

**ГЕНЕТИЧЕСКАЯ ОСНОВА МАКРОЭВОЛЮЦИОННЫХ ПРЕОБРАЗОВАНИЙ:
ИССЛЕДОВАНИЕ РЕЖИМОВ МОЛЕКУЛЯРНОЙ ЭВОЛЮЦИИ
ОРТОЛОГИЧНЫХ БЕЛКОВ ПОЗВОНОЧНЫХ И БЕСПОЗВОНОЧНЫХ**

Работа выполнена при поддержке грантом РФФИ №09-04-01641-а.

К.В. Гунбин, В.В. Суслов

Институт Цитологии и Генетики СО РАН, Новосибирск, Россия

E-mail: genkvg@bionet.nsc.ru, valya@bionet.nsc.ru

Д.А. Афонников

*Институт Цитологии и Генетики СО РАН, Новосибирский государственный
Университет, Новосибирск, Россия*

E-mail: ada@bionet.nsc.ru

В 2008 г. было четко показано, что естественный отбор белок-кодирующих генов неоднороден по своей природе и связан с двумя разными явлениями [1]: оптимизацией уровня экспрессии гена и оптимизацией структурно-функциональной организации белка. Очевидно, что второе явление связано прежде всего с аминокислотными заменами, по-разному влияющими на структуру и функцию белка. На этой основе нами был создан простой метод анализа режимов эволюции белков, основанный на Марковском моделировании эволюции белков. Используя созданный метод, в настоящей работе мы провели анализ режимов эволюции ортологичных белков позвоночных и беспозвоночных.

В анализ брались ортологичные группы белков из базы данных MetaPhOrs [2] содержащие по одному наиболее подтвержденному ортологу из 20 таксономических групп позвоночных (Amphibia, Tetraodontiformes, Smegmamorpha, Cypriniformes, Aves, Squamata, Prototheria, Metatheria, Afrotheria, Xenarthra, Insectivora, Carnivora, Equidae, Chiroptera, Artiodactyla, Catarrhini, Hominidae, низшие Primates, Lagomorpha и Rodentia) и 9 таксономических групп беспозвоночных (Lophotrochozoa, Caenorhabditis, низшие Chromadorea, Arachnida, Paraneoptera, Lepidoptera, Hymenoptera, Drosophilina и Culicomorpha). В качестве аутгруппы были использованы таксоны Cnidaria и Tunicata. На основе известной топологии деревьев позвоночных и беспозвоночных (Рис. 1) и вычисленных нами специфических для каждой ортологической группы матриц скоростей аминокислотных замен (с помощью программы MODELESTIMATOR [3]) в

каждом внутреннем узле дерева с помощью PAML [4] реконструировались предковые последовательности белков. Эти последовательности использовались для подсчета числа наблюдаемых аминокислотных замен. Для поиска нетипичных типов аминокислотных замен на каждой ветви дерева мы сравнивали число наблюдаемых замен с их ожидаемыми значениями для каждого типа замен в предположении о стационарном Марковском процессе эволюции. Ожидаемые значения замен подсчитывались на основе 1000 компьютерных симуляций молекулярной эволюции белков при помощи пакета INDELible [5], с учетом особенностей исследуемых белков. Сравнение числа ожидаемых и наблюдаемых замен каждого типа производилось с помощью перестановочного теста (10^5 перестановок). В ходе теста подсчитывалось число случайных выборок M , в которых частота ожидаемых замен определенного типа больше частоты наблюдаемых замен. Величина $M/10^5$ оценивала вероятность p , с которой встречаемость замен определенного типа, наблюдаемая в исходной выборке генов, могла возникнуть по случайным причинам.

