**Приложение 2**

К статье И.В. Быковой, Н.А. Шмакова, Д.А. Афонникова, А.В. Кочетова, Е.К. Хлесткиной «Достижения и перспективы использования методов высокопроизводительного секвенирования в генетике и селекции картофеля»

**Дополнительные материалы 1. Характеристика электронных ресурсов, содержащих данные полногеномного секвенирования картофеля**

Интернет-ресурс www.potatogenome.net Международного консорциума по секвенированию генома картофеля (The Potato Genome Sequencing Consortium – PGSC) содержит общую информацию по проекту секвенирования и сборки генома. Сборка генома предоставляется в свободном доступе в базе данных Sol Genomics (sgn.cornell.edu). Ftp этой базы данных содержит сборку генома версии v3, основанной на секвенировании гомозиготной линии DM (double monoploid – удвоенный моноплоид). Также в базе данных представлены сборки генома хлоропласта и митохондрии картофеля, сконструированные на основе секвенированных последовательностей линии RH. Кроме того, база данных содержит сборку псевдомолекул, состоящих из секвенированных фрагментов, отнесенных к определенным хромосомам картофеля, но не прошедших картирования на текущие сборки последовательностей этих хромосом. Текущая версия сборки псевдомолекул – 4.03. Наконец, база данных Sol Genomics содержит разметку сборки генома v3 в формате gff, а также последовательности CDS, генов, белков и транскриптов в формате fasta и разметку сборки псевдомолекул v4.03. Разметка сборки псевдомолекул содержит аннотированные гены, DArT-, SSR-, OPA- и DMAP-маркеры, аннотированные при помощи RepeatMasker повторы. Также включены карты SNP в формате GFF3, установленные при помощи микрочипов SolCAP Infinium, SolCAP 8303 Infinium и при помощи библиотек коротких прочтений линии RH, секвенированных с использованием платформы Illumina. Помимо этого, на ftp этой базы данных содержатся последовательности в формате fasta и разметка в формате GFF3 генов, CDS и белков *S. tuberosum*, обнаруженных в ходе проекта International Tomato Annotation Group (ITAG) на основании сходства с последовательностями генов, CDS и белков томата.

База данных Spud (solanaceae.plantbiology.msu.edu) (Hirsch et al., 2014; DOI 10.3835/plantgenome2013.12.0042) содержит сборку генома DM, основанную на данных секвенирования, предоставленных в базе данных NCBI Nucleotide под номером AEWC00000000.1. Представленная сборка имеет длину 726 МБ, 86 % из которых привязано к генетической карте, и содержит 39031 аннотированный ген. Spud DB предоставляет геномный браузер для сборки v4.03 псевдомолекул, как и для предыдущих версий сборки псевдомолекул. В браузерах содержится информация по локусам PGSC и ITAG, а также гомологиям с последовательностям *Arabidopsis thaliana*, винограда, тополя, томата и сборкам транскриптов для 12 других видов Solanaceae из базы данных PlantGDB. Кроме того, представлены данные по 56 RNA-seq библиотекам линий DM и RH и информация по SNP (ресурс SolCAP). В базе данных Spud также доступны для скачивания последовательности псевдомолекул сборки генома картофеля версии 4.04 и 4.03 в формате fasta и разметки генов и различного вида геномных маркеров в формате GFF3. Кроме того, приведены данные по уровням экспрессии генов, полученные на основании транскриптов из 40 библиотек коротких прочтений линии DM и 16 библиотек линии RH. Наконец, для скачивания доступны последовательности BAC-клонов линии RH и концевые сиквенсы BAC-клонов линий RH и DM и фосмид линии DM.

В базе данных Plaza (http://bioinformatics.psb.ugent.be/plaza/versions/plaza_v3_dicots/organism/view/Solanum+tuberosum) представлена информация по сборке генома *S. tuberosum*, в которой приводится 35130 генов, среди которых 45 РНК-кодирующих генов, а остальные – белок-кодирующие. Указано 29436 терминов Gene Ontology и 29049 доменов Interpro. Для скачивания доступна разметка генома, последовательности в fasta формате ДНК и белков, термины Gene Ontology для каждого гена и данные по 35085 семействам генов.

База данных Ensembl Plants (plants.ensembl.org) также содержит сборку генома картофеля версии 3.

В репозитории ENA (European Nucleotide Archive; www.ebi.ac.uk/ena/) представлена геномная сборка SolTub_3.0, также аннотированная в базе данных NCBI Assembly (www.ncbi.nlm.nih.gov/assembly/). Характеристики сборки следующие: общая длина собранных последовательностей – 705.7 млн п. н., количество скаффолдов – 14853, N50 скаффолдов – 1.3 млн п. н., количество контигов – 60068, N50 кон-

тигов – 31.9 т. п. н. Кроме того, сборка включает в себя хлоропластный геном длиной 155 т. п. н. Аннотация сборки генома, представленной в базе данных NCBI, включает в себя 37 882 белок-кодирующих последовательности, 768 tRNA генов, 33 269 генов, 1 960 псевдогенов. Также 84 белок-кодирующих гена, 8 рРНК генов, 45 тРНК генов локализованы на хлоропластной ДНК.

