

КОНФЕРЕНЦИЯ МОЛОДЫХ УЧЕНЫХ И СПЕЦИАЛИСТОВ

**Информационные  
ТЕХНОЛОГИИ И СИСТЕМЫ**  
**ИТиС'09**

СБОРНИК ТРУДОВ КОНФЕРЕНЦИИ

**ISBN 978-5-901158-11-1**

пос. д/о Бекасово,  
15 – 18 декабря 2009 г.

Информационные технологии и системы (ИТиС'09):  
сборник трудов конференции. – М.: ИППИ РАН, 2009. – 463 с.

Издание содержит труды ежегодной конференции молодых ученых и специалистов «Информационные технологии и системы» (ИТиС'09), традиционно организуемой Советом молодых ученых и специалистов ИППИ РАН.

Основная цель Конференции ИТиС'09 – дать возможность молодым ученым и специалистам различных подразделений ИППИ РАН, а также студентам, аспирантам и молодым ученым других институтов РАН, отраслевых институтов, университетов и вузов, познакомиться с коллегами и обменяться научными достижениями по основным для ИППИ РАН направлениям научной деятельности: теория передачи и защиты информации; математическая теория информации и управления, многокомпонентные случайные системы; информационно-коммуникационные технологии и их применение в сложных системах и сетях; информационные процессы в живых системах и биоинформатика; компьютерная лингвистика и моделирование естественного языка; высшая математика.

Все включенные в данный сборник работы прошли рецензирование и опубликованы в том виде, в котором они были представлены авторами, среди которых молодые ученые и специалисты ИППИ РАН, ИПУ РАН, ИККИ РАН, ИПМ РАН, ИПТМУ РАН, ИМБ РАН, ИОГен РАН, МГУ им. М.В. Ломоносова, МФТИ (ГУ), МАИ, НГУ, СПбГУ ИТМО, ЛГТУ, МПГУ, МТУСИ, РХТУ и др.

Труды Конференции могут представлять интерес для ученых, студентов и аспирантов, специализирующихся в областях науки, связанных с перечисленными выше научными направлениями.

Конференция проведена при финансовой поддержке ИППИ РАН, Отделения нанотехнологий и информационных технологий РАН, Целевой программы Президиума РАН «Поддержка молодых ученых 2009 г.» и РФФИ (грант 09-07-06810моб\_г)



**ISBN 978-5-901158-11-1**

© Коллектив авторов, 2009

© Учреждение Российской академии наук Институт проблем передачи информации им. А.А. Харкевича РАН, 2009

## Реконструкция эволюции белковых семейств

К.Ю. Горбунов, Е.В. Любецкая

ИППИ РАН

lyubetsk@iitp.ru

### Аннотация

*Проведен анализ эволюции ряда белковых семейств на основе модели и алгоритма, предложенного в [1]. Алгоритм позволяет построить сценарий эволюции семейства генов, описанного данным деревом генов, относительно эволюции набора видов, описанного данным деревом видов. В частности, получен результат об эволюции NAD-зависимой ДНК-лигазы (COG0272), эволюция характеризуется: 5 дупликациями, 23 потерями и 2 горизонтальными переносами с сохранением. Эти события привязаны к определенным вершинам в деревьях генов и видов. В частности, перенос гена *ycf* из *E. coli* с источником в геноме спирохет совпадает с предсказанием в [2].*

### 1. Введение

Изучение совместной эволюции генов (белков) и соответствующих видов (первые эволюционируют «внутри» вторых) является фундаментальной проблемой эволюции. Центральная трудность (или одна из центральных) состоит в том, что эволюция гена описывается своим деревом, а эволюция вида своим, и отсутствует «единое время», в котором можно было бы рассмотреть события на том и другом деревьях. Традиционный способ преодоления этой трудности состоит в использовании так называемого отображения «альфа» дерева генов в дерево видов, введенного в [3] и широко используемого с того времени. Однако оно позволяет говорить об эволюционных событиях с генами внутри видов очень косвенным образом. Например, горизонтальный перенос гена на основе отображения «альфа» описывается в терминах окрестностей листьев в первом и втором деревьях [2], что не кажется естественным. Потеря гена не может быть показана на дереве генов, а число потерь вообще вычисляется очень косвенно на основании некоторой теоремы [4] через числа односторонних дупликаций и пропусков. Нами развивается другой подход: на основе «прямого

изображения» эволюции гена внутри дерева видов. В этой работе мы ставим целью показать, что наш подход согласуется с традиционным и в то же время дает больше информации об эволюционных событиях молекулярного уровня. Для этого мы рассмотрели эволюцию нескольких КОГов.

