

МОЛЕКУЛЯРНАЯ ЭВОЛЮЦИЯ, ФИЛОГЕНИЯ И СТРАТЕГИЯ ЭВОЛЮЦИОННОГО АНАЛИЗА БЕЛКОВ

**Андрей ШУТОВ, Ирина КАХОВСКАЯ, Анжела РУДАКОВА, Наталия ЛАПТЕВА,
Анна ШАЛЛАУ***

НИЛ биохимии растений

** Институт генетики растений (Гатерслебен, Германия)*

În lucrare este prezentată o strategie rațională de analiză a evoluției unei familii de proteine, bazată pe orientarea succesiunilor de aminoacizi componenți. Principiile de bază ale strategiei sunt axate pe alegerea celor mai adecvate regiuni de succesiuni, a seturilor de succesiuni și a derivaților succesiunilor. Eficacitatea strategiei a fost demonstrată prin studierea evoluției familiei de proteine Ariadne Ring-Ibr-Ring și compararea cu datele publicate de alți autori.

An expedient strategy for evolutionary analysis of protein families based on alignment of their amino acid sequences is described. General principles of the strategy consists of selection of most suitable sequence region, sequence set and out-group sequences. The efficiency of the strategy was demonstrated using evolutionary analysis of Ariadne family of Ring-IBR-Ring proteins in comparison with the data previously published by other authors.

Введение

Результаты анализа молекулярной эволюции любого из разнообразных семейств белков, основанного на элайнменте их аминокислотных последовательностей, по крайней мере для эукариот, обычно строго согласуются с филогенией видов, установленной по совокупным признакам. Тем не менее, такую согласованность не всегда можно обнаружить в опубликованных исследованиях эволюции белковых семейств. Причиной этого может являться либо неадекватная стратегия проведенного анализа, либо формирование на уровне древних эукариот отдельных эволюционных ветвей, развивавшихся впоследствии независимо друг от друга в каждом из исследуемых видов. К последнему случаю, собственно, и сводятся различия между классической филогенией организмов и дивергентной молекулярной эволюцией их белков. Однако очевидно, что и в пределах каждой из независимых ветвей молекулярная эволюция белков должна следовать филогении видов.

Нами описана универсальная стратегия анализа эволюции семейств белков, применимая для исследования белков как узкого круга родственных видов, так и широкого круга представителей всех трех царств в системе организмов – архебактерий, бактерий и эукариот. Рациональность описанной стратегии иллюстрируется сопоставлением результатов исследования эволюции белков семейства Ariadne [1-3], полученных нами, а также другими авторами.

Методы и стратегия анализа

Были использованы следующие программы: BLAST сайта <http://www.ncbi.nih.gov/> для поиска аминокислотных последовательностей белков Ariadne, гомологичных белкам *Arabidopsis thaliana* [1], ClustalW2 для их элайнмента и TREECON [4] для их эволюционного анализа.

Для получения статистически достоверной топологии эволюционного дерева, описывающего последовательное развитие первичных структур белков того или иного семейства, мы руководствовались описанной ниже стратегией.

1. Выбор анализируемого участка последовательностей. Насколько возможно, такой участок должен быть протяженным. При этом гомологичность всех анализируемых последовательностей белков в пределах всего выбранного участка должна быть безусловной. В идеале, безусловно консервативные позиции элайнмента должны определять границы анализируемого участка.

2. Создание коллекции аминокислотных последовательностей для анализа (наиболее критичный этап). Из широкого предварительного набора последовательностей должны быть выбраны только те из них, что занимают позицию, наиболее близкую к главному стволу эволюционного дерева. Такой выбор должен опираться на оценку статистической достоверности формирования кластеров и отдель-

ных ветвей при предварительном анализе широкого набора последовательностей. В пределах каждого из кластеров должно быть отражено также таксономическое разнообразие видов, которым принадлежат анализируемые белки. В идеале, в пределах каждого из основных кластеров, последовательность эволюционных ветвей должна следовать филогении видов.

3. Выбор одной или нескольких последовательностей, пригодных к использованию в качестве корней эволюционного дерева. Такие последовательности в обязательном порядке должны принадлежать анализируемому семейству, но при этом проявлять наименьшее сходство со всеми другими последовательностями. В идеале, корневые последовательности должны принадлежать филогенетически наиболее древним из организмов, белки которых входят в анализируемую коллекцию.

