Minireview

Where is the difference between the genomes of humans and annelids?

Alexei Fedorov*† and Larisa Fedorova*

Addresses: *Department of Medicine and †Program in Bioinformatics and Proteomics/Genomics, Medical University of Ohio, Toledo, OH 43614, USA.

Correspondence: Alexei Fedorov. Email: afedorov@meduohio.edu

Published: 1 February 2006

Genome ***Biology*** 2006, **7:**203 (doi:10.1186/gb-2006-7-1-203)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2006/7/1/203

© 2006 BioMed Central Ltd

Abstract

The first systematic investigation of an annelid genome has revealed that the genes of the marine worm *Platynereis dumerilii* are more closely related to those of vertebrates than to those of insects or nematodes. For hundreds of millions of years vertebrates have preserved exon-intron structures descended from their last common ancestor with the annelids.

Among the millions of invertebrate species, the genomes of insects (particularly fruit flies of the genus *Drosophila*) and nematodes (from the genus *Caenorhabditis*) have come under the closest scrutiny. Now it is time for annelids - the segmented worms - to reveal their DNA sequences and gene structures. Last November, Raible and co-authors reported in *Science* the initial investigation of 30 genes from the marine annelid *Platynereis dumerilii* [1]. And it revealed a big surprise. The sequences of the annelid proteins were found to be more closely related to their human orthologs than to the insect and nematode orthologs. Moreover, among the species compared, the exon-intron structure of *P. dumerilii* genes was also most similar to that of humans: the human and the marine worm genomes have the highest number of introns per gene (7.8 for annelid and 8.4 for human and other mammalian genes) and more than 60% of annelid introns divide protein-coding sequences at exactly the same positions as human introns. By comparison, insects have 2.4 to 5.4 introns per gene and the plant representative *Arabidopsis thaliana* has 4.4, whereas fungi have the broadest spread from 0.0075 to 6.8 [2]. Thus, as far as shared introns are concerned, *P. dumerilii* is more similar to humans than to any insect or nematode.

The similarities in intron numbers and positions between *P. dumerilii* and humans does not imply that annelids should be combined with the vertebrates into a sister clade and distanced from nematodes and insects. Despite known uncertainties in the exact positioning of the segmented worms on the animal evolution tree (reviewed in [3]), nobody has ever grouped annelids with vertebrates. Molecular evolution is an intricate nonlinear process that can be interpreted in many different ways and it cannot be inferred from a set of equations. Conflicting facts and opposing opinions are common in the field and several alternative phylogenetic trees have been proposed for the animal kingdom. In their short article, Raible *et al.* [1] present only one animal phylogeny, whereas a subsequent comment by Kumar and Hedges [4] in *Cell* points out other well-recognized possible alternative relationships between flies, worms and humans. There has been a long and fierce debate about the phylogenetic relationships between arthropods, nematodes and vertebrates [5]. The 'Ecdysozoa hypothesis' groups arthropods and nematodes into a monophyletic clade and distances them from the vertebrates. The alternative 'Coelomata hypothesis' considers arthropods to be more closely related to the vertebrates than to the nematodes. We are still very far from resolving this dilemma. Even the whole-genome phylogenetic analyses of *Drosophila*, *Caenorhabditis elegans* and humans have not brought much clarity because, depending on the algorithms used, support can be found for both the Coelomata hypothesis [6] and the Ecdysozoa hypothesis [7].

Undoubtedly, the results in the *Science* article by Raible *et al.* [1] will serve as important additional, yet non-decisive, evidence in these endless debates on the origin of species.

The new data of Raible *et al.* [1] further complicate our perception of evolution. We used to think that humans have significantly advanced in complexity compared to the various worm phyla. It is, however, unclear where this complexity is encoded within our genome. *Homo sapiens* has only 21% more protein-coding genes than the microscopic nematode worm *Caenorhabditis elegans*, whose entire neural system is composed of 300 neurons [8]. Moreover, Raible *et al.* [1] have now shown that evolutionary changes in human protein sequences have occurred more slowly than in insects and nematodes. Thus, our proteins cannot be superior to those of invertebrates. Many biologists used to think that vertebrates had perfected their gene structures by acquiring thousands of new introns which, in turn, increased their protein diversity via alternative splicing. Yet according to Raible *et al.*, the last common ancestor between vertebrates and annelids had nearly as many introns as humans. So, evolution has hardly affected the gene structure of the vertebrate lineage.

