

Rare Coding Sequence Changes are Consistent with Ecdysozoa, not Coelomata

Manuel Irimia,*† Ignacio Maeso,† David Penny,‡ Jordi Garcia-Fernàndez,† and Scott William Roy*‡§

†Departament de Genètica, Universitat de Barcelona, Barcelona, Spain; ‡Allan Wilson Centre for Molecular Evolution and Ecology, Massey University, Palmerston North, New Zealand; and §National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland

There is growing interest in the use of alternative, more slowly-evolving RGCs (rare genomic changes). Recently, Rogozin and coauthors (Rogozin *et al.* 2007) proposed a novel phylogenetic method employing rare amino acid changes, RGC-CAMs (rare genomic changes-conserved amino acids-multiple substitutions). They applied their method to 694 sets of eukaryotic orthologs in order to distinguish the relationship between nematodes, arthropods and deuterostomes. They concluded that such rare amino acid changes were consistent with the Coelomata hypothesis, which groups arthropods and deuterostomes to the exclusion of nematodes. Here we use newly available genomic sequences from *Nematostella vectensis*, a basal metazoan, and from *Brugia malayi*, an additional nematode. We show that the apparent support for Coelomata is likely to be the result of the rapid rate of evolution leading to *Caenorhabditis* nematodes. Including the additional species paints a very different picture, with 13 remaining characters consistent with Ecdysozoa versus only 1 consistent with Coelomata.

Recently Rogozin and coauthors (2007) studied protein sequences from 694 sets of eukaryotic orthologs from 10 species to determine the evolutionary relationship between deuterostomes (represented by mouse and human), arthropods (represented by *Drosophila melanogaster* and *Anopheles gambiae*), and nematodes (represented by *Caenorhabditis elegans* and *briggsae*). They studied highly conserved amino acid positions, defined as positions in ungapped regions of alignment at which the 4 studied non-animals (the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, the plant *Arabidopsis thaliana* and the apicomplexan *Plasmodium falciparum*) shared the same amino acid. As these positions are potentially slowly evolving, they are promising phylogenetic characters (Philippe *et al.* 2000). They found 76 such positions for which (i) members of one of the animal groups shared the non-animal amino acid; (ii) members of the other 2 groups shared a different amino acid, consistent with a single change in a putative ancestor; and (iii) the 2 observed amino acids were separated by multiple nucleotide changes, decreasing the chance of homoplasy. This scenario is illustrated in figure 1A and (in simplified form) in figure 1B. These new characters join a host of alternative genomic characters which may retain phylogenetic signal for deep divergences for which traditional sequence data may be saturated (Rokas and Holland 2000; Gugerli *et al.* 2001; Henz *et al.* 2005; Roy and Gilbert 2005; Boore 2006; Kriegs *et al.* 2006).

The authors found similar numbers of positions supporting 2 alternative relationships: 34 positions supported grouping arthropods and deuterostomes (consistent with the “Coelomata” hypothesis; Field *et al.* 1988; Wolf *et al.* 2004; Philip *et al.* 2005), while 26 supported grouping arthropods with nematodes (consistent with “Ecdysozoa”; Aguinaldo *et al.* 1997; Philippe *et al.* 2005. Note that while the previous and present studies test the more general hy-

pothesis of a clade including nematodes and arthropods but not deuterostomes, we nonetheless use the more specific term “Ecdysozoa” for simplicity. The same is true for “Coelomata”). These numbers are identical at a $P = 0.2$ level by a chi-square homogeneity test. However, they argued that due to the very long external branch leading to *Caenorhabditis*, homoplastic forward mutations (fig. 1C) are more likely to artificially support Ecdysozoa (double changes in nematodes and arthropods) than Coelomata (double changes in deuterostomes and arthropods); thus the authors concluded that the data strongly supported Coelomata.

However, another important source of homoplasy is changes in an ancestor of metazoans or bilaterians (at any point along the long branch leading from the fungus-animal ancestor) and backmutation in one of the 3 bilaterian branches (fig. 1D). Due to the longer nematode branch, such mutations are more likely to lend artificial support to Coelomata (backmutation in nematodes) than to Ecdysozoa (backmutation in deuterostomes).