### 2. Методы

В работе [1] развита следующая модель эволюции гена (белкового семейства) внутри соответствующих видов. Рассматриваются корневые растущие вниз деревья генов  $G$  и видов  $S$ , к которым добавлены выходящие из корня вверх ребра. Эти ребра называются корневыми, а новая самая высокая вершина называется суперкорнем. Множество ребер в дереве видов  $S$  разбито на части, называемые временными слоями; при этом исходные ребра иногда разбиваются добавленными вершинами. Предполагается, что горизонтальный перенос гена может происходить с ребра  $a$  дерева  $S$  на несравнимое с ним ребро  $b$  этого дерева из того же временного слоя. Пусть каждый лист дерева генов  $G$  помечен именем некоторого гена, а каждый лист дерева  $S$  помечен именем некоторого вида и множеством имен генов, взятых из этого вида, это множество может быть и пустым. Деревья  $G$  и  $S$  бинарные, вспомогательное также бинарное дерево  $G'$ , которое мы сейчас опишем, имеет у некоторых из его листьев крестик, который означает потерю гена, приписанного этому листу. Оно называется внутренним деревом.

Дано семейство (ортологических или гомологичных) генов, имена которых представлены в листьях дерева видов  $S$ . Типично это – множество генов, приписанных листьям некоторого дерева генов  $G$ . Представим себе ребра дерева  $S$  в виде полых труб и будем называть их трубами. Вершины дерева  $S$  будем называть началом или концом соответствующей трубы. Рассмотрим некоторое дерево  $G'$ , листьям которого взаимно однозначно приписаны имена генов из данного их семейства. Оно расположено внутри труб дерева  $S$  следующим образом. Дерево  $G'$  имеет суперкорень в корневой трубе и имеет ребра, идущие вниз

целиком внутри труб, его листья вместе с именами совпадают с соответствующими листьями дерева  $S$ . В трубе могут проходить несколько ребер, ребро должно разветвиться в развилке труб (что означает: ген удвоился в связи с видообразованием), но может разветвиться и внутри самой трубы (что означает дубликацию гена); листу этого дерева может быть приписан крестик (потеря гена). Все разветвления в дереве  $G'$ , которые не относятся к видообразованию (или к горизонтальным переносам, ниже), считаются дубликациями. Задача состоит в том, чтобы найти внутреннее дерево  $G'$  в данном дереве видов  $S$ , которое после удаления ребер, помеченных крестиками, было бы изоморфно  $G$ . Для описания горизонтальных переносов в бинарном внутреннем дереве  $G'$  разрешается провести ребро от одной его вершины  $a$ , лежащей внутри некоторой трубы, к другой его вершине  $b$ , лежащей уже в несравнимой трубе (из того же временного слоя) или в ее конце. Переносом гена (точнее, переносом гена с сохранением копии) называется перенос, при котором в трубе, содержащей вершину  $a$ , сохраняется копия гена из этой вершины, в то время как другая копия перенеслась по стрелке в несравнимую трубу. Иногда рассматривают событие переноса гена без сохранения его копии. Это – последовательность двух событий: переноса гена с сохранением копии с последующей потерей этой копии. Рассматриваются только те деревья  $G'$ , на которых достигается минимума функция (цена вложения), равная сумме с некоторыми положительными весами значений чисел дубликаций, потерь и горизонтальных переносов с сохранением исходного гена. В счете, результаты которого приведены ниже, веса принимают следующие значения: за каждую потерю полагаем цену 2, за каждую дубликацию – цену 3, за каждый перенос с сохранением – цену 11 (переносу без сохранения соответствует цена 13). Таким образом, авторы этой работы продолжили построение и анализ внутренних деревьев и соответственно эволюции гена внутри соответствующих видов. Из этих результатов ниже приводится случай COG0272 из базы данных GenBank, NCBI.

