Rare Genomic Characters Do Not Support Coelomata: Intron Loss/Gain

Scott William Roy* and Manuel Irimia†

*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD; and †Departament de Genètica, Universitat de Barcelona, Barcelona, Spain

Recently, a new phylogenetic method employing intron position sharing across species was proposed and support for a Coelomata clade reported (Zheng et al. 2007). A rigorous analysis of the pattern of intron conservation supports the Coelomata clade of animals, Mol Biol Evol. 24:2583–2592.). Here, we show that the previous analysis depends on: 1) an idiosyncratic definition of “conserved” introns, 2) exclusion of all phylogenetically informative introns present in outgroups, 3) incorrect inference of change along the critical branch, and 4) lack of variation in rates across branches. The method thus seems unlikely to give accurate results. In addition, we address differences in rates of loss across intron sites, which Zheng et al. claimed invalidates our previous analysis that supported Ecdysozoa (Roy and Gilbert. 2005a). Resolution of a deep animal divergence by the pattern of intron conservation. Proc Natl Acad Sci USA. 102:4403–4408.). Instead, we show that our conclusions are likely to be robust to such concerns.

The phylogenetic position of nematodes continues to be debated. Many studies have concluded support for the grouping of nematodes and arthropods, consistent with the Ecdysozoa hypothesis (Aguinaldo et al. 1997; Dopazo H and Dopazo J 2005; Philippe et al. 2005; Roy and Gilbert 2005a; Delsuc et al. 2006; Irimia et al. 2007). Other studies have instead suggested the grouping of arthropods and deuterostomes to the exclusion of nematodes (consistent with the Coelomata hypothesis; Field et al. 1988; Wolf et al. 2004; Brinkmann et al. 2005; Philip et al. 2005; Rogozin et al. 2007a, 2007b). In several cases, support for Coelomata may reflect long-branch attraction and/or insufficient taxon sampling (Philippe et al. 2004; Baurain et al. 2007; Irimia et al. 2007; Lartillot et al. 2007).

Recently Zheng et al. (2007) studied splicesomal intron positions across 11 eukaryotic genomes (fig. 1). For the studied species, the implications of the Ecdysozoa/Coelomata distinction reduce to whether arthropods (represented by Anopheles gambiae and Drosophila melanogaster) group with deuterostomes (represented by humans) or nematodes (represented by Caenorhabditis elegans). Their manuscript presents 1) a new method, which they claim supports Coelomata and 2) evidence for differences in the rate of intron loss across sites, which they claim nullifies our previous support for Ecdysozoa (Roy and Gilbert 2005a). We address these 2 lines of argument in turn.

The new method performs parsimony on a subset of “conserved” intron positions, defined as those present in at least 3 out of 5 taxa: D. melanogaster, A. gambiae, C. elegans, humans, and collective nonanimal outgroups. Because most phylogenetic patterns fulfilling this criterion are not informative (table 1), this reduces to introns present in both dipteran species and either C. elegans or humans (but not both). Thus, the authors’ major finding is: parsimony infers more intron gains (excluding those subsequently lost in a dipteran) in a hypothetical Coelomata ancestor than in an Ecdysozoan ancestor (as previously seen, e.g., compare figure 3b of Rogozin et al. [2003] and figure 2 of Roy and Gilbert [2005b]), which they interpret as evidence for Coelomata. However, there are several problems.

First, support for Coelomata is entirely dependent on their idiosyncratic definition of “groups.” Under their division whereas each dipteran species (D. melanogaster and A. gambiae, diverged 250–300 MYA; Holt et al. 2002) constitutes a separate group, the deeply diverged nonanimals (plants, apicomplexans, and fungi) share a single group. As such, all phylogenetically informative conserved positions are animal specific, whereas all phylogenetically informative introns shared with nonanimals are deemed “variable” and excluded. When more intuitive phylogenetic groupings are used (e.g., 3 main animal groups, fungi, and plants + apicomplexans or animals, fungi, plants, and apicomplexans), and only introns present in at least three groups considered, the results consistently support Ecdysozoa over both Coelomata and the third alternative (arthropods as an outgroup to a clade including nematodes and deuterostomes, termed “Bizarre” by Rogozin et al. [2007a]). For every phylogenetically reasonable alternative grouping we could think of, more characters support Ecdysozoa (table 2). (Note that these results do not by themselves provide strong support for Ecdysozoa because parallel losses in relatively intron-poor nematodes and dipterans must be corrected for [as in previous methods; see below].)

