Application Note

DrosOmics: the comparative genomics browser to explore omics data in natural strains of *D. melanogaster*

Marta Coronado-Zamora, Judit Salces-Ortiz and Josefa González*

Institute of Evolutionary Biology, CSIC, Universitat Pompeu Fabra, Barcelona, Spain

*To whom correspondence should be addressed.

Abstract

Summary: The advent of long-read sequencing technologies has allowed the generation of multiple high-quality *de novo* genome assemblies for multiple species including well-known model species such as *Drosophila melanogaster*. Genome assemblies for multiple individuals of the same species are key to unravel the genetic diversity present in natural populations, especially the one generated by transposable elements, the most common type of structural variant. Despite the availability of multiple genomic datasets for *D. melanogaster* populations, we lack an efficient visual tool to display different genomes assemblies simultaneously and compare how homologous regions differ in terms of the structural variation composition. In this work, we present DrosOmics, a comparative genomics-oriented browser for 52 high-quality reference genomes of *D. melanogaster*, spanning 32 locations worldwide and including annotations from a highly reliable set of transposable elements. DrosOmics is based on JBrowse 2, a highly scalable and user-friendly platform that allows the visualization of multiple assemblies at once. In this browser, we have also compiled functional -omics data for half of the populations, including transcriptomics and epigenomics data. We believe that this resource will be key for unraveling structural and functional features of *D. melanogaster* natural populations and will open the door for exploring new evolutionary biology hypotheses.

Availability and implementation: DrosOmics browser is open and freely available at http://gonzalezlab.eu/drosomics.

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Contact: Dr. Josefa González (josefa.gonzalez@csic.es).

1 Introduction

The advent of technological developments in DNA sequencing have led to a high number of whole-genome sequences for multiple individuals of the same species, revolutionizing the field of comparative population genomics (Mitsuhashi and Matsumoto, 2020; Sakamoto *et al.*, 2020). Among these developments, long-read sequencing technologies allowed to improve the quality and completeness of reference genomes because of their ability to span repetitive regions, leading to a better detection and annotation of structural variants such as transposable elements (TEs) (Solares *et al.*, 2018; Du and Liang, 2019; Miga *et al.*, 2020; Rech *et al.*, 2022). *Drosophila melanogaster*, one of the central model species for molecular population genomics studies (Casillas and Barbadilla, 2017) and for studying TEs (Mccullers *et al.*, 2017), have benefited from these technological developments by currently having several *de novo* reference genomes representative of natural populations sampled worldwide (Chakraborty *et al.*, 2019; Long *et al.*, 2018; Rech *et al.*, 2022).

Complementary to these population genome sequence data, there has been a large increase in the generation of other functional -omics data, such as transcriptomics and epigenomics. While the genome sequence represents a static component of the organism, functional -omics data vary across different body parts or tissues (spatial), life stages (temporal), or conditions (*e.g.* different environments, experimental treatments). Capturing this information is relevant for *e.g.* understanding how organisms adapt to changing environments, plant and animal breeding, and medical genetics.

One major challenge for genomic studies is to visualize and retrieve these amounts of genome and functional -omics data. For that, web-based genome browsers, such

as JBrowse (Buels *et al.*, 2016), are the most suitable platform for browsing, visualizing and retrieving data through an efficient and user-friendly graphical interface. Previous population genomics-oriented browsers in *D. melanogaster* based on JBrowse (*e.g.,* PopFly [https://popfly.uab.cat; Hervas *et al.*, 2017] or DEST [https://dest.bio; Kapun *et al.*, 2021]), compile data from several worldwide natural populations which were aligned to the *D. melanogaster* reference genome and, thus, the structural variation particular of each population was lost. However, the recent version of JBrowse, JBrowse 2, allows multiple genome assemblies to be loaded at once and thus to explore the genome and structural variation of several populations simultaneously. Furthermore, including functional -omics data such as RNA-seq or ChIP-seq can be very informative as they allow genomic variants, *e.g.*, single nucleotide polymorphisms (SNPs) and structural variants, to be linked to variation in the gene expression and epigenetic modifications.

In this work, we present DrosOmics, a comparative genomics-oriented browser that allows the visualization and retrieval of genome sequences and functional -omics data for 52 natural populations of *D. melanogaster* collected in 32 locations world-wide. These genomes have been annotated with a high-quality and manually-curated set of TEs accounting for the variability observed in these natural populations (Rech *et al.*, 2022), and thus, outperforming and improving the current gold-standard TE annotation of FlyBase (Larkin *et al.*, 2021). DrosOmics also compiles functional annotations for half of the genomes, including transcriptomics (RNA-seq) and epigenomics datasets (ChIP-seq and ATAC-seq) performed under different experimental conditions, for whole body individuals or adult body parts. DrosOmics also incorporates tools to visualize custom genomes and tracks and to download the data. This platform is highly scalable, allowing to add new genomes

and new functional -omics datasets for *D. melanogaster* or other *Drosophila* species provided by the community. We believe that DrosOmics will be key for unraveling structural and functional features of *D. melanogaster* natural populations, representing a great opportunity to explore and test new evolutionary biology hypotheses. DrosOmics is open and freely available at http://gonzalezlab.eu/drosomics.

