

## О некоторых способах отбора горизонтально перенесенных генов на основе филогенетических данных

В.В.Вьюгин, В.А.Любецкий

*Институт проблем передачи информации РАН  
101447, Москва, Большой Каретный переулок, 19, Россия  
e-mail: lyubetsk@iitp.ru, тел. (095) 299-83-54 доб. 148, факс: (095) 209-0579*

Поступила в редколлегию 24.05.2002

**Аннотация**—В работе предлагается некоторые способы отбора генов, которые возможно попали в геном организма в результате горизонтального переноса генов. Отбор таких генов производится на основе вызываемой ими рассогласованности между деревьями эволюции генов и видов. Горизонтальный перенос генов определяется как перемещение генов между организмами без участия механизмов репродукции. Предлагаемая методика применяется к данным о 40 микроорганизмах и 132 КОГах — комплексов ортологических групп генов, объединенных по функциональному признаку из базы данных GenBank Национального центра биотехнологической информации США.

### 1. ВВЕДЕНИЕ

### 2. БИОЛОГИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ

Известно, что филогенетические деревья, построенные по различным семействам белков, часто не совпадают друг с другом, а также с известными в эволюционной биологии деревьями эволюции видов соответствующих организмов. Причиной этого являются как неточности в построении деревьев эволюции 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. Подробнее о деревьях эволюции генов и видов и событиях дубликации и потери генов см. в [9, 10].

Нашей целью является разработка методов получения информации о таких событиях на основе филогенетических данных. В этой работе мы предлагаем некоторые подходы для отбора 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 [13, 11, 14]: 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. Некоторые авторы [1, 3, 4, 5] считают, что горизонтальный перенос генов является основной силой эволюции микроорганизмов. Геном *Escherichia coli* может содержать до 18 процентов горизонтально перенесенных генов [6]. В геноме *Thermotoga maritima* 25 процента генов ближе к генам семейства Archae, чем к бактериальным генам, имеется предположение, что они попали в этот геном в результате горизонтального переноса [7, 8].

Постановка задача компьютерного анализа горизонтального переноса генов на основе филогенетических данных рассматривалась в [15, 16]. Однако в этих работах не предлагались какие-либо методы ее решения. В данной работе предлагаются некоторые методы отбора генов — кандидатов на горизонтальный перенос. Компьютерная программа 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.

Методы применяются к данным, включающим 40 микроорганизмов, составляющих 13 групп и отдельных организмов: Archae (10 микроорганизмов), Gamma-proteobacteria (7), Gram-positive bacteria (8), Alpha-proteobacteria (3), Epsilon-proteobacteria (2), Chlamidia (2), Spirochetes (2), Beta-proteobacteria (1), Cyanobacteria (1), Deinococcus radiodurans (1), Aquifex aeolicus (1), Thermotoga maritima (1), Mycobacterium tuberculosis (1). Деревья эволюции белков (генов) были построены методом максимально правдоподобия на основе выравненных белковых последовательностей из 132 КОГов — комплексов ортологических групп генов, объединенных по функциональному признаку. Полная информация о КОГах содержится в базе данных GenBank Национального центра биотехнологической информации (НЦБИ) США

### 3. ТАКСОНОМИЧЕСКИЕ ДЕРЕВЬЯ И ИХ ГОМОМОРФИЗМЫ

В качестве кандидатов на горизонтальный перенос рассматриваются гены (или соответствующие им белки), которые являются основной причиной рассогласованности между деревьями эволюции генов (белков) и деревом происхождения микроорганизмов. В связи с этим рассмотрим некоторые количественные характеристики, на основе которых определяется степень рассогласованности между деревьями.

Задача классификации исходит из некоторого набора исходных операционных таксономических единиц  $I$ . Элементы  $I$  не 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 [17].

Из 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.

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 [17]. 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  $\alpha$  произвольного дерева генов  $T'_G$ , порожденного множеством генов  $G' \subseteq G$  в произвольное дерево видов  $T_S$ , порожденное множеством видов  $S$ . Свойства такого отображения будут служить основой для расчета численных характеристик рассогласованности деревьев генов и деревьев видов. Впервые этот метод сравнения деревьев рассматривался в работе [2].

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. выполнено свойство: 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. Количественная характеристика стоимости сравнения деревьев генов и видов приведена в Приложении 1.

## 4. МЕТОДИКА РАСЧЕТОВ

### 4.1. Методика селекции горизонтально перенесенных генов

В качестве исходных данных были использованы деревья происхождения белков (генов), построенные методом максимального правдоподобия на основе выравненных белковых последовательностей, принадлежащих семействам белков, образованных по функциональному принципу. В данной работе в качестве таких семейств рассматривались КОГи – комплексы ортологических групп из базы данных GenBank НЦБИ

При разработке метода определения горизонтального переноса мы исходим из того, что наличие гена, перенесенного извне в геном организма на достаточно позднем этапе эволюции, нарушает согласованность между деревом генов, содержащим этот ген и деревом видов. Эта рассогласованность возникает 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  $A$ <sup>1</sup>. 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  $A$ .

<sup>1</sup> Расстояние  $\rho(a, b)$  между двумя вершинами  $a$  и  $b$  дерева равно числу ребер на пути между ними.

Мы используем две различные характеристики для учета подобной рассогласованности. Первая характеристика основана на сравнении окрестности гена в дереве гена и ее образа при гомоморфизме  $\alpha$  в дереве видов. Пусть  $v$  – некоторый ген и  $v_1, v_2, \dots, v_n$  – все гены, расположенные на расстоянии не более  $r$  от  $v$ . Пусть также  $s = \alpha(v)$  и  $s_i = \alpha(v_i)$ , где  $i = 1, 2, \dots, n$  — все образы гена  $v$  и его окрестности в дереве видов. Вычисляем средние расстояния от гена  $v$  до генов из его окрестности

$$r(v) = \frac{1}{n} \sum_{i=1}^n \rho(v, v_i)$$

и среднее расстояние от соответствующего вида  $s$  до соответствующих видов

$$\hat{r}(s) = \frac{1}{n} \sum_{i=1}^n \rho(s, s_i).$$

Тогда величина

$$p(v) = \hat{r}(s)/r(v)$$

характеризует степень рассеянности образа окрестности гена  $v$  в дереве видов. Большие значения этой величины характеризуют степень патологии в положении гена  $v$  в виде  $s$  и служат критерием отбора гена  $v$  в качестве кандидата на горизонтальный перенос. Далее, производится экспертный анализ более детальной информации вида

$$(\rho(s, s_1)/\rho(v, v_1), \dots, \rho(v, v_n)/\rho(s, s_n)),$$

характеризующей рассеянность окрестности гена  $v$  в дереве видов.

Вторая характеристика основана на том, что в случае горизонтального переноса гена при гомоморфизме  $\alpha$  дерева генов в дерево видов возникает серия мнимых дубликаций (определение дубликации см. в Приложении 1 (Раздел 6)). 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. 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  $T_{G-\{g\}}$ , и рассматриваем изменение стоимости согласования деревьев

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

Рассматривается среднее значение величины  $v(g)$

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

где  $N$  – число листьев  $G$ , и несмещенная оценка среднего квадратичного отклонения

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

этой величины от среднего. Гены  $g$ , для которых выполнено

$$v(g) - E(G) \leq -t\sigma,$$

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 большие значения величины  $t$  (см. Приложение 3).

## 4.2. Оценка качества КОГов

Оценка качества деревьев генов производилась двумя способами.

Первый способ основан на сравнении величин цены вложения отдельных деревьев происхождения белков со средними величинами. Пусть  $G_1, \dots, G_m$  – все деревья происхождения генов,  $S$  – дерево видов,  $c(G_i, S)$  – стоимость различия  $i$ -го дерева генов и дерева видов. Определялась средняя стоимость различия  $c = \frac{1}{m} \sum_{i=1}^m c(G_i, S)$  и соответствующая оценка среднего квадратического отклонения  $\sigma$ . Оценка качества дерева белков  $G_i$  производилась на основе величин  $v_i = c(G_i, S) - c$  и  $p_i = v_i/\sigma$ .

При втором способе для каждого дерева белков  $G_i$  рассчитывалась средняя “рассеянность” окрестностей генов  $l_i = \frac{1}{n} \sum_{g \in G} p(g)$  (заданного радиуса), где сумма берется по всем генам  $g$  из дерева генов  $G_i$ . Большие отклонения величин  $l_i$  от среднего  $E = \frac{1}{m} \sum_{i=1}^m l_i$  характеризуют большую степень рассогласованности дерева генов и дерева видов.