The last line of defense in our ambition for pre-eminence is the total size of our genomes. Indeed, the human haploid genome contains more than 3 billion nucleotides, or 3.5 picograms (pg) of DNA. According to the Animal Genome Size Database [9] this is several times more than the haploid genome size of the vast majority of invertebrates, including *Drosophila melanogaster* (0.18 pg), *C. elegans* (0.10 pg), and *P. dumerilii* (0.89 pg). There are, however, many exceptions to the correlation between an organism's morphological and functional complexity and the absolute size of its genome. Some invertebrates have a genome size comparable to that of humans [10,11]. The well-known example of the unicellular *Amoeba dubia,* with a genome 200 times larger than that of humans [9], demonstrates that mere DNA length does not determine an organism's complexity. The major fraction of an animal genome is represented by non-protein-coding sequences: for humans, the noncoding regions comprise 98.5% of the genome. For years, noncoding DNA (including introns) was largely ignored and was frequently referred to as 'junk' or 'selfish' DNA. But only a few years ago it became generally appreciated that three quarters of our genome is transcriptionally active and produces at least 16,000 non-protein-coding RNAs, many of which have important cellular functions [12,13]. Nevertheless, a considerable fraction of transcribed noncoding RNA from species with extra-large genomes probably represents random transcripts without any valuable role for the organism [14]. Therefore, organismal complexity cannot be simply determined by the genome size, the number of protein-coding genes, the number of introns, or the total number of genomic transcripts. Rather, we should assess these parameters all together, and estimate how efficiently a particular species utilizes its genomic machinery, and the proportion of nonfunctional genomic ballast to functional elements.

Introns, for example, are ambivalent elements that create several problems for cells and perform various functions [15,16]. Jeffares *et al.* [17] have recently reviewed the process of intron gain and loss. They and others have demonstrated that intron evolution is not a simple stochastic process, and that a number of biological factors have important influences on change or preservation of exon-intron gene structures. Despite the fact that marine annelids and humans have the same number of introns, commonly in the same positions, the importance of these introns for the host cells could be different. Right now we only know for sure that human introns are longer than those of *P. dumerilii* and contain hundreds of noncoding RNAs as well as a number of valuable cellular signals [14]. All in all, introns are not trivial elements, and the comparison of their positions in different species is only the first step in understanding their intricate evolution in animals and other eukaryotic taxa. Biologists have recently gained access to genomic information from dozens of eukaryotic and hundreds of prokaryotic species. This has only brought us to the embryonic stage of genome biology theory and numerous surprises are to be expected along the road ahead.

References

1. Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, *et al.*: **Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*.** *Science* 2005, **310:**1325-1326.
2. Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE: **Patterns of intron gain and loss in fungi.** *PLoS Biol* 2004, **2:**e422.
3. McHugh D: **Molecular phylogeny of the Annelida.** *Can J Zool* 2000, **78:**1873-1884.
4. Kumar S, Hedges SB: **Pushing back the expansion of introns in animal genomes.** *Cell* 2005, **123:**1182-1184.
5. Telford MJ: **Animal phylogeny: back to the coelomata?** *Curr Biol* 2004, **14:**R274-R276.
6. Wolf YI, Rogozin IB, Koonin EV: **Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis.** *Genome Res* 2004, **14:**29-36.
7. Dopazo H, Dopazo J: **Genome-scale evidence of the nematode-arthropod clade.** *Genome Biol* 2005, **6:**R41.
8. **Wormbase** [http://www.wormbase.org/db/searches/neuron_display]
9. **Animal Genome Size Database** [http://www.genomesize.com]
10. Gregory TR: **Synergy between sequence and size in large-scale genomics.** *Nat Rev Genet* 2005, **6:**699-708.
11. Vinogradov AE: **Evolution of the genome size: multilevel selection, mutation bias or bynamical chaos?** *Curr Opin Genet Dev* 2004, **14:**620-626.
12. Suzuki M, Hayashizaki Y: **Mouse-centric comparative transcriptomics of protein coding and non-coding RNAs.** *BioEssays* 2004, **26:**833-843.
13. Frith MC, Pheasant M, Mattick JS: **The amazing complexity of the human transcriptome.** *Eur J Hum Genet* 2005, **13:**894-897.
14. Fedorova L, Fedorov A: **Puzzles of the human genome: why do we need our introns?** *Curr Genomics*, in press.
15. Lynch M, Richardson AO: **The evolution of spliceosomal introns.** *Curr Opin Genet Dev* 2002, **12:**701-710.
16. Rogozin IB, Sverdlov AV, Babenko VN, Koonin, EV: **Analysis of evolution of exon-intron structure of eukaryotic genes.** *Brief Bioinform* 2005, **6:**118-134.
17. Jeffares DC, Mourier T, Penny D: **The biology of intron gain and loss.** *Trends Genet* 2006, **22:**16-22.