Vertebrate-Type Intron-Rich Genes in the Marine Annelid *Platynereis dumerilii*

Florian Raible,[1,2]* Kristin Tessmar-Raible,[1] Kazutoyo Osoegawa,[4] Patrick Wincker,[5] Claire Jubin,[5] Guillaume Balavoine,[6] David Ferrier,[7] Vladimir Benes,[3] Pieter de Jong,[4] Jean Weissenbach,[5] Peer Bork,[2] Detlev Arendt[1]*

Previous genome comparisons have suggested that one important trend in vertebrate evolution has been a sharp rise in intron abundance. By using genomic data and expressed sequence tags from the marine annelid *Platynereis dumerilii*, we provide direct evidence that about two-thirds of human introns predate the bilaterian radiation but were lost from insect and nematode genomes to a large extent. A comparison of coding exon sequences confirms the ancestral nature of *Platynereis* and human genes. Thus, the urbilaterian ancestor had complex, intron-rich genes that have been retained in *Platynereis* and human.

The vast majority of bilaterally symmetrical animals stem from the Urbilateria, the last common ancestors of humans and flies (*1*) (Fig. 1). Despite the sequencing of several bilaterian genomes, the complexity of the urbilaterian gene structure remains unknown. Complex genes enhance the potential of a genome to modify transcripts by intron-mediated editing and to encode alternative splice variants, two sources of proteome complexity that affect cellular diversity (*2*). Marked differences exist between orthologous genes of ecdysozoans (*3*) (such as *Drosophila* and *Caenorhabditis elegans*), which have fewer introns, and deuterostome vertebrates (such as mouse and human), which have many introns. This is explained either by intron gain in the evolutionary lineage leading to vertebrates (*4*) or by intron loss in ecdysozoans (*5*). To decide between these alternatives on the basis of new sequence data, we analyzed genomic loci and transcripts in the marine annelid *Platynereis dumerilii* (*6*). This species belongs to the Lophotrochozoa (*7*) (Fig. 1), for which no complete genome sequence is yet available. We also included data from the recently sequenced honeybee (*Apis mellifera*) genome (*8*) for a more general picture of insect genomic characteristics (*9*). From an initial comparison between 1000 randomly selected orthologs shared between insects, *Caenorhabditis elegans*, the ascidian *Ciona intestinalis*, and humans, we found that *Apis* shares 25% of human introns, exceeding the fraction conserved in other ecdysozoans (fig. S1). This ratio is reproduced in random 25-gene subsets of the data (fig. S1), implying that our limited *Platynereis* data set should yield a suitable estimate of intron conservation in this species relative to that in deuterostomes or in ecdysozoans (*10*).

From 2.3 megabases (Mb) of available bacterial artificial chromosome (BAC) sequence, we identified 30 *Platynereis* gene loci with orthologs in other Bilateria. Transcripts were validated by reverse transcription polymerase chain reaction and mapped onto these loci. From this, we inferred the position of introns in the resulting proteins (*8*). These *Platynereis* genes contain 233 introns, or 7.8 introns per gene, similar to their human orthologs (8.4 introns per gene) but exceeding the values for the ecdysozoans (2.4 to 5.4 introns per gene). Three-quarters of the *Platynereis* introns are found in one or more of the other tested Bilateria (Fig. 2A). Most of these are shared with humans or with the teleost fish *Fugu*, but far fewer with insects or *C. elegans*. Thus, in our data set *Platynereis* is more similar to vertebrates than to any ecdysozoan as far as shared introns are concerned. We then assessed how many of the human introns were shared with other species and found that the fraction of human introns conserved in *P. dumerilii* is more than twice as high as the fraction conserved in *Apis* and larger than the fraction of human introns shared with any of the four ecdysozoans (Fig. 2B). Given that introns shared between distant taxa are most likely due to common inheritance (*11*), we conclude that at least two-thirds of the compared human introns already existed in the urbilaterian ancestor at precisely the same amino acid position and phase. This indicates that urbilaterian genes were already rich in introns and that the apparent differences in intron abundance between insects and nematodes and vertebrates are in large part due to intron loss in the ecdysozoan lineages (Fig. 2C). This also correlates with elevated rates of other forms of genome evolution ascribed to the ecdysozoan model species, such as gene loss (*12*, *13*).