## 5. РЕЗУЛЬТАТЫ РАСЧЕТОВ

Экспериментальные расчеты проводились на основании данных из базы GenBank (<http://www.ncbi.nlm.nih.gov/COG/>).

В Приложении 2 приведен список из 40 микроорганизмов (Раздел 7.1), для которых проводились расчеты. Каждый микроорганизм обозначен тремя буквами латинского алфавита (приведены также их полные латинские названия).

Молекулярные данные по данным организмам организованы в GenBank в виде комплексов ортологических групп КОГов. Каждый такой комплекс определяется группой генов (соответствующих им белков), имеющих общее происхождение (ортологическое сходство) и ответственных за некоторую функцию. Каждый комплекс содержит выравнивания молекулярной и белковой последовательности, по которым можно используя различные методы построить варианты деревьев происхождения генов (белков) данного комплекса.

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 3 (Раздел 8.1). Использовалось дерево эволюции 40 микроорганизмов, построенное пятым методом в работе [18]. Деревья эволюции белков построенные по указанным выше 132 КОГам, а также дерево эволюции видов, были любезно предоставлены авторами работы [18]. Графические изображения деревьев эволюции белков некоторых КОГов, а также дерево эволюции видов проведены в Приложении 2 (Раздел 7.2).

Эти деревья в количестве 132 имеют длины ветвей, отражающие время эволюции составляющих белков.

Определение горизонтального переноса производилось на дереве видов, построенном пятым методом в работе [18].

Результаты расчетов приведены в Приложении 3. В разделе 8.2 приведен общий список генов – вероятных кандидатов на горизонтальный перенос, представленных компьютерной программой при некоторых параметрах. Количественная и экспертная проверка показывают достаточно высокую корреляцию двух методов отбора генов – кандидатов на горизонтальный перенос.

В разделе 8.3 эти же гены сгруппированы по геномам организмов. Кандидаты на перенос были обнаружены в 22 организмах.

В разделе 8.4 они же сгруппированы по предполагаемым источникам горизонтального переноса. В качестве источника горизонтального переноса гена обычно рассматривается семейство (из 13 указанных выше), в которое попадают гены – соседи данного гена из его окрестности заданного радиуса (в случае данных расчетов радиус равен 4).

## 6. ПРИЛОЖЕНИЕ 1: СТОИМОСТЬ РАЗЛИЧИЯ ДЕРЕВЬЕВ

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 [12] 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 первой из этих характеристик.

В работах [9, 10] предложен некоторый алгоритм построения 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.

## 7. ПРИЛОЖЕНИЕ 2: ИСХОДНЫЕ ДАННЫЕ

## 7.1. Список организмов

Afu	Archaeoglobus fulgidus
Hbs	Halobacterium sp. NRC-1
Mja	Methanococcus jannaschii
Mth	Methanobacterium thermoautotrophicum
Tac	Thermoplasma acidophilum
Tvo	Thermoplasma volcanium
Pho	Pyrococcus horikoshii

Pab	Pyrococcus abyssi
Ape	Aeropyrum pernix
Sso	Saccharomyces cerevisiae
Aae	Aquifex aeolicus
Tma	Thermotoga maritima
Dra	Deinococcus radiodurans
Mtu	Mycobacterium tuberculosis
Mle	Mycobacterium leprae
Spy	Streptococcus pyogenes
Bsu	Bacillus subtilis
Bha	Bacillus halodurans
Syn	Synechocystis
Eco	Escherichia coli K12
Buc	Buchnera sp. APS
Pae	Pseudomonas aeruginosa
Vch	Vibrio cholerae
Hin	Haemophilus influenzae
Pmu	Pasteurella multocida
Xfa	Xylella fastidiosa
Nme	Neisseria meningitidis MC58
NmA	Neisseria meningitidis Z2491
Hpy	Helicobacter pylori 26695
Cje	Campylobacter jejuni
Mlo	Mesorhizobium loti
Ccr	Caulobacter crescentus
Rpr	Rickettsia prowazekii
Ctr	Chlamydia trachomatis
Cpn	Chlamydia pneumoniae
Tpa	Treponema pallidum
Bbu	Borrelia burgdorferi
Uur	Ureaplasma urealyticum
Mpn	Mycoplasma pneumoniae
Mge	Mycoplasma genitalium

### 7.2. Графические изображения деревьев

## 8. ПРИЛОЖЕНИЕ 3: РЕЗУЛЬТАТЫ РАСЧЕТОВ

### 8.1. Комплексы ортологических групп генов

Оценка качества когов: отмеченные \*\* не использовались из-за низкого качества (сравнительно большой величины переменных  $nc$  и  $E$ ).

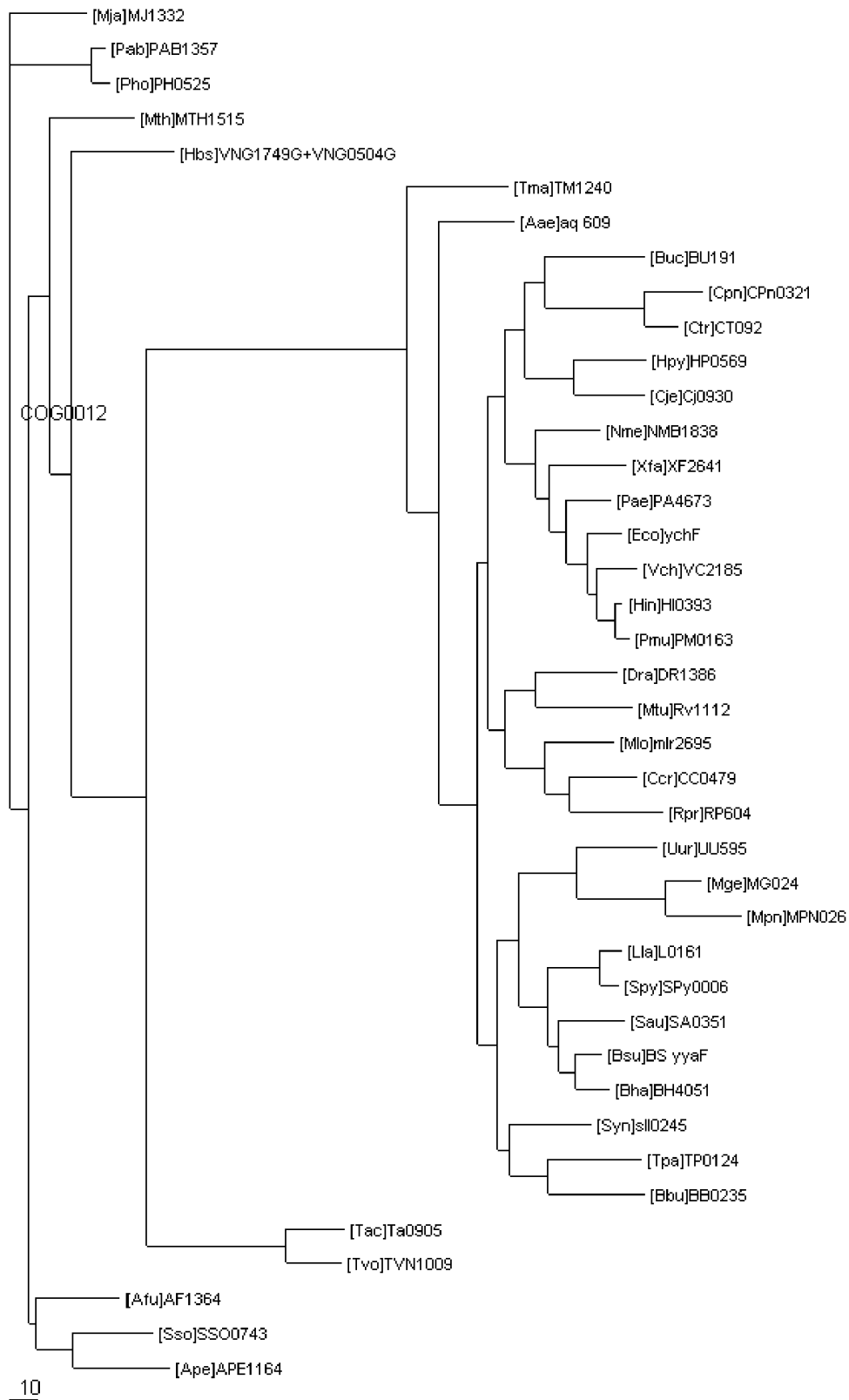
Статистика по ценам согласования с деревом видов.