Notably, all the support for Coelomata relies on the “Coelomata” introns (those specific to humans and dipterans) having been gained along the internal branch (otherwise these introns are ancestral bilaterian introns that have been lost along an external branch and so are not informative). However, it has been repeatedly shown that such introns are not gains in Coelomata, but ancestral bilaterian ancestors that have been lost in C. elegans. Of the 87 Coelomata gains inferred by parsimony in a data set directly related to that of Zheng et al. (Rogozin et al. 2003), maximum likelihood reanalysis estimated than none (0) were true gains (Nguyen et al. 2005). More recently Carmel et al. (2007) estimated no gains along the Coelomata branch, even though parsimony infers 42 (calculated from their raw data, provided to us by Liran Carmel). Therefore, it seems most unlikely that many of these introns represent changes in a Coelomata ancestor.

That the Coelomata intron positions are not true Coelomata synapomorphies becomes even clearer when the intron-rich basal animal Nematostella vectensis is considered. In a similar data set (Sullivan et al. 2006), 69% of
intron positions shared between dipterans and humans (but absent from *C. elegans* and nonanimals and thus seemingly supportive of a Coelomata clade) are present in *N. vectensis* (calculated from their raw data, provided to us by Jim Sullivan), clear evidence that these introns are largely not true Coelomata synapomorphies. (We were unable to perform the analogous analysis on the Zheng et al. data set as the raw protein alignments were not preserved by the original authors.)

Further, the Zheng et al. method is expected to systematically place the taxon that has experienced the most intron loss as the outgroup. Regardless of topology, in order to be included in the study, an ancestral bilaterian intron must be retained all the way to dipterans (implying presence in the Coelomata or Ecdysozoan ancestor), as well as either 1) lost in humans and retained in *C. elegans* (supporting Ecdysozoa) or 2) lost in *C. elegans* and retained in humans (Coelomata). Because the *C. elegans* branch has experienced far more intron loss than humans (Rogozin et al. 2003; Banyai and Patthy 2004; Csuros 2005; Nguyen et al. 2005; Raible et al. 2005; Roy and Gilbert 2005a, 2005b; Carmel et al. 2007; Csuros et al. 2007), we a priori expect more ancestral bilaterian introns to artifactually support the Coelomata topology. The relative probabilities of loss along the 2 branches appear to be around 3 to 1: Among introns shared between a dipteran and a nonanimal, 24.0% and 63.5% are absent in humans and *C. elegans*, respectively. Thus, the relative probabilities of a character artifactually supporting Coelomata or Ecdysozoa are expected to be perhaps an order of magnitude different, consistent with the findings of Zheng et al.

The final point is more subtle. Zheng et al. introduce a modified Dollo parsimony argument that takes into account differences in probabilities of loss along branches—formally, differences in the cost for an intron loss along different branches. Different relative costs for loss along the human and *C. elegans* branches lead to different relative probabilities of Coelomata and Ecdysozoa. They show that the data support Coelomata only as long as the probability of loss in nematodes is at least 9.3 times higher than for humans (for details, see Zheng et al. 2007). However, the observed probability of loss along the nematode branch is only around 3 times higher than for humans (see above), suggesting a similar relative cost, well within the range over which the data support Ecdysozoa. In total, then, for several reasons the Zheng et al. results may not provide strong support for Coelomata.

The second claim of Zheng et al. was that variations in rates of intron loss across sites fully explain and nullify our previous findings (Roy and Gilbert 2005a). Zheng et al. first confirm previous findings of such differences (Sverdlov et al. 2004; Roy and Gilbert 2005c), which violate an assumption of our proposed intron position–based phylogenetic method (Rogozin et al. 2005; Roy and Gilbert 2005a). They conclude that our findings of support of Ecdysozoa are “fully explained by parallel loss of introns in nematodes and arthropods” (Zheng et al. 2007). However,
Table 2

