

structures of analogous glycosyltransferases (LanGT1/LndGT1/Orf27, LanGT2/LndGT2/Orf29 and LanGT4/LndGT4/Orf32) synthesizing carbohydrate chain of landomycin E all three producers were identical more than 75 %, meanwhile there was less than 33 % identity of enzymes structures (LanGT1/LanGT2/LanGT4, LndGT1/LndGT2/LndGT4, Orf27/Orf29/Orf32) within each particular microorganism. **Conclusions.** Evolutionary relationship between analogous enzymes was revealed: they are orthologs. Two enzymes (LanGT3 and LanGT1) of *S. cyanogenus S136* were identified as paralogous ones. *Key words:* structure, *in silico* analysis, homology, ortholog, paralog, glycosyltransferase.

УДК 633.52:577.21:632.165

РАБОКОНЬ А.Н., ПОСТОВОЙТОВА А.С., ПИРКО Я.В., БЛЮМ Я. Б.

Институт пищевой биотехнологии и геномики НАН Украины,

Украина, 04123, г. Киев, Осиповского, 2а, e-mail: nastya-rabokon@rambler.ru

АНАЛИЗ ГОМОЛОГОВ ГЕНОВ ОСНОВНЫХ БЕЛКОВ ЦИТОСКЕЛЕТА У РАЗЛИЧНЫХ ВИДОВ ВЫСШИХ РАСТЕНИЙ

Гены цитоскелетных белков, часть из которых относятся к генам «домашнего хозяйства» (housekeeping genes), необходимы для поддержания важнейших жизненных функций как отдельно взятых клеток, так и всего организма в целом [1]. Недавние исследования показали, что длина интронов некоторых из них, точнее ILP (Intron Length Polymorphism), может быть удобным и надежным инструментом для изучения генетического разнообразия и филогенетических связей между различными видами растений. Также ILP неплохо зарекомендовал себя для предварительной характеристики и распознавания различных генотипов растений [2]. В связи с этим изучение экзон-интронной структуры генов, кодирующих цитоскелетные белки (прежде всего актина, α - и β -тубулина) у различных видов растений представляет практический интерес с точки зрения расширения возможного спектра молекулярных маркеров за счет более глубокого анализа именно экзон-интронной структуры этих генов [3]. В связи с этим, целью работы было проведение биоинформационного анализа последовательностей генов актина (основного белка микрофиламентов), α - и β -тубулина (базовый белок микротрубочек) у видов ризовидки Таля (*Arabidopsis thaliana*), риса (*Oriza sativa*), картофеля (*Solanum tuberosum*) и томата (*Solanum lycopersicum*) для изучения их интрон-экзонной структуры.

Материалы и методы

Полные последовательности генов актина и тубулина *A. thaliana* были взяты из базы данных GenBank (www.ncbi.nlm.nih.gov/genbank/). В геноме этого растения

аннотировано 8 последовательностей генов актина (At2g37620(Act1), At3g18780(Act2), At3g53750(Act3), At5g59370(Act4), At5g09810(Act7), At1g49240(Act8), At3g12110(Act11), At3g46520(Act12)), 6 последовательностей генов α -тубулина (At1g64740(Tua1), At1g50010(Tua2), At5g19770(Tua3), At1g04820(Tua4), At5g19780(Tua5), At4g14960(Tua6)) и 9 последовательностей генов β -тубулина (At1g75780(Tub1), At5g62690(Tub2), At5g6270(Tub3), At5g44340(Tub4), At1g20010(Tub5), At5g12250(Tub6), At2g29550(Tub7), At5g23860(Tub8), At4g20890(Tub9)). В дальнейшем полные последовательности, отдельно кодирующие области, а также их продукты были использованы для поиска и анализа гомологов у *O. sativa*, *S. tuberosum* и *S. lycopersicum*. Геномы выше упомянутых видов полностью секвенированы, но в базе данных GenBank до сих пор отсутствует их аннотация. Для множественного выравнивания исследуемых нуклеотидных секвенсов были использованы программы Clustal 2.0 и UGENE [5, 6].

В качестве матрицы для поиска потенциальных гомологов были использованы гены: актина – actin-1 (ACT1_ARATH), α -тубулина – tubulin alpha-1 (TBA1_ARATH) и β -тубулина tubulin beta-1 (TBB1_ARATH) из *A. thaliana*. С помощью инструмента BLASTN был проведен поиск в базе данных Phytozome v9.1 (www.phytozome.net). Отбор гомологов был основан на процентных показателях идентичности и сходства генов, а также на полноте нуклеотидных последовательностей и транслируемых продуктов.

