

GENOME RESEARCH

Genomes in Motion

David L. Baillie

Genome Res. 2002 12: 843-

Access the most recent version at doi:[10.1101/gr.293102](https://doi.org/10.1101/gr.293102)

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *Genome Research* go to:
<http://www.genome.org/subscriptions/>



Genomes in Motion

David L. Baillie

Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

The examination of a sequenced genome produces many fascinating insights into how genes function, as well as tantalizing hints about the importance of gene orders and orientation. Indeed a static view of any single genome leads to a number of hypotheses regarding the history of its organization. Since the genome projects began, it has been clear that many of the questions arising from the examination of any single genome might well be resolved by having other genomes with which to compare. We now have a small collection of fairly complete eukaryotic genome sequences to examine (two distantly related yeasts, a nematode, an insect, and a higher plant). Although other sequences are near completion, they are not yet of sufficiently high quality that they can be confidently used in this type of comparison. Existing genome sequences are evolutionarily widely separated and the organisms are morphologically very different. Thus they are not yet very helpful when one wants to consider the forces and mechanisms that have led to the present state. Recognition of this by the genome community has resulted in efforts to sequence genomes that will fill in the phylogenetic gaps and that are evolutionarily close to existing sequenced genomes.

A prominent example is the mouse genome as a complement to the human genome. Efforts are also underway to produce genome sequences of close relatives of some of the more tractable model organism genomes (worm, fly, and yeast). In the case of the worm *Caenorhabditis elegans*, a sister species with very similar morphology has been selected, *Caenorhabditis briggsae*. The sequencing effort thus far has produced ~15 million bases of genome sequence (~15%–18% of the total). This sequence is available at

E-MAIL baillie@sfu.ca; **FAX** 604-291-5583.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.293102>.

the Genome Sequencing Center, Washington University School of Medicine (<http://genome.wush.edu/gsc/>). There is a concerted effort between the Washington University Genome Center and the Sanger Centre to complete the *C. briggsae* genome. The availability of these two high-quality data sets has proved irresistible to bioinformatics researchers. Kent and Zahler (2000) and Webb et al. (2002) have used this data to show the usefulness of newly developed tools for teasing information out of genomic sequence data from these closely related species.

In this issue of *Genome Research*, data from these two species has again been used in an extensive analysis of genome rearrangement. Rates of rearrangement are calculated and compared to the earlier data from *Drosophila* species. Coghlan and Wolfe at Trinity College have done an extensive and elegant analysis of the genomes of *C. elegans* and *C. briggsae* genomes and made some surprising discoveries and predictions for the overall rate of rearrangement in *Caenorhabditis*. They point out that this data set is "the largest available for any pair of congenic eukaryotes." The extent and quality of the sequence data make this analysis possible.

By first using BLASTX, Coghlan and Wolfe (2002) were able to predict the locations of 1784 orthologous genes in nearly 13 million megabases of *C. briggsae* genomic DNA. These were localized to 756 segments that ranged in size from 1 to 19 genes. When rearrangements were considered these segments could be reduced to 252, some containing as many as 109 genes. Using this set of ordered orthologs they analyzed the data to deduce the number of chromosomal rearrangements that would be required to give rise to the observed order. They determined that 517 chromosomal rearrangements would be needed. Transpositions are the most common event, but inversions and translocations each contributed about half as many

breaks. This leads to the conclusion that the genomes have had some 4030 rearrangements occur since the separation. This is a remarkable rate of rearrangement, even when considering the 50–120 million years that the investigators estimate for the divergence of the two species. They point out that this is higher than that reported for *Drosophila*. However, we will have to wait for comparable sequence data to arrive for a *Drosophila* sister species for this to be confirmed. Indeed they calculate that the breakage rate in *C. elegans* is 1400–17,000 times higher than has been calculated for mammals; again we must await the comparisons based on similar high-quality sequence in pairs of mammals. It is worth noting that the length of the conserved regions is increasing, Kent and Zahler (2000) claimed they averaged 8.1 kb, whereas as this paper claims they are 53 kb. This difference is largely attributed to differences in the analytic method and assumptions made in the two papers. It is clear that much is being learned about how genomes may be compared and how information from this comparison may be used. A whole *C. briggsae* genome assembly has been completed and is currently being analyzed (R. Waterston and R. Durbin, pers. comm.), this will allow the predictions made in the Coghlan and Wolfe paper (2002) to be confirmed.

REFERENCES

- Coghlan, A. and Wolfe, K.H. 2002. *Genome Res.* **12**:
 Kent, W.J. and Zahler, A.M. 2000. *Genome Res.* **10**: 1115–1125.
 Webb, C.T., Shabalina, S.A., Ogurtsov, A.Y., and Kondrashov, A.S. 2002. *Nucleic Acid Res.* **30**: 1233–1239.

WEB SITE REFERENCES

- <http://genome.wush.edu/gsc/>; The Genome Sequencing Center, Washington University School of Medicine.