

Об одном алгоритме поиска горизонтального переноса гена на основе филогенетических деревьев белков

В.В.Вьюгин, В.А.Любецкий

*Институт проблем передачи информации РАН
101447, Москва, ГСП-4, Б.Каретный пер., 19, Россия
e-mail: lyubetsk@iitp.ru*

Поступила в редколлегию 21.08.2001

Аннотация—Целью работы является разработка методов для получения информации о событиях, происходивших в процессе эволюции в геномах организмов на молекулярном уровне, на основе филогенетических данных. В данной работе мы предлагаем некоторый метод для выявления генов, которые возможно попали в геном организма в результате горизонтального переноса, на основе вызываемой ими рассогласованности между деревьями эволюции генов и видов. Горизонтальный перенос генов определяется как перемещение генов между организмами без участия механизмов репродукции. Данный метод применяется к данным о 26 группах микроорганизмах и 132 КОГах – комплексах ортологических групп генов, объединенных по функциональному признаку из базы данных GenBank Национального центра биотехнологической информации США.

1. БИЛОГИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ

Хорошо известно, что филогенетические деревья, построенные по различным семействам белков, часто не совпадают друг с другом, а также с известными в эволюционной биологии деревьями видов соответствующих организмов. Причиной этого являются как неточности в построении деревьев генов, так и тот принципиальный факт, что деревья генов могут отличаться от дерева видов из-за событий, происходивших на молекулярном уровне в процессе эволюции и не связанных с дивергенцией видов. К числу таких событий обычно относят дубликации и потери генов, а также горизонтальные переносы генов.

Нашей целью является получение информации о таких событиях на основе филогенетических данных. В данной работе мы предлагаем некоторый метод определения потенциально горизонтально перенесенных генов на основе вызываемой ими рассогласованности между деревьями генов и видов.

Горизонтальный перенос генов определяется как перемещение генов между организмами без участия механизмов репродукции. Не вызывает сомнения наличие горизонтального переноса между бактериями и другими микробами. Бактерия может получить в свою хромосому ген извне следующими путями [4, 2, 5]: 1) непосредственно из окружающей среды; 2) путем вирусной инфекции бактериофагами; 3) от другой бактерии посредством плазмид или “прыгающих” генов. Значение горизонтального переноса для медицины подчеркивается тем, что большинство плазмид являются переносчиками генов устойчивых к антибиотикам.

Мы предлагаем метод выявления генов – кандидатов на горизонтальный перенос. Компьютерная программа лишь составляет списки генов, послуживших причиной значительной рассогласованности между деревьями эволюции генов и видов. Дальнейший отбор должен производиться экспертным путем на основе анализа их функций и сходства с генами организма.

Постановка задачи компьютерного анализа горизонтального переноса генов на основе филогенетических данных рассматривалась в [6, 7]. Однако в этих работах не предлагались какие-либо методы ее решения.

Данный метод применяется к данным, включающим 26 групп микроорганизмов и 132 КОГа – комплексов ортологических групп генов, объединенных по функциональному признаку, полученным из базы данных GenBank Национального центра биотехнологической информации (НЦБИ) США.

2. ТАКСОНОМИЧЕСКИЕ ДЕРЕВЬЯ И ИХ ГОМОМОРФИЗМЫ

Задача классификации исходит из некоторого набора исходных операционных таксономических единиц I . Элементы I не будут наделены какой-либо структурой, поэтому для простоты множество I будет иметь вид $I = \{1, 2, \dots, N\}$. По множеству I можно построить таксономическое дерево T_I (т.е. связный граф без циклов) следующим образом. Дерево T_I имеет N листьев, каждый из которых помечен одноэлементным подмножеством из числа $\{1\}, \{2\}, \dots, \{N\}$. Внутренние вершины дерева T_I помечены подмножествами I , образованными из элементов I по следующему правилу. Если непосредственные потомки некоторой вершины дерева помечены подмножествами A и B , то сама вершина помечена их объединением $A \cup B$. Таким образом, корень дерева T_I будет помечен множеством I . Вершина дерева и множество его помечающее отождествляются. Таким образом, мы рассматриваем только бинарные деревья. Подробнее см. в [8].