2 DrosOmics browser content overview

2.1 *D. melanogaster* reference genomes

DrosOmics compiles 52 high-quality genome assemblies that have been obtained using long-read sequencing technologies (Table 1 and Table S1).

Table 1. 52 *D. melanogaster* reference genome assemblies compiled in DrosOmics.

Genomes	Source locations	Data source
32 genomes	12 localities from 8 countries: America (1), Europe (7)	Rech *et al.* (2022)
14 genomes	14 localities from 11 countries: Africa (2), Europe (2), North America (1), North Atlantic Ocean (1), South America (2), Asia (3)	Chakraborty *et al.* (2019)
5 genomes	5 localities from 5 countries: Africa (1), Europe (1), America (1), Africa (1), Oceania (1)	Long *et al.* (2018)
1 genome (ISO-1)	America	Solares *et al.* (2018)

2.2 Genome browser implementation

DrosOmics contains TE and gene annotations for the aforementioned 52 *D. melanogaster* reference genomes. Note that the current version of DrosOmics provides gene annotations that were transferred to each genome from the reference release 6 of *D. melanogaster* (Larkin *et al.*, 2021). Functional -omics data for 26

genomes were retrieved from previous research (Everett *et al*., 2020; Salces-Ortiz *et al.,* 2020; Green *et al.*, 2021; Horváth *et al.*, 2022; Table 2 and Table S1) or are reported here for the first time (Supplemental Material, Table S1). Genome and functional annotations can be graphically displayed along the chromosome arms for multiple genomes simultaneously (Figure 1 and Table 2). DrosOmics is built on JBrowse 2 (Buels *et al.*, 2016) and is currently running under a docker container using Apache on a CentOS 7.9.2009 Linux x64 server with 4 vCores processors and 16 GB RAM.

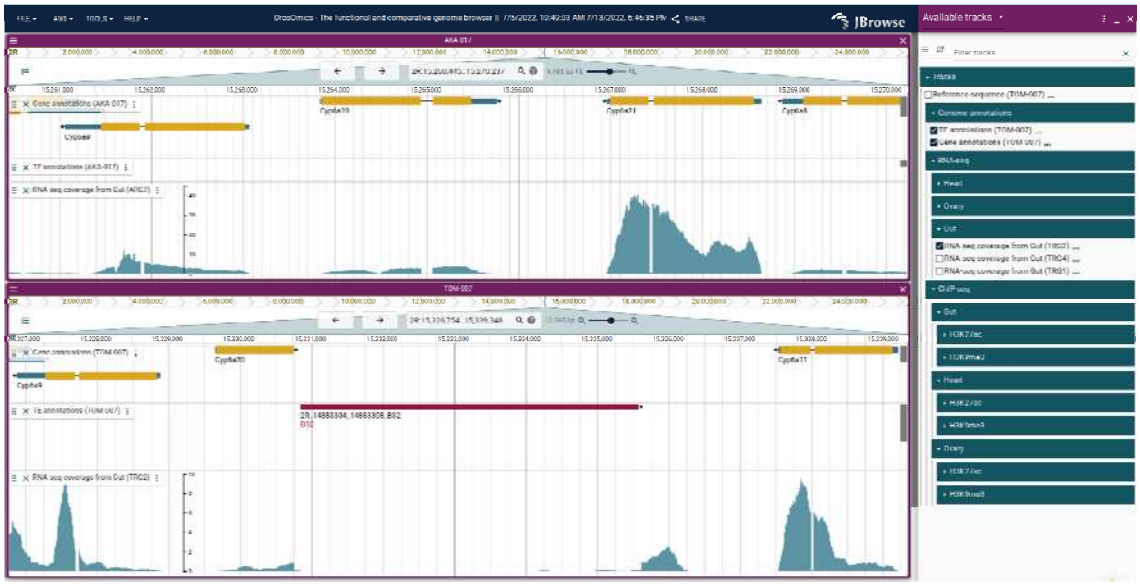

Figure 1. DrosOmics snapshot displaying a homologous region containing genes from the *cyp* family in two different genomes: AKA-017 (Akaa, Finland) on top and TOM-007 (Tomelloso, Spain) on the bottom. Gene, TE annotations, and RNA-seq tracks are activated for both genomes. On the left, the track selector menu for TOM-007 available annotations is displayed.

Table 2. Available tracks in DrosOmics: >500 available tracks. Further details on raw data number accessions, data processing and publications associated are available on Supplemental Material and Table S1. Symbols indicate the publication source for the functional data: □This publication; ■Everett *et al.* (2020); ▲Salces-Ortiz *et al*. (2020); ●Green *et al*. (2021); ►Horváth *et al*. (2022). The number of replicates (repl) for each dataset is given. FC: fold change.