Результаты и обсуждение

В результате анализа генома риса выявлено 11 нуклеотидных последователь-

ностей, кодирующих актин: Os01g73310, Os02g38340, Os03g50885, Os03g56970, Os03g61970, Os05g01600, Os05g36290, Os08g03440, Os10g36650, Os11g06390, Os12g06660. В среднем последовательности имели 5 экзонов (90–2280 п.н.) и 4 интрона (длиной от 74 п.н. до 1694 п.н.). Обнаружено 4 гена тубулина: Os07g38730, Os03g11970, Os03g51600, Os11g14220, которые содержат в среднем по 4 экзона (200–1010 п.н.) и 3 интрона (90–990 п.н.); 9 генов β -тубулина: Os03g56810, Os03g45920, Os03g01530, Os01g59150, Os01g18050, Os06g46000, Os02g07060, Os05g34170, Os06g07280 – 3 экзона (250–1080 п.н.) и 2 интрона (90–1110 п.н.) соответственно.

В геноме томата также обнаружено 11 гомологов гена актина: Solyc00g017210, Solyc01g104770, Solyc03g078400, Solyc04g011500, Solyc04g071260, Solyc05g054480, Solyc06g076090, Solyc10g080500, Solyc10g086460, Solyc11g005330, Solyc11g065990. Они содержали в среднем 4 экзона (40–2700 п.н.) и 3 интрона (68–860 п.н.). В геноме *S. lycopersicum* было 4 гена α -тубулина: Solyc08g006890, Solyc04g077020, Solyc02g087880, Solyc02g091870 – в среднем 4 экзона (90–895 п.н.) и 3 интрона (100–1150 п.н.); и 9 генов β -тубулина: Solyc04g081490, Solyc10g085620, Solyc10g085020, Solyc10g080940, Solyc12g089310, Solyc06g035970, Solyc06g005910, Solyc03g025730, Solyc03g118760 – 4 экзона (110–2020 п.н.) и 3 интрона (50–1630 п.н.).

В геноме картофеля было выявлено 9 нуклеотидных последовательностей потенциальных генов актина: PGSC0003DMG400003985, PGSC0003DMG400000439, PGSC0003DMG400018449, PGSC0003DMG400027746, PGSC0003DMG400023708, PGSC0003DMG400019204, PGSC0003DMG400008912, PGSC0003DMG400030319, PGSC0003DMG400023429; среднее количество интронов составило 3 (71–

479 п.н.), а экзонов – 4 (30–2600 п.н.). Найдено 5 генов α -тубулина: PGSC0003DMG400004272, PGSC0003DMG400011537, PGSC0003DMG400001320, PGSC0003DMG400030627, PGSC0003DMG400008752 – 4 экзона (100–1245 п.н.) и 3 интрона (90–4410 п.н.); 9 генов β -тубулина:

PGSC0003DMG400009938, PGSC0003DMG400011088, PGSC0003DMG400028193, PGSC0003DMG400019131, PGSC0003DMG400029337, PGSC0003DMG400014296, PGSC0003DMG400029926, PGSC0003DMG400020850, PGSC0003DMG400030431 – 3 экзона (100–2050 п.н.) и 2 интрона (79–1266 п.н.).

Следует отметить, что у *A. thaliana* в нуклеотидных последовательностях, кодирующих актин, выявлено 4 экзона и 3 интрона (кроме АСТ3 – 3 экзона, 2 интрона); в генах α -тубулина в среднем также 4 экзона и 3 интрона; все гены β -тубулина содержат 3 экзона (270–690 п.н.) и 2 интрона (80–790 п.н.).

В результате проведенного биоинформационного анализа установлено, что уровни идентичности кодирующих областей (экзонов) у изучаемых растений составляли в среднем: для генов актина – 77%, α -тубулина – 75%; β -тубулина – 78%. Степень идентичности в процентах приведена в табл. 1–3. Поскольку интроны и регуляторные участки генов являются более вариабельными как по длине, так и по нуклеотидному составу, уровни идентичности полных последовательностей генов оказались значительно меньшими.

Таким образом, идентичность экзонов данных генов у всех четырех изученных видов растений оказалась достаточно высокой, что указывает на высокое сходство кодирующих областей цитоскелетных генов у разных видов растений.

