Molecular Biology and Evolution, 2025, **42**, 1–28
https://doi.org/10.1093/molbev/msaf132
Advance access publication 18 August 2025
Discoveries

MBE OXFORD

Footprints of Worldwide Adaptation in Structured Populations of *Drosophila melanogaster* Through the Expanded DEST 2.0 Genomic Resource

Joaquin C. B. Nunez,[1,2,*†‡] **Marta Coronado-Zamora**,[3,4,*†‡] **Mathieu Gautier**,[5,‡] **Martin Kapun**,[6,‡] **Sonja Steindl**,[6,‡] **Lino Ometto**,[7,‡] **Katja Hoedjes**,[8,‡] **Julia Beets**,[8,‡] **R. Axel W. Wiberg**,[9,‡] **Giovanni R. Mazzeo**,[2] **David J. Bass**,[2,10,11] **Denys Radionov**,[12] **Iryna Kozeretska**,[13,‡] **Mariia Zinchenko**,[14] **Oleksandra Protsenko**,[13,15,‡] **Svitlana V. Serga**,[5,13,‡] **Cristina Amor-Jimenez**,[16,17] **Sònia Casillas**,[16,17,‡] **Alejandro Sánchez-Gracia**,[18,19,‡] **Aleksandra Patenkovic**,[20,‡] **Amanda Glaser-Schmitt**,[21,‡] **Antonio Barbadilla**,[16,17,‡] **Antonio J. Buendia-Ruíz**,[22] **Astra Clelia Bertelli**,[6,7] **Balázs Kiss**,[23,‡] **Banu Sebnem Önder**,[24,‡] **Bélen Roldán Matrín**,[25] **Bregje Wertheim**,[26,‡] **Candice Deschamps**,[5,‡] **Carlos E. Arboleda-Bustos**,[27,‡] **Carlos Tinedo**,[18,‡] **Christian Feller**,[28] **Christian Schlötterer**,[29,‡] **Clancy Lawler**,[30] **Claudia Fricke**,[31,‡] **Cristina P. Vieira**,[32,‡] **Cristina Vieira**,[33,‡] **Darren J. Obbard**,[34,‡] **Dorcas Juana Orengo**,[18,19,‡] **Doris Vela**,[35] **Eduardo Amat**,[36] **Elgion Loreto**,[37] **Envel Kerdaffrec**,[38] **Esra Durmaz Mitchell**,[38,‡] **Eva Puerma**,[39,‡] **Fabian Staubach**,[40] **M. Florencia Camus**,[41,‡] **Hervé Colinet**,[42,‡] **Jan Hrcek**,[43,‡] **Jesper Givskov Sørensen**,[44,‡] **Jessica Abbott**,[45,‡] **Joan Torro**,[46] **John Parsch**,[21,‡] **Jorge Vieira**,[32,‡] **Jose Luis Olmo**,[47] **Khalid Khfif**,[48,‡] **Krzysztof Wojciechowski**,[49] **Lilian Madi-Ravazzi**,[50] **Maaria Kankare**,[51,‡§] **Mads F. Schou**,[44,‡] **Emmanuel D. Ladoukakis**,[52,‡] **M. Josefa Gómez-Julián**,[22] **M. Luisa Espinosa-Jimenez**,[22] **Maria Pilar Garcia Guerreiro**,[16,‡] **Maria-Eleni Parakatselaki**,[52] **Marija Savic Veselinovic**,[53,‡] **Marija Tanaskovic**,[20,‡] **Marina Stamenkovic-Radak**,[53,‡] **Margot Paris**,[38,‡] **Marta Pascual**,[18,19,‡] **Michael G. Ritchie**,[54,‡] **Michel Rera**,[55,‡] **Mihailo Jelić**,[53,‡] **Mina Hojat Ansari**,[40,‡] **Mina Rakic**,[53] **Miriam Merenciano**,[4,‡] **Natalia Hernandes**,[30] **Nazar Gora**,[56] **Nicolas Rode**,[5,‡] **Omar Rota-Stabelli**,[57,‡] **Paloma Sepulveda**,[58] **Patricia Gibert**,[33,‡] **Pau Carazo**,[59,‡] **Pinar Kohlmeier**,[26] **Priscilla A. Erickson**,[2,60] **Renaud Vitalis**,[5] **Jorge Roberto Torres**,[61,‡] **Sara Guirao-Rico**,[18,19,‡] **Sebastian E. Ramos-Onsins**,[62] **Silvana Castillo**,[63] **Tânia F. Paulo**,[64,‡] **Venera Tyukmaeva**,[65,‡] **Zahara Alonso**,[66] **Vladimir E. Alatortsev**,[67,‡] **Elena Pasyukova**,[67,‡] **Dmitry V. Mukha**,[68,‡] **Dmitri A. Petrov**,[69,70,*‡§] **Paul Schmidt**,[71,*‡§] **Thomas Flatt**,[38,*‡§¶] **Alan O. Bergland**,[2,*‡§¶] **Josefa Gonzalez**[3,4,*‡§¶]