Средняя нормированная цена согласования = 9.13381

Оценка дисперсии нормированной цены согласования = 3.9048

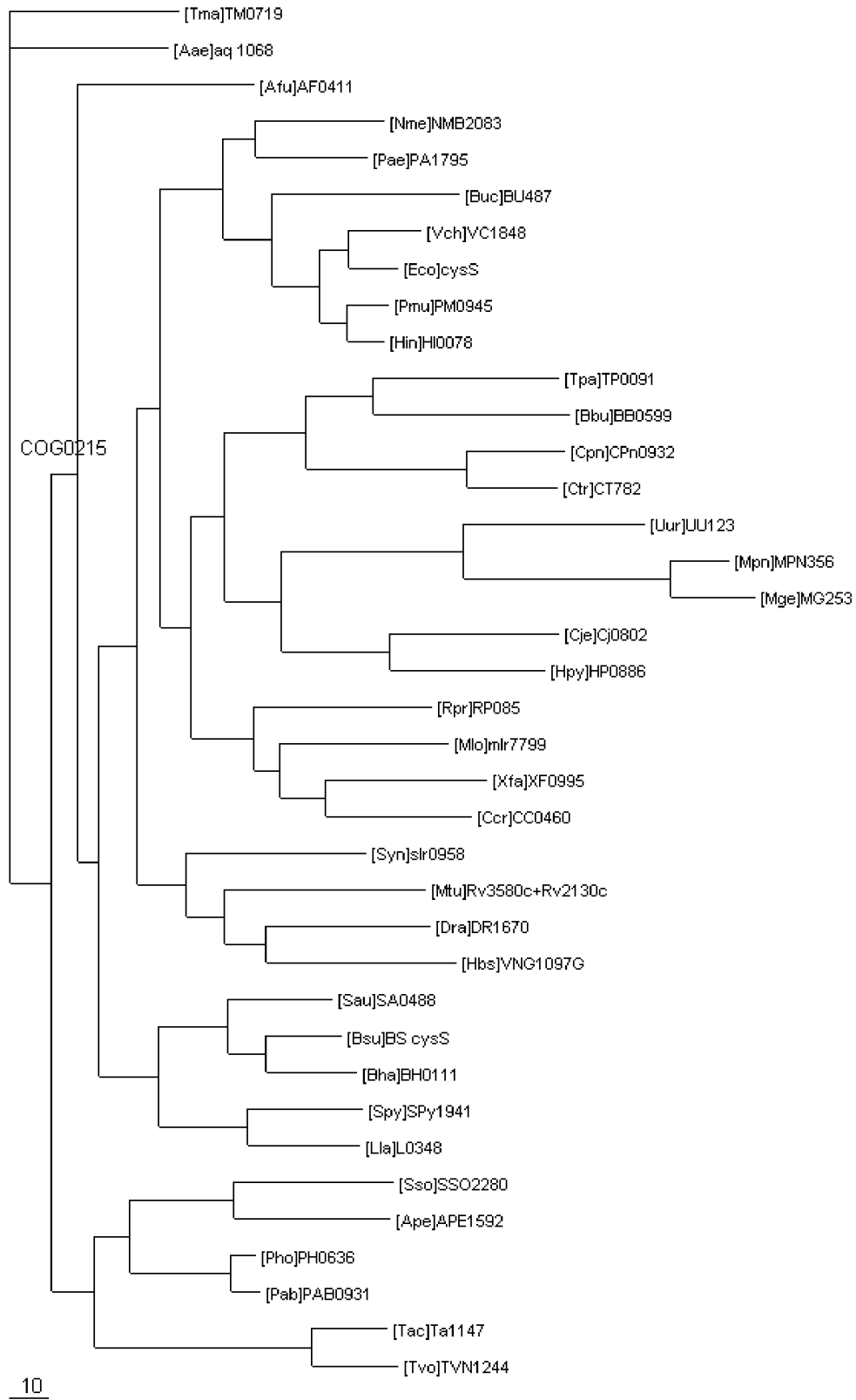
Информация о цене согласования деревьев КОГов в форме:

