

СРАВНИТЕЛЬНАЯ И ЭВОЛЮЦИОННАЯ ГЕНОМИКА ЛАНЦЕТНИКОВ, ХОРДОВЫХ И ПОЗВОНОЧНЫХ ЖИВОТНЫХ

Томпсон С.,¹ Романов М.Н.,^{1,2} Гриффин Д.К.¹

¹Университет Кента, Кентербери, Великобритания;

²ФГБОУ ВО «Московская государственная академия
ветеринарной медицины и биотехнологии – МВА имени
К.И. Скрябина», Москва, Россия

E-mail: m.romanov@kent.ac.uk

Аннотация

Ланцетники (*Amphioxus*), обитающие в теплых морях, являются единственными живыми представителями подтипа головохордовых (*Cephalochordata*). Исследования их генома помогают сделать выводы об эволюции и филогении всех хордовых. Ветвь головохордовых животных, окаменелости которых относятся к кембрийскому периоду, характеризуется медленной скоростью эволюции, что сначала было выяснено на фенотипическом, а затем и на геномном уровне. Именно эта медленная дивергенция ланцетников, а также их филогенетическое положение в качестве базальной хордовой группы, делают их незаменимыми для исследования эволюции хордовых. Полногеномное сравнение трех групп хордовых, включая ланцетников, оболочников и позвоночных, способствует уточнению филогенетических отношений между ними. В то время как сравнение консервативных семейств генов показало, что в тех случаях, когда геном ланцетников содержит базовый набор генов (например, для передачи сигналов клеток), позвоночные животные часто имеют два-четыре паралога, возникших в результате двух событий дупликации всего генома, которые являются консервативными на очень большом филогенетическом расстоянии.

Ключевые слова: геном, сравнительная геномика, эволюционная геномика, филогения, позвоночные животные, хордовые, ланцетники

Введение

Ланцетники (род *Branchiostoma*, или *Amphioxus*) – это мелкие фильтрующие морские животные, которые проводят большую часть своего времени на морском дне, частично зарываясь в песок. Их основной источник пищи – одноклеточные водоросли, получаемые путем фильтрации морской воды через их лишенную челюстей ротовую полость.

Эти примитивные хордовые животные относятся к подтипу бесчерепных, или головохордовых (*Cephalochordata*). В настоящее время насчитывается 29 видов ланцетников [1].

Морфология ланцетника

Ключевые анатомические сходства между ланцетником и другими хордовыми включают хорду, полую спинную нервную трубку, жаберные щели, пронизывающие глотку, сегментированное тело, сложенное из сомитов, эндостиль и постанальный хвост на определенной стадии их жизненного цикла [2] (рис. 1). Однако, хотя ланцетник очень похож на позвоночных, у него отсутствуют некоторые ключевые новообразования позвоночных, такие как мигрирующие клетки нервного гребня, сильно обособленная кость головного мозга и эндоскелет [3,4].

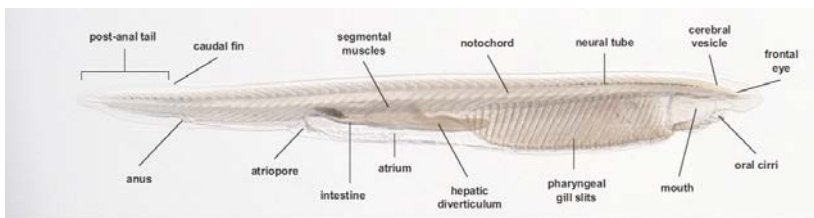


Рис. 1. Анатомия ланцетника [5]

Данные из описания окаменелостей [6] и эмбриологии [7] указывают на то, что морфология ланцетника во многих отношениях очень похожа на то, что можно ожидать от хордовых предков.

Структура генома ланцетника

Геном ланцетника, вероятно, не испытал больших эволюционных изменений, наблюдаемых в других ветвях хордовых [8,9]. Однако он не характеризуется эволюционным застоем за 500 миллионов лет с момента его расхождения от остальной части хордовой ветви и развил некоторые собственные производные особенности, включая дупликацию гена опсина и появление новых генов, которые, как предполагается, функционируют в качестве генов врожденного иммунитета [9].

Геном ланцетника размером 520 млн пар оснований (Mb) был расшифрован группой во главе с Николасом Патнэмом [10] с использованием полногеномной стратегии секвенирования методом дробовика [11]. Геномную ДНК выделяли из гонад самца вида *Branchiostoma floridae*. Несмотря на то, что эта работа была опубликована в 2008 году, только в 2020 году появилась физическая и генетическая карты ланцетника, благодаря чему удалось реконструировать геном его 19 пар хромосом [12]. Первоначально более половины предсказанных генов были помещены в скаффолды с более чем 130 или более генами, поэтому предварительная сборка генома была достаточно большой, что позволило осуществить сравнительный анализ консервативной синтении относительно геномов других видов.

