

ОРФАННЫЕ ГЕНЫ КАК ДВИГАТЕЛЬ ЭВОЛЮЦИИ

Журавлева А.А.

*Журавлева Анастасия Александровна — младший научный сотрудник,
Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр
Институт цитологии и генетики Сибирского отделения Российской академии наук»
г. Новосибирск*

Аннотация: в статье анализируются гены, встречающиеся только у одного вида или в группе близкородственных видов, называют орфанными. Они имеют короткие открытые рамки считывания, низкий уровень экспрессии, и специфичны для разных тканей, стадий развития и стрессовых условий. Белки орфанных генов могут выполнять различные функции в клетке, обеспечивая возможность для развития новых функций и признаков, таких как метаболизм сахаров, реакции на стресс и адаптация к условиям окружающей среды. Они могут возникать различными способами: посредством дупликации и дивергенции генов, а также из мобильных элементов. Важную роль в их возникновении и эволюции играют эпигенетические изменения. Помимо всего прочего, митохондриальный геном также вносит вклад в появление орфанных генов. Однако идентификация и аннотирование этих генов представляют собой сложную задачу. На сегодняшний день проблемы идентификации орфанных генов включают в себя отсутствие сборки генома в масштабе хромосом, доступность геномных аннотаций, ограниченное количество высококачественных эталонных геномов и технологий секвенирования, необходимых для обнаружения орфанных генов.

Ключевые слова: орфанные гены, мобильные элементы, эволюция.

В эволюции растительного генома важную роль играет формирование генов *de novo*, так называемых орфанных генов [1]. Орфанные гены (ORF) не имеют гомологии с какими-либо известными кодирующими последовательностями отдаленно родственных видов [2]. Такие гены могут развиваться из уже существующих последовательностей, кодирующих белок (дупликация и дивергенция, горизонтальный перенос генов), или возникать *de novo* из предковых некодирующих последовательностей. Гены *de novo* характеризуются короткими открытыми рамками считывания, низким уровнем экспрессии и специфичны для разных тканей, стадий развития и стрессовых условий [3]. По этим причинам их трудно идентифицировать и отличить от псевдогенов или артефактов аннотаций [4]. Орфанные гены имеют узкое филогенетическое распространение, при этом до 30% из них обнаруживаются во всех геномных каталогах каждого анализируемого вида [5].

Орфанные гены кодируют белки, не имеющие значительного сходства последовательностей у близкородственных видов. Они отличаются от белков неорфанных генов относительно небольшими (по количеству аминокислот) размерами, что указывает на минимальное число экзонов [6].

Орфанные гены в растениях обладают функциональными характеристиками, сходными с неорфанными генами, что говорит об их схожем эволюционном происхождении [1]. Остается неясным, какова связь между адаптивной эволюцией и появлением новых агрономических признаков. Можно сделать предположение, что орфанные гены могли эволюционировать вместе с генами белков, с которыми они взаимодействуют, выполняя определенные функции, а также с регуляторами/факторами для реализации предполагаемых функций, включая модуляцию морфологических признаков [1]. Несмотря на их функциональную важность, немного известно о лежащих в основе молекулярных механизмах, с помощью которых орфанные гены участвуют в разнообразных биологических процессах.

Систематические исследования на приматах [7] и растениях показывают, что орфанные гены появились посредством дупликации и дивергенции генов, а также из мобильных элементов (TE) [8]. Мобильные элементы составляют значительную часть всех известных геномов. Считается, что мобильные элементы играют важную роль в эволюции генов *de novo*: предоставляют новые промоторы для уже существующих ORF, создают новые кодирующие последовательности посредством ретропозиции, захвата генов или слияния экзонов [9]. Было обнаружено, что мобильные элементы связаны примерно с половиной идентифицированных генов, специфичных для линии, у приматов и риса [10]. Это позволяет предположить, что мобильные элементы играют важную роль в эволюции генов всех эукариот, в том числе именно мобильные элементы являются основными факторами геномного разнообразия *Triticeae* [11]. Тем не менее, до сих пор неизвестна подлинная частота появления генов *de novo* [9].

Существует две основные гипотезы эволюции генов *de novo* [9]. В «*expression- first model*», транскрибируемые некодирующие области (т.е. протогены) развиваются нейтрально и подвергаются генетическим вариациям давлению отбора. Таким образом, протогены могут приобретать адаптивные мутации и постепенно превращаться в более сложные и функциональные гены *de novo*. С другой

стороны, модель «ORF-first» постулирует, что потенциальные ORF уже присутствуют в геноме и транслируются после приобретения сайтов связывания транскрипционных факторов [9].

