BIOINFORMATICS APPLICATIONS NOTE

Vol. 28 no. 4 2012, pages 595–596
doi:10.1093/bioinformatics/btr691

Genetics and population analysis

Advance Access publication December 15, 2011

PopDrowser: the Population Drosophila Browser

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

[1]Institut de Biotecnologia i de Biomedicina and Departament de Genètica i de Microbiologia (Facultat de Biociències), Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona) and [2]Departament de Genètica and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain

Associate Editor: Jeffrey Barret

ABSTRACT

Motivation: The completion of 168 genome sequences from a single population of *Drosophila melanogaster* provides a global view of genomic variation and an understanding of the evolutionary forces shaping the patterns of DNA polymorphism and divergence along the genome.

Results: We present the 'Population Drosophila Browser' (PopDrowser), a new genome browser specially designed for the automatic analysis and representation of genetic variation across the *D. melanogaster* genome sequence. PopDrowser allows estimating and visualizing the values of a number of DNA polymorphism and divergence summary statistics, linkage disequilibrium parameters and several neutrality tests. PopDrowser also allows performing custom analyses on-the-fly using user-selected parameters.

Availability: PopDrowser is freely available from http://PopDrowser.uab.cat.

Contact: miquel.ramia@uab.cat

Received on October 6, 2011; revised on November 29, 2011; accepted on December 8, 2011

1 INTRODUCTION

Population genetics studies have been so far based on fragmentary and non-random samples of genomes, providing a partial and often biased view of the population genetics processes (Begun *et al.*, 2007). A new dimension to genetic variation studies is provided by the new availability of within-species genomes. Next-generation sequencing technologies are making affordable genome-wide population genetics data, not only for humans and the main model organisms, but also for most organisms on which research is actively carried out on genetics, ecology or evolution (Pool *et al.*, 2010).

Genome browsers are very useful tools to query and visualize disparate annotations at different genomic locations using a web user interface (Schattner, 2008). A number of web-based genome browsers displaying genetic variation data are already available (Benson *et al.,* 2002; Dubchak and Ryaboy, 2006; Frazer *et al.*, 2007; Hubbard *et al.*, 2002; Kent *et al.*, 2002; Stein *et al.*, 2002). Such browsers, however, are not well suited to deal with population genomics sequence information. For example, HapMap (International HapMap Consortium, 2003), the most comprehensive genome browser of variation data so far, contains information on single nucleotide polymorphisms (SNPs), Copy Number Variations (CNVs) and linkage disequilibrium of human populations. It does not offer, however, genetic variation estimates along sliding-windows or neutrality-based tests.

The Drosophila Genetic Reference Panel (DGRP) (T.Mackay *et al.*, accepted for publication) has recently sequenced and analyzed the patterns of genome variation in 168 inbred lines of *Drosophila melanogaster* from a single population of Raleigh (USA), and conducted a genome-wide association analysis of some phenotypic traits. A major goal of this project is to create a resource of common genetic polymorphism data to aid further population genomics analyses. As a part of this DGRP project, here we present a modified Gbrowse specifically designed for the automatic estimation and representation of population genetic variation in *D. melanogaster*, the 'Population Drosophila Browser' (PopDrowser). Unlike other population analysis tools (Hutter *et al.*, 2006; Kofler *et al.*, 2011), the PopDrowser is a genome browser, which can be customized to create analogous resources for any other species with within-species polymorphism data.

2 IMPLEMENTATION

2.1 Input data

The initial input data are a set of 168 aligned intraspecific *D. melanogaster* sequences from the DGRP project, and also include the genome sequences of *Drosophila yakuba* and *Drosophila simulans,* which were used as outgroup species.

2.2 Interface and implementation

PopDrowser allows reporting precomputed estimates of several DNA variation measures along each chromosome arm through the combined implementation of the programs PDA 2 (Casillas and Barbadilla, 2006), MKT (Egea *et al.*, 2008) and VariScan 2 (Hutter *et al.*, 2006). The data and summary statistics are graphically displayed along the chromosome arms on a web-based user interface using the Gbrowse software.

PopDrowser also includes an innovative capability that allows performing custom analyses on-the-fly. After selecting a chromosome region and a particular track, the user can conduct exhaustive analyses by defining their own custom input parameters. Furthermore, users can choose to either visualize the output of their analyses graphically in the browser—as a new track—or to

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

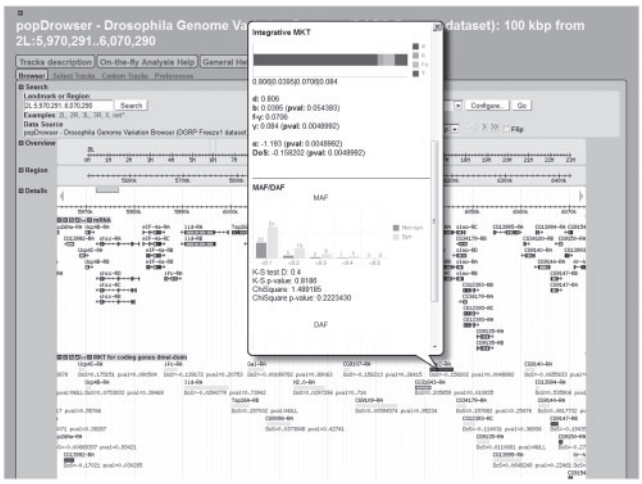

Fig. 1. PopDrowser snapshot showing the results of the McDonald–Kreitman tests in the *ade2-RA* gene within its genome context.

download it in a tabulated text file. The estimates available for on-the-fly analyses are specified in Table1.