To probe this possibility, we studied additional animal species of key phylogenetic relevance. First, we studied genomic sequence from a basal animal, the cnidarian *Nematostella vectensis*. If a position truly reflects a single change in the ancestor of 2 bilaterian groups, *N. vectensis* would be expected to have the putatively ancestral (non-animal) amino acid (fig. 2A). If instead a position reflects a change before the animal ancestor and a subsequent backmutation, *N. vectensis* should exhibit the other amino acid (fig. 2B).

The results are very different for Ecdysozoa and Coelomata (fig. 3). Among the 34 positions previously supporting Coelomata, more than half (19) are consistent with backmutation events (i.e. *N. vectensis* matches deuterostomes/arthropods; fig. 2B). In only one-fifth of cases (7/34) did *N. vectensis* exhibit the supposedly ancestral animal amino acid (i.e. matching nematodes; fig. 2A). In 7 other cases *N. vectensis* exhibited a third amino acid, and in the final case, *N. vectensis* had a 2 amino acid indel at the position. For positions supporting Ecdysozoa, the situation was the reverse: 20/26 positions were consistent with single mutation in a nematode-arthropod ancestor (*N. vectensis* matches deuterostomes), while only one position was consistent with backmutation (*N. vectensis* matches nematodes/arthropods). The remaining 5 cases showed an ambiguous

*These authors contributed equally to this work.

Key words: Ecdysozoa, Coelomata, phylogeny, long branches, rare genomic changes.

E-mail: royscott@ncbi.nlm.nih.gov.

Mol. Biol. Evol. 24(8):1604–1607. 2007

doi:10.1093/molbev/msm105

Advance Access publication May 24, 2007

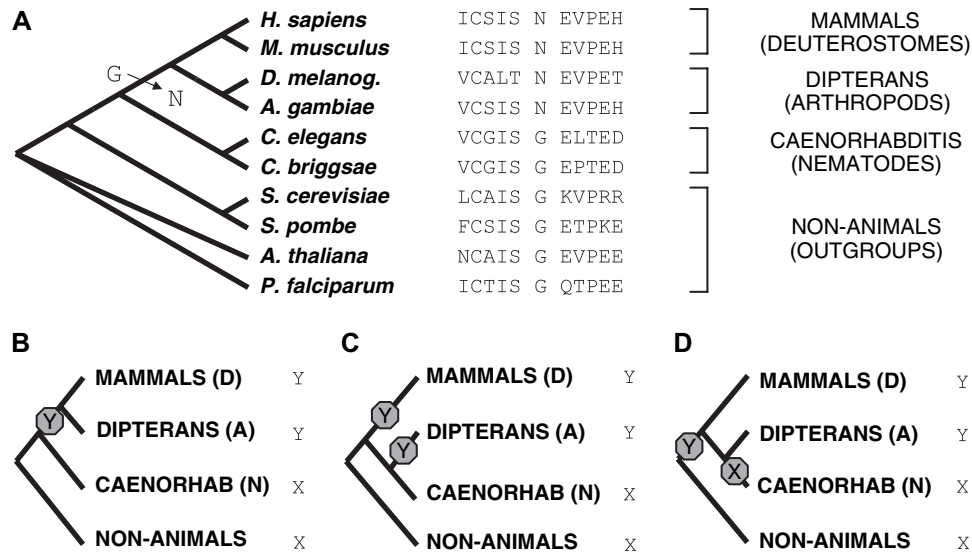


FIG. 1.—Rare amino acid changes as phylogenetic characters. (A) An example of a rare amino acid change supporting Coelomata, from Rogozin et al. (2007). The phylogenetic pattern for the middle position in the alignment (G/N) is consistent with a single change from G to N in a putative ancestor of deuterostomes and arthropods (the Coelomata hypothesis). (B) General schematic for such an explanation, with a single change from an amino acid X-to-Y (Y hexagon). (C) Alternative explanation assuming alternative Ecdysozoa topology: double change from X-to-Y in both deuterostomes and arthropods, as discussed by Rogozin et al. (2007). (D) Alternative explanation assuming alternative Ecdysozoa topology: a forward X-to-Y change in the animal ancestor and a backmutation in nematodes.

alignment or a third amino acid. The case for the third possible topology, called ‘Bizarre’ by Rogozin and coauthors (Rogozin et al. 2007), was intermediate between the other 2 alternatives, with 10/16 characters consistent with a single forward mutation and 5 consistent with backmutation (and 1 showing a third amino acid).