[1]Department of Biology, University of Vermont, Burlington, VT, USA
[2]Department of Biology, University of Virginia, Charlottesville, VA, USA
[3]Institut Botànic de Barcelona (IBB) CSIC-CMCNB, Barcelona, Spain
[4]Institute of Evolutionary Biology, CSIC, UPF, Barcelona, Spain
[5]CBGP, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
[6]Natural History Museum, Vienna, Austria
[7]Department of Biology and Biotechnology, University of Pavia, Pavia, Italy
[8]Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Received: November 13, 2024. **Revised:** April 25, 2025. **Accepted:** May 14, 2025
© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026

[9]Department of Zoology, Stockholm University, Stockholm, Sweden
[10]Department of Biology, Johns Hopkins University, Baltimore, MD, USA
[11]Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA
[12]Department of Zoology, Hydrobiology and General Ecology, Odesa I.I. Mechnikov National University, Odesa, Ukraine
[13]National Antarctic Scientific Center of Ukraine, Kyiv, Ukraine
[14]Faculty of Biology and Forestry, Lesya Ukrainka Volyn National University, Lutsk, Ukraine
[15]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[16]Department of Genetics and Microbiology, Facultat de Biociencies, Universitat Autònoma de Barcelona, Barcelona, Spain
[17]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Barcelona, Spain
[18]Departament de Genetica, Microbiologia i Estadistica, Universitat de Barcelona, Barcelona, Spain
[19]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain
[20]Institute for Biological Research, National Institute of the Republic of Serbia, University of Belgrade, Belgrade, Serbia
[21]Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität München, Munich, Germany
[22]Instituto de Enseñanza Secundaria Eladio Cabañero, Tomelloso, Spain
[23]HUN-REN Plant Protection Institute, Centre for Agricultural Research, Budapest, Hungary
[24]Genetic Variation and Adaptation Laboratory, Department of Biology, Hacettepe University, Ankara, Turkey
[25]Instituto de Enseñanza Secundaria Alonso de Ercilla, Toledo, Spain
[26]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands
[27]Neuroscience Group, Genetics Institute, Universidad Nacional de Colombia, Bogotá, Colombia
[28]Landesgymnasium Sankt Afra zu Meißen, Meißen, Germany
[29]Institute of Population Genetics, Vetmeduni Austria, Vienna, Austria
[30]Department of Biosciences, The University of Melbourne, Victoria, Australia
[31]Institute for Zoology, Institute for Biology, Martin-Luther University Halle-Wittenberg, Halle, Germany
[32]Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal
[33]Laboratoire de Biométrie et Biologie Evolutive, CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France
[34]Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK
[35]Faculty of Exact, Natural and Environmental Sciences, Pontificia Universidad Católica del Ecuador, Quito, Ecuador
[36]Bioforense Research Group, Faculty of Law and Forensic Sciences, Tecnológico de Antioquia, Medellin, Colombia
[37]Department of Biochemistry and Molecular Biology, Federal University of Santa Maria, Santa Maria, RS, Brazil
[38]Department of Biology, University of Fribourg, Fribourg, Switzerland
[39]Vall d'Hebron Institute of Oncology, Barcelona, Spain
[40]Department of Evolution and Ecology, University of Freiburg, Freiburg, Germany
[41]Research Department of Genetics, Evolution & Environment, University College London, London, UK
[42]University of Rennes, CNRS, ECOBIO, Rennes, France
[43]Biology Centre of the Czech Academy of Sciences, Institute of Entomology, Ceske Budejovice, Czech Republic
[44]Department of Biology, Aarhus University, Aarhus, Denmark
[45]Department of Biology, Lund University, Lund, Sweden
[46]Instituto de Enseñanza Secundaria Benjamín Jarnés, Zaragoza, Spain
[47]Instituto de Enseñanza Secundaria Azuer, Ciudad Real, Spain
[48]Entomology Laboratory, URTNEQ, INRA, Tangier, Morocco
[49]Administration of Regional Landscape Parks of Lublin, Voivodeship, Chelm, Poland
[50]Institute of Biosciences, Humanities, and Exact Sciences, Sao Paulo State University, Sao José do Rio Preto, Brazil
[51]Department of Biological and Environmental Science, University of Jyvaskyla, Jyvaskyla, Finland
[52]Department of Biology, University of Crete, Heraklion, Greece
[53]Faculty of Biology, University of Belgrade, Belgrade, Serbia
[54]School of Biology, University of St Andrews, St Andrews, UK
[55]Institut Jacques Monod, Paris, France
[56]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Lab, Berkeley, CA, USA
[57]Center Agriculture Food Environment (C3A), University of Trento, Trento, Italy
[58]Instituto de Enseñanza Secundaria Carpetania, Toledo, Spain
[59]Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Valencia, Spain
[60]Department of Biology, University of Richmond, Richmond, VA, USA
[61]La Ciència Al Teu Món, Barcelona, Spain
[62]Centre for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Barcelona, Spain
[63]Instituto de Enseñanza Secundaria Jose de Mora, Granada, Spain
[64]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[65]Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, UK
[66]Centro de Educación Infantil y Primaria Ramón y Cajal, Zaragoza, Spain
[67]Institute of Molecular Genetics of Russian Academy of Sciences, Moscow, Russia
[68]Vavilov Institute of General Genetics of Russian Academy of Sciences, Moscow, Russia
[69]Department of Biology, Stanford University, Stanford, CA, USA
[70]CZ Biohub, San Francisco, CA, USA
[71]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