Анализ предварительной сборки с использованием информации открытых рамок считывания (ORF), полученных из тегов экспрессируемых последовательностей (EST), подтвердил, что сборка охватывала более 95% содержимого генома ланцетника, которое кодирует белки. Одновременно с помощью методов прогнозирования генов (*ab initio*) [7] предположили, что гаплонидный геном ланцетника содержит оценочно 21900 локусов, кодирующих белки. Специфические приспособления ланцетника, которые возникли за 500 млн лет с момента его дивергенции от остальной хордовой ветви, сравнительно небольшие по сравнению со значительной вторичной потерей генов, наблюдаемой в геноме оболочника – асцидии *Ciona intestinalis* (размером 155 Mb) [9]. Примером такой вторичной потери генов можно считать наличие многих древних генов гомеобокса билатеральных

животных, включая *Hox7*, *Hox8*, *NK1* и *Vax*, в геноме *B. floridae* и их отсутствие в геноме *C. intestinalis* [9].

Филогенетические взаимоотношения в группе вторичноротых (Deuterostomia)

Традиционно представители подтипа Cephalochordata (ланцетники) считались ближайшими живыми родственниками позвоночных животных, а представители Urochordata (оболочники) рассматривались самым ранним дивергирующим подтипом хордовых [13]. Эта теория, хотя и получила всеобщее признание, подтверждается лишь ограниченным количеством морфологических особенностей, которые довольно неоднозначны, например, метамерная сегментация [14].

В последние годы эта точка зрения была поставлена под сомнение при проверке данных о последовательностях белков в сочетании с методами статистической вероятности [15]. Однако ограниченный выбор видов хордовых в этих исследованиях означал невозможность сделать какие-либо твердые выводы, поскольку недостаточное число анализируемых таксонов вероятно отрицательно влияло на результаты филогении [16]. В другом исследовании с использованием более крупной выборки таксонов из 13 хордовых были обнаружены доказательства того, что оболочники являются сестринской группой позвоночных, а также высказано предположение (хотя и с незначительной статистической вероятностью), что ланцетник более тесно связан с иглокожими, что делает хордовых парафилетической группой [17].

В более поздних исследованиях было подтверждено, что оболочники являются наиболее близким подтипом к позвоночным, но также были получены доказательства того, что Cephalochordata относятся к группе базальных хордовых [10,18], отвергая более ранние предположения [17], что ланцетник обладает наиболее близким сходством с иглокожими. Было также показано, что сравнение динамики интронов ланцетника и других животных обеспечивает независимую поддержку новой модели филогении хордовых (рис. 2) [10].

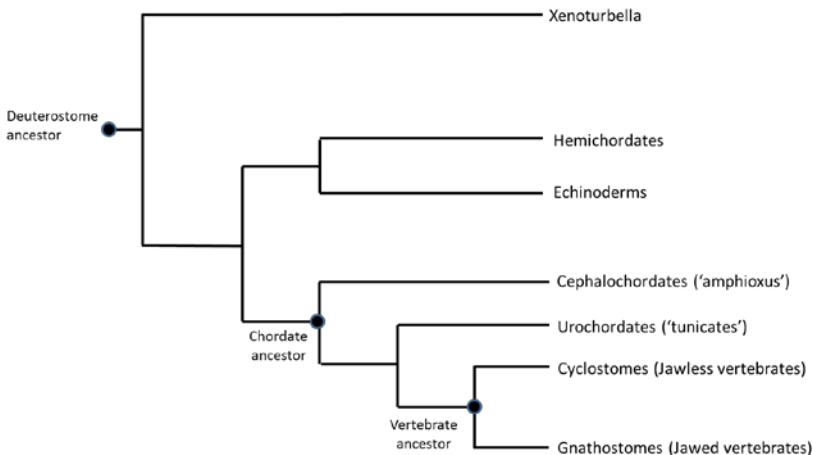


Рис. 2. Филогения вторичноротых (Deuterostomia).