Studying additional bilaterian species further clarified the picture (fig. 3). Among the remaining 7 positions supporting Coelomata, in 4 cases either the arthropod

Apis mellifera or the deuterostome *Strongylocentrotus purpuratus* exhibited the ‘wrong’, i.e. putatively ancestral (non-animal/*N. vectensis*/nematode) amino acid, consistent with late-occurring (convergent) changes having occurred within mammals and dipterans (fig. 2C, D). In 2 other cases, the nematode *Brugia malayi* matches deuterostomes and arthropods, consistent with a single change between the animal and bilaterian ancestor and a backmutation in *Caenorhabditis* after the *Caenorhabditis-Brugia* divergence

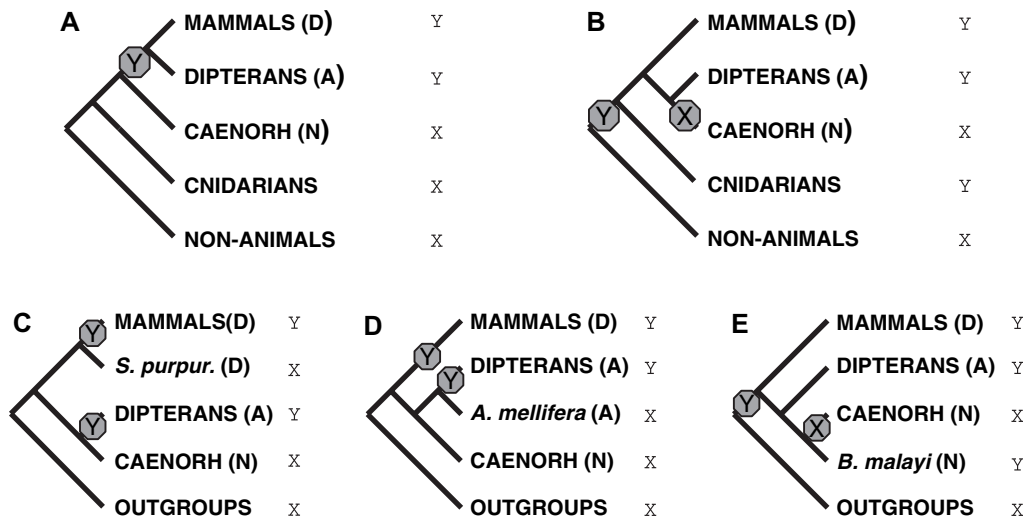


FIG. 2.—Different phylogenetic patterns for positions that previously supported Coelomata, after including additional animal species. First, we added the cnidarian *N. vectensis*. (A) Cnidarians share the non-animal (putatively ancestral) amino acid, consistent with a single X-to-Y mutation in a putative deuterostome-arthropod ancestor. (B) Cnidarians match deuterostomes/arthropods, consistent with an X-to-Y mutation in the animal ancestor and a backmutation in nematodes. Next, we added additional bilaterian species. (C) In some cases, the deuterostome *S. purpuratus* matches non-animals, consistent with double forward mutation (X-to-Y) in mammals and arthropods. (D) In some cases, the arthropod *A. mellifera* matches non-animals, consistent with double forward mutation in deuterostomes and dipterans. (E) In some cases, the nematode *B. malayi* matches deuterostomes/arthropods, consistent with backmutation in *Caenorhabditis*.

