

УДК 575.852.112

ЧИСЛО ГОМОЛОГОВ НЕКОТОРЫХ ФЕРМЕНТОВ БИОСИНТЕЗА ТРИПТОФАНА У РАСТЕНИЙ КОРРЕЛИРУЕТ С ДОЛЕЙ БЕЛКОВ, АССОЦИИРОВАННЫХ С ТРАНСКРИПЦИЕЙ

© 2014 г. И.И. Турнаев¹, И.Р. Акбердин¹, В.В. Суслов¹, Д.А. Афонников^{1,2}

¹ Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: turn@bionet.nsc.ru;

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 17 октября 2014 г. Принята к публикации 1 декабря 2014 г.

Путь биосинтеза триптофана (ПБТ) универсален у большинства известных организмов, хотя и отсутствует у животных и некоторых зубактерий. У растений этот путь консервативен, но для разных видов наблюдается различное количество паралогов ферментов, участвующих в этом пути. В настоящей работе исследована возможная роль изменения числа паралогов ПБТ в процессе эволюции. Для этого проведена идентификация паралогов ферментов этого пути в известных полногеномных последовательностях и оценена статистическая связь между числом паралогов ПБТ и сложностью организмов. Показано, что сложность организмов достоверно коррелирует с числом гомологов ферментов синтеза триптофана у растений как для всех гомологов ферментов этого пути суммарно, так и для гомологов трех из шести ферментов этого пути ASA/ASB, PAI и IGPS. Выявленные зависимости могут быть обусловлены тем, что рост сложности организации растений и увеличение числа гомологов ферментов синтеза триптофана являются механизмами эволюционной адаптации к изменчивым условиям наземной среды обитания.

Ключевые слова: путь биосинтеза триптофана, филогенетические сети, морфологическая сложность организмов, адаптация, изменчивость внешних условий.

ВВЕДЕНИЕ

Триптофан синтезируется бактериями, грибами и растениями и необходим для биосинтеза белков. Кроме того, у растений триптофан служит предшественником таких веществ, как фитоалексины, глюкозинолаты, ряд алкалоидов, участвующих в процессах защиты от патогенов и вредителей, а также ауксина, ключевого гормона морфогенеза растений (Radwanski, Last, 1995). Кроме того, триптофан, как аминокислота, является субстратом для синтеза белков. Следует отметить, что реакция синтеза белков двадцатисубстратная, по числу аминокислот. При этом ситуация, когда аминокислоты за счет их низких концентраций могли бы быть регуляторами синтеза белка, является аминокислотным голоданием, в ответ на который у эукариот

клетка снижает скорости синтеза рРНК и тРНК примерно в 10–15 раз, тем самым искусственно создавая избыток аминокислот для реакции белкового синтеза (Картель и др., 2011).

В то же время в одной или двух субстратных реакциях субстраты часто становятся регуляторами собственных реакций. Соответственно триптофан, будучи в пути биосинтеза ауксина у растений субстратом односубстратной реакции, играет роль ее регулятора. Таким образом, триптофан участвует в специфической регуляции экспрессии генов, так как ауксин – регулятор транскрипции ряда генов (Zhao, 2012).

У растений ПБТ из хоризмата включает шесть последовательных реакций, контролируемых ферментами ASA/ASB (антранилат синтаза α /антранилат синтаза β), TRP (ПАТ, фосфорибозил антранилат-трансфераза), PAI

(фосфорибозил-антранилат-изомеразы), IGPS (индол-3-глицерол фосфат синтаза), TSA (триптофан синтеза α), TSB (триптофан синтеза β) (Radwanski, Last, 1995). Гены, кодирующие ферменты ПБТ, несмотря на участие в таком консервативном процессе, как биосинтез белка, у ряда таксонов претерпели дубликации, роль которых до конца не выяснена.

Число генов-компонентов генных сетей в ходе эволюции увеличивается за счет дубликаций генов, с их последующей дивергенцией (Teichmann, Babu, 2004). В ряде случаев за счет дубликации генов образуются мультигенные семейства, включающие гены, кодирующие в одном организме белки с перекрывающимися функциями. К таким семействам относятся гемоглобины, иммуноглобины, антигены гистосовместимости, актины, тубулины, кератины, коллагены, белки теплового шока, клейкие белки слюны, белки хориона, белки кутикулы, желточные белки, фазеолины (запасной гликопротеин семян фасоли), белки YUCCA растений, также как гены гистонов, рибосомных и транспортных РНК (Feliner, Rossello, 2012). Таким образом, ситуация множественности генов, кодирующих белки с перекрывающимися функциями, широко распространена, и паралоги ферментов путей биосинтеза, подобных ПБТ, относятся к таким белкам.