Сборка генома, содержащаяся в базе данных Ensembl Plants (http://plants.ensembl.org/index.html), имеет общую длину 727.4 млн п. н., содержит 39 021 белок-кодирующий ген, 60 163 транскрипта, 3 621 некодирующий ген. На ftp базы данных находятся в свободном доступе последовательности хромосом картофеля и соответствующие разметки. Кроме того, в базе данных Ensembl представлены примерно 194 тыс. EST, картированных на геном при помощи Exonerate.

В ENA SRA (http://www.ebi.ac.uk/ena) хранятся данные 16 последовательностей транскриптомов разных тканей картофеля. В базе данных NCBI SRA (sequence read archives) приведены 1 932 библиотеки коротких рядов, относящиеся к *S. tuberosum*.

База данных SolCAP (solcap.msu.edu) предоставляет данные по SNP и микрочипам картофеля и томата. Доступны для скачивания массивы из 8 303 и 69 011 SNP картофеля, полученных из линий Atlantic, Premier Russet и Snowden. Указывается положение SNP на суперскаффолдах сборки генома и контекст из 50 нуклеотидов. Массив из 8 303 SNP включен в микрочип SolCAP Infinium SNP platform. Кроме того, с использованием данных по сортам картофеля Bintje, Kennebec и Shepody были созданы перечни межсортовых SNP. Для этого были сконструированы сборки перечисленных разновидностей, основанных на опубликованных в NCBI GenBank EST. Для каждого контига, собранного из EST, приводятся все обнаруженные в нем SNP, с указанием локализации и частоты встречаемости отдельных нуклеотидов. Также в этой базе данных приведены различные данные по микрочипам, используемым для поиска SNP в геноме картофеля, доступные для скачивания.

База данных по генным сетям, регуляторным и метаболическим путям растений PMN (Plant Metabolic Network; www.plantcyc.org) содержит следующие данные, относящиеся к *S. tuberosum*: 11 553 гена, 493 метаболических или регуляторных пути, 2 883 ферментативных и 77 транспортных реакций, 5 785 белков.

*S. tuberosum* также представлен в базе данных KEGG (Kyoto Encyclopedia of genes and genomes; http://www.genome.jp/kegg/kegg2.html) под кодовым именем организма ‘sot’. При этом KEGG опирается на сборку генома SolTub_3.0, представленную в базе данных NCBI. В базе данных KEGG присутствует 28 464 белок-кодирующих гена и 2 437 РНК-кодирующих генов. Также представлено 130 метаболических путей, относящихся к *S. tuberosum*.

**Дополнительные материалы 2. Характеристика электронных ресурсов, содержащих данные о микроРНК картофеля**

Данные по микроРНК картофеля аннотированы в нескольких базах данных. Одна из них – miRBase (Kozomara, Griffiths-Jones, 2013), интернет-ресурс, посвященный микроРНК. Ресурс содержит репозиторий, в котором представлены данные по микроРНК, в том числе и 224 микроРНК *S. tuberosum* (http://www.mirbase.org/cgi-bin/mirna_summary.pl?org=stu). Для них приведены первичная и вторичная структура, локализация в геноме картофеля сборки SolTub3.0, семейства генов, к которым они относятся.

Интернет-ресурс miRNEST 2.0 (Szcześniak, Makałowska, 2014) содержит интегрированные из нескольких баз данных сведения по микроРНК, в том числе микроРНК картофеля (http://rhesus.amu.edu.pl/mirnest/copy/browse.php). Всего в базе данных представлено 87 микроРНК *S. tuberosum*, для которых указаны первичная и вторичная структура, семейство генов. База данных дает возможность поиска близких по структуре микроРНК. Для 64 представленных микроРНК наличествуют гомологии с микроРНК, имеющимися в базе данных miRBase, приводится статистическая значимость гомологии и выравнивание двух микроРНК друг относительно друга.

RNAcentral (The RNAcentral Consortium, 2017) – база данных некодирующих РНК, в которой представлены в том числе микроРНК *S. tuberosum* (http://rnacentral.org/search?q=solanum%20tuberosum%20AND%20rna_type:%22miRNA%22). Для картофеля в базе данных представлены 424 микроРНК или прекурсоров микроРНК.

В базе данных spudDB (Hirsch et al., 2014) аннотирована разметка сборки генома картофеля, указывающая положение в геноме микроРНК. Также в разметке генома *S. tuberosum*, приведенной в базе данных Ensembl Plants, указаны микроРНК и их расположение на сборке генома.