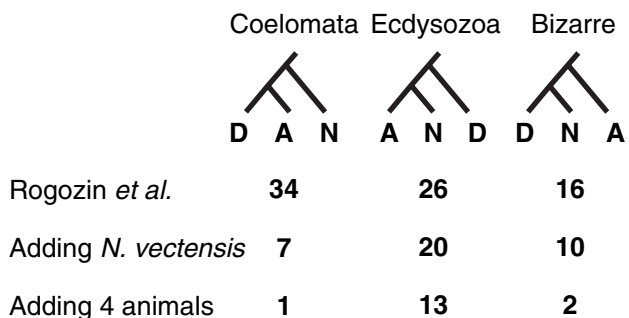


FIG. 3.—Summary of data. Number of RGC-CAM characters supporting the 3 phylogenies given the original data (Rogozin et al. 2007), adding the cnidarian *N. vectensis*, and adding all 4 additional animal species.

(fig. 2E). In total, then, only a single position remained consistent with a single change assuming the Coelomata phylogeny.

More support remains for Ecdysozoa. Only 7/20 of the characters supporting Ecdysozoa were similarly eliminated

by any of the 3 additional bilaterians. Thus, 13 characters continued to support Ecdysozoa, versus only 1 for Coelomata ($P = 0.0009$ level by a binomial distribution), and 2 for ‘Bizarre’. Thus, while the number of remaining characters is small the remaining data supports Ecdysozoa, not Coelomata.

How good are these rare amino acid changes as phylogenetic characters? The positions used here do seem to be potentially slowly evolving: of 41,451 total alignment positions for which all 4 non-animals have the same amino acid, the same amino acid is also conserved across all 6 animals in 79.5% of cases. Among the 6470 that are not universally conserved, 76.0% show only 2 total amino acids across the ten species. Thus less than 5% of positions that are conserved across non-animals show more than 2 total amino acids; for only 3.2% of positions that are conserved among non-animals, any of the 6 animals have a qualifying amino acid. Thus the chance for multiple confounding changes may be low.

Nonetheless, the present results underscore that multiple changes do occur—indeed the vast majority of

KOG0188

```

Mmu AIEELR-AKGLEATDDSPKYNY-QSDSSGS---YVFECTVATVLAALRREKMFVDEVVTGQ
Hsa AIEELR-ARGLEVTDDSPKYNY-HLDSSGS---YVFENTVATVMALRREKMFVEEVSTGQ
Spu AINELQEKRRVPPPTNDSAKYKY-ESND-GS---YTFESTVGTVIGLRKDRQFVEEVNHGE
Ame AITELQ-NEKIKPTNDIPKYNY-KVISNKIYEEYEFAPCFSTIIALRRAKTFVDEVSSE
Dme AISQLQ-EQGVPPNTDFKYKY-EAVSDERDSAYNYGVCNSKIIVALRFENQFVNEITSGQ
Aga GISELQ-ERKVPATDDSFKYRY-KAESIDPLAQYVFEPCTGKIIVALRFNNAFVEEVQAGQ
Bma AISELK-NKGIPTDDSPKYKYALSSSEQTLYNFEKCEGIILALRKDKIFLDTLNSGD
Cel ALAELQ-QKGVPTTDDSPKYAY-TFTGEGSDAVYKFEPCVKGKILAIRRDKGFVDQLAAGE
Cbr ALAELQ-QKGIPTTDDSPKYAY-SFTGEGSEAVYKFEPCVKGKILAIRRDKGFVDQLVAGE
Nve LLDKLNQVMNLPPTIDLPKYSY-SSDALGS---YVFQPVSGTIKALLHEKEFVNEVPGGE
: : * : * * * * * * : . : : * : : * :
    
```