"#. 'name' ( $c/N$ )= $nc$ ,  $v$ ,  $p$  --  $E$ ",

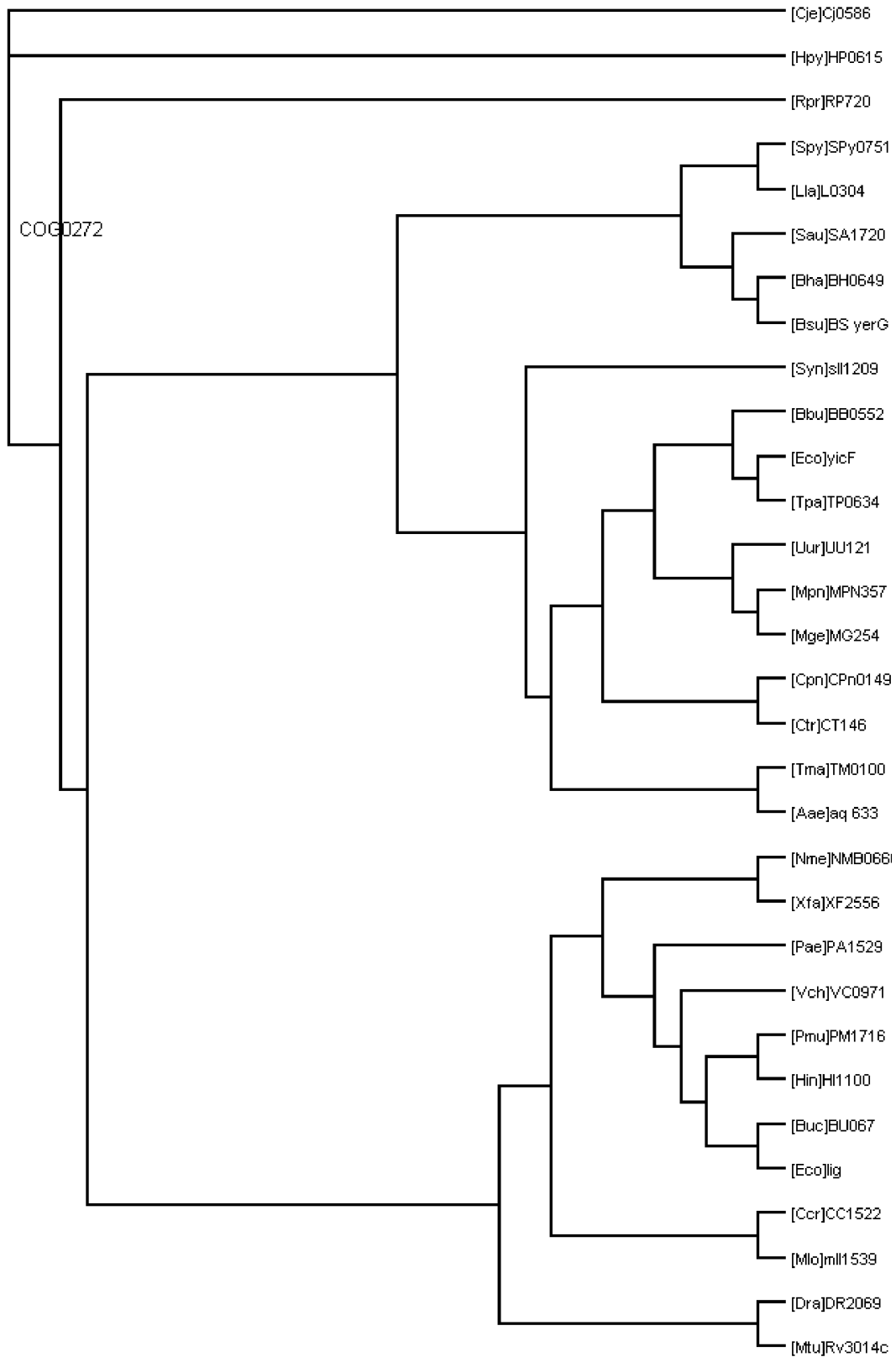


**Рис. 1.** COG0012, горизонтально перенесенные гены BU191, slI0245.

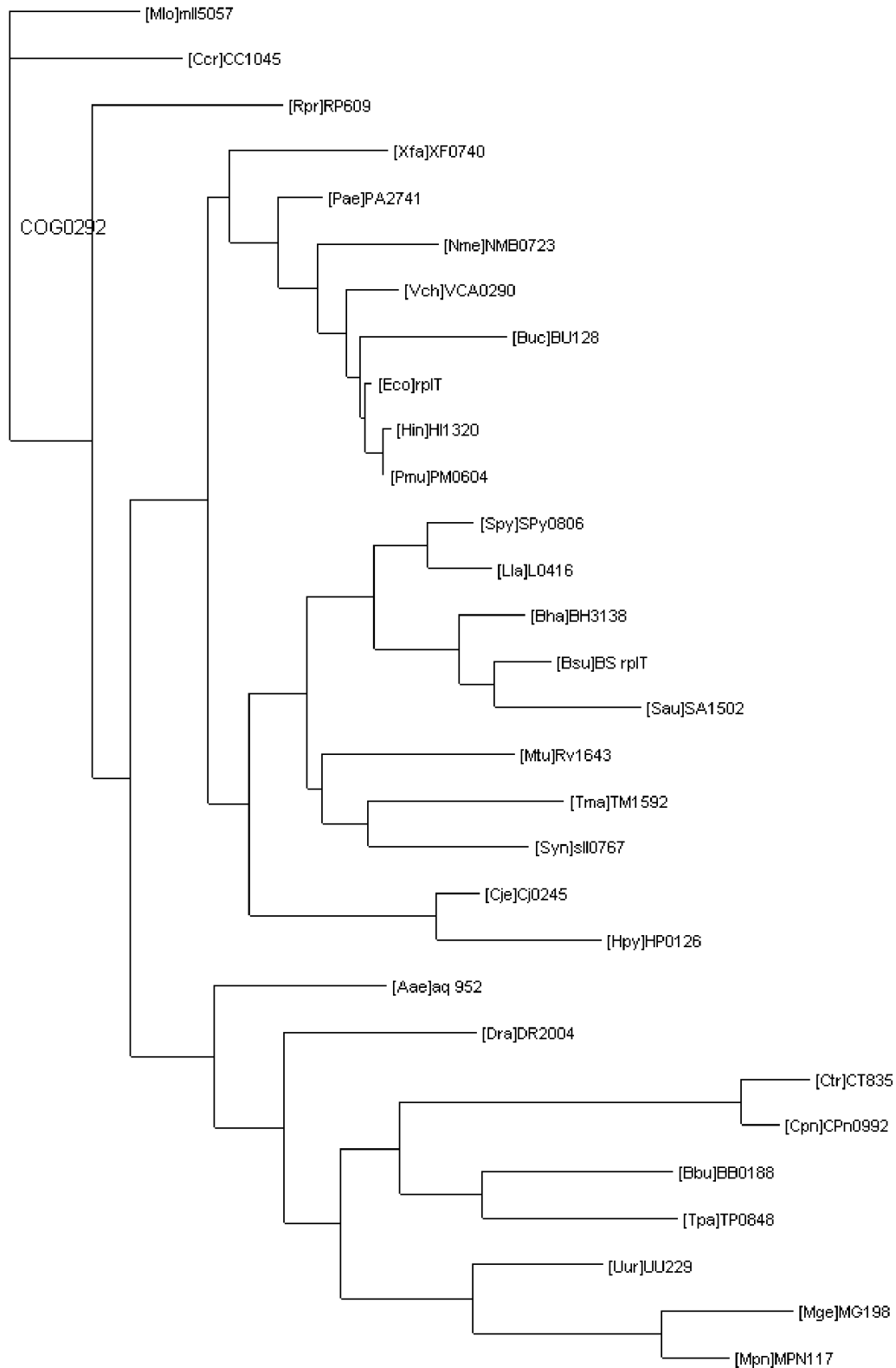




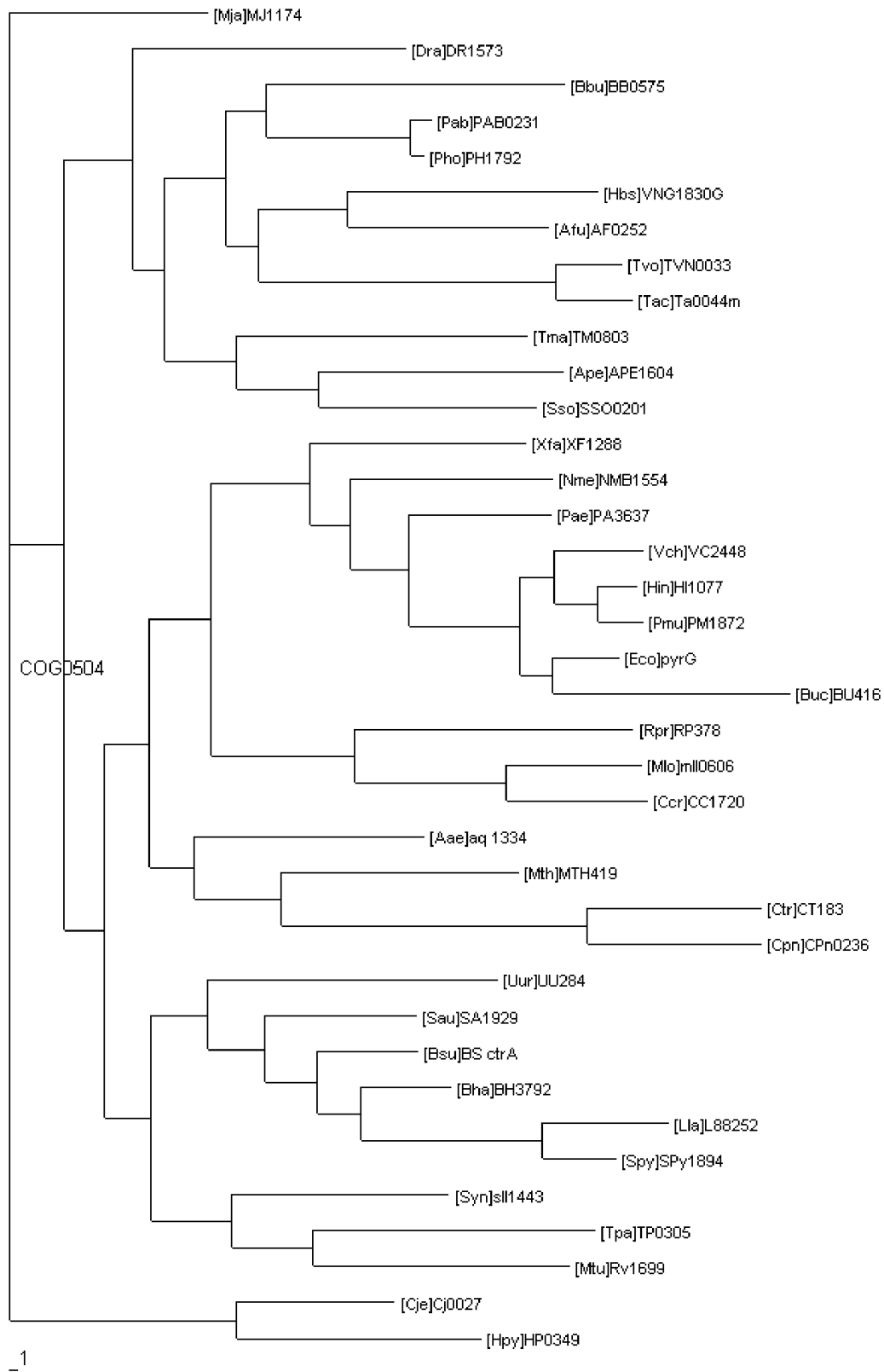
**Рис. 2.** COG0215, горизонтально перенесенный ген XF0995.