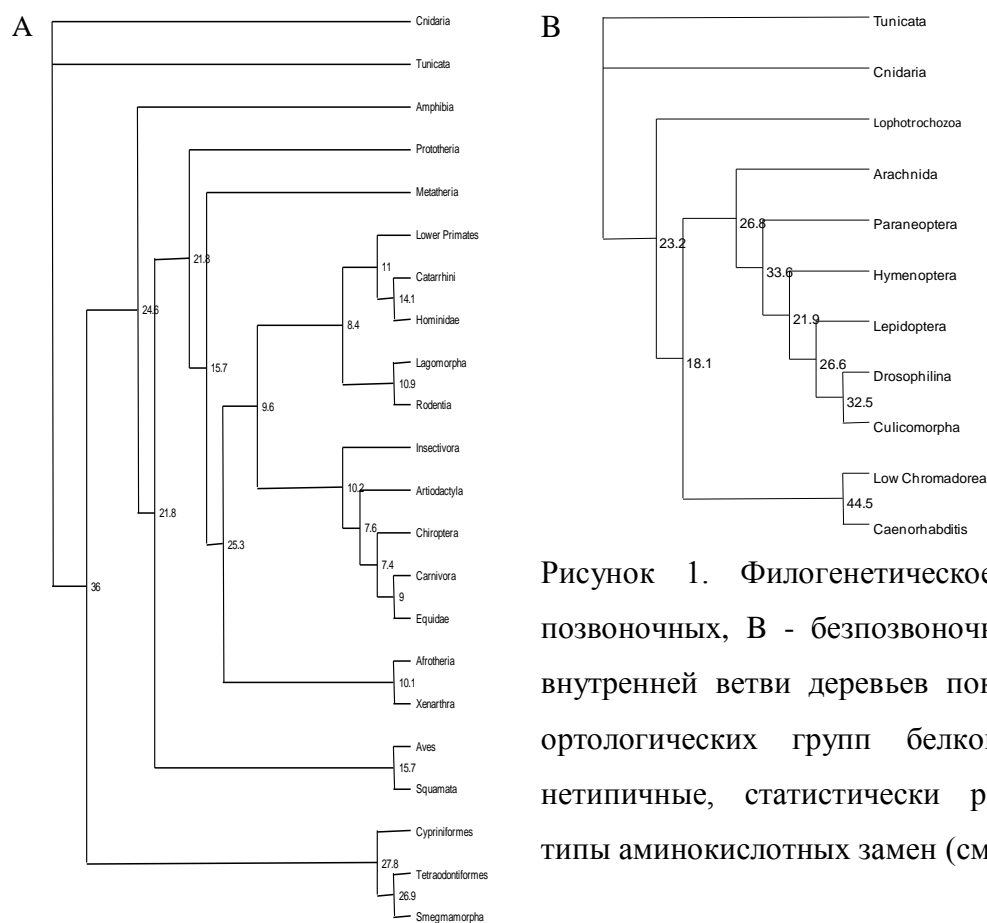


Рисунок 1. Филогенетическое дерево А - позвоночных, В - беспозвоночных. На каждой внутренней ветви деревьев показана доля (%) ортологических групп белков, содержащих нетипичные, статистически редкие ($p < 0.01$), типы аминокислотных замен (см. текст).

В результате проведенного анализа было показано, что внутренние ветви, содержащие >20% ортологичных групп белков содержащих нетипичные ($p < 0.01$), аминокислотные замены соответствуют основным ароморфозам позвоночных (Рис 1.): 1) периодам полногеномных дупликаций при формировании позвоночных и рыб, 2) выходу на сушу, 3) формированию амниот, 4) дивергенции млекопитающих, а также группы плацентарных. Также было показано значительное по сравнению с позвоночными (>12%) и с кембрийскими дивергенциями (>7%) увеличение числа ортологичных групп белков, содержащих нетипичные, аминокислотные типы замен ($p < 0.01$) в группах насекомых и нематод (Рис 1.). Абсолютный максимум этого числа характерен для формирования группы нематод, интенсивно теряющей доменное разнообразие белков [6]. Показано, что формирование групп Insecta и Diptera сопровождаемое увеличением числа ортологичных групп белков, содержащих нетипичные, аминокислотные типы замен может быть связано с формированием тесного экологического сообщества насекомых и покрытосеменных растений [7]. В настоящей работе также проведен сравнительный анализ функций ортологичных групп белков содержащих нетипичные замены на каждой внутренней ветви дерева позвоночных и беспозвоночных, что позволило связать режимы молекулярной эволюции определенных ортологичных групп с морфологическими преобразованиями в палеонтологической истории таксонов Bilateria (Рис. 1).

