *et al.*, 2005; Chen *et al.*, 2006; Balakin *et al.*, 2006; Jenkins *et al.*, 2006; Muster *et al.*, 2008; и др.).

PASS является «инструментом» для виртуальной хемогеномики, позволяющим идентифицировать *in silico* наиболее вероятные лиганды для известных мишеней и, наоборот, наиболее вероятные мишени для известных лигандов, а также предсказывать эффекты, обусловленные взаимодействием конкретных лигандов с конкретными мишенями.

В настоящей работе представлено краткое описание PASS и приведены примеры ее практического применения для установления соответствия между лигандами, их возможными биологическими мишенями и вызываемыми этими взаимодействиями эффектами.

### Методика прогнозирования спектров биологической активности

Компьютерное прогнозирование биологической активности осуществляется в программе PASS на основе анализа взаимосвязей «структура–активность» для обширной выборки известных биологически активных веществ (БАВ).

Основные компоненты прогнозирующей системы, помимо обучающей выборки, включают представление биологической активности, описание химической структуры, математический метод анализа зависимостей «структура–активность» и прогноза биологической активности для новых веществ.

Поскольку поиск новых и исследование известных БАВ происходят постоянно, для повышения эффективности компьютерного прогноза необходимы регулярное пополнение и коррекция содержащейся в обучающей выборке информации. Так, в 1995 г. обучающая выборка содержала 9314 БАВ, число прогнозируемых наименований биологической активности составляло 114, а средняя точность прогноза при скользящем контроле с исключением по одному (см. ниже) для всех веществ и всех наименований активности обучающей выборки – 76 %; а в 2007 г. – 117332 БАВ, 3300 наименований активности и 94 % соответственно. Наиболее современная версия программы (бета-версия PASS 2008, июль 2008 г.) содержит 200126 БАВ в обучающей выборке и прогнозирует примерно 4000 наименований биологической активности со средней точностью около 94 %.

Биологическая активность представлена в программе PASS качественно («неактивно»/«активно»), что позволяет объединять в одной обучающей выборке информацию о БАВ, собранную из многочисленных источников. Вместе с тем при наличии соответствующих экспериментальных данных, полученных в одинаковых условиях, можно классифицировать вещества и с учетом величины активности (например «нет

активности»/«слабая активность»/«умеренная активность»/«сильная активность»).

Список прогнозируемых PASS (версия 2008 г.) биологических активностей включает около 4000 наименований<sup>1</sup>, из них 140 наименований, характеризующих блокирование эндогенных биорегуляторов (например Acetylcholine antagonist); 71 наименование, характеризующее активирование эндогенных биорегуляторов (например Acetylcholine agonist); 5 наименований, характеризующих блокирование высвобождения эндогенных биорегуляторов (например Acetylcholine release antagonist); 9 наименований, характеризующих стимуляцию высвобождения эндогенных биорегуляторов (например Acetylcholine release agonist); 9 наименований, характеризующих блокирование обратного захвата эндогенных биорегуляторов (например Adenosine uptake inhibitor); 2219 наименований, характеризующих ингибирование ферментов (например 12 Lipoxygenase inhibitor); 41 наименование, характеризующее активацию ферментов (например ATPase stimulant); 268 наименований, характеризующих блокирование рецепторов (например 5 Hydroxytryptamine 1 antagonist); 121 наименование, характеризующее активацию рецепторов (например 5 Hydroxytryptamine 1 agonist); 28 наименований, характеризующих блокирование ионных каналов (например Chloride channel antagonist); 5 наименований, характеризующих активацию ионных каналов (например Calcium channel agonist); 28 наименований, характеризующих блокирование транспортеров (например GABA transporter 1 inhibitor); 128 наименований, характеризующих вещества в качестве субстратов ферментов метаболизма<sup>2</sup> (например CYP3A4 substrate); 24 наименования, характеризующих ингибирование ферментов метаболизма (например CYP3A4 inhibitor); 24 наименования, характеризующих индукцию ферментов метаболизма (например CYP3A4 inducer); 28 наименований, характеризующих ингибирование белков, не являющихся

<sup>1</sup> В программе PASS используются англоязычные наименования видов биологической активности, широко распространенные в базах данных по фармакологически активным соединениям (см., например <http://integrity.prouos.com/>).

<sup>2</sup> В фармакологии различают ферменты, являющиеся мишенями фармакологических веществ, воздействие на которые приводит к определенным фармакологическим эффектам, и ферменты, осуществляющие метаболизм лекарств (фармакодинамика и фармакокинетика взаимодействия вещества с организмом). Поскольку программа PASS изначально создавалась для компьютерной оценки новых фармакологических веществ, в ней используется соответствующая фармакологическая терминология, различающая ферменты-мишени и ферменты метаболизма лекарств.

ферментами (например Collagen inhibitor); 8 наименований, характеризующих ингибирование экспрессии транскрипционных факторов (например Transcription factor Rho Inhibitor); 2 наименования, характеризующих усиление экспрессии транскрипционных факторов (например Transcription factor TP53 expression enhancer); 72 наименования, характеризующих блокирование биологических процессов (например Apoptosis antagonist); 40 наименований, характеризующих стимулировать биологических процессов (например Apoptosis agonist); 66 наименований, характеризующих воздействие веществ на инфекционные агенты (например Antileishmanial); 261 наименование, характеризующее фармакотерапевтическое действие (например Anxiolytic); 389 наименований, характеризующих побочное/токсическое действие веществ (например Carcinogen).

Для описания структуры органических соединений в PASS выбрана в качестве основы структурная формула. Это соответствует задаче прогноза спектров активности новых, планируемых к синтезу (сконструированных *in silico*), органических соединений. Стереохимические особенности молекул не учитываются, несмотря на очевидную зависимость от них биологической активности. Это обусловлено невозможностью обеспечить полноту такой информации для достаточно большой выборки структур, а также тем, что единой универсальной стереохимической номенклатуры нет, и наиболее адекватным способом представления стереохимической информации является на самом деле пространственная структура молекул. На основе структурной формулы вещества PASS генерирует так называемые MNA-дескрипторы (Multilevel Neighborhoods of Atoms – многоуровневые атомные окрестности), которые строятся рекурсивно для каждого атома молекулы следующим образом:

– MNA-дескриптор 0-го уровня – метка *A* самого атома;

– MNA-дескриптор любого следующего уровня – условное обозначение структурного фрагмента  $A(D_1D_2...D_i...)$ , где  $D_i$  – MNA-дескриптор предыдущего уровня для *i*-го непосредственного соседа данного атома с меткой *A*. Дескрипторы соседей  $D_1D_2...D_i...$  записываются в каком-либо однозначном порядке, например лексикографическом.

В настоящее время в PASS используется бесповторное множество MNA-дескрипторов 1-го и 2-го уровней; при этом в дескрипторах 2-го уровня используется индикатор « $\leftrightarrow$ » для обозначения атомов в цепях.

Детальное описание MNA-дескрипторов приведено в публикации Filimonov *et al.* (1999), а методология их разработки описана в статье Д.А. Филимонова и В.В. Поройкова (2006).

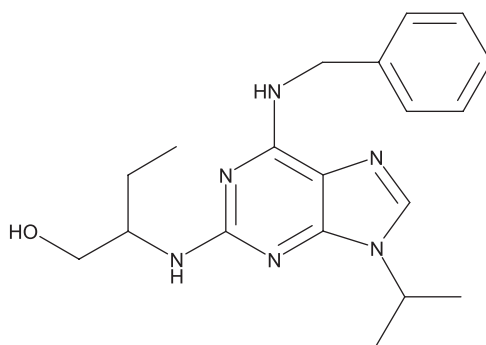
Используемый в PASS математический подход детально описан в публикациях (Филимонов, Поройков, 2006; Filimonov, Poroikov, 2008).

Результат прогноза спектра биологической активности представлен в PASS в виде упорядоченного списка названий соответствующих активностей и вероятностей  $P_a$  «быть активным» и  $P_i$  «быть неактивным» для прогнозируемого соединения. Упорядочение выполнено по убыванию разности  $P_a - P_i$ , соответственно более вероятные виды активности находятся в начале спрогнозированного спектра. Прогноз спектра активности может анализироваться в соответствии с потребностями пользователя, но по умолчанию в него включаются активности, для которых  $P_a > P_i$ .

На рис. 1 в качестве примера приведены структурная формула и часть прогнозируемого PASS (версия 2007) спектра биологической активности для препарата Seliciclib (R-Roscovitine, Cyclacel Ltd.).

Как видно из рис. 1, известное для данного препарата ингибирование CDK1 прогнозируется с вероятностью, превышающей 50 % (выделено полужирным шрифтом). В качестве дополнительных фармакотерапевтических эффектов прогнозируется Systemic lupus erythematosus treatment, Multiple sclerosis treatment и Myocardial ischemia treatment, а в качестве действия на дополнительную мишень – Hydrolase inhibitor.

В процессе обучения программы реализуется процедура скользящего контроля с исключением по одному. При этом каждое вещество обучающей выборки «исключается» из нее со всей сопутствующей информацией о биологической активности, и прогноз для него осуществляется так же, как для нового соединения. Процедура повторяется для всех веществ обучающей выборки, в результате чего мы получаем оценки точности прогноза,



```

> <PASS_RESULT_COUNT>
6 of 3300 Possible Activities at Pa > 0.500

> <PASS_ACTIVITY_SPECTRUM>
0.785 0.002 Cyclin-dependent kinase inhibitor
0.677 0.051 Systemic lupus erythematosus treatment
0.625 0.001 Cyclin-dependent kinase 1 inhibitor
0.555 0.065 Multiple sclerosis treatment
0.600 0.122 Hydrolase inhibitor
0.591 0.148 Myocardial ischemia treatment

```

Рис. 1. Пример прогноза спектра биологической активности для препарата Seliciclib.

усредненные по всей обучающей выборке и по подвыборкам веществ, обладающих конкретными видами биологической активности. Как уже говорилось выше, средняя точность прогноза составляет около 94 %. Минимальная точность прогноза составляет около 70 %, например для Antineoplastic (Renal cancer); максимальная – 100 % (для ряда активностей, включая 3C-Like protease (Human coronavirus) inhibitor, Angiotensin AT1 A receptor antagonist, Antidote (organophosphates) и др.

В специальном компьютерном эксперименте было показано, что алгоритм прогноза PASS является устойчивым по отношению к неполноте информации о структурах и спектрах биологической активности химических соединений в обучающей выборке. Сокращение реально известной информации о структурах или активности химических соединений в обучающей выборке на 50 % лишь незначительно уменьшало точность прогноза в перекрестном контроле (Poroikov *et al.*, 2000).

Величины  $P_a$  и  $P_i$  интерпретируются как оценки меры принадлежности вещества к классам активных и неактивных соединений либо как оценки ошибок первого и второго рода. Чем больше для конкретной активности величина  $P_a$  и чем меньше величина  $P_i$ , тем больше шанс обнаружить данную активность в эксперименте.

В дальнейшем мы будем рассматривать ситуации, когда значение  $P_a$  достаточно высоко и значительно превосходит  $P_i$ . Если при анализе прогноза спектра активностей для исследования выбираются те виды активности, для которых  $P_a > 90\%$ , то мы рискуем пропустить около 90 % действительно активных соединений, но вероятность ложноположительных прогнозов при этом ничтожно мала; для  $P_a > 80\%$  пропустим уже только 80 % активных соединений, но и вероятность ложноположительных прогнозов будет выше, наконец, для  $P_a > P_i$  вероятности ошибок первого и второго рода равны. Этот порог используется «по умолчанию» для отсекаания вероятных видов активности при прогнозе.

Если прогноз спектра активности обширен, то структура молекулы данного вещества довольно проста и не содержит каких-либо особенностей, обеспечивающих высокую селективность его биологического действия.

#### Возможности и примеры использования программы PASS

Как уже говорилось выше, PASS позволяет предсказывать *in silico*: 1) вероятные мишени для известных лигандов; 2) вероятные лиганды для известных мишеней; 3) вероятные фарма-



котерапевтические эффекты, а также побочное и токсическое действие веществ.

С целью расширения возможностей использования PASS научным сообществом мы реализовали веб-сервис PASS INet (<http://www.ibmc.msk.ru/PASS>), позволяющий осуществить прогноз спектра биологической активности веществ через Интернет (Lagunin, *et al.*, 2000; Садым и др., 2002; Sadym *et al.*, 2003). Используя стандартный Интернет-браузер, пользователь может направить на сайт <http://www.ibmc.msk.ru/PASS/> структурную формулу вещества, представленную в виде MOL-файла, и автоматически получить на дисплее своего компьютера прогноз наиболее вероятных для данной молекулы наименований биологической активности. В 2007 г. число зарегистрированных пользователей составило около 3000, а число структур молекул, направленных на прогноз спектра биологической активности, более 70000.

На основе прогнозируемых PASS INet спектров биологической активности синтезировано и протестировано около 20 новых молекул, принадлежащих к различным химическим классам; результаты прогноза подтверждены в эксперименте. В частности, показано наличие ГАМК-эргической и анксиолитической активности у производных хиназолинона, ноотропной активности у циклических нитронов, ЦНС-модулирующей активности у производных азетидин-2-она, противовоспалительной активности у производных 1-ациламиноалкил-3,4-диалкоксибензола, антибактериальной активности у замещенных амидов и гидразидов дикарбоновых кислот, антилейшманиозной активности у 9-амино-2-метоксиакридинов и др. (Geronikaki *et al.*, 2008).

Обширный список прогнозируемых PASS наименований биологической активности позволяет вести поиск как моноагетных лигандов, действующих на отдельные мишени, так и мультитагетных веществ, действующих на две или более мишеней. Виртуальный скрининг может проводиться в базах данных (БД) коммерчески доступных образцов химических веществ, поставляемых рядом фирм (см., например, <http://www.chembridge.com>, <http://www.ibscreen.com>, <http://www.asinex.com>).

Рассмотрим в качестве примера поиск моно- и мультитагетных веществ среди веществ из

БД ChemBridge DIVERSet (версия 2006 г.). С помощью PASS (версия 2007 г.) был выполнен прогноз спектров биологической активности для 39856 из 40080 молекул, содержащихся в этой БД (224 структуры являются двухкомпонентными, и для них прогноз не был получен). Время прогноза на стандартном персональном компьютере – около 15 минут, т. е. около 0,02 секунды/молекулу.

Результаты прогноза были введены в компьютерную программу PharmaExpert (Poroikov *et al.*, 2005), которая дает возможность поиска веществ с заданным профилем биологической активности. Мы выполнили поиск молекул, прогнозируемых как ингибиторы циклин-зависимой киназы 1, с помощью программы PharmaExpert. Оказалось, что на уровне  $Pa > Pi$  такая активность прогнозируется для 165 веществ; при  $Pa > 10\%$  – для 164 веществ; при  $Pa > 20\%$  – для 8 веществ; при  $Pa > 30\%$  – для 3 веществ; при  $Pa > 40\%$  – не прогнозируется ни для одной молекулы. Таким образом, вероятность выявления ингибиторов циклин-зависимой киназы 1 среди веществ из БД ChemBridge DIVERSet не велика. Аналогичный поиск может быть проведен в других БД, например в ChemNavigator (<http://www.chemnavigator.com>), содержащей информацию о более чем 25 млн уникальных структур.

Далее мы проанализировали возможность выявления в этой БД мультитагетных противоопухолевых веществ. Оказалось, что число веществ, прогнозируемых с  $Pa > 70\%$  как действующих одновременно на 2 мишени, составляет 395 молекул; на 3 – 334 молекулы; на 4 – 248 молекул; на 5 – 93 молекулы; на 6 – 18 молекул; на 7 – 1 молекулу. Таким образом, на основе компьютерного прогноза спектров биологической активности можно отбирать потенциальные мультитагетные противоопухолевые препараты. Ранее такого рода поиск был нами проведен для антигипертензивных (Lagunin *et al.*, 2003) и противовоспалительных веществ (Geronikaki *et al.*, 2008) с дуальными механизмами действия, и результаты прогноза были подтверждены экспериментально (Lagunin *et al.*, 2003; Geronikaki *et al.*, 2008).

На основе прогноза спектров биологической активности с помощью PASS были выявлены вещества с противоязвенным, гепатопротекторным, противоопухолевым, местноанестезирующим,



противовоспалительным, антиаритмическим, антиоксидантным, анксиолитическим, ноотропным, антибактериальным, антилейшманиозным и рядом других эффектов (Филимонов, Поройков, 2006; Geronikaki *et al.*, 2008). Нами также были показаны возможности применения PASS для прогноза удлинения QT-интервала (Filz *et al.*, 2007) и канцерогенности (Lagunin *et al.*, 2005).

Приведенные примеры убедительно демонстрируют, что с использованием PASS можно успешно предсказывать как основные фармако-терапевтические, так и побочные эффекты уже на стадии конструирования молекул *in silico* до проведения реального химического синтеза и биологического тестирования.

### Заключение

Применение PASS позволяет решать некоторые задачи хемогеномики *in silico*, однако обязательным условием этого является предварительное наличие информации о некотором числе лигандов, которые могут составить обучающую выборку. В случае новых мишеней, для которых такого рода лиганды еще не известны, сначала должен быть проведен их поиск экспериментальными методами.

Представляется весьма перспективным применение PASS для поиска мультитагетных лигандов путем виртуального скрининга баз данных коммерчески доступных образцов химических соединений, поскольку такие препараты могут обладать значительными преимуществами в сравнении с монотагетными лекарствами (Wermuth, 2004).

Совместное использование программ PASS и PharmaExpert дает возможность анализа межлекарственных взаимодействий, что весьма актуально как при разработке комбинированных синтетических препаратов, так и при оценке действия природных субстанций, полученных на основе лекарственных растений.

### Благодарности

Работа выполнена при частичной финансовой поддержке грантов РФФИ (05-07-90123, 06-03-08077, 06-03-39015), CRDF (RC1-2064), ИНТАС (03-55-5218), МНТЦ (574, 3197, 3777), FP6 (LSHB-CT-2007-037590), FP7 (200787).

### Литература

- Поройков В.В. Компьютерное предсказание биологической активности веществ: пределы возможного // Химия в России. 1999. № 2. С. 8–12.
- Филимонов Д.А., Поройков В.В. Прогноз спектров биологической активности органических соединений // Рос. химич. журнал. 2006. Т. 50. № 2. С. 66–75.
- Филимонов Д.А., Поройков В.В., Караичева Е.И. и др. Компьютерное прогнозирование спектра биологической активности химических соединений по их структурной формуле: система PASS // Эксперим. клинич. фармакология. 1995. Т. 58. № 2. С. 56–62.
- Садым А.В., Лагунин А.А., Филимонов Д.А. и др. Интернет-система прогноза спектра биологической активности химических соединений // Хим.-фарм. журнал. 2002. Т. 36. № 10. С. 21–26.
- Balakin K.V., Tkachenko S.E., Kiselyov A.S. *et al.* Focused chemistry from annotated libraries // Drug Discov. Today: Technologies. 2006. V. 3. № 4. P. 397–403.
- Caron P.R., Mullican M.D., Mashal R.D. *et al.* Chemogenomic approach to drug discovery // Cur. Opin. Chem. Biol. 2001. V. 5. № 4. P. 464–470.
- Chen X., Liang Y., Xu J. Toward automated biochemotype annotation for large compound libraries // Mol. Diversity. 2006. V. 10. № 3. P. 495–509.
- Geronikaki A., Druzhilovsky D., Zakharov A. *et al.* Computer-aided predictions for medicinal chemistry via Internet // SAR and QSAR in Environ. Res. 2008. V. 19. № 1/2. P. 27–38.
- Geronikaki A.A., Lagunin A.A., Hadjipavlou-Litina D.I. *et al.* Computer-aided discovery of anti-inflammatory thiazolidinones with dual cyclooxygenase/lipoxygenase inhibition // J. Med. Chem. 2008. V. 51. № 6. P. 1601–1609.
- Filimonov D.A., Poroikov V.V. Probabilistic approach in activity prediction // Chemoinformatics Approaches to Virtual Screening / Ed. A. Varnek, A. Tropsha. Cambridge: RSC Publishing, 2008.
- Filimonov D., Poroikov V., Borodina Yu. *et al.* Chemical similarity assessment through multilevel neighbourhoods of atoms: definition and comparison with the other descriptors // J. Chem. Inf. Comput. Sci. 1999. V. 39. № 4. P. 666–670.
- Filz O., Lagunin A., Filimonov D. *et al.* Computer-aided prediction of QT-prolongation // SAR and QSAR Environ. Res. 2007. V. 19 № 1/2. P. 81–90.
- Jenkins J.L., Bender A., Davies J.W. *In silico* target fishing: predicting biological targets from chemical structure // Drug Discov. Today: Technologies. 2006. V. 3. № 4. P. 413–421.

- Jonsdottir S.O., Jorgensen F.S., Brunak S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates // *Bioinformatics*. 2005. V. 21. № 10. P. 2145–2160.
- Kubinyi H. Chemogenomics in drug discovery // *Ernst Schering Res. Found. Workshop*. 2006. V. 58. P. 1–19.
- Lagunin A.A., Dearden J., Filimonov D.A. *et al.* Computer-aided rodent carcinogenicity prediction // *Mutat. Res.* 2005. V. 586. № 2. P. 138–146.
- Lagunin A.A., Gomazkov O.A., Filimonov D.A. *et al.* Computer-aided selection of potential antihypertensive compounds with dual mechanisms of action // *J. Med. Chem.* 2003. V. 46. № 15. P. 3326–3332.
- Lagunin A., Stepanchikova A., Filimonov D. *et al.* PASS: prediction of activity spectra for biologically active substances // *Bioinformatics*. 2000. V. 16. № 8. P. 747–748.
- Muegge I. Selection criteria for drug-like compounds // *Med. Res. Rev.* 2003. V. 23. № 3. P. 302–321.
- Muster W., Breidenbach A., Fisher H. *et al.* Computational toxicology in drug development // *Drug Disc. Today*. 2008. V. 13. № 7. P. 303–310.
- Poroikov V.V., Filimonov D.A. How to acquire new biological activities in old compounds by computer prediction // *J. Comput. Aid. Mol. Des.* 2002. V. 16. № 11. P. 819–824.
- Poroikov V., Filimonov D. PASS: Prediction of Biological Activity Spectra for Substances // *Predictive Toxicology / Ed. Christoph Helma*. London: Taylor and Francis, 2005. P. 459–478.
- Poroikov V.V., Filimonov D.A., Borodina Yu.V. *et al.* Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds // *J. Chem. Inform. Comput. Sci.* 2000. V. 40. № 6. P. 1349–1355.
- Poroikov V., Lagunin A., Filimonov D. PharmaExpert: diseases, targets and ligands – three in one // *QSAR and Molecular Modelling in Rational Design of Bioactive Molecules / Ed. Esin Aki Sener, Ismail Yalcin*. Ankara (Turkey): CADD & D Society, 2005. P. 514–515.
- Rognan D. Chemogenomic approach to rational drug design // *British J. Pharmacol.* 2007. V. 152. № 1. P. 38–52.
- Sady M., Lagunin A., Filimonov D. *et al.* Prediction of biological activity spectra via Internet // *SAR and QSAR Environ. Res.* 2003. V. 14. № 5/6. P. 339–347.
- Wermuth C.G. Multitargeted drugs: the end of the «one-target-one-disease» philosophy? // *Drug Discov. Today*. 2004. V. 9. № 19. P. 826–827.

## COMPUTER-AIDED PREDICTION OF BIOLOGICAL ACTIVITY SPECTRA FOR SUBSTANCES: VIRTUAL CHEMOGENOMICS

V.V. Poroikov, D.A. Filimonov, T.A. Glorizova, A.A. Lagunin,  
D.S. Druzhilovsky, A.V. Stepanchikova

The Orekhovich Institute of Biomedical Chemistry of Russian Academy of Medical Sciences,  
Moscow, Russia, e-mail: vladimir.poroikov@ibmc.msk.ru

### Summary

On the basis of predicted with computer program PASS biological activity spectra it is possible to reveal the most probable molecular targets, with which chemical compounds interact in biological objects, and biological effects caused by these interactions. Average accuracy of prediction is about 94 %, and the method is robust in relation to the incompleteness of information in the training set. Wide utilization of this approach as a «tool» for virtual chemogenomics (<http://www.ibmc.msk.ru/PASS>) provides the opportunities for independent experimental confirmation of predictions in a dozen of cases.

## THE WITHIN-INDIVIDUAL BASIS OF BETWEEN-INDIVIDUAL DIFFERENCES

V.G. Cherdantsev, V.A. Scobeyeva

Department of Biological Evolution, Faculty of Biology, Moscow State University, Moscow, Russia,  
e-mail: VGCherdantsev@yandex.ru; skobei-khanum@yandex.ru

From the analysis of quantitative morphological dynamics of amphibian gastrulation, we infer a general law that, in the developmental series of repeatable embryonic areas subject to shaping, the proportion of within- to between-individual variation in different areas depends primarily only on a spatiotemporal order of their formation. The younger («distal») areas are more variable and develop at higher rates than the older («proximal») ones. Despite the absence of a notion of fitness, the general principle of morphogenetic dynamics is mathematically the same as that of selection, since the system tends to minimization of within-individual variance at a rate proportional to its amount. Thus, the within-individual component of variation seems to provide a driving force of morphogenesis, which means that the concept of developmental canalization is a bias.

**Key words:** variability, morphogenesis, repeatability, selection.

### Introduction

When in a sample of the adult organisms the amount of variance in metric characters expressed in the values of the coefficients of variation (*CV*) is about 20 % and more the biologist would suspect that something is wrong with sampling. Meanwhile, when we consider the embryonic structures, the values of the *CV* at a level of 15–20 % correspond not to the upper, but rather to the lower limit of variation in morphological characters that are subject to change in the course of normal development (Cherdantsev, Scobeyeva, 1994; Scobeyeva, Cherdantsev, 1999; Cherdantsev, 2003; Scobeyeva, 2006). Thus, the *CV* values of egg sizes in a single clutch of the common frog (*Rana temporaria*) spawn in nature never exceed 6–7 %. In the beginning of gastrulation, the yolk plug diameter varies in the same clutch at a 15–18 % level while characteristics of the dorsal blastopore lip shape vary in a range from 27 to 67 % (Cherdantsev, Scobeyeva, 1996; Cherdantsev, 2003). Variation of this order of magnitude is common to genetically homogeneous groups of embryos (siblings) developing in optimal environmental conditions and having the same developmental age (Scobeyeva, 2006). It follows that

the normal developmental variation can be referred neither to genetic nor environmental components of variance. It merits being considered as a special developmental component of natural phenotypic variance – as the inherent developmental variability (IDV, see Cherdantsev *et al.*, 1996).

There is a principal difference between the IDV and so-called «intangible» developmental variation (Falconer, 1981). The last refers to «accidents» or «errors» of development and, consequently, should increase with time. In contrast to that, the IDV is subject to increase with initiating of each new morphogenetic shaping of the embryo and decreases, as the developing structure acquires its eventual shape. This holds both for the shaping itself, such as the dorsal blastopore lip shaping in amphibian embryos (Cherdantsev, Scobeyeva, 1994; Cherdantsev, 2003; Scobeyeva, 2006), and positioning of gene expression domain boundaries, such as those between anteroposterior segments of *Drosophila* embryos (Houchmandzadeh *et al.*, 2002; Jaeger *et al.*, 2004). In both cases, instead of interfering with developmental pathways, the IDV seems to drive them, selecting and differentially enhancing the «appropriate» fluctuations, which suggests that identification of the IDV dynamics

with Waddington's canalization (Waddington, 1940) is not a matter of fact.

The nature of IDV seems to be the same as that of within-individual differences between the repetitively (serially) homologous structures of the adult organism. The repetitive homology means that a given series of structures, irrespective of the degree of their morphological divergence, arise with modification of parameters of the same developmental (parametric) program (van Valen, 1982). By definition, if more than one measurement of a morphological character can be made in each individual in space and time, the phenotypic variance can be partitioned into variance within individuals and variance between individuals (Falconer, 1981).

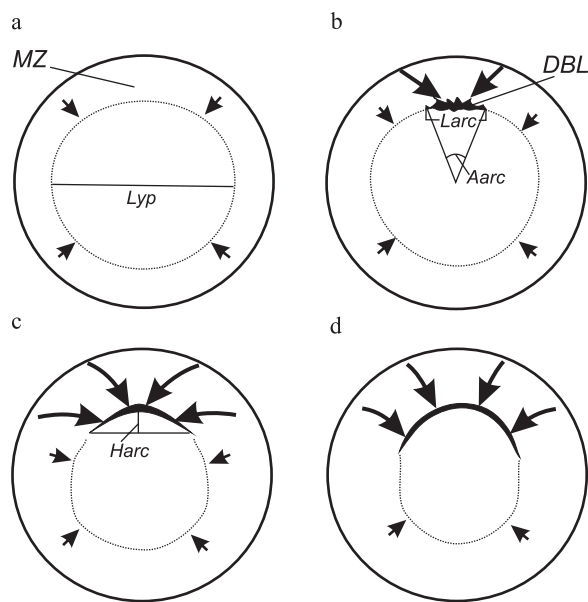
One of the basic principles of morphogenesis is that of spatial unfolding stating that the shape differences of embryonic areas display the succession of shaping of the same area (Cherdantsev, 2003, 2006). Therefore, a measurement made in the structure can be partitioned into components corresponding to the consecutive stages of its shaping. This is sufficient for considering the developing structure subject to shaping as an analogue of repetitively homologous structures, which provides an opportunity of partitioning of their variance into the between- and within-individual components.

Our aim here is to show that lying at the heart of the IDV is the within-individual variation. The point is simply that, if the differences between repetitively homologous parts form a one-parametric set whose elements are at consecutive stages of the same developmental transformation, then their own variances are added by positive covariation being proportional to the degree of within-individual spatial differences. The between-individual variance increases at the expense of the within-individual one, that is, at the expense of spatial heterogeneity of the developing structures in each given individual. This explains why variation in quantitative characters connected with embryonic shaping proves not a hindrance, but rather an impetus of their directional change. We will purpose to show that in the developing structure consisting of repetitively homologous elements each element moves towards minimization of within-individual variance of the whole structure at a rate proportional to the amount of variance.

### Spatiotemporal repeatability in amphibian gastrulation and its implications

**Analysis of variance in synchronously developing embryos fixed at consecutive stages of morphogenesis.** The hierarchical commitment inherent to determination of embryonic rudiments is a particular case of the spatiotemporal series of repeatable serially homologous structures. In amphibian gastrulation this series begins with outlining of the blastopore circumference, the so-called marginal zone (MZ), whose inner diameter (*Lyp*, see Fig. 1, a) is the only metric character that can be measured at the vegetal view of the whole gastrula prior to the appearance of the dorsal blastopore lip (DBL). The further succession of morphological events develops as a series of local shaping of the blastopore circumference initiating at the dorsal pole of the MZ and traveling to the ventral pole (Figs. 1, b–d).

The DBL appears as a cleft whose orientation is that of the tangent to the blastopore circumference (Fig. 1, b). It follows that at this stage the DBL has no planar curvature on its own and the only measurements characterizing the shape of DBL are that



**Fig. 1.** Spatiotemporal repeatability of morphogenesis in frog gastrulation.

a – blastopore circumference prior to the onset of gastrulation, b – initiation of DBL, c – DBL develops its own curvature, d – the DBL curvature forces out that of the blastopore circumference; *Bl* – blastopore, for other designations see text. Pointers – cell flows at the blastopore circumference, arrows – cell flows shaping the DBL arc.



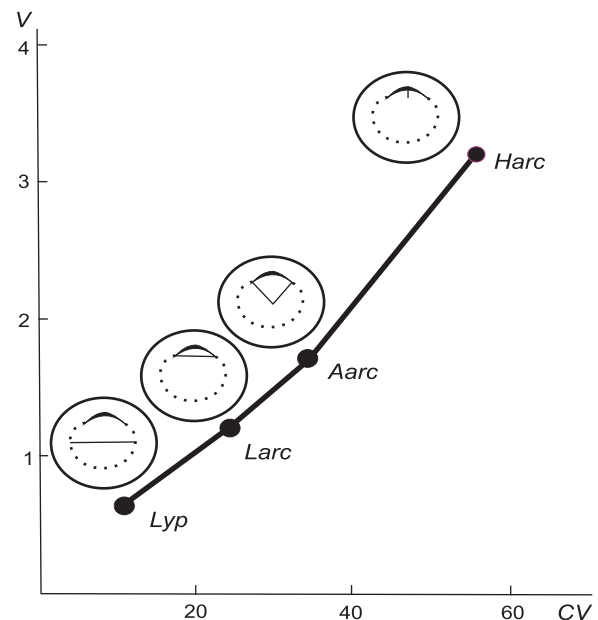
of its linear (*Larc*, see Fig. 1, b) and angular (*Aarc*, see Fig. 1, b) lengths. Later on the DBL forms its own curvature which is distinct from that of the blastopore circumference and, starting from this moment, the measurements can be added by the linear height of the DBL arc (*Harc*, see Fig. 1, c), which is a measure of the arc curvature. We consider an interval between the arc formation and its conversion into the semicircular DBL (Figs. 1, b–d). At this interval, one can measure all listed characters and each of them does not change the direction of its temporal change: *Larc*, *Aarc* and *Harc* are subject to increase while the *Lyp* is subject to decrease up to the blastopore closure.

The repetitive homology of changes that occur over the MZ both in space in time becomes evident when we are considering patterns of the mass cell flows (arrows in Fig. 1) molding new shapes of the blastopore circumference. The MZ shape prior to the onset of gastrulation implies that cell flows have normal orientation, that is, the cells move along the tracks that are coincident with the normal radii of curvature of the blastopore circumference (Fig. 1, a). The initial (straight) shape of the DBL means (for the details see Cherdantsev, 2003, 2006; Scobeyeva, 2006) that cell flows deviate from normal orientation to converge in the centre of the DBL line (Fig. 1, b). Their convergence means that they shape the DBL line converting it into the arc with its own curvature. The arc formation normalizes the orientation of cell flows at new radii of curvature, which means deviating from normal orientation of the bordering flows of cells and, consequently, spreading of the DBL in space (Fig. 1, c). New curvature of the blastopore circumference arising with shaping of the DBL arc gradually forces out the initial one (Fig. 1, d). Thus, the DBL arc develops as a spatial unfolding of a single shaping process generating a dorsoventral series of repetitively homological DBL shapes. In a correspondence to the unfolding principle, at one pole of the series is an «older» DBL fragment, whose own developmental history is identical to that of the whole series, and at opposites (lateral) poles are «younger» DBL fragments having no developmental history on their own. Older fragments are «distal» in the sense that their curvature is most remote from the initial one, while younger fragments are «proximal» in the sense that their own curvature is close to the initial curvature of MZ (see Fig. 1).

The relationships between the coefficients of variation of metric characters and their specific rates were studied in fixed embryos of the common frog, *Rana temporaria*, developed synchronously from the eggs of the same clutches spawn in nature. The embryos were reared in the lab at optimal aeration conditions and temperature (18 °C). The first fixation was made just after the DBL had formed its own curvature (see Fig. 2).

The next fixation was made in three hours, when the DBL acquired the crescent-like shape. Each fixation consisted of 300 embryos taken from the same egg clutch. The specific rates of changes were calculated as  $V = (M_1 - M_2)/M_1$ , where  $M_1$  and  $M_2$  were the mean values of measurements made in the first and second fixations, and matched to the values of the coefficients of variation (*CV*) calculated for the first fixation.

The arrangement of quantitative characters in the phase space  $\{V, CV\}$  shown in Fig. 2 demonstrates a clear-cut positive connection between *CV* and *V* values. The DBL arc height (*Harc*) leading in the amount of both variance and developmental rates is a youngest character, as it arises at later gastrulation stages by bending of the straight DBL



**Fig. 2.** Interdependence between variation and specific rates in the DBL shape characteristics in the frog gastrula.

*CV* (abscissa) – coefficients of variation at the crescent-like DBL arc stage, *V* (ordinate) – specific rates of changes between the crescent-like and semicircular DBL arc stages. Designation of characters is the same as in Fig. 1.



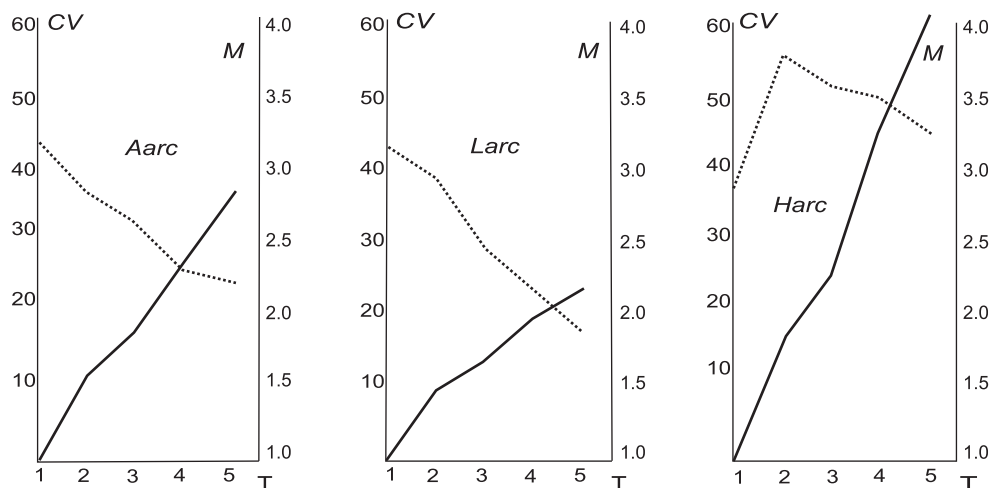
line and, consequently, can develop under the constancy of the characters *Aarc* and *Larc*. The last two characters arise by shaping of a fragment of the blastopore circumference and, as referred to the tangent of this circumference, can develop on their own under the constancy of the MZ diameter (*Lyp*). As for the character *Lyp* having the least variance and slowest developmental rate, it refers to the «oldest» part of the MZ zone whose shape has formed prior to the gastrulation onset.

The additional variance of younger characters, especially *Harc*, is a direct consequence of the convergence of cell flows, which provides an opportunity of changing of the DBL arc curvature under the constancy of its linear and angular lengths. This, in turn, provides a positive feedback between the arc shaping in both space and time. In fact, both the lateral spreading of DBL and the increase in its curvature at the dorsal pole lead to an increase of spatial heterogeneity of the DBL as a whole. This stimulates lateral (youngest) DBL fragments to take over the dorsal ones until the new curvature becomes uniform over the whole DBL. Earlier we have shown (Cherdantsev, 2006; Scobeyeva, 2006) that this occurs just at a transition of DBL from the crescent-like to semicircular shape when the orientation of cell flows normalizes in all DBL fragments (cf. orientation of the cell flows shown in Figs. 1, b–d). It follows that the decrease of variation is not ca-

nalization in Waddington’s sense. Rather, this is a direct and natural consequence of vanishing of the positive feedback between the local and non-local behavior of the DBL parts.

**Analysis of variance in the individual pathways by repeated measurements of quantitative characters in living embryos.** Up to this point, we have been considering the spatial repeatability of the developing structure. In order to evaluate its temporal component in the explicit form, one needs to have a series of measurements made at the same individuals, which presumes measuring of the living embryos rearing in standardized lab conditions. From the same egg clutches of *R. temporaria* spawned in nature, we selected embryos with the synchronous onset of gastrulation. Selected embryos develop in individual plastic wells at 18 °C. At 40 min intervals, we oriented each individual embryo the vegetal pole upwards and made the photographs at which we measured the metric characters *Lyp*, *Larc*, *Aarc* and *Harc* (see Figs. 1, a–c). In total, we measured 150 embryos from seven egg clutches, each character in each embryos being subject of 5 successive measurements. The mean values of characters we expressed in conventional units equal to 0.05 mm.

The time dynamics of the coefficients of variation (*CV*) and mean values (*M*) of the characters *Harc*, *Larc* and *Aarc* is shown in Fig. 3, the number of a measurement (1–5, abscissa) being the measure of time.



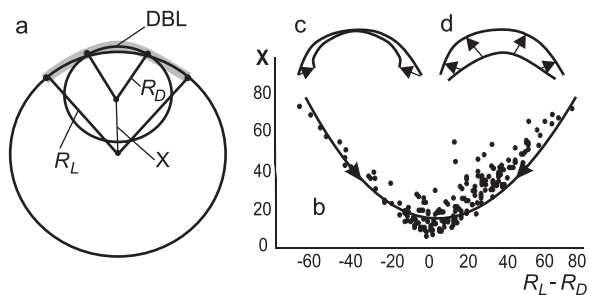
**Fig. 3.** The time dynamics of variation and means in the frog gastrulation.

Abscissa (T): measurements numbers (1–5). Ordinates: coefficients of variation (*CV*, dotted line) and mean values (*M*, solid line) of characters (for designations see text) are shown at corresponding graphics. These data are those of measuring of the living embryos in 40 min intervals from the appearance of the gastral cleft to the semicircular DBL arc. For each character, their mean values are those normalized to the initial means.

The variation is at a highest level at the onset of the DBL arc formation and decreases, as the arc acquires the spatially homogeneous shape, the semicircular one, whose formation roughly corresponds to a measurement number five. Note that for the characters *Aarc* and *Larc* the maximum of variance corresponds to the very beginning of gastrulation when the DBL line has no a curvature on its own. As for the character *Harc*, its variability reaches the maximum just at a stage (see Fig. 1, c) at which the change in the DBL arc curvature becomes the preponderant morphogenetic process. In total, the general variation of the character *Harc* is much higher than that of the characters *Larc* and *Aarc* (cf. Figs. 2, 3).

Thus, as it concerns with the proportion of variability of the characters, the study of successive developmental stages yields the same general results as that of variation in synchronously fixed embryos. The same refers to matching of the variation of characters to their developmental rates. The DBL arc curvature (*Harc*) changes at a much higher rate than the arc linear and angular lengths (*Aarc* and *Larc*).

**A model of morphogenesis based on within-individual selection analog.** Our starting point is that the between-individual differences are reproducible, if (*and only if*) they have a within-individual analog. One can infer an appropriate model from the unfolding principle, as it matches between-individual differences to within-individual ones. We make use of a simple geometric construction shown in Fig. 4, a.



**Fig. 4.** A model of the DBL shaping based on the analysis of variability.

a – partitioning of the DBL arc into younger ( $R_L$ ) and older ( $R_D$ ) circles, b – plot, c and d – DBL shaping modes corresponding to two parabola branches; X – a distance between their centers; pointers – movement vectors of the dynamical system, arrows – cell flows allowing for the corresponding DBL shaping.

In each individual embryo having the crescent-like DBL, we partitioned the arc into three fragments of equal angular lengths, a dorsal (median) one and two lateral fragments that flank the dorsal region. Then, as shown in Fig. 4, a, we constructed two circles, one having a radius close to that of the dorsal region ( $R_D$ ), and the other circle with a radius close to that of the lateral DBL regions ( $R_L$ ). For the simplicity, we assume that both lateral fragments have equal curvature.

The reason for such a construction is the unfolding principle, which permits to consider these circles as corresponding to consecutive stages of the DBL shaping (see Fig. 1). Then, the evolution of shape fits to one-dimensional dynamics of a variable  $X$ , which is the distance between the centers of «younger» ( $R_L$ ) and «older» ( $R_D$ ) circles shown in Fig. 4, a.

The empirical dependence of  $X$  on  $(R_L - R_D)$  obtained from the analyses of variance of the DBL shaping and shown in Fig. 4, b perfectly fits to a parabolic curve in the phase space  $\{X, (R_L - R_D)\}$ . It follows from the distribution of data that a point at which  $(R_L - R_D)$  and  $X$  are also close to zero is an attractor of the dynamical system. This means that the system moves towards minimization of spatial variance of the DBL shape (see pointers in Fig. 4, b), which, owing to the unfolding principle, is equivalent the achievement of a state at which  $dX/dt$  is also equal to zero. The DBL moves to this state by adapting  $R_L$  to  $R_D$  (Fig. 4, c, a dominating shaping mode), or vice versa (Fig. 4, d, a minor shaping mode) as shown by arrows in corresponding figures.

It follows that, in a rough approximation,  $dX/dt = X(R_L - R_D)$ ,

where  $(R_L - R_D)$  is obviously proportional to the amount of within-individual variance in all quantitative characters whose dynamics concerns with the DBL shaping and tends to zero when the DBL acquires a uniform curvature. In general, if, in amphibians, the dorsal pole were not beforehand fixed, we would have had  $dX/dt = X(R - R^*)$ ,

where  $R$  were a radius of curvature of a given DBL fragment, and  $R^*$  – mean radius of curvature of the whole DBL. Again,  $(R - R^*)$  is obviously proportional to the amount of variance.

Thus, we can conclude that the morphogenetic system moves towards minimization of within-individual variance at a rate proportional to the

amount of variance, which is mathematically indistinguishable from what Fisher's «Fundamental Theorem of Natural Selection» states.

### Conclusion

The above data show that within-individual variability inherent to normal dynamics of morphogenesis is the main source of between-individual differences that are highly reproducible with no respect to a contribution of both the additive genetic variance and fitness. As a dynamical system, morphogenesis is subject to the same mathematical principles as those of selection, where the notion of fitness is absent being replaced by the account of relationships between the shapes of embryonic areas involved into the same shaping process. Among a variety of shaping trends, each embryonic area selects those that can be readily continued by neighboring areas. The situation is very similar to that arising in fallacious genetic experiments in which the animals with better characteristics get the better diet. In our case, this is a premise to within-individual selection based exclusively on differential capacity of different shapes of the areas to involve surroundings into their own shaping. An opportunity of this kind of selection means, first, that selection is a wider concept than fitness and, second, that the shape of a developing area carries information on its developmental trend.

It follows, in particular, that it is erroneous to identify the dynamics of inherent developmental variation (IDV) with developmental canalization. The reason why variation decreases with approximating of the form to the eventual state lies not in the enhancement of regulation, but rather in the exhausting of within-individual variation component allowing for both the direction and rate of morphological changes.

What is worthy to stress is that the within-individual component of variation, owing to the unfolding principle, is reproducible despite its inheritance and contribution to fitness may be close to zero. Then, insofar as within-individual differences manifest at individual level, they might

correspond to what Darwin understated under the vague term of «indefinite» variation (cf. Cherdantsev *et al.*, 1996).

### Acknowledgements

This study was supported by the Russian Foundation for Basic Research, grant 08-04-00057-a.

### References

- Cherdantsev V.G. Morphogenesis and evolution. M.: KMK Ltd Press, 2003. 360 p. (In Russian).
- Cherdantsev V.G. The dynamic geometry of mass cell movements in animal morphogenesis // *Int. J. Dev. Biol.* 2006. V. 50. P. 169–182.
- Cherdantsev V.G., Scobeyeva V.A. The morphological basis of self-organization // *Rivista Biol. Forum.* 1994. V. 87. P. 57–85.
- Cherdantsev V.G., Kreslavsky A.G., Severtsov A.S. Episelective Evolution // *Evolutionary Theory* / Ed. L. van Valen. 1996. V. 11. P. 69–87.
- Cherdantsev V.G., Scobeyeva V.A. On Gurwitsch's Morphogenetic Law // *Biophotonics* / Ed. L.V. Belousov, F.A. Popp. Moscow: Bioinformatics, 1996. P. 53–70.
- Falconer D.S. *Introduction to Quantitative Genetics*. 2nd ed. Longman, London; N.Y., 1981. 450 p.
- Jaeger J., Surkova S., Blagov M. *et al.* Dynamic control of positional information in the early *Drosophila* embryo // *Nature*. 2004. V. 430. P. 368–371.
- Houchmandzadeh B., Wieschaus E., Leibler S. Establishment of developmental precision and proportions in the early *Drosophila* embryo // *Nature*. 2002. V. 415. P. 798–802.
- Scobeyeva V.A. Cherdantsev V.G. Dorsoventral polarization of gastrulation movements in spade-footed toad embryos (*Pelobates fuscus* L.) // *Ontogenez* (Russ. J. Dev. Biol.). 1999. V. 30. P. 436–447.
- Scobeyeva V.A. The natural variability of morphogenesis: a tool for exploring the mechanics of gastrulation movements in amphibian embryos // *Int. J. Dev. Biol.* 2006. V. 50. P. 315–322.
- Valen van L. Homology and causes // *J. Morphol.* 1982. V. 8. P. 101–107.
- Waddington C.H. *Organizers and Genes*. Cambridge Univ. Press, 1940. 320 p.