В работе Vogel и Chortia (2006) на 36 видах, представляющих разные таксоны эукариот, было показано, что расширение белковых семейств (оценивалось по изменению общего числа белковых доменов в семействе) коррелирует с ростом числа клеточных типов. Так, положительная корреляция была показана для 194 из 1 219 исследованных семейств белков. Известно, что в процессе эволюции усложнение организма связано с увеличением сложности регуляторной компоненты его генома (Колчанов и др., 2004). Поэтому, если функции триптофана как субстрата для синтеза регуляторных низкомолекулярных соединений существенны для растений, можно предположить, что увеличение числа гомологов ферментов ПБТ, т. е. сложности этого пути, будет коррелировать с увеличением сложности растений в процессе их эволюции. В настоящей работе проведен анализ зависимости между сложностью организации растений и числом паралогов генов ПБТ в геномах

24 видов растений, принадлежащих к разным таксонам. Показано, что между числом паралогов ферментов ПБТ и сложностью организации растений существует значимая положительная корреляция, что свидетельствует о важной регуляторной функции триптофана у растений. Выявленная зависимость может быть связана с необходимостью реализации дифференциальной экспрессии генов-паралогов ПБТ для динамического изменения уровня триптофана в процессе ответа растений на изменяющиеся условия внешней среды.

МАТЕРИАЛЫ И МЕТОДЫ

Формирование выборок гомологов ферментов пути биосинтеза триптофана у растений

Для анализа были взяты последовательности белков ферментов ПБТ у *Arabidopsis thaliana*: ASA (идентификаторы TAIR (Lamesch *et al.*, 2012) AT5G05730, AT2G29690); ASB (AT1G25220, AT5G57890); PAT (TRP) (AT5G17990); PAI (AT1G07780, AT5G05590, AT1G29410); IGPS (AT2G04400, AT5G48220); TSA (AT4G02610, AT3G54640); TSB (AT5G54810, AT4G27070). С помощью программы BLASTP 2.2.29+, (e -value $< 10^{-40}$) проведен поиск гомологов этих ферментов в 24 полностью секвенированных геномах растений из базы данных (БД) PLAZA (Van Bel *et al.*, 2012): пяти зеленых водорослей, одного мха, одного плауна, четырех однодольных растений и 13 двудольных. Для каждого фермента проводили реципрокный поиск, при котором объединялись результаты поиска по всем его паралогам.

Идентификация паралогов ферментов пути биосинтеза триптофана у растений

Множественное выравнивание последовательностей выявленных гомологов проводили с помощью программы Mafft 7.110 (Kato, Toh, 2008). Проверяли наличие и целостность в последовательностях ключевых консервативных доменов, информация о которых была взята из БД CDD версии 3.10 (Marchler-Bauer *et al.*, 2013). Последовательности со значительными нарушениями доменов (крупные вставки или

делеции) удаляли из выборки. Для более точной идентификации ортологов и паралогов в последовательностях белковых семейств мы реконструировали филогенетические сети методом ProteinMLdist/NeighborNet из пакета SplitsTree4 (v. 4.12.8) (Huson, Bryant, 2006), поскольку известно, что филогенетические сети более адекватно отражают филогенетические отношения при наличии в эволюции горизонтального переноса генов, что можно ожидать при анализе эволюции белковых семейств у растений (Richardson, Palmer, 2007).

Для распознавания гомологов ферментов ПБТ нами намеренно был выбран порог распознавания BLASTP 2.2.29+, ($e\text{-value} < 10^{-40}$), позволяющий идентифицировать как белки ферментов ПБТ, так и родственные им белки, выполняющие функции, не связанные с синтезом триптофана. Чтобы в построенных филогенетических сетях определить границу, отделяющую белки ПБТ от родственных им белков с другими функциями, мы для каждого из отдельных кластеров филогенетической сети определяли белки с известными функциями на основе имеющихся литературных данных и информации из БД CDD (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>). Последовательности белков, принадлежащие к кластерам, в которых (1) оказались белки с функциями, не относящимися к синтезу триптофана, и (2) не было белков ПБТ, удаляли из выборки, после чего по новым выборкам заново строили филогенетические сети.