Track type	Track name	Description	Genomes	Datasets
Genome sequence	Reference sequence	Displayed by chromosome arm (2R, 2L, 3L, 3R and X)	All	52
Genome annotations	TE annotations Gene annotations	- *De novo* TE annotations - Gene annotations transferred from *D. melanogaster* release 6 annotation	All	104
RNA-seq	RNA-seq coverage	- Basal conditions: head, gut, ovary - Basal conditions: whole-body - Basal and malathion treatment: gut - Basal and copper treatment: whole body - Basal and desiccation treatment: whole body	5 genomes ■ 8 genomes ▲ 4 genomes ● 6 genomes ► 6 genomes	15 (3 repl) 8 (2 repl) 8 (3 repl) 12 (3 repl) 12 (3 repl)
ChIP-seq enrichment	ChIP-seq FC signal ChIP-seq peak calling	- H3K9me3 (repressive mark): head, gut, ovary - H3K27ac (active mark): head, gut, ovary	5 genomes 5 genomes	15 (3 repl) 15 (3 repl)
ATAC-seq enrichment	ATAC-seq FC signal ATAC-seq peak calling	- Basal and malathion treatment whole-body	▲4 genomes	8 (3 repl)

DrosOmics also contains utilities and support resources: a tool for downloading all the compiled data and a Help section including useful documentation and a Tutorial with step-by-step worked examples.

3 Concluding remarks

DrosOmics browser is highly scalable, allowing the continuous update by the addition of new annotations, populations, or even new *Drosophila* species. We aim that DrosOmics will become a reference tool for future comparative genomics studies in *D. melanogaster*.

Data availability

All processed data is available to download from http://gonzalezlab.eu/drosomics. The RNA-seq and ChIP-seq raw data for five genomes that are reported here for the first time are available in the NCBI Sequence Read Archive (SRA) database under BioProject PRJNA643665.

Acknowledgements

We would like to thank the UPF Scientific IT Core Facility for support on implementing the JBrowse platform, S. Radío (from González Lab) for providing the TE and gene annotations for the five genomes of Long *et al*. (2018), and the DrosEU consortium for sharing flies with us.

Funding

This work has been supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (H2020-ERC-2014-CoG-647900). DrosEU is funded by an Special Topic Network award from the European Society for Evolutionary Biology (ESEB). J.S.-O. was funded by a Juan de la Cierva-Formación fellowship (FJCI-2016-28380).

Conflict of Interest: none declared.

References

Adams, M. D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, 287, 2185–95.

Buels, R. *et al.* (2016) JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.*, 17, 1–12.

Casillas, S. and Barbadilla, A. (2017) Molecular population genetics. *Genetics*, 205, 1003–1035.

Chakraborty, M. *et al.* (2019) Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat. Commun.*, 10, 1–11.

Du, H. and Liang, C. (2019) Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. *Nat. Commun.,* 10, 1–10.

Everett, L. J. *et al.* (2020) Gene expression networks in the *Drosophila* Genetic Reference Panel. *Genome Res.*, 30, 485–496.

Green, L. *et al.* (2021) Natural variation in copper tolerance in *Drosophila melanogaster* is shaped by transcriptional and physiological changes in the gut. *bioRxiv*, 2021.07.12.452058.

Hervas, S. *et al.* (2017) PopFly: the *Drosophila* population genomics browser. *Bioinformatics*, 33, 2779–2780.

Horváth, V. *et al.* (2022) Basal and stress-induced expression changes consistent with water loss reduction explain desiccation tolerance of natural *Drosophila melanogaster* populations. *bioRxiv*, 2022.03.21.485105.

Kapun, M. *et al.* (2021) *Drosophila* Evolution over Space and Time (DEST): A new population genomics resource. *Mol. Biol. Evol.*, 38, 5782–5805.

Larkin, A. *et al.* (2021) FlyBase: updates to the *Drosophila melanogaster* knowledge base. *Nucleic Acids Res.*, 49, D899–D907.

Long, E. *et al.* (2018) Genomic structural variations within five continental populations of *Drosophila melanogaster*. *G3 Genes|Genomes|Genetics*, 8, 3247–3253.

Mccullers, T. J. and Steiniger, M. (2017) Transposable elements in *Drosophila*. *Mob. Genet. Elements*, 7, 1–18.

Miga, K. H. *et al.* (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature.,* 585, 79–84.

Mitsuhashi, S. and Matsumoto, N. (2020) Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.*, 65, 11–19.

Rech, G. E. *et al.* (2022) Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in *Drosophila*. *Nat. Commun.,* 13, 1–16.

Sakamoto, Y. *et al.* (2020) A new era of long-read sequencing for cancer genomics. *J. Hum. Genet.*, 65, 3–10.

Salces-Ortiz, J. *et al.* (2020) Transposable elements contribute to the genomic response to insecticides in *Drosophila melanogaster*. *Phil. Trans. R. Soc. B,* 375, 20190341–20190341.

Solares, E. A. *et al.* (2018) Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. *G3 Genes|Genomes|Genetics*, 8, 3143–3154.