## МОДЕЛИРОВАНИЕ ЭВОЛЮЦИИ ТРОФИЧЕСКИ ЗАМКНУТЫХ СООБЩЕСТВ С КОМПЕНСАТОРНЫМ И НЕКОМПЕНСАТОРНЫМ МЕТАБОЛИЗМОМ

С.А. Лашин<sup>1,2</sup>, В.В. Суслов<sup>1</sup>, Ю.Г. Матушкин<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: lashin@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Проведено сравнительное моделирование динамики адаптивности и биоразнообразия популяций симбиотических одноклеточных гаплоидных асексуальных организмов, обитающих в биотопе конечного объема, в который извне поступает неспецифический субстрат (НС), необходимый всем особям. Симбиотичность заключается в обмене между популяциями специфическими субстратами. С помощью оригинального комплекса программ «Эволюционный конструктор», позволяющего варьировать и генетическую, и трофическую структуру модели, показано, что при сублетальном голодании по НС популяции с компенсаторным метаболизмом (недостаток НС может быть частично компенсирован благодаря коэволюции симбионтов) обладают лучшей адаптируемостью (мутация в одной из популяций спасает все популяции симбионтов), но постепенно теряют биоразнообразие. Популяции с некомпенсаторным метаболизмом, напротив, сохраняют биоразнообразие вплоть до вымирания. Для популяций с компенсаторным метаболизмом выявлено критическое соотношение численности популяции и притока НС, при котором субституционный груз замены нормального аллеля на мутантный выдержать невозможно. Обсуждаются недостатки и преимущества эволюционной стратегии формирования биоты на базе немногих таксонов – универсалов и стратегии, связанной с поддержанием высокого биоразнообразия.

**Ключевые слова:** моделирование эволюции, горизонтальный перенос.

### Введение

Жизнь на Земле существует в форме взаимодействующих сообществ двух типов: популяций и экосистем. Популяции обеспечивают возможность воспроизводства и осуществления дарвинова отбора. Обмен веществом и энергией как с внешней средой, так и между популяциями – одна из основных функций экосистем, в результате которой обеспечиваются условия для воспроизводства организмов (Камшилов, 1974; Маргалев, 1992; Жерихин, 1994). Значит, любой эволюционный процесс логично рассматривать взаимодополнительно: как с популяционной, так и с экосистемной точек зрения. Следовательно, эволюционный успех носителя того или иного генотипа связан не только с фиксацией в популяции, но и с влиянием такой популяции на функционирование экосистемы. Однако в

отличие от популяции принцип конкурентного исключения (принцип Гаузе) не позволяет однозначно связать успех с численностью или частотой генотипа в экосистеме. Определение эволюционного успеха генотипа (вида, таксона) можно дать лишь через успех экосистемы в целом (Камшилов, 1974; Маргалев, 1992; Жерихин, 1994).

Так как внешняя среда постоянно меняется, эволюционный успех в общем виде можно определить как снижение вероятности гибели сообщества от неопределенных, непредвиденных причин (Красилов, 1986). Можно выделить два универсальных способа достижения такого успеха: способность к постоянному поиску новых потоков энергии и вещества (Левченко, 1993, 2003) и повышение стабильности и замкнутости системы (Одум, 1986). Если исключить миграцию извне, у экосистем остается



единственный способ достижения такого успеха – дарвинов отбор в популяциях. Таким образом, эволюционный успех связан с балансом двух характеристик: устойчивости (способности сохранять структуру и гомеостаз несмотря на изменения внешней среды) и адаптируемости (способности отвечать на неконтролируемые изменения среды<sup>1</sup> за счет неопределенной изменчивости).

Из вышесказанного следует, что программная среда для моделирования эволюции должна не только обеспечить анализ определенного статического (структурно) состояния экосистемы, но и давать возможность расчета практически любых вариантов экосистем, могущих возникнуть на базе такого эталона за счет его динамики (обратимые изменения), развития (закономерные необратимые изменения) и эволюции (спонтанные изменения в закономерностях развития, тестируемые отбором).

В настоящее время создано около 4 тыс. моделей динамики экосистем (Jørgensen, 1995, 1999). Их можно разбить на портретные (индивидуально-ориентированные), обобщенные и структурно-динамические (оптимизационные) (Левченко, 1993; Jørgensen, 1999; Fog, 2000; Суховольский, 2004). Подробные портретные модели наглядны, но вычислительно сложны (в идеале каждая особь требует описания), трудно расширяемы и статичны – их структура в ходе численных расчетов меняется лишь в преддетерминированных экспериментатором пределах. Обобщенные модели вычислительно проще, за что приходится платить меньшей подробностью описания, опуская детали структуры экосистем или задавая их в неявном виде путем взаимозависимых коэффициентов. В структурно-динамическом (оптимизационном) моделировании проблема сложности описания реального биоразнообразия решается путем анализа обобщенных целевых функций (максимум используемой энергии, минимальная энтропия, максимум биомассы и др.) с ним связанных.

<sup>1</sup> Дж. Хаксли (Huxley, 1943, 1954) выделял третий путь к эволюционному успеху – контроль над средой, за что его упрекали в антропоцентризме. Этот путь часто реализуется в паразитарных экосистемах: паразит изменяет половозрастную структуру популяции хозяина (Горячева, 2004; Юрлова, 2003) и даже его поведение (Bakker *et al.*, 1997; Hansen, Poulin, 2005), не говоря уже о подавлении иммунитета (Супотницкий, 2005).

Так, в работе G. Vendoricchio и S.E. Jørgensen (1997) предложена целевая функция «максимальная запасенная экзергия». Она позволяет оценивать для каждого таксона в экосистеме не только биомассу, но и количество информации, отталкиваясь от среднего количества генов в его геноме. Такой способ позволяет хорошо моделировать эволюцию экосистем, состоящих из эволюционно далеких таксонов.

Важной особенностью развиваемой авторами методики пошагового моделирования, реализованной в комплексе программ «Эволюционный конструктор» (ЭК), является возможность гибкого варьирования степени детальности описания моделируемых объектов. Это достигается посредством арифметики генетических спектров (см. Материалы и методы). Таким образом, модели, реализованные на ЭК, могут по желанию пользователя приближаться как к обобщенным, так и к портретным моделям, не являясь в чистом виде ни одной из них (Lashin *et al.*, 2007; Лашин и др., 2008). ЭК также позволяет в ходе моделирования изменять трофическую структуру взаимоотношений между популяциями и трофический спектр отдельных популяций (что может выражаться в изменении числа переменных модели) за счет мутаций, горизонтального переноса генов, внутри- и межвидовой конкуренции и отбора. Под трофической структурой подразумеваем организацию сообщества, основанную на трофических взаимоотношениях популяций, под трофическим спектром – набор кормов, необходимый для нормальной жизнедеятельности представителей отдельных популяций, и их количественную пропорцию, а в некоторых случаях – замещаемость кормов (Одум, 1986; Бигон и др., 1989). Спектр мутаций принципиально не преддетерминирован, хотя может быть ограничен, например, исходя из вычислительных мощностей. ЭК позволяет варьировать степень портретности описания модели, что соответствует реальному эволюционному процессу (отбор может быть как индивидуальным, так и групповым) и позволяет при необходимости оценивать поведение целевых функций.

В природе часто наблюдается компенсация используемых ресурсов (Одум, 1986; Бигон и др., 1989), когда недостаток одного из них вид частично компенсирует за счет другого (закон



компенсации ресурсов Рюбеля (Rubel, 1935)), либо взаимодействие ресурсов изменяет потребность вида в одном из них (кондиционирование ресурсов) (Беклемишев, 1970; Одум, 1986).

В данной работе проведено сравнительное моделирование динамики устойчивости, адаптируемости и биоразнообразия трофически замкнутых сообществ с компенсаторным и некомпенсаторным типами метаболизма. Нужно подчеркнуть, что все обсуждение проводится в предположении о стабильном состоянии экосистемы (стационарном или циклически изменяющемся), без которого говорить о длительном существовании экосистемы вообще нет смысла. Обычно в моделях экосистем описание метаболизма особей не отличается от описания химической реакции, имеющей постоянные стехиометрические коэффициенты. Набор таких коэффициентов задает условия существования вида, подчиняющиеся закону минимума Либиха, сформулированному в 1840 г. (Одум, 1986).

### Материалы и методы

При разработке ЭК мы исходили из того, что отбору в подавляющем большинстве случаев все равно, каким образом в онтогенезе формируются признаки, нужные для выполнения какой-либо биологической функции (Гаузе, 1941; Шмальгаузен, 1968; Черданцев, 2003). Многочисленные примеры конвергенции подтверждают, что это допущение соответствует реальности. Действительно, отбор может влиять на онтогенетический процесс лишь в случае неустойчивости такового процесса (Шмальгаузен, 1968; Иорданский, 2004; Гунбин и др., 2007), что опять-таки проявляется в высокой вариабельности признаков – в аномально широкой норме реакции на популяционном уровне (Беляев, 1979) или на уровне высоких таксонов – при отсутствии выраженного архетипа<sup>2</sup> (Рожнов, 2006). Эволюционное формирование взаимной скоррелированности признаков (архетипа) ограничивает их варьирование. Как генетические, так и палеонтологические данные подтверждают, что период высокой вариабельности

<sup>2</sup> Архетип – своего рода стандарт таксона, инвариантная комбинация признаков, определяющая основные характеристики представителей таксона (Заварзин, 1973; Любарский, 1993; Vasilyeva, 1999).

фенотипа в филогенезе относительно краток – большую часть своей жизни таксон имеет устойчивый архетип. С момента возникновения такой корреляционной решетки филогенетическая классификация начинает хорошо совпадать с типологической (сходство = родство) (Заварзин, 1973; Vasilyeva, 1999). Этот факт позволил существенно упростить подход к описанию биоразнообразия в ЭК.

В ЭК проблема сложности описания биоразнообразия элементов экосистем решена путем отказа от априорного выделения как классов элементов, так и филетических отношений между классами. Любой элемент (особь, популяция, вид) не выделяется априорно, а определяется (и переопределяется в ходе расчетов) по комбинации значений нескольких признаков (Lashin *et al.*, 2007; Лашин и др., 2008). В текущей версии ЭК это: 1) множество потребляемых субстратов; 2) множество продуцируемых продуктов; 3) множество скоростей (эффективностей) утилизации соответствующих субстратов и синтеза соответствующих продуктов.

Считается, что каждому признаку соответствует конкретный ген (потенциально подверженный мутациям), благодаря чему признак может наследоваться и изменяться. Таким образом, *ген в ЭК представляет собой единицу наследования*.

В текущей версии ЭК моделирует коэволюцию популяций трофически связанных одноклеточных асексуальных гаплоидных организмов (особей или клеток), обитающих в *окружающей среде* конечного объема с идеальным перемешиванием. С внешним миром окружающая среда связана *процессами притока и оттока* субстратов и особей. Особи могут потреблять *субстраты* и секретировать *продукты*. Генотип особи состоит из генов потребления субстратов, генов производства продуктов, при этом варианты (аллели) генов различаются по обеспечиваемой ими эффективности соответствующих процессов. Гены могут мутировать, вследствие чего количественно меняется соответствующая эффективность (Lashin *et al.*, 2007; Лашин и др., 2008)<sup>3</sup>. Баланс между группами генов определяет фенотип особи – ее потребность в субстратах,

<sup>3</sup> Мутации по генам неспецифического субстрата в текущей версии летальны. Также возможен горизонтальный перенос генов между организмами (Lashin *et al.*, 2007; Лашин и др., 2008).

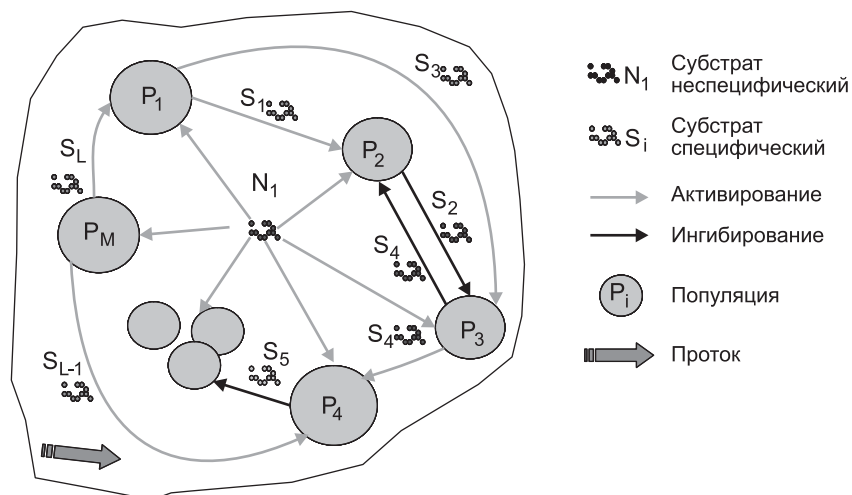
способность конкурировать за них и использовать их для производства других субстратов и размножения. Особи объединяются в *популяции* по принципу генетической близости. Популяции образуют *трофически связанные сети*. Продукты, секретлируемые одной популяцией, могут служить субстратами для другой популяции (популяций), влияя на ее размножение. Продукты/субстраты, секретлируемые популяциями, будем называть *специфическими*. Концентрация специфических субстратов в окружающей среде определяется отношением скорости их синтеза к скорости их потребления, а также к скорости и концентрации оттока. Существуют также *неспецифические общие субстраты*, чье присутствие необходимо для жизни всех особей. Концентрации неспецифических субстратов в окружающей среде зависят от соотношения притока и оттока и от интенсивности их потребления. На рис. 1 представлена схема основных объектов и процессов предлагаемой методики моделирования (показан случай с единственным неспецифическим субстратом  $N_1$ ).

Сочетание портретного и обобщенного подходов моделирования позволяет вычислительно эффективно моделировать полиморфизм по многим генам, изменения генетической

структуры популяций, ведущие к значительной перестройке структуры модели (вплоть до изменения числа уравнений и переменных). К таким изменениям относятся как изменения частот аллелей вследствие отбора и мутаций, так и горизонтальный перенос генетического материала (Lashin *et al.*, 2007; Лашин и др., 2008). Также ЭК позволяет задавать различные стратегии размножения и трофические спектры для разных популяций.

## Результаты

Сравнивается эволюция двух трофических колец (ТК) при сублетальном дефиците большинства субстратов. В обоих случаях популяции, выделяя специфические субстраты, активируют развитие соседа по ТК. Для простоты на рисунках количество популяций в ТК ограничивалось тремя: популяция  $P_1$  утилизирует специфический субстрат  $S_1$ , продуцируемый  $P_3$ , и сама продуцирует специфический субстрат  $S_2$ ;  $P_2$  утилизирует  $S_2$  и продуцирует  $S_3$ , который в свою очередь потребляется популяцией  $P_3$  (рис. 2). Кроме того, все популяции нуждаются в едином для всех неспецифическом субстрате  $N$ , вводимом



**Рис. 1.** Схема основных объектов и процессов, моделируемых в ЭК.

В проточной системе находятся популяции ( $P_1, P_2, \dots, P_n$ ) и поступает неспецифический субстрат  $N_1$ . Популяции потребляют специфические субстраты, которые синтезируются и секретлируются соответствующими популяциями. Например популяция  $P_1$  потребляет один специфический субстрат  $S_L$  и секретлирует два специфических субстрата:  $S_1$  (потребляется популяцией  $P_2$ ) и  $S_3$  (потребляется популяцией  $P_3$ ). Специфические субстраты могут активировать или ингибировать рост популяций, например, субстрат  $S_4$  активует рост популяции  $P_4$  и ингибирует рост  $P_2$ .

в систему извне. Рассматривались два типа ТК с разной трофической популяций. В одном случае дефицит (но не полное отсутствие) неспецифического субстрата в среде частично может быть компенсирован высокой концентрацией специфических субстратов (закон компенсации Рюбеля – ТК-Р). В другом случае отсутствие или недостаток неспецифического или специфического субстрата не могут быть компенсированы и являются критическими для популяции (закон минимума Либиха – ТК-Л).

Расчет прироста популяции ТК-Р проводился по формуле:

$$F_1(n_0, r_0, \vec{S}, \vec{C}, P) = \Delta P = \left( r_0 n_0(P) \cdot \sum_{i \in I_{consumed}} c_i s_i \right)^{1/2} - k_{flow} \cdot P - k_{death} \cdot P^2, \quad (1)$$

где  $P$  – размер популяции (количество клеток);  $n_0$  – количество неспецифического субстрата, потребленного популяцией из среды;  $r_0$  – эффективность утилизации неспецифического субстрата (значение признака, определяемое одним геном);  $\vec{S}$  – вектор значений специфических субстратов, потребленных популяцией из среды;  $\vec{C}$  – вектор скоростей утилизации соответствующих специфических субстратов (вектор значений признаков, определяемых соответствующими генами);  $I_{consumed}$  – множество индексов специфических субстратов, утилизируемых данной популяцией;  $k_{flow}$  – скорость

$$F_2(n_0, r_0, \vec{S}, \vec{C}, P) = \Delta P = P \cdot \frac{\left( \frac{n_0/P}{K_{01}(r_0)} \right)}{1 + \left( \frac{n_0/P}{K_{02}(r_0)} \right)} \cdot \prod_{i \in I_{consumed}} \frac{\left( \frac{s_i/P}{K_{i1}(c_i)} \right)^2}{1 + \left( \frac{s_i/P}{K_{i2}(c_i)} \right)^2} - k_{flow} \cdot P - k_{death} \cdot P^2, \quad (2)$$

где  $K_{01}$ ,  $K_{02}$ ,  $K_{i1}$ ,  $K_{i2}$  – константы эффективности потребления неспецифического и специфических субстратов, которые зависят от соответствующих признаков вектора  $\vec{C}$ . Остальные обозначения аналогичны формуле (1).

Параметры модели (скорость притока/оттока субстратов, смертность, эффективность потребления/продукции) и начальные данные (размеры популяций и концентрации субстратов) были подобраны так, чтобы в сублетальных условиях окружающей среды короткий период

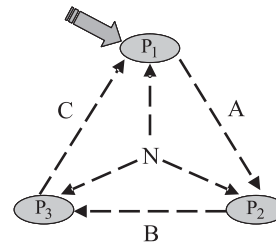


Рис. 2. Схема моделируемого ТК.

Пунктирные стрелки – активирующие субстрат-продуктные отношения; толстая стрелка – мутация в клетке популяции  $P_1$ , увеличивающая эффективность утилизации субстрата С.

протока в окружающей среде;  $k_{death}$  – коэффициент смертности.

Согласно формуле, каждая популяция ТК-Р утилизирует один неспецифический субстрат  $n_0$  и один или несколько специфических субстратов. Количество утилизируемых специфических субстратов определяется количеством положительных компонент вектора  $\vec{C}$  (с номерами из множества  $I_{consumed}$ ). Остальные компоненты вектора  $\vec{C}$  – нулевые, т. е. популяция неспособна их потреблять.

Расчет прироста популяции ТК-Л проводился по формуле:

роста численности популяций сменялся постепенным снижением численности, вплоть до полного вымирания всего ТК (в разных симуляциях – в течение 100–300 поколений). Мутация приводила к росту эффективности утилизации специфического субстрата. В результате из популяции  $P_k$  выделялась популяция  $P'_k$ , конкурентно вытеснявшая материнскую популяцию.

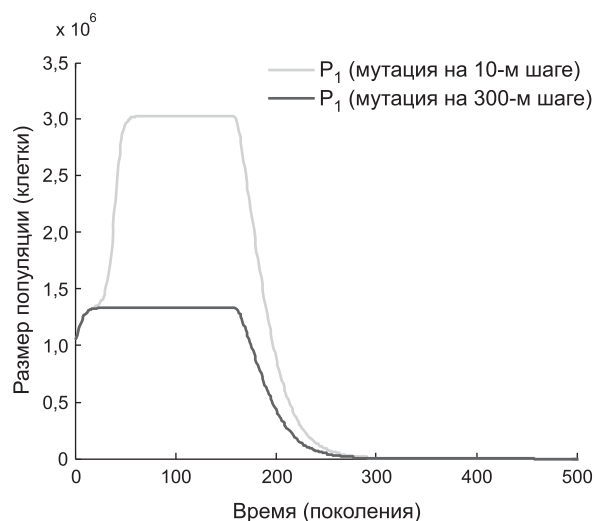
Моделирование показало, что без таких мутаций гибнут как ТК-Р, так и ТК-Л, причем

время их жизни было одного порядка. Фиксация одной из популяций ТК-Л мутации ведет к кратковременному росту численности всех популяций ТК, но не меняет предельный режим функционирования – время жизни кольца растет незначительно (рис. 3). Лидирующую по биомассе популяцию  $P'_k$  быстро догоняют популяции-соседи, так что соотношение биомасс меняется незначительно. Фиксация мутаций в ТК-Р либо на порядок увеличивает время жизни популяций ТК, либо спасает ТК от вымирания (рис. 4). Популяция  $P'_k$  обгоняет по численности своих соседей. Таким образом, ТК-Л вплоть до вымирания сохраняет биоразнообразие – популяции синхронно снижают численность и гибнут. Напротив, ТК-Р, вымирая, постепенно снижает разнообразие. Следовательно, ТК-Л потенциально может восстановить и совокупную биомассу, и биоразнообразие (например, если случайно вырастет поступление  $N$ ), а восстановление совокупной биомассы ТК-Р вовсе не значит восстановления биоразнообразия.

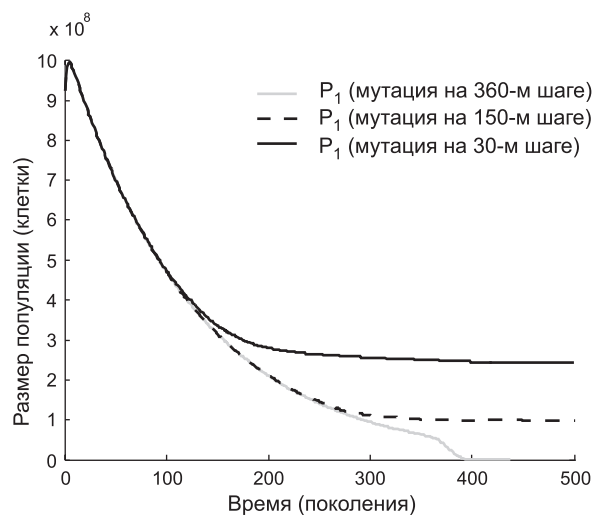
Моделирование также показало, что для ТК-Р характерно критическое соотношение численности популяции и притока  $N$ , при котором никакие мутации не спасут популяцию, а с ней и ТК от вымирания (рис. 4). При этом в трофическом кольце популяций складывается интересная ситуация: сразу после появления адаптивной мутации  $P'_k$  кольцу, чтобы выжить, необходимо поддерживать высокую численность инеадаптивной популяции  $P_k$ . Только в таком случае будет обеспечен спасительный быстрый рост численности  $P'_k$ . Когда же популяция  $P'_k$  достигнет достаточной численности, трофическому кольцу станет выгодна элиминация  $P_k$  – она становится генетическим грузом. Но быстро сбросить этот груз уже невозможно, и трофическое кольцо вымирает.

### Обсуждение

Моделирование показало существенно лучшую адаптируемость ТК-Р по сравнению с ТК-Л. Фиксация полезных мутаций даже в одной популяции ТК-Р увеличивала устойчивость всей системы, существенно продлевая время ее жизни, т. е. давая ей дополнительный шанс дождаться конца голодовки, либо вовсе спасала ТК от вымирания за счет оптимизации



**Рис. 3.** Динамика популяции  $P_1$  в 2 ТК-Л после мутаций, произошедших на 10-м и 300-м шагах соответственно. Ни одна из мутаций не смогла «спасти» популяции вне зависимости от времени возникновения.



**Рис. 4.** Динамика популяции  $P_1$  в 3 ТК-Р после мутаций, произошедших на 360-м, 150-м и 30-м шагах. Мутации на 30-м и 150-м шагах приводят к замедлению процесса вымирания (хотя тенденция все равно остается), а мутация на 360-м шаге не может «спасти» популяцию.

метаболизма. Тем не менее в мире существует множество таксонов, не способных в широких масштабах компенсировать один ресурс другим. Наиболее яркий пример – неспособность усваивать свободный азот у многих фотосинтезирующих эукариот (Анисимов и др., 1985). Эту особенность нельзя объяснить ни сложностью биохимической машины азотфиксации,

ни физиологическими запретами на создание такой машины, ни краткостью времени для эволюции. Наконец, ее нельзя объяснить и узкой специализацией эукариот-фотосинтетиков. Первыми фотосинтезирующими организмами были цианобактерии (Заварзин, 2001), они же явились предками хлоропластов у растений в ходе симбиогенеза (Степанова, 1985; Марков, Куликов, 2005). Несмотря на прокариотическую организацию и, следовательно, ограниченные размеры генома, цианобактерии сумели освоить азотфиксацию (хотя для этого в ряде случаев им пришлось пожертвовать способностью гетероцист к размножению) (Meeks *et al.*, 2001). Напротив, растения, имеющие несравненно большие геномы, утратили эту возможность, хотя впоследствии ряд таксонов независимо приобрел ее вторично, совершенно иным путем, вновь пойдя на симбиоз с прокариотами (Schultze, Kondorosi, 1998). Также можно привести незначительный процент видов животных, симбиогенно освоивших фотосинтез, хотя возможность компенсировать недостаток пищи автотрофией имеет очевидное адаптивное преимущество и независимо развивается у многих групп беспозвоночных (Назаров, 1981). Жесткая конкуренция с растениями вряд ли могла ограничить экспансию таких видов хотя бы потому, что кишечнополостные, освоившие такой симбиоз, успешно формируют коралловые рифы, занимающие огромные площади (Шепард, 1987). В периоды глобальных биосферных кризисов интенсивность конкуренции временно снижается. Несмотря на геологическую кратковременность таких периодов, времени оказывается достаточно, чтобы угнетенные таксоны, сформировав собственные экосистемы, добились доминирующего положения в биосфере (Каландадзе, Раутиан, 1993) (что и произошло с современными кораллами – рифовые экосистемы, первые из которых создали губки в кембрии, надолго исчезали из геологической летописи после каждого великого вымирания, чтобы затем постепенно возрождаться на базе других таксонов – мшанок, рудистов и, наконец, кораллов (Михайлова и др., 1989)).

Палеонтологические данные свидетельствуют о том, что первые автотрофные биоценозы появились более ~ 2 млрд лет назад и в их основе лежали организмы-универсалы – ци-

анобактерии. Это определило консерватизм цианобактериальных матов, которые дожили до нашего времени практически без изменений в трофической структуре, а также таксономическом (по крайней мере, на уровне порядка) составе (Заварзин, 2001). Однако около 850 млн лет назад цианобактериальные маты были существенно потеснены сообществами водорослевых лугов, чьими продуцентами были менее универсальные эукариоты. Эукариотические экосистемы отличались на порядок более высоким биоразнообразием (Закруткин, 1993; Федонкин, 2000). Аналогичная ситуация повторилась при освоении суши (Бурзин, 1998; Каратыгин, 2005). Почему же основным трендом эволюции жизни на Земле стало увеличение биоразнообразия за счет прогрессивной специализации и, как следствие – смены доминирующих таксонов, а не формирование биоты на базе немногих таксонов-универсалов?

Известно, что в смене условий на Земле прослеживается определенная циклика. Ее источником служат внешние эндогенные (геофизические) и экзогенные (астрономические) циклические процессы (Дажо, 1976; Dobretsov *et al.*, 2008). В результате поступление вещества и энергии в экосистему меняется. Чтобы сохранить общее значение этого потока постоянным, экосистема должна вести эволюционный поиск новых источников вещества и энергии (Левченко, 2003). В этом смысл принципа максимального давления жизни, согласно которому любой таксон, стремясь к экологической экспансии, постоянно увеличивает свое давление на среду (Вернадский, 1987). Таким образом, критическим параметром становится «способность к обучению» экосистемы. Ее можно оптимизировать, либо увеличив исходное биоразнообразие экосистемы, либо удлинив ее время жизни в условиях голодания.

Через определенный период времени внешняя циклика восстановит исходное значение потока вещества и энергии. Если этот период сравним с временем жизни долгоживущих экосистем (в нашей модели это ТК-Р), то при их низком разнообразии новые источники энергии и вещества, вполне возможно, так и не будут найдены. Напротив, в экосистемах, сохраняющих высокое биоразнообразие вплоть до вымирания, вероятность найти такой источник выше.



Тогда при восстановлении исходного значения потока новый источник прибавится к нему. Общее значение потока повысится (Левченко, 1993, 2003), предоставив экосистеме ресурсы для дальнейшей прогрессивной эволюции (II биогеохимический принцип-постулат В.И. Вернадского – эволюция идет в направлении роста интенсивности биогенной миграции атомов биосферы) (Вернадский, 1987). Таким образом, в долговременной перспективе у Либиховских систем появляется преимущество перед Рюбелевскими. Следует сразу отметить, что преимущество это носит, как и все эволюционные процессы, недетерминистский, вероятностный характер, тогда как преимущество по устойчивости детерминировано. Следовательно, по чисто стохастическим причинам даже при высоком биоразнообразии система может погибнуть. Это согласуется с постоянной сменой ведущих биот в ходе эволюции (Михайлова и др., 1989).

### Благодарности

Работа частично поддержана грантом РФФИ № 06-04-49556; проектами № 10.7 по программе РАН «Молекулярная и клеточная биология» и № 18.13 по программе РАН «Происхождение и эволюция биосферы»; госконтрактом № 10104-37/П-18/110-327/180608/015 с ПИН РАН по подпрограмме 2 программы фундаментальных исследований Президиума РАН «Происхождение и эволюция биосферы» и грантом НШ-2447.2008.4. Научная школа Н.А. Колчанова «Биоинформатика и системная компьютерная биология».

### Литература

- Анисимов А.А., Олонина Л.Н., Лукина Е.В. Эволюция систем минерального питания растений // Эволюция функций в растительном мире. Л.: ЛГУ, 1985. С. 109–122.
- Беклемишев Н.В. О классификации биоценологических (симфизиологических) связей // Биоценологические основы паразитологии. М.: Наука, 1970. С. 90–138.
- Беляев Д.К. Дестабилизирующий отбор как фактор изменчивости при доместикации животных // Природа. 1979. № 2. С. 36–45.
- Бигон М., Харпер Дж., Таунсенд К. Экология. Особи, популяции и сообщества. М.: Мир, 1989. Т. 1. 667 с.; Т. 2. 477 с.
- Бурзин М.Б. Докембрийские предтечи «пионеров суши» // Природа. 1998. № 3. С. 83–95.
- Вернадский В.И. Химическое строение биосферы Земли и ее окружения. 2 изд., М.: Наука, 1987. 339 с.
- Гаузе Г.Ф. Проблема стабилизирующего отбора // Журн. общ. биологии. 1941. Т. II. № 2. С. 193–209.
- Горячева И.И. Бактерии рода *Wolbachia* – репродуктивные паразиты членистоногих // Усп. соврем. биологии. 2004. Т. 124. № 3. С. 246–259.
- Гунбин К.В., Суслов В.В., Колчанов Н.А. Молекулярно-генетические системы развития: динамика функционирования и молекулярная эволюция // Биохимия. 2007. Т. 73. № 2. С. 219–230.
- Дажо А. Основы экологии. М.: Прогресс, 1976. 450 с.
- Жерихин В.В. Эволюционная биоценология. Проблема выбора моделей // Экосистемные перестройки и эволюция биосферы. М.: Недра, 1994. С. 13–20.
- Заварзин Г.А. Становление биосферы // Вестн. РАН. 2001. Т. 71. № 11. С. 988–1001.
- Заварзин Г.А. Несовместимые признаки в системе родов бактерий // Журн. общ. биологии. 1973. Т. 34. № 4. С. 530–538.
- Закруткин В.Е. О масштабах накопления органического вещества в докембрии и фанерозое // Проблемы доантропогенной эволюции биосферы. М.: Наука, 1993. С. 202–212.
- Иорданский Н.Н. Макроэволюция: макрогенез и типогенез // Журн. общ. биологии. 2004. Т. 65. № 451. С. 451–463.
- Каландадзе Н.Н., Раутиан А.С. Симптоматика экологических кризисов // Стратиграфия. Геологическая корреляция. 1993. Т. 1. № 5. С. 3–8.
- Камшилов М.М. Эволюция биосферы. 2 изд., М.: Наука, 1974. 256 с.
- Каратыгин И.В. Грибы как компоненты экосистем прошлого // Ботан. журнал. 2005. Т. 90. № 9. С. 1297–1318.
- Красилов В.А. Нерешенные проблемы теории эволюции. Владивосток: ДВНЦ АН СССР, 1986. 138 с.
- Лашин С.А., Суслов В.В., Лихошвай В.А. и др. Моделирование эволюции сообществ: программа «Эволюционный конструктор» // Системная компьютерная биология. Новосибирск: СО РАН, 2008. С. 662–676.
- Левченко В.Ф. Модели в теории биологической эволюции. СПб.: Наука, 1993. 383 с.
- Левченко В.Ф. Эволюция биосферы до и после появления человека. СПб.: Ин-т эволюционной физиологии и биохимии РАН, 2003. 166 с.
- Любарский Г.Ю. Метод общей типологии в биологических исследованиях. II. гипотетико-дедуктивный метод // Журн. общ. биологии. 1993.

- Т. 54. № 5. С. 515–531.
- Маргалеф Р. Облик биосферы. М.: Наука, 1992. 214 с.
- Марков А.В., Куликов А.М. Происхождение эвкарриот: выводы из анализа белковых гомологий в трех надцарствах живой природы // Палеонтол. журнал. 2005. Т. 30. № 4. С. 345–357.
- Михайлова И.А., Бондаренко О.Б., Обручева О.П. Общая палеонтология. М.: МГУ, 1989. 334 с.
- Назаров В. За порогом вражды. М.: Мысль, 1981. 240 с.
- Одум Ю. Экология. М.: Прогресс, 1986. Т. 1/2. 620 с.
- Рожнов С.В. Закон гомологических рядов Н.И. Вавилова и архаическое многообразие по данным палеонтологии // Эволюция биосферы и биоразнообразия. М.: КМК, 2006. С. 134–147.
- Степанова А.М. Симбиотическая гипотеза происхождения хлоропластов и митохондрий // Эволюция функций в растительном мире. Л.: ЛГУ, 1985. С. 98–109.
- Супотницкий М.В. Микроорганизмы, токсины и эпидемии. 2-е изд. М.: Вузовская книга, 2005. 376 с.
- Суховольский В.Г. Экономика живого. Новосибирск: Наука, 2004. 140 с.
- Федонкин М.А. Холодная заря животной жизни // Природа. 2000. № 9. С. 3–11.
- Черданцев В.Г. Морфогенез и эволюция. М.: КМК, 2003. 360 с.
- Шеппард Ч. Жизнь кораллового рифа. Л.: Гидрометеиздат, 1987. 183 с.
- Шмальгаузен И.И. Факторы эволюции (теория стабилизирующего отбора). М.: Наука, 1968. 452 с.
- Юрлова Н.И. Влияние паразитирования трематод на репродуктивный потенциал природной популяции *Lymnaea stagnalis* (Gastropoda, Lymnaeidae) // Зоол. журнал. 2003. Т. 120. № 9. С. 1027–1037.
- Bakker T.C.M., Mazzi D., Zola S. Parasite induced changes in behavior and color make *Gammarus pulex* more prone to fish predation // Ecology. 1997. V. 78. P. 1098–1104.
- Bendoricchio G., Jørgensen S.E. Exergy as goal function of ecosystems dynamic // Ecological Modelling. 1997. V. 102. P. 5–15.
- Dobretsov N.L., Kolchanov N.A., Suslov V.V. On important stages of Geosphere and Biosphere evolution // Biosphere origin and evolution. Springer Science+Business Media, 2008. P. 3–24.
- Fog A. Simulation of evolution in structured populations: the open source software package altruist // Biotech Software & Internet Report. 2000. V. 1. № 5. P. 226–229.
- Hansen E.K., Poulin R. Impact of a microphallid trematode on the behaviour and survival of its isopod intermediate host: phylogenetic inheritance? // Parasitol. Res. 2005. V. 97. № 3. P. 242–246.
- Huxley J.S. The Evolutionary Process // Evolution as a Process / Eds J.S. Huxley, A.C. Hardy, E.B. Ford. London: G. Allen & Unwin, 1954. P. 1–23.
- Huxley J.S. Evolution. The modern synthesis. 2-nd ed. London: Publisher, NY.: Harper & Brothers, 1943. 645 p.
- Jørgensen S.E. State-of-the-art of ecological modelling with emphasis on development of structural dynamic models // Ecological Modelling. 1999. V. 120. P. 75–96.
- Jørgensen S.E. State-of-the-art of ecological modeling in limnology // Ecological Modelling. 1995. V. 78. P. 101–115.
- Lashin S.A., Suslov V.V., Kolchanov N.A., Matushkin Yu.G. Simulation of coevolution in community by using the «Evolutionary Constructor» program // In Silico Biol. 2007. V. 7. P. 261–275.
- Meeks J.C., Elhai J., Potts M. *et al.* An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium // Photosyn. Res. 2001. V. 70. № 1. P. 85–106.
- Rubel E. The replaceability of ecological factors and the law of the minimum // Ecology. 1935. V. 16. № 3. P. 336–341.
- Schultze M., Kondorosi A. Regulation of symbiotic root nodule development // Annu. Rev. Genet. 1998. V. 32. P. 33–57.
- Vasilyeva L.N. Systematics in mycology // Bibl. Mycol. 1999. Bd. 178. S. 1–253.

**EVOLUTION MODELING OF TROPHIC WEB-LIKE COMMUNITIES  
WITH COMPENSATORY AND NON-COMPENSATORY METABOLISM**

**S.A. Lashin<sup>1,2</sup>, V.V. Suslov<sup>1</sup>, Yu.G. Matushkin<sup>1,2</sup>**

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: lashin@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

**Summary**

Comparative modeling of populations' adaptivity and biodiversity dynamics has been carried out. Populations consist of symbiotic unicellular haploid asexual organisms inhabiting a biotope of limited volume in which the nonspecific substrate (NS) inflows externally. The organisms consume NS, and it is necessary for every of them. Symbiosis is provided by intra-populational exchange of specific substrates. Using the original software package «Evolutionary Constructor» allowing varying both genetic and trophic structure of a model, we have shown that under sub-lethal conditions (NS deficiency) the populations with compensatory metabolism (NS deficiency can be compensated by symbiont coevolution) have better adaptivity (mutation in a separate population can save all symbiont populations) but gradually lose biodiversity, whereas the populations with non-compensatory metabolism keep biodiversity down to their extinction. The critical balance limit between the population size and the NS flow was obtained. Wherein it is outreached, the substitutional load of normal allele by the mutant one can't be hold. Advantages and disadvantages of two evolutionary strategies – the strategy of biota formation on the basis of several universal taxons and the strategy of high biodiversity conservation are discussed.

## TOPOLOGICAL INDEX OF THE *P53-MDM2* CIRCUIT

V.P. Golubyatnikov<sup>1</sup>, E. Mjolsness<sup>2</sup>, Yu.A. Gaidov<sup>3</sup>

<sup>1</sup>Institute of Mathematics, SB RAS, Novosibirsk, Russia, e-mails: glbtn@math.nsc.ru; emj@uci.edu gja22@ngs.ru; <sup>2</sup>University of California, Irvine, CA 92697-3435, USA;

<sup>3</sup>State Pedagogical University, Novosibirsk, Russia

Analysis of oscillating regimes in regulatory networks models with multistability properties and with switch-oscillator mechanisms is an important task both from the biological and from the mathematical viewpoints. Unification of discrete and continuous approaches in gene network modeling should be useful for understanding the functioning of the regulation mechanisms of natural and artificial biological systems. Our investigations of the nonlinear dynamical systems as models of the gene networks are based on topological methods elaborated in (Golubyatnikov *et al.*, 2006; Gaidov, Golubyatnikov, 2007). Here we study the phase portrait of a system of nonlinear differential equations proposed in (Chikarmane *et al.*, 2007), see also (Lahav *et al.*, 2004) as a model of oscillations in the *p53-Mdm2* DNA damage repair network. We construct a domain  $Q$  in the phase space such that the topological index of the velocity vector field in  $Q$  is nontrivial in presence of the DNA damage and vanishes for small amounts of this damage. This demonstrates a direct relationship between the index of the corresponding velocity vector field and the presence of the DNA damage in this model and exhibits a topological approach to the unification of discrete and continuous gene network modeling.

**Key words:** gene network models, stationary points, topological index, oscillations.

### Introduction

We continue our studies (Golubyatnikov *et al.*, 2006; Likhoshvai *et al.*, 2008) of kinetic dynamical systems serving as gene network models. Here, our main aim is to describe multistability in the regulatory network of the DNA damage response system as modeled in (Chikarmane *et al.*, 2007), and to outline important connections of studies of discrete and continuous gene network models. A short exposition of our considerations was presented in (Golubyatnikov, Mjolsness, 2008).

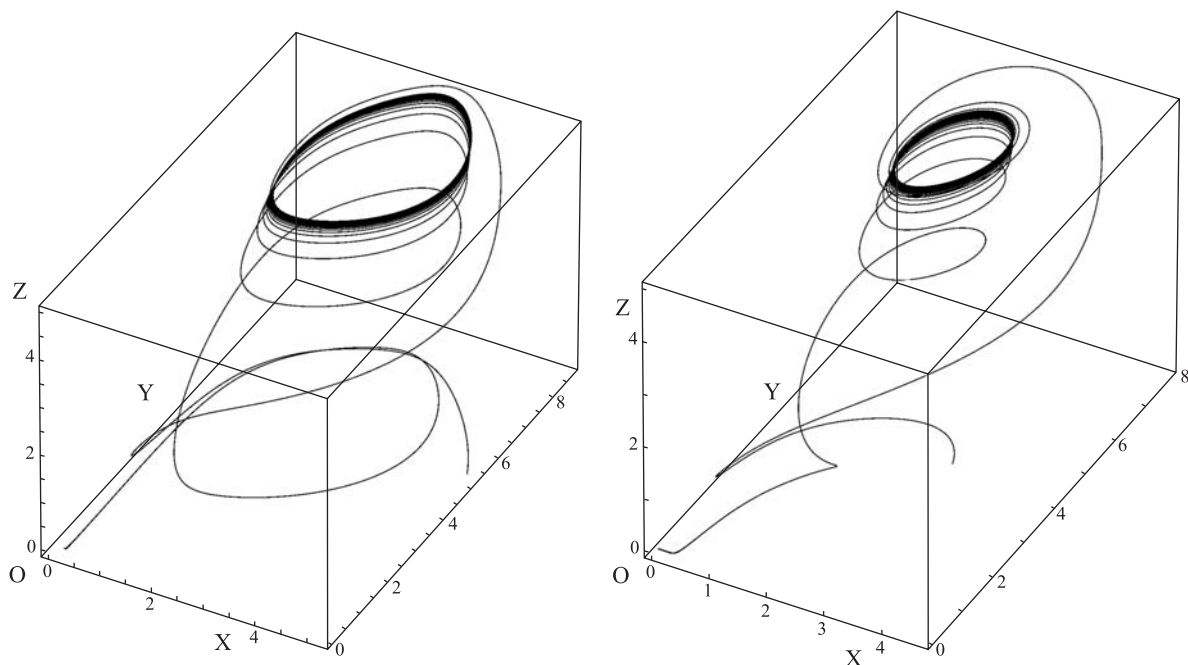
The following kinetic system (Chikarmane *et al.*, 2007) represents a model of oscillations in the *p53-Mdm2* circuit. This model contains positive and negative feedback loops in its regulation mechanism. Its phase space is the positive octant  $R^4_+$  in the space of the variables  $\{x, y, z, D\}$ . Here the positive feedback is related to autocatalysis in *p53*, which activates *Mdm2* as well, and the negative one corresponds to inhibition of *p53* by *Mdm2*.

$$\begin{aligned} \frac{dx}{dt} &= \Phi_1(x, y, z) = \alpha_0 z + \frac{\alpha_1 \cdot x^6}{k_1 + x^6} - \gamma_1 x y - \gamma_2 x; \\ \frac{dy}{dt} &= \Phi_2(x, y) = \alpha_2 + \frac{\alpha_3 \cdot x^4}{k_2 + x^4} - \gamma_3 y; \\ \frac{dz}{dt} &= z \cdot \Phi_3(z, D) = \\ &= \frac{\alpha_{1s} \cdot z \cdot (B - z)}{k_{1s} + B - z} - \frac{\alpha_{2s} \cdot z}{(k_{0d} + D)(k_{2s} + z)}; \\ \frac{dD}{dt} &= -\alpha_d \cdot D \cdot z x. \end{aligned} \quad (1)$$

Most of the variables here denote dimensionless concentrations of corresponding species:  $x=[p53]$ ,  $y=[Mdm2]$ ,  $z=[Atm-P]$ ;  $z$  is the switch variable.  $D(t)$  denotes the amount of the DNA damage in the network,  $B = [Atm] + [Atm-P]$  appears from the conservation law.

### Results

1. For each fixed  $t_0$  consider intersection of 3-D plane  $(x, y, z, D(t_0))$  with the positive octant  $R^4_+$ .



**Fig. 1.** Behavior of trajectories of the system (1).

$B = 5, \varepsilon = 4.4$ . Initial data:  $z_0 = 0, 1; D_0 = 10$ . Left  $n = 5; x_0 = 5; y_0 = 5$ . Right  $n = 4; x_0 = 4; y_0 = 3$ .

We call the points where  $\frac{dx}{dt} = \frac{dy}{dt} = \frac{dz}{dt} = 0$  **3-stationary** and consider a domain  $Q = [0, C_1] \times [0, C_2] \times [\varepsilon, B] \subset R^3(x, y, z)$  with the boundary  $\partial Q$ . Here  $\gamma_2 \cdot C_1 \geq \alpha_0 \cdot B + \alpha_1, \gamma_3 \cdot C_2 \geq \alpha_2 + \alpha_3$ , and  $\varepsilon$  is positive. For small  $t$  all trajectories of (1) of the points of  $\partial Q$  enter  $Q$ . According to (1),  $\Phi_3(B, D) < 0$  and  $\frac{dD}{dt}$  is strictly negative for all  $t$  and positive  $x$  and  $z$ . If  $\Phi_3(0, 0) > 0$ , then there exists unique  $z_1 \in (0, B)$  such that  $\Phi_3(z_1, D) = 0$  for all  $D$ .  $\Phi_2(x, y) = 0; \Phi_1(x, y, z_1) = 0$  imply that the summed indices of 3-stationary points in  $Q$  coincides with the signed number of roots  $x$  of the equation

$$\alpha_0 \cdot z_1 + \frac{\alpha_1 \cdot x^6}{k_1 + x^6} - \frac{\gamma_1 \cdot x}{\gamma_3} \left( \alpha_2 + \frac{\alpha_3 \cdot x^4}{k_2 + x^4} \right) - \gamma_2 \cdot x = 0.$$

Direct calculations show that for small values of  $t$  the sum of 3-dimensional indices of all 3-stationary points in  $Q$  equals -1. This corresponds to large values of the DNA damage. For large  $t$  and any  $\varepsilon \geq 0$  this index vanishes. In this case the DNA damage is small.

The values of all parameters in our numerical experiments were taken close to ones listed in (Chikarmane *et al.*, 2007). It was indicated there that for some intermediate value  $D(t)$  during the

repair process, the variable  $x$  passes through an Andronov–Hopf bifurcation.

Fig. 1 above shows typical behavior of trajectories of the system (1). Here for small values of  $z_0$  and large  $D_0$  trajectories jump to the domain  $4.4 < z < 5$  and remain there till  $D > 0.58$ .

2. We have seen that the geometric properties of the graphs of the right hand sides of the equations (1), such as spectra of the Jacobian matrices evaluated at the stationary points, conditions of the type  $\Phi_3(0, 0) > 0$ , and the slopes of these graphs, are much more important for bifurcations and other dynamical characteristics of these models than is the concrete analytic representation of  $\Phi_1, \Phi_2, \Phi_3$ . Similar observations were made in (Golubyatnikov *et al.*, 2006; Gaidov, Golubyatnikov, 2007) for other models of asymmetric gene networks. General constructions related to the Conley index, see (Salamon, 1990), can be used in the cases of more complicated gene network models.

### Conclusion and Discussions

Our analysis of the *p53-Mdm2* oscillatory dynamics is connected with decomposition of high-dimensional models to lower-dimensional ones.



This procedure can be considered as a discretization of a continuous model. Similar decompositions can be used in mathematical modeling of more complicated regulatory networks. The direct relation of the topological index of the repair system (1) with the presence of the DNA damage demonstrates a new approach to gene networks studies. This approach is based on the comparison of properties of continuous models and discrete models of the same gene network. Such a dualism can yield new possibilities in the study of problems in system biology.

### Acknowledgements

The work was partially supported by the leading scientific schools grant 8526.2006.1, RFBR grant 09-01-00070, SB RAS grant 119, and by US NIH grant P50-GM76516. Authors are indebted to E.G. Balahnina and I.V. Stepanova for assistance and to E.P. Volokitin, T.D. Kolesnikova and V.A. Likhoshvai for discussions.