4. Межкластерный множественный элайнмент. Присутствие вставок, специфичных для индивидуальной последовательности в пределах того или иного кластера, часто приводит к неадекватному совмещению последовательностей вставок с последовательностями других кластеров. В идеале, когда специфичность вставок доказана, их следует удалять перед межкластерным элайнментом.

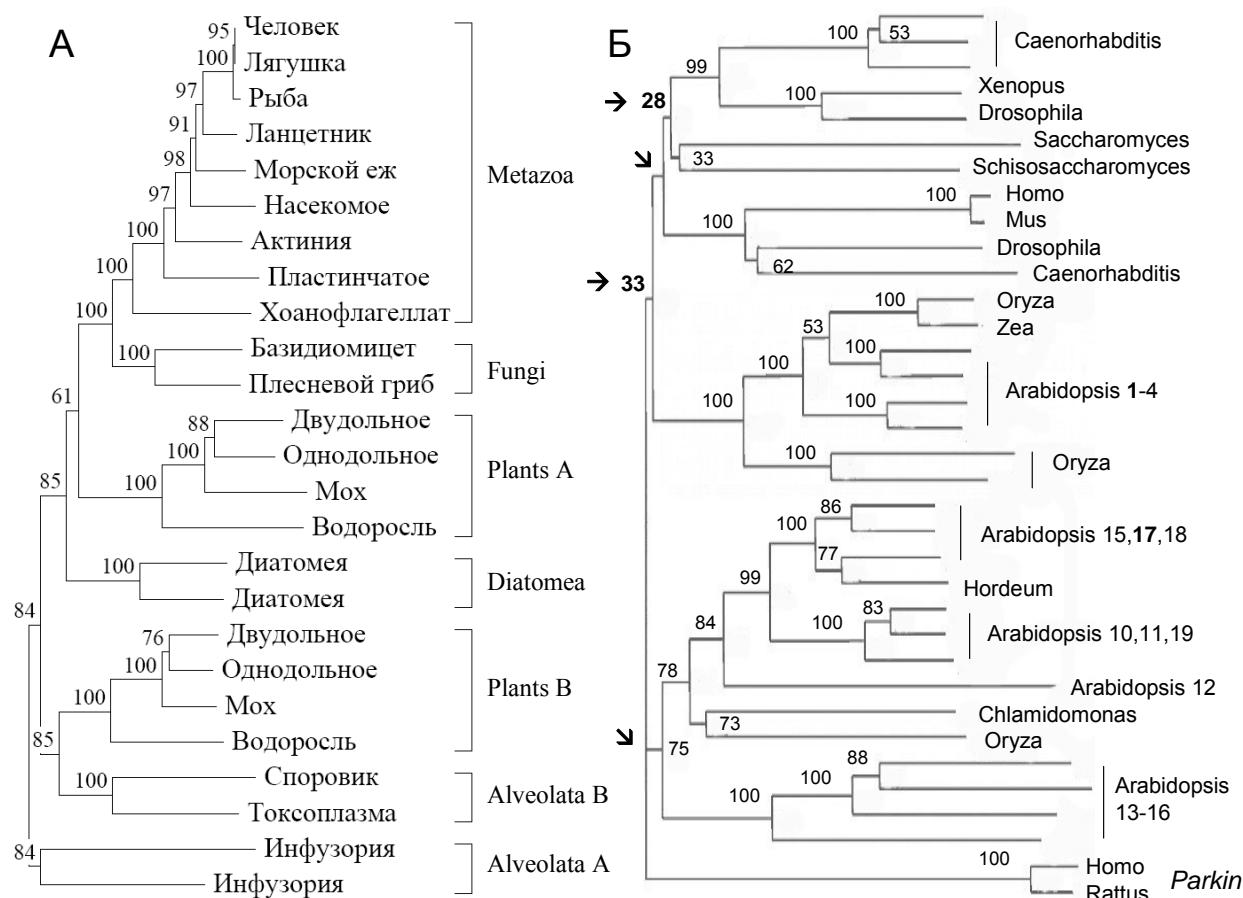
Результаты исследования и их обсуждение

Семейство белков Agiadne выбрано нами в качестве объекта эволюционного анализа по нескольким причинам. Это семейство, являющееся членом обширного суперсемейства белков, ответственных за многие жизненно важные процессы [1-3] и характерных для всех эукариот, привлекает внимание многих исследователей. Тем не менее, лишь в работе Mladek et al. [2] предпринята попытка описать эволюцию Agiadne как самостоятельного семейства. Последовательности Agiadne достаточно консервативны и содержат Ring-IBR-Ring домен, в котором присутствуют чередующиеся глобально консервативные остатки Cys и His [1-3]. Поэтому, по крайней мере в пределах Ring-IBR-Ring домена, результаты элайнмента, проведенного нами и другими авторами, практически совпадают. Это дает возможность прямого сопоставления двух стратегий эволюционного анализа – описанной выше и использованной Mladek et al. [2]. Наконец, известны многие сотни аминокислотных последовательностей Agiadne из разнообразных источников: от древних до высокоразвитых эукариот.

