The Evolution of the Hh-Signaling Pathway
Genes: A Computer-Assisted Study

Konstantin V. Gunbin\textsuperscript{a,}\textsuperscript{*}, Dmitrij A. Afonnikov\textsuperscript{a,\textsuperscript{b} and Nikolay A. Kolchanov\textsuperscript{a}}

\textsuperscript{a}Institute of Cytology and Genetics SB RAS, Lavrentyev Ave. 10, Novosibirsk, 630090, Russia
\textsuperscript{b}Novosibirsk State University, Pirogova Str. 2, Novosibirsk, 630090, Russia

Edited by S. Rodin (guest editor) and N. Kolchanov; received 20 January 2007; revised and accepted 29 March 2007; published August 2007

ABSTRACT: Positive selection of genes that comprise signaling cascades and play the paramount role in the development of multicellular organisms is critical to our understanding of the reasons for the evolution of embryonic development. In this work, we analyze the evolution of 9 genes involved in the function of the Hh signaling cascade. We demonstrated that positive selection is a characteristic feature of the protein domains, encoded by gene regions, whose functions are related to the molecular mechanisms of development. We also found that the positive selection of Hh-signaling cascade transcription factors, morphogens, their development-related receptors and intracellular signal transduction factors are related to the divergence of the Bilateria taxonomic types.

KEYWORDS: Hh-signaling cascade, positive selection, Bilateria, compensatory changes

INTRODUCTION

The evolution of the fundamental molecular mechanisms underlying the embryonic animal development has become a key issue in the modern theory of evolution [Sole et al., 2003; Shu et al., 2006]. Allometric growth and symmetry of the bilaterian body plans, which appeared at about the same time (~570 MYA), have suggested that the bilaterian ancestors possessed a new type of embryogenesis (compared with the one that arose in the vendobionts, the antique multicellular animals) giving rise to the formation of new tissues and organs [Knoll and Carroll, 1999; Erwin and Davidson, 2002]. The embryonic type of development specific to the now existing bilaterian animals has supposedly passed during the Precambrian evolution of Eumetazoa (~50–80 MYA geological period) and morphologically it involved the following crucial events [Davidson et al., 1995; Peterson et al., 2000]: 1) the emergence of molecular mechanisms for cell differentiation; 2) the rise of primitive embryogenesis with cell and tissue differentiation immediately after fertilization; 3) the integration of the molecular mechanisms of cell proliferation – differentiation and the appearance of cells capable of embryonic induction; 4) the development a common molecular mechanism for the differentiation of these cells. It has been assumed that the rise of new tissues and organs during evolution was associated with a drastic reorganization of the molecular developmental mechanisms and also with the appearance of genes expressed in particular

\textsuperscript{*}Corresponding author. E-mail: genkvg@bionet.nsc.ru.

Electronic publication can be found in In Silico Biol. 7, 0047 <http://www.bioinfo.de/isb/2007/07/0047/>, August 2007.

1386-6338/07/$17.00 © 2007 – IOS Press and Bioinformation Systems e.V. and the authors. All rights reserved
organs and tissues [Sole et al., 2003; Salazar-Ciudad et al., 2003; Salazar-Ciudad and Jernvall, 2004]. Thus, it appears that the adaptive mode of the evolution of genes defining the formation of new tissue and organ may generate the type of Bilateria embryogenesis and give rise to different bilaterian body plans.

An important molecular mechanism that provides cell differentiation is signal transduction [Pires-daSilva and Sommer, 2003]. The signal transduction pathways ensure the reception of the concentration gradients of morphogens and their transformation into the differentiated state of cells within tissues and organs. Hence it may be assumed that at the molecular level the key molecular rearrangements may be related to changes in genes that participate in signal transduction pathways. A study of the events that had occurred during the stages of evolution corresponding to the emergence of Bilateria may provide useful information on the evolution of the molecular mechanisms controlling cell differentiation [Raff, 1996].

The Hedgehog (Hh) cascade, its sensors, switches and routers have been amply described and studied [Lum and Beachy, 2004]. It is involved in diverse morphogenetic processes [Ingham and McMahon, 2001; Nybakken and Perrimon, 2002]. A set of proteins accomplishing versatile functions can be distinguished in the Hh signal cascade. These include the Hh morphogen, the Ptc and Smo cell receptors, the Cos2, PKA, Slmb, Su(Fu) and Fu proteins that make up the high molecular weight complex transducing the Hh signal within the cell nucleus, and the Ci transcription factor. It is known how these proteins interact in the functioning Hh-cascade and the interactions can be described as a gene network. A computer-based description has been made and deposited in the Gene Net database [Gunbin et al., 2004a]. In the current study, we analyze the evolution of proteins involved in the function of the Hh-cascade at the steps that match with the evolutionary events at the morphological level related to the appearance of the Bilateria taxa and the divergence of the Bilateria taxonomical types on the phylogenetic tree (the divergence between Ecdysozoa, Lophotrochozoa and Deuterostomia groups of taxonomical types) [Sepkoski and Markov, 1999; Peterson et al., 2004; Benton and Donoghue, 2006].