Из такого определения таксономического дерева легко следует, что для любых двух таких множеств, либо одно из них является подмножеством другого (если соответствующая ему вершина дерева является потомком другой вершины), либо имеют пустое пересечение (в противоположном случае). Таким образом, каждое таксономическое дерево T_I состоит из кластеров, составленных из элементов множества I . Для любой внутренней вершины дерева или его корня g , через cg обозначается его левый непосредственный потомок, а через $\hat{c}g$ – его правый непосредственный потомок. Если g не корень, то pg обозначает непосредственного предка g .

В качестве множеств таксономических единиц I будут рассматриваться множества двух типов – множества генов (выделенных из молекулярных последовательностей ДНК; таким генам однозначно соответствуют последовательности аминокислот – белки) и множества видов (названия современных организмов или их групп).

Пусть G – некоторое конечное множество, элементы которого будут называться генами (или белками). Мы не будем рассматривать какую-либо структуру на генах, поэтому для простоты считаем, что $G = \{1, 2, \dots, N\}$. Мы будем рассматривать системы попарно непересекающихся подмножеств G_1, G_2, \dots множества генов (белков), которые будут называться семействами генов (или белковыми семействами). Обычно объединение в семейства происходит по некоторой выделенной функции, которую выполняет в организме данное белковое семейство. В данной работе в качестве таких семейств будут рассматриваться КОГи – кластеры ортологических групп из базы данных GenBank НЦБИ (<http://www.ncbi.nlm.nih.gov/COG/>).

Каждое множество генов G_i , как множество таксономических единиц, может порождать различные таксономические деревья T_{G_i} . Методы построения таких деревьев на основе молекулярных последовательностей ДНК (или соответствующих белковых последовательностей) рассматриваются в [8]. Как правило, в практических приложениях такие деревья являются бинарными и имеют корень. Такие деревья будут называться деревьями генов.

Пусть S – конечное множество видов. Каждый вид характеризуется некоторым набором генов. Поэтому мы будем считать, что элементы S – виды, являются непустыми подмножествами множества генов G , причем для любых двух различных видов $s, s' \in S$ должно выполняться условие $s \cap s' = \emptyset$. Мы будем также считать, что каждый элемент множества G содержится в одном из видов $s \in S$ (хотя в практических приложениях это условие может нарушаться). Мы будем строить из множества таксономических единиц S различные деревья T_S . Такие деревья будут называться деревьями видов.

Определим операцию отображения α произвольного дерева генов T'_G , порожденного множеством генов $G' \subseteq G$ в произвольное дерево видов T_S , порожденное множеством видов S . Свойства тако-

го отображения будут служить основой для расчета численных характеристик рассогласованности деревьев генов и деревьев видов.

Пусть даны два бинарных дерева – дерево видов T_S и дерево генов T_G , где $G' \subseteq G$. По этим деревьям однозначно строится отображение

$$\alpha : T_G \rightarrow T_S$$

следующим образом: для каждого $g \in T_G$ (по определению $g \subseteq G$) значение $\alpha(g)$ определяется как минимальное по теоретико-множественному включению $s \in T_S$ такое, что $g \subseteq s$. Легко видеть, что такое отображение является *гомоморфизмом* деревьев, т.е. если $g \subseteq g'$, то $\alpha(g) \subseteq \alpha(g')$.

Если бы дерево генов T_G и дерево видов T_S имели одинаковую структуру, то отображение $\alpha(g)$ было бы изоморфизмом, т.е. любая вершина s из T_S была бы образом некоторой вершины g из T_G : $s = \alpha(g)$; и $g \subseteq g'$ было бы равносильно $\alpha(g) \subseteq \alpha(g')$ для всех вершин g, g' дерева генов.