In our data set, the ascidian *C. intestinalis* also shares fewer introns with humans than does *Platynereis* (Fig. 2B). This is again counterintuitive, considering the phylogeny (Fig. 1). Although this may be partially due to the relatively fragmentary *Ciona* genome assembly (*14*), it more likely reflects rapid genome evolution in the tunicate lineage (*15*). Illustrative of this trend, *Platynereis* shares with humans an intron in the N-terminal region of the Pax6 Paired domain that is crucial for generating two functionally divergent Pax6 isoforms. This intron has been considered a vertebrate innovation (*16*), because it is absent in *Ciona* and in another ascidian, *Phallusia* (*17*). The apparent intron loss in *Ciona* is not simply due to genome compaction (~6% of the human genome size), because ancestral introns are largely retained in the small, intron-rich genome (*18*) of the teleost fish *Fugu* (~10% of human genome size) (Fig. 2).

Next, we tested whether the retention of ancestral gene features in *Platynereis* is also apparent from the evolution of exon sequences. Systematic comparisons already indicated that the human proteome has

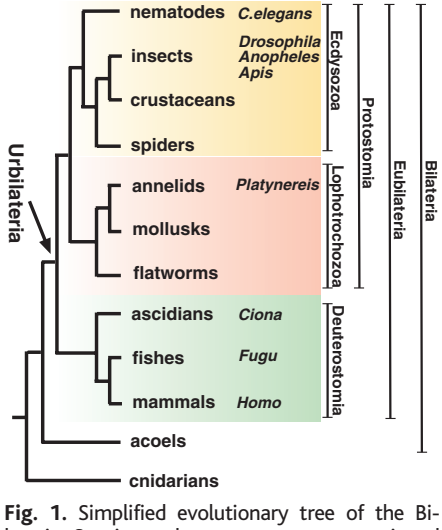

Fig. 1. Simplified evolutionary tree of the Bilateria. Species and group names as mentioned in the text.

[1]Developmental Unit, [2]Computational Unit, [3]Genomics Core Facility, European Molecular Biological Laboratory (EMBL), Meyerhofstraße 1, D-69117 Heidelberg, Germany. [4]Children's Hospital, Oakland Research Institute, 747 52nd Street, Oakland, CA 94609, USA. [5]Génoscope, Centre National de Séquençage and CNRS UMR 8030, 2 Rue Gaston Crémieux, CP 5706 91057 Evry Cedex, France. [6]Centre de Génétique Moléculaire du CNRS, Bâtiment 26, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France. [7]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

*To whom correspondence should be addressed. E-mail: raible@embl.de (F.R.); arendt@embl.de (D.A.)