### 3. Результаты

В эволюции COG0272 (NAD-зависимая ДНК-лигаза) модель и алгоритм из [1] указали на следующие эволюционные события (перед скобками указан номер трубы, который по определению совпадает с номером ее конца, а в скобках перечислены вершины внутреннего дерева, соответствующие генам, которые в этой трубе дублировались или потерялись; число штрихов указы-

вает номер копии исходного гена, которая потерялась):

пять дубликаций: 0(0); 1(20,22,32); 5(35); 23 потери: 2(21',23'); 3(2'); 5(57'); 7(38'); 18(39'); 19(39''); 21(57''); 24(50'); 25(34'); 28(34'',56'); 35(2'',56''); 36(23'); 38(21''); 39(55'); 40(21'''); 41(21''''',53'); 42(23'''); 56(20'); 57(17'); и два горизонтальных переноса с сохранением: гена 9 из трубы 32 в трубу 42 и гена 7 из трубы 33 в трубу 12. Этот последний перенос является переносом гена *yicF*. Дерево этого КОГа показано на рис. 1.

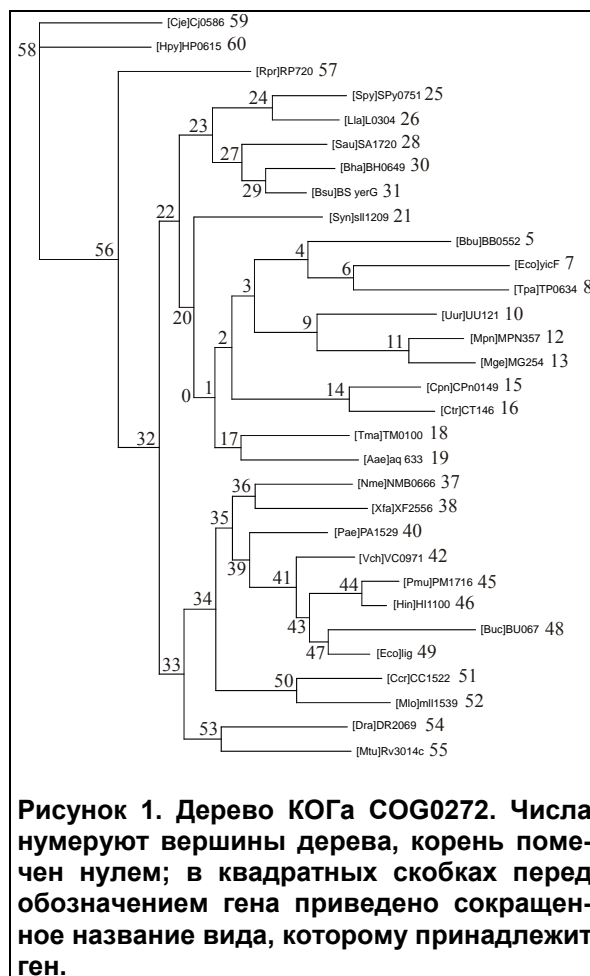
На рис. 2 приведено использованное нами дерево видов.

На рис. 3 показан фрагмент построенного нами вложения (т.е. внутреннего дерева  $G'$ ) дерева данного КОГа в данное дерево видов (для удобства на рисунке не показаны древние виды Tma и Aae, кроме того в этом КОГе нет представителей архей).