Таблица 1. Анализ сходства гомологов кодирующих областей генов актина (%)

Вид растения	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. tuberosum</i>	<i>S. lycopersicum</i>
<i>A. thaliana</i>	100	71	80	80
<i>O. sativa</i>	71	100	71	71
<i>S. tuberosum</i>	80	71	100	84
<i>S. lycopersicum</i>	80	71	84	100
В среднем	77			

Таблица 2. Анализ сходства гомологов кодирующих областей генов α -тубулина (%)

Вид растения	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. tuberosum</i>	<i>S. lycopersicum</i>
<i>A. thaliana</i>	100	72	69	74
<i>O. sativa</i>	72	100	72	77
<i>S. tuberosum</i>	69	72	100	79
<i>S. lycopersicum</i>	74	77	79	100
В среднем	75			

Таблица 3. Анализ сходства гомологов кодирующих областей генов β -тубулина (%)

Вид растения	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. tuberosum</i>	<i>S. lycopersicum</i>
<i>A. thaliana</i>	100	78	79	79
<i>O. sativa</i>	78	100	76	76
<i>S. tuberosum</i>	79	76	100	82
<i>S. lycopersicum</i>	79	76	82	100
В среднем	78			

Выводы

Полученные данные свидетельствуют о том, что последовательности экзонов исследованных генов у различных видов высших растений весьма схожи, а последовательности интронов имеют

существенные различия, то есть обладают высокой степенью полиморфизма. Таким образом, интроны цитоскелетных генов могут рассматриваться в качестве потенциальных маркеров при изучении генетического разнообразия видов растений.

Литература

1. Chen R., Gyokusen M., Nakazawa Y., Gyokusen K. Selection of housekeeping genes for transgene expression analysis in *Eucommia ulmoides* Oliver using real-time RT-PCR [Электронный ресурс] // J. Bot. – 2010. – 2010. – 7 p. – Article ID 230961, – Режим доступа: <http://dx.doi.org/10.1155/2010/230961>
2. Braglia L., Manca A., Mastromauro F., Breviario D. cTBP: A successful intron length polymorphism (ILP)-based genotyping method targeted to well defined experimental needs // Diversity. – 2010. – 2. – P. 572–585.
3. Nicot N., Hausman J.F., Hoffmann L., Evers D. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress // J. Exp. Bot. – 2005. – 56. – P. 2907–2914.
4. Yamada K., Lim J., Dale J.M. et al. Empirical analysis of transcriptional activity in the *Arabidopsis* genome // Science. – 2003. – 302, N 5646. – P. 842–846.
5. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. Clustal W and Clustal X version 2.0 // Bioinformatics. – 2007. – 23. – P. 2947–2948.
6. Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit // Bioinformatics. – 2012. – 28. – P. 1166–1167.

RABOKON A.N., POSTOVOYTOVA A.S., PIRKO YA.V., BLUME YA.B.

Institute of Food Biotechnology and Genomics, Nat. Acad. of Sci. of Ukraine, Ukraine, 04123, Kyiv, Osipovskogo str., 2A, e-mail: nastya-rabokon@rambler.ru

ANALYSIS OF GENE HOMOLOGY OF SOME CYTOSKELETAL PROTEINS IN DIFFERENT SPECIES OF HIGHER PLANTS

Aims. The genes of cytoskeletal proteins are important to maintain essential life processes and functions at different stages of the ontogenesis. Intron length polymorphism (ILP) of some of them can be used as convenient and reliable tool for the study of genetic diversity and phylogenetic relationships between different plant species. ILP also can be used for genotyping different plants. In this connection, analysis of exon-intron structure of some cytoskeletal protein genes in different plant species represents considerable interest to extend the possible range of molecular markers. **Methods.** Using *Arabidopsis* actin and tubulin genes we have performed the similarity search of genes via TBLASTN predicted proteins of the *Oriza sativa*, *Solanum tuberosum* and *Solanum lycopersicum* genomes available at Phytozome v9.1. **Results.** We conducted out selection of actin, α - and β -tubulin gene homologues in the genomes of analyzed plants; studied its exon-intron structure; established the percent identity of CDS in these genes. **Conclusions.** Obtained data show that the exon sequences of these cytoskeleton genes in various species of higher plants are very similar, and intron sequences have significant differences, that reflects a high degree of polymorphism. Thus, the introns of these genes may be used as molecular markers in ILP-analysis.

Key words: gene, cytoskeletal proteins, actin, tubulin, homologue, exon, intron polymorphism.