ՀԱՅԱՍՏԱՆԻ ԿԵՆՍԱԲԱՆԱԿԱՆ ՀԱՆԴԵՍ
БИОЛОГИЧЕСКИЙ ЖУРНАЛ АРМЕНИИ
XXXIII, 9, 955—959, 1980

УДК 577.158

АЛГОРИТМ ПОСТРОЕНИЯ ЭВОЛЮЦИОННЫХ ДРЕВ ФЕРМЕНТОВ НА ПРИМЕРЕ ЛАКТАТДЕГИДРОГЕНАЗЫ

И. М. ЗАРАФЯН, Г. П. КАЗАНЧЯН, К. С. ДАНИЕЛЯН, Ж. И. АКОПЯН

Разработан усовершенствованный метод построения филогенетических древ с учетом первичных структур белков на всех этапах построения на примере пяти структур лактатдегидрогенезы. Алгоритм дает возможность находить участки конвергентных замен в процессе эволюции, а также учитывает замены функционально близких аминокислот.

Ключевые слова: ферменты, эволюционные древа, лактатдегидрогеназа.

В данной работе мы попытались разработать подход к изучению эволюции мономеров лактатдегидрогеназы (ЛДГ), одного из ключевых ферментов метаболизма в клетке. Структура молекул ЛДГ претерпевает небольшие изменения в процессе эволюции, особенно в области активного центра. Естественно, что в такой системе, где число допустимых мутаций сравнительно невелико, роль конвергентной эволюции в топологии филогенетического древа существенна. Нахождение участка структур белков, в которых имеют место конвергентные замены, является трудной задачей, поскольку исследования проводятся только с их современными формами. Если построение дендрограмм условно разбить на три этапа—построение матрицы расстояний между современными формами белков, построение древа по данной матрице расстояний, выявление вероятностной предковой структуры по топологии построенного древа—то можно заметить, что известные алгоритмы [1, 6, 7] обычно на втором этапе отвлекаются от первичных структур белков. Мы же рассматриваем такой подход к построению древа, при котором на всех этапах алгоритма учитывается первичная структура белков, в нашем случае—пяти структур ЛДГ. Подобные исследования в совокупности с изучением вопроса о филогенетической близости различных семейств белков помогут пролить свет на молекулярные основы эволюции в целом.

Исходя из построенной матрицы расстояний (рис. 1) на каждом шаге объединения структур белков проводится корректировка матрицы с учетом возможности параллельных или обратных замен. Выбирается минимальный элемент, соответствующий паре белков с наиболее сходными структурами, объединяющейся далее вершиной, соответствующей их предковой структуре. Пусть в i-той позиции объединяемой пары белков находятся аминокислоты A и B соответственно. Предполагая,

что в их предковой последовательности с равной вероятностью i-тую позицию может занимать и та и другая аминокислота, для каждого случая последовательно рассмотрим i-тую позицию всех оставшихся бел-

	1	2	3	4	5
1	—	44	109	106	125
2		—	101	105	118
3			—	53	96
4				—	108
5					—

Рис. 1.

	1	2	3	4	5
1	—	44	114	115	133
2		—	110	111	129
3			—	53	96
4				—	108
5					—

Рис. 2.

Рис. 1. Матрица расстояний между пятью структурами Н-и М-мономеров ЛДГ. Расстояние между структурами белковых субъединиц представляет сумму минимальных мутационных расстояний между аминокислотами, находящимися в гомологичных позициях белков. 1—Н-субъединица ЛДГ цыпленка, 2—Н-субъединица ЛДГ свиньи, 3—М-субъединица ЛДГ свиньи, 4 — М-субъединица ЛДГ цыпленка, 5 — М-субъединица ЛДГ акулы

Рис. 2. Скорректированная матрица расстояний. Разность между элементами первой и второй строк постоянна и равна 4.

