

Minireview

The latest buzz in comparative genomics

Rob J Kulathinal and Daniel L Hartl

Address: Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

Correspondence: Daniel L Hartl. E-mail: dhartl@oeb.harvard.edu

Published: 4 January 2005

Genome Biology 2005, **6**:201

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/6/1/201>

© 2005 BioMed Central Ltd

Abstract

A second species of fruit fly has just been added to the growing list of organisms with complete and annotated genome sequences. The publication of the *Drosophila pseudoobscura* sequence provides a snapshot of how genomes have changed over tens of millions of years and sets the stage for the analysis of more fly genomes.

The genus *Drosophila* is no stranger to the spotlight. With over 2,000 known species, *Drosophila* offers a useful investigative platform for biologists of all sorts. Its interesting and diverse biology and ease of breeding in a variety of conditions has made *Drosophila* a favorite laboratory model organism. As the leading player in its genus, *Drosophila melanogaster* has enjoyed a long and distinguished tenure in biological research, particularly because it has become an indispensable model system for genetics. Ultimately, *D. melanogaster* was among the first eukaryotes to be sequenced [1] and the genome sequence triggered much excitement in terms of novel approaches and new-found collaborations.

New fly on the block

Although bottled 'populations' of *D. melanogaster* genetic mutants quickly became the standard resource for geneticists, these lab strains were at first not useful to those researchers studying evolutionary processes. *D. melanogaster* and its sibling species *Drosophila simulans*, although currently distributed worldwide, arrived only recently from Africa and are, therefore, not the most ideal material for understanding historical mechanisms. To study a more natural situation, Theodosius Dobzhansky, a naturalist and geneticist, began to work with the then little-known species *Drosophila pseudoobscura*, whose natural habitat range largely covers the western part of North America. Dobzhansky believed that

the genetics of speciation could be successfully understood only by studying natural genetic variation within populations, and he and others spent years developing genetic tools for *D. pseudoobscura*. Dobzhansky thought of *D. melanogaster* as a 'garbage species' whose human commensal activity was problematic for investigating microevolutionary processes involved in reproductive isolation. Much of his species choice was fortuitous - Dobzhansky taught at Caltech (Pasadena, USA) and was captivated by the large and ecologically stable levels of variation that he found among chromosome inversions in nearby populations of *D. pseudoobscura*. As a consequence of Dobzhansky's pioneering research, *D. pseudoobscura* and its sibling species *Drosophila persimilis* have become an important pair for geneticists interested in the evolution of reproductive isolation and speciation. Owing to its population-specific variation, *D. pseudoobscura* also became one of the most important population-genetic models [2-4] as well as an important reference species in comparison to *D. melanogaster* for studying evolution.

So it was with great interest that the research community recently welcomed *D. pseudoobscura* as the second fruit fly with a completely sequenced genome, providing a unique opportunity to systematically investigate the molecular evolution of two genomes from the same genus. The comparative approach enables evolutionary biologists to study precisely the types of changes that occur among nucleotides, genes, syntenic groups and genomes as a whole. The rate at

which proteins and chromosomes evolve is a direct consequence of the processes involved in the divergence of both genomes and species. And for those interested in annotating regulatory and coding regions of *D. melanogaster*, the direct comparison of orthologous regions between the two species provides an important resource for further curation of the *D. melanogaster* genome.

Time flies

To a good first approximation, the recent publication of the genome sequence of *D. pseudoobscura* [5] addresses many of these questions. For example, how different are the genomes of two congeneric species that diverged approximately 35 million years ago? Of nearly 14,000 genes annotated in a recent release of the *D. melanogaster* genome, more than 90% show evidence of orthology to the assembled *D. pseudoobscura* genome. Using a conservative reciprocal best-hits criterion, 10,516 orthologs were identified and their gene structures compared. Average nucleotide identities are relatively low in functionally less-constrained parts of genes - 40% among introns, 45-50% among untranslated regions (UTRs) and 49% among the third-position base pairs of codons. Not surprisingly, mean identity is higher among first and second position codon base-pairs (70%) as well as among protein-binding sites (63%).