***Corresponding authors:** E-mails: joaquin.nunez@uvm.edu; marta.coronado@csic.es; dpetrov@stanford.edu; schmidtp@sas.upenn.edu; thomas.flatt@unifr.ch; aob2x@virginia.edu; josefa.gonzalez@csic.es.
[†]These authors contributed equally to the study as co-first authors.
[‡]The European Drosophila Population Genomics Consortium (DrosEU).
[§]The Drosophila Real-Time Evolution Consortium (DrosRTEC).
[¶]These authors contributed equally to the study as co-senior authors.
Associate editor: Yuseob Kim

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026

Abstract

Large-scale genomic resources can place genetic variation into an ecologically informed context. To advance our understanding of the population genetics of the fruit fly *Drosophila melanogaster*, we present an expanded release of the community-generated population genomics resource *Drosophila Evolution over Space and Time* (DEST 2.0; https://dest.bio/). This release includes 530 high-quality pooled libraries from flies collected across six continents over more than a decade (2009 to 2021), most at multiple time points per year; 211 of these libraries are sequenced and shared here for the first time. We used this enhanced resource to elucidate several aspects of the species' demographic history and identify novel signs of adaptation across spatial and temporal dimensions. For example, we showed that the spatial genetic structure of populations is stable over time, but that drift due to seasonal contractions of population size causes populations to diverge over time. We identified signals of adaptation that vary between continents in genomic regions associated with xenobiotic resistance, consistent with independent adaptation to common pesticides. Moreover, by analyzing samples collected during spring and fall across Europe, we provide new evidence for seasonal adaptation related to loci associated with pathogen response. Furthermore, we have also released an updated version of the DEST genome browser. This is a useful tool for studying spatiotemporal patterns of genetic variation in this classic model system.

Keywords: *Drosophila melanogaster*, dataset, population structure, seasonal selection, local adaptation, ecological genomics

Introduction

Drosophila melanogaster is a foundational model system in biology. Seminal studies in this species have played important roles in the development of modern population genetics, from empirical tests of genetic drift to classic examples of adaptation (Buri 1956; Lewontin 1974; Parsons 1975; McDonald and Kreitman 1991; Powell 1997; Casillas and Barbadilla 2017; Flatt 2020). Beyond its role as a model genetic system (Hales et al. 2015), *D. melanogaster* has a fascinating natural history in its own right. The species originated in southern-central Africa (Lachaise et al. 1988; Lachaise and Silvain 2004; Sprengelmeyer et al. 2020), splitting from its sister taxon, *Drosophila simulans*, between 1.4 and 3.6 Ma (Obbard et al. 2012; Suvorov et al. 2022). While the species may have originally been a marula fruit specialist in the seasonal woodlands of southern-central Africa (Mansourian et al. 2018; Sprengelmeyer et al. 2020), it later adapted as a human commensal, ultimately developing a cosmopolitan distribution across all human-inhabited continents (Kapun et al. 2021; Chen et al. 2024).