The aim of the present study was to identify positive selection in the various functional regions of proteins involved in the Hh-cascade, on the one hand, and on phylogenetic tree branches that match with the divergence of the main Bilateria taxonomical types during the evolution of the multicellular organisms, on the other hand.

Our analysis is based on identification of gene regions showing an excess of nonsynonymous over synonymous fixations [Kimura, 1983]. An excess of nonsynonymous substitutions over synonymous substitutions is an unambiguous indicator of positive natural selection in the various regions of genes and it may be connected with functional changes in protein domains [Kimura, 1983; Suzuki, 2004; Merl et al., 2005]. There are many approaches for searching of the positive selection mode. One approach defines the ratio of the fixation rate of the nonsynonymous ($K_a$) to the synonymous ($K_s$) mutations in various gene regions or in discrete codons. From the results it may be concluded that positive selection acts with different pressures on the different gene regions. Another approach defines the $K_a/K_s$ ratios on separate branches of the phylogenetic tree, allowing to answer the question: At what evolutionary step genes subject to positive selection (as a rule, this selection might have been related to the acquirement of a new function by a protein after gene duplication)? [Yang and Nielsen, 2000; Tzeng et al., 2004; Tang and Wu, 2006].

Here we search and analyze the events of Darwinian positive selection in the genes involved in the Hh-cascade. We started by searching regions in genes which are participants in the cascade and for which selection is positive at the codon level. We proceeded on the assumption that the protein domains, encoded by these gene regions, may be of functional importance with regard to morphogenesis and cell
differentiation. Next, we searched positive selection of the protein domains at the evolutionary step matching the divergence of Bilateria taxonomical types. We also assumed that the genes under positive selection at this stage of evolution may be related to the emergence of the diverse bilaterian body plans.

As a result, we demonstrated that positive selection is a characteristic feature of the protein domains, encoded by gene regions whose functions are related to the molecular mechanisms of morphogenesis. Our salient finding was that amino acid substitutions become fixed at those stages of evolution that match with the divergence of the Bilateria taxonomical types [Gunbin et al., 2005].

SUBJECT OF INQUIRY

The classical model of the Hh signaling pathway in invertebrates includes the following series of events (Fig. 1) [Ingham and McMahon, 2001; Nybakken and Perrimon, 2002; Lum and Beachy, 2004]: The Disp gene codes for the enzyme responsible for exocytosis of the Hh morphogen [Burke et al., 1999; Gallet et al., 2003; Peters et al., 2004]. In the absence of Hh, the Ptc morphogen receptor inhibits the transduction of the signal into the cell through the Smo protein. The Cos2, Fu, and Ci proteins form a high-molecular-weight complex. In the absence of Hh, the complex is bound to the cytoskeleton, and the Ci transcription factor is phosphorylated by PKA protein kinase and cleaved within this complex. After the cleavage, the N-terminal Ci fragment is transported to the nucleus and represses a gene cassette. Interaction between Hh and its receptor Ptc eliminates the inhibitory effect of Ptc on Smo activity. The activated Smo triggers a series of reactions resulting in removal of Fu/Cos2/Ci from the cytoskeleton and hyperphosphorylation of Fu and Cos2. This results in stabilization of the full-sized Ci and its transport to the nucleus. The Su(Fu) protein is bound to Fu and Ci. Su(Fu) is involved in retaining Ci in the cytoplasm. The Slmb protein is responsible for ubiquitination of Ci.

We have compiled a computerized description of the gene network for this pathway, it includes a description of the interactions of 33 genes and 46 proteins and protein complexes, based on analysis of 150 publications [Gunbin et al., 2004a]. Here we present the results of an analysis of sequences of the following gene families: Hh, Ptc, Smo, Disp, PKA, Slmb, Su(Fu), Fu, and Ci (data is available from the authors upon request). The Cos2 family was omitted from consideration, because there were no experimentally annotated protein sequences of vertebrate species [Katoh, Y. and Katoh, M., 2004]. We also examined published experimental data on the functions of domains of the proteins encoded by genes under study (protein domains are indicated with the same numerals in Table 1 and in Fig. 2).

METHOD

Analysis of amino acid and nucleotide sequences included the following steps: (1) search for similar sequences for protein families involved in the operation of the Hh pathway; (2) alignment of amino acid sequences; (3) construction of phylogenetic trees; (4) search for regions of genes for the proteins found evolving under adaptive selection; (5) identification of the positive selection mode for branches of the phylogenetic tree corresponding to the stage of divergence of the major groups of Bilateria phyla.