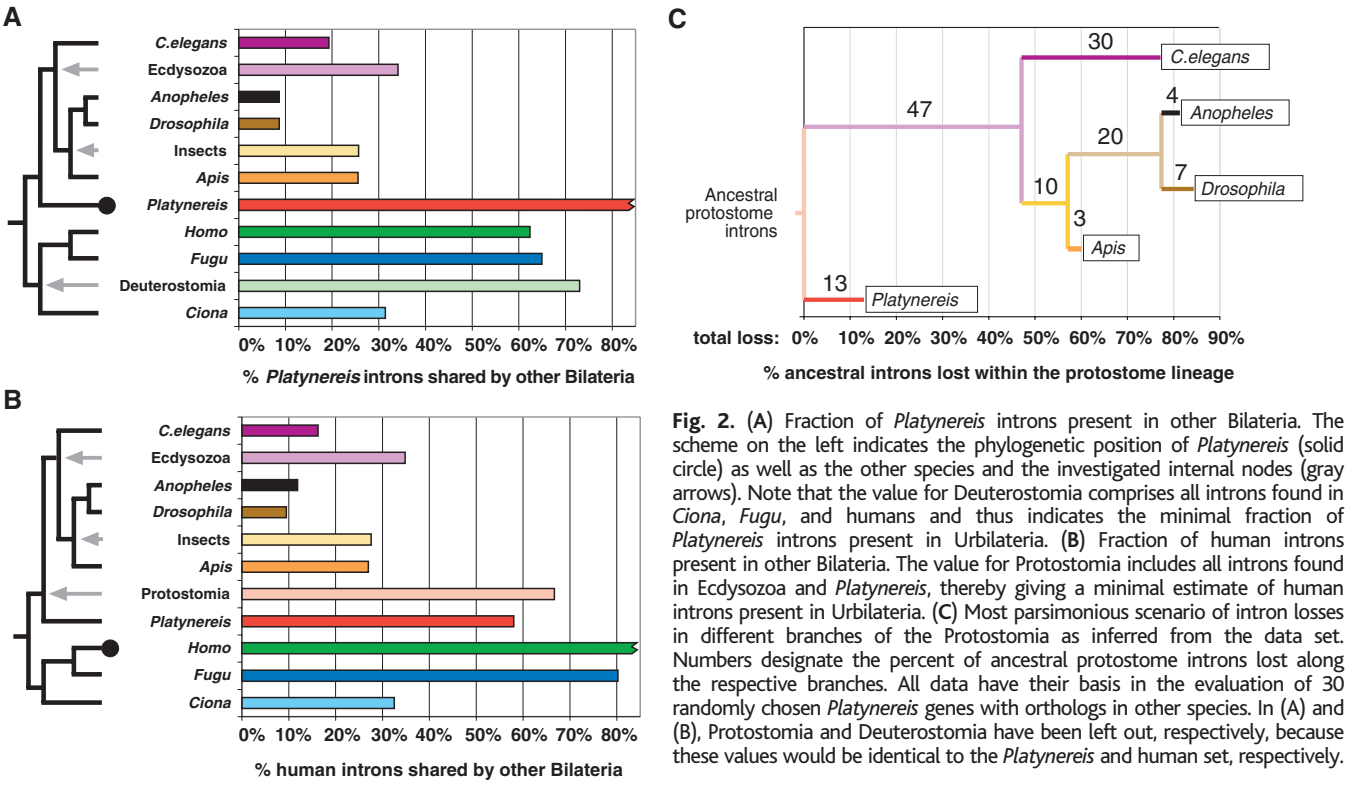

Fig. 2. (**A**) Fraction of *Platynereis* introns present in other Bilateria. The scheme on the left indicates the phylogenetic position of *Platynereis* (solid circle) as well as the other species and the investigated internal nodes (gray arrows). Note that the value for Deuterostomia comprises all introns found in *Ciona*, *Fugu*, and humans and thus indicates the minimal fraction of *Platynereis* introns present in Urbilateria. (**B**) Fraction of human introns present in other Bilateria. The value for Protostomia includes all introns found in Ecdysozoa and *Platynereis*, thereby giving a minimal estimate of human introns present in Urbilateria. (**C**) Most parsimonious scenario of intron losses in different branches of the Protostomia as inferred from the data set. Numbers designate the percent of ancestral protostome introns lost along the respective branches. All data have their basis in the evaluation of 30 randomly chosen *Platynereis* genes with orthologs in other species. In (A) and (B), Protostomia and Deuterostomia have been left out, respectively, because these values would be identical to the *Platynereis* and human set, respectively.

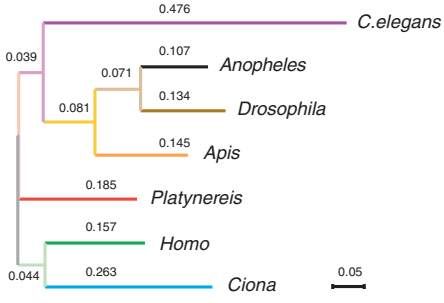

Fig. 3. Global distance tree based on concatenated alignments between conserved blocks of orthologous proteins found in *C. elegans*, *Anopheles*, *Drosophila*, *Apis*, *Platynereis*, humans, and *Ciona intestinalis*. Distances are indicated next to individual branches. The tree was calculated from 38,303 noninvariant amino acid positions according to the BLOSUM62 substitution model by using a maximum likelihood approach (*8*). All internal branches are fully supported.