The recent development of genomic resources for *D. melanogaster* has led to key discoveries about its phylogeography. For example, demographic inference has revealed that fruit flies expanded out of Africa after the last glacial maximum ~10,000 ya (Kapopoulou et al. 2020), entering Asia around 3 to 4 kya (Chen et al. 2024), and Europe around ~1,800 ya (Sprengelmeyer et al. 2020). European populations split into spatially defined clusters across Europe ~1,000 ya (Kapun et al. 2020, 2021). In the past two centuries, African and European populations experienced a secondary contact event in North America and Australia, likely due to mercantile activities and immigration (Capy et al. 1986; David and Capy 1988; Caracristi and Schlötterer 2003; Kao et al. 2015; Bergland et al. 2016). Unlike its sister species *D. simulans*, *D. melanogaster* is capable of overwintering across a broad swath of temperate habitats (Izquierdo 1991; Machado et al. 2016; but see Serga et al. 2015) and can establish resident populations across its range (Ives 1945, 1970; Machado et al. 2016; Kapun et al. 2021; Nunez et al. 2024). In temperate regions, *D. melanogaster* populations peak in size at least twice during the growing season (June and September; Atkinson and Shorrocks 1977) and drastically decrease upon the onset of winter. These yearly boom-and-bust cycles are in part responsible for estimates of "local" effective population size that are orders of magnitude smaller than the "global" effective population size (Duchen et al. 2013; Sprengelmeyer et al. 2020; Nunez et al. 2024).

Over the past two decades, *D. melanogaster* has been the subject of numerous population genomics studies, which have collectively illuminated our general understanding of the evolution, the demography, and the genetic basis of adaptation (e.g. reviewed in Casillas and Barbadilla 2017; Haudry et al. 2020; Guirao-Rico and González 2019). Like many other cosmopolitan drosophilids, *D. melanogaster* populations commonly occur along spatially distributed environmental gradients (e.g. latitudinal and altitudinal), leading to the formation of clines, with a large body of work providing evidence for spatially varying (clinal) selection (De Jong and Bochdanovits 2003; Hoffmann and Weeks 2007; Fabian et al. 2012; Adrion et al. 2015; Mateo et al. 2018; Flatt 2020). Moreover, populations of *D. melanogaster* are known to experience strong fluctuating selection regimes across the changing seasons (e.g. Schmidt and Conde 2006; Bergland et al. 2014; Behrman et al. 2015; Rajpurohit et al. 2018; Erickson et al. 2020; Machado et al. 2021; Rudman et al. 2022; Nunez et al. 2024; reviewed in Johnson et al. 2023). For example, worldwide analyses of genetic variation have found that chromosomal inversion polymorphisms are often involved in both clinal and seasonal adaptation (Lemeunier and Aulard 1992; Kapun et al. 2016a, 2023; Kapun and Flatt 2019; Nunez et al. 2024). Likewise, several studies have successfully linked clinally or seasonally varying polymorphisms in *D. melanogaster* to fitness-relevant phenotypes (Lemeunier and Aulard 1992; Hoffmann and Weeks 2007; Schmidt et al. 2008; Pitchers et al. 2013; Cogni et al. 2014; Paaby et al. 2014; Kapun et al. 2016a, 2016b, 2023; Durmaz et al. 2019; Kapun and Flatt 2019; Betancourt et al. 2021; Yu and Bergland 2022; Glaser-Schmitt et al. 2024; Nunez et al. 2024). Populations of *D. melanogaster* can thus be thought of as powerful "natural laboratories" to study adaptation across time and space, and to disentangle the contributions of selection and demography (Jensen et al. 2005; Ometto et al. 2005; Teshima et al. 2006; Thornton and Jensen 2007; Pavlidis et al. 2010).

Despite the status of *D. melanogaster* as a model organism, generating genomic datasets that capture the breadth and depth of genetic and phenotypic variation across the cosmopolitan range of the species is a complex task for single research groups. Furthermore, existing data for this species are heterogeneous across studies: several studies use resequenced inbred or isofemale lines (Langley et al. 2012; Mackay et al. 2012; Lack et al. 2015, 2016; Coughlan et al. 2022), while others use sequencing of outbred individuals sequenced as a pool (i.e. Pool-Seq, Schlötterer et al. 2014; e.g. Bergland et al. 2014; Machado et al. 2021; Nunez et al. 2024). For these reasons, we have previously developed the *Drosophila Evolution over Space and Time* (DEST; https://dest.bio/) resource, with the aim of facilitating collaborative population genomic studies in *D. melanogaster* (Kapun et al. 2021). The DEST resource is the result of the collaborative efforts of the European *Drosophila* Population Genomics Consortium

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026

(DrosEU, https://droseu.net/; Kapun et al. 2020) and the *Drosophila* Real-Time Evolution Consortium, DrosRTEC (Machado et al. 2021). DEST represents both a tool for mapping genomic data and an open-access data repository of worldwide genetic variation in the fruit fly. As a bioinformatics tool, DEST is a pipeline for mapping Pool-Seq reads to a hologenome reference of fly (i.e. *D. simulans* and *D. melanogaster*) and microbial genomes, as well as for removing contamination from other species, such as *D. simulans*. The tool is a modular mapping pipeline that uses a Docker image (Boettiger 2015) and *Snakemake* (Köster and Rahmann 2012) to ensure independence of operating systems. As a genomic panel, the original release of the dataset (DEST 1.0) consisted of 271 Pool-Seq *D. melanogaster* samples (>13,000 flies) collected in >20 countries on four continents at different seasons and across multiple years. Using these data, we had previously described general patterns of phylogeographic structure across four continents, developed a panel of geographically informative markers (GIMs) to assess the provenance of fly samples with 90% accuracy, and inferred some basic demographic features of population subdivision in Europe (Kapun et al. 2021).