KOG0712

```

Mmu CVLNEGMPYIRRPYEKGRLLIEFKVNFPENGFLSPDKLSLLEKLLPERKEVEETDEMDOV
Hsa CVLNEGMPYIRRPYEKGRLLIEFKVNFPENGFLSPDKLSLLEKLLPERKEVEETDEMDOV
Spu MVVGEGLMPYKKNPFERGRLLIQFQINFPENNAIQEKNLEKLEAIMPAREDCIVTDDMEMV
Ame CILNEGMPYKDPFTHGRLLIQFVNFNPKS--MDPSVIPTLEQCLPPREVEIPEGAEDC
Dme CIAEEGMPYKKNPMEKGTLLIQFEVIFPEV--INPSVVPVTLKQCLPPAPEVDIPIDAEQT
Aga CVYGEGLMPLMNDPTEKGRLLIQFVVGFPDS--LPPEVVPEIRKYLPPTQPDPIPEDHETV
Bma TIIIGEGMPHYKKNPFDKGDLIIQFAVRFPPK--IM--EVEQLKNLLPNGTEPLVSDDAEVV
Cel VIHNEGMPMRRASSDKGDLVQFDVKFPDK--INPDAAKKLADLLPGKREEIIDEDAQV
Cbr VIHNEGMPMRRAPSDRGDLVQFDVKFPDK--INPDAAKKLADLLPGKREEIIDEDADV
Nve SIEDEGMPHRRNPFHKGRLLIQFVQVFPENGVLNPKNMDKLEKLLPPRPEIIPDETEDV
: * * * . . : * * : * : * : * : * : * : * : * : * : * :
    
```

KOG1057

```

Mmu EYGITKAEKLEIAKGYCTPLVRKIRSDLQR----TQDDDTVNKLHPVYSRGLVSPERHVR
Hsa EYGITKAEKLEIAKGYCTPLVRKIRSDLQR----TQDDDTVNKLHPVYSRGLVSPERHVR
Spu EYGITTEKLEIGKGICTPLLRKIRSDLHRI---NSNDETTYRLNPLYSGVMSPDRQVR
Ame EYGLTVQEKLTIGQICTPLLRKIRADLQRNIEESGE-ETVNRNLNPRYSHGVSSPGRHVR
Dme EYGLTPQEKLAIGQGICSPLLRKKIKGDLQRNIDEVED-EFMNRLNPHYSHGVASPRHVR
Aga EYGLTMHEKLTIGQICTPLLRKIRADLQRNIEELGGEESVNRLNPRYSHGVSSPGRHVR
Bma EYGISSENSKVIIGQHVCTPLLRKIKSDLYHCVENPNEDDTQTRLDPRASQGIATPFRHVR
Cel EYGIKTENKMVIAQRVCTPLLRKIRNDLHRCLLENKESEETQTRLDPRASQGIATPFRHVR
Cbr EYGIKTENKMVIAQRVCTPLLRKIRNDLHRCLLENKESEETQTRLDPRASQGIATPFRHVR
Nve EYGLSAEEKVIAKMKCIRLLRKKIQGDLKH----ADTEDTHRLNPEYSQSVITPFRHVR
* * * : . * : * : * * : * : * : * : * : * : * : * : * : * : * :
    
```

FIG. 4.—Three indel positions lending support to Ecdysozoa. In the 694 groups of orthologs used in this study, only 3 clear cases of lineage-specific multiple amino acid indels were found. All 3 are consistent with Ecdysozoa. The KOGs represented in the figure correspond to the following proteins: (i) KOG0188: Alanyl-tRNA synthetase (corresponding to amino acid position 496 in the human sequence), (ii) KOG0712: Molecular chaperone (DnaJ superfamily) (position 331), and (iii) KOG1057: Arp2/3 complex-interacting protein VIP1/Asp1 (position 792).

characters supporting Coelomata turned out to involve homoplasy. In general, the 4-fold reduction in the number of characters (from 76 to 16) with the addition of only 4 additional taxa is troublesome—clearly many of these sites have experienced multiple changes within animals. Backmutation might pose a particular challenge—given the ancestral amino acid's widespread conservation, backmutation might be quite common. Further, under covarion behavior, ancestrally slow-evolving sites may not be so in animals (Fitch and Markowitz 1970; Lockhart et al. 1998; Penny et al. 1998; Lockhart and Steel 2005).