The current implementation is running in an Ubuntu 10.04 Linux x64 server, 2 IntelXeon 3Ghz processors, 32GB RAM with Apache.

2.3 Output

Along with reference genome annotations, the genome browser output includes measures of a number of nucleotide summary statistics, such as the levels of nucleotide diversity (π and θ), DNA divergence between species (K), different measures of linkage disequilibrium and genome-wide neutrality tests. Such analyses are computed along each chromosome arm in non-overlapping sliding windows of 0.05, 0.1, 0.5, 1, 10, 50 and 100 kb. For each gene, the browser also provides a single track including information of the generalized and the integrative McDonald–Kreitman tests (McDonald and Kreitman, 1991; T.Mackay *et al.*, accepted for publication) along with minor and derived allele frequency (MAF, DAF) spectrums (Fig. 1). All the tracks included in the PopDrowser are summarized in Table 1.

Table 1. Summary of the PopDrowser tracks

Category	Gene annotations and estimates
Drosophila melanogaster reference annotations (build 5.13) and recombination	Gene structure, mRNA, CDS (Coding Sequence), ncRNA, tRNA, orthologous genes, phastCons, GC content. local recombination rate (Fiston-Lavier *et al.*, 2010)
Density tracks	Genes, microsatellites, transposons, CDS, SNPs
Nucleotide variants	SNPs, single nucleotide fixations
Measures of nucleotide variation and LD)[a]	Number of segregating sites (S), total minimum number of mutations (η), number of singletons (η_e), nucleotide diversity (π), Watterson's estimator of nucleotide diversity per site (θ), number of haplotypes (h), haplotype diversity (Hd), nucleotide divergence per site (corrected by Jukes–Cantor) (K). LD: D, absolute D ($\|D\|$), D', absolute D' ($\|D'\|$), r^2
Neutrality tests[a]	Fu and Li's D, D^a, F, F^a, Fay and Wu's H, Tajima's D, Fu's F_S statistics. MKT (per gene)

LD, linkage disequilibrium; CDS, coding sequence.
[a]Estimates available for on-the-fly analyses (except MKT per gene).

ACKNOWLEDGEMENTS

We thank Raquel Egea for helping in implementing the MKT software and Albert Vernon Smith from HapMap for his help in the implementation of the pie graph glyph. We also thank Dave Clements, Lincoln Stein and Scott Cain for technical support, and John H. Werren for valuable discussions on the browser. This paper was prepared with full knowledge and support of the DGRP Consortium.

Funding: Ministerio de Ciencia e Innovación (Spain) (BFU2009-09504 to A.B., BFU2010-15484 to J.R.); Catalonian Comissió Interdepartamental de Recerca i Innovació Tecnològica (2009SGR-88 to A. Ruiz; 2009SGR-1287 to M. Aguadé); Departament de Genètica i de Microbiologia of the Universitat Autònoma de Barcelona (409-04-2/08 to M.R.); ICREA Academia (Generalitat de Catalunya to J.R., in part).

Conflict of Interest: none declared.

REFERENCES

Begun,D.J. *et al.* (2007) Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biol.*, **5**, e310.

Benson,D.A. *et al.* (2002) GenBank. *Nucleic Acids Res.*, **30**, 17–20.

Casillas,S. and Barbadilla,A. (2006) PDA v.2: improving the exploration and estimation of nucleotide polymorphism in large datasets of heterogeneous DNA. *Nucleic Acids Res.*, **34**, W632–W634.

Dubchak,I. and Ryaboy,D.V. (2006) VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods Mol. Biol.* 2006, **338**, 69–89.

Egea,R. *et al.* (2008) Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res.*, **36**, W157–W162.

Fiston-Lavier,A.S. *et al.* (2010) *Drosophila melanogaster* recombination rate calculator. *Gene*, **463**, 18–20.

Frazer,K.A. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.

Hubbard,T. *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.

Hutter,S. *et al.* (2006) Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics*, **7**, 409.

International HapMap Consortium (2003) The International HapMap Project. *Nature*, **426**, 789–796.

Kent, W.J. *et al.* (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.

Kofler,R. *et al.* (2011) PoPoolation, a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, **6**, e15925.

McDonald,J.H. and Kreitman,M. (1991) Adaptive protein evolution at the Adh locus in Drosophila. *Nature*, **351**, 652–654.

Pool,J.E. *et al.* (2010) Population genetic inference from genomic sequence variation. *Genome Research*, **20**, 291–300.

Schattner,P. (2008) *Genomes, Browsers and Databases. Data-Mining Tools for Integrated Genomic Databases*. Cambridge University Press, New York.

Stein,L.D. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.