Затем из выборок удаляли последовательности, положение которых на графе филогенетической сети нарушало общепринятую топологию филогенетического дерева растений. Проводили несколько итераций этой процедуры, до тех пор пока нарушения не устранялись. Отметим, что среди удаленных таким образом последовательностей многие содержали делеции и вставки среднего размера в консервативных участках белков. Описанный отбор последовательностей позволил выявить среди гомологов ферментов устойчивые группы белков, с высокой вероятностью выполняющие сходную функцию. Если в полногеномных данных гомологи какого-либо фермента не обнаруживались, для определения их возможного наличия в геноме мы проводили дополнительный поиск в последовательностях

EST организма, представленных в БД GenBank и Ensemble (EST последовательности в большинстве представляют фрагменты мРНК, кодирующих белки, что делает неэффективным их использование для точного определения количества гомологов и реконструкции филогении).

Анализ корреляций между сложностью организмов и количеством паралогов ферментов пути биосинтеза триптофана

Для оценки сложности организмов мы использовали параметр $F_{\text{БАТ}}$ – отношение количества белков, ассоциированных с транскрипцией (БАТ), к общему числу белков организма (Lang *et al.*, 2010). К числу БАТ относятся транскрипционные факторы и другие регуляторы транскрипции. $F_{\text{БАТ}}$ может быть точно оценен на основе полногеномных данных и хорошо коррелирует с такой широко известной характеристикой сложности организмов, как число клеточных типов (Там же). Данные по количеству белков БАТ растений вышеуказанных таксонов (зеленые водоросли, мхи, плауны, одно- и двудольные) взяты из статьи Lang с соавт. (2010). Для оценки значимости взаимосвязи между числом паралогов и количеством белков БАТ мы использовали коэффициент корреляции Пирсона.

РЕЗУЛЬТАТЫ

Количество выявленных гомологов ферментов ПБТ в геномах растений представлено в табл. 1. Оказалось, что водоросли и плауны имеют по одному паралогу каждого фермента, мхи и сосудистые растения – от 1 до 9 паралогов. Девять паралогов ASA выявлено в геноме *Malus domestica*. Для ферментов ПБТ, у которых в полногеномных данных не было найдено гомологов, был проведен анализ библиотек EST. В некоторых случаях (PAI у *Mallus domestica* и ASA у *Chlamidomonas reinhardtii*) гомологичные последовательности не обнаружены ни в полногеномных данных, ни в библиотеках EST, что, вероятно, связано с неполной представленностью генов геномов данных организмов как в полногеномных проектах, так и в библиотеках EST этих видов. Анализ взаимосвязи между

параметром F_{БАТ} растений и числом гомологов ферментов ПБТ выявил статистически значимую ($p = 0,0005$) положительную корреляцию

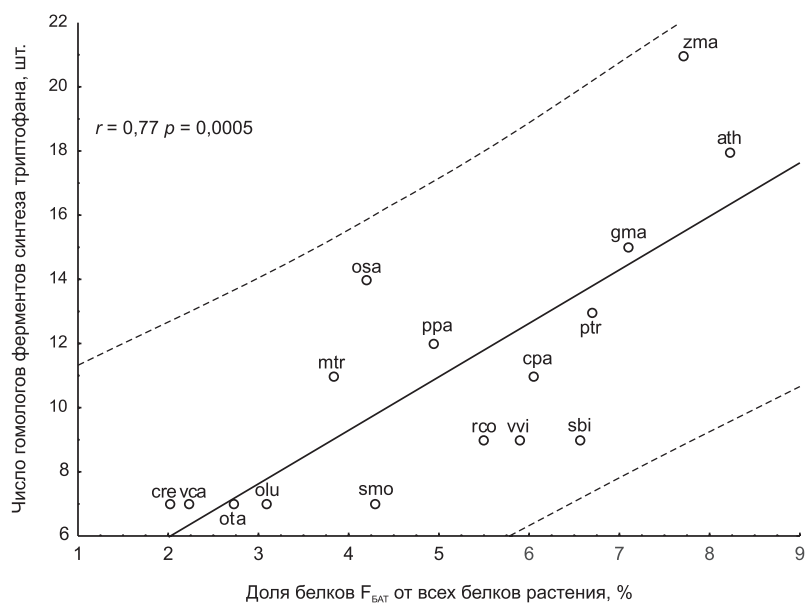
как для отдельных ферментов (ASA/ASB (обе субъединицы), PAI и IGPS; табл. 2), так и для их суммарного числа (см. рисунок).

Таблица 2

Зависимость числа гомологов пути биосинтеза триптофана от параметра F_{БАТ} для каждого из ферментов этого пути

Семейство белков	r	p
ASA	0,56	0,0025*
ASB	0,69	0,031*
TRP	0,29	0,27
PAI	0,79	0,0033*
IGPS	0,59	0,015*
TSA	0,43	0,094
TSB	0,38	0,15

Примечание. Представлены коэффициенты корреляции (r) и уровни статистической значимости (p) для этих двух параметров. * достоверные значения ($p < 0,05$).