Использованная литература

1. Drummond D.A., Wilke C.O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008. Vol. 134, № 2. P. 341-352.
2. Prysycz L.P. et al. MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucl. Acids Res.* 2010. doi: 10.1093/nar/gkq953.
3. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood, *Mol. Biol. Evol.* 2007. Vol. 24, № 8. P. 1586-1591.
4. Arvestad L. Efficient methods for estimating amino acid replacement rates, *J. Mol. Evol.* 2006. Vol. 62, № 6. P. 663-673.
5. Fletcher W., Yang Z. INDELible: a flexible simulator of biological sequence evolution, *Mol. Biol. Evol.* 2009. Vol. 26, № 8. P. 1879-1888.

6. Zmasek C.M., Godzik A. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires, *Genome Biol.* 2011. Vol. 12, № 1. P. R4.

7. Grimaldi D., Engel M.S. *Evolution of the insects.* NY: Cambridge University Press, 2005. 755p.

**DEEP INSIDE INTO VERTEBRATES AND INVERTEBRATES
MACROEVOLUTION: THE MOLECULAR EVOLUTION MODES OF STRICT
ORTHOLOGOUS PROTEIN SEQUENCES**

This work was supported by RFBR grant No. 09-04-01641-a.

K.V. Gunbin, V.V. Suslov

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

E-mail: genkvg@bionet.nsc.ru, valya@bionet.nsc.ru

D.A. Afonnikov

Institute of Cytology and Genetics SB RAS, Novosibirsk State University, Novosibirsk, Russia

E-mail: ada@bionet.nsc.ru

In 2008, it was clearly demonstrated that the natural selection of protein-coding genes is heterogeneous in nature and associated with two different phenomena [1]: optimization of the gene expression level and optimization of protein structural and functional organization. It is known, that the second phenomenon is primarily associated with amino acid substitutions having different effects on the protein structure and function. Thus, simple method for the analysis of protein evolution modes was developed. This method is based on Markov simulation of protein sequence evolution. This study, we analyzed the evolution modes of vertebrate and invertebrate orthologous protein sequences using our novel method.

The molecular evolution modes of 752 orthologous protein groups (OPGs) of vertebrates and of 1736 OPGs of invertebrates were analyzed. In our analysis the OPGs were taken from MetaPhOrs database [2]. The OPGs containing at least one strictly confirmed orthologous sequence without ambiguous characters from each of 20 vertebrate species groups and 9 invertebrate species groups were taken into analysis (Figure 1). Cnidaria and Tunicata were served as outgroup taxa. The ancestral protein reconstruction in each internal tree node of OPG was made using PAML 4 CODEML program [3] (marginal reconstruction) on the basis of the known tree topology (Figure 1) and amino acid replacement matrices calculated by MODELESTIMATOR [4] program. These ancestral sequences were used to calculate the number of observed amino acid substitutions. For each amino acid replacement type we compared the observed number of changes with expected ones under the assumption of a stationary Markov process of protein evolution. Expected replacement numbers were calculated using 1000 computer simulations of OPG molecular evolution by the INDELible

program [5], taking into account the peculiarities of the investigated OPGs. Comparison of expected and observed numbers of each replacement type was performed using permutation tests (10^5 permutations). We count the number of random samples, M , in which the frequency of expected changes of a certain type higher than the frequency of observed ones. Thus, the value $M/10^5$ is the occurrence probability, p , of a certain amino acid replacement type observed by chance.