Рассмотрим основные характеристики, которые задают отличие гомоморфизма деревьев $\alpha(g)$ от изоморфизма деревьев. Это *дубликации* в области определения, т.е. вершины g и g' , такие, что g' является непосредственным потомком g и $\alpha(g) = \alpha(g')$, а также *пропуски* в области значений, т.е. наличие вершин s в дереве видов T_S , таких что $\alpha(g) \subset s \subset \alpha(pg)$ (вершина s расположена строго между вершинами $\alpha(g)$ и $\alpha(pg)$). Пара (g, s) , где $s = \alpha(g)$ называется *односторонней дубликацией*, если $\alpha(g) = \alpha(cg)$ или $\alpha(g) = \alpha(\hat{c}g)$, но не одновременно. Если выполнены два этих условия одновременно, то (g, s) называется *двухсторонней дубликацией*. Множество односторонних дубликаций обозначается $O(T_G, T_S)$.

Вершина $s \in T_S$ называется *g-промежуточной*, если она находится строго между $\alpha(g)$ и $\alpha(pg)$. Пусть I_g – множество всех *g-промежуточных* вершин. Общее множество промежуточных вершин определяется

$$M(T_S, T_G) = \cup_{g \in T_G} I_g.$$

В [3] вводится мера различия дерева генов T_G и дерева видов T_S – *функция стоимости различия*

$$c(T_G, T_S) = |M(T_G, T_S)| + |O(T_G, T_S)|,$$

которая одновременно является количественной характеристикой степени отличия гомоморфизма $\alpha(g)$ от изоморфизма. Здесь $|A|$ обозначает число элементов конечного множества A .

Некоторые алгоритмы построения деревьев по генетическим последовательностям вычисляют также численные характеристики, выражающие расстояние от одной последовательности до другой. В некоторых случаях они могут интерпретироваться как время в предположении, что скорости эволюции рассматриваемых генов одинаковы.

Пусть заданы длины $c(a, b)$ ребер дерева генов. Тогда можно ввести функцию стоимости различия отличия деревьев T_G и T_S

$$L(T_G, T_S) = \sum_{(g, \alpha(g)) \in O(T_G, T_S)} c(g, pg) + \sum_{(g, \alpha(g)) \in M(T_G, T_S)} c(g, pg) |I_g|,$$

в которой первый член характеризует потери от дубликаций, а второй – потери от пропущенных вершин. Причем веса представляют собой преобразованный показатель сходства последовательностей, приписанных вершинам этого ребра, или бутстрэп поддержки соответствующего кластера (части дерева под этим ребром). В данной статье мы будем пользоваться второй из этих характеристик.

В работе [1] предложен некоторый алгоритм построения консенсусного дерева видов, использующий приведенную выше функцию стоимости различия деревьев, и проведено сравнительное исследование эффективности этого алгоритма. Производилось сравнение результатов работы данного алгоритма и еще двух известных компьютерных методов построения дерева видов. На примере генов позвоночных и митохондриальных геномов показано, что этот алгоритм дает приемлемые результаты.

3. МЕТОДИКА РАСЧЕТОВ

Целью расчетов в данной работе является выделение множества генов, которые возможно были перенесены в геномы различных организмов извне путем горизонтального переноса.

Построение дерева видов является первым этапом цикла расчетов. Пусть заданы деревья генов $T_{G_1}, T_{G_2}, \dots, T_{G_n}$. Задача построения дерева видов решается на основе подбора дерева T_S , для которого величина

$$c(T_S) = c(T_{G_1}, T_S) + c(T_{G_2}, T_S) + \dots + c(T_{G_n}, T_S) \quad (1)$$

достигает минимума. По-видимому, в общем случае решение этой задачи требует экспоненциального времени. Программа TIQMAX используем эвристический алгоритм, поиска локального минимума величины (1). Начальные деревья видов T_S^0 можно генерировать с помощью датчика случайных чисел. Алгоритм заключается в последовательной локальной перестройке текущего дерева видов T_S в окрестности каждой вершины с целью поиска дерева T_S , дающего локальный минимум величине $c(T_S)$. Подробнее алгоритм описан в [1].