Sampling of Amino Acid and Nucleotide Sequences


**Data Processing (Multiple Alignment and Construction of Phylogenetic Trees)**

Preliminary alignment of protein sequences was performed using algorithms CLUSTALW (http://www.ebi.ac.uk/clustalw/) [Thompson et al., 1994], T-COFFEE (http://www.tcoffee.org/Projects_home_page/t_cooffee_home_page.html) [Notredame et al., 2000], and MAFFT (http://align.bmr.kyushu-u.ac.jp/mafft/software/) [Katoh, K. et al., 2002; 2005]. The alignments were improved with the RASCAL package (ftp://ftp-igbmc.u-strasbg.fr/pub/RASCAL) [Thompson et al., 2003]. The quality of the resulting preliminary alignments was assessed as follows. Regions corresponding to conserved domains of proteins stored in NCBI CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [Marchler-Bauer et al., 2005] were identified in the sequences of proteins aligned; column score was estimated for these regions as in [Thompson et al., 1999]. The alignment with the highest score was chosen from three preliminary alignments constructed with different programs. The best alignment was used for the subsequent analysis and as the source for constructing aligned cDNA sequences for positive selection analysis.

Phylogeny was reconstructed from the multiple alignment of amino acid sequences by the maximum likelihood method implemented in the PHYML program (http://atgc.lirmm.fr/phyml/) [Guindon and Gascuel, 2003]. We used the replacement matrix that satisfied best the AIC criterion implemented in the PROTTEST program (http://darwin.uvigo.es/software/prottest.html) [Abascal et al., 2005]. For all families, it was the WAG matrix [Whelan and Goldman, 2001] that satisfied the criterion. For tracing errors of phylogeny reconstruction, which could result from sequencing errors or nonuniform evolution of protein regions, we compared the topologies of the phylogenetic trees with those of trees stored in the Tree of Life database (http://www.tolweb.org/tree/) [Maddison and Schulz, 1996-2006]. Sequences whose positions on the tree we reconstructed did not match with the taxonomic positions of the corresponding species in the Tree of Life were excluded as recommended in [Nei et al., 1998; Sanderson and Shaffer, 2002].

**Detection of Positive Selection in Gene Sequences**

Detection of gene regions evolving under positive selection was an essential part of our analysis. First, this analysis allows determination of protein regions where amino acid replacements accumulated faster than expected. This may be related to functional divergence of the regions and, as a consequence, to their specific role in the operation of the Hh signaling pathway. Second, comparison of such gene regions appeared promising in revealing selection events on phylogenetic tree branches.

There are many methods allowing search for gene regions under adaptive selection. However it is known that different methods are differently sensitive to specific features of nucleotide sequences and to their similarity degree [Suzuki, 2004; Merl et al., 2005]. We sought for regions evolving in the adaptive mode using algorithms: PSWIN (available from S.D. Reid) [Whittam and Nei, 1991], ADAPTSITE (http://www.cib.nig.ac.jp/dda/yossuzuk/welcome.html) [Suzuki et al., 2001; Suzuki, 2004], PLATO (http://evolve.zoo.ox.ac.uk/software.html?id=plato) [Grassly and Holmes, 1997], and PAML (http://abacus.gene.ucl.ac.uk/software/paml.html) [Yang, 1997]. These programs implement various measures and algorithms for identification of gene regions of one codon and more in length, where nucleotides are replaced in the course of positive selection. The final conclusion on the existence of
Fig. 1. Schematic presentation of the Hh signaling pathway. (A) Exocytosis of the Hh morphogen. Succession of Hh-pathway events: (B) without Hh; (C) with Hh.
Fig. 2. Detection of regions evolving under positive selection in nine genes involved in the Hh pathway. X-axis: amino acid numbers in the protein multiple alignments, starting from the N-end. Y-axis: for the ADAPTSITE (black line), PAML (black dotted line), and PSWIN (grey line) programs, the number of adaptively evolving codons in a 30 aa window; for the PLATO program (grey dotted line), the Z-score for a 30 codons stretch, significantly deviating from random values. Below the X-axes, gene regions considered here to evolve under positive selection are indicated in red, protein domains are indicated in grayscale and labeled with numbers (see Table 1).
positive selection was based on the voting rule. If three of four programs detect this evolution mode in a 30-codon long stretch of a gene, its evolution is considered to have been driven by positive selection. In our view, the approach allows a more reliable detection of positive selection in gene sequences, being less sensitive to drawbacks of single methods. We ran the programs with default parameters for both in multiple cDNA alignments (gapless) and phylogenies. We chose the length of 30 codons as the approximate size of the functional protein domain.