### References

- Chikarmane V., Ray A., Sauro H.M., Nadim A. A model for p53 triggered by DNA damage // *SIAM J. Appl. Dynamic Systems*. 2007. V. 6. № 1. P. 61–76.
- Gaidov Yu.A., Golubyatnikov V.P. On some nonlinear dynamical systems modeling asymmetric gene networks // *Bul. of Novosibirsk State University*. 2007. V. 7. № 2. P. 23–32. (In Russ.).
- Golubyatnikov V.P., Gaidov Yu.A., Kleshchev A.G., Volokitin E.P. Modeling of asymmetric gene networks functioning with different types of regulation // *Biophysics*. 2006. V. 51. Suppl. 1. P. 61–65.
- Golubyatnikov V.P., Mjolsness E. Topological index of a model of p53 dynamics triggered by DNA damage // *Abstracts of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008)*. Novosibirsk, 23–28 June 2008. Novosibirsk: Inst. Cytol. Genet., 2008. P. 89.
- Lahav G., Rosenfeld N., Sigal A. *et al.* Dynamics of the p53-Mdm2 feedback loop in individual cell // *Natural Genet*. 2004. V. 36. № 1. P. 147–150.
- Likhoshvai V.A., Golubyatnikov V.P., Demidenko G.V. *et al.* Theory of Gene Networks // *Computational Systems Biology, (Russian) / Eds N.A. Kolchanov, S.S. Goncharov, V.A. Likhoshvai, V.A. Ivanisenko*. Novosibirsk: SB RAS. 2008. P. 397–480.
- Salamon D.A. Morse theory, the Conley index and Floer homology // *Bul. of Lond. Math. Soc.* 1990. V. 22. № 1. P. 113–140.

## СИСТЕМА АВТОМАТИЗИРОВАННОЙ ГЕНЕРАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ГЕННЫХ СЕТЕЙ

Ф.В. Казанцев<sup>1</sup>, И.Р. Акбердин<sup>1</sup>, К.Д. Безматерных<sup>1</sup>, В.А. Лихошвай<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: kazfdr@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Одной из актуальных проблем системной биологии является изучение законов функционирования генных сетей как динамических систем, обеспечивающих формирование фенотипических, морфологических и иных характеристик живых организмов. Для решения этой проблемы необходимо разрабатывать средства компьютерного анализа динамики функционирования генных сетей с учетом их структурно-функциональной организации. В работе мы представляем систему MGSgenerator, которая, опираясь на информацию о строении генной сети, автоматически генерирует математическую модель в терминах биохимических реакций и обобщенных функций Хилла. Входной информацией для представленной системы служит база данных GeneNet. Готовая модель может быть экспортирована в форматы SiBML, SBML и Step+ для проведения численного анализа.

**Ключевые слова:** генные сети, математические модели, автоматическая генерация, обобщенные функции Хилла.

### Введение

Одним из центральных вопросов системной биологии является выяснение молекулярно-генетических механизмов функционирования живых систем. Для изучения их динамических характеристик все шире применяются методы математического и компьютерного моделирования. Математические модели строятся на основании информации о структурно-функциональной организации генных сетей (ГС) и их динамических свойствах, которая сосредоточена по сотням и тысячам научных публикаций. В связи с этим получили развитие компьютерные технологии, позволяющие накапливать экспериментальные данные в специализированных базах данных: KEGG (Kanehisa, Goto, 2000), Ecosys (Keseler *et al.*, 2005), Sabio-RK (Wittig *et al.*, 2006), Biomodels (Novère *et al.*, 2006). Это позволило стандартизировать хранение и представление многих типов экспериментальных данных и разрабатывать методы их системного анализа. Одной из таких компьютерных технологий является система GeneNet,

ориентированная на реконструкцию генных сетей (Ananko *et al.*, 2005). Данная система позволяет аккумулировать информацию из научных публикаций о структурно-функциональной организации молекулярно-генетических систем и представлять ее пользователю в форме графических диаграмм. Представленная в GeneNet информация отражает статическую часть сведений о генных сетях. Однако генные сети являются динамическими системами. Их состояние меняется с течением времени, и эти изменения во многом определяются структурно-функциональной организацией генных сетей. Эффективным инструментом для установления взаимосвязей между структурно-функциональной организацией генных сетей и их динамическими свойствами является метод математического моделирования. Разработка систем автоматической генерации математических моделей на основе информации, хранящейся в базах по структурно-функциональной организации генных сетей, является актуальной проблемой системной биологии. В 2008 г. Draeger с соавторами разработали одну из

первых таких систем – SBMLsqueezer (Draeger *et al.*, 2008). Данная система осуществляет автоматическую генерацию математических моделей генных сетей, реконструированных в среде CellDesigner (Funahashi *et al.*, 2003), используя стандарт SBML. Текущая версия стандарта SBML не позволяет хранить полную информацию о структурно-функциональной организации ГС (тип объекта, вид регуляторного взаимодействия и др.), что является необходимым условием для автоматического построения адекватной математической модели. В результате SBMLsqueezer настроен на работу только с ГС, реконструированными в системе CellDesigner.

В настоящей работе мы представляем программу MGSgenerator, которая автоматизирует процесс построения модели в терминах обыкновенных дифференциальных уравнений на основе исходной (первичной) информации о структурно-функциональной организации генной сети, хранящейся в базе данных GeneNet. Программа MGSgenerator анализирует элементарные подсистемы генной сети, генерирует их математические модели и на их основе с учетом структурной организации генной сети строит общую математическую модель. Модель может быть экспортирована в форматы SiBML (Лихошвай и др., 2001), SBML (Hucka *et al.*, 2003) и Step+ (Fadeev *et al.*, 2006) для проведения численного анализа.

### Материалы и методы

Конструирование моделей проводится на основе информации о строении генных сетей, накапливаемой в базе данных GeneNet (<http://www.mgs.bionet.nsc.ru/mgs/gnw>). Входная информация подается в виде XML-файла, автоматически генерируемого для каждой генной сети средствами GeneNet. Для конструирования моделей используется представление в GeneNet генных сетей в виде совокупности элементарных событий (подсистем), связанных между собой входами и выходами. Элементарные подсистемы в GeneNet описывают молекулярно-генетические (транскрипция, сплайсинг, транспорт, трансляция и т. д.) и биохимические (ферментативные реакции, фосфорилирова-

ние, формирование комплексов белков и т. д.) процессы, протекающие в различных компартаментах организма. Для идентификации модели, соответствующей элементарному событию в генной сети, поданной на вход, производится анализ наборов входных и выходных реагентов, а также наборов регуляторных взаимодействий, которые описаны в стандарте базы GeneNet.

Математические модели генных сетей генерируются на основе обобщенного химико-кинетического подхода (Лихошвай и др., 2001).

На первом шаге каждой элементарной подсистеме  $S_p$ ,  $I = 1, \dots, n$  ставится в соответствие собственная математическая модель – система обыкновенных дифференциальных уравнений, которая описывает закон изменения концентраций управляемых веществ в зависимости от текущих концентраций управляющих веществ и параметров системы:

$$\frac{dX_i}{dt} = V_i(Y_i, K_i), \quad (1)$$

где  $X_i$  – вектор управляемых веществ,  $Y_i$  – вектор управляющих веществ,  $K_i$  – вектор параметров.

Вид функций  $V_i$  конструируется в терминах обобщенных функций Хилла на основе информации о структурно-функциональной организации элементарной подсистемы, представленной в системе GeneNet. Обобщенные функции Хилла являются естественным обобщением стационарных биохимических моделей, поэтому структура ОФХ отражает механизм, а параметры – кинетические характеристики участников описываемого процесса. Более того, метод ОФХ эффективен при недостатке кинетических экспериментальных данных и может быть применен при построении математических моделей молекулярно-генетических систем разных уровней детализации (Likhoshvai, Ratushny, 2007). Конструирование функций  $V_i$  проводится на основе оригинального алгоритма, разработанного в рамках настоящей работы.

Генерация модели  $M$  из моделей элементарных подсистем осуществляется по правилу суммирования скоростей протекания элементарных процессов, которое формулируется исходя из биохимической природы молекулярно-генетических процессов (Лихошвай и др., 2001):

$$\begin{array}{c}
 S_1 : \frac{dX_1}{dt} = V_1(Y_1, K_1), \quad S_2 : \frac{dX_2}{dt} = V_2(Y_2, K_2), \quad \dots \quad S_n : \frac{dX_n}{dt} = V_n(Y_n, K_n) \\
 \swarrow \quad \downarrow \quad \dots \quad \searrow \\
 M : \begin{cases} \frac{dX}{dt} = V(Y, K), & V(Y, K) = \sum_{i=1}^n V_i(Y_i, K_i), \\ X = \cup X_i, Y = \cup Y_i, K = \cup K_i. \end{cases}
 \end{array}$$

Рис. 1. Правило суммирования скоростей по всем непрерывным моделям элементарных подсистем  $S_i$ .

### Результаты и обсуждение

Представленная в настоящей работе программа MGSgenerator использует в качестве входных данных структурно-функциональную организацию генных сетей, реконструированных в системе GeneNet. В настоящей версии программы алгоритм автоматической генерации математической модели реализован в интерактивном варианте, так как на определенных этапах требуется принятие экспертного решения. Связано это с тем, что алгоритм реализован для ограниченного набора элементарных подсистем, который не исчерпывает возможного их разнообразия. Каждую такую подсистему программа автоматически распознает в составе генной сети и генерирует для нее математическую модель. Правила генерации описаны в следующих разделах статьи. Если в генной сети встречается подсистема, для которой построение элементарной модели невозможно, то выдается сообщение об этом. В этом случае пользователь может либо модифицировать структуру генной сети и подать ее на вход программы MGSgenerator повторно, либо вручную добавить в конструируемую модель математическую модель элементарной подсистемы.

**Сопоставление математических моделей элементарным подсистемам.** В MGSgenerator генная сеть рассматривается как совокупность взаимодействующих элементарных подсистем  $S_1, \dots, S_n$ . В представленной версии программы элементарными подсистемами являются элементарные события (реакции), описанные в стандарте базы данных GeneNet. Сопоставле-

ние определенной математической модели для каждой выделенной элементарной подсистемы осуществляется на основе анализа наборов входных и выходных реагентов и наборов управляющих воздействий. Данная информация экстрагируется из стандарта представления реакций базы данных GeneNet, и на ее основе определяются тип подсистемы, а также структура модели.

В текущей версии программы MGSgenerator генерация математических моделей поддерживается для следующих типов элементарных подсистем ГС: репликация, транскрипция, обратная транскрипция, трансляция, ферментативный синтез, транспорт, мультимеризация, сплайсинг, фосфорилирование, ацетилирование и метилирование, а также подсистем притока веществ и их деградации.

Для генерации математических моделей элементарных подсистем всех типов используется формула общего вида, состоящая из 4 сомножителей:

$$\begin{aligned}
 V_i(Y_i, K_i) &= \\
 &= k \cdot V_{i1}(Y_{i1}) \cdot V_{i2}(Y_{i2}, K_{i2}) \cdot V_{i3}(Y_{i3}, K_{i3}) \\
 Y_i &= Y_{i1} \cup Y_{i2} \cup Y_{i3} \\
 K_i &= \{k\} \cup K_{i2} \cup K_{i3}.
 \end{aligned} \tag{2}$$

Первый сомножитель является константой скорости синтеза и имеет размерность, обратную времени. Второй сомножитель  $V_{i1}$  соответствует входному веществу, которое определяет тип подсистемы и имеет размерность концентрации. Сомножитель  $V_{i1}$  может принимать одно из двух значений, в зависимости от типа подсистемы:

$$V_{i1}(Y_{i1}) = \begin{cases} 1, & Y_{i1} = \emptyset, \text{ если процесс – приток вещества,} \\ y, & Y_{i1} = \{y\}, \text{ иначе.} \end{cases} \tag{3}$$

Здесь  $y$  – концентрация ключевого вещества подсистемы. В случае транскрипции или репликации – это ген, в случае трансляции или сплайсинга – РНК, а в случае ферментативного синтеза – фермент. Третий ( $V_{i2}$ ) и четвертый ( $V_{i3}$ ) сомножители являются безразмерными. Сомножитель  $V_{i2}$  описывает влияние регуляторов на скорость протекания рассматриваемого процесса. В слу-

чае отсутствия последних множитель  $V_{i2}$  полагается равным 1. Если же список регуляторов не пуст, то вид данного сомножителя зависит от их типа. Различаются регуляторы, выполняющие функцию активаторов ( $a_j$ ), и регуляторы, выполняющие функцию ингибиторов ( $r_l$ ). В соответствии с конкретной спецификацией выписывается формула вида:

$$V_{i2}(Y_{i2}, K_{i2}) = \left( \delta + \sum_j \left( \frac{a_j}{k_{1,a,j}} \right)^{h_{1,a,j}} \right) / \left( 1 + \sum_j \left( \frac{a_j}{k_{2,a,j}} \right)^{h_{2,a,j}} + \sum_l \left( \frac{r_l}{k_{r,l}} \right)^{h_{r,l}} \right)$$

$$Y_{i2} = (\{a_j\}_{j=1,\dots,n}, \{r_l\}_{l=1,\dots,m}) \quad (4)$$

$$K_{i2} = (\delta, \{k_{1,a,j}, k_{2,a,j}, h_{1,a,j}, h_{2,a,j}\}_{j=1,\dots,n}, \{k_{r,l}, h_{r,l}\}_{l=1,\dots,m}),$$

где  $k_{1a}$ ,  $k_{2a}$  – константы активатора,  $k_r$  – константа ингибитора,  $h_{1a}$ ,  $h_{2a}$ ,  $h_r$  – коэффициенты Хилла.

Множитель  $V_{i3}$  описывает зависимость скорости реакции от концентраций субстратов и продуктов. В случае таких подсистем, как репликация, транскрипция, трансляция, сплайсинг или транспорт, полагается, что  $V_{i3} = 1$ .

Для подсистем, имеющих тип «ферментативный синтез», «фосфорилирование», «ацетилирование» и «метилирование», вид функции  $V_{i3}$  зависит от концентрации субстратов и продуктов и определяется структурой входного

и выходного списков. Тип распознается по наличию во входном списке сущности «фермент», который осуществляет выбранный процесс. Для подсистем фосфорилирования, метилирования, ацетилирования (в текущей версии системы эти процессы объединены в один класс и не различаются друг от друга, так как нет достаточной информации во входных данных для идентификации) входной список содержит сущность «белок» ( $s$ ), который подвергается модификации. Сущность, которая отвечает модифицированной форме белка, входит в выходной список ( $p$ ). В этом случае сомножитель  $V_{i3}$  имеет вид:

$$V_{i3}(Y_{i3}, K_{i3}) = \frac{k_1 \cdot x \cdot \frac{s}{k_{ms}} - k_{-1} \cdot x \cdot \frac{p}{k_{mp}}}{1 + \frac{s}{k_{ms}} + \frac{p}{k_{mp}}} \quad (5)$$

$$Y_{i3} = \{p, s, x\}; K_{i3} = \{k_1, k_{-1}, k_{ms}, k_{mp}\},$$

где  $k_1$  – константа прямой реакции,  $k_{-1}$  – константа обратной реакции,  $k_{ms}$  и  $k_{mp}$  – константы Михаэлиса. Для подсистем данного типа по умолчанию полагается, что субстрат и продукт конкурируют между собой за активный сайт фермента.

Для подсистем типа «ферментативный синтез» рассматривается несколько вариантов. Пусть входной список содержит  $n$  субстратов ( $s_j, j = 1, \dots, n$ ), из которых синтезируется один продукт, который содержится в выходном списке. Тогда  $V_{i3}$  имеет вид:

$$V_{i3}(Y_{i3}, K_{i3}) \equiv \frac{\prod_{j=1}^n \frac{s_j}{k_{ms,j}} - \omega_{p,s} \frac{p}{k_{mp}}}{\prod_{j=1}^n \left( 1 + \frac{s_j}{k_{ms,j}} \right) + \frac{p}{k_{mp}}} \quad (6)$$

$$Y_{i3} = \{p, s_j, j = 1, \dots, n\}; K_{i3} = \{\omega_{p,s}, k_{mp}, k_{ms,j}, j = 1, \dots, n\},$$



где  $k_{ms}$  и  $k_{mp}$  – константы Михаэлиса,  $\omega_{p,s}$  – отношение константы прямой реакции к константе обратной реакции. Для подсистем данного подтипа по умолчанию полагается, что субстраты могут независимо связываться с фер-

ментом, но при этом конкурируют с продуктом.

Пусть входной список содержит  $n$  субстратов ( $s_j, j = 1, \dots, n$ ), из которых синтезируется  $n$  продуктов ( $p_j, j = 1, \dots, n$ ). Тогда применяется формула:

$$V_{i3}(Y_{i3}, K_{i3}) = \frac{k_s \cdot x \cdot \prod_j \frac{s_j}{k_{ms,j}} - k_p \cdot x \cdot \prod_j \frac{p_j}{k_{mp,j}}}{\prod_i \left( 1 + \frac{s_j}{k_{ms,j}} + \frac{p_j}{k_{mp,j}} \right)} \quad (7)$$

$$Y_{i3} = \{p_j, s_j, x, j = 1, \dots, n\}; K_{i3} = \{k_s, k_p, k_{ms,j}, k_{mp,j}, j = 1, \dots, n\},$$

где  $k_s$  – константа прямой реакции,  $k_p$  – константа обратной реакции,  $k_{ms}$  и  $k_{mp}$  – константы Михаэлиса. В формуле (7) реализовано предположение, что  $j$ -й субстрат конкурирует с  $j$ -м продуктом за связывание с ферментом, а пары субстрат–продукт не конкурируют между собой.

**Пример конструирования модели генной сети.** Рассмотрим генную сеть, содержащую ген *sup79b3*, кодирующий транскрипционный фактор, который является автоингибитором инициации транскрипции (рис. 2).

Данная генная сеть содержит 4 элементарные подсистемы: подсистему синтеза и транспорта мРНК из ядра в цитоплазму, подсистему синтеза и доставки белка из цитоплазмы в ядро, подсистемы деградации мРНК и белка (подсистемы деградации не отображаются в

структуре генной сети). В результате разбора информации система MGSgenerator генерирует математическую модель, содержащую два дифференциальных уравнения:

$$\begin{cases} \frac{dr}{dt} = k_r \cdot g \cdot \frac{k_{ba}}{1 + \left(\frac{p}{k_i}\right)^{k_g}} - k_{dr} \cdot r \\ \frac{dp}{dt} = k_p \cdot r - k_{dp} \cdot p \end{cases}, \quad (8)$$

где  $g$  – концентрация гена *sup79b3*,  $r$  – концентрация мРНК *sup79b3*,  $p$  – концентрация белка *Sup79b3*,  $k_r$  и  $k_p$  – константы активации,  $k_{ba}$  – константа базальной активности,  $k_{dr}$  и  $k_{dp}$  – константы деградации,  $k_i$  – константа ингибирования,  $k_g$  – Хилловский коэффициент.

Программа MGSgenerator была использована для генерации математической модели, контролирующей метаболизм ауксина в меристеме побега *Arabidopsis thaliana* L., и показала свою эффективность при решении поставленных теоретических задач (Акбердин и др., 2009).

**Расширяемость программы MGSgenerator.** Программа MGSgenerator реализована на языке программирования Java, что обеспечивает ее переносимость на различные платформы. Программа поддерживает две точки расширения функциональности:  $G$  – точку подключения источников данных и  $\Sigma$  – точку подключения модулей экспорта готовой математической модели, что позволяет сформировать модульную компьютерную систему моделирования и численного анализа динамики функционирования генных сетей (рис. 3).

В качестве источника данных в представленной компьютерной модульной системе для

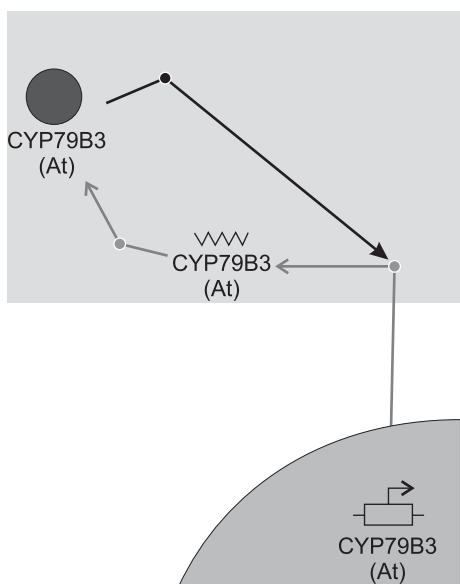


Рис. 2. Структура генной сети процесса автоингибирования.

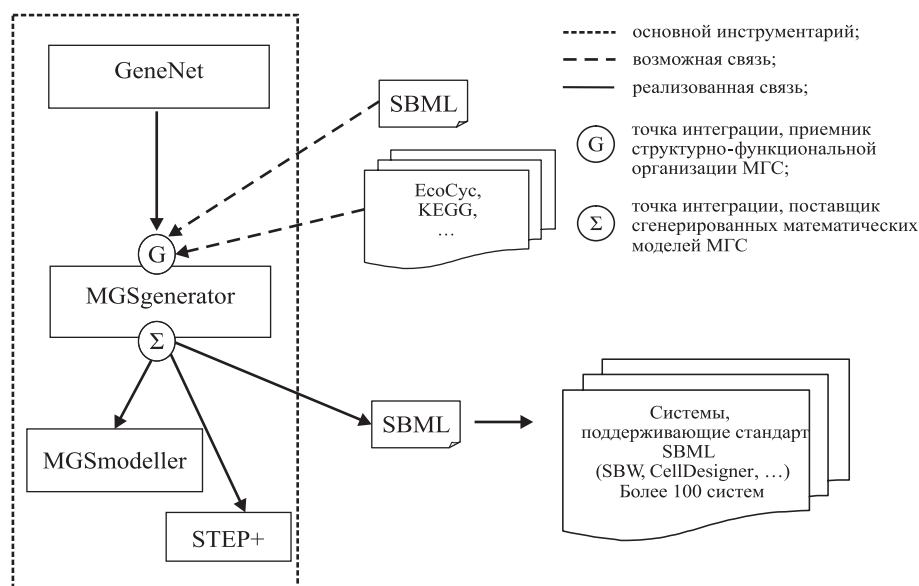


Рис. 3. Схема организации компьютерной модульной системы моделирования на основе программы MGSgenerator.

настоящей версии программы MGSgenerator служит база GeneNet, которая предоставляет информацию о структурно-функциональной организации ГС. В будущем функциональность системы может быть расширена другими источниками данных, в частности, подключением баз KEGG и ЭкоСус. В качестве форматов представления математической модели в настоящей версии программы MGSgenerator используются: формат SiBML системы MGSmodeller (Kazantsev *et al.*, 2008), формат системы STEP+ (Fadeev *et al.*, 2006) и формат SBML (Hucka *et al.*, 2003). В случае необходимости функциональность по представлению математической модели также может быть расширена.

### Заключение

Представленная в работе программа MGSgenerator является ключевым элементом при разработке эффективных компьютерных модульных систем, которые в перспективе способны обеспечить полный технологический цикл изучения закономерностей строения и функционирования генных сетей с применением методов математического моделирования. Предложенная версия генератора математических моделей генных сетей является начальным этапом в разработке автоматизированных систем генерации математических моделей. Включенные в нее математи-

ческие модели покрывают лишь незначительную часть возможных механизмов. В дальнейшем планируется существенно расширить базу шаблонных моделей элементарных подсистем, а также возможности представления математических моделей ГС в других форматах.

### Литература

- Акбердин И.Р., Казанцев Ф.В., Омелянчук Н.А., Лихошвай В.А. Математическое моделирование метаболизма ауксина в клетке меристемы побега растения // Информ. вестник ВОГиС. 2009. Т. 13. № 1. С. 170–175.
- Лихошвай В.А., Матушкин Ю.Г., Ратушный А.В. и др. Обобщенный химико-кинетический метод моделирования генных сетей // Молекуляр. биология. 2001. Т. 3. № 6. С. 1072–1079.
- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. V. 33. P. D425–D427.
- Draeger A., Hassis N., Supper J. *et al.* SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks // BMC Systems Biol. 2008. V. 2.
- Fadeev S.I., Korolev V.K., Gainova I.A., Medvedev A.E. The package Step+ for numerical study of autonomous systems arising when modeling dynamics of genetic-molecular systems // Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2006). Novosibirsk, 16–22 July 2006. Novosibirsk: Inst. Cytol. Genet., 2006. V. 2. P. 118–120.

- Funahashi A., Morohashi M., Kitano M., Tanimura N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks // BIOSILICO. 2003. V. 1. P. 159–162.
- Hucka M., Finney A., Sauro H. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models // Bioinformatics. 2003. V. 19. P. 524.
- Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D. *et al.* MGSmodeller – a computer system for reconstruction, calculation and analysis of mathematical models of molecular genetic system // Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008). Novosibirsk, 22–28 June 2008. Novosibirsk: Inst. Cytol. Genet., 2008. P. 113.
- Kanehisa M., Goto S. KEGG: Kyoto encyclopedia of genes and genomes // Nucl. Acids Res. 2000. V. 28. P. 27–30.
- Keseler I., Collado-Vides J., Gama-Castro S. *et al.* EcoCyc: a comprehensive database resource for *Escherichia coli* // Nucl. Acids Res. Database issue. 2005. V. 33. P. D334–D337.
- Likhoshvai V., Ratushny A. Generalized Hill function method for modeling molecular processes // J. Bioinformatics and Comp. Biol. 2007. V. 5. № 2. P. 521–531.
- Novère N., Bornstein B., Broicher A. *et al.* BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems // Nucl. Acids Res. 2006. № 34. P. D689–D691.
- Wittig U., Golebiewski M., Kania R. *et al.* SABIO-RK: Integration and Curation of Reaction Kinetics Data // Proc. of the 3rd Intern. Workshop on Data Integration in the Life Sciences 2006 (DILS'06). Hinxton, UK. Lecture Notes in Bioinformatics. P. 94–103.

## THE TOOL FOR AUTOMATIC GENERATION OF GENE NETWORKS MATHEMATICAL MODELS

F.V. Kazantsev<sup>1</sup>, I.R. Akberdin<sup>1</sup>, K.D. Bezmaternykh<sup>1</sup>, V.A. Likhoshvai<sup>1,2</sup>

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: kazfdr@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

### Summary

Nowadays there are a plenty of reconstructed gene networks that describe different systems in live organism critical for forming phenotypical, morphological and others properties of individuals. The analysis of gene networks in terms of dynamical systems is one of the key challenges of system biology. To facilitate this analysis we have proposed MGSgenerator system that provides automatic generation of mathematical models on the basis of gene networks structure. The mathematical models are generated in biochemical and generalized Hill function terms. The GeneNet database is used as data source of gene network structures. The generated model could be exported in SiBML, SBML and Step+ formats for further analysis.

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МЕТАБОЛИЗМА АУКСИНА В КЛЕТКЕ МЕРИСТЕМЫ ПОБЕГА РАСТЕНИЯ

И.Р. Акбердин<sup>1</sup>, Ф.В. Казанцев<sup>1</sup>, Н.А. Омелянчук<sup>1</sup>, В.А. Лихошвай<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: akberdin@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Разработана структурная модель генной сети метаболизма ауксина в клетке растения с учётом генетической регуляции (<http://www.mgs.bionet.nsc.ru/systems/mgl/genenet/>). На основе структурной модели и литературных данных разработана база математических моделей 235 элементарных подсистем: ферментативных реакций метаболизма ауксина и генетической регуляции экспрессии кодирующих их генов. Проведен компьютерный анализ динамики режимов нормального и аномального функционирования системы регуляции метаболизма ауксина; продемонстрирована адекватность теоретических расчетов к экспериментальным данным.

**Ключевые слова:** ауксин, математическая модель, компьютерный анализ, конъюгация, *Arabidopsis thaliana* L.

### Введение

Гормон ауксин (класс низкомолекулярных соединений, относящихся к производным индола) является одним из важнейших веществ, участвующих в регуляции различных биологических процессов растений, особенно их морфогенеза. Координация ауксином процессов роста и развития растений осуществляется через контролируемое формирование паттернов распределения ауксина и дозозависимых механизмов его действия. Изучение регуляции, действия и взаимодействий ауксина с другими фитогормонами и рецепторами является актуальной проблемой биологии развития растений. Применение технологий математического моделирования позволяет провести теоретический анализ движения и распределения паттернов ауксина в развивающейся ткани; исследовать *in silico* альтернативные гипотезы относительно роли ауксина как одного из сигналов развития и сравнить их с данными, полученными *in vivo*; проанализировать влияние и роль мутаций, связанных с изменением функционирования системы метаболизма и транспорта фитогормона, регулирующих рост и развитие растений.

В настоящей работе реконструируется математическая модель и проводится теоретический анализ механизмов регуляции метаболизма ауксина при развитии меристемы побега *Arabidopsis thaliana* L.

### Материалы и методы

Для реконструкции генной сети внутриклеточной регуляции метаболизма ауксина использованы база данных и модуль визуализации генных сетей *GeneNet* (Ananko *et al.*, 2005), разработанные в Институте цитологии и генетики СО РАН.

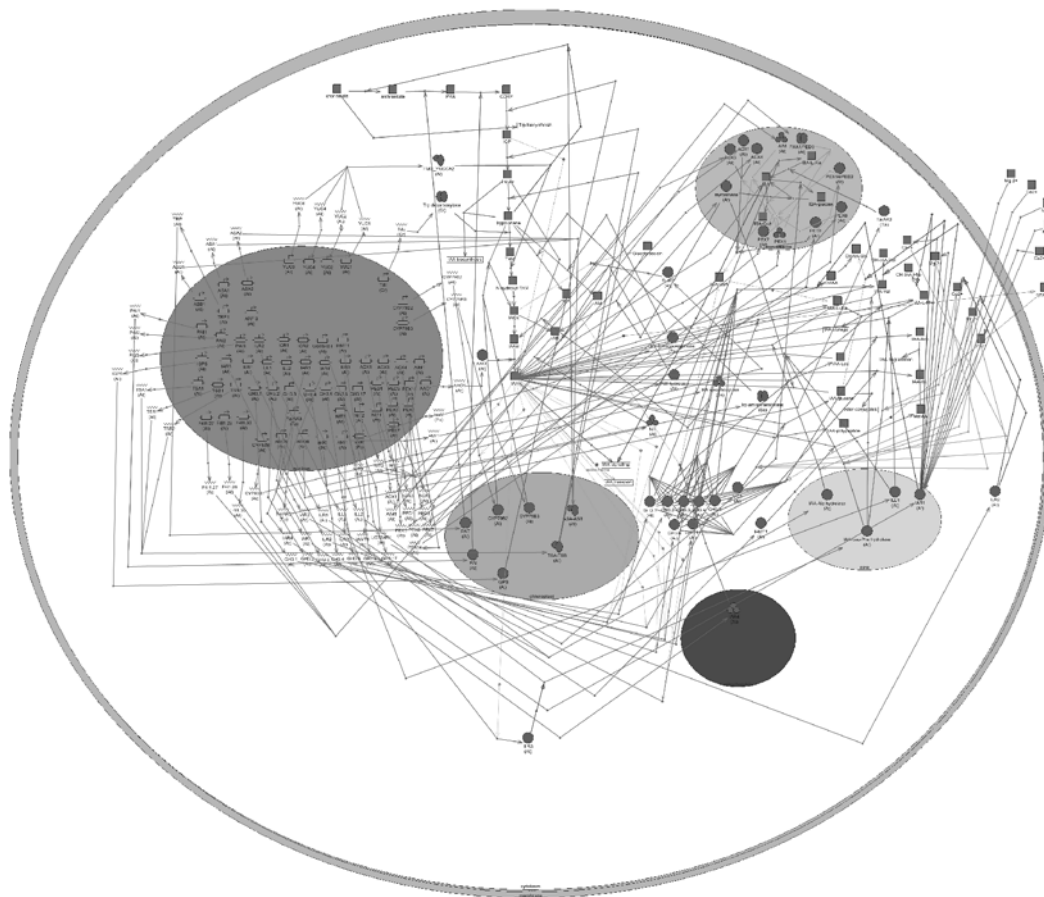
Автоматическая генерация базы элементарных моделей молекулярно-генетических систем осуществлялась на основе программного модуля MGSgenerator (Kazantsev *et al.*, 2008a), который является промежуточным звеном в процессе генерации математических моделей. Его цель – построение математической модели по имеющемуся структурно-функциональному описанию генной сети (ГС) и некоторой дополнительной информации. Модуль имеет собственный стандарт организации данных и программный интерфейс (API), с помощью

которого возможно разрабатывать модули генерации математических моделей для различных источников, описывающих ГС. Используемая в исследовании версия генератора полностью построена по технологии «Plug-in». Такая организация позволяет минимальными усилиями наращивать и изменять функциональность существующей системы и интегрировать ее с новыми. Для реконструкции и анализа комплексной математической модели генной сети метаболизма ауксина использовалась компьютерная система MGSmodeller (Kazantsev *et al.*, 2008b), которая разработана на языке Java. Данная система содержит следующие модули: «конструктор/редактор моделей», «расчет моделей», «обратная задача», «оптимальное управление». Многофункциональность системы основывается на возможностях каждого модуля. Модуль

«конструктор/редактор моделей» позволяет создавать и редактировать модели, используя оригинальный стандарт спецификации моделей «SiBML». Пользовательский интерфейс системы MGSmodeller позволяет иерархически отображать данные, редактировать их и наглядно отображать результаты расчета и анализа математических моделей.

### Результаты и обсуждение

**Разработка и адаптация модели.** На основе разработанной в Институте цитологии и генетики СО РАН базы данных и модуля визуализации генных сетей *GeneNet* (Ananko *et al.*, 2005) реконструирована структурная модель генной сети внутриклеточной регуляции метаболизма ауксина (рис. 1). Для реконструкции исполь-



**Рис. 1.** Структурная модель генной сети метаболизма ауксина. Структура генной сети реконструирована на основе 107 литературных источников и содержит 62 гена (□), 56 мРНК (∞) и 44 белка (●). Она включает 245 молекулярно-генетических процессов, распределенных по 8 компартментам: цитоплазма, клеточная мембрана, ядро, пероксисома, эндоплазматический ретикулум, митохондрия, хлоропласт и внеклеточное пространство.



зованы данные из 107 научных публикаций. В табл. 1 приведена статистика компонентов структурной модели генной сети метаболизма ауксина.

**Таблица 1**  
Компоненты генной сети  
метаболизма ауксина\*

Название	Количество
Белок	44
РНК	56
Ген	62
Компартменты	8
Реакция	245
Метаболит	41

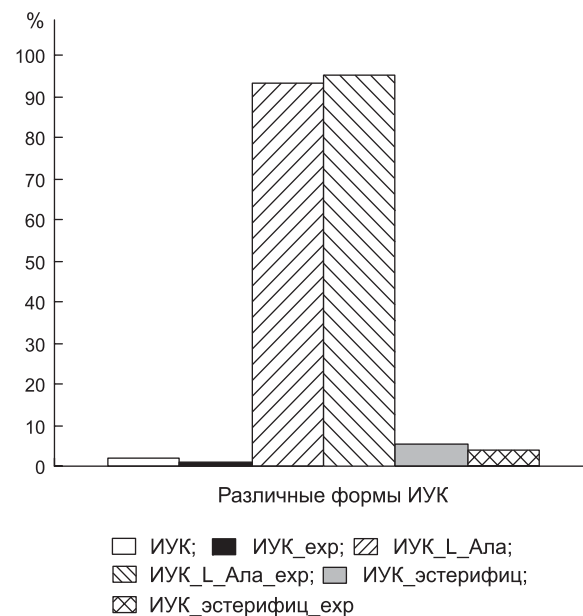
\* Комментарии к таблице в тексте.

Структурная модель описывает основной триптофан-зависимый путь биосинтеза ауксина и триптофан-независимый путь его биосинтеза из индола, а также включает описание основных путей конъюгации и деградации ауксина в клетках меристемы побега и корня. Основной путь биосинтеза ауксина содержит две отрицательные обратные связи: 1) триптофан аллостерически ингибирует ферментативную реакцию превращения хоризмата в антранилаат (предшественники триптофана); 2) ауксин активирует транскрипцию мРНК GH3 белков, которые являются ферментами реакций конъюгации ауксина.

Метаболизм ауксина распределен по многим компартментам клетки (рис. 1). Это позволяет предположить, что молекулярно-генетическая система гомеостаза ауксина в клетке растения сформировалась до образования некоторых органелл клетки. В результате эволюционных процессов происходило усложнение и компартментное распределение метаболических реакций. Данная гипотеза частично подтверждается экспериментальными наблюдениями (Cooke *et al.*, 2002).

Математическая модель регуляции внутриклеточного метаболизма ауксина *Arabidopsis thaliana*, автоматически реконструированная на основе структурной модели генной сети (рис. 1) с использованием технологии, представленной в разделе «Материалы и методы»,

описывает внутриклеточный уровень реализации морфогенетических механизмов регуляции развития АМП (апикальная меристема побега). Реконструированная математическая модель содержит порядка двухсот параметров. Значения большинства параметров модели взяты из литературных источников для *A. thaliana*, кукурузы и риса. Оставшаяся часть оптимальных значений параметров была подобрана численно. В результате было достигнуто соответствие модели экспериментальным данным относительно содержания различных форм ауксина в клетке (рис. 2). В экспериментах *in silico* так же, как и в экспериментах *in vivo* в клетках побега арабидопсиса, свободная форма ауксина и его эстрифицированная форма составляют около 1 % и меньше 4 % от общего пула ауксина соответственно. Оставшийся пул ауксина (95 %) составляют его амидные формы (Park *et al.*, 2001). Одним из свойств рассматриваемой молекулярно-генетической системы (МГС) является ее способность к быстрому увеличению концентрации ауксина в клетке растения при внешних стрессовых условиях. Вероятно, за эту способность как раз и отвечает пул амидных



**Рис. 2** Результаты моделирования процентного содержания различных форм ауксина в клетке меристемы побега.

По оси y – проценты, по оси x – форма ИУК (индол-3-уксусная кислота, или ауксин); эстр – процентное содержание формы ауксина из экспериментальных данных.

форм ауксина, из которого происходит быстрое высвобождение ауксина в результате реакций деконъюгации. С другой стороны, низкий процент (всего 1 %) внутриклеточного содержания свободной формы ауксина в нормальных условиях предполагает ключевую роль сигнальных путей и транспорта данного фитогормона в формировании наблюдаемых паттернов роста и развития всего растения.

**Теоретический анализ молекулярно-генетических механизмов гомеостаза ауксина в клетке.** В данной молекулярно-генетической системе особую роль в поддержании гомеостаза внутриклеточного содержания ауксина занимают механизмы регуляции реакций конъюгации. Как уже говорилось выше, именно пул конъюгированных форм ауксина определяет изменения концентрации ауксина в клетке при внешних воздействиях. Ферменты, которые осуществляют реакции конъюгации ИУК (индол-3-уксусной кислоты) с аминокислотами, кодируются белками семейства GH3, которые в свою очередь кодируются ауксин-индуцируемыми генами. В результате накапливание GH3 транскриптов индуцируется ауксином (Hagen *et al.*, 1984). Помимо этого, существует огромное число видов растений, которые имеют GH3 гены, реагирующие на присутствие ауксина. По крайней мере, некоторые ИУК-индуцируемые гены GH3 кодируют ферменты, осуществляющие реакции конъюгации ИУК в ИУК-аминокислоту (Staswick *et al.*, 2005). Таким образом, индукция ауксином GH3 генов, вероятно, обеспечивает ослабление сигнала ауксина инактивацией ИУК через конъюгацию.

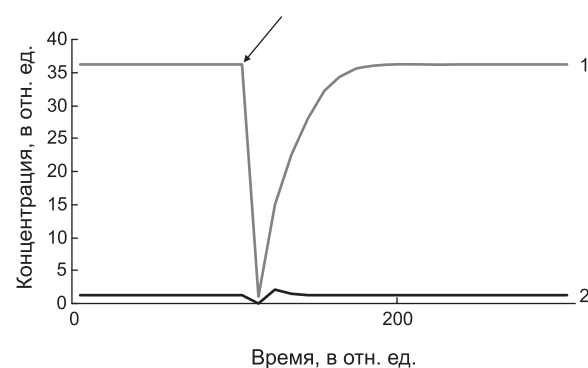
Эксперименты *in silico* по индукции ауксином GH3 генов проводили с помощью следующей процедуры: в системе, находящейся в стационарном состоянии, изменяли значения концентрации ИУК и GH3 белка и рассчитывали значения этих концентраций вновь до стационарной фазы. На рис. 3 представлен один из экспериментов *in silico*. Как видно из рис. 3, происходит качественное повышение концентрации GH3 белка в клетке, который в свою очередь опосредует уменьшение концентрации ауксина. Аналогичные графики распределения получены при различных начальных значениях.

Из полученных результатов можно сделать вывод, что независимо от начальных значений

концентраций ауксина и продукта GH3 транскриптов в заданном диапазоне относительных единиц концентрации происходит активация ауксином наработки продукта GH3 транскриптов, который ослабляет сигнал ауксина инактивацией ИУК через конъюгацию.

Несомненно, что на поддержание гомеостаза внутриклеточного содержания ауксина влияют механизмы, регулирующие ферментативные реакции метаболического пути биосинтеза ИУК. Мутанты арабидопсиса *trp3-1* и *trp2-1* имеют нарушения в триптофансинтазах  $\alpha$  и  $\beta$  соответственно (Last *et al.*, 1991; Radwanski *et al.*, 1996), которые в свою очередь осуществляют реакции превращения триптофана и его предшественников в индол. Мутанты данных типов накапливают амидные формы ауксина (Normanly *et al.*, 1993; Ouyang *et al.*, 2000), несмотря на низкий уровень триптофана (Muller, Weiler, 2000; Ouyang *et al.*, 2000). Однако триптофансинтаза  $\alpha$  способна производить индол без участия  $\beta$ -субъединицы. Индол в свою очередь в последующем преобразуется в определенные «защитные» вещества (Frey *et al.*, 1997, 2000; Melanson *et al.*, 1997) или в ауксин.

Мутантные фенотипы *trp3-1* и *trp2-1* так же, как и фенотип двойного мутанта *trp3-1 trp2-1*, были имитированы в экспериментах *in silico* путем уменьшения значений констант скорости синтеза мРНК и их продуктов триптофансинтаз  $\alpha$  и  $\beta$  в 100 000 раз по сравнению с исходными. В табл. 2 суммированы результаты расчета рав-



**Рис. 3.** Результаты моделирования динамики изменения концентрации ИУК и GH3 белка в норме.

По оси Y – концентрация в отн. ед., по оси X – время расчета в отн. ед. Стрелкой указан момент перевода системы из стационарного состояния в состояние с измененными концентрациями GH3 белка и ИУК (1 и 0,01 соответственно); 1 – динамика изменения концентрации GH3 белка, 2 – динамика изменения концентрации ИУК.

**Таблица 2**  
Концентрации ИУК-аланина (ИУК\_Ала)  
в норме и при мутациях (*trp2-1*, *trp3-1*)

Фенотип	<i>trp2-1</i>	<i>trp2-1 trp3-1</i>
Концентрация ИУК_Ала (в процентном содержании от нормы)	102,45	103,67

новесных концентраций ИУК-аланина (ИУК\_Ала) в норме и при мутациях (*trp2-1*, *trp3-1*).

Из представленных в табл. 2 результатов можно сделать вывод, что несмотря на резкое уменьшение скорости наработки ферментов, осуществляющих реакции превращения промежуточных веществ в ИУК, в одиночном и двойном мутантах наблюдается заметное увеличение стационарной концентрации конъюгатной формы ауксина, что качественно соответствует экспериментальным данным. Эксперименты *in silico* подтверждают, что незначительное изменение стационарной концентрации ИУК\_Ала в мутантах по сравнению с нормой принципиально связано со стратегией хранения различных форм ауксина, компенсирующих нарушения основного пути биосинтеза ИУК у высших растений (Sztein *et al.*, 2000). В экспериментах *in silico* также выявляется более значимый вклад в изменение содержания амидной формы ауксина, вносимый нарушением *trp3-1* в суммарный вклад двойного мутанта *trp3-1 trp2-1*, что соответствует экспериментальным представлениям о важности триптофансинтазы  $\alpha$ , которая может функционировать в данной МГС без  $\beta$ -субъединицы (Park *et al.*, 2001).

Математическая модель регуляции внутриклеточного метаболизма ауксина *A. thaliana*, реконструированная на основе структурной модели соответствующей геномной сети, отражает внутриклеточный иерархический уровень морфогенетических механизмов регуляции развития АМП. Полученные результаты подтверждают адекватность разработанной математической модели МГС регуляции гомеостаза ауксина внутри клетки. Проведённые эксперименты *in silico* свидетельствуют о качественно быстрых процессах молекулярно-генетической регуляции поддержания гомеостаза исследуемой системы в ответ на изменение концентрации ауксина в клетке.

Ранее нами была разработана математическая модель развития меристемы побега *A. thaliana* в терминах клеточного автомата, которая является пространственно-распределенной моделью (Akberdin *et al.*, 2007). Несмотря на «простоту» принципов (типы клеток, параметры клеток и т. д.), которые являются основой пространственно-распределенной модели, клеточный автомат позволяет описать ряд феноменологических экспериментальных наблюдений и предоставляет возможности изучения мутантных фенотипов *in silico*. С другой стороны, модель метаболизма ауксина в клетке меристемы побега растения описывает внутриклеточный иерархический уровень развития растения в целом и меристемы побега в частности. Следующим этапом в применении методов математического моделирования к исследованию законов роста и развития меристемы побега *A. thaliana* будет объединение пространственно-распределенной модели и модели внутриклеточного метаболизма ауксина. Результаты этого этапа позволят исследовать причинно-следственные связи между молекулярно-генетическими событиями в клетках растения и природой морфологических изменений в структуре и функционировании отдельных органов и всего растения.

### Благодарности

Авторы выражают благодарность К.Д. Безматерных, В.В. Мироновой, И.А. Гайновой и С.И. Фадееву за обсуждение и консультации в ходе данной работы. Работа поддержана грантом НШ-2447.2008.4.

### Литература

- Akberdin I.R., Ozonov E.A., Mironova V.V. *et al.* A cellular automaton to model the development of primary shoot meristems of *Arabidopsis thaliana* // J. of Bioinformatics and Computational Biol. 2007. V. 5. 02B. P. 641–650.
- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. V. 1. № 33. Database Issue: P. 425–427.
- Cooke T.J., Poli D., Sztein A.E., Cohen J. Evolutionary patterns in auxin action // Plant Mol. Biol. 2002. V. 49. P. 319–338.
- Frey M., Chomet P., Glawischnig E. *et al.* Analysis of a chemical plant defense mechanism in grasses // Science. 1997. V. 277. P. 696–699.

- Frey M., Stettner C., Parer P.W. *et al.* An herbivore elicitor activates the gene for indole emission in maize // Proc. Natl Acad. Sci. USA. 2000. V. 97. P. 14801–14806.
- Hagen G., Kleinschmidt A., Guilfoyle T. Auxin-regulated gene expression in intact soybean hypocotyl and excised hypocotyl sections // Planta. 1984. V. 162. P. 147–153.
- Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D., Likhoshvai V.A. MGSgenerator – the tool for automatic generation of molecular genetic system mathematical models on basis of gene networks structure // Abstr. of the Sixth Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008), Novosibirsk, Russia, June 22–28, 2008a. P. 114.
- Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D. *et al.* A computer system for reconstruction, calculation and analysis of mathematical models of molecular genetic system // Abstr. of the Sixth Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008). Novosibirsk, Russia, June 22–28, 2008b. P. 113.
- Last R.L., Bissinger P.H., Mahoney D.J. *et al.* Tryptophan mutants in Arabidopsis: The consequences of duplicated tryptophan synthase b genes // The Plant Cell. 1991. 3. P. 345–358.
- Melanson D., Chilton M.-D., Masters-Moore D., Chilton W.S. A deletion in an indole synthase gene is responsible for the DIMBO Adeficient phenotype of bxbx maize // Proc. Natl Acad. Sci. USA. 1997. V. 94. P. 13345–13350.
- Muller A., Weiler E.W. Indolic constituents and indole-3-acetic acid biosynthesis in the wild-type and a tryptophan auxotroph mutant of *Arabidopsis thaliana* // Planta. 2000a. 211. P. 855–863.
- Normanly J., Cohen J.D., Fink G.R. *Arabidopsis thaliana* auxotrophs reveal a tryptophan-independent biosynthetic pathway for indole-3-acetic acid // Proc. Natl Acad. Sci. USA. 1993. V. 90. P. 10355–10359.
- Ouyang J., Shao X., Li J. Indole-3-glycerol phosphate, a branchpoint of indole-3-acetic acid biosynthesis from the tryptophan biosynthetic pathway in *Arabidopsis thaliana* // Plant J. 2000. V. 24. P. 327–333.
- Park S., Walz A., Momonoki Y.S., Slovin J.P. *et al.* Partial characterization of major amide-linked conjugates of IAA in Arabidopsis seed (Abstr. #321) // Final Program July 2001, Amer. Soc. of Plant Biologists, Canad. Soc. of Plant Physiol. meeting, Providence, Rhode Island, 2001. P. 81–82.
- Radwanski E.R., Barczak A.J., Last R.L. Characterization of tryptophan synthase alpha subunit mutants of *Arabidopsis thaliana* // Mol. General Genet. 1996. V. 253. P. 353–361.
- Staswick P.E., Serban B., Rowe M. *et al.* Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid // The Plant Cell. 2005. V. 17. P. 616–627.
- Sztein A.E., Cohen J.D., Cooke T.J. Evolutionary patterns in the auxin metabolism of green plants // Int. J. Plant Sci. 2000. V. 161(6). P. 849–859.