Основная идея, лежащая в основе предлагаемого метода определения горизонтального переноса основывается на том, что наличие гена, перенесенного извне в геном организма на достаточно позднем этапе эволюции, нарушает согласованность между деревом генов (КОГом), содержащим этот ген и деревом видов по следующим причинам. Допустим, что ген g , находящийся в организме A , был перенесен в процессе эволюции в другой организм B , находящийся в дереве видов на достаточно большом расстоянии от организма A ¹. Тогда в дереве генов ген g будет близок к группе генов организмов, окружающих организм A . В то же время группа генов из геномов организмов, окружающих организм B будет находится на достаточно большом расстоянии от группы генов из геномов организмов, окружающих организм A . Таким образом, при гоморфизме дерева генов в дерево видов возникает серия мнимых дубликаций достаточно большого суммарного веса, возникшая из-за необходимости отобразить ген g в организм B . Эта серия дубликаций увеличит стоимость согласования дерева генов и дерева видов. Данная серия дубликация пропадет при простом удалении гена g из дерева видов и мы ожидаем, что стоимость согласования заметно уменьшится. Порог этого заметного уменьшения можно выбирать экспертным путем.

Методика расчетов основана на указанном выше простом соображении. Для каждого дерева генов (КОГа) T_G мы по очереди удаляем гены $g \in G$ из листьев дерева генов T_G , пусть при этом получается дерево генов $T_{G-\{g\}}$, (вместе с этими листьями) и рассматриваем изменение стоимости согласования деревьев

$$e(g) = L(T_{G-\{g\}}, T_S) - L(T_G, T_S).$$

Рассматривается среднее значение величины $e(g)$

$$E(e) = \frac{1}{N} \sum_{g \in G} e(g),$$

где N – число элементов G , и среднее квадратичное отклонение этой величины от среднего

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{g \in G} (e(g) - E(e))^2}.$$

Гены g , для которых величина $e(g) - E(e) \leq -t\sigma$, относим к числу “подозрительных” генов – кандидатов на горизонтальный перенос. Здесь t – величина, определяющая порог отклонения от среднего. В случае гипотезы о нормальном распределении отклонений стоимости от среднего заведомо подходит порог $t = 3$. Практические расчеты показали, что можно выбирать в качестве порога значительно

¹ Здесь под расстоянием между вершинами графа понимается длина самого короткого пути на графе, соединяющего эти вершины.

большие значения величины t (см. Приложение 2). Видимо разумно выбирать порог по результатам экспертного анализа отобранных генов.

4. РЕЗУЛЬТАТЫ РАСЧЕТОВ

Экспериментальные расчеты проводились на основании данных из базы GenBank (версии июня 2001 г.), предоставленных авторам настоящей работы сотрудниками НЦБИ.

В Приложении 1 приведен список из 26 групп микроорганизмов, для которых проводились расчеты. Группы обозначены буквами латинского алфавита.

Молекулярные данные по данным организмам организованы в GenBank в виде кластеров ортологических групп КОГов (COG). Кластер определяется группой генов (соответствующих им белков), имеющей общее происхождение (ортологическое сходство) и ответственных за некоторую функцию. Каждый кластер содержит выравнивания молекулярной и белковой последовательности, по которым можно используя различные методы построить варианты деревьев происхождения белков данного кластера. На данный момент (август 2001г.) в базе данных имеется 3166 таких кластеров.

Настоящие расчеты проводились на основе 132 КОГов, выбранных специалистами НЦБИ для анализа наличия горизонтального переноса составляющих их генов. Их список и нумерация приведены в Приложении 1. КОГи были использованы для построения филогенетическим деревьев соответствующих белковых семейств.

Были использованы два набора исходных данных. Первый набор данных – деревья генов, предоставленные специалистами НЦБИ. Эти деревья генов (в количестве 132) содержат длины ветвей, отражающие время эволюции составляющих белков. Вторая группа деревьев (также в количестве 132) была построена специалистами Института проблем передачи информации РАН на основе готового выравнивания COG-ов с помощью пакета PHYLIP (с использованием метода бутстрепов и метода максимальной экономии: программы SEQBOOT, PROTPARSE, CONSENSE). Данный метод в качестве результата выдает только топологические характеристики построенных деревьев. С одной стороны, он не учитывает время эволюции, и поэтому выдает менее информативные деревья, с другой стороны, такой метод гарантирован от ошибок, возникающих вследствие различной скорости эволюции различных генов.

Кластеры в деревьях снабжены весом, представляющим собой процент бутстрепа, который выражает степень доверия к соответствующему ребру (кластеру).