Результаты исследования эволюции белков семейства Agiadne, полученные нами в соответствии с описанной выше стратегией и Mladek et al. [2], сопоставлены на рис. 1. Последовательное формирование кластеров и отдельных ветвей на рис. 1А получено при высокой статистической поддержке и строго согласуется с филогенией таксонов [5] и отдельных организмов эукариот (рис. 1В) при единственном исключении, упомянутом в начале настоящей статьи и обсуждаемом ниже.

Из топологии эволюционного дерева следует, что на уровне Alveolata от основного эволюционного ствола, развивающегося в строгом соответствии с канонами классической дивергентной эволюции (от Alveolata А до Fungi/Metazoa), отделилась ветвь (Alveolata В и Plants В). В соответствии с высокой статистической поддержкой вероятность такого события достаточно велика. Поскольку эволюционное расстояние между Alveolata В и Plants В огромно, можно предположить, что их объединение в общий кластер отражает древнее событие горизонтального переноса генов, давшее начало формированию существенно различающихся вариантов А и В генов, одновременно присутствующих в геномах растений на четырех показанных на рис. 1А эволюционных уровнях (зеленые водоросли, мхи, однодольные, двудольные). В этом контексте следует упомянуть о том, что коллекция последовательностей Agiadne в геномах животных и грибов настолько велика, что можно утверждать отсутствие в них генов В-типа. Тем не менее, не исключено, что связующее эволюционное звено между Alveolata В и Plants В может быть обнаружено впоследствии.

Несмотря на низкую статистическую поддержку формирования кластеров (или даже на ее отсутствие) Mladek et al. [2] также наблюдали два типа растительных белков семейства Agiadne, из которых только один, соответствующий кластеру Plants А на рис. 1А, обнаруживал родство с белками животных и грибов (рис. 1Б). В остальном различия между результатами анализов, приведенными на рис. 1, с очевидностью свидетельствуют в пользу описанной нами стратегии эволюционного анализа. Упомянем о трех источниках этих различий, обусловленных различиями в стратегии анализа.



В

Metazoa: Человек *Homo* aah51877, Лягушка *Xenopus* aai60576, Рыба *Danio* aah67684, Ланцетник *Branchiostoma* een48283, Морской еж *Strongylocentrotus* XP_784503, Насекомое *Drosophila* saa66953, Актиния *Nematostella* edo36616, Пластинчатое *Trichoplax* edv21022, Хоанофлагеллат *Monosiga* edq88427. **Fungi:** Базидиомицет *Ustilago* eak84310, Плесневой гриб *Aspergillus* eed56180. **Plants A:** Двудольное *Arabidopsis* At4g34370, Однодольное *Oryza* eee58654, Мох *Physcomitrella* edq82307, Водоросль *Chlamidomonas* edp00463. **Diatomea:** *Phaedactylum* eec43003, *Thalassiosira* eed93169. **Plants B:** Двудольное *Arabidopsis* At2g31510, Однодольное *Oryza* cae02043, Мох *Physcomitrella* edq60605, Водоросль *Chlamidomonas* edp04392, **Alveolata B:** Споровик *Cryptosporidium* eea06700, Токсоплазма *Toxoplasma* eeb01843. **Alveolata A:** Инфузории *Paramecium* cak68077, *Tetrahymena* ear95505.

Рис. 1. Эволюция белков семейства Ariadne.

А. Результаты настоящей работы. Проанализированная область (497 позиций элайнмента) соответствует основной части последовательностей белков. Цифры над ветвями отражают статистическую поддержку кластеров (процент из 1000 репликаций). Alveolata A выбраны в качестве корневых последовательностей.