## MATHEMATICAL MODEL OF AUXIN METABOLISM IN MERISTEM CELLS OF PLANT SHOOTS

I.R. Akberdin<sup>1</sup>, F.V. Kazantsev<sup>1</sup>, N.A. Omelyanchuk<sup>1</sup>, V.A. Likhoshvai<sup>1,2</sup>

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: akberdin@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

### Summary

Indole-3-acetic acid (IAA) is physiologically active in the form of the free acid, but can also be found in conjugated forms in plant tissues. IAA can be degraded and redundant pathways lead to its synthesis. Auxin participates in regulation of cell differentiation in development of embryo, leaves, vascular tissue, fruit, primary and lateral root and in controlling apical dominance and tropisms. The regulation of the IAA metabolism (synthesis, conjugation and degradations) is an enough complex process and may explain in some aspects how this simple substance is able to influence such diverse processes. Mathematical modeling of IAA metabolic gene network can help reveal the main factors governing this complex process. To reach this aim, we first reconstructed a gene network of auxin biosynthesis, conjugation degradation by annotating experimental data from 107 published papers into GeneNet computer system. This gene network after reduction was input into converter to generate the mathematical model of auxin metabolism. We have reconstructed the gene network and develop the mathematical model of auxin metabolism in Arabidopsis shoots. The model allows to reproduce some phenomenological and molecular-genetic aspects of the auxin role in the plant development.



## МОДЕЛИРОВАНИЕ РЕГУЛЯЦИИ АУКСИНОМ ИНИЦИАЦИИ ЛАТЕРАЛЬНЫХ ОРГАНОВ У *ARABIDOPSIS THALIANA* L.

В.А. Лихошвай<sup>1,2</sup>, Н.А. Омелянчук<sup>1</sup>, В.В. Миронова<sup>1</sup>, Ф.В. Казанцев<sup>1</sup>,  
И.Р. Акбердин<sup>1</sup>, В.К. Королев<sup>3</sup>, С.И. Фадеев<sup>2,3</sup>, Н.А. Колчанов<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: likho@bionet.nsc.ru; <sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия; <sup>3</sup> Институт математики СО РАН, Новосибирск, Россия

В работе представлена математическая модель распределения ауксина в линейных и кольцевых многокомпарментных структурах растения, состоящих из чередующихся клеток и компарментов апопласта. В линейной незамкнутой в кольцо конфигурации модель соответствует структуре корня, расположенной вдоль его центральной оси. Модель воспроизводит максимум концентрации ауксина, соответствующий максимуму в кончике корня растения, наблюдаемому в экспериментах. Также модель имеет стационарные распределения, содержащие внутренние максимумы концентраций ауксина, которые объясняют случайное месторасположение придаточных корней. В кольцевой конфигурации модель соответствует клеточному ансамблю, располагающемуся в субэпидермальном слое на горизонтальном срезе меристемы побега в районе закладывающихся примордиев листьев. Модель последовательно воспроизводит 0, 1, 2 или 4 стационарных максимума концентрации ауксина при увеличении интенсивности потока ауксина в кольцо извне. Полученные результаты показывают, что четыре основных подсистемы: поступление ауксина в клетки извне, диссипация/деградация ауксина, пассивный и активный транспорт ауксина с учетом конфигурации клеток – достаточны для описания качественных особенностей распределения ауксина вдоль центральной оси корня и в субэпидермальном слое меристемы побега.

**Ключевые слова:** арабидопсис, математическое моделирование, активный транспорт ауксина, корень, меристема, примордии листьев.

### Введение

Формирование локальных максимумов гормона растений ауксина является событием, предeterminирующим закладку латеральных органов в меристемах побега и корня (Reinhardt *et al.*, 2003, De Smet *et al.*, 2007). Можно предполагать, что основную роль в создании локальных максимумов ауксина играет активный вынос ауксина из клетки с помощью белков-транспортёров семейства PIN. Мутанты по гену *PIN1* не образуют цветков (Okada *et al.*, 1991). В мутантах по гену *PIN1* и генам биосинтеза ауксина *YUC1* и *YUC4* не образуются ни листья, ни цветки (Cheng *et al.*, 2007). Микрокапли ауксина, помещенные на апекс побега ниже апикальной меристемы побега (АМП) у *pin1* мутантов, вызывают образование при-

мордиев цветка строго выше точки нанесения и всегда на точно фиксированном расстоянии от меристемы побега (Reinhardt *et al.*, 2003). В случае если микрокапля ауксина наносится у *pin1* мутанта на верхнюю точку меристемы, то равномерно диффундирующий вниз ауксин вызывает формирование кольцевого примордия, опоясывающего АМП (Reinhardt *et al.*, 2003). Это позволяет предполагать наличие определенной кольцевой зоны в меристеме побега, в которой повышение уровня ауксина, локальное в первом случае и общее по кольцу во втором, вызывает образование примордиев.

Верхняя точка меристемы – это зона стволовых клеток, называемая центральной зоной. Центральная зона окружена кольцом периферической зоны меристемы (ПЗ), где клетки делятся быстрее и происходит инициация



примордиев (Laufs *et al.*, 1998, Reinhardt *et al.*, 2003). Клеточные деления и изменения в ориентации клеточных делений в субэпидермальном слое АМП (L2) предшествуют инициации листовых примордиев (Medford *et al.*, 1992). У арабидопсиса инициация первых (ювенильных) листьев происходит в эмбриогенезе, а развитие примордиев этих листьев – в апексе проростка. Примордии первых двух листьев располагаются напротив друг друга. Примордий первого листа немного более развит, что позволяет предполагать, что инициация двух первых примордиев листа является не одновременной, а поочередной. Примордий третьего листа располагается перпендикулярно линии между двумя первыми листьями. Примордий четвертого листа появляется противоположно примордию третьего листа (угол между ними близок к 180°). Инициация пятого и последующих листьев, называемых взрослыми листьями, идет уже в меристеме проростка, угол между примордиями этих листьев составляет 137° (спиральный филлотаксис).

Ранее нами была разработана модель распределения ауксина в клетках, расположенных вдоль центральной оси корня, на которой была продемонстрирована роль активного транспорта ауксина из клетки в образовании локальных максимумов ауксина, предeterminирующих формирование латеральных органов корня (боковых и придаточных корней) (Лихошвай и др., 2007). В то же время моделирование распределения ауксина в меристеме побега показало, что для формирования наблюдаемых максимумов необходимо преобладание скорости поступления ауксина в клетку над процессом его транспорта из клетки (Jonsson *et al.*, 2006). В результате в опубликованных моделях максимумы в корне и меристеме побега формируются на основе противоположных по знаку суммарных потоков ауксина из клетки. В корне он положительный, а в меристеме – отрицательный. В результате возникла проблема описания динамики распределения ауксина в корне и в меристеме побега на основе единой системы элементарных подсистем, управляющих движением ауксина. В настоящей работе мы представляем такую модель. Она описывает движение ауксина в одномерных линейных цепочках компартментов, замкну-

тых и не замкнутых в кольцо. Компартменты соответствуют клеткам и межклеточным пространствам апопласта. В модели движение ауксина описывается четырьмя процессами: поступлением ауксина в клетки извне, диссипацией/деградацией ауксина, пассивным и активным транспортом ауксина. Эти процессы считаются универсальными, протекающими по одним законам как в корне, так и в ядре. Поэтому формулы и значения параметров, описывающие данные процессы, являются одинаковыми для всех возможных конфигураций модели. Различия конфигураций клеточных ансамблей в корневой и побеговой системах моделируются на уровне формирования соответствующей структуры клеточного ансамбля в модели.

В работе продемонстрировано, что модель в линейной конфигурации воспроизводит закономерности распределения ауксина, которые соответствуют закономерностям, наблюдаемым вдоль центральной оси корня. В кольцевой конфигурации модель воспроизводит локальные максимумы концентрации ауксина, предeterminирующие инициацию латеральных органов в меристеме побега. Разработанная модель показывает, что предeterminация инициации латеральных органов в корне и побеге может быть объяснена на основе одних и тех же молекулярных механизмов распределения ауксина между клетками: диффузии, деградации/диссипации, пассивного и активного транспорта ауксина и различий в конфигурации клеточных ансамблей и локализации на мембранах клеток транспортной системы ауксина, опосредованной PIN1 белком.

## Материалы и методы

Модель, представленная в настоящей работе, разработана на основе элементарных моделей четырех подсистем, рассмотренных в работе (Лихошвай и др., 2007). К ним относятся:

Подсистема 1 (Поступление ауксина в компартмент). Скорость поступления описывается уравнением

$$V_a(a) = \alpha, \quad (1)$$

$\alpha$  – интенсивность поступления ауксина.

Подсистема 2 (диссипация/деградация ауксина). Скорость диссипации/деградации описывается уравнением

$$V_d(a, K_d) = K_d a, \quad (2)$$

$K_d$  – константа скорости диссипации/деградации.

Подсистема 3 (пассивный транспорт ауксина). Скорость пассивного транспорта описывается уравнением

$$V_t(a, K_t) = K_t a, \quad (3)$$

$K_t$  – константа скорости пассивного транспорта.

Подсистема 4 (активный транспорт ауксина). Скорость активного транспорта описывается уравнением

$$V_a(a, K_o, f(b)) = K_o f(b) a, \quad (4)$$

$K_o$  – константа скорости активного транспорта. Авторегуляция ауксином собственного активного транспорта, опосредованного PIN1-системой, описана обобщенной функцией Хилла

$$f(b) = \frac{\left(\frac{b}{q_{11}}\right)^{p_1}}{1 + \left(\frac{b}{q_{12}}\right)^{p_1}} \cdot \frac{1}{1 + \left(\frac{b}{q_2}\right)^{p_2}}. \quad (5)$$

В ней учтены активация транспорта низкими концентрациями ауксина в клетке и подавление высокими (Vieten *et al.*, 2005; Sauer *et al.*, 2006). Во всех формулах (1)–(5), через  $a$  и  $b$  обозначена концентрация ауксина в компартменте, которым в настоящей работе может быть клетка или межклеточный объем апопласта (компаратмент апопласта). В работе В.А. Лихошвай с соавт. (2007) в качестве компартмента выступает только клетка.

Для конструирования и расчета модели использовалась компьютерная система MGSmodeller (Likhoshvai *et al.*, 2008). Численное интегрирование проводилось методом Гира (Gear, 1971). При анализе модели использовалась система STEP+ (Fadeev *et al.*, 2006).

### Результаты и обсуждение

Ранее нами была опубликована математическая модель, описывающая распределение ауксина в клетках, расположенных вдоль центральной оси корня (Лихошвай и др., 2007). Для краткости обозначим данную модель как ИМ (исходная модель). ИМ позволила объяснить механизм формирования максимума концентрации ауксина в кончике корня, а также случай-

ность позиций закладки придаточных корней (Лихошвай и др., 2007).

Известно, что в меристеме побега закладка листовых примордиев также преддетерминируется формированием концентрационных пиков ауксина. Возникает естественный вопрос о возможности описания процессов распределения ауксина в корне и побеге на основе одних и тех же процессов. В ИМ рассмотрено 4 типа процессов (1)–(4), ответственных за поступление ауксина извне, его диссипацию/деградацию, пассивный и активный транспорт. Авторегуляция активного транспорта описана формулой (5). ИМ в соответствии со строением корня вдоль центральной оси имеет линейную конфигурацию (Лихошвай и др., 2007). Очевидно, данная конфигурация не подходит для моделирования распределения ауксина в меристеме побега. Простейшей конфигурацией, пригодной для моделирования закономерностей распределения ауксина в меристеме побега, является одномерное кольцо клеток, расположенное на поверхности меристемы побега вокруг ее центральной оси (Jonsson *et al.*, 2006). Однако преобразование исходной модели в кольцевую конфигурацию с сохранением формул (1)–(5) не позволяет получить качественное согласование ее поведения с экспериментом (данные не приводятся). В то же время в ранее опубликованной модели максимумы концентраций ауксина были получены при использовании прямой зависимости скорости поступления ауксина в клетку от его концентрации в ней (Jonsson *et al.*, 2006). В исходной модели наблюдается прямо противоположный эффект: скорость транспорта ауксина из клетки преобладает над скоростью его поступления в клетку. Это объясняет невозможность получения на основе ИМ закономерностей распределения ауксина в меристеме побега. Действительно, в ИМ в качестве компартментов рассматриваются только клетки, а апопласт не рассматривается. Поэтому в ИМ потоки ауксина из клетки в клетку не могут быть структурно отделены один от другого. В результате переход к кольцевой конфигурации ИМ не меняет знака интенсивности потока, что и не позволяет получить требуемый результат. Необходимо структурно отделить процессы поступления ауксина в клетку от процесса его транспорта из клетки. Это можно легко сделать,

расширив исходную модель добавлением в нее второго типа компартментов, соответствующих межклеточным промежуткам апопласта – компартментам апопласта. В расширенной модели ауксин из клетки поступает в примыкающие компартменты апопласта, а из апопласта ауксин поступает в примыкающие клетки. В расширенной модели клетки непосредственно ауксином не обмениваются, что более адекватно отражает строение ткани растения.

Добавление компартментов апопласта в модель позволяет описать наблюдаемые закономерности распределения ауксина и в корне, и в меристеме побега на основе одних и тех же эле-

ментарных блоков (1)–(5), описанных в разделе «Материалы и методы». В расширенной модели в линейной конфигурации автоматически устанавливается преобладание интенсивности потока ауксина из клетки над его потоком в клетку, а в кольцевой – автоматически устанавливается преобладание скорости поступления потока ауксина в клетку над его потоком из клетки. В результате модель оказывается способной описывать распределение ауксина и в корне, и в меристеме побега.

Первый этап в обосновании адекватности расширенной модели состоит в изучении ее свойств в линейной конфигурации (РМЛК):

$$\begin{aligned} \frac{da_{c,N}}{dt} &= V_{\alpha}(\alpha) + V_t(a_{p,N-1}, P_{p,t}) - V_D(a_{c,N}, K_{c,d}) - V_t(a_{c,N}, P_{c,t}) - V_{Pim1}(a_{c,N}, K_{c,o}, f(a_{c,N})), \\ \frac{da_{c,i}}{dt} &= V_t(a_{p,i}, P_{p,t}) + V_t(a_{p,i-1}, P_{p,t}) - 2V_t(a_{c,i}, P_{c,t}) - V_D(a_{c,i}, K_{c,d}) - V_{Pim1}(a_{c,i}, K_{c,o}, f(a_{c,i})), \quad i = \overline{N-1, 2}, \\ \frac{da_{p,i}}{dt} &= V_t(a_{c,i}, P_{c,t}) + V_t(a_{c,i+1}, P_{c,t}) + V_{Pim1}(a_{c,i+1}, K_{c,o}, f(a_{c,i+1})) - 2V_t(a_{p,i}, P_{p,t}) - V_D(a_{p,i}, K_{p,d}), \quad i = \overline{N-1, 2}, \\ \frac{da_{p,1}}{dt} &= V_t(a_{c,2}, P_{c,t}) + V_t(a_{c,1}, P_{c,t}) + V_{Pim1}(a_{c,2}, K_{c,o}, f(a_{c,2})) - 2V_t(a_{p,1}, P_{p,t}) - V_D(a_{p,1}, K_{p,d}), \\ \frac{da_{c,1}}{dt} &= V_t(a_{p,1}, P_{p,t}) - V_t(a_{c,1}, P_{c,t}) - V_D(a_{c,1}, K_{c,d}). \end{aligned} \quad (6)$$

РМЛК (6) так же, как и ИМ, соответствует конфигурации корня вдоль его центральной оси. Но в отличие от ИМ РМЛК содержит (2N-1) переменных, которые соответствуют концентрациям ауксина в клетках и межклеточных компартментах апопласта. N равно количеству клеток. Количество компартментов апопласта равно (N-1). Они расположены между клетками. Нумерация клеток и компартментов апопласта ведется в направлении от кончика корня к его началу. Далее следует компартмент апопласта с номером (N-1), лежащий между клетками N и N-1. Затем располагается клетка с номером N-1, компартмент апопласта с номером (N-2) и так далее вплоть до клетки № 1, находящейся на самом кончике корня. Концентрация ауксина в i-й клетке обозначается как  $a_{c,i}$ ,  $a_{p,i}$  обозначена концентрация ауксина в межклеточном компартменте за номером i,  $K_{c,d}$ ,  $K_{p,d}$  – константы скорости диссипации/деградации ауксина в клетке и апопласте, соответственно  $P_{c,t}$ ,  $P_{p,t}$  – константы скорости транспорта ауксина из клетки,  $K_{c,o}$  – константа скорости активного транспорта ауксина через клетки в апопласт,

$\alpha$  – интенсивность поступления ауксина в N-ю клетку из побега. Аналогично модели ИМ в РМЛК ауксин поступает только в клетку, имеющую номер N, что имитирует поступление ауксина из побега в корень. Так же, как и в ИМ, пассивный транспорт в каждом компартменте РМЛК является бинаправленным, т. е. может происходить как вправо, так и влево. А активный транспорт может происходить только в направлении из клетки в сторону соседнего компартмента апопласта, расположенного по направлению к кончику корня.

РМЛК (6) с параметрами из табл. 1 воспроизводит экспериментально наблюдаемый пик концентрации ауксина в кончике корня и внутренние пики (данные не приводятся). Все основные закономерности ИМ наблюдаются и в линейной конфигурации расширенной модели (6). В частности, РМЛК объясняет механизм поддержания ниши ствольных клеток в меристеме корня и случайное расположение придаточных корней.

Рассмотрим теперь расширенную модель в конфигурации, которая соответствует L2 –

Таблица 1

Значения параметров модели (6), при которых распределение ауксина вдоль вертикальной оси корня, рассчитанное *in silico*, качественно воспроизводит градиент ауксина, выявленный экспериментально

Наименование параметра	Обозначение	Размерность параметра <sup>c</sup>	Значение параметра в модели (1)
Константа интенсивности поступления ауксина в N-ю клетку	$\alpha$	$cu/tu$	1
Константа диффузии ауксина из клетки в клетку <sup>a</sup>	$P_t$	$cu/tu$	
Константа диффузии ауксина из апопласта в клетку <sup>b</sup>	$P_{pt}$	$1/tu$	1
Константа диффузии ауксина из клетки в апопласт <sup>b</sup>	$P^{ct}$	$1/tu$	0,08
Коэффициент диссипации	$K_d$	$1/tu$	0,0045
Константа скорости активного транспорта	$K_o$	$1/tu$	0,21
Константа порога активации ауксин-зависимого транспорта	$q_{11}$	$cu$	1
Константа порога насыщения процесса активации ауксин-зависимого транспорта	$q_{12}$	$cu$	10
Константа порога ингибирования ауксин-зависимого транспорта	$q_2$	$cu$	3
Коэффициент нелинейности механизмов активации	$p_1$	$dl$	2
Коэффициент нелинейности механизмов ингибирования	$p_2$	$dl$	10

<sup>a</sup> – только для ИМ, <sup>b</sup> – только для РМЛК, <sup>c</sup> – значения параметров даются в условных единицах:  $cu$  – условная единица концентрации,  $tu$  – условная единица времени.

слою периферической зоны меристемы побега, в которой находится область инициации листовых примордиев. Эта область может быть представлена в виде одномерного кольца клеток, опоясывающего домен экспрессии маркера центральной зоны гена *CLV3* (Medford *et al.*, 1992; Laufs *et al.*, 1998; Brand *et al.*, 2002; Reinhardt *et al.*, 2003). Известно, что в нем ак-

тивный транспорт ауксина может происходить через обе стенки клетки, ориентированные в обе стороны кольца. В нем наблюдается повышение концентрации ауксина в отдельных клетках, что вызывает инициацию примордиев (Reinhardt *et al.*, 2003). Для того чтобы численно исследовать данный вопрос, построим кольцевую конфигурацию расширенной модели (РМКК):

$$\begin{aligned} \frac{da_{p,N}}{dt} &= P_{c,t}(a_{c,1} + a_{c,N}) + K_{c,o}a_{c,1}f(a_{c,1}) + K_{c,o}a_{c,N}f(a_{c,N}) - (2P_{p,t} + K_{p,d})a_{p,N}, \\ \frac{da_{p,i}}{dt} &= P_{c,t}(a_{c,i+1} + a_{c,i}) + K_{c,o}a_{c,i+1}f(a_{c,i+1}) + K_{c,o}a_{c,i}f(a_{c,i}) - (2P_{p,t} + K_{p,d})a_{p,i}, \quad i = \overline{N-1, 1}, \\ \frac{da_{c,i}}{dt} &= \alpha_i + P_{p,t}(a_{p,i} + a_{p,i-1}) - (2P_{c,t} + K_{c,d} + 2K_{c,o}f(a_{c,i}))a_{c,i}, \quad i = \overline{N, 2}, \\ \frac{da_{c,1}}{dt} &= \alpha_1 + P_{p,t}a_{p,1} + P_{p,t}a_{p,N} - (2P_{c,t} + K_{c,d} + 2K_{c,o}f(a_{c,1}))a_{c,1}. \end{aligned} \quad (7)$$

РМКК конструируется из подсистем (1)–(5) так же, как и РМЛК. Структурные отличия между РМКК и РМЛК определяются отличиями в строении зон моделирования в корне и меристеме побега.

При исследовании РМКК будем использовать набор параметров из табл. 1, при которых РМЛК адекватно описывает распределение ауксина в корне. Данное ограничение вытекает из основной цели работы: построение модели,

описывающей распределение ауксина в корне и меристеме побега на основе одних и тех же молекулярных процессов.

### Стационары РМКК

В апикальной меристеме побега зародыша кольцо периферической зоны в L2 слое состоит примерно из 12 клеток (Barton, Poethig, 1993; Brand *et al.*, 2002), поэтому поведение РМКК мы будем изучать при  $N = 12$ . Цель эксперимента *in silico* состоит в исследовании стационарных распределений концентраций ауксина в клетках и апопласте, которые реализуются при разных значениях параметра  $\alpha$  и разных начальных данных. Так как РМКК (7) является симметричной по строению, то у нее существует симметричный стационар, реализующийся из нулевых начальных данных. Данный стационар удовлетворяет условиям равенства концентраций ауксина в клетках и компартментах апопласта:  $a_{c,i} = a_{c,j}$ ,  $a_{p,i} = a_{p,j}$ ,  $i, j = 1, \dots, N$ . Симметричный стационар при очень малой интенсивности потока ауксина извне является единственным и устойчивым (выкладки не приводятся), т. е. при низкой интенсивности поступления ауксина в кольцо клеток устанавливается равномерное распределение концентрации ауксина в клетках, в котором не имеется концентрационных пиков. Данная ситуация соответствует апикальной меристеме побега зародыша до инициации листовых примордиев.

Повышение интенсивности потока ауксина в клетки моделируемого кольца имитируется в РМКК увеличением значения  $\alpha$ . При этом появляются стационары, отличные от симметричного. Среди стационаров выявляются такие, которые имеют 1, 2 и 4 пика концентраций ауксина (рис. 1).

### Эволюция стационарного распределения концентрации ауксина, реализуемого в РМКК из нулевых начальных данных

Как уже отмечалось выше, при малых значениях параметра  $\alpha$  стационарное распределение концентрации ауксина, реализуемое в модели (7) из нулевых начальных данных, является устойчивым. Далее для краткости назовем его 0-стационаром. При значении  $\alpha = 0$ , 0-стационар является единственным и глобально устойчивым. В результате из любого начального распределения концентраций ауксина формируется 0-стационар. Однако при повышении значения  $\alpha > 0$  число разнообразных стационаров возрастает, и некоторые из них являются стабильными. Поэтому при повышении значения  $\alpha$  область устойчивости 0-стационара становится локальной. Анализ, проведенный методом продолжения по параметру (Fadeev *et al.*, 2006), выявляет наличие области неустойчивости 0-стационара (рис. 2).

Левая граница неустойчивости соответствует значению  $\alpha = 0,0262$ . При переходе через

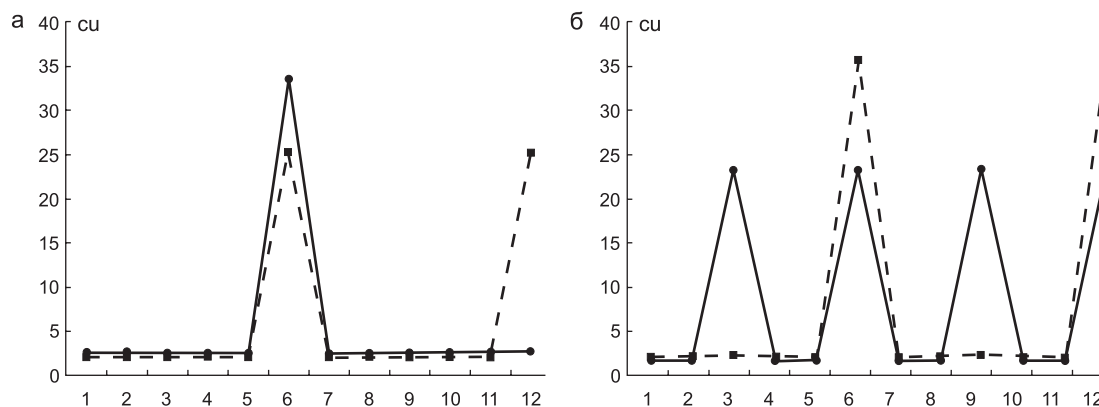
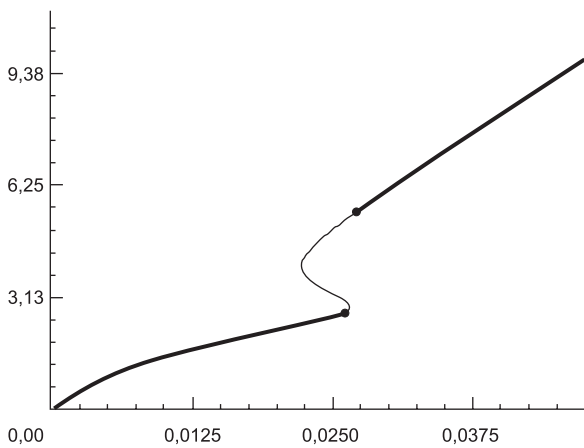


Рис. 1. Распределение концентраций ауксина в решениях модели.

а – с одним (жирная линия,  $\alpha = 0,02622$ ) и двумя (пунктир,  $\alpha = 0,03655$ ) максимумами; б – с двумя (пунктир,  $\alpha = 0,05025$ ) и четырьмя (жирная линия,  $\alpha = 0,05026$ ) максимумами. По оси абсцисс указаны номера клеток.





**Рис. 2.** Диаграмма устойчивости симметричного 0-стационара в зависимости от значения параметра  $\alpha$ .

Жирным выделены устойчивые ветки стационара, тонкой линией – неустойчивые. По оси абсцисс отложено значение  $\alpha$ , по оси ординат – значение стационарной концентрации ауксина в шестой клетке.

данную границу 0-стационар теряет устойчивость. В результате любое начальное сколь угодно малое отклонение концентрации ауксина в клетке приводит к выходу на не 0-стационар. Численные расчеты показали, что преимущественно реализуется стационар, содержащий единственный максимум (рис. 1, а).

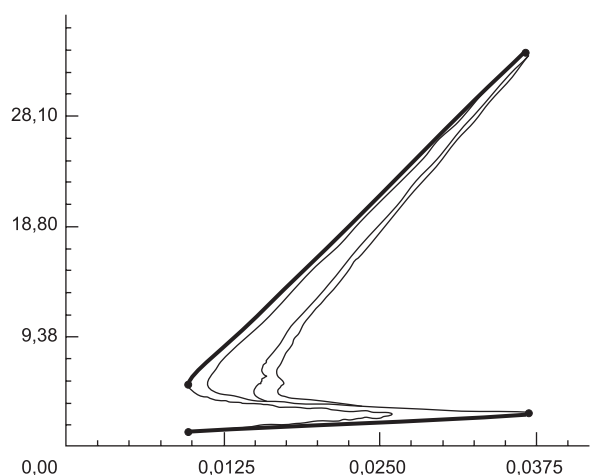
Назовем данный стационар 1-стационаром. Следует отметить, что и при более низких значениях параметра  $\alpha$  в модели (7) имеются 1-стационары, например, при  $\alpha = 0,0125$ . Однако при  $\alpha < 0,0262$  0-стационар является устойчивым и чтобы перейти из него в 1-стационар, необходимо преодолеть некий порог устойчивости. В точке  $\alpha = 0,02621$  происходит потеря устойчивости 0-стационара и самопроизвольный переход из него в 1-стационар. Мы интерпретируем данное численное наблюдение как формирование пика концентрации ауксина, предeterminирующего инициацию первого листа побега в АМП зародыша, что согласуется с экспериментальными наблюдениями (Medford *et al.*, 1992). Таким образом, РМКК объясняет механизм формирования первого пика концентрации ауксина в индивидуальной клетке увеличением притока ауксина извне до величин, при которых 0-симметричное распределение теряет устойчивость. После достижения данного порога даже незначительное увеличение интенсивности внешнего

притока ауксина приводит к переходу системы из 0-стационара (концентрация ауксина одинакова в клетках и компартментах апопласта) в новый стабильный стационар, который является 1-стационаром, характеризующимся наличием максимума концентрации ауксина в единственной клетке. Следует отметить, что на основании проведенного анализа объясняется только механизм возникновения 1-стационара, но не его позиция, которая в силу симметричности модели (7) определяется случайными процессами.

### Стационары РМКК, реализующиеся из начального 1-стационара

Проанализируем теперь стационары, которые реализуются в модели (7) из 1-стационара, полученного в расчете при  $\alpha = 0,02622$  и взятого в качестве начальных данных при увеличении значения параметра  $\alpha$ . Логическим основанием для проведения данного эксперимента *in silico* является экспериментальное наблюдение: примордий второго листа развивается в апексе побега напротив примордия первого листа (Medford *et al.*, 1992). В таком случае второй максимум концентрации ауксина должен формироваться в модели в клетке, расположенной напротив клетки, в которой сформировался первый максимум. На рис. 3 показана рассчитанная диаграмма стабильности 1-стационара при увеличении значения параметра  $\alpha$ .

1-стационар существует в интервале между  $\alpha \in [0,0097, 0,03654]$  (рис. 3). Согласно расчетам, в данном интервале существует несколько стабильных и неустойчивых 1-стационаров. Эксперименты *in silico* показали, что из 1-стационара, полученного при  $\alpha = 0,02622$  и взятого в качестве начальных значений при  $\alpha \in [0,0097, 0,03654]$ , реализующийся стационар также имеет только один максимум, который не меняет позиции (рис. 1, а). При  $\alpha > 0,03654$  1-стационары исчезают. В результате происходит качественная перестройка реализующегося стационара: напротив первого максимума, который не меняет позицию, появляется второй максимум, идентичный первому (рис. 1, а,  $\alpha = 0,03655$ ). Возникает 2-стационар с двумя пиками концентрации ауксина. Перестройка 1-стационара в 2-стационар, имеющий два



**Рис. 3.** Диаграмма устойчивости симметричного 1-стационара в зависимости от значения параметра  $\alpha$ .

Жирным выделены устойчивые ветки стационара, тонкой линией – неустойчивые. По оси абсцисс отложено значение  $\alpha$ , по оси ординат – значение стационарной концентрации ауксина в шестой клетке.

максимума концентраций ауксина в клетках, расположенных друг напротив друга, является ответом системы на рост притока ауксина извне. Таким образом, модель объясняет механизм возникновения второго локального максимума напротив первого максимума в кольце клеток периферической зоны меристемы. Согласно проведенному анализу, установление второго локального максимума следует после формирования первого максимума и является результатом усиления потока ауксина извне. Последовательность событий и расположение «примордиев» друг напротив друга полностью соответствуют экспериментальным наблюдениям (Medford *et al.*, 1992).

#### Стационары РМКК, реализующиеся из начальных данных с двумя максимумами

Проанализируем теперь стационары, которые реализуются в модели (7) из 2-стационара, полученного в расчете при  $\alpha = 0,03655$  и взятого в качестве начальных данных, при увеличении значения параметра  $\alpha$ .

Необходимость проведения данного эксперимента *in silico* вытекает из экспериментального наблюдения, что примордии третьего и

четвертого листа располагаются перпендикулярно линии между двумя первыми листьями, напротив друг друга (Medford *et al.*, 1992). Это указывает на то, что локальные максимумы ауксина, инициирующие эти примордии, также должны формироваться в клетках, расположенных перпендикулярно линии между двумя первыми максимумами напротив друг друга. Применяя метод гомотопии, устанавливаем, что при  $\alpha \in [0,01115, 0,05024]$  2-стационар является устойчивым (Fadeev *et al.*, 2006). Также численно устанавливаем, что задача Коши, решаемая с начальными данными, соответствующими начальному 2-стационару, полученному при  $\alpha = 0,03655$ , сходится к 2-стационару, если значение  $\alpha$  берется из интервала  $[0,01115, 0,05024]$ . Однако правее точки  $\alpha = 0,05024$  начальный 2-стационар исчезает, и новый стационар уже имеет 4 максимума, т. е. формируется 4-стационар. При этом позиции максимумов, которые присутствовали в начальном 2-стационаре, по-прежнему сохраняют позиции, хотя концентрации ауксина в них незначительно снижаются в итоговом 4-стационаре. Позиция двух дополнительных максимумов соответствует положению третьего и четвертого листа в апексе побега, один из максимумов (третий) располагается перпендикулярно линии между двумя первыми максимумами, четвертый максимум располагается напротив третьего (Medford *et al.*, 1992).

Таким образом, модель (7) воспроизводит последовательность формирования локальных максимумов ауксина в периферической зоне меристемы в ответ на увеличение притока ауксина.

#### Заключение

Проведенный в работе численный анализ позволяет сделать вывод, что наблюдаемые в экспериментах закономерности формирования концентрационных пиков в корне и меристеме побега на ранних стадиях развития растения могут быть воспроизведены на основе одних и тех же молекулярных и генетических процессов: поступления ауксина извне, процессов диссипации/деградации ауксина, пассивного и активного транспорта ауксина. Наблюдаемые в корне и меристеме побега индивидуальные особенности распределения ауксина определя-

ются различными конфигурациями клеточных структур. Это позволяет предположить, что локальные максимумы ауксина, предeterminирующие инициацию латеральных органов в корне и первых четырех листьев в меристеме побега, формируются на основе единого механизма распределения ауксина, основанного на диффузии и ауксин-регулируемом активном транспорте. Индивидуальные особенности в корне и меристеме побега проявляются на уровне позиционирования транспортных систем в клетках корня и меристемы, что, собственно, и определяет в них индивидуальную структуру клеточных ансамблей.

### Благодарности

Настоящая работа была поддержана грантом РФФИ 08-04-01214а, Междисциплинарным интеграционным проектом СО РАН № 115, проектом № 10.7 «Компьютерное моделирование и экспериментальное конструирование генных сетей» по программе РАН «Молекулярная и клеточная биология», проектом № 18.13 «Эволюция молекулярно-генетических систем: компьютерный анализ и моделирование» по программе РАН «Происхождение и эволюция биосферы», Госконтрактом № 10104-37/П-18/110-327/180608/015 от 18.06.2008 с ПИН РАН на выполнение НИиОКР по теме «Экосистемно-биоценоотические и генетические механизмы биологической эволюции и корреляция биологических событий» по подпрограмме 2 программы фундаментальных исследований Президиума РАН «Происхождение и эволюция биосферы», грантом НШ-2447.2008.4. Научная школа Н.А. Колчанова «Биоинформатика и системная компьютерная биология».

### Литература

- Лихошвай В.А., Омелянчук Н.А., Миронова В.В. и др. Математическая модель распределения ауксина в корне растения // *Онтогенез*. 2007. Т. 38. Вып. 6. С. 446–456.
- Лихошвай В.А., Казанцев Ф.В., Акбердин И.Р. и др. Компьютерная система для конструирования, расчета и анализа моделей молекулярно-генетических систем (МГСмоделлер) / A computer system for reconstruction, calculation and analysis mathematical models of molecular genetic system (MGSmodeller), Свидетельство № 2008612820, июнь 2008.
- Barton M.K., Poethig R.S. Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild type and in the *shoot meristemless* mutant // *Development*. 1993. V. 119. № 3. P. 821–831.
- Brand U., Grünewald M., Hobe M., Simon R. Regulation of CLV3 expression by two homeobox genes in *Arabidopsis* // *Plant Physiol*. 2002. V. 129. № 2. P. 565–575.
- Cheng Y., Dai X., Zhao Y. Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis* // *Plant Cell*. 2007. V. 19. № 8. P. 2430–2439.
- De Smet I., Tetsumura T., De Rybel B. *et al.* Auxin-dependent regulation of lateral root positioning in the basal meristem of *Arabidopsis* // *Development*. 2007. V. 134. № 4. P. 681–690.
- Gear C.W. The automatic integration of ordinary differential equations // *Communs. ACM*. 1971. V. 14. № 1. P. 176–190.
- Fadeev S.I., Korolev V.K., Gainova I.A., Medvedev A.E. The package Step+ for numerical study of autonomous systems arising when modeling dynamics of genetic-molecular systems // *Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2006)*. Novosibirsk, 16–22 July 2006. Novosibirsk: Inst. Cytol. Genet., 2006. V. 2. P. 118–120.
- Jonsson H., Heisler M.G., Shapiro B.E. *et al.* An auxin-driven polarized transport model for phyllotaxis // *Proc. Natl Acad. Sci. USA*. 2006. V. 103. № 5. P. 1633–1638.
- Laufs P., Grandjean O., Jonak C. *et al.* Cellular parameters of the shoot apical meristem in *Arabidopsis* // *Plant Cell*. 1998. V. 10. № 8. P. 1375–1390.
- Medford J.I., Behringer F.J., Callos J.D., Feldmann K.A. Normal and abnormal development in the *Arabidopsis* vegetative shoot apex // *Plant Cell*. 1992. V. 4. № 6. P. 631–643.
- Okada K., Ueda J., Komaki M.K. *et al.* Requirement of the auxin polar transport system in early stages of *Arabidopsis* floral bud formation // *Plant Cell*. 1991. V. 3. № 7. P. 677–684.
- Reinhardt D., Pesce E.R., Stieger P. *et al.* Regulation of phyllotaxis by polar auxin transport // *Nature*. 2003. V. 426. № 6964. P. 255–260.
- Sauer M., Balla J., Luschnig C. *et al.* Canalization of auxin flow by Aux/IAA-ARF-dependent feedback regulation of PIN polarity // *Genes Dev*. 2006. V. 20. № 20. P. 2902–2911.
- Vieten A., Vanneste S., Wisniewska J. *et al.* Functional redundancy of PIN proteins is accompanied by auxin-dependent cross-regulation of PIN expression // *Development*. 2005. V. 132. P. 4521–4531.

**MODELING OF AUXIN REGULATED LATERAL ORGAN INITIATION  
IN *ARABIDOPSIS THALIANA* L.**

**V.A. Likhoshvai<sup>1,2</sup>, N.A. Omelyanchuk<sup>1</sup>, V.V. Mironova<sup>1</sup>, F.V. Kazantsev<sup>1</sup>, I.R. Akberdin<sup>1</sup>,  
V.K. Korolev<sup>3</sup>, S.I. Fadeev<sup>2,3</sup>, N.A. Kolchanov<sup>1,2</sup>**

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: likho@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia;<sup>3</sup>Institute of Mathematics, SB RAS,  
Novosibirsk, Russia

**Summary**

The mathematical model of auxin distribution in root and shoot apical meristem considering diffusion and active efflux by PIN1 protein is presented. The model reproduces formation of local auxin maxima predetermining initiation of the lateral organs primordia in both root apical meristem and the peripheral zone of shoot apical meristem.

## ГЕНЕТИЧЕСКИЙ КОНТРОЛЬ РАЗВИТИЯ МЕХАНОРЕЦЕПТОРОВ У *DROSOPHILA MELANOGASTER* – ОПИСАНИЕ В БАЗЕ ДАННЫХ «NEUROGENESIS»

Т.А. Бухарина<sup>1</sup>, Д.П. Фурман<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: bukharina@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Большие щетинки (макрохеты) располагаются упорядоченным образом на голове и теле дрозофилы и образуют характерный для каждого вида щетиночный узор. Простая организация каждого щетиночного органа, который состоит всего из четырех специализированных клеток, делает макрохеты удобной моделью для исследования закономерностей формирования пространственных структур с фиксированным числом элементов и механизма клеточной дифференцировки.

В работе систематизированы экспериментальные данные об основных генах и их продуктах, контролирующих три этапа формирования щетиночного узора. На основе созданной базы данных реконструированы две генные сети – «Neurogenesis(determination)» и «Neurogenesis(asymmetric division)». Рассмотрены ключевые элементы и механизмы, обеспечивающие функционирование этих генных сетей.

**Ключевые слова:** *Drosophila melanogaster*, макрохеты, *achaete-scute* комплекс, сигнальные пути, генные сети, регуляторные контуры.

### Введение

Формирование специализированных пространственных структур занимает важное место в реализации онтогенетической программы многоклеточных организмов. Примером такой структуры является щетиночный узор *D. melanogaster*, который образуют на голове и теле насекомого макрохеты (щетинки) – внешние сенсорные органы, выполняющие функцию механорецепторов.

Дефинитивный щетиночный орган состоит из 4 клеток: триогена (щетинковой клетки), тормогена (гнездовой клетки), нейрона и теогена (клетки оболочки нерва). Все они являются производными одной родительской клетки (SOP – Sensory Organ Precursor cell). В морфогенезе макрохет выделяются три этапа. Первый этап состоит в обособлении из массы эктодермальных клеток крылового имагинального диска пронеуральных кластеров – групп из 20–30 клеток, каждая из которых обладает потенциальной возможностью стать SOP. На втором этапе детер-

минируется родительская клетка и уточняется ее позиция внутри пронеурального кластера. В ходе заключительного этапа происходят два асимметричных деления SOP и дифференцировка дочерних клеток в компоненты щетиночного органа. Каждый этап имеет собственное генетическое обеспечение. За топологию пронеуральных кластеров ответственна группа пронеуральных генов комплекса *achaete-scute* (*AS-C*), локальная экспрессия которых инициируется в определенных районах эктодермы специфичными транскрипционными факторами – продуктами генов *u-shaped*, *pannier* и комплекса *iroquois* (Gymez-Skarmeta *et al.*, 2003). Детерминацию родительской клетки и ее позиционирование в пределах кластера контролируют нейрогенные гены, входящие в Notch сигнальный путь (Lai, 2004); направление дифференцировки каждой из дочерних клеток определяют селекторные гены, в частности *numb* и *neuralized* (Knoblich *et al.*, 1995; Bardin *et al.*, 2004).

Контроль становления щетиночного узора осуществляется сложной молекулярно-генети-



ческой системой (Фурман, Бухарина, 2008). К настоящему времени накоплено значительное число экспериментальных данных по этому вопросу, и актуальность приобретает задача их интеграции, систематизации и анализа.

### Материалы и методы

Имеющиеся данные по генетическому контролю развития макрохет сведены в базу данных «Neurogenesis», созданную на основе аннотирования 339 литературных источников. В настоящее время база содержит информацию о 90 генах и 124 продуктах этих генов – 119 белках и 5 РНК.

С помощью технологии GeneNet (Ananko *et al.*, 2005) на основе базы данных «Neurogenesis» реконструированы две генные сети, описывающие два первых этапа (сеть «Neurogenesis (determination)») и третий этап (сеть «Neurogenesis (asymmetric division)») морфогенеза макрохет. Основные характеристики двух генных сетей приведены в табл. 1.

Генная сеть «Neurogenesis (determination)» доступна по адресу: [http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/Neurogenesys\\_determination.html](http://www.mgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/Neurogenesys_determination.html).

Генная сеть «Neurogenesis:determination» анализировалась на предмет поиска ключевых объектов и замкнутых регуляторных контуров – основных характеристик генных сетей. При анализе использовались средства технологии GeneNet, основанные на методах теории графов (Колчанов и др., 2000).

### Результаты и обсуждение

Анализ генной сети «Neurogenesis (determination)» показал, что она характеризуется крайне малым количеством ключевых объектов –

компонентов с максимальными степенями связанности с другими объектами. Таких объектов всего два: комплекс генов *AS-C* и гетеродимерный транскрипционный фактор *AC/DA* (Achaete/Daughterless) с 19 и 10 связями соответственно (рис. 1).

Действительно, экспериментально показано, что инактивация генов комплекса *achaete-scute* влечет утрату всех или части макрохет стандартного набора (Garcia-Bellido, Santamaria, 1978). Нейтральный путь развития клеток предопределяется содержанием в них пронеуральных белков – продуктов генов комплекса *achaete-scute*. В клетках пронеурального кластера оно

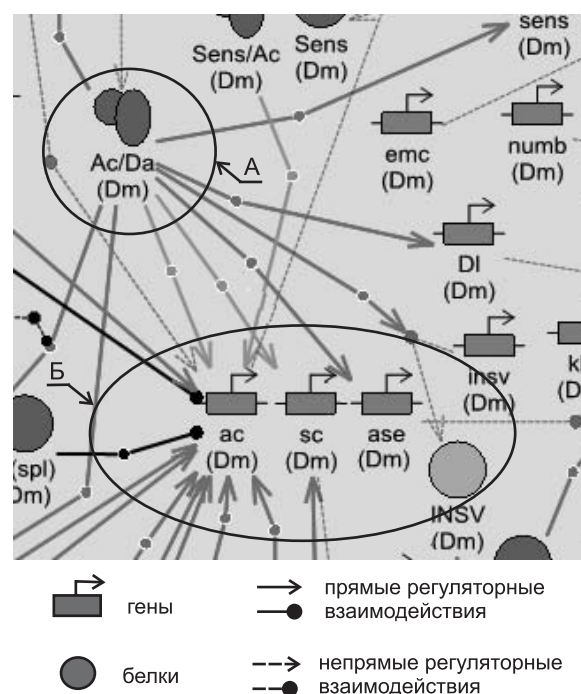


Рис. 1. Фрагмент генной сети «Neurogenesis (determination)» с ключевыми объектами: транскрипционным фактором *AC/DA* (А) и *AS-C* (Б).

*AC/DA* – Achaete/Daughterless; *ac* – *achaete*, *sc* – *scute*, *ase* – *asense* (пронеуральные гены, входящие в состав *AS-C*).

Таблица 1

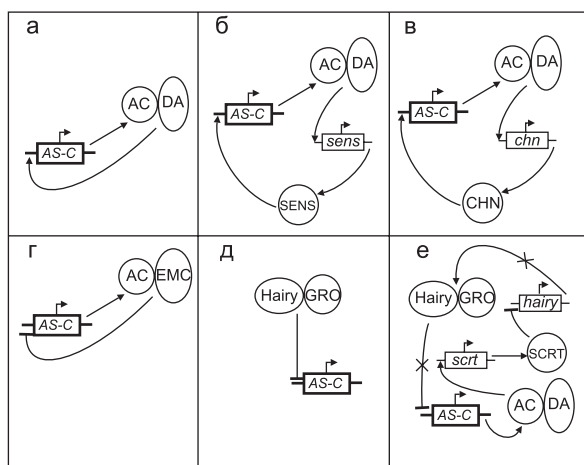
Основные характеристики генных сетей «Neurogenesis:determination» и «Neurogenesis:asymmetric division»

Название сети	Количество белков	Количество генов	Количество РНК	Количество связей	Аннотировано статей
Neurogenesis:determination	90	71	5	220	226
Neurogenesis:asymmetric division	68	46	2	121	120

выше, чем в окружающих клетках эктодермы, а в родительской клетке достигает максимального значения (Reeves, Posakony, 2005).

В генной сети «Neurogenesis (determination)» существуют два типа регуляторных взаимодействий, позволяющих изменять экспрессию генов *AS-C*. Первый тип функционирует при участии регуляторных контуров (с положительными или отрицательными обратными связями), определяющих экспрессию внутри одной клетки. Второй тип связан с межклеточной регуляцией экспрессии генов *AS-C* через регуляторные контуры, включающие сигнальные пути EGFR и Notch.

