

## NEWS AND COMMENTARY

*Drosophila* unleashed

# The 12-genome update of the genetic power horse has been released

C Schlötterer

Heredity (2008) 100, 337–338; doi:10.1038/sj.hdy.6801084;  
published online 5 December 2007

Prospective sport car owners have a difficult choice: Audi, Porsche, BMW—just to name a few. For prospective graduate students, the choice of their preferred study organism has now been facilitated by the availability of 12 completely sequenced *Drosophila* genomes (Clark *et al.*, 2007; Stark *et al.*, 2007). No matter if you study molecular biology, functional genetics or evolutionary genomics, you can now be sure that with *Drosophila*, you pick an organism that will not let you down by lacking the tools required for your project.

Two publications that recently appeared in *Nature* report the sequencing, assembly and first analyses of these 12 genomes. The authors entered uncharted territories—for the first time the genomes of multiple species of the same genus are reported, which adds the new dimension of a comparative analysis in a phylogenetic framework. While some aspects of the analyses were conceptually well laid out, it is also clear that other analyses require the input of the entire community. Hence, well before publication, the genomic sequences were made available to the community to provide many labs the opportunity of contributing to the first insights from the 12 genomes. As a result of this community effort, more than 40 manuscripts are either already published, in press or under review elsewhere. I predict that this is just the beginning of a long series of publications taking advantage of the analytical power of 12 genomes in the same genus.

As expected, the analysis of 12 genomes provides a stimulating mixture of confirmatory results and astonishing insights. The accelerated rate of molecular evolution displayed by genes with a function in male reproduction compared to those not related to reproduction was foreseeable, since this had been shown earlier in an impressive series of publications. Also, the high variability in content of transposable elements among the *Drosophila* species sequenced

could have been anticipated. Nevertheless, the differences are impressive: while 25% of the genome of *Drosophila ananassae* consist of transposable elements, *D. simulans* and *D. grimshawi* contain only 2.7%. The dynamic nature of the *homeobox* (*Hox*) gene cluster came as no surprise to a *Drosophilist*. The clustering of *Hox* genes, which is conserved across vertebrates and invertebrates, led to the hypothesis that the collinear arrangement of the genes in the *Hox* cluster is a central functional component. In *Drosophila*, however, several cluster splits have occurred without any recognized functional implications.

The real power of the 12 genomes becomes apparent when the pattern of sequence conservation, inferred from genome-wide alignments for all species, is used for the identification of functionally important regions. RNA genes form hairpins and this structure is functionally important. To maintain the base pairing in the stem region, many mutations in the stem are expected to be accompanied by a compensatory change. Searching for a pattern of sequence conservation combined with compensatory changes, 177 putative RNA genes were identified in intergenic regions. Hairpin structures were also identified in introns, coding sequence and untranslated regions. The authors propose that these hairpins serve regulatory functions.

The discovery of regulatory sequence motifs is a substantially more challenging task. As regulatory motifs are short and present at many weakly specified sites, previous analyses integrated conservation information over all motif instances across the genome. While this approach provided considerable statistical power for the identification of even weakly conserved motifs, the identification of individual motif instances was problematic. Having multiple related genomes at hand, Stark *et al.* (2007) predicted individual motif instances by accounting for the phylogenetic distance to the *D. melanogaster* reference.

The comparison of the motifs identified by sequence conservation to motifs by chromatin immunoprecipitation data is an impressive demonstration of the statistical power provided by the 12 genomes. Nevertheless, it also became apparent that several of the motifs identified by ChIP are not evolutionarily conserved. This discrepancy suggests that a considerable proportion of the functionally important sequences is highly dynamic, and requires an analysis on the population level to understand the selective forces operating.

Using my personal lunch small talk indicator, the most highly noted result is the overwhelming evidence for unusual features of protein-coding genes: despite having already been described earlier, the polycistronic transcripts and conserved 'programmed' frameshifts, a feature that was thought to be rare in eukaryotes, have certainly spiced up my recent lunches. Nevertheless, my personal favorite among these unusual features is the predicted stop codon read-through. Despite the fact that the stop codon is conserved among the various *Drosophila* species, a clear signature of coding sequences was detected after this conserved stop codon. The authors note that it is unlikely that these genes are selenoproteins, in which the stop codon UGA is recoded as selenocysteine. Rather, Stark *et al.* (2007) propose alternative mechanisms, such as regulation of ribosomal release factors, A-to-I editing or alternative splicing.

For an evolutionary biologist, the icing on the cake is undoubtedly the analysis of genes and genomes in a phylogenetic framework. Much of the past evolutionary inference in *Drosophila* relied on moderately sized data sets that were generated in individual labs with small grants. Now, the entire genome could be analyzed across multiple species providing the opportunity to detect even subtle effects by comparing groups of genes, rather than individual genes only. Using  $\omega$ , the ratio of non-synonymous to synonymous substitutions, Clark *et al.* (2007) showed that functionally similar genes (based on Gene Ontology (GO) terms) are similarly constrained. Eleven percent of the GO terms were found to have elevated  $\omega$ , among which were 'defence response', 'response to biotic stimulus', 'receptor binding' and 'odor binding', suggesting that genes falling into these categories have experienced more recurrent positive mutations. Interestingly, the two specialist species,

*D. sechellia* and *D. erecta*, show a significantly accelerated evolution of olfactory and gustatory genes, probably related to the ecological shifts of these species. Apart from function, other factors such as chromosomal location, gene expression and the presence of repetitive amino acids in close proximity were also found to influence the rate of protein evolution.

With more questions generated than answered, the 12 *Drosophila* genomes are certainly only the first step into a new area of comparative genomics. Soon the Mammalian Genome Project will provide genomic sequences for 24 additional species. The *Drosophila* community is geared up for the next important step—the analysis of intra-specific variation on the genomic scale. Seven megabase pairs have been resequenced in 50 *D. melanogaster* individuals (<http://www.dpgp.org/melanogaster/>)

and six additional *D. simulans* strains have already been sequenced (Begun *et al.*, 2007), albeit at a low coverage. The recent advances in DNA sequencing technology in combination with its small genome size, make *Drosophila* particularly suitable for intraspecific resequencing studies and soon we will have large population data sets not for genes, but for genomes—so buckle up! C Schlotterer is at Institut für Tierzucht und Genetik, Veterinärmedizinische Universität Wien, Wien, Austria.

*e-mail: christian.schlotterer@vu-wien.ac.at*

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh YP, Hahn MW *et al.* (2007). Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* 5: e310.  
Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA *et al.* (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203–218.

Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW *et al.* (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* 450: 219–232.

#### Editor's suggested reading

- Anisimova M, Liberles DA (2007). The quest for natural selection in the age of comparative genomics. *Heredity Advance Online Publication*, 12 September 2007; doi:10.1038/sj.hdy.6801052.  
Leitch AR (2007). Conserved gene order belies rapid genome turnover: the dynamic interplay between genomic DNA and the outside world. *Heredity* 98: 61–62.  
Oldham MC, Geschwind DH (2006). Comparative genomics: grasping human transcriptome evolution: what does it all mean? *Heredity* 96: 339–340.  
Sabot F, Schulman AH (2006). Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. *Heredity* 97: 381–388.  
Wittkopp PJ (2006). Evolution of *cis*-regulatory sequence and function in Diptera. *Heredity* 97: 139–147.