Here, we introduce the second release of the DEST resource (DEST 2.0), with expansions in several methodological and biological aspects. From a methodological perspective, we have broadened the utility of our Docker application to allow for single–end reads to be mapped, a change that allows for older datasets to be integrated into DEST. We have explored levels of contamination by other species in DEST pools using a highly efficient *k*-mer-based approach (Gautier 2023). We have also estimated genome-wide rates of recombination using our Pool-Seq data by applying a deep learning approach (*ReLERNN*; Adrion et al. 2020). All data on genetic variation and population genetic summary statistics can be visualized and retrieved using our new and improved genome browser, which has been built with the latest JBrowse version 2 (Diesh et al. 2023).

From a biological standpoint, DEST 2.0 includes a substantial expansion of the size and scope of the initial dataset. The current release includes 530 high-quality Pool-Seq samples (>32,000 flies), comprising a combination of the previous DEST release with newly sequenced pools, collected between 2016 and 2021 by DrosEU, as well as publicly available Pool-Seq samples from published studies of wild-derived *D. melanogaster* (Hoffmann et al. 2002; Reinhardt et al. 2014; Svetec et al. 2016; Fournier-Level et al. 2019; Lange et al. 2022; Nunez et al. 2024). To showcase the utility of DEST 2.0, we performed several analyses to infer demography and selection, powered by the rich spatial and temporal density of our dataset. Below, we divide these analyses into two general categories: "spatial insights" and "temporal insights." For each category, we highlight case studies of demographic inference and genome-wide scans for adaptive differentiation. Our analyses provide novel insights into patterns of demography and selection of natural *D. melanogaster* populations and generate hypotheses that can be tested with the power of the *Drosophila* genetics toolbox in future work. In general, our work illustrates the value of DEST 2.0 as an open resource for the *Drosophila* evolutionary genetics and genomics community.

Results

DEST 2.0, an Expanded *Drosophila* Population Genomics Resource

The current DEST release (version 2.0) includes 530 high-quality samples as well as an additional 207 pools that fell below our quality thresholds and were excluded from the analysis (see supplementary table S1, Supplementary Material online). In its totality, the 737 pooled libraries originated from multiple sources, including both releases of the DEST dataset (i.e. 1.0 and 2.0), the *Drosophila* Genome Nexus (DGN; Lack et al. 2016; including one sample from *D. simulans*; see Fig. 1a), as well as from previous publications (Hoffmann et al. 2002; Reinhardt et al. 2014; Svetec et al. 2016; Fournier-Level et al. 2019; Lange et al. 2022; Nunez et al. 2024). The 737 samples within DEST 2.0 vary in sequencing characteristics, ranging from a read depth (RD) of 4× to 300× and from an effective haploid sample size (n_e; the sample size accounting for pool size and Pool-Seq sampling effects) of 3.7 to 77.2 (supplementary fig. S1 and text S1, Supplementary Material online; see Kolaczkowski et al. 2011; Feder et al. 2012; Gautier et al. 2013).

To ensure the highest possible quality of each sample, we calculated a suite of sequencing statistics, including the PCR duplication rate, fraction of missing data, coverage, and number of private SNPs across the totality of the dataset (all 737 pools). In addition, we also estimated the p_N/p_S statistic (i.e. the ratio of the number of genome-wide nonsynonymous polymorphisms to the number of genome-wide synonymous polymorphisms, as in Kapun et al. 2021; supplementary fig. S2, Supplementary Material online), and assessed non-*D. melanogaster* contamination through competitive mapping and *k*-mer approaches (Kapun et al. 2021; Gautier 2023; supplementary fig. S3, Supplementary Material online). Next, we used a principal component analysis (PCA) on all quality control metrics to assess whether samples should be included or excluded from downstream analyses (see Fig. 2a and supplementary fig. S4, Supplementary Material online; see the "Estimation of Nucleotide Diversity" section in Materials and Methods). Finally, 136 samples that consisted of multiple replicates from the same locality, each with low coverage, were collapsed into a single sample. For a more detailed description of data filtering procedures and recommendations for users, see supplementary text S2, Supplementary Material online. Based on the results of these analyses, we obtained a final high-quality dataset of 530 samples and 4,789,696 SNPs across autosomes and the X chromosome for downstream analyses. The high-quality dataset contains representative samples from 45 countries across all continents (22 from Africa, 40 from Asia, 302 from Europe, 141 from North America, 17 from Australia, and 7 from South America; Fig. 1a) and across a time span of 12 years (2009 to 2021). In total, our 530 high-quality samples represent 164 localities; of these, 112 were sampled only in 1 year (68%), 18 were sampled across 2 years (11%), and the rest (34; 21%) were sampled multiple times across several years (Fig. 1b). Overall, descriptions and basic subsetting of SNP statistics for DEST 2.0 are shown in Table 1. Unless stated otherwise, all of the following analyses are based on the 530 high-quality samples.