Построение различных вариантов деревьев видов 26 организмов осуществлялось на основе минимизации стоимости различия деревьев генов и деревьев видов с помощью программы TIQMAX. Принципы, положенные в основу работы этой программы см. в [1]).

В каждом из двух вариантов расчета функция стоимости определялась двумя различными способами. При первом способе учитывались длины ребер, отражающие эволюционное время. При втором из них учитываются степени доверия к кластерам дерева генов (т.е. только учитывается топология деревьев генов).

При каждом из способов 5000 раз задавалось случайное начальное дерево видов, которое перестраивалось до дерева, на котором достигался локальный минимум функции стоимости различия. В целом, имеет место большая неустойчивость результатов при варьировании начального дерева видов – получено очень много типов деревьев генов даже с близким к наименьшему значению локального минимума функции стоимости различия.

Для каждого способа расчетов по первым 250 деревьям с наилучшими значениями функции стоимости различия было построено консенсусное дерево видов. Как было видно по результату, высокую степень доверия имеют только кластеры, близкие к корню и листьям. Низкая степень доверия к промежуточным кластерам отражает хаотическое разнообразие деревьев КОГов в их центральной части.

Первая петерка деревьев видов в RNYLIP-формате, соответствующая первому варианту расчетов приведена в Приложении 2. Первое из этих деревьев изображено на рисунке. Это дерево в результате экспертного анализа было отобрано как “наилучшее” дерево видов и использовалось в дальнейших расчетах.

Определение горизонтального переноса в обоих случаях производилось на отобранном дереве видов.

Изменения функции стоимости различия для части КОГов приведены для первого способа расчетов в Приложении 2. Гены в верхней части каждого списка, которым соответствует наибольшее по модулю отрицательное значение (уменьшение значения функции стоимости различия) – наиболее вероятные кандидаты на горизонтальный перенос.

5. ВЫВОДЫ

Результаты определения явления горизонтального переноса продемонстрировали их заметную устойчивость при изменении дерева видов. Гены, дающие наибольшее уменьшение стоимости вложения, попадают в группу “подозрительных” генов для достаточного высоких значений порога t при обоих использованных способах построения дерева видов и для их вариантов, полученных внутри каждого способа. Этот факт частично можно объяснить высокой степенью доверия к кластерам дерева видов, близким к корню и двухэлементным кластерам из листьев. При построении гомоморфизма деревьев горизонтальный перенос проявляется в трудности отображения одного из элементов двухэлементного кластера генов, связанной с тем, что их образы в дереве видов достаточно сильно удалены и путь между образами в дереве видов должен проходить через вершины, близкие к корню. А именно эти вершины у корня и у листьев имеют наибольшую поддержку в дереве видов.

ПРИЛОЖЕНИЕ 1: Исходные данные

Список организмов

- A Archaeoglobus fulgidus
- O Halobacterium sp. NRC-1
- M Methanococcus jannaschii + Methanobacterium thermoautotrophicum
- 4P Thermoplasma acidophilum + Thermoplasma volcanium
- K Pyrococcus horikoshii + Pyrococcus abyssi
- Z Aeropyrum pernix
- Y Saccharomyces cerevisiae + Candida albicans
- Q Aquifex aeolicus
- V Thermotoga maritima
- D Deinococcus radiodurans
- R Mycobacterium tuberculosis + Mycobacterium leprae
- L Lactococcus lactis + Streptococcus pyogenes
- B Bacillus subtilis + Bacillus halodurans
- C Synechocystis
- E Escherichia coli + Escherichia coli + Buchnera sp.
- F Pseudomonas aeruginosa
- G Vibrio cholerae
- H Haemophilus influenzae + Pasteurella multocida
- S Xylella fastidiosa
- N Neisseria meningitidis + Neisseria meningitidis