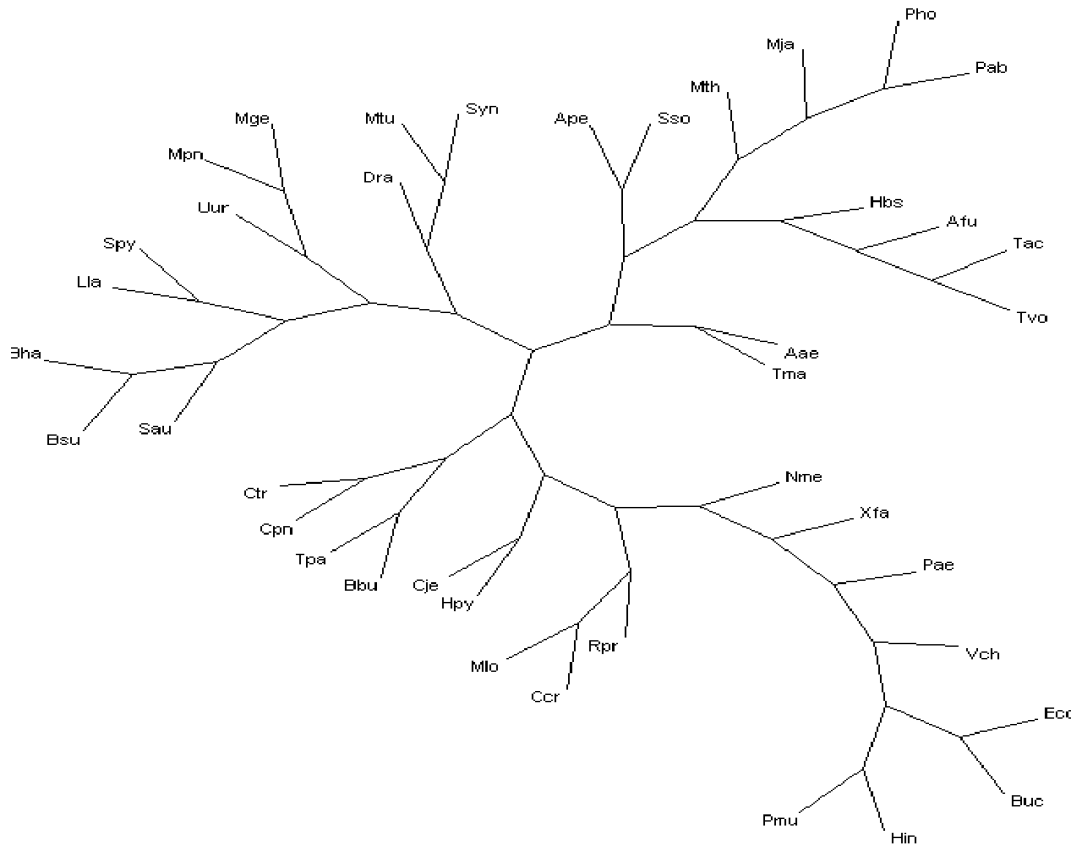
**Рис. 3.** COG0272, горизонтально перенесенный ген RP720.



**Рис. 4.** COG0292, горизонтально перенесенный ген TM1592.



**Рис. 5.** COG0504, горизонтально перенесенный ген MJ1174.



**Рис. 6.** Дерево эволюции 40 микроорганизмов.

где

- # – идентификатор КОГа (дерева генов);
- name – имя КОГа (дерева генов);
- c – стоимость согласования с деревом видов;
- N – число генов в дереве генов;
- $pc=c/N$  – нормированная цена согласования;
- v – отклонение стоимости согласования от среднего;
- p –  $v/\sigma$
- E – среднее по всему дереву генов величины "рассеянность" (см. Раздел 3.1).

1. "COG0012" (311/40)=8, -1.35854, -0.347915 -- 1.46719.
2. "COG0013" (300/40)=7, -1.64207, -0.420525 -- 1.18579.
3. "COG0016" (333/40)=8, -0.806059, -0.206428 -- 1.38127.
4. "COG0018" (526/41)=13, 3.69252, 0.945637 -- 1.7012.
5. "COG0020" (368/40)=9, 0.0595317, 0.0152458 -- 1.53958.
6. "COG0048" (232/39)=6, -3.17834, -0.813958 -- 1.24956.
7. "COG0049" (259/40)=6, -2.66142, -0.681577 -- 1.2879.
8. "COG0051" (168/40)=4, -4.94198, -1.26562 -- 1.24689.
9. "COG0052" (202/40)=5, -4.08177, -1.04532 -- 1.22479.
10. "COG0060" (321/40)=8, -1.11057, -0.284412 -- 1.50097.
11. "COG0061" (508/40)=13, 3.57054, 0.914399 -- 1.45005.
12. "COG0064" (150/30)=5, -4.13955, -1.06012 -- 1.2224.
13. "COG0072" (460/40)=12, 2.36646, 0.606038 -- 1.34853.
14. "COG0080" (204/41)=5, -4.16327, -1.0662 -- 1.26956.
15. "COG0081" (147/40)=4, -5.45614, -1.39729 -- 1.19688.
16. "COG0082" (255/33)=8, -1.39275, -0.356677 -- 1.4234.
17. "COG0085" (120/40)=3, -6.14457, -1.5736 -- 1.33321.
18. "COG0087" (214/40)=5, -3.78053, -0.968176 -- 1.32415.
19. "COG0088" (220/40)=5, -3.64578, -0.933668 -- 1.36418.
20. "COG0090" (224/40)=6, -3.52876, -0.903699 -- 1.14792.
21. "COG0091" (348/40)=9, -0.429935, -0.110104 -- 1.20602.
22. "COG0092" (285/40)=7, -2.00362, -0.513117 -- 1.2665.
23. "COG0093" (149/39)=4, -5.30619, -1.35889 -- 1.33681.
24. "COG0094" (211/40)=5, -3.85064, -0.98613 -- 1.43788.
25. "COG0096" (219/40)=5, -3.64765, -0.934145 -- 1.17856.
26. "COG0097" (260/40)=6, -2.63905, -0.675849 -- 1.42134.
27. "COG0098" (327/40)=8, -0.96671, -0.24757 -- 1.42774.
28. "COG0099" (326/40)=8, -0.992639, -0.25421 -- 1.34573.
29. "COG0100" (210/39)=5, -3.74175, -0.958244 -- 1.51314.
30. "COG0101" (335/38)=9, -0.307877, -0.078846 -- 1.09956.
31. "COG0102" (363/40)=9, -0.0712028, -0.0182347 -- 1.21586.
32. "COG0103" (323/40)=8, -1.06649, -0.273123 -- 1.28778.
33. "COG0104" (202/31)=7, -2.62778, -0.672962 -- 1.43339.
34. "COG0105" (240/33)=7, -1.86902, -0.478646 -- 1.88682.
35. "COG0126" (167/39)=4, -4.85992, -1.2446 -- 1.39857.
36. "COG0127" (334/35)=10, 0.422389, 0.108172 -- 1.39693.
- \*\*37. "COG0128" (426/34)=13, 3.39938, 0.870565 -- 1.74688.
38. "COG0130" (438/37)=12, 2.70002, 0.691462 -- 1.44981.
39. "COG0134" (343/30)=11, 2.3159, 0.593091 -- 1.64787.
- \*\*40. "COG0135" (594/30)=20, 10.6769, 2.73431 -- 2.10444.
41. "COG0143" (531/41)=13, 3.82501, 0.979566 -- 1.7076.
42. "COG0148" (257/43)=6, -3.15061, -0.806856 -- 1.26214.
43. "COG0149" (501/41)=12, 3.09425, 0.792423 -- 1.44359.
44. "COG0151" (267/30)=9, -0.228365, -0.0584832 -- 1.41347.
45. "COG0152" (407/32)=13, 3.59763, 0.921335 -- 1.6001.
- \*\*46. "COG0159" (496/31)=16, 6.86712, 1.75864 -- 1.9295.
47. "COG0162" (511/43)=12, 2.75306, 0.705045 -- 1.53438.
48. "COG0164" (372/35)=11, 1.50232, 0.384736 -- 1.48287.
49. "COG0166" (510/35)=15, 5.43163, 1.39102 -- 1.71523.
- \*\*50. "COG0167" (489/37)=13, 4.08709, 1.04669 -- 1.80796.
- \*\*51. "COG0169" (705/39)=18, 8.93655, 2.28861 -- 2.14233.
- \*\*52. "COG0171" (672/37)=18, 9.04131, 2.31544 -- 2.26388.
53. "COG0172" (497/40)=12, 3.29395, 0.843565 -- 1.43893.
54. "COG0173" (123/30)=4, -5.01757, -1.28498 -- 1.31364.