Estimates of Nucleotide Diversity and Recombination Rates

To describe patterns of genetic variation in the DEST 2.0 data, we analyzed nucleotide diversity π (Tajima 1983, 1989) estimated with *npStat* (Ferretti et al. 2013). This analysis was conducted on a subset of 504 samples with masked BAM files (see the "Masked gSYNC Files" section in Materials and Methods). As previously observed (Begun and Aquadro 1993; Andolfatto 2001; Mackay et al. 2012; Kapun et al. 2021; Coughlan et al. 2022), we found that sub-Saharan African populations had

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026

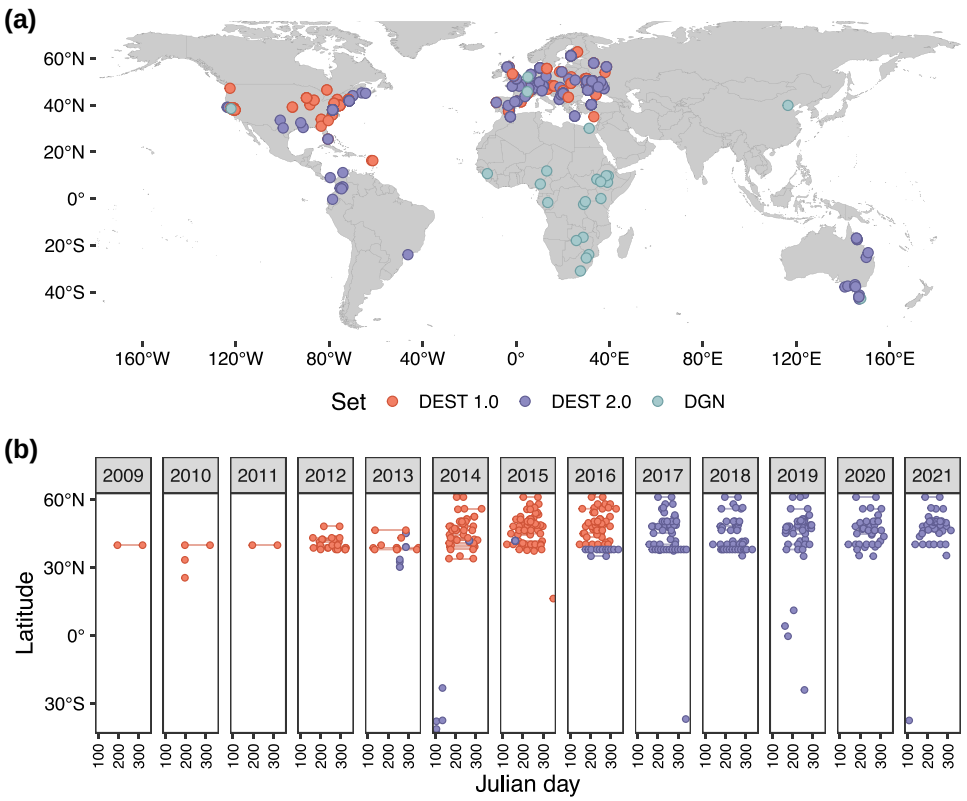


Fig. 1. Spatial and temporal scales of DEST. a) World map showing 530 high-quality samples part of DEST 1.0 (Kapun et al. 2020), DEST 2.0 (this study), and the DGN (Lack et al. 2016). b) Sampling density across years of sampling contained in the DEST dataset. The colors are consistent with (a).

higher levels of genetic variation than other populations (Fig. 2b), consistent with out-of-Africa demography (Li and Stephan 2006; Lack et al. 2016; Arguello et al. 2019; Kapopoulou et al. 2020; Kapun et al. 2021).

