Metazoa evolution: the relation between molecular evolution of orthologous protein groups and aromorphoses

K.V. Gunbin, D.A. Afonnikov

Institute of Cytology and Genetics SB RAS, Novosibirsk State University, Novosibirsk, Russia,
genkvg@bionet.nsc.ru, ada@bionet.nsc.ru

The molecular evolution of more than 139 000 compact orthologous protein groups (COPGs) containing about 2 000 000 unique proteins from 100 fully sequenced organisms belonging to 25 Metazoa species groups (Figure 1) were analyzed. Orthology relations were taken from OrthoDB 6 database [1]. To select orthologous proteins highly related with each other we define COPGs as a parts of diffuse orthologous protein groups (OPGs) containing proteins belonging to minimum 5 species groups with monophyletic origin. For example OPG composed of Tunicata, Agnata, Actinopterigii, Sarcopterigii, Amphibia, Dinosauria, and Afrotheria species groups contained only one COPG with 6 species groups (without Afrotheria). The ancestral protein reconstruction in each internal tree node of COPG was made using RAXML 7.4.2 on the basis of best COPG protein alignment (selected by AQUA 1.1) containing only columns with limited variation (selected using NOISY 1.5.12), the COPG tree (FASTTREE 2.1) corrected by species tree (using TREEFIX [2]) and calculated amino acid replacement matrices (via MODELESTIMATOR 1.1). These ancestral sequences were used to calculate the number of observed amino acid substitutions. For each observed well confirmed amino acid replacement type (by the RAXML ancestral state reconstruction probabilities) 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 simulations of COPGs molecular evolution by the INDELible 1.03, taking into account the peculiarities of the investigated COPGs (alignment length, phylogenetic tree, amino acid replacement matrix and frequencies of occurrence). Comparison of expected and observed numbers of each replacement type was performed using new highly computationally optimized permutation test [3].
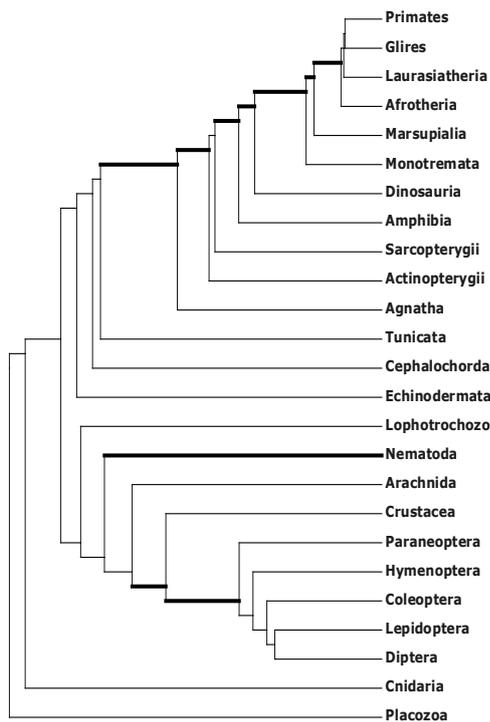
Fig. 1. Phylogenetic tree of Metazoa species groups under analysis. Bold tree branches denote cases with statistical excess of COPGs containing atypical, statistically rare ($p \leq 0.01$) types of amino acid replacements.

We showed that internal tree branches of Metazoa, containing excess of COPGs with atypical ($p < 0.01$) amino acid replacement types strictly correspond to aromorphoses in the vertebrate and invertebrate evolution: 1) the full genome duplications in early stages of vertebrate evolution, 2) the adaptation of vertebrates to terrestrial environments, 3) the origin of Amniota, 4) the divergence of primitive mammals and placental mammals; 5) the divergence of Pancrustacea and Insecta; 6) the divergence of Nematoda. We also conducted the functional enrichment permutation test of COPGs containing atypical amino acid replacement at each inner tree branch. This test allows us to uncover various features of Metazoan gene networks evolution related with Vertebrate and Invertebrate aromorphoses.

1. R.M. Waterhouse et al. (2013) OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs, *Nucl. Acids Res.*, **41**:D358-D365.

2. Y.C. Wu (2013) TreeFix: statistically informed gene tree error correction using species trees, *Syst. Biol.*, **62**:110-120.

3. K.V. Gunbin, A. Ruvinsky (2013) Evolution of general transcription factors, *J. Mol. Evol.*, **76**:28-47.