ков. Если эту позицию занимает аминокислота A, тогда реальное расстояние аминокислоты B до аминокислоты X_k каждого k-того белка из оставшихся будет не $R_i(B, X_k)$, учтенное в матрице, а $R_i(B, A) + R_i(A, X_k)$. В данном случае число мутационных событий для i-той позиции, не учтенное в матрице расстояний, равняется

$$\sum_{k=1}^{n} [R_i(A, B) + R_i(A, X_k) - R_i(B, X_k)].$$

Если i-тую позицию занимает аминокислота B, то выражение приобретает следующий вид:

$$\sum_{k=1}^{n} [R_i(A, B) + R_i(B, X_k) - R_i(A, X_k)].$$

Исходя из принципа минимизации числа эволюционных событий [3], предковой последовательности приписывается та аминокислота, для которой вычисленная сумма меньше, а к элементам соответствующей строки матрицы прибавляется число неучтенных эволюционных событий. Проделав аналогичную процедуру для всех гомологичных позиций, в которых различается объединяемая пара белков, получаем скорректи-

рованную матрицу для рассматриваемых структур ЛДГ (рис. 2). Разность между элементами первой и второй строк матрицы, соответствующих объединяемым вершинам, постоянна для всех столбцов, относящихся к оставшимся белкам, что является признаком аддитивности поддрева. Далее вычисляем расстояние структур белков до найденной предковой структуры, после чего получаем матрицу, размерность которой на единицу меньше (рис. 3). Длины ребер a_1 и a_2 от объединяемых вершин до их предковой вершины можно легко вычислить из системы уравнений:

$$\begin{cases} a_1 + a_2 = m \\ a_1 - a_2 = n, \end{cases}$$

где m—расстояние между объединяемыми вершинами, n—постоянная разность между элементами строк этих вершин. Из рис. 2 видно, что на первом шаге объединения $m=44$, а $n=4$, откуда $a_1=24$, а $a_2=20$.

Эти последовательные объединения порождают дендрограмму, приведенную на рис. 4. Следует отметить, что на всех шагах построения сохраняется признак аддитивности, т. е. построенное древо является адди-

	(1,2)	3	4	5
(1,2)	–	90	91	109
3		–	53	96
4			–	108
5				–

Рис. 3.

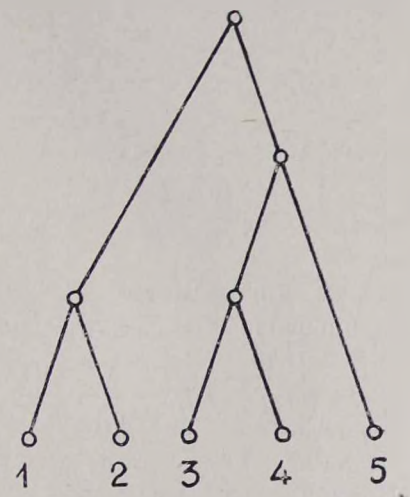

Рис. 4.

Рис. 3. Матрица расстояний после первого шага алгоритма с объединенными вершинами (1, 2).

Рис. 4. Дендрограмма пяти структур Н- и М-мономеров ЛДГ. 1—Н-субъединица ЛДГ цыпленка, 2—Н-субъединица ЛДГ свиньи, 3—М-субъединица ЛДГ свиньи, 4—М-субъединица ЛДГ цыпленка, 5—М-субъединица ЛДГ акулы.

тивным. Сравнивая скорректированную матрицу с исходной, можно выявить участки конвергентных замен: так, если в первоначальной матрице структура М-мономера ЛДГ акулы была ближе к структуре М-мономера ЛДГ свиньи, то после корректировки она стала ближе к структуре М-мономера ЛДГ цыпленка, что более соответствует общим представлениям об эволюции видов.