Суперфилюм Deuterostomia состоит из четырех типов: Xenoturbella как самый базальный [18,19], за которым следуют полухордовые (Hemichordata) и иглокожие (Echinodermata) и далее идут хордовые (Chordata) с бесчерепными (Cephalochordata), т.е. ланцетниками (Amphioxus), в качестве базальных организмов [10], за которыми следуют оболочники (Urochordata) и позвоночные (Vertebrata)

Полногеномные сравнения эукариот

Использование филогенетических анализов

Геномные данные, полученные в результате секвенирования геномов первичноротых беспозвоночных, таких как нематода *C. elegans* [20] и различных насекомых [21,22], а также вторичноротых, таких как оболочник *C. intestinalis* [8], и геномов позвоночных, в том числе человека [23], мыши [24] и шимпанзе [25], были положены в основу сравнения с геномами организмов, занимающих ключевые филогенетические позиции. К последним относят такие виды, как ланцетник [10] и в последнее время морскую миногу из класса круглоротых (Cyclostomata), который входит в группу бесчелюстных позвоночных и является внешней группой по отношению к челюстноротым (Gnathostomata; рис. 2) [26]. При этом можно получить представление о струк-

туре и составе генов предковых геномов у хордовых и позвоночных животных.

Дупликация всего генома в эволюционной линии позвоночных

Вероятно, самая популярная гипотеза возникновения сложности, наблюдаемой у высших позвоночных, восходит к идеям Сусуму Оно [27], который первым предположил, что серия полногеномных дупликаций обеспечивает дополнительные генетические вариации для развития новых генов в эволюции позвоночных. Согласно этой теории, геномы ланцетника и позвоночных были увеличены в размере по сравнению с геномами беспозвоночных в ходе одного-двух раундов дупликации всего генома (2R). Однако в более поздней работе [28] было высказано предположение о том, что дупликация произошла после дивергенции ланцетника от ветви позвоночных.

В совсем недавних исследованиях генома была предложена модель четырехкратной консервативной синтении у видов позвоночных [29], которая окончательно продемонстрировала наличие двух раундов полногеномной дупликации и расширила предыдущие представления о событиях полногеномной дупликации. Этот вывод был сделан на основании сравнительных исследований специфических, представляющих интерес участков в геномах хордовых. Эти участки включали, но не ограничивались кластером генов *Hox* [30] и область главного комплекса гистосовместимости [31]. Таким образом, было получено свидетельство, подтверждающее гипотезу о том, что увеличенная сложность высших позвоночных в организации таких систем, как нервная и эндокринная, является результатом повышенной гибкости генома, возникшей в результате 2R.

Накоплены также сведения, показывающие повышенное сохранение дублированных в процессе 2R генов, которые обогащены функциями, связанными с регуляцией транскрипции, передачей сигналов, развитием и нейронной активностью. Например, гены, участвующие в передаче сигнала, имели в два раза больше шансов сохраниться в двух или более копиях по сравнению со средней скоростью сохранения генов [10]. Это согласуется с идеей о том, что паралоги, созданные в ходе 2R,

были вовлечены в развитие новых особенностей, которые мы ассоциируем с биологией позвоночных.

Хронология событий дупликации всего генома

Четырехкратная синтения геномов ланцетника и человека демонстрирует, что события 2R произошли после дивергенции головохордовых, но до разделения костистых (лучеплавниковых) рыб и четвероногих, поскольку было обнаружено, что обе группы демонстрируют четырехкратную консервативную макросинтению, что указывает на то, что они обе подверглись 2R. Естественно, следующий вопрос, который нужно задать, касается хронологии 2R: произошли ли обе дупликации достаточно быстро одна за другой, возможно, даже одновременно, или они были значительно разделены по времени? Для ответа на этот вопрос было представлено несколько различных теорий (рис. 3).

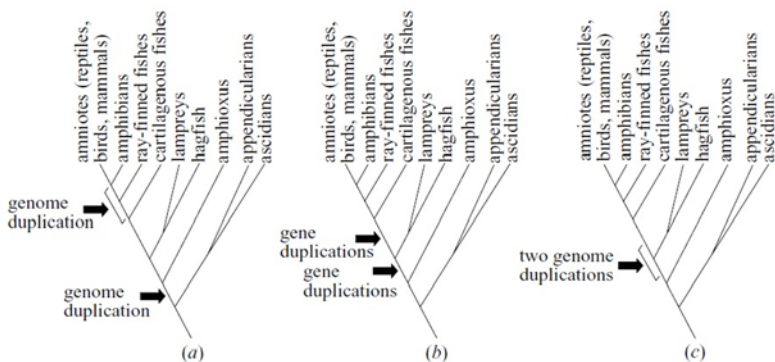


Рис. 3. Филогения хордовых и хронология геномных дупликаций [29].