departed less from the urbilaterian proteome than have the ones of *Drosophila* or *C. elegans* (*12*). To compare this to coding sequence evolution in polychaetes, we determined a set of 442 pan-bilaterian genes from 21,000 *Platynereis* expressed sequence tags. We then calculated distances to other bilaterians between aligned protein regions by using three different distance matrices (*8*). Both in single protein comparisons (fig. S2) and in a distance tree calculated from concatenated protein alignments (Fig. 3), *Platynereis* and human proteins were found to be more closely related to each other than to their ecdysozoan orthologs. Moreover, *Platynereis* proteins are closer to human than to *Ciona* orthologs, consistent with the trend observed for intron retention. We conclude that, at the intron and exon level, *Platynereis* and humans can be regarded as similarly slow-evolving representatives of protostomes and deuterostomes, respectively.

Our analyses consistently support the notion that Urbilateria possessed genes that, in both structure and sequence, were more similar to today's human or *Platynereis* genes than to those of dipterans, nematodes, or ascidians, where these initially complex genes have been secondarily simplified. Thus, in order to reconstruct the urbilaterian genome, comparisons of vertebrates with slow-evolving invertebrates will be of great benefit.

References and Notes

1. E. M. De Robertis, Y. Sasai, *Nature* **380**, 37 (1996).
2. B. R. Graveley, *Trends Genet.* **17**, 100 (2001).
3. A. M. Aguinaldo *et al.*, *Nature* **387**, 489 (1997).
4. I. B. Rogozin, Y. I. Wolf, A. V. Sorokin, B. G. Mirkin, E. V. Koonin, *Curr. Biol.* **13**, 1512 (2003).
5. S. W. Roy, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1986 (2005).
6. K. Tessmar-Raible, D. Arendt, *Curr. Opin. Genet. Dev.* **13**, 331 (2003).
7. K. M. Halanych *et al.*, *Science* **267**, 1641 (1995).
8. Materials and methods are available as supporting material on *Science* Online.
9. S. Kumar, *Nat. Rev. Genet.* **6**, 654 (2005).
10. To ensure accurate alignment of intron positions, we could only compare orthologous genes showing at least a moderate degree of coding sequence conservation. Less-conserved gene sets may show a lower degree of conservation in intron placement, but there is no a priori reason to expect that this should affect the ratio of intron conservation between species.
11. A. V. Sverdlov, I. B. Rogozin, V. N. Babenko, E. V. Koonin, *Nucleic Acids Res.* **33**, 1741 (2005).
12. R. D. Kortschak, G. Samuel, R. Saint, D. J. Miller, *Curr. Biol.* **13**, 2190 (2003).
13. A. Kusserow *et al.*, *Nature* **433**, 156 (2005).
14. P. Dehal *et al.*, *Science* **298**, 2157 (2002).
15. A. L. Hughes, R. Friedman, *Evol. Dev.* **7**, 196 (2005).
16. P. Callaerts, G. Halder, W. J. Gehring, *Annu. Rev. Neurosci.* **20**, 483 (1997).
17. S. Glardon, P. Callaerts, G. Halder, W. J. Gehring, *Development* **124**, 817 (1997).
18. S. Aparicio *et al.*, *Science* **297**, 1301 (2002); published online 25 July 2002 (10.1126/science.1072104).

19. We thank C. Burgtorf and H.-D. Pfannenstiel for providing *P. dumerilii* sperm DNA, H. Snyman, K. Willmann, and the GeneCore team for excellent help, and members of the Arendt and Bork groups for critical discussions. This work was supported by grants from the European Union (LSHG-CT-2003-503265, LSHG-CT-2003-503329, and QLRT-2001-02062 to P.B. and Marine Genomics Europe network to D.A.), the French Ministry of Research (to J.W. and G.B.), and the CNRS (to G.B.) and by a Boehringer Ingelheim Foundation Scholarship (to K.T.-R.). BAC library construction and characterization was supported by a subcontract (to K.O. and P.d.J.) from a grant (NSF IBN 0208138) awarded to L. Holland (Scripps Institution of Oceanography). The authors declare that they have no competing financial interests.

Supporting Online Material

www.sciencemag.org/cgi/content/full/310/5752/1325/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S4

19 August 2005; accepted 25 October 2005
10.1126/science.1119089