Detection of Adaptive Selection on Phylogenetic Tree Branches

We detected positive selection on phylogenetic tree branches using an approach based on pairwise comparison of coding sequences without reconstruction of ancestral codons. As in the previous case, we used several methods, and the final result was obtained by voting. The methods used were: 1) NG86 (estimation of the $P_N/P_S$ ratio, the program DISTS2 (http://ftp.dna.affrc.go.jp/pub/unix/syn/)) [Nei and Gojobori, 1986; Ina, 1995], 2) YN00 (estimation of the $K_a/K_s$ ratio, the PAML package (http://abacus.gene.ucl.ac.uk/software/paml.html)) [Yang, 1997; Yang and Nielsen, 2000], 3) TW06 (estimation of the $K_a/K_s$ ratio, the program KAKH (available from C.-I. Wu)) [Tang and Wu, 2006], and 4) the logarithm of the WAG/NED4 ratio [Peltier et al., 2000], where NED4 was calculated as the prevalence of synonymous codons for amino acids A, G, L, P, R, S, T, and V with identical positions third of the codons, and WAG was the evolutionary distance evaluated from the WAG matrix with the TREE-PUZZLE program (http://www.tree-puzzle.de/) [Schmidt et al., 2002].

The statistical significance of the results obtained using these programs for long evolutionary distances typical of the considered proteins was additionally tested by simulation of sequence evolution. Simulation of the molecular evolution of genes (proteins) using the evolver program (Monte Carlo simulation of nucleotide or amino acid substitutions) allows to take into account the effect of saturation of substitutions in the deep branches on the outcome of methods used to search positive selection [Yang, 1997].

Multiple alignments of amino acid sequences were done with the evolver program of the PAML package (http://abacus.gene.ucl.ac.uk/software/paml.html) [Yang, 1997] on the basis of phylogenies reconstructed using the WAG amino acid replacement matrix and taking into account the prevalences of amino acids in the sequences. The parameter of the gamma distribution shape, $\alpha$, was taken to be unity, and the number of categories for the gamma distribution was 8. The coding nucleotide sequences were aligned on the basis of the resulting multiple alignments of amino acid sequences using their reverse translation to codons. We applied this method, because 1) direct generation of sequences at the codon level followed by their translation to amino acid sequences yielded considerable deviations of the amino acid composition of the simulated sequences from that observed in real proteins and 2) saturation of synonymous changes in the deep branches of the phylogenetic trees (use of the different synonymous codons tends to be random). For each family, 150 alignments of gene and protein sequences were generated. The positive selection measures, described above, were calculated for sequence pairs from simulated alignments to define boundaries of the 95% confidence and tolerance intervals in the assumed absence of positive selection. The estimates of positive selection measures for real sequence pairs were compared with the boundaries of the resulting confidence intervals to provide the basis for calculation of the significance index for each measure. This index took the values: 1 (observed within the tolerance interval but below the lower boundary of the confidence interval of mean values), 2 (observed within the confidence interval of mean values), 3 (observed within the tolerance interval but above the upper boundary of the confidence interval of mean values), and 4 (observed above the upper boundary of the tolerance interval of values). Positive selection was considered to be detected if the total of the indices
### Table 1
Functional domains of studied proteins