55. "COG0178" (241/33)=7, -1.8211, -0.466376 -- 1.39458.  
56. "COG0180" (392/43)=9, -0.0160326, -0.00410588 -- 1.57603.  
\*\*57. "COG0190" (367/33)=11, 1.99333, 0.510482 -- 1.82823.  
58. "COG0193" (277/30)=9, 0.096764, 0.0247808 -- 1.49834.  
59. "COG0197" (213/40)=5, -3.8101, -0.975747 -- 1.21521.  
60. "COG0198" (431/40)=11, 1.63897, 0.419733 -- 1.53017.  
61. "COG0200" (378/40)=9, 0.315736, 0.0808585 -- 1.49277.  
62. "COG0201" (233/41)=6, -3.44868, -0.88319 -- 1.31361.  
63. "COG0202" (277/40)=7, -2.21385, -0.566957 -- 1.15665.  
64. "COG0203" (169/30)=6, -3.48409, -0.89226 -- 1.46641.  
65. "COG0215" (321/38)=8, -0.686494, -0.175808 -- 1.32159.  
66. "COG0216" (122/30)=4, -5.05279, -1.294 -- 1.1754.  
67. "COG0221" (351/31)=11, 2.18464, 0.559477 -- 1.94924.  
68. "COG0222" (216/30)=7, -1.93298, -0.495028 -- 1.53163.  
69. "COG0223" (223/34)=7, -2.58752, -0.662653 -- 1.40143.  
70. "COG0231" (523/46)=11, 2.23336, 0.571953 -- 1.28317.  
71. "COG0233" (247/30)=8, -0.890607, -0.22808 -- 1.34584.  
72. "COG0237" (574/41)=14, 4.8713, 1.24752 -- 1.56396.  
73. "COG0242" (427/35)=12, 3.06082, 0.783861 -- 1.49107.  
74. "COG0244" (296/40)=7, -1.72555, -0.441905 -- 1.15625.  
75. "COG0250" (547/43)=13, 3.59167, 0.91981 -- 1.41034.  
76. "COG0256" (293/40)=7, -1.80929, -0.463352 -- 1.30903.  
77. "COG0258" (526/47)=11, 2.05253, 0.525643 -- 1.42662.  
78. "COG0261" (213/30)=7, -2.02578, -0.518792 -- 1.32033.  
79. "COG0264" (239/30)=8, -1.17062, -0.29979 -- 1.35147.  
80. "COG0272" (277/31)=9, -0.205165, -0.0525418 -- 1.37072.  
81. "COG0275" (343/30)=11, 2.31052, 0.591713 -- 1.3921.  
82. "COG0284" (328/32)=10, 1.11701, 0.286062 -- 1.58453.  
83. "COG0290" (136/30)=5, -4.61507, -1.1819 -- 1.25615.  
84. "COG0292" (152/30)=5, -4.0589, -1.03947 -- 1.21997.  
85. "COG0294" (632/35)=18, 8.93705, 2.28874 -- 1.74487.  
86. "COG0305" (323/31)=10, 1.29233, 0.33096 -- 1.4765.  
87. "COG0313" (165/31)=5, -3.81519, -0.977053 -- 1.25129.  
88. "COG0319" (300/30)=10, 0.873376, 0.223667 -- 1.51746.  
89. "COG0335" (218/30)=7, -1.87297, -0.479659 -- 1.55333.  
90. "COG0336" (167/30)=6, -3.57569, -0.915717 -- 1.10452.  
\*\*91. "COG0340" (676/34)=20, 10.7424, 2.75108 -- 1.90066.  
92. "COG0343" (237/36)=7, -2.55078, -0.653243 -- 1.36046.  
\*\*93. "COG0351" (605/34)=18, 8.6697, 2.22027 -- 2.56713.  
94. "COG0359" (353/30)=12, 2.62384, 0.671954 -- 1.30539.  
95. "COG0441" (330/43)=8, -1.44977, -0.371279 -- 1.59777.  
96. "COG0442" (413/40)=10, 1.2007, 0.307493 -- 1.47929.  
97. "COG0452" (344/32)=11, 1.60078, 0.409953 -- 1.53021.  
98. "COG0461" (422/34)=12, 3.28162, 0.840407 -- 1.49625.  
99. "COG0462" (237/40)=6, -3.19925, -0.819312 -- 1.36404.  
100. "COG0481" (70/30)=2, -6.8051, -1.74275 -- 1.03436.  
101. "COG0495" (438/40)=11, 1.82353, 0.466996 -- 1.47613.  
\*\*102. "COG0504" (207/38)=5, -3.67374, -0.940828 -- 1.61932.  
103. "COG0519" (155/33)=5, -4.45194, -1.14012 -- 1.44228.  
104. "COG0522" (322/40)=8, -1.0725, -0.274662 -- 1.47966.  
105. "COG0525" (273/40)=7, -2.30591, -0.590532 -- 1.41409.  
106. "COG0528" (261/40)=7, -2.5969, -0.665055 -- 1.36989.  
107. "COG0532" (227/40)=6, -3.45625, -0.885129 -- 1.29522.  
108. "COG0533" (291/40)=7, -1.85166, -0.474202 -- 1.48657.  
109. "COG0536" (128/30)=4, -4.86804, -1.24668 -- 1.26151.

110. "COG0540" (331/30)=11, 1.91051, 0.489272 -- 1.52867.  
 111. "COG0541" (261/40)=7, -2.62109, -0.671248 -- 1.17246.  
 112. "COG0544" (216/30)=7, -1.93036, -0.494357 -- 1.32803.  
 \*\*113. "COG0547" (553/35)=16, 6.66709, 1.70741 -- 2.18522.  
 114. "COG0552" (193/40)=5, -4.31018, -1.10382 -- 1.29987.  
 115. "COG0556" (144/31)=5, -4.50445, -1.15357 -- 1.45038.  
 116. "COG0571" (287/30)=10, 0.432556, 0.110776 -- 1.41852.  
 \*\*117. "COG0573" (504/32)=16, 6.62598, 1.69688 -- 2.10748.  
 118. "COG0576" (457/34)=13, 4.30563, 1.10265 -- 1.53474.  
 \*\*119. "COG0581" (523/33)=16, 6.72149, 1.72134 -- 2.02637.  
 120. "COG0587" (406/35)=12, 2.46825, 0.632107 -- 1.55767.  
 \*\*121. "COG0597" (458/30)=15, 6.1451, 1.57373 -- 1.92035.  
 122. "COG0653" (132/32)=4, -5.00405, -1.28151 -- 1.15667.  
 123. "COG0682" (326/30)=11, 1.749, 0.447911 -- 1.46369.  
 124. "COG0691" (263/30)=9, -0.375059, -0.0960508 -- 1.23979.  
 125. "COG0706" (417/34)=12, 3.12737, 0.800904 -- 1.33608.  
 126. "COG0781" (417/30)=14, 4.76395, 1.22002 -- 1.4122.  
 127. "COG0858" (595/30)=20, 10.6975, 2.73957 -- 1.71543.  
 128. "COG1160" (105/30)=3, -5.64189, -1.44486 -- 1.12695.  
 129. "COG1214" (384/30)=13, 3.65963, 0.937215 -- 1.43976.  
 130. "COG1466" (341/30)=11, 2.22011, 0.568559 -- 1.21476.  
 \*\*131. "COG1488" (343/34)=10, 0.955242, 0.244633 -- 1.88579.  
 132. "COG2812" (205/30)=7, -2.29842, -0.588616 -- 1.31892.

## 8.2. Наиболее вероятные кандидаты на горизонтальный перенос

Каждый ген снабжен строкой информации вида:

"ког [вид]ген.....цена% длина рассеянность (соседи) отклонение",

где

'ког' - идентификатор КОГа (и дерева генов G), к которому отнесен ген

'вид' - идентификатор организма-владельца данного гена;

'ген' - идентификатор гена g;

'цена' - изменение  $(v(g)/L(G, S)) * 100\%$  цены согласования деревьев при удалении данного гена g из дерева генов G (в процентах от начальной цены согласования  $L(G, S)$ );

'длина' - отношение длины ветви листа-гена к средней длине по всем листьям дерева G;

'рассеянность' - обратное отношение среднего расстояния до соседей (из окрестности радиуса 4) на дереве генов к аналогичному расстоянию на дереве видов;

'соседи' - список расстояний от данного гена g до всех соседей из его окрестности длины 4 в дереве генов и аналогичных расстояний между их образами в дереве видов (расстояния разделены знаком /);

'отклонение' - величина  $(v(g) - E(G)) / \sigma$  -- изменение цены в долях среднего квадратического отклонения.