Зависимость числа гомологов ферментов пути биосинтеза триптофана от доли белков БАТ среди всех белков растения, F_{БАТ}. По оси X отложено значение параметра F_{БАТ}. По оси Y – суммарное число гомологов по всем ферментам пути биосинтеза триптофана. Коэффициент корреляции и уровень его значимости приведены на графике. Штриховыми линиями обозначены границы 95 % доверительного интервала.

cre – *Chlamidomonas reinhardtii*, ota – *Ostreococcus tauri*, olu – *Ostreococcus lucimarinus*, vca – *Volvox carteri*, ppa – *Physcomitrella patens*, smo – *Selaginella moellendorffii*, zma – *Zea mays*, sbi – *Sorghum bicolor*, osa – *Oryza sativa spp japonica*, vvi – *Vitis vinifera*, ath – *Arabidopsis thaliana*, ptr – *Populus trichocarpa*, cpa – *Carica papaya*, rco – *Ricinus communis*, gma – *Glycine max*, mtr – *Medicago truncatula*.

ОБСУЖДЕНИЕ

Сравнительный анализ функций белков в парах паралога ферментов ПБТ, выявленных с помощью нокаутного анализа их генов у некоторых видов семенных растений, свидетельствует о специфичной экспрессии разных гомологов одного фермента синтеза триптофана в зависимости от условий среды. Например, у *A. thaliana* ген *ASA2* экспрессируется на конститутивном базальном уровне, а уровень экспрессии его паралога *ASA1* в десять раз выше и может дополнительно повышаться в ответ на ранение или инфицирование бактериальными патогенами (Niyogi, Fink, 1992). Такие различия характерны и для паралога *ASA1*, *ASA2* у *Ruta graveolens* (Bohlmann *et al.*, 1995), а также *OASA1*, *OASA2* у однодольного растения *Oryza sativa japonica* (Tozawa *et al.*, 2001). Ген, кодирующий TSB2 у *A. thaliana*, продуцирует только 10 % мРНК триптофан синтазы β в тканях листа, экспрессируется конститутивно на базальном уровне и необходим для роста растения при недостатке освещения. Его паралог, ген, кодирующий TSB1, экспрессирует 90 % мРНК триптофан синтазы β , но лишь при ярком освещении (Last *et al.*, 1991). Эти данные демонстрируют важное значение дубликаций в ПБТ: один ген из пары экспрессируется на конститутивном базальном уровне и требуется для синтеза белков и вторичных метаболитов триптофана на минимальном уровне, необходимым, например, для биосинтеза белков, а второй ген пары экспрессируется индуцибельно на высоком уровне в ответ на внешние условия, благоприятные для быстрого роста и развития растения, или на стрессовые факторы (бактериальную инфекцию, ранения, недостаток освещения и др.), повышая концентрацию триптофана, часть молекул которого может служить субстратом для синтеза ауксина, участвуя в регуляторных процессах.

Растения в ходе эволюции были вынуждены приспосабливаться к увеличению изменчивости условий внешней среды (резкое увеличение пессимальности условий и амплитуды их изменений при выходе растений из воды на сушу) за счет усложнения их морфологии: увеличение количества тканей (появление корней, листьев, развитого стебля, тканей, позволивших освоить сушу папоротникам, хвощам, плаунам; переход

от спор к семенам – у семенных папоротников, позволивший растениям оторваться от экосистем с высокой влажностью; появление тканей цветов и плодов у покрытосеменных растений) и усложнения молекулярно-генетических систем растений. Это усложнение морфологии позволило растениям занимать в процессе эволюции все большее число экологических ниш, а также освоить ниши со значительными колебаниями условий (жара – холод, дожди – засуха и т. д.). Это обстоятельство может объяснить выявленные нами положительные корреляции между долей генов БАТ (мерой сложности организма) и числом гомологов ферментов синтеза триптофана как для всего пути биосинтеза в целом, так и для трех (*ASA/ASB*, *PAI* и *IGPS*) из шести его ферментов в частности.

БЛАГОДАРНОСТИ

Работа поддержана грантом РФФИ 14-14-00734.