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Domain protein composition</th>
<th>Domain function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hh</td>
<td>1) N-terminal</td>
<td>Ptc receptor binding</td>
<td>Bijlsma et al., 2004</td>
</tr>
<tr>
<td></td>
<td>2) C-terminal, intein</td>
<td>Self-excision; binding of cholesterol to Hh N-terminal domain</td>
<td>Hall et al., 1997; Bijlsma et al., 2004</td>
</tr>
<tr>
<td>Disp</td>
<td>12 transmembrane loops</td>
<td>Membrane anchoring; possibly, secretion of the Hh determined by two extracellular domains, located between loops 1 and 2 (region adjacent to the N-end of domain 5) and loops 7 and 8 (region between domains 5 and 6)</td>
<td>Burke et al., 1999; Kuwabara and Labouesse, 2002</td>
</tr>
<tr>
<td></td>
<td>3) Sterol-sensing</td>
<td>Intracellular vesicular transport</td>
<td>Tseng et al., 1999</td>
</tr>
<tr>
<td></td>
<td>4) Exporters of RND superfamily (predicted)</td>
<td>Possibly, ion transport</td>
<td></td>
</tr>
<tr>
<td>Ptc</td>
<td>12 transmembrane loops</td>
<td>Membrane anchoring; the Hh contacts with two extracellular domains, located between loops 1 and 2 (region adjacent to the N-end of domain 5) and loops 7 and 8 (region between domains 5 and 6)</td>
<td>Fuse et al., 1999</td>
</tr>
<tr>
<td></td>
<td>5) Sterol-sensing</td>
<td>Intracellular vesicular transport; interaction with the Smo protein</td>
<td>Martin et al., 2001; Karpen et al., 2001</td>
</tr>
<tr>
<td></td>
<td>6) Exporters of RND superfamily (predicted)</td>
<td>Possibly, ion transport; Contact with Smo</td>
<td>Tseng et al., 1999; Karpen et al., 2001; Jia et al., 2004</td>
</tr>
<tr>
<td>Smo</td>
<td>7 transmembrane loops</td>
<td>Membrane anchoring</td>
<td>Jia et al., 2003; Lum et al., 2003; Ogden et al., 2003; Ho et al., 2005</td>
</tr>
<tr>
<td></td>
<td>7) N-terminal, extracellular</td>
<td>Possibly, binding to Ptc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8) G-receptors</td>
<td>G proteins binding</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9) C-terminal</td>
<td>Binding to the high-molecular-weight complex Fu/Cos2/Ci</td>
<td></td>
</tr>
<tr>
<td>PKA</td>
<td>10) Serine/threonine kinases, catalytic 11) WD-domain</td>
<td>Protein phosphorylation; Protein-protein interactions; binding to the SCF ubiquitin-ligase complex</td>
<td>Petretti and Prigent, 2005; Jackson and Eldridge, 2002; Cardozo and Pagano, 2004; Nakayama and Nakayama, 2005</td>
</tr>
<tr>
<td>Slmb</td>
<td>12) F-box</td>
<td>Protein phosphorylation</td>
<td>Murone et al., 2000; Ascano et al., 2002; Monnier et al., 2002; Ascano and Robbins, 2004</td>
</tr>
<tr>
<td>Fu</td>
<td>13) Protein kinase catalytic domain</td>
<td>Protein phosphorylation</td>
<td></td>
</tr>
<tr>
<td>C-terminal</td>
<td></td>
<td>Inhibition of Hh-signaling; interaction with Su(Fu) and Cos2</td>
<td>Dunaeva et al., 2003; Merchant et al., 2004; Barnfield et al., 2005; Monnier et al., 2002; Monnier et al., 2002; Manasse et al., 1996; Akimaru et al., 1997; Chen et al., 2000</td>
</tr>
<tr>
<td>Su(Fu)</td>
<td>14) SUFU-domain</td>
<td>Interaction with Ci</td>
<td></td>
</tr>
<tr>
<td>Ci (Gli)</td>
<td>15) and 17) Cytoplasm anchoring (at C-end and N-end from Zinc finger)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16) Zinc finger</td>
<td>DNA-binding</td>
<td>Wang and Jiang, 2004</td>
</tr>
<tr>
<td></td>
<td>18) Transcription cofactors binding</td>
<td>Binding to Ci transcription cofactors</td>
<td>Alexandre et al., 1996; Akimaru et al., 1997; Chen et al., 2000</td>
</tr>
</tbody>
</table>
Fig. 3. The phylogenetic tree of Hh gene family. The topology indicated by asterisk is in agreement with [Maddison and Schulz, 1996–2006; Bourlat et al., 2006]. Branches supposedly affected by adaptive evolution events are shown by thick black line. Genes analyzed are boxed. Vertebrate paralogs are indicated with circled numerals.

of the four methods exceeded 10. Positive selection for a tree branch was considered to be detected if positive selection was detected for all sequence pairs of the subtrees. In consideration of genes belonging to vertebrate species, priority was given to species of the classes Actinopterygii, Sarcopterygii, Amphibia, and Reptilia. We focused attention on the phylogenetic tree branches corresponding to the divergence stage of the major phyllum groups of Bilateria (insects and crustaceans – mollusks and annelids – chordates and echinoderms).

The detailed results of the various analysis steps (multiple alignments, phylogenetic trees, positive selection detection) are available from the authors upon request.

RESULTS

Detection of Protein Domains Evolving under Positive Selection

At the first step, we identified regions of genes of the Hh pathway evolving under positive selection. The results for nine genes are given in Fig. 2. We found that proteins involved in the Hh pathway and
their different domains are differently subject to positive selection (Fig. 2).

- In the Hh gene, positive selection occurs in the region coding for the intein domain, in agreement with previous studies [Kumar et al., 1996].
- In the Disp gene, evolution under positive selection manifests itself in the region coding for the N-terminal peptide, whose function is as yet unknown.
- In the Ptc gene, positive selection was detected for the region coding for the protein domains binding to the Hh morphogen.
- Positive selection in the Smo gene was detected in the regions coding for: (1) the domain binding to the high-molecular-weight Fu/Cos2/Ci complex, (2) the extracellular domain, and (3) the domain binding to G proteins.
- In the PKA gene and its homologs, the effect of positive selection is observed in small region (~15 codons) coding for the N-terminal peptide.
- No positive selection was found for the Slmb gene.
- In the Fu gene, positive selection affects the region coding for the C-terminal peptide occurring outside the catalytic domain of the Fu protein.
- In the Su(Fu) gene, positive selection affects the region coding for the middle portion of the SU(FU) domain.
- In the Ci gene and its homologs Gli, the effect of positive selection is observed in regions coding for domains responsible for retaining Ci(Gli) in the cytoplasm and for binding to transcription cofactors.