We inferred levels of genome-wide recombination across 74 samples representative of the populations analyzed (see the "Recombination Landscape" section in Materials and Methods), using the deep learning method *ReLERNN* (Adrion et al. 2020; see Fig. 2c; supplementary fig. S5, Supplementary Material online). Overall, recombination rate is highly heterogeneous among samples and among chromosomes (two-way ANOVA, $F_{73,292} = 19.7$, $P < 1.0 \times 10^{-25}$, and $F_{4,292} = 1{,}599.4$, $P < 1.0 \times 10^{-25}$, respectively; Tukey's HSD tests, all pairwise comparisons between chromosomes $P < 1.0 \times 10^{-7}$, except for 3R vs. 2R, where $P = 0.074$). In most populations, there is a significant positive correlation between recombination rate and genetic diversity, consistent with recurrent genetic hitchhiking and background selection (Begun and Aquadro 1993; supplementary table S2, Supplementary Material online).

The presence of common cosmopolitan inversions had a noticeable impact on the recombination landscape. Average recombination rates were significantly lower around the inversion breakpoints for six out of the seven inversions analyzed (Wilcoxon test, $P < 0.01$; for inversions *In(2L)t*, *In(2R)NS*, *In(3L)P*, *In(3R)Payne*, *In(3R)C* and *In(3R)K*; supplementary table S3, Supplementary Material online). Recombination was also lower for those regions spanning five of these inversions than for the rest of the chromosome (Wilcoxon test, $P < 0.01$; for inversions *In(2L)t*, *In(2R)NS*, *In(3R)Payne*, *In(3R)C*, and *In(3R)K*; supplementary table S3, Supplementary Material online).

PCAs showed that populations belonging to the same geographic region share similar recombination landscapes (supplementary fig. S6 and table S1, Supplementary Material online for metadata). The geographic pattern is more evident when considering relative values of recombination, i.e. the ratio of the average recombination rate of each window to the average recombination across the respective chromosome, and is therefore informative on the recombination landscape rather than the absolute recombination rate (compare panels A and B with panels C and D in supplementary fig. S6, Supplementary Material online).

Spatial Population Structure is Defined by Latitudinal and Longitudinal Clines

To investigate patterns of population structure in the DEST 2.0 dataset, we performed PCA on all 530 samples that passed quality filters. We used biallelic SNPs from the euchromatic regions of the four major autosome arms (Fig. 3a; also see supplementary fig. S7, Supplementary Material online). When all autosomes are considered, PC1 divides samples from sub-Saharan Africa from all other continents. At the level of individual regions, PC1 is correlated with both latitude and longitude in North America ($r = -0.7$; $P < 2 \times 10^{-16}$ and $r = -0.60$; $P < 2.2 \times 10^{-16}$, respectively) and longitude in Europe ($r = -0.80$; $P < 2.2 \times 10^{-16}$; Fig. 3b and c). These patterns of population structure were consistent with previously published studies (Kapun et al. 2020, 2021; Machado et al. 2021). Notably, both PC1 and PC2 primarily divided African samples from all other clusters, and PC2 also separated samples in Europe from samples in North America, South America, and Australia.

The patterns seen across chromosome-specific PCA were strongly correlated to that of the other chromosomes for both PCs 1 and 2 ($r_{2L\text{-other chr.}} = 0.906$ to 0.957, $r_{2R\text{-other chr.}} = 0.933$

