The Drosophila Genome Variation Browser

Miquel Ràmia[1], Pablo Librado[2], Sònia Casillas[1], Julio Rozas[2] and Antonio Barbadilla[1]

[1]Departament de Genètica i Microbiologia. Facultat de Biociències. Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain.
[2]Departament de Genètica. Facultat de Biologia. Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain.

1. Abstract

The sequencing of the complete genome of 192 individuals of a single population of *D. melanogaster,* which has a high quality reference genome, will allow both a global view of genomic variation and an understanding of the forces that are responsible for the patterns of polymorphism and divergence along the genome (the DGRP project). We present a novel genome browser for the representation of genetic diversity in *D. melanogaster*, the "Drosophila Genome Variation Browser" (DGVBrowser), that automates the estimation of genetic variation along each chromosome and graphically represents all the information extracted from the data (SNP/CNV frequencies, nucleotide diversity, recombination rates, etc.), as well as other features gathered from external sources (TE insertions, mapped QTLs, etc.). Interestingly, this genome browser can be easily used to create analogous resources for any other species for which polymorphic sequences at the genome scale are being obtained. Beta version at: http://dgvbrowser.uab.es .

2. The "Drosophila Genetic Reference Panel"

The "Drosophila Genetic Reference Panel" (DGRP)[1] is a multitudinary effort to sequence the genomes of 192 inbreed lines from one natural population of *D. melanogaster.* The DGRP aims are to create:

- A community resource for association mapping of QTLs.
- A community resource of common Drosophila polymorphisms.
- A "test bench" for statistical methods used in QTL association and mapping studies for traits affecting human disease.

3. Drosophila Genome Variation Browser

To view and query the genomic polymorphism variation we have designed and implemented the DGVBrowser using the Generic Genome Browser[2] (GBrowse) from GMOD (Fig. 1). Annotations of the reference Drosophila genome have been imported from Flybase (Genes, mRNA, tRNA, Transposons,Introns, Exons, ...) along with the display of genome wide variation features. Beta version at http://dgvbrowser.uab.es .

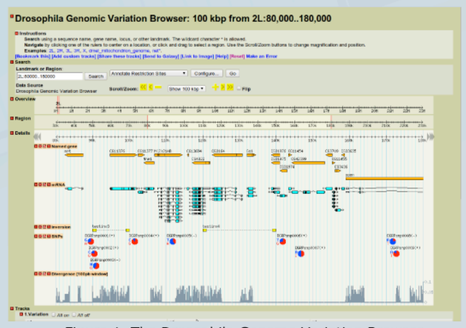


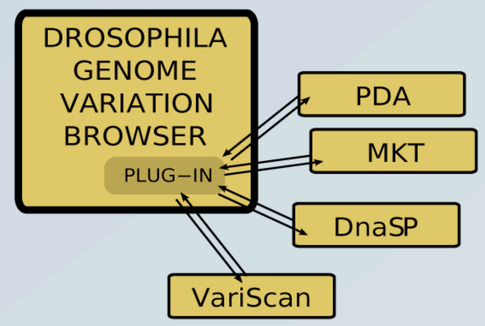
Figure 1. The Drosophila Genome Variation Browser

The Drosophila Genome Variation Browser can display diferent types of genomic variation at the nucleotide level (Table 1) using several graphic visualizations (glyphs):

- **Nucleotide Variation**: SNPs with both chromosomal location and allele frequency using frequency pie glyphs.
- **Structural Variation**: Indels, Copy Number Variation (CNV) and Chromosomal inversions. Both chromosomal location and allele frequency are reported using generic and frequency pie glyphs.
- **Quantitative data:** Displayed with xy plots of sliding windows (polymorphism [π], Tajima's D, divergence, recombination rates, neutrality estimates, Fay and Wu, etc.) which are precomputed for the whole genome.

Table 1. Genome variation annotations implemented in the DGVB with their graphical visualization (glyphs) with the GBrowse nomenclature.

Annotations	Visualization
SNPs with frequency	Allele_pie (HapMap)
Indels	Generic / Triangle
Indel frequency	Allele_pie
Inversions	Segments (with Breakpoints)
Inversion frequency	Allele_pie
CNV	Generic
Linkage disequilibrium (LD)	LDplot (HapMap)
Estimations: • Polymorphism (π) • Divergence • Tajima's D • Recombination rates • Fay & Wu • Neutrality estimates	XYplot / wiggle_plot (precomputat) *Plugin* (on-the-fly)
PhastCons	Remote (DAS - USCS)
Phenotype, Disease association	Generic + External Link
Others: • Minor Allele Frequency (MAF) • Derived Allele Frequency (DAF) • McDonald-Kreitman test (MKT)	Image/Graphic



GBrowse allows to expanding its functionality beyond the browsing by developing plug–ins. We use this option to interact with different software and obtain variation data on–the–fly.

The Pipeline Diversity Analysis (PDA)[3] and the McDonald-Kreitman Test Website (MKT)[4] have been developed by the Bioinformatics of the Genetic Variation Group at the Universitat Autònoma de Barcelona. PDA explores the polymorphic sequences available in the public databases, sorts and aligns them, and estimates genetic diversity in diferent functional regions. The MK test is used to compare online patterns of polymorphism and divergence to detect neutrality or selection in a DNA region.

DnaSP[5] estimates several measures of sequence variation and carries out several tests of neutrality. VariScan[6] conducts many population genetic analyses at the whole genome scale with a sliding window aproximation. DnaSP and VariScan have been developed by the Molecular Evolutionary Genetics Group at the Universitat de Barcelona.

4. Conclusions

• The DGVBrowser has been especifically designed to represent genome variation with a collection of tracks for its use in Population Genomics studies.

• DGVBrowser can be easily adapted to create analogous resources for any other species for which genome variational data is obtained.

5. References

1. Mackay T et al. The Drosophila Genetic Reference Panel - DGRP. http://service004.hpc.ncsu.edu/mackay/Good_Mackay_site/DBRP.html

2. Stein LD et al. The generic genome browser: a building block for a model organism system database. Genome Res (2002) 12: 1599-610

3. Casillas S, Barbadilla A. PDA v.2: improving the exploration and estimation of nucleotide polymorphism in large datasets of heterogeneous DNA. Nucleic Acids Res. (2006) Jul 1;34(Web Server issue):W632-4.

4. Egea R, Casillas S, Barbadilla A. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res. (2008) Jul 1;36(Web Server issue):W157-62.

5. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. (2009) Jun 1;25(11):1451-2.

6. Hutter S, Vilella AJ and Rozas J. Genome-wide DNA polymorphism analyses using VariScan. BMC Bioinformatics (2006) 7: 409.