Correlation between protein regions evolving under positive selection and their function indicates that positive selection occurred mainly for proteins and their domains involved in protein-protein interactions within the Hh signaling pathway, whereas proteins of “universal” functions (PKA and Slmb) were under negative selection. Neither was positive selection found for the catalytic domains of the proteins under study; this may indicate that this function was probably conserved during the evolution of the Hh pathway.

Detection of Phylogenetic Tree Branches Evolving under Positive Selection

The results of the analysis performed to detect phylogenetic branches presumably evolving under positive selection are shown in Figs 3–11 and in Supplementary Information. They indicate that positive selection unevenly affected the genes and proteins under study.

No positive selection was found on branches of the phylogenetic trees for the Ptc, PKA, and Slmb genes. This result was expected for PKA and Slmb, because we found no regions evolving under positive selection. In contrast, such regions were found in Ptc, but their analysis did not allow definite localization of positive selection stages on the phylogenetic tree, probably, because of the very weak positive selection signal observed in proteins of this family.

The stages of positive selection in the Hh family genes correspond to the divergence of the genera Artemia, Patella, Ciona, Lytechinus and Strongylocentrotus, and, possibly, Branchiostoma (appearance of arthropods and vertebrates) and to the formation of paralogous groups of the Hh gene in vertebrates, in agreement with previous studies [Kumar et al., 1996; Shimeld, 1999]. In the Disp genes, positive selection was detected for the tree branches passing from the genera of the class Insecta to those of the Craniata group. This corresponds also to the appearance times of arthropods and vertebrates. In the Smo gene family, positive selection is typical of branches passing from Insecta to the genera Danio, Tetraodon, and Fugu, corresponding to the appearance times of arthropods and vertebrates. According to our analysis, for the Su(Fu) gene family positive selection manifests on tree branches passing from the genera of class Insecta to those of the Craniata group. As for the above families, this corresponds to the
Table 2
Correlation between adaptive modes of gene evolution and appearance of major animal taxa

<table>
<thead>
<tr>
<th>Gene</th>
<th>Functional group (developmental gene or housekeeping gene)</th>
<th>Events of positive selection in genes correlated with appearance of major animal taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hh</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
<tr>
<td>Ptc</td>
<td>Developmental gene</td>
<td>−</td>
</tr>
<tr>
<td>Smo</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
<tr>
<td>Disp</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
<tr>
<td>Cos2</td>
<td>Developmental gene</td>
<td>+ *</td>
</tr>
<tr>
<td>PKA</td>
<td>Housekeeping gene</td>
<td>−</td>
</tr>
<tr>
<td>Slmib</td>
<td>Housekeeping gene</td>
<td>−</td>
</tr>
<tr>
<td>Su(Fu)</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
<tr>
<td>Fu</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
<tr>
<td>Ci</td>
<td>Developmental gene</td>
<td>+</td>
</tr>
</tbody>
</table>

Designations: “+”, gene families found to evolve in the adaptive mode; “−”, gene families found to evolve under negative selection; “*”, absence of functional homologous proteins from insects and vertebrates, low degree of similarity in vertebrate proteins.

Fig. 4. The phylogenetic tree of Disp gene family coding for transmembrane proteins. Designations as in Fig. 3.

appearance of arthropods and vertebrates. At the same evolutionary stages, positive selection affected genes of the Fu family: positive selection corresponds to the branch passing from the genus Danio and Gallus to the insect genera, corresponding to the appearance times of arthropods and vertebrates. For the genes of the Ci(Gli) family, positive selection was detected for the branch passing from the
Fig. 5. The phylogenetic tree of $Ptc$ gene family coding for transmembrane proteins. Designations as in Fig. 3. Paralogs of genes are indicated with circled numerals.

genus Branchiostoma to the genera of the class Insecta and the group Tetrapoda, corresponding to the appearance stage of major taxa of bilateral organisms (arthropods and vertebrates).

Thus, we conclude that the main stages of positive selection in the evolution of the Hh-pathway genes corresponded to the divergence of main phylum groups of Bilateria: Ecdysozoa (insects and crustaceans), Lophotrochozoa (mollusks and annelids), and Deuterostomia (chordates and echinoderms). This indicates that the evolutionary stages under positive selection and the appearance of major animal taxa are correlated (Table 2).