К первому типу относятся авто- и трансрегуляция экспрессии генов комплекса *AS-C*. Позитивная авторегуляция осуществляется гетеродимерами AC/DA, негативная – гетеродимерами AC/EMC. Трансрегуляция транскрипционной активности комплекса осуществляется через взаимодействие с активаторами (Senseless, Charlatan) и репрессорами (Hairy в комплексе с кофактором Groucho) (рис. 2). Для процессов авторегуляции и позитивной трансрегуляции показано наличие элементарных замкнутых кон-



**Рис. 2.** Схема внутриклеточной регуляции активности генов *AS-C*.

а – позитивная авторегуляция; б, в – позитивная трансрегуляция; г – негативная авторегуляция; д – негативная регуляция репрессором Hairy; е – дополнительный регуляторный контур, подавляющий воздействие репрессора Hairy. Курсивом обозначены названия генов, прямым шрифтом – названия белков. *AS-C* – *achaete-scute* комплекс, AC – Achaete, sens – senseless, GRO – Groucho, EMC – Extramacrochaetae, DA – Daughterless, chn – charlatan, scr1 – scratch. Стрелками указана активация, линиями с обрубленными окончаниями – репрессия процессов.

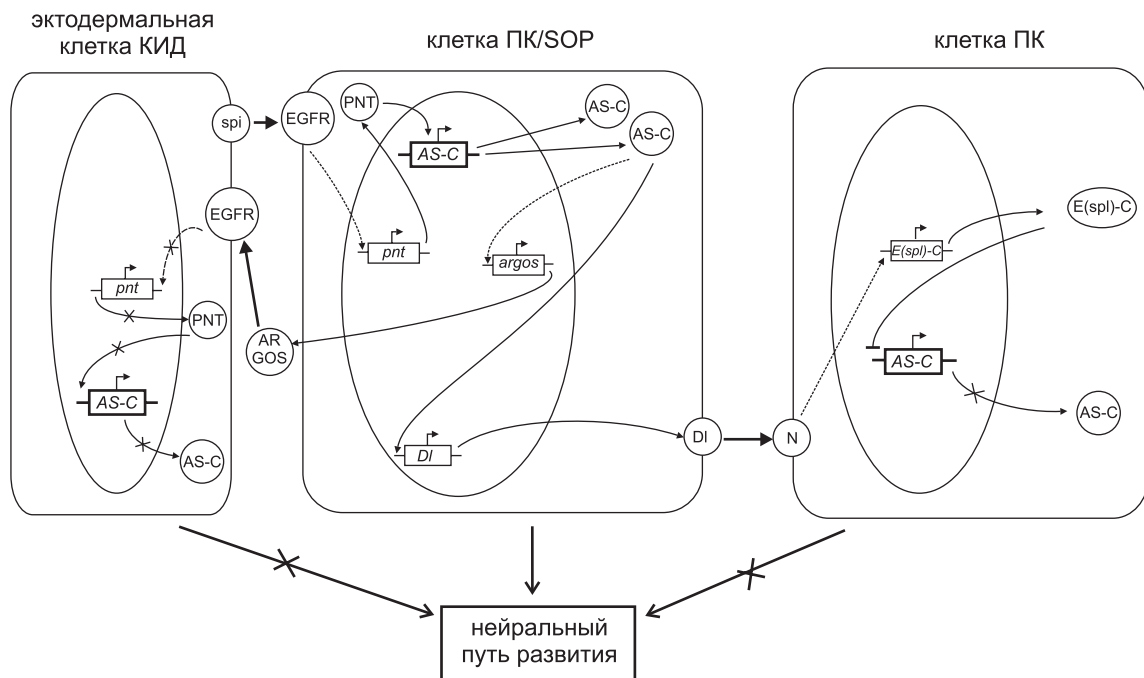
туров, тогда как для процесса негативной трансрегуляции найти такой контур не удалось.

В системе регуляции экспрессии генов *achaete-scute* выявляется более сложный замкнутый контур, регулирующий активность *AS-C* через блокировку экспрессии гена *hairy*. Блокировка является результатом воздействия на этот ген репрессора, кодируемого геном *scratch*, экспрессия которого активируется пронеуральными белками (рис. 2, е).

Благодаря активации и репрессии внутриклеточных механизмов регуляции активности генов *AS-C*, в клетках пронеурального кластера, а затем и в клетке SOP достигается и поддерживается уровень белков этого комплекса, необходимый для дальнейшего развития щетиночного органа (Cabreza, Alonso, 1991; Cabrera *et al.*, 1994; Van Doren *et al.*, 1994; Vaessin *et al.*, 1994; Roark *et al.*, 1995; Smith, Cronmiller, 2001; Escudero *et al.*, 2005).

Наряду с внутриклеточными функционируют и межклеточные механизмы регуляции активности генов *AS-C* через EGFR- и Notch сигнальные пути, которые оказывают противоположное влияние на транскрипционную активность комплекса. EGFR-сигнальный путь завершается синтезом белка Pointed – транскрипционного фактора, активирующего экспрессию генов комплекса *achaete-scute* в клетках пронеурального кластера (Culi *et al.*, 2001). В то же время пронеуральные белки активируют ген *argos*, продукт которого блокирует передачу EGFR-сигнала в окружающие эктодермальные клетки, что способствует увеличению различий по содержанию белков *AS-C* между клетками кластера и окружающими клетками эктодермы (Golembo *et al.*, 1996; Culi *et al.*, 2001) (рис. 3).

Notch-сигнальный путь опосредует процесс латерального ингибирования, закрывающий возможность дифференцировки по нейральному типу всем клеткам кластера, кроме одной (Schweisguth *et al.*, 2004). Финальными событиями в трансдукции сигнала по Notch-пути являются инициация транскрипции генов комплекса *Enhancer of split (E(spl))* и подавление репрессорными белками E(spl)-C транскрипции генов *AS-C* и/или их генов-мишеней во всех клетках пронеурального кластера, кроме будущей клетки SOP, где экспрессия генов *AS-C* продолжается в прежнем режиме (рис. 3).

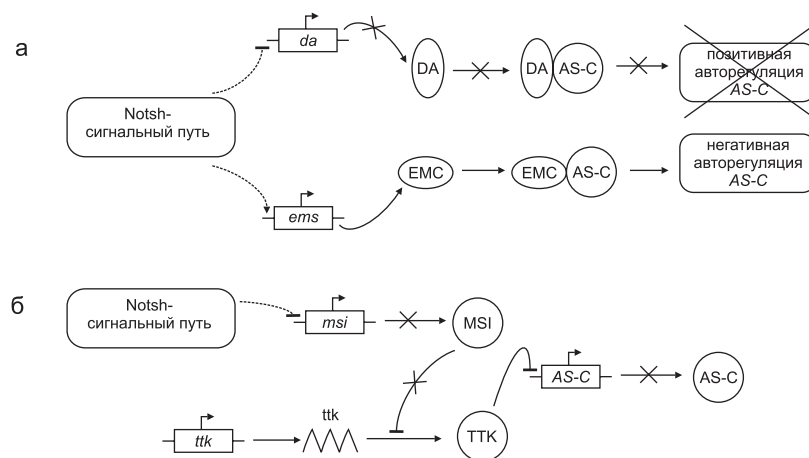


**Рис. 3.** Схема участия EGFR- и Notch-сигнальных путей в регуляции экспрессии генов комплекса *achaete-scute* в эктодермальных клетках КИД и различных типах клеток пронеурального кластера.

*pnt* – pointed, *spi* – spitz, EGFR – Epidermal Growth Factor Receptor, *DI* – Delta, *E(spl)-C* – Enhancer of split complex. Остальные обозначения соответствуют обозначениям на рис. 2.

Помимо прямого подавления активности генов *AS-C* репрессорными белками *E(spl)-C* белки Notch-каскада запускают дополнительные регуляторные механизмы в клетках пронеурального кластера, окружающих клетку SOP. В результате в этих клетках прекращается экспрессия гена

*daughterless* и активируется экспрессия гена *extramacrochaetae*, что приводит к блокировке позитивной авторегуляции экспрессии генов комплекса *AS-C* (Takahisa *et al.*, 1996; Wesley, Saez, 2000) (рис. 4, а). Одновременно с этим происходит снятие запрета на трансляцию РНК



**Рис. 4.** Дополнительные контуры, регулирующие экспрессию генов комплекса *achaete-scute*, активирующиеся с участием белков N-каскада.

*msi* – musashi, *ttk* – tramtrack. Остальные обозначения соответствуют обозначениям на рис. 2.

tramtrack, который является репрессором пронеуральных генов, через подавление транскрипции гена *musashi* (Badenhorst *et al.*, 2002) (рис. 4, б).

Таким образом, сочетанное действие внутри- и межклеточных механизмов, регулирующих экспрессию генов *AS-C*, приводит к дифференциальной активности генов этого комплекса в разных клетках пронеурального кластера. В результате достигается критический уровень содержания белков *AS-C* в единственной клетке кластера, что является необходимым условием для ее детерминации как родительской клетки щетиночного органа.

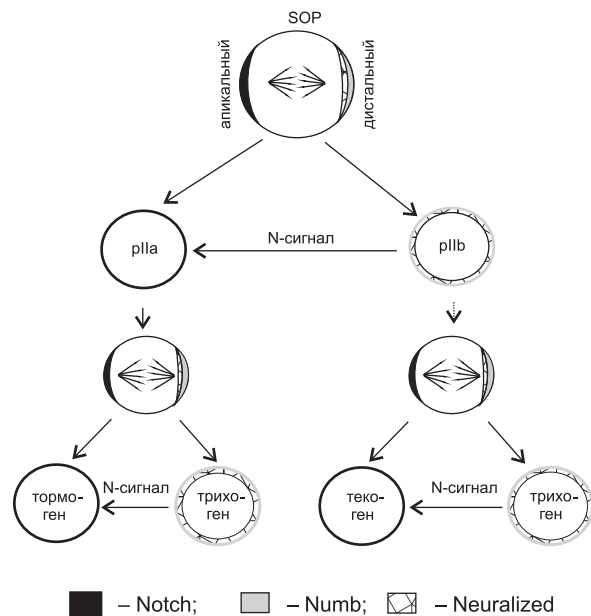
На третьем этапе морфогенеза макрохет происходят два последовательных деления клетки SOP, в результате которых возникают 4 дочерние клетки. Далее они дифференцируются в разные компоненты щетиночного органа: трихоген, тормоген, нейрон и текоген. Этот этап поддерживается генной сетью, насчитывающей около 50 генов.

Ключевыми объектами сети «Neurogenesis (asymmetric division)» являются белки Notch и Numb. Значительную роль в процессе асимметричного деления играют интернализация белков Sanpodo, Delta и Notch (Seto *et al.*, 2002), а также процессы, обеспечивающие правильную ориентацию веретена деления клетки относительно оси распределения белковых детерминант Numb, Neuralized (NEUR),  $\alpha$ -Adaptin (Bowman *et al.*, 2006).

Именно асимметричное распределение белковых детерминант задает направление передачи Notch-сигнала между дочерними клетками и определяет дальнейшее направление их дифференцировки (рис. 5) (Jan Y., Jan L., 2001; Bardin *et al.*, 2004).

После первого митотического деления распределение Numb, NEUR и Notch в дочерних клетках оказывается различным вследствие преимущественной сегрегации этих белков в одну из них (Knoblich *et al.*, 1995; Hutterer, Knoblich, 2005; Le Borgne *et al.*, 2005) (рис. 5).

Дочерняя клетка, в которую попали Numb и NEUR, пойдет по пути нейральной специализации, поскольку Numb будет блокировать передачу Notch-сигнала внутрь нее, а NEUR способствовать передаче сигнала во вторую клетку, свободную от Numb и NEUR. В последующем она даст начало нейрону и текогену.



**Рис. 5.** Схема распределения белков Numb, Neuralized, Notch и направлений передачи N-сигнала при асимметричном делении родительской клетки сенсорного органа и ее дочерних клеток.

pIIa и pIIb – дочерние клетки SOP.

Отсутствие Numb и NEUR во второй дочерней клетке предопределяет возможность нормального восприятия ею Notch-сигнала, что закроет для нее нейральную ветвь развития. После второго митотического деления она даст начало трихо- и тормогену (Frise *et al.*, 1996; Jafar-Nejad *et al.*, 2002).

Позиционирование белковых детерминант в родительской клетке определяется формированием специфических белковых комплексов на противоположных полюсах клетки SOP. В апикальной части клетки SOP образуется комплекс, состоящий из белков, связывающихся с гетеродимером Partner of inscuteable (Pins) и  $\alpha$ -субъединицы G белка (Gai). Это белки Discs Large 1 (Dlg), нефосфорилированная форма белка Lethal(2) giant larvae (Lgl), Partner of Numb (Pon) и Numb. В дистальной части клетки располагается комплекс, состоящий из атипичной протеинкиназы C (aPKC), Par6, Bazooka и фосфорилированного LGL. Экспериментально показано, что именно фосфорилирование LGL с помощью aPKC нарушает связь этого белка с мембраной клетки и препятствует кортикальной локализации белковых детерминант (Betschinger *et al.*, 2003).



Обязательным условием асимметричного деления является определенное расположение веретена деления относительно полярно распределенных белковых детерминант. Комплекс Pins, Gai и белка Mushroom body defect (Mud), ответственный за связывание с микротрубочками веретена, усиливает их полимеризацию и обеспечивает правильную ориентацию веретена деления в клетке (Bowman *et al.*, 2006).

Общая схема регуляции асимметричного деления клеток представлена на рис. 6.

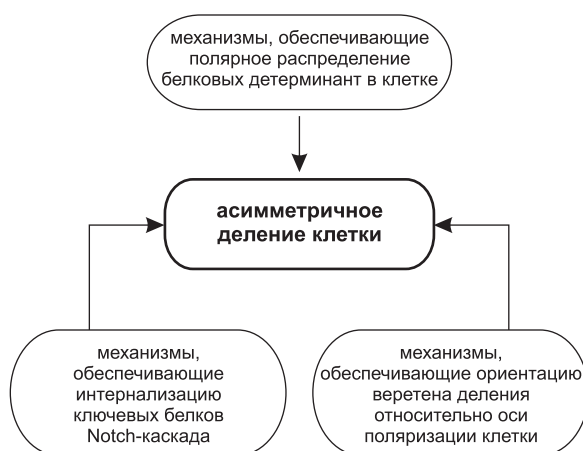


Рис. 6. Общая схема процессов, обеспечивающих асимметричное деление клеток.

Формирование дефинитивного щетиночного органа является результатом последовательного ограничения формообразовательных потенциалов эктодермальных клеток крылового имагинального диска. После определения топологии пронейральных кластеров система генетического обеспечения морфогенеза макрохет начинает работать на увеличение содержания пронейральных белков сначала в клетках пронейрального кластера, а затем в единственной родительской клетке. Контроль этого процесса обеспечивается как внутриклеточной регуляцией активности генов *AS-C*, так и межклеточными событиями, опосредуемыми сигнальными путями EGFR и Notch. Асимметричное деление родительской клетки детерминирует последующее направление специализации дочерних клеток. Асимметричность деления определяется положением специфических белковых комплексов на полюсах родительской клетки,

ориентацией веретена деления относительно этих белковых комплексов, а также процессом интернализации ряда мембранных белков.

В молекулярно-генетическую систему контроля морфогенеза макрохет вовлечено несколько десятков генов.

### Благодарности

Работа была поддержана проектом № 10.7 «Компьютерное моделирование и экспериментальное конструирование генных сетей» по программе РАН «Молекулярная и клеточная биология»; проектом № 18.13 «Эволюция молекулярно-генетических систем: компьютерный анализ и моделирование» по программе РАН «Происхождение и эволюция биосферы»; государственным контрактом 10104-37/П-18/110-327/180608/015 на выполнение НИиОКР по теме «Экосистемно-биоценологические и генетические механизмы биологической эволюции и корреляция биологических событий» по подпрограмме 2 программы фундаментальных исследований Президиума РАН «Происхождение и эволюция биосферы» и грантом НШ-2447.2008.4 «Биоинформатика и системная компьютерная биология».

### Литература

- Колчанов Н.А., Ананьо Е.А., Колпаков Ф.А. и др. Генные сети // Молекуляр. биология. 2000. Т. 34. № 4. С. 533–544.
- Фурман Д.П., Бухарина Т.А. Генетический контроль развития макрохет // Онтогенез. 2008. Т. 39. № 4. С. 245–258.
- Ananko E.A., Podkolodny N.L., Stepanenko I.L. *et al.* GeneNet in 2005 // Nucl. Acids Res. 2005. V. 33. P. D425–D427.
- Badenhorst P., Finch J.T., Travers A.A. Tramtrack cooperates to prevent inappropriate neural development in *Drosophila* // Mech. Dev. 2002. V. 117. P. 87–101.
- Bardin A.J., Le Borgne R., Schweisguth F. Asymmetric localization and function of cell-fate determinants: a fly's view // Curr. Opin. Neurobiol. 2004. V. 14. P. 6–14.
- Betschinger J., Mechtler K., Knoblich J.A. The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl // Nature. 2003. V. 422. P. 326–330.
- Bowman S.K., Neumüller R.A., Novatchkova M. *et al.* The *Drosophila* NuMA Homolog Mud regulates



- spindle orientation in asymmetric cell division // *Dev. Cell*. 2006. V. 10. P. 731–742.
- Cabrera C.V., Alonso M.C. Transcriptional activation by heterodimers of the *achaete-scute* and *daughterless* gene products of *Drosophila* // *EMBO J*. 1991. V. 10. P. 2965–2973.
- Cabrera C.V., Alonso M.C., Huikeshoven H. Regulation of scute function by extramacrochaete *in vitro* and *in vivo* // *Development*. 1994. V. 120. P. 3595–3603.
- Culi J., Martin-Blanco E., Modolell J. The EGF receptor and N signaling pathways act antagonistically in *Drosophila* mesothorax bristle patterning // *Development*. 2001. V. 128. P. 299–308.
- Escudero L.M., Caminero E., Schulze K.L. *et al.* Charlatan, a Zn-finger transcription factor, establishes a novel level of regulation of the proneural *achaete/scute* genes of *Drosophila* // *Development*. 2005. V. 132. P. 1211–1222.
- Frise E., Knoblich J.A., Younger-Shepherd S. *et al.* The *Drosophila* Numb protein inhibits signaling of the Notch receptor during cell-cell interaction in sensory organ lineage // *Proc. Natl Acad. Sci. USA*. 1996. V. 93. P. 11925–11932.
- Garcia-Bellido A., Santamaria P. Developmental analysis of the *achaete-scute* system of *Drosophila melanogaster* // *Genetics*. 1978. V. 88. P. 469–486.
- Golembo M., Schweitzer R., Freeman M., Shilo B.Z. *argos* transcription is induced by the *Drosophila* EGF receptor pathway to form an inhibitory feedback loop // *Development*. 1996. V. 122. P. 223–230.
- Gymez-Skarmeta J.L., Campuzano S., Modolell J. Half a century of neural prepatterning: the story of a few bristles and many genes // *Nat. Rev. Neurosci*. 2003. V. 4. P. 587–598.
- Hutterer A., Knoblich J.A. Numb and alpha-Adaptin regulate Sanpodo endocytosis to specify cell fate in *Drosophila* external sensory organs // *EMBO Rep*. 2005. V. 6. P. 836–842.
- Jafar-Nejad H., Norga K., Bellen H. Numb: «Adapting» notch for endocytosis // *Dev. Cell*. 2002. V. 3. P. 155–156.
- Jan Y.N., Jan L.Y. Asymmetric cell division in the *Drosophila* nervous system // *Nat. Rev. Neurosci*. 2001. V. 2. P. 772–779.
- Knoblich J.A., Jan L.Y., Jan Y.N. Asymmetric segregation of Numb and Prospero during cell division // *Nature*. 1995. V. 377. P. 624–627.
- Lai E.C. Notch signaling: control of cell communication and cell fate // *Development*. 2004. V. 131. P. 965–973.
- Le Borgne R., Remaud S., Hamel S., Schweisguth F. Two distinct E3 ubiquitin ligases have complementary functions in the regulation of delta and serrate signaling in *Drosophila* // *PLoS Biol*. 2005. V. 3. P. e96.
- Reeves N., Posakony J.W. Genetic programs activated by proneural proteins in the developing *Drosophila* PNS // *Dev. Cell*. 2005. V. 8. P. 413–425.
- Roark M., Sturtevant M.A., Emery J. *et al.* *Scratch*, a pan-neural gene encoding a zinc finger protein related to snail, promotes neuronal development // *Genes Dev*. 1995. V. 9. P. 2384–2398.
- Schweisguth F. Regulation of Notch signaling activity // *Curr. Biol*. 2004. V. 14. P. R129–138.
- Seto E.S., Bellen H.J., Lloyd T.E. When cell biology meets development: endocytic regulation of signaling pathways // *Genes Dev*. 2002. V. 16. P. 1314–1336.
- Smith J.E., Cronmiller C. The *Drosophila daughterless* gene autoregulates and is controlled by both positive and negative cis regulation // *Development*. 2001. V. 128. P. 4705–4714.
- Takahisa M., Togashi S., Suzuki T. *et al.* The *Drosophila tamou* gene, a component of the activating pathway of *extramacrochaetae* expression, encodes a protein homologous to mammalian cell-cell junction-associated protein ZO-1 // *Genes Dev*. 1996. V. 10. P. 1783–1795.
- Vaessin H., Brand M., Jan L.Y., Jan Y.N. *daughterless* is essential for neuronal precursor differentiation but not for initiation of neuronal precursor formation in *Drosophila* embryo // *Development*. 1994. V. 120. P. 935–945.
- Van Doren M., Bailey A.M., Esnayra J. *et al.* Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of *achaete* // *Genes Dev*. 1994. V. 8. P. 2729–2749.
- Wesley C.S., Saez L. Analysis of notch lacking the carboxyl terminus identified in *Drosophila* embryos // *J. Cell Biol*. 2000. V. 149. P. 683–696.

**GENETICAL CONTROL OF MECHANORECEPTORS FORMATION  
IN *D. MELANOGASTER* – THE DESCRIPTION  
IN DATABASE «NEUROGENESIS»**

**T.A. Bukharina<sup>1</sup>, D.P. Furman<sup>1,2</sup>**

<sup>1</sup> Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: bukharina@bionet.nsc.ru;

<sup>2</sup> Novosibirsk State University, Novosibirsk, Russia

**Summary**

Large bristles (macrochaetae) are orderly located on drosophila head and body and form the species-specific bristle pattern. A simple organization of each bristle organ, comprising only four specialized cells, makes macrochaetes a convenient model for studying the development patterns of spatial structures with a fixed number of elements and the mechanism of cell differentiation.

The experimental data on the main genes and their products that are involved in the control of three stages of bristle pattern formation were systematized in the work. Two gene networks «Neurogenesis(determination)» and «Neurogenesis(asymmetric division)» were reconstructed on the basis of generated data base. The crucial elements and mechanisms acting in these gene networks are discussed.

## МОДЕЛИРОВАНИЕ МЕХАНИЗМОВ РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНА *EVEN-SKIPPED* У ДРОЗОФИЛЫ

А.Д. Матвеева<sup>1</sup>, Дж. Ионидес<sup>1</sup>, М.Г. Самсонова<sup>1</sup>, Дж. Райнитц<sup>2</sup>

<sup>1</sup> Отдел компьютерной биологии, Центр перспективных исследований Санкт-Петербургского государственного политехнического университета, Россия, e-mail: anya@odd.bio.sunysb.edu; <sup>2</sup>

Отдел прикладной математики и статистики, Центр генетики развития, Университет  
Стони Брук, Нью-Йорк 11794-3600, США

Работа посвящена предсказанию организации 5'-проксимального регуляторного района гена *even-skipped* (*eve*) исходя из картин экспрессии репортерного конструкта *p1.7eve-lacZ*. Для предсказания была использована математическая модель, предложенная в работе Янсенса с соавторами (Janssens *et al.*, 2006). Численные эксперименты, выполненные в этой работе, показали, что наилучшее совпадение решений уравнений модели с экспериментальными данными дает набор из 34 сайтов связывания активаторов Bcd, Hb, Cad и репрессоров Kr, Gt, Kni и Tll. Нами показано, что решения уравнений модели можно улучшить, введя два дополнительных предположения о том, что в силу удаленности сайта N1 от кластера сайтов связывания Bcd фактор Hb, связывающийся с этим сайтом, действует как репрессор и о функциональной значимости сайта связывания транскрипционного фактора Sloppy-paired в исследуемом регуляторном районе гена *eve*.

**Ключевые слова:** дрозофила, гены сегментации, регуляция транскрипции, моделирование, синергия.

### Введение

В основе детерминации сегментов у плодовой мушки дрозофилы на стадии бластодермы лежит точный контроль транскрипции генов сегментации во времени и пространстве. Программа транскрипционного контроля закодирована в регуляторных районах этих генов. Она определяет формирование сложных пространственных картин экспрессии генов сегментации, которые намечают границы будущих сегментов.

Большинство генов сегментации кодируют транскрипционные факторы (ТФ). Регуляторные районы этих генов содержат несколько энхансеров, каждый из которых отвечает за экспрессию на определенном этапе развития и в определенной области эмбриона. Например, на стадии бластодермы экспрессия гена *even-skipped* (*eve*) в виде 7 полос, расположенных перпендикулярно anteriорнопостериорной (переднезадней, А-Р) оси эмбриона, регулируется 4 энхансерами, каждый из которых контролирует экспрессию этого гена в 1 или 2 полосах (Levine, Tjian, 2003).

Наиболее хорошо изучены энхансеры 5'-проксимального регуляторного района гена *eve*, MSE2 и MSE3 (Minimal Stripe Elements) длиной 400 п.н. и 500 п.н. соответственно. Термин «Minimal» означает, что если эти фрагменты сделать меньше, то экспрессия либо значительно уменьшится, либо совсем прекратится. В составе репортерного конструкта *lacZ* энхансер MSE2 определяет формирование картины экспрессии, повторяющей вторую полосу экспрессии нативного гена *eve*, а MSE3 – третью и седьмую полосы.

Однако в эксперименте с конструктом, содержащим нативный ген *eve* без энхансера MSE2, все еще наблюдалась остаточная экспрессия в области второй полосы (Ludwig *et al.*, 2005), а при удалении MSE3 наблюдалась остаточная экспрессия в области седьмой полосы (Small *et al.*, 1992). Эти факты ставят под сомнение достаточность энхансеров MSE2 и MSE3 для регулирования экспрессии в области второй, третьей и седьмой полос и указывают на необходимость более тщательного исследования этих энхансеров и прилегающих к ним областей.

В настоящее время общепринятым подходом к идентификации регуляторных районов генов является *in silico* предсказание сайтов связывания ТФ с последующей экспериментальной проверкой. Однако в случае больших геномов эукариот наиболее распространенные методы предсказания, основанные на использовании позиционно-весовых матриц, не позволяют дискриминировать настоящие сайты связывания от фонового шума и поэтому дополняются поиском либо кластеров транскрипционных сайтов, либо эволюционно консервативных сайтов.

В последнее время был предложен новый подход, заключающийся в предсказании организации регуляторных районов эукариот исходя из картин экспрессии соответствующих генных конструкций (Janssens *et al.*, 2006; Segal *et al.*, 2008). В этой работе мы использовали такой подход с целью предсказания структурной организации, т. е. взаимного расположения и соотношения сайтов связывания ТФ и их возможных кооперативных эффектов, в регуляторном районе гена *eve* длиной 1,7 т.п.н. на основе количественных характеристик экспрессии репортерного конструкта p1.7eve-lacZ. Этот район содержит энхансер MSE2 с прилегающими регуляторными районами и запускает экспрессию репортерного гена в области второй полосы и отчасти в области седьмой полосы (Small *et al.*, 1992).

## Материалы

### Изображения картин экспрессии

В нашем распоряжении имелись 110 латерально ориентированных эмбрионов дрозофилы, трансформированных репортерным конструктом p1.7eve-lacZ, содержащим энхансер MSE2 (рис. 1). Все эмбрионы находятся в циклах деления ядер 13 и 14. У каждого из эмбрионов с помощью конфокальной микроскопии были получены изображения картин экспрессии: нативного гена *eve* на уровне



Рис. 1. Конструкт p1.7eve-lacZ.

белка, мРНК *lacZ* и гистонов. Для регистрации экспрессии гена *eve* использовали метод непрямого иммунофлуоресцентного маркирования, мРНК *lacZ* регистрировали методом гибридизации *in situ*, для выявления гистонов использовали PicoGreen (Molecular Probes) (Wu *et al.*, 2001).

### Данные по концентрациям ТФ, взаимодействующих с энхансером MSE2

Данные по относительным концентрациям ТФ Bicoid (Bcd), Hunchback (Hb), Caudal (Cad), Kruppel (Kr), Giant (Gt), Knirps (Kni) и Tailless (Tll) в каждом временном классе и в каждом ядре эмбриона были получены из базы данных FlyEx (<http://urchin.spbcas.ru/flyex/>).

## Методы

### Получение количественных данных об экспрессии

Обработка изображений картин экспрессии гена *eve* и конструкта p1.7eve-lacZ, а также изображений гистонов осуществлялась с помощью пакета ProStack (Processing Stacks) по ранее разработанному алгоритму (Matveeva *et al.*, 2006). В процессе обработки изображения были приведены к стандартному виду и формату, а информация, заключенная в них, была переведена в текстовые файлы, содержащие значения относительных концентраций каждого из сканированных продуктов в каждом ядре эмбриона. В дальнейшем для построения одномерных картин экспрессии мы учитывали только ядра из центральной 10 %-й полосы, вырезанной в направлении А-Р оси эмбриона.

Принадлежность эмбрионов к циклу деления 13 или 14А определяли подсчетом числа ядер. Эмбрионы, принадлежащие циклу деления 14, дополнительно классифицировались по 8 временным классам, продолжительностью 6,5 мин развития каждый (Т1-Т8), на основе визуального анализа картин экспрессии гена *eve* (Myasnikova *et al.*, 2002). Одномерные данные внутри каждого временного класса регистрировали и усредняли (Myasnikova *et al.*, 2001).

Удаление неспецифического сигнала из количественных данных по экспрессии *lacZ*

было осуществлено по разработанной ранее методике (Myasnikova *et al.*, 2005). Для этого все картины экспрессии *lacZ* были сглажены квадратичными сплайнами (Myasnikova *et al.*, 2001).

### Поиск сайтов связывания

Поиск сайтов связывания ТФ в регуляторном районе конструкта p1.7*eve-lacZ* осуществлен при помощи программы patser-v3d (Hertz, Stormo, 1999) с использованием позиционно-весовых матриц (ПВМ) (Berman *et al.*, 2002; Rajewsky *et al.*, 2002).

### Модель транскрипционного контроля экспрессии гена *eve*

Модель транскрипционного контроля основана на предположении, что инициация транскрипции – ферментативный процесс, катализируемый адаптерами транскрипции. В ходе катализа с увеличением общего числа связавшихся адаптеров уменьшается активационный энергетический барьер по закону Аррениуса и экспоненциально увеличивается скорость транскрипции. Для активации адаптера требуется  $1/C_a$  активирующих ТФ, где величина  $C_a$  характеризует активационную эффективность. Посадке ТФ на ДНК могут помешать две причины: конкурентное связывание и репрессия близкого радиуса действия.

Вероятность связывания ТФ  $a$  с сайтом  $i$  с координатами  $[m_i, n_i]$  относительно старта транскрипции записывается в виде:

$$f_{i[m_i, n_i; a]} = \frac{K_{i[m_i, n_i; a]} v^a}{1 + K_{i[m_i, n_i; a]} v^a}, \quad (1)$$

где  $v^a$  – концентрация ТФ  $a$ , а  $K_i$  – константа связывания с сайтом  $i$ . Константа связывания определенного ТФ с сайтом, обладающим наивысшей оценкой по ПВМ, является свободным параметром, а для последующих сайтов определяется из соотношения ПВМ оценки данного сайта к наивысшей оценке.

Если сайт связывания ТФ  $a$ ,  $i[m_i, n_i; a]$  перекрывается с сайтом связывания ТФ  $b$ ,  $j[m_j, n_j; b]$ , возникает конкурентное связывание, и вероятность связывания ТФ  $a$  с сайтом  $i$  запишется в виде:

$$f_{i[m_i, n_i; a]} = \frac{K_{i[m_i, n_i; a]} v^a}{1 + K_{i[m_i, n_i; a]} v^a + K_{j[m_j, n_j; b]} v^b}. \quad (2)$$

Если ТФ  $a$  является активатором, то обозначаем  $f$  как  $f_{i[m_i, n_i; a]}^A$ , если репрессором близкого радиуса действия (РБД), то как  $f_{i[m_i, n_i; a]}^Q$ . При-сутствие РБД влияет на связывание активаторов со своими сайтами, поэтому полная вероятность связывания ТФ  $a$  с сайтом  $i$  с учетом действия РБД записывается в виде:

$$F_{i[m_i, n_i; a]}^A = f_{i[m_i, n_i; a]}^A \prod_k (1 - q(d_{ik}) E_b f_{k[m_k, n_k; b]}^Q), \quad (3)$$

где  $d_{ik} = m_i - m_k$  – расстояние между сайтами и  $q(d_{ik})$  – эмпирическая функция, равная 1 при  $d_{ik} < 100$  п.н., 0 при  $d_{ik} > 150$  п.н., а при  $d_{ik}$  в интервале от 100 до 150 п.н.  $q(d_{ik})$  находят с помощью линейной интерполяции между 0 и 1.  $E_b$  – эффективность репрессии.

Количество сайтов для посадки адаптера выражается как:

$$N = \sum_a C_a \sum_i F_{i[m_i, n_i; a]}^A. \quad (4)$$

Тогда количество адаптеров, связавшихся через активаторы с этими сайтами, запишется:

$$M = f^{AF} N. \quad (5)$$

В бластодерме дрозофилы концентрация адаптеров фиксирована, поэтому величину  $f^{AF}$  считают фиксированным параметром модели.

Таким образом, по закону Аррениуса скорость транскрипции можно выразить как:

$$d[mRNA]/dt = R_0 \exp(-(\Theta - QM)), \quad \text{при } QM < \Theta \quad (6)$$

$$d[mRNA]/dt = R_0, \quad \text{при } QM \geq \Theta,$$

где свободный параметр  $\Theta$  определяет максимальное значение активационного энергетического барьера в отсутствии активаторов,  $Q$  – фиксированная величина, на которую уменьшается величина энергетического барьера с присоединением каждого из адаптеров, свободный параметр  $R_0$  – максимальная скорость транскрипции (Janssens *et al.*, 2006). Рассчитываемая скорость транскрипции пропорциональна концентрации мРНК, поскольку ее время жизни невелико (6 мин) по сравнению с рассматриваемыми в модели интервалами времени (Reinitz *et al.*, 2003).



Параметры модели находили параллельным методом численного отжига Лама (Lam, Delosme, 1988).

### Результаты

С использованием ПВМ в регуляторном районе гена *eve* длиной 1.7 т.п.н. было найдено большое количество сайтов. Наибольшим сходством с консенсусами, описанными соответствующими весовыми матрицами, обладают 34 сайта связывания активаторов Bcd, Hb, Cad и репрессоров Kr, Gt, Kni и TII (рис. 2).

Как было показано в работе Янссенс с авторами (Janssens *et al.*, 2006), именно этот набор сайтов дает хорошее совпадение решений уравнений модели с экспериментальными данными. Среднеквадратичное отклонение решений уравнений от экспериментальных данных (rms) составило 9,20.

### Коактивация

Качество модели можно улучшить, если помимо информации об относительной локализации сайтов связывания и их аффинности к соответствующим ТФ учитывать возможность возникновения кооперативных эффектов. Экспериментальные данные указывают на существование синергетического взаимодействия между транскрипционными факторами Bcd и Hb в энхансере MSE2 (Simpson-Brose *et al.*, 1994). В MSE2, содержащем несколько сайтов связывания активатора Bcd (рис. 2, B1-B5), фактор Hb проявляет себя в роли активатора транскрипции, тем самым усиливая активационную функцию Bcd. В соседнем энхансере MSE3, не содержащем сайты связывания Bcd, фактор Hb оказывает репрессирующее воздействие на транскрипцию (Small *et al.*, 1991).

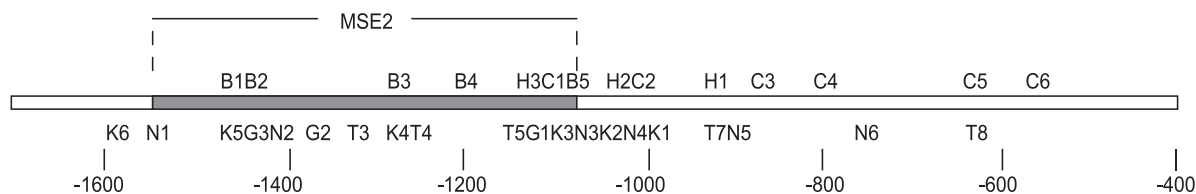
Мы предположили, что в регуляторном районе конструкта *p1.7eve-lacZ* степень синергии между факторами Bcd и Hb напрямую зависит от расстояния между их сайтами связывания на ДНК. Таким образом, в нашей модели было учтено, что на фактор Hb, ассоциированный с сайтом, максимально удаленным от кластера сайтов Bcd (рис. 2, H1), синергетическое воздействие не распространяется, и этот фактор действует как репрессор.

Среднеквадратичное отклонение картин экспрессии конструкта, рассчитанных по такой модели, от экспериментальных данных составило  $rms = 8,66$ . По сравнению с моделью, опубликованной в работе Янссенс (Janssens *et al.*, 2006), решения уравнений новой модели лучше описывают формирование области экспрессии в районе седьмой полосы нативного гена *eve* в ранних временных классах T1 и T2, а также в поздних временных классах T5 и T6. Более того, решения модели очень хорошо согласуются с экспериментальными данными во временном классе C13.

Лучшее согласие решений уравнений новой модели с экспериментальными данными указывает на правомочность гипотезы о том, что в силу удаленности сайта H1 от кластера сайтов связывания Bcd фактор Hb, связывающийся с этим сайтом, действует как репрессор.

### Влияние репрессора Sloppy-paired1

На ранних стадиях, предшествующих циклу 14A, ген *sloppy-paired1* (*slp1*) экспрессируется в виде непрерывной области в anteriорной части эмбриона. Действуя в роли репрессора внутри этой области и в ее окрестностях, Slp1 предотвращает эктопическую активацию некоторых энхансеров генов pair-rule (в том числе энхансера MSE2 гена *eve*), тем самым напрямую влияя на формирование границ полос экспрессии этих генов (Andrioli *et al.*, 2002).



**Рис. 2.** Схема расположения сайтов связывания ТФ Bcd (B), Hb (H), Cad (C), Kr (K), Gt (G), Kni (N) и TII (T) в 5'-проксимальном регуляторном районе гена *eve* длиной 1724 п.н.

В энхансере MSE2 была найдена высококонсервативная последовательность (GTTT)<sub>4</sub>, являющаяся специфическим сайтом связывания репрессора Slp1 (Andrioli *et al.*, 2002). Так как этот сайт расположен рядом с сайтами связывания активатора Bcd, Андриоли с соавторами (Andrioli *et al.*, 2002) предположили, что связывание Slp1 может мешать присоединению активаторов, что объясняет высокую чувствительность энхансера MSE2 к действию Slp1.

Для того чтобы проверить это предположение, результирующая модель была дополнена сайтом связывания Slp1 с координатами от -1221 до -1205.

Среднеквадратичное отклонение картин экспрессии конструкта, рассчитанное по такой модели, от экспериментальных данных составило 8,48.

По сравнению с моделью, учитывающей принцип коактивации, решения уравнений новой модели лучше описывают формирование области экспрессии в районе второй полосы нативного гена *eve* во временном классе T4. Также наблюдается небольшое улучшение в предсказании уровня экспрессии в седьмой полосе во временном классе T6 (рис. 3).

Лучшее согласие решений уравнений модели с экспериментальными данными указывает на

правильность предположения о том, что сайт связывания Slp1 с координатами от -1221 до -1205 важен для контроля экспрессии конструкта *p1.7eve-lacZ*, а значит и для правильной экспрессии гена *eve*.

### Выводы

Данная работа посвящена предсказанию организации 5'-проксимального регуляторного района гена *eve* исходя из картин экспрессии репортерного конструкта *p1.7eve-lacZ*. Этот район содержит энхансер MSE2 с прилегающими регуляторными районами и запускает экспрессию репортерного гена в области второй полосы и отчасти в области седьмой полосы (Small *et al.*, 1992).

Для предсказания была использована математическая модель, предложенная в работе Янссенс с соавторами (Janssens *et al.*, 2006). Численные эксперименты, выполненные в этой работе, показали, что наилучшее совпадение решений уравнений модели с экспериментальными данными дает набор из 34 сайтов связывания активаторов Bcd, Hb, Cad и репрессоров Kf, Gt, Kni и Tll. Эти сайты обладают наибольшим сходством с консенсусами, описанными соответствующими весовыми матрицами.

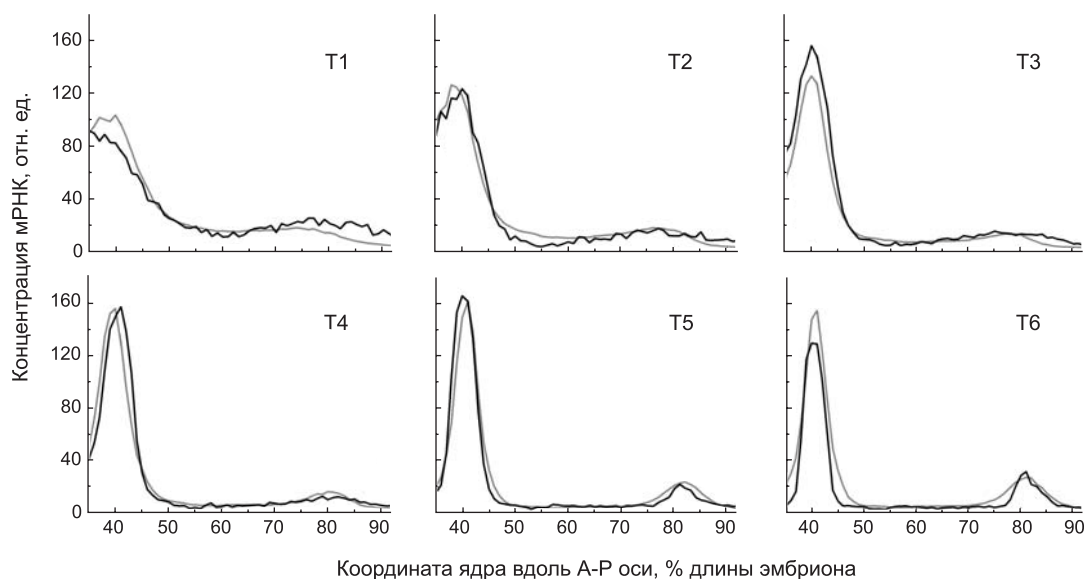


Рис. 3. Картины экспрессии конструкта *p1.7eve-lacZ* во временных классах T1 – T6, построенные по уравнениям модели (серый цвет) и соответствующие экспериментальные данные (черный цвет).

Нами показано, что решения уравнений модели можно улучшить, введя два дополнительных предположения о том, что в силу удаленности сайта H1 от кластера сайтов связывания Vcd фактор Hb, связывающийся с этим сайтом, действует как репрессор (1), и о функциональной значимости сайта связывания Slp1 с координатами от -1221 до -1205 (2). Лучшее согласие решений уравнений модели с экспериментальными данными указывает на правомерность таких гипотез и целесообразность их экспериментальной проверки.

Следует подчеркнуть, что хорошая предсказательная способность модели может быть дополнительно улучшена путем учета нуклеосомной организации 5'-проксимального района, а также путем дополнения модели математическим описанием механизмов коактивации, ко-репрессии и кооперативного взаимодействия.

В конечном счете эти теоретические исследования в сочетании с направленным экспериментом для проверки сформулированных гипотез позволят лучше понять принципы организации регуляторных районов генов у эукариот.

Работа поддержана грантом RR07801 Института здоровья США; программой поддержки грантов GAP Американского фонда гражданских исследований и развития, грант RUB1-1578-ST-05; грантом NWO-RFBR 047.011.2004.013; грантами РФФИ 08-01-00315-а и 08-04-00712-а.

### Литература

- Andrioli L.P., Vasisht V., Theodosopoulou E. *et al.* Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms // *Development*. 2002. V. 129. P. 4931–4940.
- Berman B.P., Nibu Y., Pfeiffer B.D. *et al.* Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome // *PNAS*. 2002. V. 99. P. 757–762.
- Hertz G.Z., Stormo G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences // *Bioinformatics*. 1999. V. 15. P. 563–577.
- Janssens H., Hou S., Jaeger J. *et al.* Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even-skipped* gene // *Nature Genet*. 2006. V. 38. P. 1159–1165.
- Lam J., Delosme J.M. An efficient simulated annealing schedule: Derivation // Technical Report 8816 (Yale Electrical Engineering Department, New Haven, Connecticut). 1988.
- Lam J., Delosme J.M. An efficient simulated annealing schedule: Implementation and evaluation // Technical Report 8817 (Yale Electrical Engineering Department, New Haven, Connecticut). 1988.
- Levine M., Tjian R. Transcription regulation and animal diversity // *Nature*. 2003. V. 424. P. 147–150.
- Ludwig M.Z., Palsson A., Alekseeva E. *et al.* Functional evolution of a *cis*-regulatory module // *PLoS Biol*. 2005. V. 3. I. 4. e93.
- Matveeva A., Kozlov K., Samsonova M. Methodology for building of complex workflows with PROSTAK package and iSIMBioS // Proc. of the 5th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2006). Novosibirsk, 16–22 July 2004. Novosibirsk: Inst. Cytol. Genet., 2006. V. 2. P. 204–208.
- Myasnikova E., Samsonova M., Kosman D., Reinitz J. Removal of background signal from *in situ* data on the expression of segmentation genes in *Drosophila* // *Dev. Genes. Evol.* 2005. V. 215. P. 320–326.
- Myasnikova E., Samsonova A., Kozlov K. *et al.* Registration of the expression pattern of *Drosophila* segmentation genes by two independent methods // *Bioinformatics*. 2001. V. 17. P. 3–12.
- Myasnikova E., Samsonova A., Samsonova M., Reinitz J. Support vector regression applied to the determination of the developmental age of a *Drosophila* embryo from its segmentation gene expression patterns // *Bioinformatics*. 2002. V. 18. P. 87–95.
- Rajewsky N., Vergassola M., Gaul U., Siggia E.D. Computational detection of genomic *cis*-regulatory modules applied to body patterning in the early *Drosophila* embryo // *BMC Bioinformatics*. 2002. V. 3. Article N. 30.
- Reinitz J., Hou S., Sharp D. Transcriptional control in *Drosophila* // *ComplexUs*. 2003. V. 1. P. 54–64.
- Segal E., Raveh-Sadka T., Schroeder M. *et al.* Predicting expression patterns from regulatory sequence in *Drosophila* segmentation // *Nature*. 2008. V. 451. P. 535–540.
- Simpson-Brose M., Treisman J., Desplan C. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in *Drosophila* // *Cell*. 1994. V. 78. P. 855–865.
- Small S., Blair A., Levine M. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo // *EMBO J*. 1992. V. 11. P. 4047–4057.
- Small S., Kraut R., Hoey T. *et al.* Transcriptional regulation of a pair-rule stripe in *Drosophila* // *Genes and Development*. 1991. V. 5. P. 827–839.
- Wu X., Vasisht V., Kosman D. *et al.* Thoracic patterning by the *Drosophila* gap gene *hunchback* // *Developmental Biol.* 2001. V. 237. P. 79–92.

## MODELING THE EXPRESSION OF THE DROSOPHILA EVEN-SKIPPED (EVE) GENE

A.D. Matveeva<sup>1</sup>, J.M.C. Ionides<sup>1</sup>, J. Reinitz<sup>2</sup>, M.G. Samsonova<sup>1</sup>

<sup>1</sup>Department of Computational Biology, Center of Advanced Studies, St. Petersburg State Polytechnical University, St. Petersburg, Russia, e-mail: anya@odd.bio.sunysb.edu; <sup>2</sup>Department of Applied Mathematics and Statistics, and Center for Developmental Genetics, Stony Brook University, Stony Brook, NY, USA

### Summary

In this work we use mathematical model introduced in Janssens *et al.* (2006) to predict the structural organization of the 5'-proximal regulatory region of the *even-skipped* (*eve*) gene from expression pattern of the p1.7*eve-lac* construct. Numerical experiments of Janssens *et al.* (2006) have shown that a set of 34 binding sites, in which Bcd, Hb, Cad act as activators and Kr, Gt, Kni, Tll act as repressors gives the best correspondence between the model solutions and experimental data.

We have shown that two additional assumptions can substantially improve the model accuracy. The first assumption is that Hb bound to H1 site acts as a repressor, because this site is located far away from the cluster of Bcd sites. The second one is about the functional importance of the Sloppy-paired binding site, which is found in the regulatory region of the *eve* gene.