In contrast to patterns of nucleotide divergence, chromosome arms, known as Muller's elements, are known to have remained very conserved throughout the evolution of the genus *Drosophila* [6]. In *D. melanogaster*, these six elements are arranged on the two arms of each of two metacentric autosomes, one dot autosome and one acrocentric sex chromosome (Figure 1a). In *D. pseudoobscura*, these six arms are retained, but the corresponding arms are rearranged into three acrocentric autosomes, plus one dot autosome and one metacentric sex chromosome (Figure 1b). Interestingly, most elements are almost one fifth larger in *D. pseudoobscura* than in *D. melanogaster* because of larger unclustered intergenic regions [5]. Whereas gene content within each Muller's element is remarkably conserved, gene order is not. In other words, while genes are retained in syntenic groups (on the same chromosome), they are not necessarily maintained in a continuous syntenic block (in the same order). The study by Richards *et al.* [5] reveals a history of extensive paracentric inversions (an average syntenic block is less than 100 kilobases (kb) in length, containing ten or so genes), very small pericentric inversions and a handful of single-gene transpositions. As the authors note [5], some reshuffling is not surprising. Because of the geometry of female meiosis and the lack of recombination in males, paracentric inversions are not terribly detrimental to the organism and an extensive set of inversions is found segregating, mainly on the X and third chromosomes, in natural populations of *D. pseudoobscura*. In fact, in some of his famous experiments Dobzhansky found that fitness

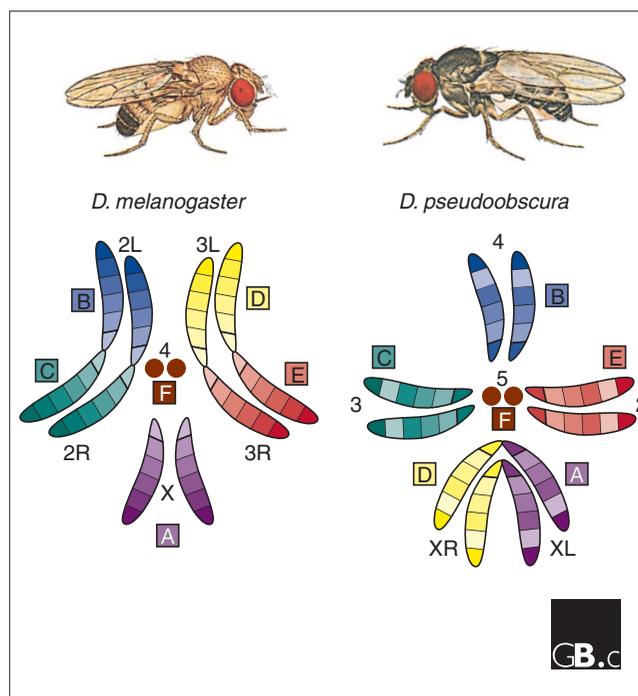


Figure 1
Arrangement of Muller's elements (chromosome arms) in *D. melanogaster* and *D. pseudoobscura*. The chromosomal arms (A-F) are highly conserved between the two species, but their organization into chromosomes differs. The chromosome number corresponding to each element is indicated. Gene content is conserved between elements, but gene order is not. The rearrangement of gene order is represented by shading within each chromosome arm.

differences between inversion types are correlated with environment [2]. But the ability precisely to identify regions of conserved gene order (a total of approximately 1,300 syntenic blocks were identified) demonstrates the power of this sort of comparative analysis [5].

The Richards *et al.* study [5] also provides an interesting causal explanation for the origin of the large number of pericentric inversions. After identifying the breakpoints of Arrowhead, one of the best-studied polymorphic inversions in *D. pseudoobscura*, the authors searched for similar instances of this short block of repeat-containing sequence among the approximately 1,300 identified interspecific syntenic breakpoints and found, remarkably, that this breakpoint motif shows homology to a large subset. In fact, these breakpoint motifs are, on average, 85% identical to each other and together constitute the largest family of repeats in *D. pseudoobscura*. Although they are significantly enriched at junctions between syntenic blocks, these breakpoint motifs share no homology to any *Drosophila* genes or known transposable elements from *D. melanogaster*.

Another interesting, but perhaps not so surprising, result demonstrates the presence of rapidly evolving male genes.