Показана относительная хронология согласно воззрениям:

a) Susumu Ohno [27], b) Holland et al. [28] и c) Furlong, Holland [29]

В дополнение к этому позднее был осуществлен филогенетический анализ с использованием секвенированных образцов австралийского каллоринха *Callorhynchus milii*, который продемонстрировал значительную консервативную макросинтению между этой хрящевой рыбой и человеком [32]. С этим согласуются данные другого исследования филогенетических тополо-

гий, которое датирует все дубликации до разделения между хрящевыми и костными позвоночными [33], что согласуется со временем 2R на рис. 3b и 3c.

Вместе с тем остается вопрос, какая хронология 2R из предложенных интерпретаций более точна. Putnam et al. [10] попытались определить время событий 2R относительно дивергенции хрящевых рыб, оболочников и бесчелостных позвоночных (миног). Результат явно отличался в зависимости от того, случилась ли дивергенция миноги задолго до 2R (как для асцидии) или после (как для иглобрюха). Это говорит о том, что либо минога отошла от челюстных позвоночных в период между двумя четко разделенными событиями дубликации (рис. 3b), либо одно или оба события дубликации совпали с дивергенцией миноги (рис. 3c). Однако исследователи до сих пор не смогли собрать достаточных доказательств того, какой из этих сценариев мог иметь место. Временные интервалы, отделяющие сценарий совпадения (рис. 3c) от четкого разделения событий дубликации (рис. 3b), определяются процессом редиплоидизации, который включает потерю большинства дублицированных генов и расхождение выживших паралогов [29,34].

Синтения ланцетника и позвоночных

Было обнаружено, что синтения между черновым геномом ланцетника (*B. floridae*) [8,10] и геномами позвоночных (человека, костистых рыб и курицы) широко наблюдается в масштабе целых хромосом (макросинтении) при ограниченном консерватизме локальной генной организации (микросинтении) [10]. Интересно, что, несмотря на более длительное время дивергенции между ланцетником (в сравнении с оболочником *C. intestinalis*) и позвоночными, синтения ланцетника с позвоночными намного выше. Так, 74% скаффолдов ланцетника имеют значительную концентрацию ортологов из одной и той же хромосомы человека по сравнению с 9%, наблюдаемыми в скаффолдах *C. intestinalis*, тем самым подтверждая приемлемость использования ланцетника для замещения последнего общего хордового предка.

Сравнительный анализ этих консервативных геномных характеристик позволил реконструировать наборы генов в 17 группах сцепления, рассматриваемых в качестве протохромосом, которые, как считается, напоминают таковые у последнего общего предка хордовых. Если сравнить эти протохромосомы с геномами позвоночных, обнаруживается четкая структура четырехкратной консервативной макросинтезии, что указывает на два раунда полногеномной дупликации (2R) в эволюционной линии позвоночных [10].

В результате анализа генома ланцетника было также открыто большое число (1299) функционально консервативных некодирующих элементов, половина из которых, как было показано, обладала энхансерной активностью *in vivo* [35]. Эти консервативные элементы не обнаруживаются в геномах других базальных хордовых (таких как *C. intestinalis*), хотя они сильно дуплицированы у позвоночных. Поэтому кажется вероятным, что регуляция генов у ланцетника является более репрезентативной для предкового состояния (с точки зрения факторов транскрипции). Эта гипотеза подтверждается исследованиями, которые показали, что в этой роли не могут выступать ни оболочники (из-за потери и перестройки генов), ни позвоночные (из-за дупликации генома) [9].

Список литературы

1. Poss S.G., Boschung H.T. Lancelets (Cephalochordata: Branchiostomatidae): how many species are valid? *Isr J Zool*, 1996, 42(Suppl): S13–S66.
2. Rychel A.L. et al. Evolution and development of the chordates: collagen and pharyngeal cartilage. *Mol Biol Evol*, 2006, 23: 541–549.
3. Holland N. et al. Sequence and developmental expression of *AmphiDll*, an amphioxus *Distal-less* gene transcribed in the ectoderm, epidermis and nervous system: insights into evolution of craniate forebrain and neural crest. *Development*, 1996, 122: 2911–2920.
4. Shimeld S.M., Holland P.W. Vertebrate innovations. *Proc Natl Acad Sci U S A*, 2000, 97: 4449–4452.