Finally, we investigated a second type of rare genomic change: multiple amino acid indels. Across the 694 sets of orthologs, we identified only 3 such clear indels (2 insertions and 1 deletion) in which all 6 members of 2 of the animal groups had an indel relative to members of the other group as well as *N. vectensis* (fig. 4). All 3 cases were consistent with Ecdysozoa.

These results show that, with additional data, the method developed by Rogozin *et al.* provides support for Ecdysozoa, and not Coelomata, consistent with the notion that the previous Coelomata signal is likely to have resulted from long branch artifacts (Aguinaldo et al. 1997; Dopazo and Dopazo 2005; Philippe et al. 2005; Delsuc et al. 2006; Baurain et al. 2007; Lartillot et al. 2007).

Methods

We downloaded the original data from the authors' website (ftp://ftp.ncbi.nlm.nih.gov/pub/koonin/RGC_CAM/). Reciprocal BLASTP searches between the 694 orthologous genes were performed for *N. vectensis* (using data downloaded from the genome project (<http://genome.jgi-psf.org/Nemvel/Nemvel.home.html>) and for *Brugia malayi* (performed on the genome project website: <http://tigrblast.tigr.org/er-blast/index.cgi?project=bma1>)). We first verified the numbers and identity of the positions identified by Rogozin et al. (i.e. the 76 positions) using custom Perl programs. We then performed alignments for all 14 species using ClustalW with default parameters, and analyzed the site patterns for each of the 76 positions using both custom Perl programs and manual inspection. Each position was checked multiple times by multiple different authors. Indels were manually searched over alignments for the 10 animal species, using the 694 groups of orthologs used by Rogozin et al. (2007).

Literature Cited

Aguinaldo AMA, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature*. 387:489–493.

Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: closely spaced Cladogeneses or undetected systematic errors? *Mol Biol Evol*. 24:6–9.

Boore JL. 2006. The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol Evol*. 21:439–446.

Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*. 439:965–968.

Dopazo H, Dopazo J. 2005. Genome-scale evidence of the nematode-arthropod clade. *Genome Biol*. 6:R41.

Field K, Olsen G, Lane D, Giovannoni S, Ghiselin M, Raff E, Pace N, Raff R. 1988. Molecular phylogeny of the animal kingdom. *Science*. 239:748–753.

Fitch W, Markowitz E. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. *Biochem Genet*. 4:579–593.

Gugerli F, Sperisen C, Büchler U, Brunner I, Brodbeck S, Palmer JD, Qiu Y-L. 2001. The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. *Molec Phylogeny Evol*. 21:167–175.

Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC. 2005. Whole-genome prokaryotic phylogeny. *Bioinform*. 21:2329–2335.

Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius Jr, Schmitz Jr. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol*. 4:e91.

Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol*. 7:S4.

Lockhart P, Steel M. 2005. A tale of two processes. *Syst Biol*. 54:948–951.

Lockhart P, Steel M, Barbrook A, Huson D, Charleston M, Howe C. 1998. A covarion model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol Biol Evol*. 15:1183–1188.

Penny D, Murray-McIntosh R, Hendy M. 1998. Estimating times of divergence with a change of rate: the orangutan/African ape divergence. *Mol Biol Evol*. 15:608–610.

Philip GK, Creevey CJ, McInerney JO. 2005. The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. *Mol Biol Evol*. 22:1175–1184.

Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, Protostomia. *Mol Biol Evol*. 22:1246–1253.

Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Müller M, Le Guyader H. 2000. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc Biol Sci*. 267:1213–1221.

Rogozin IB, Wolf YI, Carmel L, Koonin EV. 2007. Ecdysozoan clade rejected by genome-wide analysis of rare amino acid replacements. *Mol Biol Evol*. 24:1080–1090.

Rokas A, Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol*. 15:454–459.

Roy SW, Gilbert W. 2005. Resolution of a deep animal divergence by the pattern of intron conservation. *Proc Natl Acad Sci USA*. 102:4403–4408.

Wolf YI, Rogozin IB, Koonin EV. 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res*. 14:29–36.

Herve Philippe, Associate Editor

Accepted May 21, 2007