**DISCUSSION**

**Rearrangement of the Hh-Cascade Structure During Evolution and Positive Selection**

We have demonstrated that certain gene regions that encode the domains of proteins whose function may be specific precisely to developmental processes (“the developmental genes”) are under positive selection. It is of importance that these domains are involved in the internal interactions of the Hh-cascade...
proteins, either self-processing and interacting with their own domains, like in Hh, or with other proteins of the cascade, such as Ptc, Smo, Fu, Su, Ci and, presumably, Disp. As to the developmental genes of the Hh-cascade, the time span when positive selection acted matches with the divergence of the two main bilaterian groups, Ecdysozoa and Deuterostomia. This indicates that amino acid substitutions were rapidly accumulating at the time of the fundamental structural changes in the Hh-cascade. In contrast, positive selection is not characteristic of the PKA and Slmb genes that encode the proteins involved in the interaction with the other signaling pathways (phosphorylation and protein degradation processes) or the catalytic domains of the protein under study. These components perform possibly a constant function that provides the vital activities of the cells. Therefore, PKA and Slmb evolved under negative selection. The different selection pressures acted upon Hh-cascade genes can be explained by the structural changes in the Hh-signaling pathway via compensatory changes in the mechanisms of protein-protein interactions.

These changes in the cascade were possibly compensated by alterations in the dynamics of the morphogen production or sensing of the morphogen. For example, the rate and the remoteness of the response to the morphogen may be altered by modification of the mobility of the Hh morphogen or of the morphogen receptors (Ptc and Smo) [Nybakken and Perrimon, 2002]. It is of interest that in the Hh protein positive selection occurred in the intein domain controlling the excision of the Hh part, which is a morphogen. It may be suggested that the spreading rate of the morphogen (and accordingly the distance
it acts) may be controlled by its excision efficiency from the zymogen. If so, the positive selection in the excision-controlling domain would provide the reorganization of the Hh-cascade functioning by the appropriate changes in the kinetics of the yield and spread of the active Hh form.

The positive selection in the proteins of the middle part of the Hh-cascade might have ensured the reorganization of the cell response to the morphogen. It is known that the different Bilateria types differ in the composition of proteins involved in the work of the middle part of the Hh-cascade. To illustrate, the N-end part of the Drosophila Smo protein is required for transduction of Hh-signal, while in vertebrates its deletion does not produce significant changes in the Hh-cascade [Nakano et al., 2004].
In invertebrates, Cos2 and Fu proteins form a functional complex that controls the penetration of Ci transcription factor into the nucleus [Jia et al., 2003; Lum et al., 2003; Ruel et al., 2003; Ascano and Robbins, 2004]. However, the experimental evidence for the Cos2 presence in vertebrates is indirect only [Tay et al., 2005].

Positive selection in transcription factors Ci(Gli) may provide changes in the cell response type (for example, passage from the unstable-stable response) to the Hh morphogen. Bilateria taxonomic types differ greatly in the composition and function of the Ci(Gli) transcription factors [Huangfu and Anderson, 2006]. In contrast to invertebrates, vertebrates have three Gli proteins that activate the transcription of the Hh-cascade effector genes, (Gli1, Gli2 and unprocessed Gli3) and two proteins inhibiting the transcription of Hh-cascade effector genes (modified Gli2 and processed Gli3); activating Gli enhances the expression of not only the effector genes, but also the Gli1 and Gli2 genes, thereby forming the positive feedback circuit nonexistent in invertebrates [Dai et al., 1999; Aza-Blanc et al., 2000; Ikram et al., 2004; Tyurina et al., 2005]. In vertebrates transcription factor Gli3 is probably the most important of the Gli proteins because only this one, in the absence of active Smo, can be subject to proteolysis like Ci vertebrate protein [Müller and Basler, 2000; Aza-Blanc et al., 2000; Tyurina et al., 2005]. Thus, vertebrates have additional feedback circuits that establish a definite expression level of the Hh-cascade effector genes. The presence of the transcription factors Gli1 and Gli2 in vertebrates broadens the range of genes whose expression can regulate the Hh-cascade, thereby affording more opportunities of obtaining specific responses of
tissues to the same morphogen [Huangfu and Anderson, 2006].

Thus, this comparative analysis of the structure of the Hh-cascade proteins (or their functional domains) in different Bilateria suggests that the drastic changes in the Hh-cascade pathway occurred at the time of the divergence of the main groups of Bilateria taxonomic types (Ecdysozoa – Deuterostomia divergence).