U Helicobacter pylori + Helicobacter pylori + Campylobacter

J Mesorhizobium loti + Caulobacter crescentus37372628

X Rickettsia prowazekii

I Chlamydia trachomatis + Chlamydia pneumoniae

T Treponema pallidum + Borrelia burgdorferi

W Ureaplasma urealyticum + Mycoplasma pneumoniae + Mycoplasma genitalium

Список КОГов: 1. COG0016 2. COG0018 3. COG0020 4. COG0048 5. COG0049 6. COG0051
 7. COG0052 8. COG0060 9. COG0061 10. COG0064 11. COG0072 12. COG0080 13. COG0081 14. COG0082
 15. COG0085 16. COG0087 17. COG0088 18. COG0090 19. COG0091 20. COG0092 21. COG0093
 22. COG0094 23. COG0096 24. COG0097 25. COG0098 26. COG0099 27. COG0100 28. COG0101
 29. COG0102 30. COG0103 31. COG0104 32. COG0105 33. COG0126 34. COG0127 35. COG0128
 36. COG0130 37. COG0134 38. COG0135 39. COG0143 40. COG0148 41. COG0149 42. COG0151
 43. COG0152 44. COG0159 45. COG0162 46. COG0164 47. COG0166 48. COG0167 49. COG0169
 50. COG0171 51. COG0172 52. COG0173 53. COG0178 54. COG0180 55. COG0190 56. COG0193
 57. COG0197 58. COG0198 59. COG0200 60. COG0201 61. COG0202 62. COG0203 63. COG0215
 64. COG0216 65. COG0221 66. COG0222 67. COG0223 68. COG0231 69. COG0233 70. COG0237
 71. COG0242 72. COG0244 73. COG0250 74. COG0256 75. COG0258 76. COG0261 77. COG0264
 78. COG0272 79. COG0275 80. COG0284 81. COG0290 82. COG0292 83. COG0294 84. COG0305
 85. COG0313 86. COG0319 87. COG0335 88. COG0336 89. COG0340 90. COG0343 91. COG0351
 92. COG0359 93. COG0441 94. COG0442 95. COG0452 96. COG0461 97. COG0462 98. COG0481
 99. COG0495 100. COG0504 101. COG0519 102. COG0522 103. COG0525 104. COG0528 105. COG0532
 106. COG0533 107. COG0536 108. COG0540 109. COG0541 110. COG0544 111. COG0547 112. COG0552
 113. COG0556 114. COG0571 115. COG0573 116. COG0576 117. COG0581 118. COG0587 119. COG0597
 120. COG0653 121. COG0682 122. COG0691 123. COG0706 124. COG0781 125. COG0858 126. COG1160
 127. COG1214 128. COG1466 129. COG1488 130. COG2812 131. COG0012 132. COG0013

ПРИЛОЖЕНИЕ 2: Результаты расчетов

Деревья видов 26 микроорганизмов. Первая пятёрка лучших деревьев видов при первом варианте расчета:

(((A,O),(K,M)),(P,Z)),(((B,L),W),(Q,V)),((I,U),T)),((C,(D,R)),(((E,(G, H)),F),S),N),(J,X))))),Y)260674;
 (((A,O),(K,M)),(P,Z)),(((B,L),W),(Q,V)),((I,U),T)),((C,(D,R)),(((E,(G, H,Y)),F),S),N),(J,X))))),262534;
 (((A,O),(K,M)),(P,Z)),(((B,L),W),U),((I,T),(Q,V))),((C,(D,R)),(((E,(G, H,Y)),F),S),N),(J,X))))),262688x3;
 (((A,O),(K,M)),(P,Z)),(((B,L),W),U),(Q,V)),(C,(I,T))),((D,R)),(((E,(G, H)),F),S),N),(J,X))))),Y)263262x2;
 (((A,P),O),(K,M),Z)),(((B,L),W),U),((I,T),(Q,V))),((C,(D,R)),(((E,(G, H,Y)),F),S),N),(J,X))))),263351;

Наиболее вероятные кандидаты на горизонтальный перенос и их КОГи

COG0272 (30 генов).

Цена вложения : 366.935.