## МОДЕЛИРОВАНИЕ РЕГУЛЯТОРНЫХ СЕТЕЙ ДЛЯ ВЫЯВЛЕНИЯ ПРОТИВООПУХОЛЕВЫХ МИШЕНЕЙ НА ПРИМЕРЕ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ

О.Н. Коборова<sup>1</sup>, Д.А. Филимонов<sup>1</sup>, А.В. Захаров<sup>1</sup>, А.А. Лагунин<sup>1</sup>, А. Кель<sup>2</sup>,  
Ф. Колпаков<sup>3</sup>, Ю.В. Кондрахин<sup>3</sup>, Р. Шарипов<sup>3,4</sup>, В.В. Поройков<sup>1</sup>

<sup>1</sup>ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН, Москва, Россия, e-mail: okoborova@gmail.com; <sup>2</sup>BIOBASE GmbH, Halchtersche Strasse 33, D-38304 Wolfenbüttel, Germany; <sup>3</sup>ООО «Институт системной биологии», Новосибирск, Россия; <sup>4</sup>Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия

Актуальность поиска новых мишеней противоопухолевых препаратов обусловлена низкой эффективностью существующей терапии. Нами предложен и реализован в компьютерной программе NetFlowEx алгоритм, имитирующий поведение регуляторной сети на основе дихотомической модели. Путем инактивации отдельных вершин или их парных комбинаций моделируется воздействие лекарственных препаратов, ингибирующих отдельные белки-мишени в регуляторной сети либо их комбинации. Для генерализованного рака молочной железы мы применили метод к участку регуляторной сети, содержащему белки, связанные с регуляцией клеточного цикла, апоптоза и вовлеченные в прогрессию рака молочной железы, а также в формирование самой железы. В результате были идентифицированы две группы перспективных мишеней, одна из которых связана с инициацией апоптоза (фибронектиновый рецептор alpha5 beta1, фибронектин, белки Bcl-2, c-Jun, Jak2, Jak3, STAT3), а вторая – с остановкой клеточного цикла (Cyclin A, Cyclin B, Cyclin E, Cyclin D1, CDK1, CDK2, пары Histone H1 и Cyclin B, PDK1 и Cyclin B, PKAC и Cyclin B). Для некоторых из идентифицированных мишеней известны ингибиторы, изучаемые как потенциальные препараты для терапии злокачественных новообразований, и в частности рака молочной железы.

**Ключевые слова:** рак молочной железы, моделирование, системная биология, регуляторные сети, перспективные мишени.

### Введение

Актуальность поиска новых мишеней противоопухолевых препаратов обусловлена низкой эффективностью существующей терапии (Bange *et al.*, 2001; Kaue, Workman, 2002). В последние годы происходит накопление обширной геномной, транскриптомной и протеомной информации о структурной организации и функционировании регуляторных сетей в норме и при патологиях. Путем анализа этой информации можно идентифицировать белки, участвующие в возникновении и развитии злокачественных новообразований, и отобрать среди них наиболее перспективные мишени для будущих лекарств (Hill, Khalil, 2005).

Существующие на данный момент статические методы анализа (Spainhour, 2005; Ulitsky, Shamir, 2007) дают возможность работать с большими регуляторными сетями, однако не позволяют моделировать процессы в динамике. А динамические методы, наиболее перспективные из которых гибкие методы моделирования, где в зависимости от изученности системы применяются либо неравновесный (Лихошвай и др., 2004; Zhang *et al.*, 2006; Gea *et al.*, 2008), либо квазистационарный подход (Лихошвай и др., 2004; Millata *et al.*, 2007), либо метод обобщенных функций Хилла (Likhoshvai, Ratushny, 2007), используются пока только для отдельных путей регуляции. В связи с возможными изменениями в динамике поведения живых систем



большинство параметров для детализованного моделирования должно быть взято из экспериментальных данных (Millata *et al.*, 2007), которые отсутствуют для большого числа белков (Gea *et al.*, 2008), а имеющиеся данные, приведенные в различных публикациях, порой противоречат друг другу (Materi, Wishart, 2007). Дискретные методы моделирования могут быть применены к большой регуляторной сети, отображающей такой процесс, как клеточный цикл, где основную информацию несут последовательность состояний клетки и влияние активности отдельных компонентов на выбор дальнейшего пути (Tyson *et al.*, 2001). В этом случае время, которое занимают протекающие процессы, не столь важно. Недавние исследования показывают, что некоторые молекулярные сети сконструированы настолько устойчиво, что время не является критическим фактором (Braunewell *et al.*, 2006).

Целью нашей работы являются разработка алгоритма выявления наиболее перспективных мишеней для противоопухолевых препаратов на основе дискретных подходов моделирования регуляторных сетей и апробация данного алгоритма на примере генерализированного рака молочной железы.

### Материалы и методы

Нами применен алгоритм, имитирующий поведение регуляторной сети на основе дихотомической модели, в рамках которой отдельные вершины сети (белки и/или гены) могут быть в одном из двух состояний: активном и неактивном. Состояния изменяются в дискретные моменты времени. Активное состояние гена соответствует экспрессии его белка, активное состояние белка – его способности связывать свои субстраты. При таком подходе каждый ген и каждый белок представлены в виде вершины направленного графа, в котором ребра проводятся между вершинами тогда и только тогда, когда между ними имеется взаимодействие. Новое состояние вершины на следующем шаге вычисляется как функция от состояний вершин, с которыми она связана входящими в нее ребрами.

Входными данными модели являются состояния вершин (белков и/или генов) в начальный момент времени. Состояние узла обозначено

$S_i$  – для активного узла  $S_i = 1$ , а для неактивного  $S_i = 0$ . Свойства ребер опишем матрицей  $\mathbf{B}$ , элемент которой  $b_{ik} = 1$ , если ребро выходит из узла  $i$  и активирует узел  $k$ ;  $b_{ik} = -1$ , если ребро выходит из узла  $i$  и инактивирует узел  $k$ ;  $b_{ik} = 0$  во всех остальных случаях, а  $b_{ii} = 0$  означает, что узел не активирует сам себя; только одно из двух чисел:  $b_{ik}$  или  $b_{ki}$  может быть отличным от нуля.

В разработанном нами варианте дихотомических сетей моделирование процесса регуляции состоит в вычислении последовательности состояний узлов сети  $\mathbf{S}(0), \mathbf{S}(1), \mathbf{S}(2), \dots, \mathbf{S}(k)$  из заданного начального состояния  $\mathbf{S}(0)$  посредством набора дихотомических функций:

$$S_i(k+1) = F_i(S_1(k), S_2(k), \dots, S_n(k)), \quad (1)$$

где  $i = 1, \dots, n$ ;  $S_i(k)$  – состояние узла  $i$  сети из  $n$  узлов на шаге  $k$ ,  $F_i$  – функция перехода, определяемая структурой сети и правилами комбинирования состояния узлов, связанных с узлом  $i$  входящими в него ребрами. Например, если количество активных узлов, связанных с узлом  $i$ , больше количества связанных с ним неактивных узлов, то узел  $i$  будет активен; если количество активных узлов меньше количества неактивных, то узел  $i$  будет неактивен; если это количество одинаково, то предпочтение отдается активному состоянию узла  $i$ .

Функции перехода дихотомической сети представлены в нашей модели в виде:

$$F_i(S_1, S_2, \dots, S_n) = \Theta(a_i + \sum_k S_k \cdot b_{ki}), \quad (2)$$

где пороговая функция  $\Theta(z) = 0$  при  $z \leq 0$  и  $\Theta(z) = 1$  при  $z > 0$ . Параметры  $a_i$  могут отражать неизвестный участок регуляторной сети. В данной работе все  $a_i \equiv 0$ . В дальнейшем мы планируем варьировать  $a_i$  для более точной настройки модели по известным данным о процессе регуляции.

Последовательность  $\mathbf{S}(0), \mathbf{S}(1), \mathbf{S}(2), \dots, \mathbf{S}(k)$  называется траекторией. Подмножества состояний узлов сети  $\{\mathbf{S}\} \subset 2^n$  называются событиями. Вычисление траектории останавливается после наступления некоторого события, интересующего исследователя, например апоптоза, когда запускается каскад каспаз, и процесс становится необратимым, или остановки клеточного цикла на одной из его фаз.

Молекулы-мишени в нашей модели – это узлы регуляторной сети, воздействие на ко-

которые наиболее существенно изменяет траектории желательным образом: наступление апоптоза или остановка клеточного цикла. Для выявления молекул-мишеней состояние узлов, соответствующих определенным белкам, фиксируется согласно виду предполагаемого воздействия, например  $S_i \equiv 0$  при ингибировании соответствующего белка.

В качестве входной информации для моделирования были взяты данные об экспрессии генов из Cyclonet (<http://cyclonet.biouml.org>). База данных Cyclonet также содержит разнообразную информацию о регуляции клеточного цикла млекопитающих в норме и при патологиях (Kolpakov *et al.*, 2007). Участки регуляторных сетей были взяты из базы данных TRANSPATH® (<http://www.biobase.de>), содержащей информацию о сигнальных молекулах, их взаимодействиях и путях, которые складываются из этих взаимодействий (Krull *et al.*, 2006).

Описанная дихотомическая модель реализована нами в компьютерной программе NetFlowEx на платформе Java.

### Результаты и обсуждение

Ранее нами было показано, что модель воспроизводит процессы регуляции деления клетки в норме, процесс апоптоза в здоровой клетке и не противоречит патологическим процессам при опухоли молочной железы (Коборова и др., 2008). В данной работе метод был применен к большому участку регуляторной сети, состоящему из 828 вершин и 1849 ребер и содержащему различные фрагменты путей, связанных с регуляцией клеточного цикла и апоптоза, такие, как каскады каспаз и MAP киназ, пути TNFalpha, IL-6, TGFbeta, HIF-1alpha, VEGF, Fas, EGF, Wnt/beta-catenin и другие пути и входящие в них белки, которые вовлечены в прогрессию рака молочной железы, а также в формирование самой железы. Кроме того нами были использованы данные об экспрессии для генерализованного случая рака молочной железы, содержащие 42 гипер- и 48 гипэкспрессированных белков, входящих в выбранную сеть.

Процесс регуляции клеточного цикла был промоделирован с помощью NetFlowEx. За 9 шагов моделирования происходит одно деление клетки. На большем количестве шагов

также происходят клеточные деления, и моделируемый процесс остается стабильным и итеративным, поэтому для удобства расчетов было взято количество шагов, соответствующее одному делению.

**Инициация апоптоза.** Результаты моделирования с использованием данных об экспрессии в случае генерализованного рака молочной железы показали, что апоптоз заблокирован, и происходят последующие генерации новых делений. В качестве конечного эффекта, который относится к инициации апоптоза, была выбрана активация цитохрома C. При ингибировании вершин было выявлено, что инактивация фибронектинового рецептора alpha5 beta1, фибронектина и белков Bcl-2, c-Jun, Jak2, Jak3, STAT3 по отдельности ведет к выбранному нами эффекту. На рис. 1 представлен фрагмент сети, содержащий некоторые мишени.

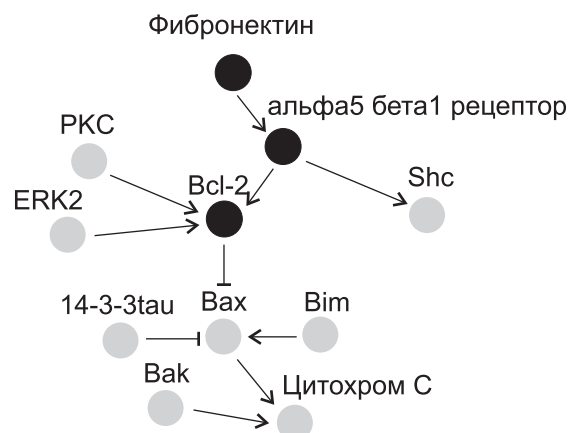


Рис. 1. Фрагмент регуляторной сети, содержащий некоторые выявленные мишени.

Темным цветом обозначены белки – выявленные мишени, светлым – некоторые связанные с ними белки.

**Остановка клеточного цикла.** В качестве конечного желаемого эффекта нами было выбрано ингибирование комплексов Cyclin/CDK, так как это ведет к остановке клеточного цикла. Данные комплексы могут быть ингибированы непосредственно или непрямым воздействием через ингибирование или активацию других белков/генов.

Сначала ингибирование каждого комплекса Cyclin/CDK было рассмотрено отдельно, и промоделировано ингибирование всех других вершин в сети по одной и попарно. Для комплекса

Cyclin A:CDK1, который обладает повышенной экспрессией во время S фазы, были найдены следующие пары: Cyclin A и Cyclin B, Cyclin B и Cyclin E, Cyclin B и PDK1, Cyclin B и HISTONE H1, Cyclin B и PKAC. Для комплексов Cyclin E: CDK2 и Cyclin B: CDK1 пары не были найдены, но независимое ингибирование составных частей этих комплексов Cyclin E и CDK2, Cyclin B и CDK1 приводит к остановке клеточного цикла.

Список перспективных мишеней, выявленных в данном исследовании, включает циклины, важные для осуществления клеточного деления. Выявленные пары Cyclin B и PKAC, Cyclin B и PDK1 более интересны по сравнению с парами, содержащими только циклины, так как их комбинации не настолько очевидны.

При рассмотрении ингибирования всех комплексов Cyclin/CDK одновременно в качестве желаемого эффекта нами не были найдены нетривиальные комбинации мишеней, блокирование которых ведет к такому эффекту. Это связано с тем, что одновременно множество различных белков участвуют в регуляции комплексов Cyclin/CDK, блокада одного и даже пары таких регуляторов не приводит к единовременному ингибированию комплексов Cyclin/CDK.

**Оценка выявленных мишеней.** Дополнительными критериями для ранжирования мишеней являются: 1) относительная важность требуемого состояния (например остановка клеточного деления или переход клетки в апоптоз) для достижения конечного эффекта (терапия опухоли); 2) наличие измененной экспрессии данной мишени при рассматриваемой патологии; 3) количество ребер, входящих и выходящих из данной вершины, соответствующее количеству взаимодействий с другими белками и/или генами. Ингибирование белка с малым числом функций по сравнению с остальными выявленными мишенями может вносить несущественный вклад в исследуемый процесс. Белки с большим числом функций могут относиться к процессам жизнедеятельности клетки, влияние на которые было бы нежелательным. В связи с этим белки со средним количеством функций считаются нами более приоритетными.

Инициация апоптоза в раковых клетках, т. е. их гибель, нами рассматривается как более важный процесс для лечения опухолевых заболеваний, чем остановка клеточного цикла опу-

холевых клеток, т. е. их сохранение, поскольку остановленный клеточный цикл при различных условиях может запуститься вновь, и рост опухоли продолжится. Таким образом, мишени, ингибирование которых приводит к апоптозу, были нами отнесены к первой группе наиболее перспективных (табл. 1). Среди данной группы мишени ранжировались на основе перечисленных выше критериев 2 и 3.

Например, фибронектину присвоен первый ранг, так как он гиперэкспрессирован в генерализованном случае рака молочной железы (Kolpakov *et al.*, 2007), имеет среднее число функций и соответствующее им малое число взаимодействий с другими белками. Кроме того, фибронектин вовлечен в механизм клеточной адгезии, процессы миграции клеток, включая эмбриогенез, заживление ран, свертывание крови, метастазирование.

Ко второй, менее приоритетной, группе отнесены мишени, воздействие на которые приводит к остановке клеточного цикла на одной из его фаз. Например, Cyclin A имеет первый ранг в данной группе, так как он обладает только одной функцией, являясь участником перехода из фазы S в фазу G2 клеточного цикла. Для пары необходимо искать либо лиганд, который воздействует на оба белка в паре, либо пару лигандов, что сложнее по сравнению с поиском одного лиганда для одной мишени.

Нами была проведена проверка полученных результатов на нашей модели в норме и при патологии (данные Cyclonet – Kolpakov *et al.*, 2007). Для этого были взяты две выявленные мишени: фибронектин, который гиперэкспрессирован при патологии, и Jak2 киназа, которая не гиперэкспрессирована при патологии, однако выявлена нами как перспективная мишень. При ингибировании домена фибронектина, отвечающего за проведение ростового сигнала, в норме цитохром C остается неактивным, каскад каспаз не запускается, комплексы Cyclin/Cdk активны, клетка претерпевает деление (рис. 2, а). При патологии в нашей модели активируется цитохром C, далее следует каскад каспаз, и клетка уходит в апоптоз, несмотря на активные комплексы Cyclin/Cdk (рис. 2, б).

При ингибировании киназы Jak2 в норме клетка претерпевает деление и не уходит в

Таблица 1

Оценка перспективных мишеней для генерализированного случая рака молочной железы

Эффект	Название мишени		Количество входящих/выходящих вершин	Количество функций
Инициация апоптоза	Одиночные мишени	1. Фибронектин*	1/3	9
		2. Фибронектиновый рецептор (alpha5 beta1)	1/2	8
		3. Bcl-2	8/3	3
		4. Jak3	1/2	10
		5. Jak2	6/6	17
		6. c-Jun	13/28	5
		7. STAT3	18/32	16
Остановка клеточного цикла	Одиночные мишени	1. Cyclin A	2/1	1
		2. Cyclin B	3/1	3
		3. Cyclin E	2/1	5
		4. Cyclin D1	4/2	1
		5. CDK1	3/4	9
		6. CDK2	1/1	11
	Пары	7. Histone H1 (с Cyclin B)	1/1	3
		8. PDK1 (с Cyclin B)	4/4	7
		9. PKAC (с Cyclin B)	2/16	7

\* Данный белок гиперэкспрессирован в начальный момент времени.

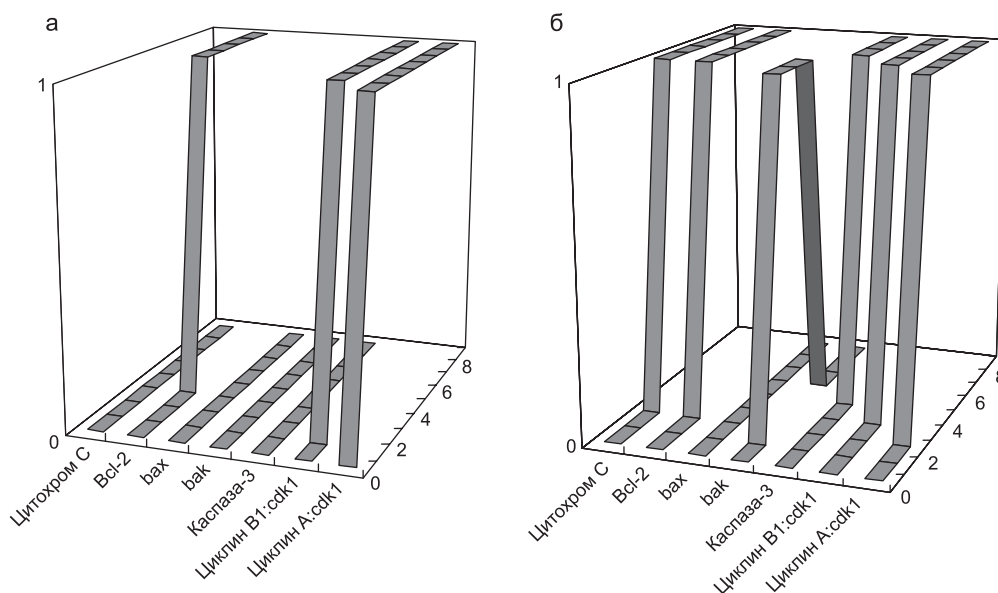
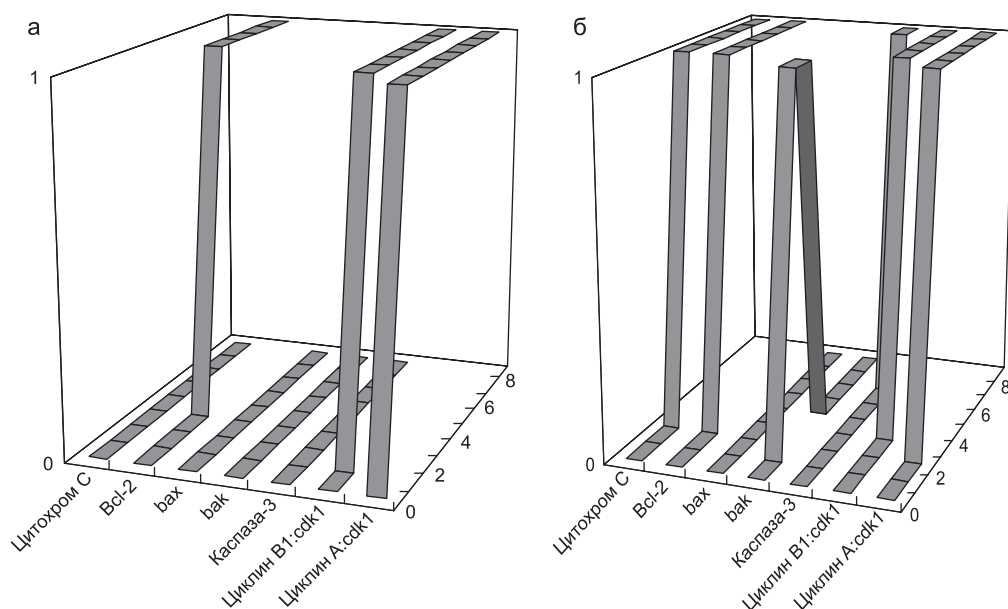


Рис. 2. Результаты моделирования для некоторых белков, отвечающих за апоптоз и деление при ингибировании фибронектина.

По оси X отложены белки, по оси Y – состояния белков, по оси Z – дискретные единицы времени. а – ингибирование фибронектина в норме; б – при патологии.



**Рис. 3.** Результаты моделирования для некоторых белков, отвечающих за апоптоз и деление при ингибировании Jak2.

По оси X отложены белки, по оси Y – состояния белков, по оси Z – дискретные шаги времени. а – ингибирование Jak2 киназы в норме; б – при патологии.

апоптоз (рис. 3, а). При патологии видно, что белок bak становится активным, способствует выходу цитохрома С из митохондрии, а Bcl-2 становится активным на третьем шаге и уже не способен сдерживать выход цитохрома С (рис. 3, б).

Как видно из результатов моделирования, ингибирование выявленных нами мишеней в норме не приводит к апоптозу, однако в опухоли это происходит. Кроме того, по данным Integrity (Prous Science – [www.prous.com](http://www.prous.com)), существуют ингибиторы, например, фибронектина, которые относятся к онколитической группе препаратов и уже используются в клинической практике. Кроме того, ингибиторы циклинзависимых киназ и циклинов также используются для терапии рака. Например, UCN-01 является ингибитором CDK1, CDK2, CDK6, CDK4, PI3K, CHK1, CHK2, PKC, PDK1, т. е. индуктором апоптоза и ингибитором некоторых сигнальных путей; находится на второй фазе клинических испытаний (Welch *et al.*, 2007).

### Заключение

Нами разработан алгоритм выявления перспективных противоопухолевых мишеней и их сочетаний, основанный на существующих

дискретных подходах моделирования, который обеспечивает анализ регуляторных сетей и доступных для большого количества белков данных об их генной экспрессии. Результаты апробации данного алгоритма для генерализованного рака молочной железы показали, что для выявленных нами мишеней известные ингибиторы принадлежат к группе противоопухолевых препаратов, находящихся на различных стадиях доклинических и клинических исследований. Ингибиторы для STAT3, Cyclin E, Cyclin D1 уже исследуются как препараты для терапии рака молочной железы.

### Благодарности

Работа выполнена при поддержке гранта Европейской комиссии № 037590 (FP6-2005-LIFESCIHEALTH-7).

### Литература

- Коборова О.Н., Филимонов Д.А., Захаров А.В. и др. Выявление противоопухолевых мишеней с использованием биоинформационных технологий // Рос. биотерапевт. журнал. 2008. Т. 7. № 2. С. 54–56.
- Лихошвай В.А., Фадеев С.И., Матушкин Ю.Г. и др. Математическое моделирование регулярных



- контуров генных сетей // Журн. вычисл. математики и матем. физики. 2004. Т. 44. № 10. С. 1921–1940.
- Bange J., Zwick E., Ullrich A. Molecular targets for breast cancer therapy and prevention // *Nature medicine*. 2001. V. 7. № 5. P. 548–552.
- Braunewell S., Bornholdt S. Superstability of yeast cell-cycle dynamics: ensuring casuality in the presence of biochemical stochasticity // *J. Theor. Biol.* 2006. V. 245. № 4. P. 638–643.
- Gea H., Qian H., Qian M. Synchronized dynamics and non-equilibrium steady states in a stochastic yeast cell-cycle network // *Math. Biosci.* 2008. V. 211. № 1. P. 132–152.
- Hill C., Khalil I.G. Systems biology for cancer // *Curr. Opin. Oncol.* 2005. V. 17. № 1. P. 44–48.
- Kaye S.B., Workman P. Translating basic cancer research into new cancer therapeutics // *Trends Mol. Med.* 2002. V. 8. № 4. P. 1–9.
- Kolpakov F., Poroikov V., Sharipov R. et al. CYCLONET—an integrated database on cell cycle regulation and carcinogenesis // *Nucl. Acids Res.* 2007. № 35. P. 550–556.
- Krull M., Pistor S., Voss N. et al. TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations // *Nucl. Acids Res.* 2006. № 34. P. 546–551.
- Likhoshvai V., Ratusny A. Generalized hill function method for modeling molecular processes // *J. Bioinform. Comput. Biol.* 2007. V. 5. № 2B. P. 521–531.
- Materi W., Wishart D.S. Computational systems biology in drug discovery and development: methods and applications // *Drug Discov. Today*. 2007. V. 12. № 7/8. P. 295–303.
- Millata T., Bullinger E., Rohwer J., Wolkenhauer O. Approximations and their consequences for dynamic modelling of signal transduction pathways // *Mathem. Biosci.* 2007. V. 207. № 1. P. 40–57.
- Spainhour C.B. *Drug Discovery Handbook* / Ed. S.C. Gad. N.Y.: Wiley-InterScience, 2005. P. 41–43.
- Tyson J., Chen K., Novak B. Network dynamics and cell physiology // *Nature Rev. Mol. Cell. Biol.* 2001. № 2. P. 908–916.
- Ulitsky I., Shamir R. Identification of functional modules using network topology and high-throughput data // *BMC Systems Biol.* 2007. V. 1. № 8. P. 1–17.
- Welch S., Hirte H.W., Carey M.S. et al. UCN-01 in combination with topotecan in patients with advanced recurrent ovarian cancer: a study of the Princess Margaret Hospital Phase II consortium // *Gynecol. Oncol.* 2007. V. 106. № 2. P. 305–310.
- Zhang Y., Yu H., Deng M., Qian M. Nonequilibrium Model for Yeast Cell Cycle // *Computational Intelligence and Bioinformatics*. 2006. V. 4115. P. 786–791.

## MODELLING OF REGULATORY NETWORKS IN IDENTIFICATION OF PROMISING DRUG TARGETS FOR BREAST CANCER THERAPY

O.N. Koborova<sup>1</sup>, D.A. Filimonov<sup>1</sup>, A.V. Zakharov<sup>1</sup>, A.A. Lagunin<sup>1</sup>, A. Kel<sup>2</sup>, F. Kolpakov<sup>3</sup>, R. Sharipov<sup>3,4</sup>, Y. Kondrachin<sup>3</sup>, V.V. Poroikov<sup>1</sup>

<sup>1</sup>The Orekhovich Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Moscow, Russia, e-mail: okoborova@gmail.com; <sup>2</sup>BIOBASE GmbH, Germany; <sup>3</sup>LTD «Institute of Systems Biology», Novosibirsk, Russia; <sup>4</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

### Summary

Low efficacy of current therapy is the reason for investigation of new anticancer drug targets. We proposed an algorithm for anticancer drug target identification, which is implemented in NetFlowEx program. The algorithm simulates behavior of a regulatory network on the basis of dichotomy model. The effect of drugs, which inhibit particular protein or combination of proteins in the regulatory network, is simulated by inhibition of single nodes or their pairwise combinations.

The method was applied to the case of generalized breast cancer using fragment of the regulatory network, which contains proteins involved in cell cycle regulation, apoptosis, breast cancer progression and normal formation of breast. As a result, two groups of promising targets were identified: the first – a number of targets, inhibition of which is connected with initiation of apoptosis (fibronectin receptor alpha5 beta1, fibronectin, Bcl-2, c-Jun, Jak2, Jak3, STAT3) and the second – a number of targets, inhibition of which is connected with cell cycle arrest (Cyclin A, Cyclin B, Cyclin E, Cyclin D1, CDK1, CDK2, pairs Histone H1 and Cyclin B, PDK1 and Cyclin B, PKAC and Cyclin B). Inhibitors of some identified targets are known and considered as potential drugs for therapy of malignant diseases and, in particular, for breast cancer.

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДЕЙСТВИЯ ПОТЕНЦИАЛЬНЫХ ПРОТИВОВИРУСНЫХ ПРЕПАРАТОВ НА РЕПЛИКАЦИЮ СУБГЕНОМНОГО РЕПЛИКОНА ВИРУСА ГЕПАТИТА С В КЛЕТКЕ

Е.Л. Мищенко<sup>1</sup>, К.Д. Безматерных<sup>1,2</sup>, В.А. Иванисенко<sup>1</sup>,  
В.А. Лихошвай<sup>1,2</sup>, Н.А. Колчанов<sup>1,2</sup>

<sup>1</sup> Институт цитологии и генетики СО РАН, Новосибирск, Россия, e-mail: elmish@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Математическое моделирование является полезным инструментом для изучения действия противовирусных препаратов на сложные системы, одной из которых является вирус–клетка. Построена математическая модель, описывающая действие потенциальных противовирусных препаратов на репликацию РНК репликона вируса гепатита С в Huh-7 клетке. Рассчитаны динамики вирусных компонент (РНК, полипротеина, активной репликазы, супрамолекулярных комплексов плюс- и минус-цепей РНК и активной репликазы) в клетке в присутствии потенциальных противовирусных препаратов – ингибиторов инициации трансляции. Для ингибиторов разных типов – NS3/NS4A протеазы, NS5B полимеразы и инициации трансляции рассчитаны зависимости минимального времени, требуемого для удаления репликона из клетки, от аффинности ингибитора к мишени. Сравнение этих зависимостей показало, что наиболее эффективной мишенью действия препаратов является NS5B полимеразы.

**Ключевые слова:** математическое моделирование, субгеномный репликон вируса гепатита С, предсказание эффектов противовирусных препаратов, специфические ингибиторы белков ВГС.

### Введение

Инфекция вирусом гепатита С (ВГС) способна вызывать тяжелое хроническое поражение печени, часто приводящее к циррозу и (или) злокачественному перерождению клеток печени. В настоящее время около 2–3 % мирового населения инфицировано ВГС и, согласно прогнозам, распространенность этой инфекции в ближайшем будущем будет увеличиваться (Hoofnagle, 2002; Chou *et al.*, 2004). Поиск новых и эффективных анти-ВГС препаратов чрезвычайно актуален, так как вакцина против ВГС не получена, а современная терапия, основанная на комбинированном применении интерферона альфа (IFN) и нуклеозидного аналога рибавирина, малоэффективна (Dudley *et al.*, 2006; El-Gogo *et al.*, 2008).

Знание жизненного цикла ВГС предоставляет новые перспективы для разработки эффектив-

ных противовирусных препаратов, направленно действующих на мишени ВГС. В настоящее время в качестве таких потенциальных препаратов проходят испытания высокоспецифичные ингибиторы NS3/NS4A протеазы (Pause *et al.*, 2003; Tsantrizos *et al.*, 2003), NS5B полимеразы (Tomei *et al.*, 2004), а также ингибиторы инициации трансляции генома ВГС (Tallet-Lopez *et al.*, 2003). Клеточная линия Huh-7, способная реплицировать субгеномный фрагмент ВГС (РНК репликон ВГС, кодирующий неструктурные белки NS2-NS5B), является удобной модельной системой для изучения действия анти-ВГС препаратов. После трансфекции Huh-7 клеток репликоном ВГС концентрация репликона в клетках растет и через 2–3 дня выходит на стационарный уровень (470–5000 молекул репликона на клетку). Стационарная концентрация репликона в клетках сохраняется в процессе культивирования клеток (Lohmann

*et al.*, 1999; Quinkert *et al.*, 2005). Новые анти-ВГС препараты тестируют по эффективности подавления стационарной концентрации РНК репликона ВГС в клетках.

Математическое моделирование может быть полезным инструментом для изучения действия противовирусных препаратов на такие сложные системы, как вирус–клетка. Математическая модель позволяет численно исследовать поведение молекулярно-генетической системы, проводить оценки действия потенциальных противовирусных препаратов, выявлять оптимальные мишени действия препаратов, а также вырабатывать оптимальную терапевтическую стратегию. Одной из проблем при разработке математической модели репликации репликона ВГС в клетке является описание предполагаемых механизмов, ответственных за установление стационарного уровня репликона, так как эти механизмы до сих пор не ясны. Были получены две разные математические модели репликации репликона ВГС в клетке (Dahari *et al.*, 2007; Mishchenko *et al.*, 2007). В модели (Dahari *et al.*, 2007) фактором, ограничивающим репликацию репликона ВГС, является количество активных рибосом (0,1 % общего количества рибосом в эукариотической клетке). В модели, описанной нами ранее (Mishchenko *et al.*, 2007), в качестве фактора, лимитирующего репликацию репликона ВГС, был взят некоторый обобщенный клеточный фактор  $F$ , критичный для формирования активной репликазы ВГС. Экспериментальные данные подтверждают участие некоторых клеточных факторов в формировании этой репликазы, включающей менее 0,1 % всех неструктурных белков ВГС (Gao *et al.*, 2004; Quinkert *et al.*, 2005; Randall *et al.*, 2007). Кроме того, ранее сконструированная модель (Mishchenko *et al.*, 2007) включает действие потенциальных анти-ВГС препаратов – специфических ингибиторов вирусных ферментов NS3/NS4A протеазы и NS5B полимеразы.

IRES (internal ribosomal entry site) область генома ВГС высококонсервативна и критична для инициации трансляции (Sarnov, 2003). В последнее время с целью получения потенциальных анти-ВГС препаратов исследуются антисенс-олигомеры к доменам IRES, так как они способны блокировать инициацию трансляции генома ВГС (Tallet-Lopez *et al.*, 2003;

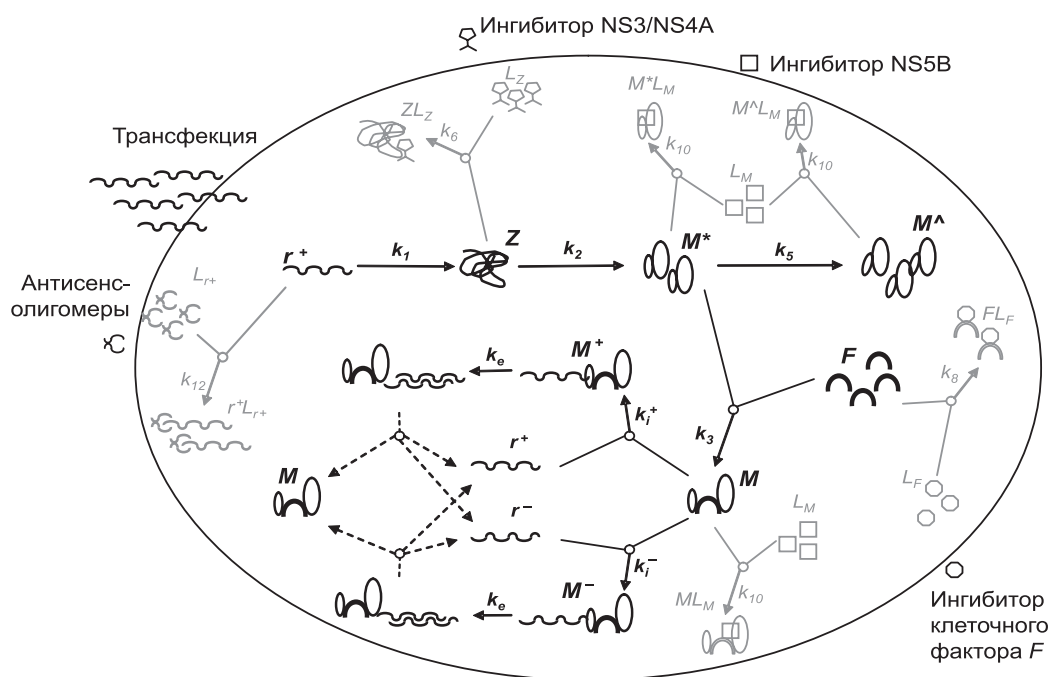
Kikuchi *et al.*, 2005). Препараты такого типа могут оказаться весьма перспективными, так как в отличие от ингибиторов вариабельных вирусных ферментов вирус не будет проявлять устойчивости к антисенс-олигомерам IRES. Модель (Mishchenko *et al.*, 2007) чувствительна к параметру, контролирующему скорость трансляции репликона. В этой связи представляло интерес ввести в модель описание механизма действия ингибиторов инициации трансляции и исследовать действие этого типа ингибиторов на репликацию репликона ВГС в клетке.

Математическая модель, представленная в данной работе, расширена введением механизма действия ингибиторов инициации трансляции. Рассчитаны кинетики РНК ВГС и других важных вирусных компонент в присутствии препаратов этого типа. Для ингибиторов разных типов (NS3/NS4A протеазы, NS5B полимеразы и инициации трансляции) оценено минимальное время, необходимое для удаления репликона из клетки, в зависимости от аффинности ингибитора к мишени. NS5B полимеразы определена как наиболее результативная мишень действия потенциальных анти-ВГС препаратов.

## Материалы и методы

**Описание метода.** Для построения математической модели и выполнения расчетов использовался пакет MGSMoeller (Kazantsev *et al.*, 2008). Оценку параметров проводили, как описано в подразделе «Оценка параметров». Концентрации клеточных компонент выражали числом молекул в клетке.

**Описание модели.** В математическую модель (Mishchenko *et al.*, 2007) был дополнительно включен механизм действия ингибиторов инициации трансляции (рис. 1). Модель была построена на основании следующих основных предположений: трансляция РНК репликона ВГС с образованием полипротеина ( $Z$ ) происходит с константой скорости  $k_1$ ,  $Z$  деградирует с константой скорости  $\mu_z$ , уравнение (1). Процессинг  $Z$  приводит к образованию неактивной репликазы ( $M^*$ ) с константой скорости  $k_2$ , уравнения (1) и (2). Интегральный клеточный фактор ( $F$ ) взаимодействует с  $M^*$  с константой скорости  $k_3$  с образованием минорной фракции активной репликазы ( $M$ ), способной реплици-



$r^+$  — плюс-цепь РНК репликона ВГС ( $r^+$ );  $r^-$  — минус-цепь РНК репликона ВГС ( $r^-$ );  $Z$  — полипротеин ( $Z$ );  
 $M^*$  — неактивная репликаза ( $M^*$ );  $M^\wedge$  — форма неактивной репликазы ( $M^\wedge$ ), не способная взаимодействовать с клеточным фактором  $F$ ;  
 $M$  — активная репликаза ( $M$ );  $F$  — клеточный фактор ( $F$ );  $M^+$  и  $M^-$  — супрамолекулярный комплекс ( $M^+$ ) или ( $M^-$ ) плюс- или минус-цепи РНК со свободным сайтом инициации и активной репликазы

**Рис. 1.** Схема процессов, включенных в модель репликации субгеномного репликона ВГС в Huh-7 клетке в присутствии потенциальных анти-ВГС препаратов.

ровать плюс- и минус-цепи РНК ( $r^+$  или  $r^-$ , соответственно). Кроме того,  $M$  может диссоциировать на  $F$  и  $M^*$  с константой скорости  $k_4$ , в то же время  $M^*$  может необратимо превращаться в форму ( $M^\wedge$ ), не способную взаимодействовать с  $F$ , с константой скорости  $k_5$ , уравнения (2)–(4). Мы предполагали, что  $F$  экспрессируется в клетке с константой скорости  $k_{in}$  и деградирует с константой скорости  $k_{out}$ , уравнение (9);  $M^*$ ,  $M$ , и  $M^\wedge$  деградируют с константой скорости  $\mu_M$ , уравнения (2)–(4),  $r^+$  и  $r^-$  деградируют с константами скорости  $\mu_{r^+}$  и  $\mu_{r^-}$  соответственно, уравнения (5) и (7). Мы также предполагали, что репликация плюс- или минус-цепей РНК иницируется при взаимодействии  $M$  с плюс- или минус-цепями РНК, имеющими свободные сайты инициации ( $R^+$  или  $R^-$ , соответственно), с образованием продуктивных супрамолекулярных комплексов ( $M^+$ ) или ( $M^-$ ) с константами скорости  $k_i^+$  и  $k_i^-$ , соответственно, уравнения

(4), (6) и (8).  $M^+$  и  $M^-$  осуществляют элонгацию репликации с константой скорости  $k_e$ , уравнения (4)–(8). Уравнения (1)–(9) при  $L_z = 0$ ,  $L_M = 0$ ,  $L_F = 0$ ,  $L_{r^+} = 0$  описывают кинетику репликации репликона ВГС в Huh-7 клетке при отсутствии анти-ВГС препаратов.

**Учет механизмов действия ингибиторов в модели.** Ранее (Mishchenko *et al.*, 2007) нами были описаны механизмы действия специфических ингибиторов NS3/NS4A протеазы, NS5B полимеразы и потенциального клеточного фактора  $F$  на репликацию репликона ВГС в клетке.

В модель дополнительно было введено описание механизма действия ингибиторов инициации трансляции. Известно, что антисенс-олигомеры к 3' домену IRES ВГС способны конкурировать с 40S субъединицей рибосомы за связывание 3' домена и, образуя комплекс с мишенью, блокировать IRES-зависимую инициацию трансляции

(Tallet-Lopez *et al.*, 2003). Мы предположили, что при взаимодействии этого типа ингибитора  $L_{r^+}$  с плюс-цепью РНК  $r^+$  образуется нетранслируемый комплекс  $r^+L_{r^+}$  с константой скорости  $k_{12}$ , который способен реплицироваться с образованием  $r^-$ . Комплекс  $r^+L_{r^+}$  диссоциирует на  $r^+$  и  $L_{r^+}$ , уравнения (5), (18), (19). Эффект ингибиторов инициации трансляции рассчитывали при концентрации  $L_{r^+} = 100$  мсМ или  $6 \times 10^7$  молекул в клетке. Концентрацию РНК ВГС при действии ингибитора инициации трансляции выражали как  $(r^+ + L_{r^+} r^+)$ .

Минимальное время, необходимое для необратимого подавления репликации РНК ВГС

в клетке в присутствии ингибиторов NS3/NS4A протеазы, NS5B полимеразы и инициации трансляции, рассчитывали при концентрации ингибиторов 100 мсМ.

Мы предполагали, что приток ингибиторов в клетку постоянен и начинается после выхода концентрации РНК ВГС на стационарный уровень (20 дней после трансфекции), нуклеозидтрифосфаты, аминокислоты, АТФ имеются в клетке в достаточном количестве. Математическая модель действия потенциальных анти-ВГС препаратов на репликацию субгеномного репликаона ВГС включает 19 дифференциальных уравнений:

$$\frac{dZ}{dt} = k_1 \cdot r^+ - k_2 \cdot Z - \mu_Z \cdot Z - k_6 \cdot L_z \cdot Z + k_7 \cdot ZL_z \quad (1)$$

$$\begin{aligned} \frac{dM^*}{dt} = & k_2 \cdot Z - k_5 \cdot M^* - k_3 \cdot M^* \cdot F + k_4 \cdot M - \\ & - \mu_M \cdot M^* - k_{10} \cdot L_M \cdot M^* + k_{11} \cdot M^* L_M \end{aligned} \quad (2)$$

$$\frac{dM^\wedge}{dt} = k_5 \cdot M^* - \mu_M \cdot M^\wedge - k_{10} \cdot L_M \cdot M^\wedge + k_{11} \cdot M^\wedge L_M \quad (3)$$

$$\begin{aligned} \frac{dM}{dt} = & k_3 \cdot M^* \cdot F - k_4 \cdot M - k_i^+ \cdot M \cdot R^+ - k_i^- \cdot M \cdot R^- + k_e \cdot M^+ + \\ & + k_e \cdot M^- - \mu_M \cdot M - k_{10} \cdot L_M \cdot M + k_{11} \cdot ML_M \end{aligned} \quad (4)$$

$$\frac{dr^+}{dt} = k_e \cdot M^- - \mu_r \cdot r^+ - k_{12} \cdot L_{r^+} \cdot r^+ + k_{13} \cdot r^+ L_{r^+} \quad (5)$$

$$\frac{dM^+}{dt} = k_i^+ \cdot M \cdot R^+ - k_e \cdot M^+ - \mu_M \cdot M^+ \quad (6)$$

$$\frac{dr^-}{dt} = k_e \cdot M^+ - \mu_r \cdot r^- \quad (7)$$

$$\frac{dM^-}{dt} = k_i^- \cdot M \cdot R^- - k_e \cdot M^- - \mu_M \cdot M^- \quad (8)$$

$$\begin{aligned} \frac{dF}{dt} = & k_{in} - k_{out} \cdot F - k_3 \cdot F \cdot (M^* + M^* L_M) + \\ & + k_4 \cdot (M + ML_M) - k_8 \cdot L_F \cdot F + k_9 \cdot FL_F \end{aligned} \quad (9)$$

$$\frac{dL_z}{dt} = k_{L_z in} - k_{L_z out} \cdot L_z - k_6 \cdot L_z \cdot Z + k_7 \cdot ZL_z \quad (10)$$

$$\frac{dZL_z}{dt} = k_6 \cdot L_z \cdot Z - k_7 \cdot ZL_z - \mu_{ZL_z} \cdot ZL_z \quad (11)$$

$$\frac{dL_F}{dt} = k_{L_F in} - k_{L_F out} \cdot L_F - k_8 \cdot L_F \cdot F + k_9 \cdot FL_F \quad (12)$$

$$\frac{dFL_F}{dt} = k_8 \cdot L_F \cdot F - k_9 \cdot FL_F - \mu_{FL_F} \cdot FL_F \quad (13)$$



$$\frac{dL_M}{dt} = k_{L_{M_{in}}} - k_{L_{M_{out}}} \cdot L_M - k_{10} \cdot L_M \cdot (M + M^* + M^\wedge) + k_{11} \cdot (ML_M + M^*L_M + M^\wedge L_M) \quad (14)$$

$$\frac{dM^*L_M}{dt} = k_{10} \cdot L_M \cdot M^* - k_{11} \cdot M^*L_M - k_5 \cdot M^*L_M - k_3 \cdot M^*L_M \cdot F + k_4 \cdot ML_M - \mu_{ML_M} \cdot M^*L_M \quad (15)$$

$$\frac{dM^\wedge L_M}{dt} = k_{10} \cdot L_M \cdot M^\wedge - k_{11} \cdot M^\wedge L_M + k_5 \cdot M^*L_M - \mu_{ML_M} \cdot M^\wedge L_M \quad (16)$$

$$\frac{dML_M}{dt} = k_{10} \cdot L_M \cdot M - k_{11} \cdot ML_M + k_3 \cdot M^*L_M \cdot F - k_4 \cdot ML_M - \mu_{ML_M} \cdot ML_M \quad (17)$$

$$\frac{dL_{r^+}}{dt} = k_{L_{r^+_{in}}} - k_{L_{r^+_{out}}} \cdot L_{r^+} - k_{12} \cdot L_{r^+} \cdot r^+ + k_{13} \cdot r^+ L_{r^+} \quad (18)$$

$$\frac{dr^+ L_{r^+}}{dt} = k_{12} \cdot L_{r^+} \cdot r^+ - k_{13} \cdot r^+ L_{r^+} - \mu_r \cdot r^+ L_{r^+} \quad (19)$$

Принятые обозначения концентраций:  $Z$  – полипротеин;  $M$  – активная репликаза;  $M^*$  – неактивная репликаза;  $M^\wedge$  – форма неактивной репликасы, не способная переходить в активную репликасу;  $F$  – клеточный фактор;  $r^+$  ( $r^-$ ) – плюс- (минус-) цепь РНК;  $M^+$  ( $M^-$ ) – супрамолекулярный комплекс плюс- (минус)- цепи РНК со свободным сайтом инициации и активной репликасы;  $L_Z$  – ингибитор NS3/NS4A протеазы;  $ZL_Z$  – комплекс «полипротеин– $Z$ -ингибитор  $L_Z$ »;  $L_F$  – ингибитор клеточного фактора  $F$ ;  $FL_F$  – комплекс «клеточный фактор– $F$ -ингибитор  $L_F$ »;  $L_M$  – неконкурентный ингибитор NS5B полимеразы;  $ML_M$ ,  $M^*L_M$  и  $M^\wedge L_M$  – комплексы ингибитора  $L_M$  с формами репликасы  $M$ ,  $M^*$ ,  $M^\wedge$  соответственно.  $R^+$  или  $R^-$  – средняя концентрация плюс- или минус-цепи РНК ( $r^+$  или  $r^-$ ) со свободным сайтом инициации (см. формулы ниже) в предположении, что активные репликасы ( $M^+$  или  $M^-$ ) равномерно распределены по всей длине цепи.  $N$  – длина РНК ВГС (~ 8,000 нуклеотидов),  $n$  – размер сайта связывания активной репликасы с РНК (~ 100 нуклеотидов).  $R^+$  и  $R^-$  рассчитывали по формулам:

$$R^+ = (r^+ + r^+ L_{r^+}) \cdot [1 - (n \cdot M^+) / (N \cdot (r^+ + r^+ L_{r^+}))]$$

$$R^- = r^- \cdot [1 - (n \cdot M^-) / (N \cdot r^-)].$$

Константы скоростей процессов приведены в табл. 1 (см. Оценка параметров) и табл. 1–4 (Mishchenko *et al.*, 2007).