Митохондриальный геном также участвует в появлении новых орфанных генов [12]. Высокая скорость геномных перестроек в нем приводит к появлению новых последовательностей, которые впоследствии встраиваются в ядерную ДНК, что может быть причиной появления орфанных генов. Этот процесс может включать перенос митохондриальной ДНК, содержащей сформированные гены, либо открытые рамки считывания, посредством транспозиции или с помощью других геномных механизмов [12].

Значительно влияют на возникновение орфанных генов эпигенетические факторы. [13]. На ранних стадиях формирования эпигенетический ландшафт орфанных генов *de novo* может значительно различаться среди популяций и внутри них. Регуляторные области орфанного гена обладают эпигенетическими сигнатурами, характерными для энхансеров, в отличие от не-орфанных генов, которые имеют классические промоторы. Более того, некоторые неэкспрессируемые орфанные гены демонстрируют репрессивные модификации гистонов, приводящие к снижению транскрипции [13]. Некоторые группы исследователей также предполагают, что эпигенетическое ремоделирование может способствовать появлению новых орфанных генов [14].

Экспрессия орфанных генов может отличаться в различных тканях и органах. Анализ кластеров генов показывает, что орфанные гены в первую очередь демонстрируют тканеспецифическую экспрессию [14]. Так, *Ms1* у *Roaseae* кодирует фосфолипидсвязывающий белок, который играет важную роль в гаметогенезе, а также в развитии экины пыльца [15]. Мужская стерильность наблюдалась у растений пшеницы, ячменя и *Brachypodium*, лишенных *Ms2*, еще одного орфанного гена, который демонстрирует специфическую для пыльника экспрессию, обусловленную вставкой ретротранспозона в промотор [16]. Таким образом, экспрессия орфанных генов может различаться в разных тканях [14].

Идентификация орфанных генов имеет решающее значение для понимания их эволюции и функциональной роли в различных организмах. Такой метод сравнительной генетики, как BLAST был одним из первых методов, использовавшихся для обнаружения орфанных генов путем сравнения последовательностей геномов разных видов [17]. Этот метод имеет ограничения, поскольку некоторые орфанные гены могли быстро эволюционировать и не сохраняться у разных видов, что затрудняет их обнаружение. Однако в таком случае можно использовать такие методы, как BLASTp, Филостратиграфия и ORFan-Finder, для которых данные ограничения отсутствуют [18]. Текущие проблемы идентификации орфанных генов включают отсутствие сборки генома в масштабе хромосом, доступность геномных аннотаций, ограниченное количество высококачественных эталонных геномов и технологий секвенирования, необходимых для обнаружения орфанных генов [19]. Так, после вычислительного и статистического анализа геномов риса открыто только 1926 из 18 398 потенциальных орфанных генов [20]. Чтобы выяснить взаимодействуют ли орфанные гены с другими клеточными компонентами и центральными регуляторами на молекулярном и геномном уровнях во время роста и развития тканеспецифичных реакций, а также получить данные о новых генах используются подходы включающие анализ экспрессии [21] и паттернов коэкспрессии, такие как WGCNA (анализ сети взвешенной корреляции генов, также известный как анализ сети взвешенной коэкспрессии генов), за которым следует анализ терминов онтологии генов (GO) с использованием базы данных KEGG (Kyoto Encyclopedia of Genes and Genomes - Киотская энциклопедия генов и геномов) [22] и другие инструменты. Так, исследование на дыне показало, что 18 коэкспрессируемых орфанных генов чрезмерно представлены в биохимических процессах, включая взаимные превращения пентозы и глюкуроната, биосинтез кутина, суберина и воска, а также метаболизм крахмала и сахарозы [23].

Заключение

Хотя орфанные гены растений обычно не имеют сходства последовательностей, они обладают схожими функциональными характеристиками с не-орфанными генами, что позволяет им функционировать вместе. Эти гены играют важную роль в процессах метаболизма сахаров, в реакциях на биотические и абиотические стрессы, а также во множестве адаптационных процессов, специфических для отдельных видов [24]. Кроме того, орфанные гены широко распространены во всех царствах, но фундаментальным барьером для определения их функции является отсутствие функциональных мотивов и складок взаимозависимых компонентов или эффектов, которые все еще различимы [1, 14]. Наконец, орфанные гены являются одним из факторов эволюции и могут действовать как катализатор видообразования.

Список литературы

1. Poretti M., Praz C.R., Sotiropoulos A.G., Wicker T. A survey of lineage specific genes in *Triticeae* reveals *de novo* gene evolution from genomic raw material // Plant Direct., 2023. №7 (3).