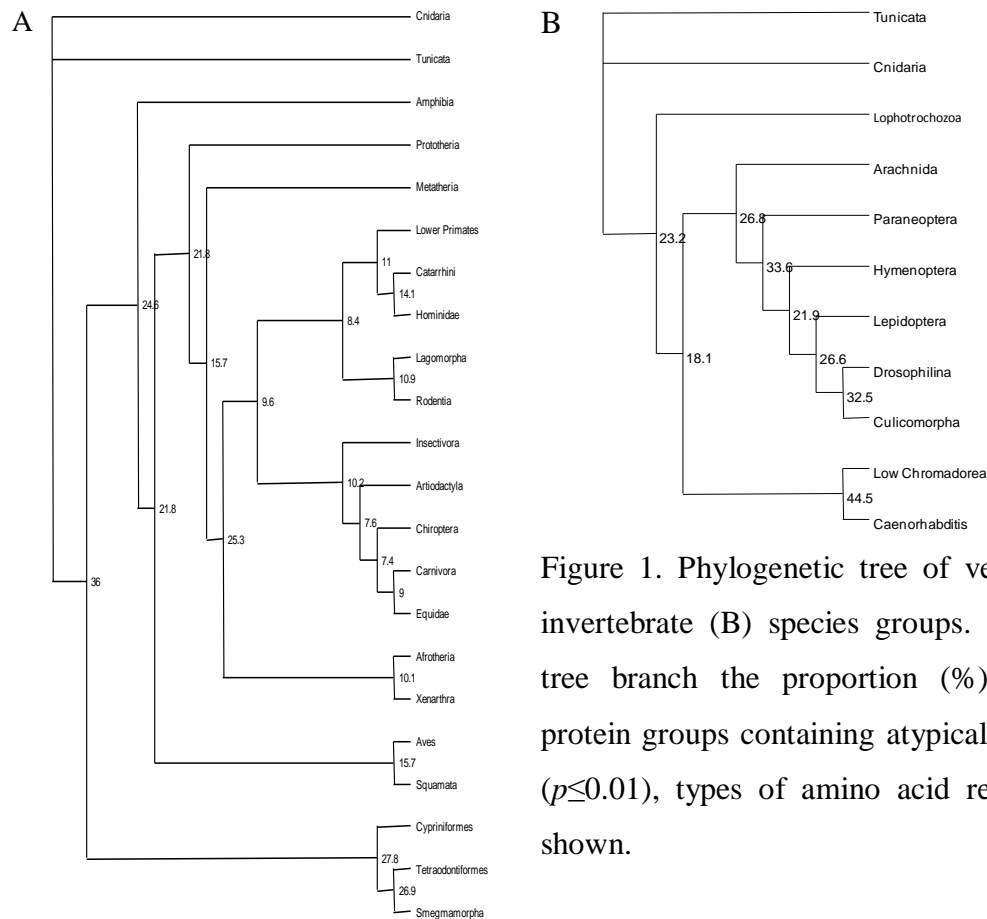


Figure 1. Phylogenetic tree of vertebrate (A) and invertebrate (B) species groups. On each internal tree branch the proportion (%) of orthologous protein groups containing atypical, statistically rare ($p \leq 0.01$), types of amino acid replacements were shown.

The analysis of R value, the proportion (%) of OPGs with atypical amino acid replacements ($p \leq 0.01$), showed that internal tree branches of vertebrate tree with $R \geq 20\%$ strictly correspond to aromorphoses in the vertebrate evolution (Figure 1): 1) the full genome duplications in early stages of vertebrate evolution and actynopterigian fish evolution, 2) the adaptation to terrestrial environments, 3) the origin of Amniota, 4) the divergence of primitive mammals and placental mammals. It is of big interest that the molecular evolution of insects and nematodes were characterized by significant increase in R value in comparison with its mean value for vertebrates (Figure 1). The absolute maximum of R value is typical for nematodes - a group with intensive lost of protein domain diversity [6]. It was shown that the

divergence of Insecta and Diptera accompanied by increasing of R (Figure 1) may be due to the emergence of insects-angiosperms ecosystems and to the formation of the characteristic Diptera morphology [7]. This study we also conducted the functional comparison of OPGs containing atypical amino acid replacements that allowed us to uncover various features of gene networks molecular evolution on each internal branches of vertebrate and invertebrate trees (Figure 1).

References

1. Drummond D.A., Wilke C.O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008. Vol. 134, № 2. P. 341-352.
2. Prysycz L.P. et al. () MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucl. Acids Res*. 2010. doi: 10.1093/nar/gkq953.
3. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood, *Mol. Biol. Evol*. 2007. Vol. 24, № 8. P. 1586-1591.
4. Arvestad L. Efficient methods for estimating amino acid replacement rates, *J. Mol. Evol*. 2006. Vol. 62, № 6. P. 663-673.
5. Fletcher W., Yang Z. INDELible: a flexible simulator of biological sequence evolution, *Mol. Biol. Evol*. 2009. Vol. 26, № 8. P. 1879-1888.
6. Zmasek C.M., Godzik A. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires, *Genome Biol*. 2011. Vol. 12, № 1. P. R4.
7. Grimaldi D., Engel M.S. *Evolution of the insects*. NY: Cambridge University Press, 2005. 755p.