Среднее приращение : -1.9901

Ср.-кв. отклонение : 1.75374

Приращения цены вложения по генам:

[E]yicF.....-49.5266

[T]TP0634.....-7.57115

[X]RP720.....-4.9911

[U]Cj0586.....-4.26419

[T]VB0552.....-4.02911

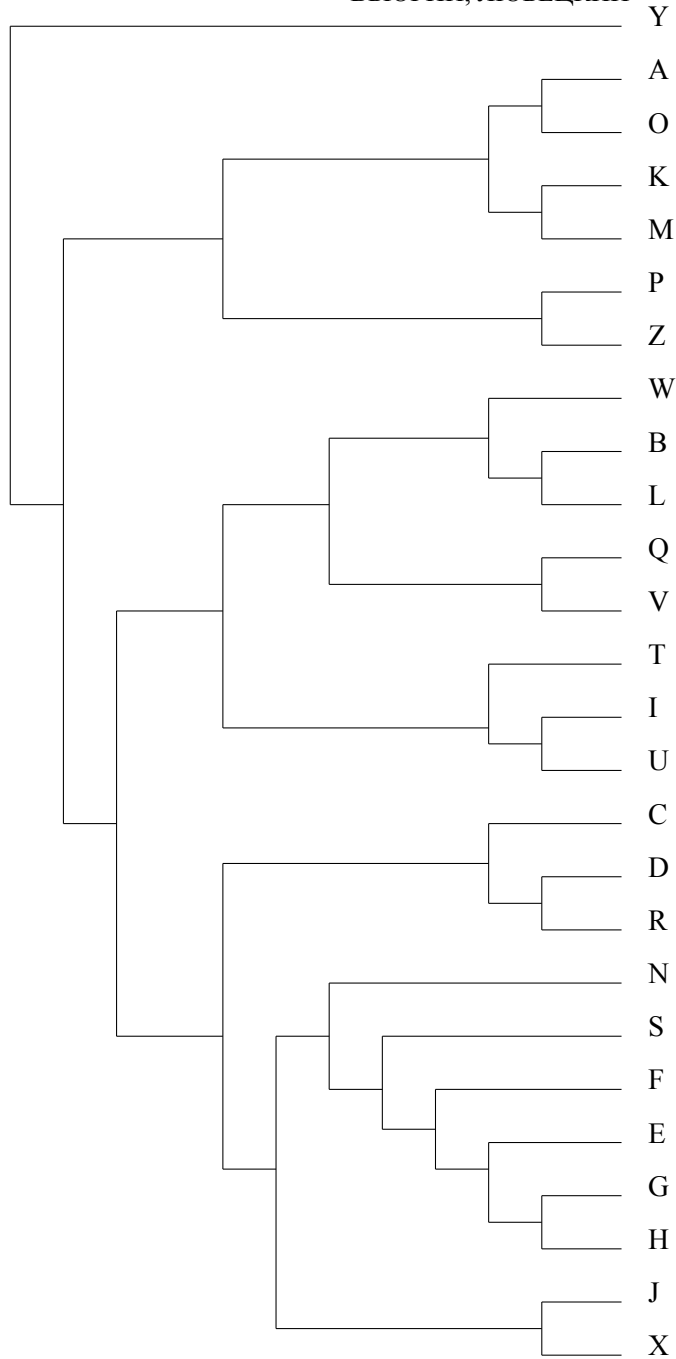


Рис. 1. Дерево эволюции 26 микроорганизмов.

[N]NMB0666.....-3.67466

[S]XF2556.....-3.66926

[U]HP0615.....-3.61347

[C]sl11209.....-3.1247

[G]VC0971.....-2.05617

COG0525 (38 генов).

Цена вложения : 349.351.

Среднее приращение : -1.14609

Ср.-кв. отклонение : 1.14766

Приращения цены вложения по генам:

[X]RP687.....-40.2158
[R]Rv2448с.....-7.14994
[S]XF0134.....-5.55816
[Z]APE1805.....-5.44512
[N]NMB0174.....-4.43061
[Q]aq_1413.....-4.34634
[G]VC2503.....-2.64857
[V]TM1817.....-1.84966
COG0215 (37 генов).

Цена вложения : 326.191.