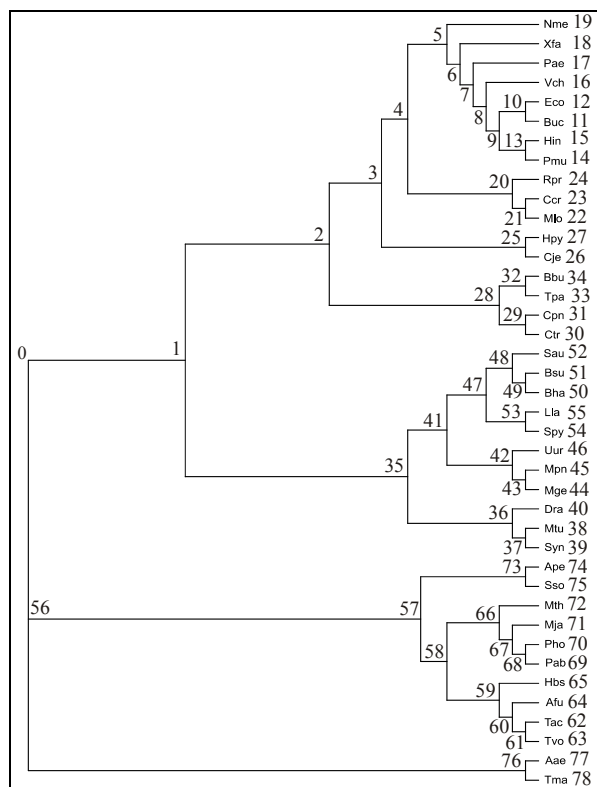
В данном случае такое дерево  $G'$  единственно.

### 4. Обсуждение

Алгоритм из [2] показал, что ген *yicF* из *E. coli* горизонтально перенесен из генома спирохет.



**Рисунок 1. Дерево КОГа COG0272. Числа нумеруют вершины дерева, корень помечен нулем; в квадратных скобках перед обозначением гена приведено сокращенное название вида, которому принадлежит ген.**

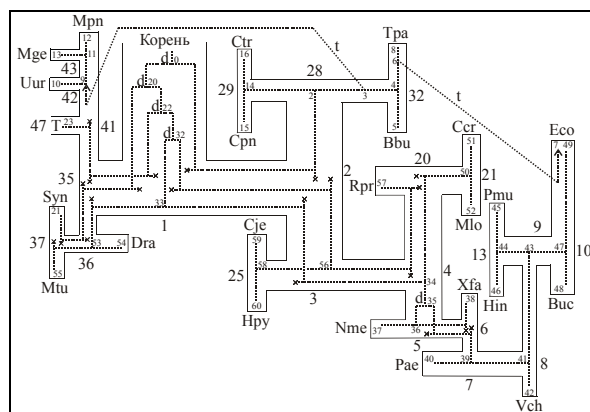


**Рисунок 2. Дерево видов. Числа нумеруют вершины дерева, корень помечен нулем; в листьях стоят сокращенные названия видов.**

Таким образом, этот и другие полученные нами результаты согласуются с известными предсказаниями горизонтальных переносов, а в ряде случаев дают новые предсказания. При этом возникает более полная картина эволюции, чем та, которая возникает на основе широко применяемого вложения «альфа» дерева генов в дерево видов, предложенного в [3].

## 5. Список литературы

[1] К.Ю. Горбунов, В.А. Любецкий, “Реконструкция эволюции генов вдоль дерева видов”, *Молекулярная биология*, 2009, т. 43, №5, сс. 1-14.



**Рисунок 3. Фрагмент вложения дерева КОГа COG0272 в дерево видов. Внутреннее дерево показано пунктирной линией внутри дерева видов. Его вершины, соответствующие дубликации гена, помечены буквой “d”, соответствующие видообразованию не помечены, а соответствующие потере помечены крестиком. Горизонтальные переносы помечены буквой “t”. Вложение поддерева дерева КОГа с корнем 23 в поддерево дерева видов с корнем 47 оказалось тождественным и не показано (корень этого поддерева отмечен буквой “Т”).**

[2] V.A. Lyubetsky, K.Yu. Gorbunov, L.Y. Rusin, V.V. V'yugin. “Algorithms to reconstruct evolutionary events at molecular level and infer species phylogeny”. In: *Bioinformatics of Genome Regulation and Structure II*, Springer Sci. & Business Media, Inc., 2006, pp. 189–204.

[3] R. Guigo, I. Muchnik, T.F. Smith, “Reconstruction of ancient molecular phylogeny”, *Mol. Phylogenet. Evol.*, 1996, vol. 6, pp. 189–213.

[4] O. Eulenstein, B. Mirkin, M. Vingron, “Duplication-based measures of difference between gene and species trees”, *J. Comput. Biol.*, 1998, vol. 5, pp. 135–148.