COG0012	[Buc]BU191.....	-10.5%	1.61	3.14	(3/12 3/12 4/10 4/10)	-22.6799
COG0012	[Syn]sl10245.....	-3.7%	1.31	2.67	(3/8 3/8)	-8.01258
COG0018	[Syn]sl10502.....	-7.07%	1.07	2.64	(3/8 3/8 4/10 4/11)	-10.9758



COG0018	[Mtu]Rv1292.....-11.3%	1.38	2.9	(2/10 4/10 4/9)	-17.5096
COG0061	[Xfa]XF2090.....-9.29%	0.927	1.89	(2/6 3/6 4/5)	-17.1741
COG0072	[Mtu]Rv1650_2....-11.5%	1.52	2.16	(3/9 4/3 4/10 4/9 4/10)	-21.7698
COG0080	[Bha]BH0409.....-6.91%	1.85	2.17	(4/6 4/10 4/10)	-17.5127
COG0085	[Aae]aq_1939.....-15.9%	1.76	2.33	(3/7 3/7)	-23.5508
COG0102	[Dra]DR0174.....-12.3%	1.73	2.8	(2/9 4/8 4/11)	-27.6861
COG0105	[Bbu]BB0463.....-5.01%	1.64	2.72	(3/11 3/11 4/9 4/10 4/8)	-7.85545
COG0126	[Aae]aq_118.....-11.6%	1.34	2	(3/7 3/7:9 4/7 4/7)	-17.7019
COG0127	[Cje]Cj1374c....-4.33%	1.42	2	(3/6 3/6)	-6.31601
COG0143	[Mtu]Rv1007c....-8.95%	1.12	3.8	(2/10 3/9)	-14.6436
COG0143	[Mlo]mlr5926....-12.9%	1.53	4.6	(2/11 3/12)	-21.8269
COG0149	[Pmu]PM1640_1...-1.91%	1.44	3.4	(2/10 4/12 4/12)	-5.12847
COG0162	[Pae]PA4138.....-8.12%	1.47	2.22	(2/7 3/7 4/6)	-19.7497
COG0173	[Mtu]Rv2572c....-17.5%	1.5	2.37	(3/9:9 4/10 4/10 4/8:9 4/80)	-15.5617
COG0178	[Hbs]VNG2636G...-8.38%	1.47	2.6	(2/8 4/9 4/9)	-12.3595
COG0193	[Dra]DR2372.....-16.2%	1.26	2.67	(2/70 4/9)	-12.9523
COG0198	[Bbu]BB0489.....-15.8%	1.75	2.63	(3/7 4/9 4/10 4/12:0 4/12:0)	-26.4288
COG0200	[Bbu]BB0497.....-6.64%	1.55	2.36	(3/8 4/9 4/9)	-17.1705
COG0203	[Mtu]Rv3456c....-11.9%	1.53	2.67	(3/8 3/8)	-14.6785
COG0215	[Hbs]VNG1097G...-23%	1.53	2.89	(2/8 3/9 4/9)	-24.9558
COG0215	[Xfa]XF0995.....-13.2%	1.28	1.89	(2/6 3/6 4/5)	-13.848
COG0221	[Mlo]mlr8562.....-9.12%	1.67	3.29	(3/10 4/13:0)	-12.28
COG0222	[Dra]DR2043.....-10.3%	1.74	3.4	(2/9 3/8)	-12.0103
COG0242	[Syn]slr1549....-10.1%	1.47	3.2	(2/8 3/8)	-15.4158
COG0250	[Aae]aq_1931.....-9.29%	1.73	2.2	(3/7 3/7 4/8)	-22.9896
COG0272	[Rpr]RP720.....-8.8%	1.4	1.67	(3/5 3/5)	-6.37324
COG0272	[Eco]yicF.....-29.1%	1.64	4.8	(2/12 3/120)	-23.0953
COG0292	[Tma]TM1592.....-20.2%	1.76	2.8	(2/7 3/7)	-16.3817
COG0294	[Ccr]CC3224.....-9.04%	0.901	3.8	(2/10 3/9)	-11.3965
COG0335	[Dra]DR0755.....-6.19%	1.77	3.6	(2/9 3/9)	-9.14553
COG0343	[Afu]AF1485.....-8.58%	1.6	3.33	(3/10 3/10)	-13.0213
COG0359	[Aae]aq_2042.....-13.2%	1.39	2.8	(2/7 3/7)	-14.2312
COG0441	[Ape]APE0809....-14.1%	1.79	3	(2/8 3/7)	-21.0898
COG0452	[Bbu]BB0812.....-8.69%	1.35	2.33	(2/7 4/7)	-13.0443
COG0504	[Mja]MJ1174.....-4.41%	1.14	2.9	(3/10 3/10 4/9)	-4.04877
COG0504	[Mth]MTH419.....-7.4%	1.38	2.67	(3/6 3/9 3/9)	-7.54575
COG0504	[Mth]MTH419.....-7.4%	1.38	2.67	-2 (3/6 3/9 3/9)	-7.54575
COG0504	[Tpa]TP0305.....-21.5%	1.64	3.2	(2/8 3/8)	-24.053
COG0525	[Rpr]RP687.....-16.2%	1.62	3.67	(3/9 3/12 3/12)	-24.9209
COG0547	[Cje]Cj0346_2....-12.7%	1.28	3.6	(2/9 3/9)	-15.1408
COG0552	[Dra]DR2260.....-13.7%	1.52	2.27	(3/8 4/8 4/9 4/9)	

					-19.5807
COG0556	[Hbs]VNG2390G....-9.23%	1.53	2.33	(3/4 3/9 3/8)	
					-10.5939
COG0571	[Syn]slr0346....-19.7%	1.56	2.67	(3/82 3/8 3/8)	
					-13.0702
COG0587	[Bsu]BS_yorL....-10.3%	1.6	3.11	(3/8 3/10 3/10)	
					-14.5798

### 8.3. Горизонтальный перенос по геномам

#### [Syn]:

COG0012	[Syn]sl10245....-3.7%	1.31	2.67	(3/8 3/8)
COG0018	[Syn]sl10502....-7.07%	1.07	2.64	(3/8 3/8 4/10 4/11)
COG0242	[Syn]slr1549....-10.1%	1.47	3.2	(2/8 3/8)
COG0571	[Syn]slr0346....-19.7%	1.56	2.67	(3/82 3/8 3/8)

#### [Xfa]:

COG0061	[Xfa]XF2090.....-9.29%	0.927	1.89	(2/6 3/6 4/5)
COG0215	[Xfa]XF0995.....-13.2%	1.28	1.89	(2/6 3/6 4/5)

#### [Aae]:

COG0085	[Aae]aq_1939....-15.9%	1.76	2.33	(3/7 3/7)
COG0126	[Aae]aq_118....-11.6%	1.34	2	(3/7 3/7:9 4/7 4/7)
COG0250	[Aae]aq_1931....-9.29%	1.73	2.2	(3/7 3/7 4/8)
COG0359	[Aae]aq_2042....-13.2%	1.39	2.8	(2/7 3/7)

#### [Mtu]:

COG0018	[Mtu]Rv1292.....-11.3%	1.38	2.9	(2/10 4/10 4/9)
COG0072	[Mtu]Rv1650_2....-11.5%	1.52	2.16	(3/9 4/3 4/10 4/9 4/10)
COG0143	[Mtu]Rv1007c....-8.95%	1.12	3.8	(2/10 3/9)
COG0173	[Mtu]Rv2572c....-17.5%	1.5	2.37	(3/9:9 4/10 4/10 4/8:9 4/80)
COG0203	[Mtu]Rv3456c....-11.9%	1.53	2.67	(3/8 3/8)

#### [Bha]:

COG0080	[Bha]BH0409.....-6.91%	1.85	2.17	(4/6 4/10 4/10)
---------	------------------------	------	------	-----------------

#### [Dra]:

COG0102	[Dra]DR0174.....-12.3%	1.73	2.8	(2/9 4/8 4/11)
COG0193	[Dra]DR2372.....-16.2%	1.26	2.67	(2/70 4/9)
COG0222	[Dra]DR2043.....-10.3%	1.74	3.4	(2/9 3/8)
COG0335	[Dra]DR0755.....-6.19%	1.77	3.6	(2/9 3/9)
COG0552	[Dra]DR2260.....-13.7%	1.52	2.27	(3/8 4/8 4/9 4/9)

#### [Bbu]:

COG0105	[Bbu]BB0463.....-5.01%	1.64	2.72	(3/11 3/11 4/9 4/10 4/8)
COG0198	[Bbu]BB0489.....-15.8%	1.75	2.63	(3/7 4/9 4/10 4/12:0 4/12:0)
COG0200	[Bbu]BB0497.....-6.64%	1.55	2.36	(3/8 4/9 4/9)
COG0452	[Bbu]BB0812.....-8.69%	1.35	2.33	(2/7 4/7)

#### [Cje]:

COG0127	[Cje]Cj1374c....-4.33%	1.42	2	(3/6 3/6)
---------	------------------------	------	---	-----------