Grouping of Species under Most Phylogenetic Criteria
Support Ecdysozoa

<table>
<thead>
<tr>
<th>Groups</th>
<th>Ecdysozoa</th>
<th>Coelomata</th>
<th>Bizarre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag, Dm, Ce, Hs, nonanimals</td>
<td>6</td>
<td>68</td>
<td>17</td>
</tr>
<tr>
<td>Ag, Dm, Ce, Hs, fungi + plants + apicomplexans</td>
<td>192</td>
<td>82</td>
<td>28</td>
</tr>
<tr>
<td>Ag, Dm, Ce, Hs, fungi + plants, apicomplexans</td>
<td>94</td>
<td>77</td>
<td>21</td>
</tr>
<tr>
<td>Ag, Dm, Ce, Hs, fungi, plants, apicomplexans</td>
<td>218</td>
<td>84</td>
<td>28</td>
</tr>
<tr>
<td>Diptera, Ce, Hs, fungi, plants + apicomplexans</td>
<td>186</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Diptera, Ce, Hs, fungi + plants, apicomplexans</td>
<td>88</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Diptera, Ce, Hs, fungi, plants, apicomplexans</td>
<td>212</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Animals, fungi, plants, apicomplexans</td>
<td>212</td>
<td>16</td>
<td>15</td>
</tr>
</tbody>
</table>

Note.—Ag, Anopheles gambiae; Dm, Drosophila melanogaster; Ce, Caenorhabditis elegans; and Hs, Homo sapiens.

Table 3

Predictions and Results from Roy and Gilbert (2005a)

<table>
<thead>
<tr>
<th>Ratios Compared</th>
<th>Coelomata</th>
<th>Predictions Ecdysozoa</th>
<th>Bizarre (D + N)</th>
<th>Results</th>
<th>Conclusion</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADN/DN/ADP[N]/DP[N]</td>
<td>=</td>
<td>&gt;</td>
<td>&lt;</td>
<td>104/246:177/610</td>
<td>&gt;</td>
<td>Ecdysozoa</td>
</tr>
<tr>
<td>DAN/AN:ADP[N]/AP[N]</td>
<td>=</td>
<td>&lt;</td>
<td>&gt;</td>
<td>104/36:177/34</td>
<td>&lt;</td>
<td>Ecdysozoa</td>
</tr>
<tr>
<td>AND/ND:AP[D]/NP[D]</td>
<td>&gt;</td>
<td>=</td>
<td>&lt;</td>
<td>104/246:75/165</td>
<td>=</td>
<td>—</td>
</tr>
<tr>
<td>NAD/AD:NP[D]/AP[D]</td>
<td>&lt;</td>
<td>=</td>
<td>&gt;</td>
<td>104/244:75/136</td>
<td>=</td>
<td>—</td>
</tr>
<tr>
<td>NDA/DA:NDP[A]/DP[A]</td>
<td>&lt;</td>
<td>=</td>
<td>&gt;</td>
<td>104/244:174/613</td>
<td>&gt;</td>
<td>Ecdysozoa</td>
</tr>
</tbody>
</table>

Note.—In the left-hand column, variables give numbers of introns with a given phylogenetic pattern among arthropods, nematodes, deuterostomes, and plants. For instance, ADN is the number of introns shared between Arthropods, Deuterostomes, and Nematodes (but not plants). Brackets indicate that either presence or absence is acceptable: For instance, DP[N] is the number of introns shared between Deuterostomes and Plants, present or absent in nematodes, and absent in arthropods. Results give actual values for the phylogenetic patterns in 684 sets of eukaryotic orthologs. See Roy and Gilbert (2005a) for details.
Two recent major studies make significant theoretical and empirical progress (Csuros et al. 2007; Carmel et al. 2007). An important remaining issue involves the prevalence of rate differences across sites, which have been repeatedly found (Sverdlov et al. 2004; Roy and Gilbert 2005b; Zheng et al. 2007) but which were not recovered by the most statistically sophisticated reconstruction to date (Carmel et al. 2007). Availability of large numbers of metazoan genomes affords the opportunity for significant model improvements.

In total, it is not clear what Zheng et al. have shown about the performance of our method. On the other hand, owing to several concerns raised here, in our opinion, their analyses do not provide convincing evidence for Coelomata.

Acknowledgments

M.I. was funded by the Spanish Ministerio de Educación y Ciencia, through the FPI grant (BFU2005-00252), and S.W.R. by the Intramural Research Program of the National Library of Medicine at National Institutes of Health/DHHS. We thank Eugene Koonin and Jordi Garcia-Fernandez and their groups for intellectual support and stimulation, for financial support, and for fostering environments of open intellectual exploration in their respective groups.

Literature Cited


Peter Lockhart, Associate Editor

Accepted January 20, 2008