2. Tautz D., Domazet-Lošo T. The evolutionary origin of orphan genes // *Nat Rev Genet.*, 2011. №12(10). P. 692–702.
3. Schlötterer C. Genes from scratch – the evolutionary fate of de novo genes // *Trends Genet.*, 2015. №31(4). P. 215–9.
4. Singh U., Syrkín Wurtele E. How new genes are born // *Elife*, 2020. №19(9).
5. Khalturin K., Hemmrich G., Fraune S. et al. More than just orphans: are taxonomically-restricted genes important in evolution? // *Trends Genet.*, 2009. №25(9). P. 404–13.
6. Heames B., Schmitz J., Bornberg-Bauer E.A. Continuum of Evolving De Novo Genes Drives Protein-Coding Novelty in *Drosophila* // *J Mol Evol.*, 2020. №88(4). P. 382–98.
7. Toll-Riera M., Bosch N., Bellora N. et al. Origin of Primate Orphan Genes: A Comparative Genomics Approach // *Mol Biol Evol.*, 2008. №26(3). P. 603–12.
8. Donoghue M.T., Keshavaiah C., Swamidatta S., et al. Evolutionary origins of Brassicaceae specific genes in *Arabidopsis thaliana* // *BMC Evol Biol.*, 2011. №11(1). P. 47.
9. Van Oss S.B., Carvunis A.R. De novo gene birth // *PLOS Genet.*, 2019. №15(5). P. 100-160.
10. Jin G., Zhou Y., Yang H., et al. Genetic innovations: Transposable element recruitment and de novo formation lead to the birth of orphan genes in the rice genome // *J Syst Evol.*, 2021. №59(2). P. 341–51.
11. Wicker T., Gundlach H., Spannagl M., et al. Impact of transposable elements on genome structure and evolution in bread wheat // *Genome Biol.*, 2018. №19(1). P. 103.
12. O'Conner S., Li L. Mitochondrial Fostering: The Mitochondrial Genome May Play a Role in Plant Orphan Gene Evolution // *Front Plant Sci.*, 2020. №11.
13. Werner M.S., Sieriebriennikov B., Prabh N., et al. Young genes have distinct gene structure, epigenetic profiles, and transcriptional regulation // *Genome Res.*, 2018. №28(11). P. 1675–87.
14. Fakhar A.Z., Liu J., Pajerowska-Mukhtar K.M., et al. The ORFans' tale: new insights in plant biology // *Trends Plant Sci.*, 2023. №28(12). P. 1379–90.
15. Wang Z., Li J., Chen S., Heng Y., et al. Poaceae-specific *MS1* encodes a phospholipid-binding protein for male fertility in bread wheat // *Proc Natl Acad Sci.*, 2017. №114(47). P. 12614–9.
16. Ni F., Qi J., Hao Q., et al. Wheat *Ms2* encodes for an orphan protein that confers male sterility in grass species // *Nat Commun.*, 2017. №8(1). P. 15121.
17. Albà M.M., Castresana J. On homology searches by protein Blast and the characterization of the age of genes // *BMC Evol Biol.*, 2007. №7(1). P. 53.
18. Qi M., Zheng W., Zhao X., et al. QQS orphan gene and its interactor *NF-YC4* reduce susceptibility to pathogens and pests // *Plant Biotechnol J.*, 2019. №17(1). P. 252–63.
19. Zhang W., Gao Y., Long M., et al. Origination and evolution of orphan genes and de novo genes in the genome of *Caenorhabditis elegans* // *Sci China Life Sci.*, 2019. №62(4). P. 579–93.
20. Guo W.J., Li P., Ling J., et al. Significant Comparative Characteristics between Orphan and Nonorphan Genes in the Rice (*Oryza sativa* L.) Genome // *Comp Funct Genomics*, 2007. №2007. P. 1–7.
21. Mishra B., Kumar N., Mukhtar M.S. Network biology to uncover functional and structural properties of the plant immune system // *Curr Opin Plant Biol.*, 2021. №62. P. 102-57.
22. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes // *Nucleic Acids Res.*, 2000. №28(1). P. 27–30.
23. Ma D., Lai Z., Ding Q., et al. Identification, Characterization and Function of Orphan Genes Among the Current Cucurbitaceae Genomes // *Front Plant Sci.*, 2022. №13.
24. Arendsee Z.W., Li L., Wurtele E.S. Coming of age: orphan genes in plants // *Trends Plant Sci.*, 2014. №19(11). P. 698–708.