COG0547	[Cje]Cj0346_2....-12.7%	1.28	3.6	(2/9 3/9)
[Mlo]:				
COG0143	[Mlo]mlr5926....-12.9%	1.53	4.6	(2/11 3/12)
COG0221	[Mlo]mlr8562....-9.12%	1.67	3.29	(3/10 4/13:0)
[Pmu]:				
COG0149	[Pmu]PM1640_1....-1.91%	1.44	3.4	(2/10 4/12 4/12)
[Pae]:				
COG0162	[Pae]PA4138.....-8.12%	1.47	2.22	(2/7 3/7 4/6)
[Hbs]:				
COG0178	[Hbs]VNG2636G...-8.38%	1.47	2.6	(2/8 4/9 4/9)
COG0215	[Hbs]VNG1097G...-23%	1.53	2.89	(2/8 3/9 4/9)
COG0556	[Hbs]VNG2390G...-9.23%	1.53	2.33	(3/4 3/9 3/8)
[Rpr]:				
COG0272	[Rpr]RP720.....-8.8%	1.4	1.67	(3/5 3/5)
COG0525	[Rpr]RP687.....-16.2%	1.62	3.67	(3/9 3/12 3/12)
[Eco]:				
COG0272	[Eco]yicF.....-29.1%	1.64	4.8	(2/12 3/120)
[Tma]:				
COG0292	[Tma]TM1592.....-20.2%	1.76	2.8	(2/7 3/7)
[Ccr]:				
COG0294	[Ccr]CC3224.....-9.04%	0.901	3.8	(2/10 3/9)
[Afu]:				
COG0343	[Afu]AF1485.....-8.58%	1.6	3.33	(3/10 3/10)
[Ape]:				
COG0441	[Ape]APE0809.....-14.1%	1.79	3	(2/8 3/7)
[Mja]:				
COG0504	[Mja]MJ1174.....-4.41%	1.14	2.9	(3/10 3/10 4/9)
[Mth]:				
COG0504	[Mth]MTH419.....-7.4%	1.38	2.67	(3/6 3/9 3/9)
[Tpa]:				
COG0504	[Tpa]TP0305.....-21.5%	1.64	3.2	(2/8 3/8)
[Bsu]:				
COG0587	[Bsu]BS_yorL....-10.3%	1.6	3.11	(3/8 3/10 3/10)

#### 8.4. Предполагаемые источники горизонтального переноса

[Clamidia]:				
COG0012	[Buc]BU191.....-10.5%	1.61	3.14	(3/12 3/12 4/10 4/10)
COG0018	[Syn]sl10502....-7.07%	1.07	2.64	(3/8 3/8 4/10 4/11)
COG0335	[Dra]DR0755.....-6.19%	1.77	3.6	(2/9 3/9)

COG0343	[Afu]AF1485.....	-8.58%	1.6	3.33	(3/10 3/10)
COG0504	[Mth]MTH419.....	-7.4%	1.38	2.67	(3/6 3/9 3/9)
COG0504	[Mth]MTH419.....	-7.4%	1.38	2.67	(3/6 3/9 3/9)
COG0504	[Mth]MTH419.....	-7.4%	1.38	2.67	-2 (3/6 3/9 3/9)
COG0571	[Syn]slr0346.....	-19.7%	1.56	2.67	(3/82 3/8 3/8)
[alpha-proteobacteria]:					
COG0061	[Xfa]XF2090.....	-9.29%	0.927	1.89	(2/6 3/6 4/5)
COG0072	[Mtu]Rv1650_2....	-11.5%	1.52	2.16	(3/9 4/3 4/10 4/9 4/10)
COG0143	[Mtu]Rv1007c....	-8.95%	1.12	3.8	(2/10 3/9)
COG0173	[Mtu]Rv2572c....	-17.5%	1.5	2.37	(3/9:9 4/10 4/10 4/8:9 4/80)
COG0215	[Xfa]XF0995.....	-13.2%	1.28	1.89	(2/6 3/6 4/5)
COG0222	[Dra]DR2043.....	-10.3%	1.74	3.4	(2/9 3/8)
COG0552	[Dra]DR2260.....	-13.7%	1.52	2.27	(3/8 4/8 4/9 4/9)
[gamma-proteobacteria]:					
COG0102	[Dra]DR0174.....	-12.3%	1.73	2.8	(2/9 4/8 4/11)
COG0193	[Dra]DR2372.....	-16.2%	1.26	2.67	(2/70 4/9)
COG0198	[Bbu]BB0489.....	-15.8%	1.75	2.63	(3/7 4/9 4/10 4/12:0 4/12:0)
[epsilon-proteobacteria]:					
COG0085	[Aae]aq_1939....	-15.9%	1.76	2.33	(3/7 3/7)
COG0504	[Mja]MJ1174.....	-4.41%	1.14	2.9	(3/10 3/10 4/9)
[spirochetes]:					
COG0080	[Bha]BH0409.....	-6.91%	1.85	2.17	(4/6 4/10 4/10)
COG0242	[Syn]slr1549.....	-10.1%	1.47	3.2	(2/8 3/8)
COG0272	[Eco]yicF.....	-29.1%	1.64	4.8	(2/12 3/120)
[Syn+Dra+Mtu]:					
COG0178	[Hbs]VNG2636G...	-8.38%	1.47	2.6	(2/8 4/9 4/9)
COG0215	[Hbs]VNG1097G...	-23%	1.53	2.89	(2/8 3/9 4/9)
COG0292	[Tma]TM1592.....	-20.2%	1.76	2.8	(2/7 3/7)
COG0294	[Ccr]CC3224.....	-9.04%	0.901	3.8	(2/10 3/9)
COG0359	[Aae]aq_2042....	-13.2%	1.39	2.8	(2/7 3/7)
COG0441	[Ape]APE0809....	-14.1%	1.79	3	(2/8 3/7)
COG0504	[Tpa]TP0305.....	-21.5%	1.64	3.2	(2/8 3/8)
COG0556	[Hbs]VNG2390G...	-9.23%	1.53	2.33	(3/4 3/9 3/8)
[Archae]:					
COG0105	[Bbu]BB0463.....	-5.01%	1.64	2.72	(3/11 3/11 4/9 4/10 4/8)

## СПИСОК ЛИТЕРАТУРЫ

1. Boucher Y., Doolittle W.F. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microb*, 2000, vol. 37, pp. 703-716.
2. Goodman M., Czelusniak J., Moore G.W., Romero-Herrera A.E., Matsuda G. Fitting the gene lineage into the species lineage: a parsimony strategy illustrated by cladograms constructed from globin sequences. *Syst. Zool.*, 1979, vol. 28, pp. 132-163.
3. Lawrence J.G. Selfish operons and speciation by gene transfer. *Trends Microb*, 1997, vol. 5, pp. 355-359.

4. Lawrence J.G. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr. Opin. Microb.*, 1999, vol. 2, pp. 519-523.
5. Doolittle W.F. Lateral genomics. *Trends Cell Biol.*, 1999, vol. 9, pp. 5-8.
6. Lawrence J.G., Ochman H. Molecular archaeology of the *Escherichia Coli* genome. *Proc. Nat. Acad. Sci. Usa*, 1998, vol. 95, pp. 9413-9417.
7. Logsdon J.M., Faguy D.M. *Thermotoga* heats up lateral gene transfer. *Curr. Biol.*, 1999, vol. 9, pp. 747-751.
8. Nelson K.E., Clayton R.A., Gill S.R., et al. Evidence for lateral gene transfer between *Archaea* and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, 1999, no. 399, pp. 323-329.
9. Гельфанд М.С., Вьюгин В.В., Любецкий В.А. Об одном способе построения деревьев эволюции видов по множественным генетическим данным. *Информационные процессы*, 2001, т. 1, no. 1, стр. 64-77.
10. Вьюгин В.В., Гельфанд М.С., Любецкий В.А. Согласование деревьев: реконструкция эволюции видов по филогенетическим деревьям генов. *Молекулярная биология* (в печати).
11. *Bacterial Conjugation*, ed Clewell D.B. Plenum Press, New York, 1993.
12. Eulenstein O, Vingron M. Arbeitspapiere der GMD. 1995, 936. Bonn. Germany.
13. Lorencz M.G, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbial Reviews*, 1994, vol. 58, pp. 563-602.
14. Bergh O., et al. High abundance of viruses found in aquatic environments. *Nature*, 1989, no. 340, pp. 467-468.
15. Page R.D.M., Charlstone M.A. From gene to organismal phylogeny: reconciled trees and gene tree/species tree problem. *Mol. Phylogenet. Evol.*, vol. 7, pp. 231-240.
16. Page R.D.M. GeneTree: comparing gene and species phylogenies using reconciled trees. *Bioinformatics application notes*, 1998, vol. 14, pp. 819-820.
17. Waterman M.S. *Introduction to computational biology*. Chapman and Hall, 1995.
18. Wolf Y., Rogozin I, Grishin N., Tatusov R., Koonin E. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.*, 2001, vol. 1, p. 8.

*Статью представил к публикации член редколлегии В.А. Любецкий*