5. Garcia-Fernández, J. et al. From the American to the European amphioxus: towards experimental Evo-Devo at the origin of chordates. *Int J Dev Biol*, 2009, 53: 1359–1366.
6. Chen, J.-Y. et al. A possible Early Cambrian chordate. *Nature*, 1995, 377: 720–722.
7. Yu J.-K. et al. Axial patterning in cephalochordates and the evolution of the organizer. *Nature*, 2007, 445: 613–617.
8. Dehal P. et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*, 2002, 298: 2157–67.
9. Holland L.Z. et al. The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res*, 2008, 18: 1100–1111.
10. Putnam N.H. et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 2008, 453: 1064–1071.
11. Weber J.L., Myers E.W. Human whole-genome shotgun sequencing. *Genome Res*, 1997, 7: 401–409.
12. Howell W.M., Boschung H.T. Chromosomes of the lancelet, *Branchiostoma floridae* (order Amphioxi). *Experientia*, 1971, 27: 1495–1496.
13. Rowe T. Chordate phylogeny and development. *Assembling the Tree of Life* (eds Cracraft J., Donoghue M.J., Donoghue M.M., Diversitas program), 2004, pp.384–409.
14. Schaeffer B. Deuterostome monophyly and phylogeny. *Evolutionary Biology* (eds Hecht M.K., Wallace B., Prance G.T.), 1987, pp.179–235.
15. Blair J.E., Hedges S.B. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol*, 2005, 22: 2275–2284.
16. Delsuc F. et al. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*, 2005, 6: 361–375.
17. Delsuc F. et al. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, 2006, 439: 965–968.
18. Bourlat S.J. et al. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature*, 2006, 444, 85–88.
19. Perseke M. et al. The mitochondrial DNA of *Xenoturbella bocki*: genomic architecture and phylogenetic analysis. *Theory Biosci*, 2007, 126: 35–42.
20. Stein L.D. et al. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol*, 2003, 1: E45.
21. Adams M.D. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287: 2185–2195.
22. Touré Y.T. et al. (2004). The *Anopheles gambiae* genome: next steps for malaria vector control. *Trends Parasitol*, 20: 142–149.

23. Lander E.S. et al. Initial sequencing and analysis of the human genome. *Nature*, 2001, 409: 860–921.
24. Mouse Genome Sequencing Consortium et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 2002, 420: 520–562.
25. Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 2005, 437: 69–87.
26. Smith J.J. et al. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nature Genet*, 2013, 45: 415–421, 421e1–2.
27. Ohno S. *Evolution by Gene Duplication*. Springer, Berlin, 1970.
28. Holland P.W. et al. Gene duplications and the origins of vertebrate development. *Dev Suppl*, 1994: 125–133.
29. Furlong R.F., Holland P.W.H. Were vertebrates octoploid? *Philos Trans R Soc Lond B Biol Sci*, 2002, 357: 531–544.
30. Garcia-Fernández J., Holland P.W. Archetypal organization of the amphioxus Hox gene cluster. *Nature*, 1994, 370: 563–566.
31. Castro L.F.C. et al. An antecedent of the MHC-linked genomic region in amphioxus. *Immunogenetics*, 2004, 55: 782–784.
32. Venkatesh B. et al. Survey sequencing and comparative analysis of the elephant shark (*Callorhynchus milii*) genome. *PLoS Biol*, 2007, 5: e101.
33. Robinson-Rechavi M. et al. Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. *Mol Biol Evol*, 2004, 21: 580–586.
34. Wolfe K.H. Yesterday's polyploids and the mystery of diploidization. *Nat Rev Genet*, 2001, 2: 333–341.
35. Hufton A.L. et al. Deeply conserved chordate noncoding sequences preserve genome synteny but do not drive gene duplicate retention. *Genome Res*, 2009, 19: 2036–2051.

Comparative and evolutionary genomics of lancelet, chordates and vertebrate animals

Thompson S.,¹ Romanov M.N.,^{1,2} Griffin D.K.¹

¹University of Kent, Canterbury, UK;

²K. I. Skryabin Moscow State Academy of Veterinary Medicine and Biotechnology, Moscow, Russia

Abstract

Lancelets, or amphioxi, found in warm seas are the only living representatives of the subphylum Cephalochordata. Their genome research is helpful in making inferences about chordate evolution and phylogeny. The cephalochordate lineage, with a fossil record dating back to the Cambrian period, has been characterized as having a slow evolutionary rate, first noticed at a phenotypic level and later at a genomic level. It is this slow divergence, as well as its phylogenetic position as the basal chordate group, that makes amphioxus a useful proxy for investigating chordate evolution. Whole genome comparisons for three chordate groups including lancelets, tunicates and vertebrates have elucidated their phylogenetic relationships. Whilst comparison of conserved gene families has shown that where the amphioxus genome contains a basic set of genes (e.g., for cell signaling), vertebrates often have 2, 3 or 4 paralogs derived from two whole genome duplication events, conserved across this very wide phylogenetic distance.

Key words: genome, comparative genomics, evolutionary genomics, phylogeny, vertebrates, chordates, lancelets