Среднее приращение : -1.9194

Ср.-кв. отклонение : 0.9947

Приращения цены вложения по генам:

[O]VNG1097G.....-28.0041
[S]XF0995.....-11.9382
[D]DR1670.....-7.73889
[E]cysS.....-6.48868
[V]TM0719.....-5.67409
[C]slr0958.....-5.08781
[Q]aq_1068.....-5.01853
[A]AF0411.....-4.71781
[F]PA1795.....-4.14327
[Z]APE1592.....-4.08846
[N]NMB2083.....-3.95982
[J]CC0460.....-3.857
[X]RP085.....-3.56922
[E]BU487.....-2.88395
[G]VC1848.....-2.68002
[J]mlr7799.....-1.43198
COG0540 (28 генов).

Цена вложения : 285.642.

Среднее приращение : -1.38465

Ср.-кв. отклонение : 1.14333

Приращения цены вложения по генам:

[V]TM1642_1.....-27.5178
[E]BU369.....-6.29074
[M]MTH1413.....-5.01509
[G]VC2510.....-4.93429
[O]VNG6309G.....-4.20463
[M]MJ1581.....-4.10059
[A]AF0106.....-3.8236

[N]NMB0106.....-3.64498
 [Z]APE1663.....-2.81125
 [E]pугВ.....-2.19992
 COG0190 (32 генов).
 Цена вложения : 367.445.
 Среднее приращение : -1.12681
 Ср.-кв. отклонение : 1.12017
 Приращения цены вложения по генам:
 [O]VNG1416G.....-27.4863
 [D]DR0867.....-8.98289
 [Q]aq_1898.....-8.04166
 [T]BB0026.....-7.00766
 [R]Rv3356с.....-5.88866
 [U]HP0577.....-4.70884
 [U]Cj0855.....-4.25756
 [P]TVN1018.....-1.3147
 [P]Ta0898.....-0.916926
 [W]UU337.....-0.605669
 [X]RP515.....-0.403706
 [G]VC1942.....-0.163943
 [J]mll6921.....-0.0836043
 COG0143 (39 генов).
 Цена вложения : 565.467.
 Среднее приращение : -1.53533
 Ср.-кв. отклонение : 0.726225
 Приращения цены вложения по генам:
 [J]mlr5926.....-24.954
 [R]Rv1007с.....-7.16021
 [X]RP683.....-5.91713
 [J]mll0419.....-4.386
 [E]BU109.....-4.09327
 [M]MTH587_1.....-3.52636
 [D]DR1433_1.....-3.51835
 [N]NMB0030_1.....-3.06403
 [M]MJ1263_1.....-2.72504
 [V]TM1085_1.....-2.70175
 [S]XF0549_1.....-2.62673
 [J]CC1480.....-2.10214
 [C]slr0649.....-1.84702
 [I]CT032.....-1.52292
 [Z]APE1129.....-1.40776
 [I]CPn0122.....-1.35852
 [F]PA3482_1.....-0.99019

[G]VC1036_1.....-0.938764

[Q]aq_1257.....-0.417513

[E]metG_1.....-0.350418

СПИСОК ЛИТЕРАТУРЫ

1. Гельфанд М.С., Вьюгин В.В., Любецкий В.А. Об одном способе построения деревьев эволюции видов по множественным генетическим данным. *Информационные процессы*, 2001, том 1, № 1, стр. 64–77.
2. *Bacterial Conjugation*, Clewell D.B., Ed., New York: Plenum Press, 1993.
3. Eulenstein O, Vingron M. Arbeitspapiere der GMD. 1995, no. 936. Bonn. Germany.
4. Lorencz M.G., Wackernagel W. Bacterial Gene Transfer by Natural Genetic Transformation in the Environment. *Microbial Reviews*, 1994, 58, pp. 563–602.
5. Bergh O. *et al.* High Abundance of Viruses Found in Aquatic Environments, *Nature*, 1989, 340, pp. 467–468.
6. Page, R.D.M., Charlstone, M.A. From Gene to Organismal Phylogeny: Reconciled Trees and Gene Tree/Species Tree Problem, *Mol. Phylogenet. Evol.*, 1998, 7, pp. 231–240.
7. Page R.D.M. GeneTree: Comparing Gene and Species Phylogenies Using Reconciled Trees, *Bioinformatics Application Notes*, 1998, 14, pp. 819–820.
8. Waterman M.S. *Introduction to Computational Biology*, Chapman and Hall, 1995.

Статью представил к публикации член редколлегии В.А. Любецкий