Ранее нами было показано, что, когда в матрице расстояний имеется множество элементов с одинаковыми или очень близкими значениями, большое влияние на топологию древа оказывает выбор минимального элемента [2]. Изначально мы выбирали те элементы, которые отражают замены функционально близких аминокислот, поскольку именно такие замены способствуют сохранению структур высшего порядка в процессе эволюции [5]. Следует отметить, что особое значение этот подход приобретает при работе с семантической частью молекулы, где доля замен по сравнению с полной структурой молекулы значительно меньше. Так, если в семантической части молекулы процент замен равен 26, то по всей структуре—45. При этом, если по всей длине молекулы число функциональных и нефункциональных замен приблизительно одинаково, то в семантической части ее превалируют замены функциональ-

Таблица

Число аминокислотных замен в структурах Н- и М-мономеров ЛДГ

В пределах объединяемых групп	Н- и М-мономеры по всей структуре	Н- и М-мономеры в семантической части молекул	Н-группа субъединиц	М-группа субъединиц
Общее число замен	135	21	35	96
Из них функционально близких аминокислот	75	15	23	59
Число аминокислотных остатков рассматриваемой структуры	333	77	333	333

но близких аминокислот (столбец 2 в табл.). При рассмотрении нагрузки функциональных замен отдельно в группах Н- и М-субъединиц ЛДГ можно заметить, что в каждой группе процент функциональных замен равен 66 и 61 соответственно.

Таким образом, предложенный алгоритм позволяет строить филогенетическое древо с учетом первичных структур на всех этапах построения. Принцип корректировки матрицы расстояний на каждом шаге алгоритма дает возможность одновременно с построением древа вычислять длины ребер и выявлять вероятностную предковую структуру всех вершин древа. Нахождение возможных участков конвергентных замен также существенно для реального отображения эволюции рассматриваемых белков.

Институт экспериментальной биологии
АН АрмССР

Поступило 28.III 1980 г.

ՖԵՐՄԵՆՏՆԵՐԻ ԷՎՈԼՅՈՒՑԻՈՆ ԾԱՌԵՐԻ ԿԱՌՈՒՑՄԱՆ ԱԼԳՈՐԻԹՄԸ ԼԱԿՏԱՏՈՂԵՀԻԴՐՈԳԵՆԱԶԻ ՕՐԻՆԱԿԻ ՎՐԱ

Ի. Մ. ՁՈՐԱՖՅԱՆ, Գ. Պ. ՂԱԶԱՆՉՅԱՆ, Կ. Ս. ԴԱՆԻԵԼՅԱՆ, Ժ. Ի. ՀԱԿՈԲՅԱՆ

Մշակված է ֆիլոգենետիկական ծառի կառուցման կատարելագործված մեթոդ՝ հաշվի առնելով կառուցման բոլոր էտապներում սպիտակուցների

առաջնային կառուցվածքները՝ լակտատդեհիդրոգենազի 5 կառուցվածքների օրինակի վրա։

Ալգոռիթմը հնարավորություն է տալիս գտնելու էվոլյուցիայի պրոցեսում կոնվերգենտային փոխարինումները, ինչպես նաև՝ հաշվի է առնում ֆունկցիոնալ մոտ կանգնած ամինաթթուների փոխարինումները։

AN ALGORYTHM OF CONSTRUCTION OF ENZYME EVOLUTIONARY TREES ON THE MODEL OF LACTATEDEHYDROGENASE

I. M. ZARAPHIAN, G. P. KAZANCHIAN, K. S. DANIELIAN, Zh. I. AKOPIAN

An improved method of construction of philogenetic tree with regard for protein primary structures on all the steps of construction on the model of 5 LDH structures has been worked out.

Algorythm allows to find section of convergent substitutions in the evolution process and takes into account substitutions of functionally related aminoacids.

ЛИТЕРАТУРА

1. *Жарких А. А.* Математические модели эволюции и селекции, Новосибирск, 1977.
2. *Зарафян И. М., Даниелян К. С., Акопян Ж. И.* Биолог. ж. Армении, *32*, 12, 1979.
3. *Camin J. H., Sokal R. R.* Evolution, *19*, 3, 1973.
4. *Cavalli-Sforza L. B., Edwards A. W. F.* Evolution, *21*, 3, 1967.
5. *Chirpich T. P.* Science, *188*, 4192, 1975.
6. *Dayhoff M. O.* Atlas of Protein Seguence and Structure, Natl. Biomed. Res. Found., Washington, 1972.
7. *Fitch W. M., Farris J. S.* J. Mol. Biol., *84*, *3*, 1974.
8. *Ycas M. H.* J. Moles. Evol., *7*, 3, 1976.