**Оценка параметров.** Верификация параметров проведена так, как описано в работе Е. Мищенко с соавт. (Mishchenko *et al.*, 2007): параметры  $k_1$ ,  $k_2$ ,  $k_e$ ,  $\mu_{r^+}$ ,  $\mu_{r^-}$ ,  $\mu_{M^*}$ ,  $k_6$ ,  $k_7$ ,  $k_{10}$ ,  $k_{11}$  были верифицированы на основе литературных данных вручную; параметры  $k_3$ ,  $k_4$ ,  $k_5$ ,  $k_i^+$ ,  $k_i^-$ ,  $k_m$  и  $k_{out}$  были подобраны так, чтобы модель соответствовала а) времени выхода вирусных белков и РНК на стационарные концентрации, б) значениям стационарных концентраций белков и РНК, в) кинетике подавления стационарного уровня вирусной РНК в присутствии ингибитора NS3/NS4A протеазы. При введении в модель описания механизма действия ингибиторов инициации трансляции были сделаны следующие оценки параметров: константы скорости  $k_{12}$  (образование комплексов  $r^+ L_{r^+}$ ) были приняты равными константе скорости образования комплекса «РНК–аптамер»  $5,9 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$  (Kikuchi *et al.*, 2005). Константы скорости  $k_{13}$  (диссоциация комплексов  $r^+ L_{r^+}$ ) были определены из значений  $K_d$  (табл. 1).

## Результаты

Представленная модель описывает механизмы действия ряда потенциальных противовирусных препаратов, таких, как ингибитор NS3/NS4A протеазы, ингибитор NS5B полимеразы и ингибитор инициации трансляции, что позволяет оценить их влияние на репликацию репликаона ВГС.

Таблица 1

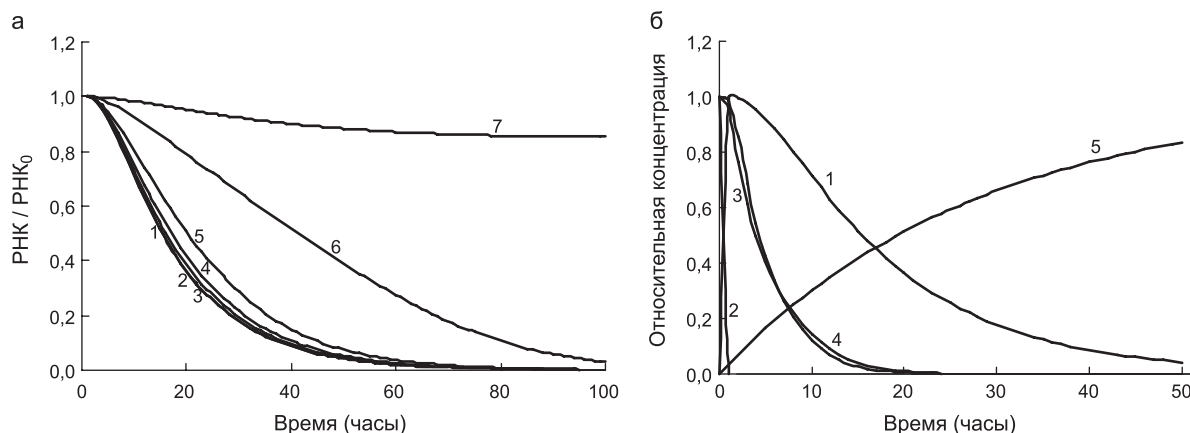
Оценки констант скоростей для ингибитора инициации трансляции и его комплексов с плюс-цепью РНК ВГС в Huh-7 клетке, несущей репликон ВГС

Константа скорости	Процесс	Значение [s <sup>-1</sup> ]
$k_{12}$	Образование комплекса $r^+L_{r+}$	$0,88 \times 10^{-7}$ [молекула <sup>-1</sup> ]
$k_{13}$	Диссоциация комплекса $r^+L_{r+}$	$k_{12} \times K_d^a$
$\mu_{r+L}$	Деградация комплекса $r^+L_{r+}$	$2 \times 10^{-5}$
$k_{Lr+in}$	Приток $L_{r+}$	$L_{r+} \times k_{Lr+out}^b$
$k_{Lr+out}$	Деградация $L_{r+}$	$10^{-5}$

<sup>a</sup> Определено для различных  $r^+L_{r+}$  комплексов как  $k_{12} \times K_d$ ; <sup>b</sup> определено как  $k_{Lr+out} \times L_{r+} = 100$  мсМ.

**Расчет динамик концентраций компонент ВГС в клетке в присутствии ингибиторов инициации трансляции.** Динамики концентраций вирусных компонент (РНК, полипротеина, активной репликазы, супрамолекулярных комплексов плюс- и минус-цепей РНК и активной репликазы) в клетке в присутствии ингибиторов инициации трансляции ( $L_r$ ) рассчитывали при значениях константы диссоциации нетранслируемого комплекса «плюс-цепь РНК–ингибитор»,  $K_d = 0,074 - 640$  нМ. Расчеты показали (рис. 2, а), что динамика уменьшения стационарной концентрации РНК ВГС (в виде суммарной концент-

рации свободной и связанной с ингибитором плюс-цепи РНК) для высокоаффинных ингибиторов  $L_r$  ( $K_d = 0,074 - 3$  нМ) была эффективной и практически не зависела от  $K_d$  (время уменьшения концентрации РНК на 50%,  $t_{1/2} = 16,1 - 16,8$  ч). Динамики изменения стационарной концентрации РНК для ингибиторов  $L_r$  меньшей аффинности ( $K_d = 7,5 - 120$  нМ) коррелировали с аффинностью  $L_r$  к мишени ( $t_{1/2} = 17,4 - 43,5$  ч). Низкоаффинный ингибитор  $L_r$  ( $K_d = 640$  нМ) незначительно влиял на стационарную концентрацию РНК. Экспериментально действие ингибиторов инициации трансляции было изучено на Huh-7



**Рис. 2.** Рассчитанные кинетики изменения концентраций компонент ВГС в Huh-7 клетке в присутствии ингибиторов инициации трансляции (100 мсМ).

Значения концентраций приведены в относительных единицах (нормализованы к максимальной концентрации). а – кинетики подавления стационарной концентрации РНК репликона ВГС в присутствии ингибиторов: 1 – I ( $K_d = 0,074$  нМ); 2 – II ( $K_d = 0,24$  нМ); 3 – III ( $K_d = 3$  нМ); 4 – IV ( $K_d = 7,5$  нМ); 5 – V ( $K_d = 21$  нМ); 6 – VI ( $K_d = 120$  нМ); 7 – VII ( $K_d = 640$  нМ); б – кинетики изменения концентраций компонент ВГС в присутствии высокоаффинного ингибитора инициации трансляции ( $K_d = 0,074$  нМ): 1 – образования нетранслируемого комплекса «ингибитор–плюс-цепь РНК ( $r^+L_r$ )»; изменения стационарных концентраций: 2 – свободной плюс-цепи РНК ( $r^+$ ); 3 – активной репликазы ( $M$ ); 4 – супрамолекулярного комплекса плюс-цепи РНК, имеющей свободный сайт инициации, и активной репликазы ( $M^+$ ); 5 – кинетика равновесной концентрации ингибитора инициации трансляции ( $K_d = 0,074$  нМ) в клетке (в относительных единицах).

клетках, трансфицированных РНК транскриптом, содержащим IRES ВГС и кодирующую последовательность люциферазы *Renilla* (Tallet-Lopez *et al.*, 2003), поэтому мы не могли сравнить расчетные и экспериментальные данные.

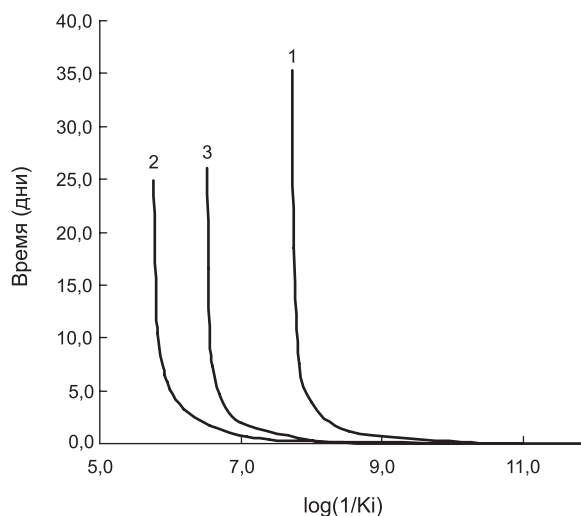
Динамики концентраций других важных вирусных компонент (связанной в комплекс с ингибитором плюс-цепи РНК  $L_r r^+$ , свободной плюс-цепи РНК  $r^+$ , полипротеина  $Z$ , активной репликазы  $M$ , супрамолекулярных комплексов плюс- и минус-цепей РНК и активной репликазы  $M^+$  и  $M^-$ ) были рассчитаны для высокоаффинного ингибитора  $L_r$  ( $K_d = 0,074$  нМ). Расчеты показали (рис. 2, б), что одновременно с образованием нетранслируемого комплекса  $L_r r^+$ , концентрация которого увеличивалась в течение 1 ч, происходило падение стационарной концентрации  $r^+$  ( $t_{1/2} = 0,4$  ч). Изменение стационарных концентраций  $M$  и  $M^+$  было более медленным ( $M$ ,  $t_{1/2} = 4,2$  ч;  $M^+$ ,  $t_{1/2} = 4,6$  ч). Динамика падения стационарной концентрации  $Z$  была такой же, как динамика падения  $r^+$ , динамика падения стационарной концентрации  $M^-$  была такой же, как динамика падения  $M^+$ .

**Расчет минимального времени, необходимого для удаления РНК ВГС из клетки в присутствии ингибиторов, направленных на вирусные мишени, NS3/NS4A протеазу, NS5B полимеразу и IRES область генома.** Минимальное время, необходимое для необратимого подавления репликации РНК ВГС в клетке, рассчитывали для ингибиторов разных типов в зависимости от аффинности ингибиторов к мишеням. Расчеты показали (рис. 3), что для высокоаффинных ингибиторов NS3/NS4A протеазы ( $K_i = 1$  пМ – 0,63 нМ), NS5B полимеразы ( $K_i = 1$  пМ – 31 нМ) и инициации трансляции ( $K_i = 1$  пМ – 10 нМ) это время очень мало (около 1 ч) и практически не зависит от  $K_i$ . Уменьшение аффинности ингибиторов: NS3/NS4A протеазы ( $K_i = 0,63$  нМ – 12 нМ), NS5B полимеразы ( $K_i = 31$  нМ – 2,0 мсМ) и инициации трансляции ( $K_i = 10$  нМ – 200 нМ) приводило к увеличению этого параметра от 1 ч до 5 дней. Дальнейшее сравнительно небольшое уменьшение аффинности ингибиторов: NS3/NS4A протеазы ( $K_i = 12$  нМ – 16 нМ), NS5B полимеразы ( $K_i = 2,0$  мсМ – 2,5 мсМ) и инициации трансляции ( $K_i = 200$  нМ – 250 нМ) вызывало резкое возрастание минимального времени от 5 дней до бесконечности.

## Обсуждение

Препятствием для получения эффективных анти-ВГС препаратов является высокая гетерогенность и мутационная скорость ВГС. Мутации в связывающих центрах вирусных ферментов NS3/NS4A протеазы и NS5B полимеразы способны значительно уменьшать аффинность ингибиторов к этим мишеням, приводя к устойчивости ВГС к потенциальным противовирусным препаратам (Lu *et al.*, 2004; Kukulj *et al.*, 2005). Устойчивые варианты вируса могут возникать даже в течение 14–25 дней применения препаратов (Trozzi *et al.*, 2003; Lu *et al.*, 2004), что соизмеримо с однократным курсом применения одного препарата.

Антисенс-олигомеры к консервативным доменам IRES ВГС способны блокировать IRES-зависимую инициацию трансляции генома ВГС и могут быть перспективными анти-ВГС препаратами, так как вирус не будет проявлять к ним устойчивость (Tallet-Lopez *et al.*, 2003; Kikuchi *et al.*, 2005). При введении описания механизма действия такого типа ингибиторов в математическую модель мы предположили, что при взаимодействии ингибитора с плюс-цепью РНК ВГС образуется нетранслируемый комплекс, который, как и свободная плюс-цепь



**Рис. 3.** Зависимость минимального времени, необходимого для необратимого подавления репликации РНК ВГС в клетке, от аффинности ингибиторов (при концентрации 100 мсМ).

1 – NS3/NS4A протеазы; 2 – NS5B полимеразы; 3 – инициация трансляции.

РНК, способен реплицироваться с образованием минус-цепи РНК. Оценки показали, что эффективность подавления стационарного уровня РНК ВГС (в виде суммарной концентрации свободной и связанной в комплексе с ингибитором плюс-цепи) в присутствии ингибиторов определяется аффинностью ингибиторов к мишени (рис. 1, а). К сожалению, экспериментально действие олигонуклеотидов, комплементарных IIIId домену IRES ВГС, было изучено по подавлению экспрессии люциферазы *Renilla* в Huh-7 клетках, трансфицированных РНК транскриптом, содержащим IRES ВГС и кодирующую последовательность люциферазы *Renilla* (Tallet-Lopez *et al.*, 2003), поэтому мы не могли сравнить расчетные и экспериментальные данные.

Расчеты временных изменений вирусных компонент в присутствии ингибиторов важны для получения более полной динамической картины подавления вируса в клетке и выработки оптимального режима применения ингибиторов. Расчеты показали (рис. 2, б), что за 1 час действия высокоаффинного ингибитора инициации трансляции ( $K_i = 74$  pM) в концентрации 100 мсМ концентрация свободной мишени падала практически до нуля. Следовательно, после 1 часа действия ингибитора можно уменьшить его концентрацию или даже прекратить его поступление в клетку.

Как известно, токсический эффект противовирусных препаратов на клетку снижается при уменьшении как концентрации, так и времени действия препаратов. С целью более точной оценки времени действия ингибиторов инициации трансляции, необходимого для необратимого подавления репликации репликаона ВГС, была рассчитана зависимость этого параметра от аффинности ингибитора к мишени (рис. 3, график 3). Оценки показали, что время, необходимое для подавления инфекции для ингибиторов с  $K_i = 1$  pM – 200 nM, изменяется от 1 ч до 5 дней, а ингибиторы с  $K_i \geq 250$  nM не способны излечить клетку (минимальное время лечения клетки бесконечно). Аналогичные расчеты были произведены для ингибиторов, действующих на NS3/NS4A протеазу и NS5B полимеразу. Анализ полученных зависимостей (рис. 3, графики 1–3) показал следующее: низкоаффинные и неспособные подавить вирусную

инфекцию ингибиторы, действующие на IRES ВГС (время подавления инфекции бесконечно), способны подавить эту инфекцию при действии на мишень NS5B (время подавления инфекции – несколько дней); низкоаффинные и неспособные подавить вирусную инфекцию ингибиторы, действующие на NS3/NS4A протеазу, могут быстро подавить инфекцию при действии на мишени IRES ВГС и NS5B полимеразу (время подавления инфекции около 1 часа). Таким образом, оценки показывают, что результативность вирусных мишеней падает в ряду NS5B полимеразы, IRES ВГС, NS3/NS4A протеазы.

Оценивая результативность действия ингибиторов на мишени ВГС, важно учитывать, что NS5B полимеразы вариабельна, а мишень IRES ВГС, напротив, высококонсервативна. Так, структура домена IIIId IRES ВГС абсолютно консервативна среди всех генотипов ВГС (Collier *et al.*, 1998). Поэтому ингибиторы инициации трансляции (например высокоаффинные антисенс-олигомеры к этому домену) могут оказаться наиболее перспективными противовирусными препаратами, так как будут характеризоваться достаточно высокой противовирусной активностью и отсутствием резистентности к ним вируса.

### Благодарности

Работа выполнена при поддержке проекта № 10.7 «Компьютерное моделирование и экспериментальное конструирование генных сетей» по программе РАН «Молекулярная и клеточная биология», НШ-2447.2008.4. Научная школа Н.А. Колчанова «Биоинформатика и системная компьютерная биология».

### Литература

- Chou R., Clark E.C., Helfand M. Screening for hepatitis C virus infection: a review of the evidence for the U.S. Preventive Services Task Force // *Ann. Intern. Med.* 2004. V. 140. P. 465–479.
- Collier A., Tang S., Elliott R.M. Translation efficiencies of the 5' untranslated region from representatives of the six major genotypes of hepatitis C virus using a novel bicistronic reporter assay system // *J. Gen. Virol.* 1998. V. 79. P. 2359–2366.
- Dahari H., Ribeiro R.M., Rice C.M., Perelson A.S. Ма-

- thematical modeling of subgenomic hepatitis C viral replication in Huh-7 cells // *J. Virol.* 2007. V. 81. P. 750–760.
- Dudley T., O'Donnell K., Haydon G., Mutimer D. Disappointing results of combination therapy for HCV? // *Gut.* 2006. V. 55. P. 1362–1363.
- El-Gogo S., Staib C., Lasarte J.J. *et al.* Protective vaccination with hepatitis C virus NS3 but not core antigen in a novel mouse challenge model // *J. Gene Med.* 2008. V. 10. P. 177–186.
- Gao L., Aizaki H., He J.W., Lai M.M. Interactions between viral nonstructural proteins and host protein hVAP-33 mediate the formation of hepatitis C virus RNA replication complex on lipid raft // *J. Virol.* 2004. V. 78. P. 3480–3488.
- Hoofnagle J.H. Course and outcome of hepatitis C // *Hepatology.* 2002. V. 36. P. S21–S29.
- Kazantsev F.V., Akberdin I.R., Bezmaternykh K.D. *et al.* MGSmodeller – a computer system for reconstruction, calculation and analysis of mathematical models of molecular genetic system // *Proc. of the 6th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008).* 2008. P. 113.
- Kikuchi K., Umehara T., Fukuda K. *et al.* A hepatitis C virus (HCV) internal ribosome entry site (IRES) domain III-IV-targeted aptamer inhibits translation by binding to an apical loop of domain IIIId // *Nucl. Acids Res.* 2005. V. 33. P. 683–692.
- Kukolj G., McGibbon G.A., McKercher G. *et al.* Binding site characterization and resistance to a class of non-nucleoside inhibitors of the hepatitis C virus NS5B polymerase // *J. Biol. Chem.* 2005. V. 280. P. 39260–39267.
- Lohmann V., Korner F., Koch J. *et al.* Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line // *Science.* 1999. V. 285. P. 110–113.
- Lu L., Pilot-Matias T.J., Stewart K.D. *et al.* Mutations conferring resistance to a potent hepatitis C virus serine protease inhibitor *in vitro* // *Antimicrob. Agents Chemother.* 2004. V. 48. P. 2260–2266.
- Mishchenko E.L., Bezmaternykh K.D., Likhoshvai V.A. *et al.* Mathematical model for suppression of subgenomic hepatitis C virus RNA replication in cell culture // *J. Bioinform. Comput. Biol.* 2007. V. 5. P. 593–609.
- Pause A., Kukolj G., Bailey M. *et al.* An NS3 serine protease inhibitor abrogates replication of subgenomic hepatitis C virus RNA // *J. Biol. Chem.* 2003. V. 278. P. 20374–20380.
- Quinkert D., Bartenschlager R., Lohmann V. Quantitative analysis of the hepatitis C virus replication complex // *J. Virol.* 2005. V. 79. P. 13594–13605.
- Randall G., Panis M., Cooper J.D. *et al.* Cellular cofactors affecting hepatitis C virus infection and replication // *Proc. Natl Acad. Sci. USA.* 2007. V. 104. P. 12884–12889.
- Sarnow P. Viral internal ribosome entry site elements: novel ribosome-RNA complexes and roles in viral pathogenesis // *J. Virol.* 2003. V. 77. P. 2801–2806.
- Tallet-Lopez B., Aldaz-Carroll L., Chabas S. *et al.* Antisense oligonucleotides targeted to the domain IIIId of the hepatitis C virus IRES compete with 40S ribosomal subunit binding and prevent *in vitro* translation // *Nucl. Acids Res.* 2003. V. 31. P. 734–742.
- Tomei L., Altamura S., Bartholomew L. *et al.* Characterization of the inhibition of hepatitis C virus RNA replication by nonnucleosides // *J. Virol.* 2004. V. 78. P. 938–946.
- Trozzi C., Bartholomew L., Ceccacci A. *et al.* *In vitro* selection and characterization of hepatitis C virus serine protease variants resistant to an active-site peptide inhibitor // *J. Virol.* 2003. V. 77. P. 3669–3679.
- Tsantrizos Y.S., Bolger G., Bonneau P. *et al.* Macrocyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection // *Angew. Chem.* 2003. V. 42. P. 1355–1360.



**MATHEMATICAL MODELING OF EFFECTS OF POTENTIAL ANTIVIRALS  
ON SUBGENOMIC HEPATITIS C VIRUS REPLICON REPLICATION  
IN A CELL**

**E.L. Mishchenko<sup>1</sup>, K.D. Bezmaternykh<sup>1,2</sup>, V.A. Ivanisenko<sup>1</sup>,  
V.A. Likhoshvai<sup>1,2</sup>, N.A. Kolchanov<sup>1,2</sup>**

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: elmish@bionet.nsc.ru;

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

**Summary**

Mathematical modeling is an advantageous tool in studies of drug effects on complex systems such as the virus–cells. We have developed the mathematical model that describes the effects of drugs on the hepatitis C virus RNA replicon replication in Huh-7 cells. The dynamics of the viral components (RNA, polyprotein, active replicase, the supramolecular complexes of the plus- and minus-RNA strands and active replicase) in the cell in the presence of potential drugs acting as translation initiation inhibitors was calculated. For inhibitors of different types – NS3/NS4A protease, NS5B polymerase, and translation initiation, the dependencies of the minimal time for HCV replicon clearance from the cell on the affinity of the inhibitor for the target was studied. Comparison of these dependencies has demonstrated that NS5B polymerase is the most efficient target for the effects of these drugs.

## ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО АНАЛИЗА МИКРОИЗОБРАЖЕНИЙ ЛИСТА ДЛЯ ОЦЕНКИ ХАРАКТЕРИСТИК ОПУШЕНИЯ ПШЕНИЦЫ *TRITICUM AESTIVUM* L.

А.В. Дорошков<sup>1</sup>, С.И. Арсенина<sup>2</sup>, Т.А. Пшеничникова<sup>1</sup>, Д.А. Афонников<sup>1,2</sup>

<sup>1</sup> Учреждение Российской академии наук Институт цитологии и генетики  
Сибирского отделения РАН, Новосибирск, Россия, e-mail: ada@bionet.nsc.ru;

<sup>2</sup> Новосибирский государственный университет, Новосибирск, Россия

Опушение листа у пшеницы имеет большое биологическое значение при адаптации к факторам внешней среды. Изучение особенностей морфологии опушения, а также выявление генетических факторов, ответственных за его формирование, позволят получать сорта, обладающие устойчивостью к суровым климатическим условиям, а также невосприимчивостью к ряду вредителей. Важным шагом в этом направлении является использование компьютерных технологий для описания количественных характеристик опушения. Для оценки характеристик опушения пшеницы в работе предложено использовать компьютерный анализ микроизображения поперечного сгиба листа растения. Результаты показали, что такая методика не требует больших временных затрат, позволяет оценивать плотность опушения, распределение трихом по длине и, таким образом, является эффективным подходом для массового анализа морфологических особенностей опушения листа у отдельных растений.

**Ключевые слова:** опушение листа, трихомы, компьютерный анализ микроизображений, *Triticum aestivum* L.

### Введение

Трихомы пшеницы представляют собой одноклеточные или многоклеточные образования эпидермиса и по классификации Солередера относятся к простым кроющим волоскам (Solereder, 1908, цит. по: Мирославов, 1974). Опушение листа у пшеницы имеет большое биологическое значение в защите от вредителей и при адаптации к факторам среды. В частности, сильное «войлочное» опушение характерно для ряда засухоустойчивых сортов, относящихся к степной экологической группе; для сортов, произрастающих во влажном климате, напротив, характерно очень слабое «мохнатое» опушение (Крупнов, Цапайкин, 1990). Показано влияние опушения покровов на влагоудерживающую способность растений яровой мягкой пшеницы *Triticum aestivum* L. (Лихенко, 2007). Кроме того, существуют данные о том, что сильно опушенные сорта пшеницы значительно более устойчивы к поражению жуком-листоедом

*Oulema melanopus* L. (Schillinger, Gallun, 1968), а также гессенской мухой *Mayetiola destructor* (Roberts *et al.*, 1979). Опушение имеет также большое физиологическое значение. Предполагается, что клетки трихом, основной объем которых вынесен за плоскость эпидермы, расширяют возможности формирования локальных градиентов тургорных давлений эпидермиса, а также снимают нарастающее в онтогенезе напряжение покровной ткани, увеличивая таким образом эффективность работы устьичного аппарата (Журавлева, 1992). Изучение генетического полиморфизма и выявление генов, ответственных за формирование опушения листа пшеницы, позволят целенаправленно получать сорта, имеющие заданный тип опушения.

Одним из перспективных направлений для решения этой задачи является использование статистических подходов для выявления взаимосвязей между генотипом и фенотипом растения, успешно используемых в генетических исследованиях *Arabidopsis thaliana* (Perazza

*et al.*, 1999). Их преимущество заключается в том, что современные методы секвенирования и выявления полиморфизма генома позволяют быстро и эффективно устанавливать генотип растения и получать информацию о его мутантных аллелях (<http://1001genomes.org/index.html>). Основным затруднением на пути к массовому использованию такого подхода в настоящее время является отсутствие эффективных методов широкомасштабного количественного анализа фенотипов для многих признаков (Baenzinger *et al.*, 2008). Типичными методами определения морфологии трихом у растений разных видов являются сканирующая электронная микроскопия (Luo, Oppenheimer, 1999; Nagata *et al.*, 1999; Perazza *et al.*, 1999), качественная визуальная оценка или визуальный подсчет при помощи линзы (Lacare, Nguyen, 2005). К недостаткам этих методов можно отнести низкую точность качественных оценок и трудоемкость методов подсчета, которые осложняют их применение для анализа большого количества образцов. Получение массовых количественных оценок морфологических характеристик растений может быть эффективным при использовании цифровых технологий. Например, анализ двумерных изображений растений был успешно использован для изучения механизмов передачи сигнала (Xiong *et al.*, 1999). Анализ трехмерных моделей растения, полученных при помощи лазерного сканирования, позволяет получить оценки направления листовой пластины и ростового искривления побега у *Arabidopsis thaliana* (Kaminuma *et al.*, 2004).

В настоящей работе для оценки характеристик опушения листа пшеницы использован метод, реализованный в программе LHDetect, основанный на компьютерном анализе изображения поперечного сгиба листа (Arsenina *et al.*, 2008). Метод позволяет по оцифрованному изображению листа, полученному с использованием световой микроскопии в проходящем свете, оценивать плотность трихом на поверхности листа, а также распределение трихом по длинам. С помощью программы LHDetect проведена оценка опушения листьев 6 сортов яровой мягкой пшеницы *Triticum aestivum* L., в число которых входили как практически неопушенные сорта, так и сорта с различными вариантами опушения. Предложенный метод

позволяет получать в цифровом виде данные как по плотности опушения, так и по распределению трихом по длине. Нами проведен анализ характеристик опушения листа для сортов пшеницы Балаганка и Голубка, неотличимых как тактильно, так и визуально по микроизображению. Показано, что эти сорта имеют близкие величины плотности опушения листа, однако различаются по распределению длин трихом. Сравнение образцов этих сортов по длинам трихом позволяет провести более четкое описание разнообразия опушения листа растений, чем сравнение его только по плотности.

## Материалы и методы

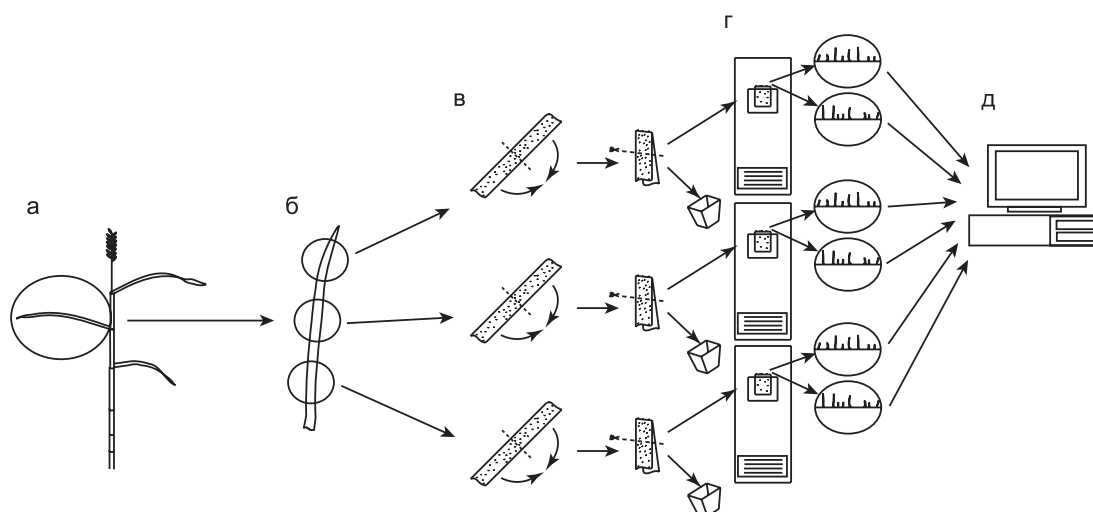
### 1. Используемые сорта

В работе было использовано 5 сортов из коллекции лаборатории хромосомной инженерии злаков (Голубка, Балаганка, Омская 9, Сibaковская 3, Дуванка 501), а также сорт Mahndorfer Viktoria, полученный из коллекции Gene Bank, IPK, Гатерслебен, Германия. Растения были выращены в полевых условиях в окрестностях Новосибирска в 2008 г.

### 2. Приготовление препаратов

Для приготовления препаратов были использованы предфлаговые листья. После сбора листья помещались в емкости с высокой влажностью для предотвращения их высыхания и последующего повреждения. В выбранной для анализа точке лист сгибался поперек таким образом, что верхняя, анализируемая, поверхность листа оказывалась снаружи (см. рис. 1). Затем при необходимости удалялись края листа и сгиб помещался на предметное стекло. Фиксация производилась покровным стеклом или липкой лентой. Такая методика позволяет получить под микроскопом изображение, на котором на контрастном фоне различимы трихомы.

Толщина сгиба по отношению к ширине листа мала. Поэтому мы предполагали, что плотность опушения (число трихом на единицу площади) прямо пропорциональна числу трихом на единицу длины сгиба листа, которое можно оценить по микроизображению.



**Рис. 1.** Основные этапы анализа опушенности листа.

а – отбор предфлаговых листьев; б – выбор точек на листе; в – приготовление микропрепаратов; г – микрофотографирование и получение цифровых изображений; д – обработка изображений и анализ данных.

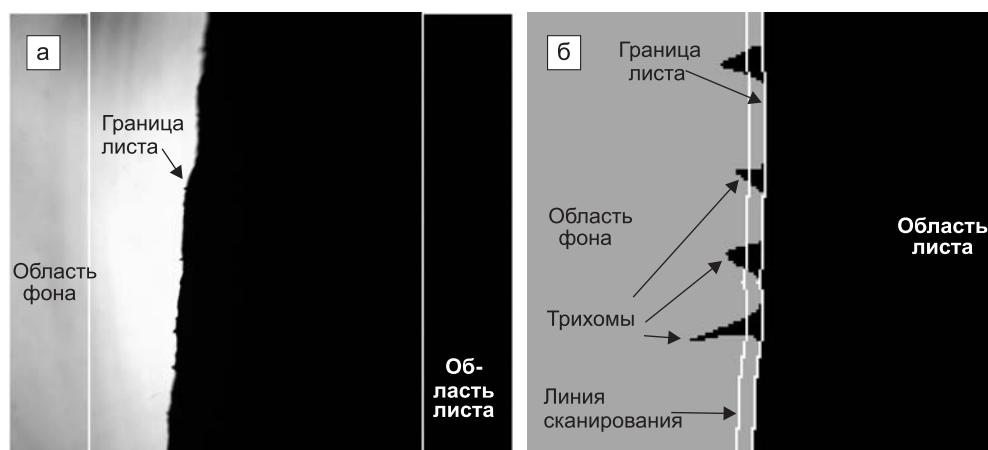
### 3. Получение изображений

Анализ препаратов производился под микроскопом Carl Zeiss Axioscop 2 plus ЦКП микроскопического анализа СО РАН, оснащенный цифровой камерой AxoCam HRc с адаптером TV2/3C 0,63x. Для обеспечения широкого поля зрения использовался объектив 5x NA=0,12. Были подобраны оптимальные условия позиционирования, освещенности листа и фона для получения микроизображений. Размер поля зрения при съемке микроизображения составил

2730 × 2163 мкм. Размер оцифрованного изображения составил 1300 × 1030 пикселей. Таким образом, высота и ширина одного пикселя на изображении соответствуют 2,1 мкм.

### 4. Обработка изображений

Алгоритм оценки степени опушенности может обрабатывать цветные изображения листа, который располагается справа, а его граница проходит приблизительно посередине изображения вертикально (рис. 2). Изображе-



**Рис. 2.** Схема работы алгоритма определения характеристик опушенности листа.

а – выбор полос для определения цветовых характеристик фона и листа; б – граница листа и работа алгоритма подсчета числа трихом.

ния подаются на вход алгоритма в формате bitmap. Работа алгоритма состоит из нескольких этапов:

- определение цветовых характеристик области фона и листа;
- определение границы листа;
- подсчет числа трихом на различной удаленности от границы листа.

Опишем принципы работы алгоритма для этих этапов.

**Определение цветовых характеристик области фона и листа.** Первый этап анализа заключается в определении принадлежности пикселя изображения листу или фону. Предварительный анализ показал, что изображения листа и фона можно хорошо различить по цвету (в том числе и по его интенсивности). Файлы в формате BMP содержат информацию об интенсивности цветовых компонент (RGB – красной, зеленой и синей) каждого пикселя изображения. Поэтому для решения первой задачи мы использовали классификацию пикселей в пространстве компонент RGB. Для этого в правой и левой части изображения определялись полосы, которые заведомо принадлежат листу и фону (рис. 2, а). Для пикселей каждой из полос определялись средние значения компонент RGB. Каждый пиксель рисунка относился к типу «фон», если в пространстве RGB его положение ближе к средним значениям компонент RGB полосы фона и к типу «лист», если в пространстве RGB его положение ближе к средним значениям компонент RGB полосы листа.

**Определение границы листа.** После классификации точек на «фон» и «лист» определялась граница листа. Для этого изображение разбивалось на горизонтальные полосы фиксированной ширины, подобранной опытным путем. В каждой из полос находилась точка фона, являющаяся крайней правой в своей полосе. Полученные в каждой полосе точки являлись узлами границы. Соединив узлы ломаной, получим линию границы листа (рис. 2, б).

**Определение числа трихом на различных расстояниях от границы листа.** Для подсчета числа трихом на разном удалении от границы листа линия границы смещалась по горизонтали от листа на 1, 2, 3, ...,  $i$ , ..., 121 количество пикселей (всего анализировалось 120 линий). Для каждого  $i$ -го смещения подсчитывалось

число пересечений линии с трихомами (т. е. число непрерывных отрезков на такой линии, пиксели которых по цветовой классификации были отнесены к типу «лист»). Это число  $s_i$  является оценкой количества трихом на удалении от поверхности листа, эквивалентном  $i$  пикселям ( $i \cdot 2,1$  мкм). На рис. 2, б представлен пример сканирования одной из линий проекции границы, которая пересекает 4 трихомы. Число трихом, имеющих на микроизображении длину  $i$ ,  $n_i$ , вычисляется как  $n_i = s_i - s_{i+1}$ . В настоящей работе мы использовали более грубую оценку распределения трихом по длине, подсчитывали числа трихом, длины которых находились в пределах  $i, i+5$ , т. е. рассчитывали  $n_i$  с шагом 5 пикселей ( $\sim 10$  мкм). Полное число трихом  $S$  на микроизображении вычислялось путем суммирования всех значений  $n_i$ . В результате такого анализа каждое изображение листа характеризовалось 25 значениями вектора  $n$  ( $i = 1, 6, 11, \dots, 121$  пикселей) и значением  $S$ , которые количественно описывают морфологию опушенности листа. Эти характеристики позволяют оценивать признак опушенности листа более точно, чем ранее использовавшиеся качественные характеристики («без опушения», «слабое опушение», «сильное опушение»), которые определялись тактильно.

**Доступность программы.** Текущая версия программы LHDetect реализована на платформе Windows32 с оконным интерфейсом и доступна по запросу авторам статьи.

## 5. Статистический анализ данных

Для сравнения средних значений числа трихом у растений пшеницы сортов Балаганка и Голубка использовался  $t$ -критерий равенства выборочных средних. Для иерархической кластеризации образцов по распределению длин трихом использовали метод UPGMA (Sneath, Sokal, 1973) и евклидово расстояние между векторами  $n$ .

## Результаты и обсуждение

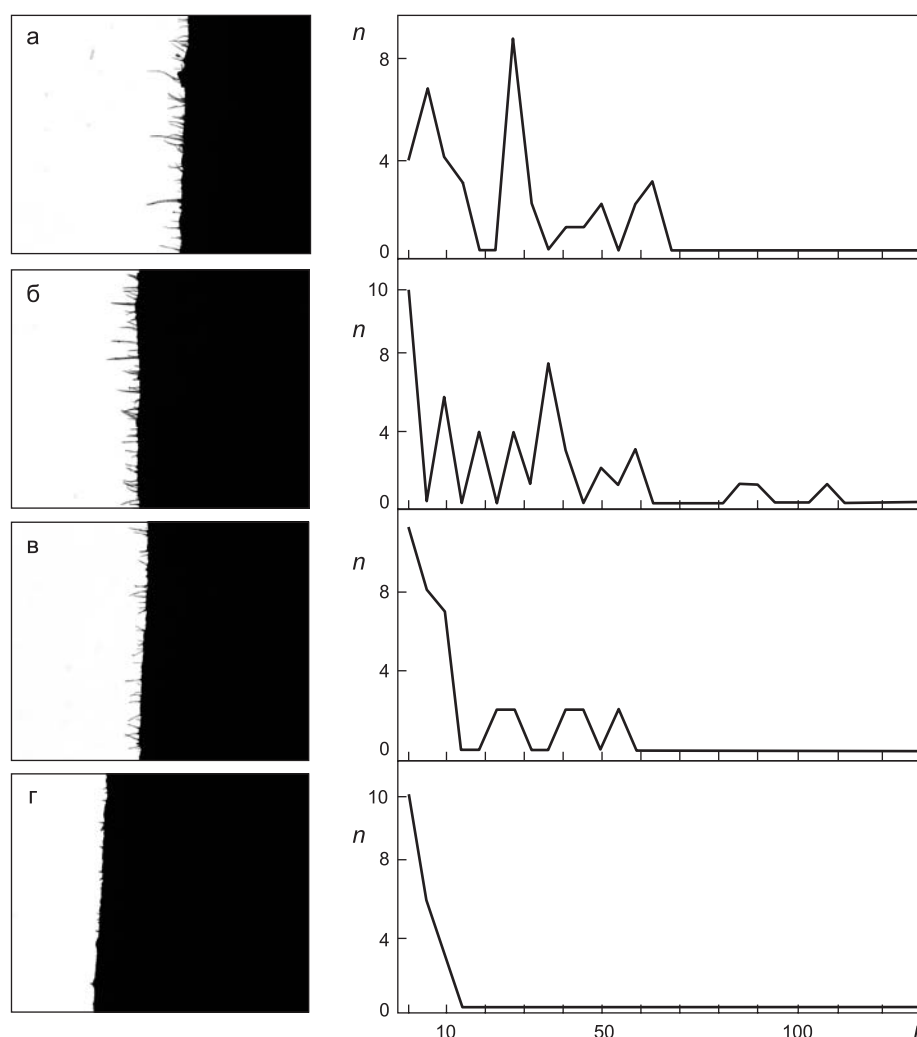
**Результаты работы программы LHDetect.** Результаты компьютерной обработки изображений листа для растений сортов Омская 9, Сибакловская 3, Дуванка 501, Mahndorfer Viktoria



представлены на рис. 3. Слева приведены микроизображения листа для растений этих сортов, справа – гистограммы распределений трихом по длинам. Из рисунка видно, что растения, которые обладают сильным опушением (рис. 3, а–в), могут иметь различные распределения трихом по длине. Так, для сортов Омская 9 и Сибаконская 3 характерно наличие большой доли длинных трихом, трихомы у этих растений имеют большой разброс по длинам. Для сорта Дуванка 501 характерно более однородное распределение трихом по длине, отсутствуют трихомы большой длины. У сорта Mahndorfer Viktoria опушение практически отсутствует. Лист имеет

небольшое количество очень коротких трихом, размер которых сравним с размерами неровностей поверхности листа. Из рисунка видно, что представление морфологии опушения листа в виде числового вектора  $n$  наглядно отражает как степень опушения, так и распределение трихом по длине.

**Сравнение морфологии опушения сортов Балаганка и Голубка.** Мы провели сравнение морфологии опушения листьев для двух сортов пшеницы Балаганка и Голубка. Эти два сорта пшеницы относятся к стародавним сортам, различным по происхождению. Эти сорта получены в первой половине прошлого века в Иркутской



**Рис. 3.** Результаты анализа микроизображений листа программой LHDetect на примере растений пшеницы сортов Омская 9 (а), Сибаконская 3 (б), Дуванка 501 (в), Mahndorfer Viktoria (г). Справа представлены микроизображения листа, слева – гистограммы распределений трихом по длине, оцененные программой. По оси X отложена длина трихом в пикселях, по оси Y – значения  $n_i$  числа трихом.

(Голубка) и Кустанайской (Балаганка) областях. Эти два сорта характерны тем, что их опушение практически неотличимо как тактильно, так и визуально по микроизображению. Для более точного сравнения морфологии опушенности этих двух сортов использовался статистический анализ числовых характеристик, полученных программой LHDetect. В выборке, состоящей из образцов листьев 18 растений сорта Балаганка и 19 растений сорта Голубка, была проведена оценка характеристик опушенности  $n$  и  $S$  для двух независимо полученных изображений. Затем для каждого листа параметры  $n$  и  $S$  были усреднены.

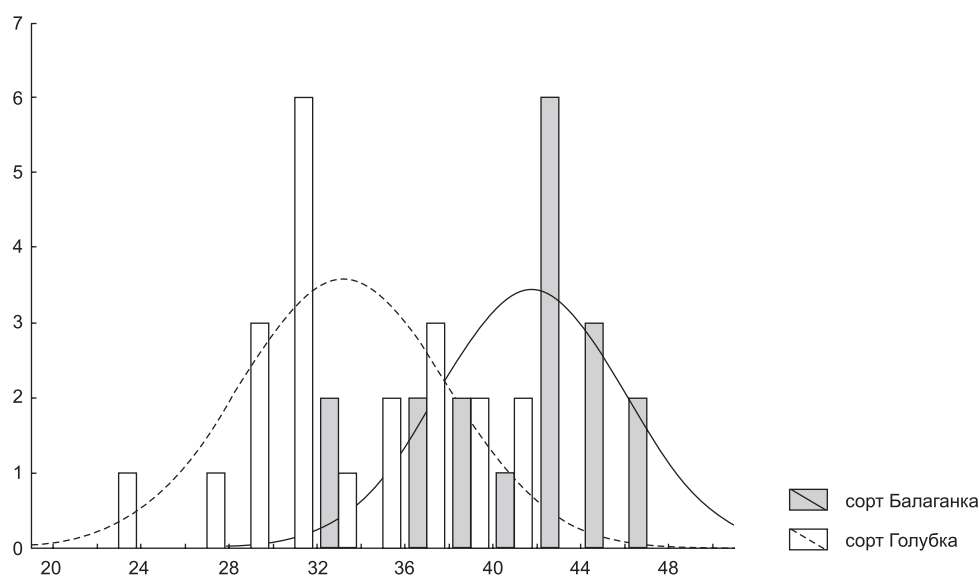
Распределение полученных таким образом значений числа трихом  $S$  для двух сортов пшеницы проведено на рис. 4. Средние значения числа трихом составили 33,17 для сорта Голубка и 41,64 для сорта Балаганка. Статистический анализ показал, что различия между этими значениями являются значимыми согласно  $t$ -критерию Стьюдента ( $p = 0,000001$ ). Таким образом, исследованные нами образцы листьев пшеницы сортов Голубка и Балаганка статистически значимо отличаются по плотности опушения листа (числу трихом на единицу площади).

Мы также провели сравнение образцов данных сортов пшеницы по распределению

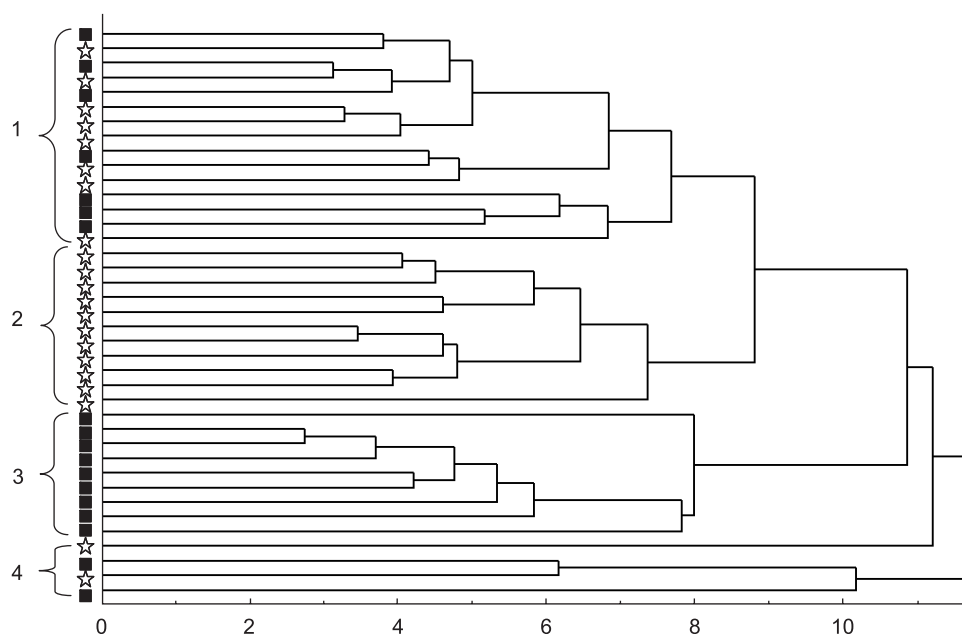
длин трихом. Для этого был использован иерархический кластерный анализ, основанный на сравнении компонент векторов  $n$ . Результаты приведены на рис. 5. Из дендрограммы видно, что все образцы разделяются на 4 группы, близкие по распределению длин трихом. Группы 2 и 3 однородны по составу. В группу 2 входят образцы растений, представленных сортом Голубка, в группу 3 – сортом Балаганка. В первую и четвертую группы входят образцы и того и другого сорта. Мы построили усредненное распределение трихом по длине для растений, входящих в эти группы (рис. 6). Отчетливо выявляется то, что соотношение трихом различной длины неодинаково. Условно можно разделить трихомы на короткие – до 44 мкм (около 21 пикс.) и длинные – более 44 мкм.

Для первой группы образцов характерны большая доля коротких трихом и наличие небольшого числа более длинных трихом. Вторая и четвертая группы характеризуются небольшой долей коротких трихом и относительно большой долей трихом большей длины. В третьей группе большое число коротких трихом, с увеличением длины доля трихом резко уменьшается.