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026

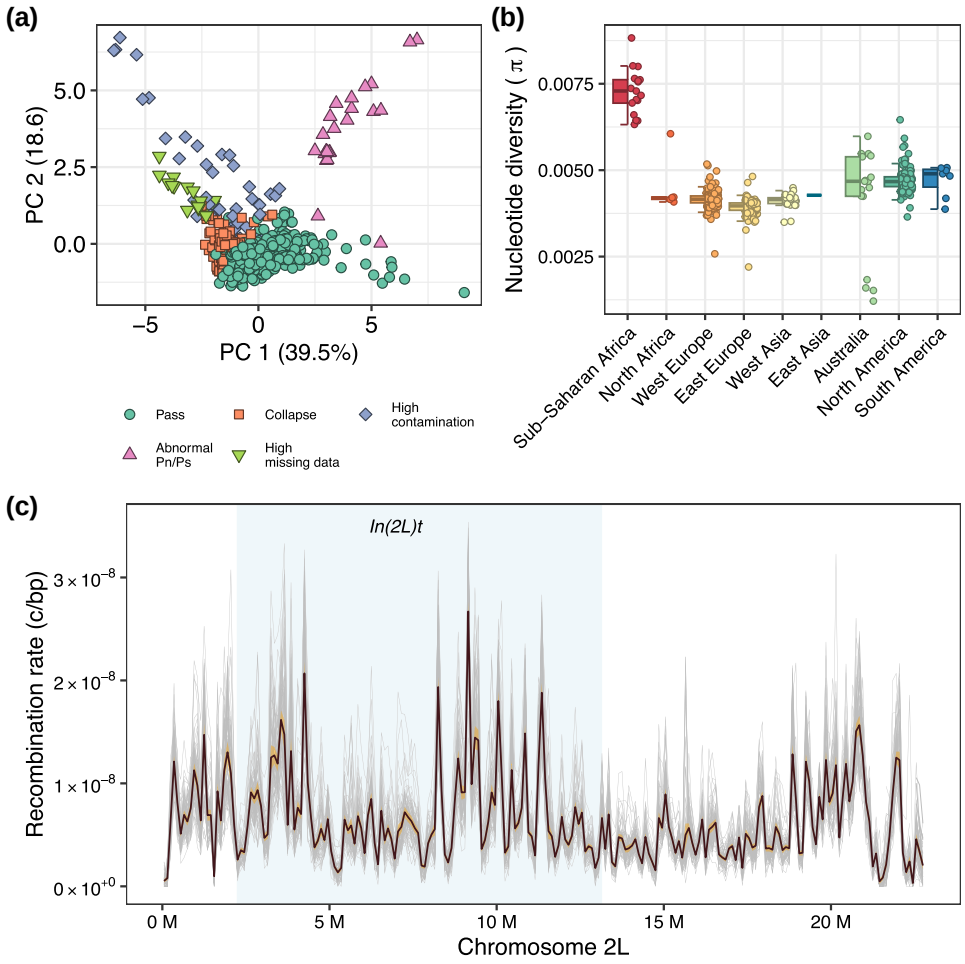


Fig. 2. Patterns of filtering, genetic variation, and recombination in DEST 2.0. a) Visualization of filtering information of samples using PCA. Each dot is a sample's QC metric, and the color indicates the filtering decision (legend: Pass: samples that pass the filter and are used in downstream analyses; Collapse: biological and/or technical replicates collapsed into a single representative sample; otherwise, samples were excluded due to abnormal p_N/p_S levels of high levels of missing data or contamination). b) Autosomal nucleotide diversity (π) calculated across continents (see the "Estimation of Nucleotide Diversity" section in Materials and Methods for details). c) Recombination landscape of chromosome 2L in samples representative of the 74 *D. melanogaster* populations analyzed (one gray line per sample). Light blue area highlights the region spanning the *In(2L)t* inversion. Average (black line) and overall distribution envelope (orange shaded ribbon; delineated by the average values ±1.96 SD) are shown.

Table 1 SNP calling information for DEST 2.0 across major autosomes and chromosome X

SNP type	*2L*	*2R*	*3L*	*3R*	*X*
Total (all)	1,080,586	901,878	1,069,441	1,212,752	525,039
Bi-allelic	1,048,510	877,852	1,039,460	1,182,310	516,077
Inside inversions	569,713	228,826	631,556	159,598	NA
In recombining regions ($c > 0$)	997,162	836,457	976,915	1,074,768	482,162
Protein coding	796,420	731,794	793,866	944,372	40,4881
Intergenic	828,039	659,966	824,903	929,539	401,586
Synonymous	95,275	91,052	90,635	101,504	49,055
Nonsynonymous	71,534	75,921	72,843	90,905	25,072
Proportion of missing data	0.0511	0.0507	0.0508	0.0493	0.0533

SNPs inside the inversion are estimated for *In(2L)t* for *2L*, *In(2R)NS* for *2R*, *In(3L)P* for *3L*, and the joint region among *In(3R)K*, *In(3R)P*, and *In(3R)Mo*. Estimated recombination rates (i.e. rate of cross-over; *c*"). Functional annotations are only reported for biallelic sites.

to 0.967, $r_{3L\text{-other chr.}} = 0.912$ to 0.967, $r_{3R\text{-other chr.}} = 0.906$ to 0.967; note that all *P*-values are $<1.0 \times 10^{-15}$). PC3 is peculiar in that the whole-genome results were similar only for the 2R-3L comparison ($r = -0.923$; $P < 1.0 \times 10^{-15}$) and the 2L-3R comparison ($r = 0.287$; $P = 1.77 \times 10^{-11}$), but not for the other comparisons (supplementary table S4, Supplementary Material online). Notably, we observe that patterns in PC3, specifically within 2L and 3R, were strongly influenced by the frequencies of *In(2L)t* and *In(3R)Payne*, two large adaptive cosmopolitan inversion polymorphisms (Kapun et al. 2023; Nunez et al. 2024).

We investigated clines in the frequencies of cosmopolitan inversion polymorphisms in DEST 2.0 using inversion-specific SNPs that are in strong linkage disequilibrium with the inversion breakpoints (Kapun et al. 2014; supplementary fig. S8, Supplementary Material online). Many inversions showed

Downloaded from academic.oup.com/mbe/article/42/8/msaf132/8237493 by guest on 09 October 2026