ЛИТЕРАТУРА

- Картель Н.А., Макеева Е.Н., Мезенко А.М. Генетика = Genetics: энциклопедический словарь. Минск: Беларуская навука, 2011. 758 с.
- Колчанов Н.А., Суслов В.В., Гунбин К.В. Моделирование биологической эволюции: регуляторные генетические системы и кодирование биологической организации // Информационный вестник ВОГИС. 2004. Т. 8. № 2. С. 86–99.
- Bohlmann J., DeLuca V., Eilert U., Martin W. Purification and cDNA cloning of anthranilate synthase from *Ruta graveolens*: modes of expression and properties of native and recombinant enzymes // *Plant J.* 1995. V. 7. No. 3. P. 491–501.
- Feliner G.N., Rossello J.A. Concerted evolution of multigene and homoelogenous recombination families. *Plant genome diversity*. Springer Wein Heidelberg, N. Y., Dordrecht, London, 2012. 171 p.
- Huson D.H., Bryant D. Application of phylogenetic networks in evolutionary studies // *Mol Biol Evol.* 2006. V. 23. No. 2. P. 254–267.
- Katoh K., Toh H. Recent developments in the MAFFT multiple sequence alignment program // *Brief Bioinform.* 2008. V. 9. No. 4. P. 286–298.
- Lamesch P., Berardini T.Z., Li D. *et al.* The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools // *Nucleic Acids Res.* 2012. V. 40. P. D1202–D1210.
- Lang D., Weiche B., Timmerhaus G. *et al.* Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and

- correlation with complexity // *Genome Biol. Evol.* 2010. V. 19. No. 2. P. 488–503.
- Last R.L., Bissinger P.H., Mahoney D.J. *et al.* Tryptophan mutants in Arabidopsis: the consequences of duplicated tryptophan synthase beta genes // *Plant Cell*. 1991. V. 3. No. 4. P. 345–358.
- Marchler-Bauer A., Zheng C., Chitsaz F. *et al.* CDD: conserved domains and protein three-dimensional structure // *Nucleic Acids Research*. 2013. V. 41. P. D348–D352.
- Niyogi K.K., Fink G.R. Two anthranilate synthase genes in Arabidopsis: defense-related regulation of the tryptophan pathway // *Plant Cell*. 1992. V. 4. No. 6. P. 721–733.
- Radwanski E.R., Last R.L. Tryptophan biosynthesis and metabolism: biochemical and molecular genetics // *Plant Cell*. 1995. V. 7. P. 921–934.
- Richardson A.O., Palmer J.D. Horizontal gene transfer in plants // *Journal experimental Botany*. 2007. V. 58. No. 1. P. 1–9.
- Teichmann S.A., Babu M.M. Gene regulatory network growth by duplication // *Nat Genet.* 2004. V. 36. No. 5. P. 492–496.
- Tozawa Y., Hasegawa H., Terakawa T., Wakasa K. Characterization of rice anthranilate synthase alpha-subunit genes OASA1 and OASA2. Tryptophan accumulation in transgenic rice expressing a feedback-insensitive mutant of OASA1 // *Plant Physiol.* 2001. V. 126. No. 4. P. 1493–1506.
- Van Bel M., Proost S., Wischnitzki E. *et al.* Dissecting plant genomes with the PLAZA comparative genomics platform // *Plant Physiology*. 2012. V. 158. No. 2. P. 590–600.
- Vogel C., Chothia C. Protein family expansions and biological complexity // *PLoS Comput Biol.* 2006. V. 2. No. 5. P. e48.
- Zhao Y. Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants // *Mol. Plant*. 2012. V. 5. No. 2 P. 334–338.

THE NUMBER OF HOMOLOGS OF SOME ENZYMES IN THE TRYPTOPHAN BIOSYNTHESIS PATHWAY CORRELATES WITH THE PROPORTION OF PROTEINS ASSOCIATED WITH TRANSCRIPTION IN PLANTS

I.I. Turnaev¹, I.R. Akberdin¹, V.V. Suslov¹, D.A. Afonnikov^{1,2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: turn@bionet.nsc.ru;

² Novosibirsk National Research State University, Novosibirsk, Russia

Summary

The tryptophan biosynthesis pathway (TBP) is ubiquitous in most known organisms, being absent only from animals and some bacteria. It is conserved in plants, although various species differ in the number of TBP enzyme paralogs. In the current work we investigated a putative possible role of changes in the number of paralogs of TBP enzymes in the course of plant evolution. We identified TBP enzyme paralogs in plant species with fully sequenced genomes and estimated the relationship between its number and organismal complexity. It is shown that organismal complexity significantly correlates with the total number of TBP paralogs and for some enzymes specifically (ASA/ASB, PAI, and IGPS). We suggest that such a relationship arises because both organismal complexity and the increasing number of paralogs may be important for the evolutionary adaptation of land plants to variable environmental conditions.

Key words: tryptophan biosynthesis pathway, phylogenetic network, morphological complexity of organisms, adaptation, variability of environmental conditions.