Таким образом, проведенный анализ распределения трихом по длине позволяет получить

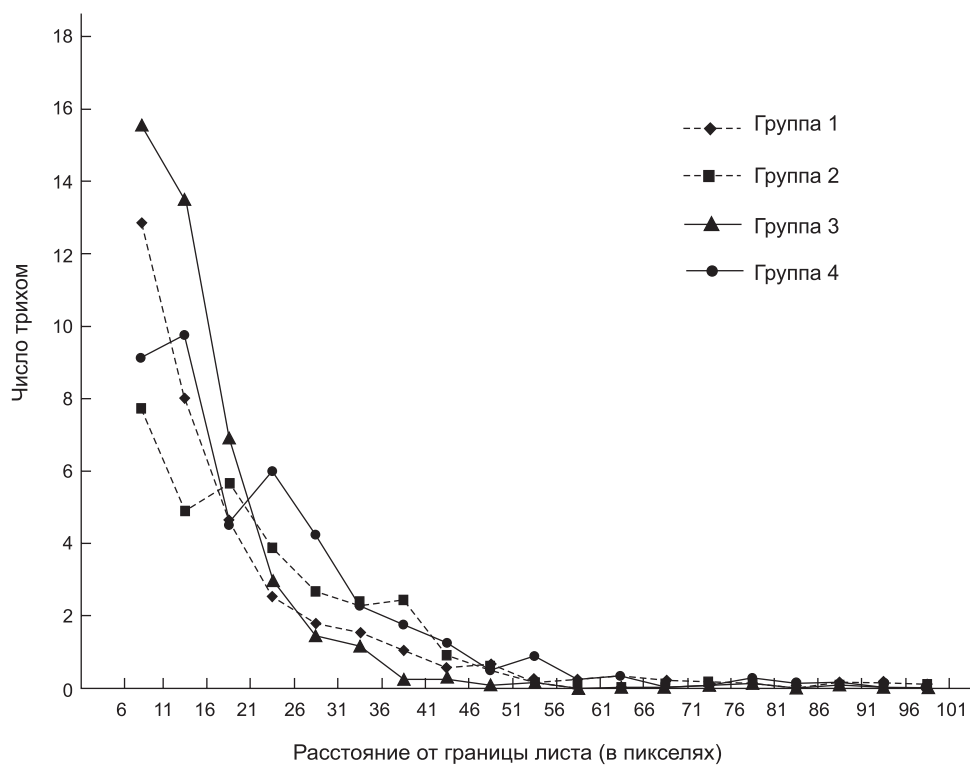


**Рис. 4.** Распределение числа трихом по длине в пикселях изображения для образцов листьев, принадлежащих сортам пшеницы Балаганка (темные столбцы, нормальная аппроксимация – сплошной линией) и Голубка (светлые столбцы, нормальная аппроксимация – пунктирной линией).



**Рис. 5.** Дендрограмма сходства образцов листьев растений пшеницы сортов Голубка и Балаганка по распределению длин трихом (компонент вектора  $n$ ).

По оси X отложено расстояние между кластерами на дендрограмме. Слева приведены обозначения образцов: черный квадрат соответствует образцу сорта Балаганка, звездочка – образцы сорта Голубка. Фигурными скобками обозначены группы образцов.



**Рис. 6.** Распределение трихом по длине, усредненное для образцов из четырех групп растений.

дополнительные данные о морфологии опушения листа и провести более детальную ее классификацию.

В настоящей работе мы использовали новую методику оценки морфологии опушения листа пшеницы, основанную на анализе микроизображения поперечного сгиба листа (Arsenina *et al.*, 2008). Практическое использование данной методики, представленное в настоящей работе на примере анализа образцов от растений разных сортов, выявило ряд ее положительных особенностей:

– с помощью предложенной методики удается количественно оценить плотность опушения листа;

– методика позволяет оценивать распределение трихом по длине и представлять результаты в виде числового вектора;

– для работы необходимы компьютер и оптический микроскоп, который позволяет сохранять микроизображение в цифровом виде;

– процедура получения изображений достаточно проста и не требует много времени, среднее время подготовки образца и получения изображения при настроенном микроскопе составляет около минуты;

– использование достаточно простого алгоритма анализа изображения в программе LHDetect позволяет выполнить операцию оцифровки за несколько секунд.

Дальнейшее исследование возможностей и недостатков предложенного метода позволит, по нашему мнению, улучшить точность подсчета трихом. Так, на практике оказалось, что при существующем алгоритме важную роль играет качество микроизображения. Часть нежелательных эффектов, которые затрудняют подсчет трихом, удается устранить путем подбора оптимальных условий съемки. Однако некоторые проблемы таким образом устранены быть не могут (например, перекрест или наложение трихом на микроизображении). По-видимому, часть нежелательных эффектов микроизображения можно будет в дальнейшем устранить за счет использования более совершенного алгоритма анализа.

В целом простота и низкая трудоемкость метода делают его выигрышным в задачах массового анализа морфологий опушения листьев пшеницы. Мы надеемся, что разработка подоб-

ных подходов в будущем позволит эффективно использовать огромный ресурс современных методов массового генотипирования растений в задачах поиска взаимосвязи генотип–фенотип.

### Благодарности

Работа выполнена при поддержке проекта № 10.7 «Компьютерное моделирование и экспериментальное конструирование генных сетей» по программе РАН «Молекулярная и клеточная биология», проекта № 18.13 «Эволюция молекулярно-генетических систем: компьютерный анализ и моделирование» по программе РАН «Происхождение и эволюция биосферы», государственного контракта 10104-37/П-18/110-327/180608/015 на выполнение НИиОКР по теме «Экосистемно-биоценоотические и генетические механизмы биологической эволюции и корреляция биологических событий» по подпрограмме 2 программы фундаментальных исследований Президиума РАН «Происхождение и эволюция биосферы» и гранта НШ-2447.2008.4 «Биоинформатика и системная компьютерная биология».

### Литература

- Журавлева Н.А. Механизм устьичных движений, продукционный процесс и эволюция. Новосибирск: Наука, 1992. 140 с.
- Крупнов В.А., Цапайкин А.П. Опушение листьев пшеницы: генетические и экологические аспекты // С.-х. биология. Сер. «Биология растений». 1990. № 1. С. 51–57.
- Лихенко И.Е. О взаимосвязи опушения органов растений яровой мягкой пшеницы с хозяйственно и биологически ценными признаками в условиях Западной Сибири // Растениеводство и селекция. 2007. № 6. С. 25–31.
- Мирославов Е.А. Структура и функция эпидермиса листа покрытосеменных растений. Л.: Наука, 1974. 120 с.
- Arsenina S.I., Afonnikov D.A., Pshenichnikova T.A. Development of the computer program for defining leaf hairiness in wheat based on its microscope image processing // VI Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2008). June 22–28, 2008, Novosibirsk, Russia. Book of abstracts. Novosibirsk: Inst. Cytol. Genet, SB RAS, 2008. P. 29.
- Baenzinger S., Outstanding in their field: the phenotype

- of the 21st century plant breeder // XI Intern. Wheat Genet. Symp., 2008 (<http://hdl.handle.net/2123/3325>).
- Kaminuma E., Heida N., Tsumoto Y. *et al.* Automatic quantification of morphological traits via three-dimensional measurement of *Arabidopsis* // *The Plant J.* 2004. V. 38. P. 358–365.
- Lacape J.-M., Nguyen T.B. Mapping quantitative trait loci associated with leaf and stem pubescence in cotton // *J. Hered.* 2005. V. 96. P. 441–444.
- Luo D., Oppenheimer D.G. Genetic control of trichome branch number in *Arabidopsis*: the roles of the *FURCA* loci // *Development.* 1999. V. 126. P. 5547–5557.
- Nagata T., Todoriki S., Hayashi T. *et al.* Gamma-radiation induces leaf trichome formation in *Arabidopsis* // *Plant Physiol.* 1999. V. 120. P. 113–119.
- Perazza D., Herzoga M., Hülskamp M. *et al.* Trichome cell growth in *Arabidopsis thaliana* can be derepressed by mutations in at least five genes // *Genetics.* 1999. V. 152. P. 461–476.
- Roberts J.J., Gallun R.L., Patterson F.L., Forster J.E. Effects of wheat leaf pubescence on the hessian fly // *J. Econ. Entomol.* 1979. V. 72. № 2. P. 211–214.
- Schillinger J.A., Gallun R.L. Leaf pubescence of wheat as a deterrent to the cereal leaf beetle, *Oulema melanopus* // *Annals Entomol. Soc. Amer.* 1968. V. 61. № 4. P. 900–903.
- Sneath P.H.A., Sokal R.R. *Numerical Taxonomy. The Principles and Practice of Numerical Classification.* San Francisco, W.H. Freeman and Co, 1973.
- Xiong L., David L., Stevenson B., Zhu J.-K. High throughput screening of signal transduction mutants with luciferase imaging // *Plant Mol. Biol. Rep.* 1999. V. 17. P. 159–170.

## THE USE OF COMPUTER-BASED IMAGE PROCESSING TO LEAF HAIRINESS ANALYSIS IN WHEAT *TRITICUM AESTIVUM* L.

A.V. Doroshkov<sup>1</sup>, S.I. Arsenina<sup>2</sup>, T.A. Pshenichnikova<sup>1</sup>, D.A. Afonnikov<sup>1,2</sup>

<sup>1</sup>Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: [ada@bionet.nsc.ru](mailto:ada@bionet.nsc.ru);

<sup>2</sup>Novosibirsk State University, Novosibirsk, Russia

### Summary

Leaf hairiness in wheat is of great importance for protection from pests and for adaptation to environmental factors. Study of the features of leaf hairiness morphology and identification of the corresponding genes will allow for obtaining varieties resistant to hard climatic conditions and certain pests. Using new computer-based technologies for descriptions of quantitative characteristics of leaf hairiness is the important step in this direction. In this study, the use of digital analysis of leaf folds images for leaf hairiness evaluation in wheat is offered. This technique is not time-consuming and allows analysis of hairiness density and trichomes length distribution. The suggested method appeared to be the effective approach for a large-scale analysis of leaf hairiness morphological peculiarities in individual plants.



## ПАМЯТИ ВАСИЛИЯ ЕФИМОВИЧА ПЕРФИЛЬЕВА (1931–2008 гг.)



31 марта 2008 г. на 77-м году жизни после продолжительной болезни скончался главный научный сотрудник лаборатории генофонда Всероссийского НИИ генетики и селекции плодовых растений им. И.В. Мичурина (ГНУ ВНИИГиСПР им. И.В. Мичурина, РАСХН) Василий Ефимович Перфильев.

В.Е. Перфильев – крупный ученый-биолог, плодовод, широко известный в научных кругах трудами по разработке методических вопросов применения математических методов в биологии, прежде всего в генетике и селекции плодовых и ягодных культур.

В.Е. Перфильев родился 23 июля 1931 г. в селе Троицком, Новохоперского района, Воронежской области. После окончания Новохоперской средней школы в 1949 г. В.Е. Перфильев поступил на плодоовощной факультет Московской сельскохозяйственной академии им. К.А. Тимирязева, которую окончил с отличием в 1954 г. Увлечение садоводством проявилось у него

еще в школьные годы под влиянием дяди Перфильева Григория Ефимовича – страстного садовода-любителя, который был знаком и переписывался с Иваном Владимировичем Мичуриным. Заложенный им в 1930-е гг. колхозный сад был одним из лучших не только в Воронежской области, там были многие сорта И.В. Мичурина и других селекционеров. Первые уроки по плодоводству Василий получил, работая в летние каникулы вместе с другими учащимися в этом саду.

После окончания ТСХА В.Е. Перфильев работал с 1954 по 1958 гг. в Балашовской области и в Башкирской АССР агрономом колхоза, заместителем главного агронома по овощеводству МТС, главным агрономом МТС. В 1958 г. он поступил в аспирантуру при кафедре плодоводства Плодоовощного института им. И.В. Мичурина. В 1962 г. защитил диссертацию на тему: «Влияние карликовых подвоев яблони на привой в связи с закладкой цветковых почек» (руководитель – профессор, доктор сельскохозяйственных наук В.И. Будаговский). Ему была присуждена ученая степень кандидата биологических наук.

С 1961 г. Василий Ефимович работал в Центральной генетической лаборатории им. И.В. Мичурина (ЦГЛ) младшим научным сотрудником, затем старшим научным сотрудником, а с 1974 г. заведовал созданной им лабораторией генетико-статистических методов. В 1989 г. защитил докторскую диссертацию на тему: «Разработка методических вопросов селекции плодовых растений на основе статистического анализа количественных признаков».

В 1991 г. В.Е. Перфильев был назначен директором ЦГЛ (с 1992 г. Всероссийский НИИ генетики и селекции плодовых растений им. И.В. Мичурина). Несмотря на сложные финансовые условия в период перестройки, институт сохранил кадровый состав и достаточно высокий уровень научных исследований. Ежегодно

проводились Мичуринские чтения, другие научные совещания и семинары, публиковались научные труды. Улучшились жилищные условия для научных сотрудников за счет строительства многоквартирного дома. В 1997 г. В.Е. Перфильев принял решение уйти с поста директора ВНИИГиСПР им. И.В. Мичурина и сосредоточить внимание на продолжении научно-исследовательской работы.

Еще в студенческие годы параллельно с изучением плодово-овощеводства и физиологии растений (дипломную работу он выполнял в ИФР им. К.А. Тимирязева) В.Е. Перфильев начал глубоко постигать теоретические и методические вопросы математики с целью применения их в биологии.

Большое методологическое значение имеет обоснованное им представление об онтогенезе растений как о вероятностном процессе, создающее теоретическую основу для использования идей и методов теории вероятностей и математической статистики при решении различных биологических вопросов.

Для математического описания динамики роста побегов древесных растений в течение вегетационного периода предложена формула роста, в основе которой лежит интеграл вероятностей. Установлено, что асимметрия эмпирических распределений растений по показателям роста тесно связана с условиями выращивания. По величине и знаку коэффициента асимметрии можно судить о том, насколько условия выращивания благоприятны для растений. Сильно выраженная положительная асимметрия свидетельствует о том, что условия выращивания неблагоприятны для роста растений, что растения далеко не полно реализовали свои возможности в отношении роста и что имеются большие возможности для усиления роста растений с помощью различных воздействий, прежде всего путем надлежащей агротехники.

Разработаны методические вопросы применения анализа главных компонент в исследованиях по генетике количественных признаков и по корреляциям между признаками. Показано, что выход перспективных гибридов в разных комбинациях скрещивания плодовых растений зависит от степени различия между собой родительских форм по комплексу количественных признаков. В качестве показателя степени

различия разных форм можно использовать евклидово расстояние.

Предложен принципиально новый подход к решению проблемы подбора родительских пар при выведении новых сортов плодовых и ягодных растений, основанный на разработанном совместно с А.И. Бутенко способе вычисления вероятностей получения гибридов с заданными признаками в разных комбинациях скрещивания.

Предложен способ определения числа учетных растений, необходимых для обеспечения заданной точности исследований в опытах по сравнительной оценке сортов плодовых культур по урожайности с учетом ошибок первого и второго рода.

Большое внимание В.Е. Перфильев уделял разработке методических рекомендаций по генетико-селекционным вопросам. Совместно с другими сотрудниками им подготовлен ряд методических рекомендаций, которые широко используются работниками многих научно-исследовательских учреждений страны.

В 1994 г. вышла в свет книга В.Е. Перфильева «Варьирование и взаимосвязь количественных признаков у плодовых растений» (Мичуринск, 1994. 188 с.), написанная на основании результатов многолетних исследований автора по указанным проблемам. Содержащиеся в книге материалы имеют немаловажное методологическое и методическое значение для решения различных вопросов генетики, селекции и возделывания плодовых растений.

Результаты исследований В.Е. Перфильева обобщены и опубликованы в ведущих научных журналах: «Физиология растений», «Доклады АН СССР», «Вестник сельскохозяйственных наук», «Научные доклады высшей школы», «Садоводство», «Вестник Российской академии сельскохозяйственных наук», а также в докладах научных и научно-методических совещаний (более 60 публикаций). Под его руководством подготовлены методические рекомендации по работе с базой данных Cerasus (2002 г.).

В течение многих лет В.Е. Перфильев преподавал в Мичуринском педагогическом государственном институте (МПГИ) и в Мичуринском аграрном университете. Ему присвоено звание профессора по кафедре селекции плодовых культур и ботаники (1994 г.) и Почетного про-

фессора Мичуринской сельскохозяйственной академии (1995 г.).

В.Е. Перфильев возглавлял Всесоюзную школу-семинар по генетике и селекции на базе руководимой им лаборатории генетико-статистических методов, являлся членом бюро Совета по научно-методическому руководству селекционными центрами, членом бюро секции садоводства Отделения растениеводства и селекции РАСХН, заместителем председателя Мичуринского отделения Вавиловского общества генетиков и селекционеров.

Им подготовлены через аспирантуру высокоспециализированные кадры по вопросам применения математических методов в генетике и селекции плодовых культур.

Светлая память о нем надолго сохранится в сердцах всех, кто его знал.

### Основные работы В.Е. Перфильева

- Перфильев В.Е. О скороплодности карликовых деревьев яблони // Науч. докл. высш. школы. Биол. науки. 1961. № 1. С. 195–197.
- Перфильев В.Е. Влияние подвоя на соотношение веса коры и древесины у яблони // Агробиология. 1961. № 2. С. 304–305.
- Перфильев В.Е. Особенности закладки плодовых почек у карликовых деревьев // Садоводство. 1961. № 5. С. 17–18.
- Перфильев В.Е. Связь между содержанием кальция в золе побегов яблони и способностью деревьев к закладке плодовых почек // Физиол. растений. 1962. № 3. С. 371–372.
- Перфильев В.Е. О причинах скороплодности карликовых деревьев яблони // Агробиология. 1962. № 4. С. 619–622.
- Перфильев В.Е. Математическое выражение роста сеянцев яблони в течение первого вегетационного периода // Докл. АН СССР. 1966. Т. 169. № 4. С. 971–974.
- Перфильев В.Е. О закладке цветочных почек у яблони // Вестн. с.-х. науки. 1966. № 12. С. 65–68.
- Перфильев В.Е. Объяснение различной продолжительности предплодоносного периода у сеянцев яблони с позиций представлений многовариантности онтогенеза // Бюл. науч. информ. Центр. генет. лаб. им. И.В. Мичурина. 1967. Вып. 13. С. 27–32.
- Перфильев В.Е. Регрессионный анализ роста сеянцев яблони // Тр. Центр. генет. лаб. им. И.В. Мичурина. 1967. Т. 9. С. 261–266.
- Перфильев В.Е. Коэффициент асимметрии статистических распределений растений как показатель условий роста // С.-х. биология. 1969. № 3. С. 451–457.
- Перфильев В.Е. Статистическая обработка опытных данных // Программа и методика изучения сортов плодовых и ягодных культур. Мичуринск, 1970. С. 212–236.
- Перфильев В.Е. Метод оценки предельных возможностей сеянцев яблони // Вестн. с.-х. науки. 1978. № 7. С. 119–122.
- Перфильев В.Е., Лебедев А.В., Тихонов В.А. О возможности использования обобщенного показателя расстояния между сортами в селекции плодовых растений // С.-х. биология. 1979. № 4. С. 482–486.
- Перфильев В.Е., Лебедев А.В., Тихонов В.А. Особенности плодоношения разных сортов яблони в зависимости от структуры кроны // Биологические основы продуктивности плодовых семечковых культур. М., 1979. С. 49–50.
- Перфильев В.Е. Изучение количественных признаков у яблони методом главных компонент // Генетика. 1980. № 8. С. 1447–1452.
- Перфильев В.Е., Лебедев А.В., Яковлев С.П. и др. Методические рекомендации по применению статистических методов в генетике и селекции плодовых растений. Мичуринск, 1980.
- Перфильев В.Е. О возможности использования статистических методов при подборе пар для гибридизации // Бюл. науч. информ. Центр. генет. лаб. им. И.В. Мичурина. 1980. Вып. 35. С. 69–74.
- Перфильев В.Е., Бутенко А.И. Исследование взаимосвязи между признаками методом главных компонент. Развитие научного наследия И.В. Мичурина. Мичуринск, 1981. С. 47–48.
- Перфильев В.Е., Лебедев А.В., Курсакова Л.А., Бутенко А.И. К вопросу о подборе родительских пар для гибридизации у сливы // Бюл. науч. информ. Центр. генет. лаб. им. И.В. Мичурина. 1981. Вып. 37. С. 20–23.
- Перфильев В.Е., Бутенко А.И., Лебедев А.В. и др. Исследование количественных признаков у плодовых и ягодных растений методом главных компонент // Тр. Центр. генет. лаб. им. И.В. Мичурина. 1981. С. 90–100.
- Перфильев В.Е., Бутенко А.И. Вычисление вероятностей получения гибридов с заданными свойствами в разных комбинациях скрещивания плодовых растений // Вестн. с.-х. науки. 1984. № 7. С. 112–117.
- Перфильев В.Е., Лебедев А.В., Чуча В.С., Боровских О.В. Зависимость показателей плодоношения яблони от признаков роста, облиственности и закладки цветковых почек // Бюл. науч. информ. Центр. генет. лаб. им. И.В. Мичурина. 1986.

- Вып. 43. С. 51–55.
- Перфильев В.Е., Бутенко А.И., Зубов А.А. Вычисление вероятностей получения гибридов с заданными признаками в разных комбинациях скрещивания плодовых и ягодных растений // Докл. ВАСХНИЛ. 1988. № 6. С. 19–20.
- Перфильев В.Е. Генетический анализ // Энциклопедия садоводства. Кишинев, 1990. Т. 1.
- Перфильев В.Е. Закон независимого распределения признаков. Энциклопедия садоводства. Кишинев, 1990. Т. 1.
- Перфильев В.Е. Сравнительная оценка сортов плодовых культур по урожайности (методические рекомендации). Мичуринск, 1990. 70 с.
- Об использовании корреляций при отборе сеянцев плодовых культур по косвенным признакам // Бюл. науч. информ. Центр. генет. лаб. им. И.В. Мичурина. 1991. Вып. 50. С. 9–14.
- Перфильев В.Е. Применение статистических методов в селекции плодовых растений // Состояние и проблемы садоводства России: сб. науч. тр. Новосибирск: НИИСС им. Лисавенко, 1997. Ч. 1. С. 63–68.
- Перфильев В.Е. Многовариантность развития и закономерности изменчивости количественных признаков у растений // С.-х. биология. 2002. № 5. С. 95–203.
- Перфильев В.Е., Лебедев А.Н., Насонов В.Н. и др. Методические рекомендации по работе с базой данных Cerasus / Гл. ред. Н.И. Савельев. Мичуринск: ГНУ ВНИИГИСПР им. Мичурина, 2002. 19 с.
- Перфильев В.Е., Чуча В.С. О зависимости между показателями роста и плодоношения у яблони // С.-х. биология. 2003. № 5. С. 28–32.
- Перфильев В.Е. Особенности преподавания дисциплины «Основы научных исследований в агрономии» по специальности «Селекция и генетика сельскохозяйственных культур» // Тр. ученых Мичуринского гос. аграрного ун-та: сб. науч. тр. Воронеж, 2005. С. 166–170.
- Перфильев В.Е. О количественных показателях, характеризующих структуру кроны яблони // С.-х. биология. 2005. № 5. С. 86–93.
- Перфильев В.Е. Некоторые вопросы применения статистических методов в генетике, селекции и сортоизучении плодовых растений в плане развития научного наследия И.В. Мичурина // Тр. Всерос. науч.-исслед. ин-та генетики и селекции плодовых растений им. И.В. Мичурина. Генетика и селекция плодовых растений: сб. науч. тр. Воронеж: Кварта, 2005. С. 167–181.
- Перфильев В.Е., Борзых Н.В. Связь генотипических особенностей структуры у груши с габитусом и плодоношением деревьев // Вестн. РАСХН. 2008. № 6. С. 39–41.

**Коллектив ГНУ ВНИИГИСПР им. И.В. Мичурина**

## ПРАВИЛА ОФОРМЛЕНИЯ РУКОПИСЕЙ ДЛЯ ПУБЛИКАЦИИ В ЖУРНАЛЕ «ИНФОРМАЦИОННЫЙ ВЕСТНИК ВОГиС»

### Общие положения

Журнал «Информационный вестник ВОГиС» публикует на русском (или английском) языке работы по всем разделам генетики, селекции, а также смежных наук. К публикации принимаются результаты оригинальных экспериментальных исследований; теоретические и обзорные статьи, представляющие интерес для научного сообщества; краткие сообщения, рецензии, письма редактору, персоналии, хроника, информация, сообщения из отделений ВОГиС, материалы и документы по истории генетики и селекции. Печатаются также материалы, касающиеся образовательных программ и методики преподавания генетики и селекции в средней и высшей школах. Журнал печатает заказанные редколлегией обзоры и проблемные статьи. Они могут быть предложены также авторами после предварительной заявки, которую рассматривает редколлегия. Заявка должна содержать резюме предлагаемого обзора или проблемной статьи. Отдельные тематические выпуски посвящаются наиболее актуальным проблемам генетики и селекции.

Хотя журнал и является официальным изданием Вавиловского общества генетиков и селекционеров, членство авторов в обществе необязательно – журнал одинаково открыт для всех.

Сайт журнала в Интернете:

<http://www.bionet.nsc.ru/vogis/>

В редакцию статьи представляются в электронном виде в формате MS WinWord 6.0 (и выше) на дискетах размером 3,5", на CD, через FTP или по электронной почте в форме присоединенных файлов. Текст статьи, включая аннотацию на русском и английском языках, таблицы, иллюстрации и подписи к ним, а также список литературы оформляются одним файлом. Иллюстрации дополнительно присылаются отдельными файлами. Если пересылаемый

материал велик по объему, следует архивировать файлы в формат \*.zip или \*.rar.

Текст статьи на бумаге обязателен.

Наш адрес:

630090 Новосибирск, пр. академика Лаврентьева 10, Институт цитологии и генетики СО РАН. Редакция журнала «Информационный вестник ВОГиС».

Электронный адрес: [vestnik@bionet.nsc.ru](mailto:vestnik@bionet.nsc.ru)

К публикации в журнале «Информационный вестник ВОГиС» принимаются статьи, прошедшие рецензирование. Решение о публикации принимается редакционной коллегией журнала на основании экспертных оценок рецензентов с учетом соответствия представленных материалов тематической направленности журнала, их научной значимости и актуальности.

Поступившая в редакцию рукопись направляется на отзыв специалистам в данной области исследований. Авторы статьи могут назвать двух-трех потенциальных рецензентов. Наряду с фамилией каждого рецензента обязательно указание его полного имени и отчества, места работы, телефона, адреса электронной почты. Выбор рецензентов остается за редколлегией журнала. Рукопись, получившая отрицательные отзывы двух независимых рецензентов, решением редколлегии отклоняется.

Статья, нуждающаяся в доработке, направляется авторам с замечаниями рецензента и научного редактора. Авторы должны учесть все замечания, сделанные в процессе рецензирования и редактирования статьи, ответить на каждое из замечаний и указать место в рукописи, где сделаны изменения. В случае несогласия с рецензентом или редактором автор должен кратко и четко обосновать свою позицию. Сделанные автором изменения в рукописи необходимо внести в электронный вариант текста и вернуть в редакцию. После доработки статья повторно рецензируется и редколлегией принимается решение о возможности публикации.



Статья, отправленная редакцией на доработку после рецензии и исправленная в соответствии с замечаниями рецензента, должна быть возвращена в редакцию в течение 15 дней с момента ее получения авторами, в этом случае сохраняется первая дата поступления. Статья, возвращенная в редакцию по прошествии месяца, будет иметь новую дату поступления.

Редакция не предоставляет авторам копии корректуры статьи на бумаге. Статья высылается автору в виде pdf-файла. На стадии корректуры не допускаются замены текста, рисунков или таблиц. Если в корректуру вносятся исправления, при возвращении ее в редакцию подробный список сделанных исправлений необходимо приложить в виде отдельного файла.

Редакция оставляет за собой право отклонить без рецензии статьи, не соответствующие профилю журнала или оформленные с нарушением правил.

На всех стадиях работы с рукописями и для общения с авторами, редакторами и рецензентами используется электронная почта, поэтому авторы должны быть внимательны при указании своего электронного адреса и своевременно извещать редакцию о его изменении.

### Требования к оформлению рукописей

Статьи должны быть написаны на русском (или английском) языке, отредактированы и оформлены в соответствии с нижеследующими требованиями.

Объем статьи – до 15 страниц формата А4, нумерация страниц сквозная. В этот объем входят текст, аннотация, список литературы, таблицы, иллюстрации и подписи к ним. Основной текст набирается шрифтом Times New Roman через 1,5 интервала с выравнением по ширине и без переносов, размер шрифта 12 pt. Поля – 3 см со всех сторон страницы.

Статьи большего объема принимаются только после предварительного согласования с редакцией.

**Латинские названия** объектов исследований в названии статьи и в тексте пишутся с соблюдением общепринятых правил таксономической номенклатуры: бинарные видовые – курсивом (*Drosophila melanogaster*), таксонов более высокого ранга – прямым шрифтом

(*Drosophila* или *Drosophilidae*). При первом упоминании в тексте родовые и видовые названия приводятся без сокращений, далее по тексту родовое название обозначается одной прописной (первой) буквой, а видовое указывается полностью (*D. melanogaster*).

**Названия и символы** генов набираются курсивом, а названия их продуктов – с прописной буквы прямым шрифтом. Например: гены *fos*, *c-мус*, *ATM*; белки Fos, с-Мус, АТМ. Курсивом выделяются обозначения мобильных элементов, например, *hobo*-элемент, а также три первых буквы названий сайтов рестрикции, например, *HindIII*. Названия фагов и вирусов пишутся в латинской транскрипции прямым шрифтом.

Следует использовать общепринятые сокращения и аббревиатуры или приводить их дополнительно в тексте. Все физические размерности рекомендуется приводить в международной системе СИ.

**Математические формулы и уравнения** набираются в редакторах MS WinWord (версия 6.0 и выше) или MathType. Уравнения располагаются по центру строки и нумеруются арабскими цифрами в круглых скобках в порядке их упоминания в тексте. Номера уравнений выравняются по правому краю строки. Уравнения отделяются от текста сверху и снизу одной пустой строкой. При написании нескольких уравнений они также разделяются пустой строкой.

**Таблицы и иллюстрации** (графики, схемы, фотографии, штриховые рисунки) представляются в черно-белом варианте.

**Таблицы, иллюстрации** и подписи к ним размещаются в тексте статьи при первом их упоминании. При этом не следует использовать опцию «обтекание текста».

Таблицы снабжаются тематическими заголовками и нумеруются арабскими цифрами в порядке их упоминания в тексте. Все графы в таблицах должны иметь заголовки. Сокращение слов в таблицах не допускается. Все аббревиатуры должны быть расшифрованы в сносках к таблице.

Число иллюстраций в статье не должно быть больше 6 (исключения согласовываются с редакцией). Максимальный размер иллюстраций или таблиц не должен превышать размера рабочего поля 15,7 × 23 см. Весь иллюстра-

тивный материал должен иметь минимум надписей.

На графиках необходимо указывать величины, значения которых даются на осях, и обозначение их размерностей.

Имеющиеся в схемах детали обозначаются арабскими цифрами или буквами русского алфавита и расшифровываются в подписях.

Иллюстрации нумеруются в порядке их упоминания в тексте. При ссылке в тексте на иллюстрацию указывается ее номер и буквенные и цифровые обозначения ее деталей, например: рис. 1, а, кривая 2.

Графики и схемы должны выполняться с помощью векторных программ (Microsoft Excel, CorelDraw 9, Microsoft PowerPoint), их следует присылать в виде отдельных файлов с сохранением форматов, использованных для их создания. Если они выполнялись в других векторных программах, то необходимо использовать формат EPS.

Фотографии представляются в виде отдельных файлов в форматах JPEG, TIFF, BMP, PNG с разрешением 300–600 dpi.

Штриховые рисунки, выполненные от руки, должны быть отсканированы в режиме bitmap с разрешением 800 dpi и сохранены в формате TIFF.

### Структура рукописи

Материалы должны быть размещены следующим образом:

1. Название статьи. Должно быть кратким и отражать содержание работы. Печатается прописными буквами прямым полужирным шрифтом без подчеркивания и разрядки. Латинские названия объектов исследований в названии статьи пишутся без сокращений, с соблюдением общепринятых правил таксономической номенклатуры.

2. Инициалы и фамилия (фамилии) автора(ов) (А.А. Иванов, Б.В. Петров...). Печатаются прямым строчным шрифтом и отделяются от названия статьи пустой строкой.

3. Полное название и адрес учреждения, где работает автор(ы) – шрифт прямой строчной; адрес(а) электронной почты автора(ов) – шрифт прямой строчной.

Отметить арабскими цифрами соответствие фамилий авторов учреждениям, в которых они работают; звездочкой пометить фамилию автора, с которым будет вестись переписка.

4. Аннотация (резюме) статьи с кратким изложением основной цели работы и ее результатов (не более 20 строк). Шрифт – прямой строчной.

5. Ключевые слова (не более 10). Шрифт – прямой строчной. В ключевых словах должны быть отражены: (1) объект; (2) метод; (3) область исследования; (4) специфика данной работы.

6. Текст статьи, оформленный в соответствии с правилами.

7. Благодарности и ссылки на источники финансирования работы.

8. Литература.

9. Аннотация (резюме) статьи на английском языке с кратким изложением основной цели работы и ее результатов (не более 20 строк).

Для экспериментальных статей рекомендуются следующие разделы: **Введение, Материалы и Методы; Результаты; Обсуждение; Литература.**

Теоретические, обзорные и проблемные статьи могут иметь произвольную структуру, но обязательно должны содержать резюме. Названия разделов в таких статьях определяются автором.

Названия разделов печатаются строчными буквами на отдельной строке без подчеркивания и отделяются от текста одной пустой строкой. Шрифт – прямой полужирный. Подзаголовки внутри разделов печатаются на отдельной строке строчными буквами. Шрифт – прямой полужирный. Заголовки и подзаголовки выравниваются по центру.

Раздел «Литература» отделяется от текста статьи пустой строкой и содержит перечень цитированных источников с обязательным указанием заглавия (см. ниже образец). Библиографические ссылки внутри текста приводятся в круглых скобках. При этом указывается фамилия автора публикации без инициалов и год публикации, например: (Иванов, 1999). Если у публикации два автора, то указываются обе фамилии и год издания, например: (Gihg, Smith, 2001). Работы трех и более авторов цитируются

следующим образом: (Gatsby *et al.*, 1998; Добров и др., 2000).

При ссылках на несколько публикаций ссылки в скобках располагаются в хронологическом порядке, например: «В ряде работ (Смирнов, 1978; Smith, Gatsby, 1998; Павлова и др., 2001)...». При этом если цитируются работы одного и того же года, ссылки располагаются в алфавитном порядке (сначала русские, потом иностранные фамилии). Если цитируются несколько работ одного и того же автора (или одной и той же группы авторов), опубликованных в одном и том же году, то к году добавляются русские или латинские строчные буквы в алфавитном порядке.

Например: Смирнов и др., 1995а, б; Smith 1997а, d; Ulrich *et al.*, 1998b. Порядок расстановки букв определяется положением статьи в разделе «Литература».

Список литературы должен содержать библиографическое описание всех литературных источников, ссылки на которые фигурируют в тексте статьи. Цитированная литература сводится в алфавитные списки сначала на русском языке, потом на иностранных языках с указанием фамилий и инициалов всех авторов каждой публикации. Работы одного и того же автора располагаются в хронологической последовательности. В случае если в списке приводятся несколько работ одного автора, опубликованных в одном и том же году, им дают буквенные обозначения: 1999а, б, в и т. д.; для иностранных авторов – 1999а, b, c и т. д.

Оформляйте список по следующему образцу, обращая внимание на знаки препинания и пробелы.

### Литература

#### ДЛЯ ЖУРНАЛОВ:

Кольцов Н.К. Проблема прогрессивной эволюции // Биол. журнал. 1933. Т. 2. Вып. 4/5. С. 475–500.

Hanahan D., Weinberg R.A. The hallmarks of cancer // Cell. 2000. V. 100, N 2. P. 57–70.

#### ДЛЯ КНИГ:

Завадский К.М. Развитие эволюционной теории после Дарвина (1859–1920-е годы). Л.: Наука, 1973. 423 с. (или конкретные страницы, например, С. 22–32).

То же для иностранных изданий.

#### ДЛЯ СБОРНИКОВ:

Розова М.А., Янченко В.И., Мельник В.М. Зависимость урожайности яровой твердой пшеницы от метеорологических факторов в Приобской лесостепи Алтайского края // Современные проблемы и достижения аграрной науки в животноводстве и растениеводстве: Сб. статей Междунар. науч.-практ. конференции. Барнаул, 2003. Ч. 1. С. 71–74.

Golygina V.V., Istomina A.G., Kiknadze I.I. Chromosomal polymorphism in natural populations of *Chironomus balatonicus* Dévai, Wülker et Scholl // Late 20th century research on Chironomidae / Ed. O. Hof-frichter. Aachen: Shaker-Verlag, 2000. P. 89–92.

#### ДЛЯ ТЕЗИСОВ ДОКЛАДОВ:

Peshkov I.M., Likhoshvai V.A., Matushkin Yu.G., Fadeev S.I. On research into hypothetical networks on ecological nature // Proc. of the 4th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2004). Novosibirsk, 25–30 July 2004. Novosibirsk: Inst. Cytol. Genet., 2004. V. 2. P. 128–130.

#### ДЛЯ ИНТЕРНЕТ-ПУБЛИКАЦИЙ:

Smith A., Green P. RepeatMasker. 1999. available at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

Schwenger G.T.F., Mordvinov V.A., Fournier R., Czabotar P., Peroni S., Sanderson C.J. (2000). Interleikin-5. In: Academic Press Cytokine Reference Database. DOI:10.1006/rwcy.2000.0902.

На отдельной странице следует привести оригинальное написание иностранных фамилий, встречающихся в тексте статьи, подписях к иллюстрациям и таблицам.

На отдельной странице следует указать:

– на русском языке – сведения об авторах (фамилии, имена, отчества полностью, ученые степени, звания, должности, место работы, полные почтовые адреса с индексами, домашние и служебные телефоны, факсы, адреса электронной почты и адреса личных страниц в Интернете);

– на английском языке – общепринятую версию названия учреждения, где выполнена работа, и транслитерацию фамилий авторов.

## NOTES FOR CONTRIBUTORS

### General information

The HERALD of Vavilov Society for Geneticists and Breeding Scientists is a journal which publishes works written in Russian or English on all aspects of genetics, breeding and related sciences. The results of original experimental research works, theoretical articles (assays), digests, short communications, reviews, letters to the Editor, current and past events, materials and documents with relevance to genetics, breeding and their educational aspects in high school and university can all be considered for publication. The HERALD publishes invited reviews and topical articles. On important occasions, The HERALD features special issues.

Although The HERALD is the official journal of the Vavilov Society of Geneticists and Breeding Scientists, it assumes no members – only attitude: anyone can contribute.

The HERALD is available online:

<http://www.bionet.nsc.ru/vogis/>

The contributions should be submitted in hard copy and in electronic form (MS Word 6.0 or higher) on 3.5-inch floppy disks, CDs, via FTP or as attachments to electronic mail. The texts of the manuscript, including the abstract, tables, figures, figure legend and references, should be submitted as a file. Additionally, each figure should be sent separately as an attached file. If the material is too large to email, archiving to .zip or .rar format should be considered.

RUSSIAN FEDERATION

630090 Novosibirsk

Prospekt Lavrentyeva 10

Institute of Cytology and Genetics, SB, RAS

The Herald of Vavilov Society for Geneticists and Breeding Scientists

Email: [vestnik@bionet.nsc.ru](mailto:vestnik@bionet.nsc.ru)

The manuscripts submitted for publication in The HERALD are peer-reviewed. The acceptance decision is made by the Editors based on the reviews.

All the compliant manuscripts submitted to The HERALD for consideration are forwarded

to reviewers. The reviewers are persons who are believed by the Editors to have sufficient expertise in the field. The author may name two or three persons who he or she wants to step down as reviewers; in this case, their last names and respective affiliations, telephone numbers and electronic addresses should be provided. The Editors' choice of reviewers is final. Any manuscript that has received two negative reviews will be declined.

The manuscript that needs to be improved is returned to the author as the proofs. The authors should take account of all the remarks made, duly respond to them and indicate in the proofs where the corrections have been made. In case of any dispute that might arise in connection with the reviewers' feedback, the author should clearly explain his or her points. The proofs should be returned to the Editors. The Editors will then forward the proofs to the reviewers for opinion.

The proofs that have been sent to the author should be returned within 15 days of receipt. If the proofs are returned to the Editors later than within one month of receipt, the manuscript will be assigned a new date of receipt.

No hard copy proofs are provided to the authors. The proofs are sent to the authors as .pdf files. No correction to the text, figures or tables should be made while the proof-reading process is in progress. After due corrections have been introduced to the proofs, a detailed list of the corrections should be provided in a separate file.

The Editors reserve the right to decline non-compliant manuscripts without review.

The Editors use electronic mail to communicate with the authors, so it is the author's responsibility to ensure that the email address they have provided is correct and current.

### Layout

The articles should be written in the English language and laid out as described below.

Any manuscript should not exceed 15 consecutively numbered standard-sized pages of



A4 size, including the text, summary, references, tables, figures and figure legends. The text should appear in Times New Roman, 12 pt, fully justified, not hyphenated, with line spacing 1.5 lines. Margins should be 3 cm on each side. The contributors of larger manuscripts should consult the Editors before submitting.

**The Latin names** in the title of the manuscript and within the text should be typed in accordance with the standard nomenclature: binary names are italicized (for example, *Drosophila melanogaster*), higher taxa are in upright type (for example, *Drosophila*, *Drosophilidae*). At the first mention within the text, the generic and species names should appear in full (as, for example, *Drosophila melanogaster*); elsewhere, the generic name is abbreviated to its first letter, and the species name is written in full (for example, *D. melanogaster*).

**The names and symbols** of genes should be italicized (for example, *fos*, *c-myc*, *ATM*), while the names of gene products should appear in upright type, with an initial capital (for example, Fos, c-Myc, ATM). The names of mobile elements should be italicized (for example, *hobo*), and so should the first three letters in the names of restriction sites (for example, *HindIII*). The names of phages and viruses should appear in upright type.

All non-standard abbreviations that may appear throughout should be explained in the text. The recommended system for the dimensions of physical units is the SI.

**Mathematical formulae and equations** should be created using MS Word 6.0 (or higher) or MathType. The equations should be centered and numbered with parenthesized Arabic numerals in order of mention within the text. The numbers of the equations are right-justified. A blank line should be inserted before and after an equation. A blank line should be inserted to separate any two consecutive equations.

**Tables and figures** (plots, diagrams, photographs, line drawings) should be provided in black and white.

Tables, figures and figure legends should appear after the first mention within the text. Text wrapping should be avoided.

Tables should be numbered with Arabic numerals in order of mention within the text and provided with telling titles. Each column in a table should be provided with a title or description. Words in the

tables should appear in full. Abbreviations used in the tables should be explained in the legend.

The number of figures in a particular manuscript is limited to 6 (exceptions should be agreed upon with the Editors). The maximum size of a figure or table is limited to 15.7 cm x 23 cm. Text labels should be kept to a minimum.

The symbols and units of quantities represented on the axes of the plots should be indicated.

Figure parts should be labeled with Arabic numerals or Roman letters and explained in the figure legend. The figures should be numbered in order of mention within the text. Referring to a figure in the text: Fig. 1. Referring to a figure part in the text: Fig. 1, a, curve 2.

Plots and diagrams should be created using the vector programs MS Excel, CorelDraw 9, and MS PowerPoint and submitted as separate files in original formats. If other vector programs have been used, images should be stored in the EPS format.

Photographs should be submitted as separate .JPEG, .TIFF, .BMP or .PNG files, with a resolution of 300-600 dpi.

Line drawings should be bitmap scanned with a resolution of 800 dpi and saved to a .TIFF file.

### Manuscript Structure

Any manuscript should contain:

1. A brief and telling title in upper-case bold-face upright type. Underlining or kerning should be avoided. Latin names in the title should be typed in full, in accordance with the standard nomenclature.

2. The author(s) name(s) in title-case upright type (R.U. Right, I.M. Fine) and separated from the title of the manuscript by a blank line.

3. The affiliation(s) and address(es) in title-case upright type; electronic address(es) in lower-case upright type. Each author's name and his or her affiliation should be matched using Arabic numerals. The name of the author to whom the correspondence will be sent should be asterisked.

4. An abstract, which is a brief (no more than 20 lines) summary of the work done and the results obtained. The abstract should appear in sentence-case upright type.

5. Key words (no more than 10). The key words should convey (1) the main object(s); (2)



the method(s) used; (3) the scope; (4) the special points of the work done. The key words should appear in lower-case upright type.

6. The main text.

7. Acknowledgements.

8. References.

9. Summary provides a brief representation of the main aim of research and the results obtained, and should not exceed 20 lines.

For experimental articles, the following sections are recommended: Introduction, Materials and Methods; Results; Discussion; References. Essays, reviews and topical articles can have an arbitrary structure and section titles; however, an abstract is required on all occasions.

Section titles should appear in upper-case bold-face upright type on a separate line. Underlining should be avoided. Any section title should be separated from the text by a blank line.

Subsection titles should appear on a separate line, in upper-case bold-face upright type. The section and subsection titles should be centered.

References in the text should appear in parentheses.

A work written by a single author:  
(Smith, 1999)

A work written by two authors:  
(Gihl, Smith, 2001)

A work written by more than two authors:  
(Gatsby *et al.*, 1998)

When more than one work is cited, references should appear in chronological order:

(Gatsby, 1978; Smith, Gatsby, 1998; Smith *et al.*, 2001)

Works written in the same year should appear in alphabetical order:

(Aleph, 1978; Gatsby, 1978; Zenith *et al.*, 1978).

Works written by the same author(s) in the same year should be listed as a, b, c and so on:

(Smith 1997a, d; Ulrich *et al.*, 1998b)

A particular letter indicates the position of the work in the References section.

The References section is a list of all the works cited in the text. Each of the works cited should appear in alphabetical order and contain the last

names and initials of all the authors. Works written by the same author should appear in chronological order. Works written by the same author in the same year should be listed as a, b, c and so on. For example: 1999a, b, c.

References should appear as follows:

### References

#### AN ARTICLE IN A JOURNAL:

Hanahan D., Weinberg R.A. The hallmarks of cancer // *Cell*. 2000. V. 100, N 2. P. 57–70.

#### A BOOK:

Searle A.G. Comparative Genetics of Coat Color in mammals. L.; N.Y.: Logos Press, Acad. Press, 1968. 303 p.

#### A CHAPTER IN A BOOK:

Golygina V.V., Istomina A.G., Kiknadze I.I. Chromosomal polymorphism in natural populations of *Chironomus balatonicus* Dévai, Wülker et Scholl // Late 20th century research on Chironomidae / Ed. O. Hof-frichter. Aachen: Shaker-Verlag, 2000. P. 89-92.

#### AN ABSTRACT IN THE PROCEEDINGS OF A CONFERENCE:

Peshkov I.M., Likhoshvai V.A., Matushkin Yu.G., Fadeev S.I. On research into hypothetical networks on ecological nature // Proc. of the 4th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2004). Novosibirsk, 25-30 July 2004. Novosibirsk: Inst. Cytol. Genet., 2004. V. 2. P. 128-130.

#### AN ONLINE PUBLICATION:

Smith A., Green P. RepeatMasker. 1999. available at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

Schwenger G.T.F., Mordvinov V.A., Fournier R., Czabotar P., Peroni S., Sanderson C.J. (2000). Interleikin- 5. In: Academic Press Cytokine Reference Database. DOI:10.1006/rwcy.2000.0902.

Журнал «Информационный вестник ВОГиС» включен ВАК Минобрнауки России в Перечень ведущих рецензируемых научных журналов и изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук (по биологическим наукам, на соискание доктора наук). (Редакция, апрель 2008 г.: <http://vak.ed.gov.ru>)

Журнал «Информационный вестник ВОГиС» включен в федеральный почтовый Объединенный каталог «ПРЕССА РОССИИ». Персональный подписной индекс № 42153.

Отредактировано и подготовлено к печати  
в редакционно-издательском отделе ИЦиГ СО РАН

Редакторы: А.А. Ончукова, И.Ю. Ануфриева  
Дизайн: А.В. Харкевич  
Компьютерная графика: А.В. Харкевич, Т.Б. Коняхина  
Компьютерная верстка: Т.Б. Коняхина, Н.С. Глазкова

Подписано в печать 4.03.2009 г.  
Формат бумаги 60×84 1/8. Усл.-печ. л. 27,5. Уч.-изд. л. 26  
Тираж 400. Заказ 65.

Отпечатано в типографии ГУ «Издательство СО РАН»  
630090, Новосибирск, Морской проспект, 2