**Evolution of the Molecular-Genetic Systems that Define the Morphogenesis and the Evolution of Hh-Signaling Cascade Genes**

During embryogenesis of the primitive animals (sponge, coelenterates) the adult tissue differentiation is determined immediately after fertilization and, consequently, fertilization directly determined adult morphogenesis [Davidson *et al.*, 1995; Peterson *et al.*, 2000]. This development type is the most primitive and supposedly the ancestral multicellular animals possessed it [Salazar-Ciudad *et al.*, 2003; Salazar-Ciudad and Jernvall, 2004]. In contrast, Bilateria form embryonic tissue types missing later in the adult. Bilateria embryonic tissues are capable, as a rule, to produce morphogens that are capable of giving rise to other tissues capable, in turn, of also producing morphogens [Kolchanov *et al.*, 2004; Suslov *et al.*, 2004; Gunbin *et al.*, 2004b; 2004c]. The ultimate result is that changes in the growth parameters of embryonic tissues, no matter how small, can generate tissues of new types, change in morphogenetic patterns, novel adult traits. Thus, Bilateria use a novel evolutionary strategy (in comparison with the primitive animals) improving their chances of coping with their environment.

This evolutionary strategy of Bilateria has been studied in detail in the exemplary case of the rodent digestive system by Vorontsov, 1967. He has shown that the different organs of the rodent digestive system, as a rule, are at different specialization levels in response to the same environmental agents.
Specialization of one organ of the system may be not related to the specialization of another organ. If so, one organ of the system may compensate the function of another. More than that, the different specialization levels of the tissues and organs of an animal organism and the compensation of the function of a tissue/organ by another may be universal for the Bilateria evolution [Hartman et al., 2001]. In this way, the uneven rates of the evolutionary reorganization of tissues and organs and the compensation of the function of a tissue/organ by another may provide the rapid and parsimonious strategy of the adaptability of bilaterian species. In the genes that are expressed in one tissue the fixation of nonsynonymous mutation is 3–5-fold greater than in the genes that are expressed in at least 16 tissues [Duret and Mouchiroud, 2000]. Therefore, the uneven rate of evolutionary reorganization is also a feature of the bilaterian genes that carry different functional loads. It would appear that the genes for tissue differentiation were expressed in a restricted tissues set during the emergence of Bilateria [Sole et al., 2003; Salazar-Ciudad et al., 2003; Salazar-Ciudad and Jernvall, 2004]. Clearly, the uneven rates of the evolutionary reorganization of tissues and the compensation of a tissue function by another could provide: 1) parsimonious and rapid adaptability of Bilateria species to various environmental agents and 2) rapid emergence of wide diversity of Bilateria, including the appearance of various Bilateria body plans (the appearance of Bilateria taxonomic types). Presumably, the appearance of precisely this novel, in principle, evolutionary strategy was the cause of the Cambrian explosion of Eumetazoa and especially of Bilateria.

However, with the increasing morphological complexity, the number of developmental stages and embryonic tissues increased and a need for an evolutionary stagnation of many molecular-genetic systems arose. For these reasons, the adaptive evolution (positive selection) of the genes involved in the molecular-genetic developmental systems derived from other fundamental molecular machineries halted. This is
what we observed. Presumably, for these reasons, the adaptive evolution of the signaling pathway genes appears to be suspended in our days.

CONCLUSION

Here, we searched and analyzed the selection mode in gene families involved in the work of the Hh-signaling cascade. We used a number of methods to detect the positive selection mode both in gene regions and on phylogenetic tree branches. The combined computer-assisted approach allowed us to increase reliability as compared with the case when each method is used alone. The resulting data were related to those on the structural-functional organization of the proteins under analysis and also related to their interactions in the functioning Hh-cascade. The results demonstrated that the function of the analyzed proteins was subject to positive selection during the short span, in geological terms, corresponding to the stages when large bilaterian taxa (Ecdysozoa, Lophotrochozoa and Deuterostomia) formed. This was manifested as an increase in the accumulation rates of amino acid substitutions in many protein domains that control the specific interactions within the signaling cascade. The results suggested that changes in substitution accumulation are causes for the structural reorganization of the Hh-cascade at the formation steps of Ecdysozoa and Deuterostomia. Their compensatory nature is presumably dual: changes in the dynamics of the Hh-cascade function due to adaptive substitutions in the Hh, Ptc and Smo proteins and/or changes in the protein domains that are responsible for the protein-protein interactions.
within the intracellular part of the Hh-signaling pathway and related to the structural reorganization of these interactions.

ACKNOWLEDGEMENTS

The work has been supported by the Ministry of Education of the Russian Federation grant “Development of the Higher School Scientific Potential” 2.1.1.4935, Russian Foundation of the Basic Research (05-04-49141-a, 05-07-98012-p), SB RAS integration projects 49, No. 10104-34/II-18/155-270/1105-06-001/28/2006-1. The computation was performed in part at the High Performance Computing Center, SB RAS.

REFERENCES

- Nei, M., Kumar, S. and Takahashi, K. (1998). The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proc. Natl. Acad. Sci. USA 95, 12390-12397.