Б. Данные, опубликованные Mladek et al. [2]. Анализируемая область последовательностей ограничивается Ring-IBR-Ring доменом (менее половины полных последовательностей). Цифры над ветвями и между ними отражают статистическую поддержку кластеров (процент из 1000 репликаций). Стрелками отмечены случаи статистически мало достоверной поддержки формирования кластеров (или её полного отсутствия). В качестве корневых использованы последовательности семейства Parkin из суперсемейства белков, содержащих Ring-IBR-Ring домен [1,3]. Нумерация последовательностей *Arabidopsis* соответствует использованной Mladek et al. [2].

В. Номенклатура последовательностей, выбранных для построения дерева А.

1. Анализируемый участок последовательностей. Mladek et al. не использовали всю доступную информацию (анализ был ограничен Ring-IBR-Ring доменом). Однако это малозначительно, поскольку приводит, по нашим данным, лишь к снижению статистической поддержки кластеров при сохранении основной топологии эволюционного дерева неизменной.

2. Коллекция последовательностей для анализа. В этой коллекции Mladek et al. собрали все 19 вариантов последовательностей *Arabidopsis* и, как можно заключить из топологии эволюционного древа, случайно выбранные последовательности других видов. Так, при полном отсутствии статистической поддержки кластер грибных белков оказывается между двумя группами белков животных. Вероятно, многие из последовательностей коллекции существенно отклоняются от основного эволюционного ствола. Это относится и к последовательностям *Arabidopsis*, многообразие которых, по нашим данным, можно свести лишь к двум последовательностям, представляющим типы А и В (выделены жирным шрифтом на рис. 1Б). Остальные либо мало отличаются от этих типов, либо, отражая специфику *Arabidopsis*, не характерны для других видов растений и, следовательно, отклоняются от основного эволюционного ствола А или ветви В.

3. Выбор в качестве корней эволюционного древа последовательностей, не принадлежащих семейству *Ariadne*, принципиально неверен (особенно если число корневых последовательностей невелико). Об этом свидетельствует полное отсутствие статистической поддержки родства двух главных кластеров: кластера, специфичного для растений, и кластера, объединяющего растения, животные и грибы.

Сформулированная в настоящей работе стратегия эволюционного анализа последовательностей белков суммирует накопленный к настоящему времени положительный опыт. Формализованное описание этой стратегии целесообразно, поскольку отклонения от нее встречаются достаточно часто и иногда приводят не только к неопределенным и малодостоверным результатам, но и к принципиально неверным выводам. Так, результаты анализа одного и того же семейства белков могут привести к прямо противоположным выводам о принадлежности *Gnetales* древним покрытосеменным [6] или древним голосеменным [7] (по совокупности имеющейся информации верен последний вывод [8]).

Самостоятельный интерес представляют результаты впервые корректно проведенного исследования эволюции семейства *Ariadne*, показавшие формирование двух типов растительных белков, различия первичных структур которых могут отражать их функциональную специфичность.

Литература:

1. Marin I., Ferrús A. Comparative genomics of the RBR family, including the Parkinson's Disease-related gene *Parkin* and the genes of the *Ariadne* subfamily // *Mol. Biol. Evol.*, 2002, vol.19, p.2039-2050.
2. Mladek C., Guger K., Hauser M.T. Identification and characterization of the *ARIADNE* gene family in *Arabidopsis*. A group of putative E3 ligases // *Plant Physiol.*, 2003, vol.131, p.27-40.
3. Eisenhaber B., Chumak N., Eisenhaber F., Hauser M.T. The ring between ring fingers (RBR) protein family // *Genome Biol.*, 2007, vol.8. Article 209, p.1-10.
4. Van de Peer Y., De Wachter R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment // *Comp. Appl. Biosci.*, 1994, vol.10, p.569-570.
5. Embley T.M., Martin W. Eukaryotic evolution, changes and challenges // *Nature*, 2006, vol.440, p.623-630.
6. Häger K.P., Müller B., Wind C., Erbach S., Fischer H. Evolution of legumin genes: loss of ancestral intron at the beginning of angiosperm diversification // *FEBS Lett.*, 1996, vol.387, p.94-98.
7. Shutov A.D., Braun H., Chesnokov Yu.V., Horstmann C., Kakhovskaya I.A., Bäumlein H. Sequence peculiarity of gnetalean legumin-like seed storage proteins // *J. Mol. Evol.*, 1998, vol.47, p.486-492.
8. Winter K.U., Becker A., Münster T., Kim J.T., Saedler H., Theissen G. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants // *Proc. Natl. Acad. Sci. USA*, 1999, vol.96, p.7342-7347.

Prezentat la 05.10.2010