The authors [5] compared a set of predicted protein-coding genes from the *D. pseudoobscura* genome with the extensive collection of expressed sequence tags (ESTs) derived from various tissues of *D. melanogaster*. Testis-specific genes are found to be the most rapidly diverged between the two species, with an average percentage of amino-acid identity roughly 15% less than that of other transcripts. Not only are testis-specific genes diverging faster, but it seems that there is a greater number of testis-specific retrotransposed genes present in *D. melanogaster*. A significantly higher number of testis-specific orphan genes also supports a male-driven process of evolutionary innovation at the molecular level. Other work has found similar patterns of male divergence [7,8], but the analysis presented by Richards *et al.* [5] represents the first systematic and genome-wide demonstration of this phenomenon.

At 35 million years, *D. pseudoobscura* was considered sufficiently divergent from *D. melanogaster* to provide an ample supply of fixed nucleotide differences, yet close enough to retain conserved regulatory signatures when compared to *D. melanogaster* [9]. It was hoped that the *D. pseudoobscura* genome could therefore be used as a tool for detecting regions important for gene regulation. The presence of a functionally important signature is highlighted in a notable study by Ludwig *et al.* [10], in which chimeric *eve stripe 2* promoters from these two fruit-fly species cause misexpression of the *eve stripe 2* gene, whereas complete transgenes remain functional in the other species' genetic background. Richards *et al.* [5] map onto the *D. pseudoobscura* genome known *cis*-regulatory elements from the literature and find, rather unexpectedly, that these elements show levels of divergence close to random. This means that more closely related species must be sequenced in order to locate *cis*-regulatory elements in the *Drosophila* genome.

Flying high

The addition of *D. pseudoobscura* to the genomic cast is a milestone in comparative genomics. Comparison of the genome of this important model of speciation and development with that of its well-annotated sister species, *D. melanogaster*, will quickly become an indispensable tool for biologists. By using this genomic resource [5], we will be closer to tackling problems such as cracking the regulatory code and understanding the genetic basis of speciation given that, unlike *D. melanogaster*, *D. pseudoobscura* can hybridize with closely related species to generate fertile and viable offspring. At a broader level, this exploratory analysis represents the beginning of a larger chapter as other species of *Drosophila* are currently in various stages of genome sequencing. Thanks to the landmark efforts of a strong fruit-fly community, a dozen *Drosophila* species will be sequenced, assembled and eventually annotated during the coming year. The Richards *et al.* [5] comparative analysis of congeneric genomes is only a preview of exciting things to come.

Acknowledgements

We thank Brian Bettencourt and Stephen Richards for keeping us continually informed about the status of the *D. pseudoobscura* project.

References

1. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, *et al.*: **The genome sequence of *Drosophila melanogaster***. *Science* 2000, **287**:2185-2195.
2. Dobzhansky T: *Genetics and the Origin of Species*. New York: Columbia University Press; 1937.
3. Lewontin RC, Hubby JL: **A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura***. *Genetics* 1966, **54**:595-609.
4. Lewontin RC: *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press; 1974.
5. Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, *et al.*: **Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene and cis-element evolution**. *Genome Res* 2005, **15**:1-18.
6. Muller HJ: **Bearing of *Drosophila* work on systematics**. In *The New Systematics*. Edited by Huxley J. Oxford: Oxford University Press; 1940:185-268.
7. Singh RS, Kulathinal RJ: **Sex gene pool evolution and speciation: a new paradigm**. *Genes Genet Syst* 2000, **75**:119-130.
8. Swanson W, Vacquier V: **The rapid evolution of reproductive proteins**. *Nat Rev Genet* 2002, **3**:137-144.
9. Bergman CM, Pfeiffer BD, Rincon-Limas DE, Hoskins RA, Gnirke A, Mungall CJ, Wang AM, Kronmiller B, Pacle J, Park S, *et al.*: **Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome**. *Genome Biol* 2002, **3**:research0086.1-0086.20.
10. Ludwig MZ, Bergman C, Patel NH, Kreitman M: **Evidence for stabilizing selection in a eukaryotic enhancer element**. *Nature* 2000, **403**:564-567.