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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

[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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

(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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

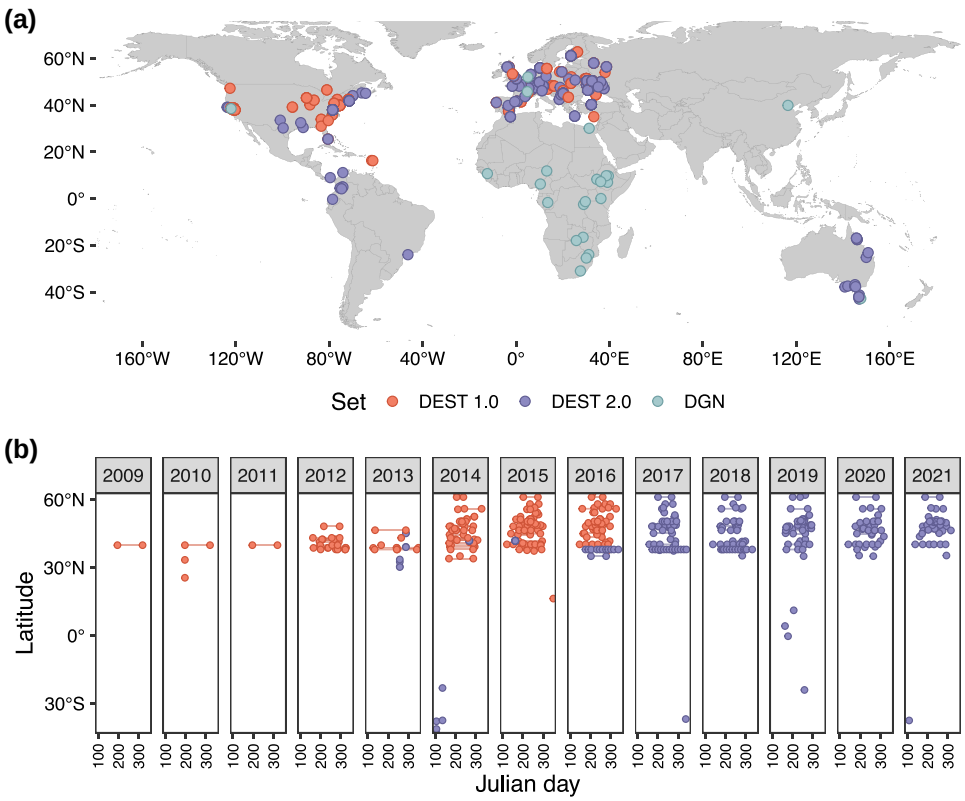

Fig. 1. Spatial and temporal scales of DEST. a) World map showing 530 high-quality samples part of DEST 1.0 (Kapun et al. 2020), DEST 2.0 (this study), and the DGN (Lack et al. 2016). b) Sampling density across years of sampling contained in the DEST dataset. The colors are consistent with (a).

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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

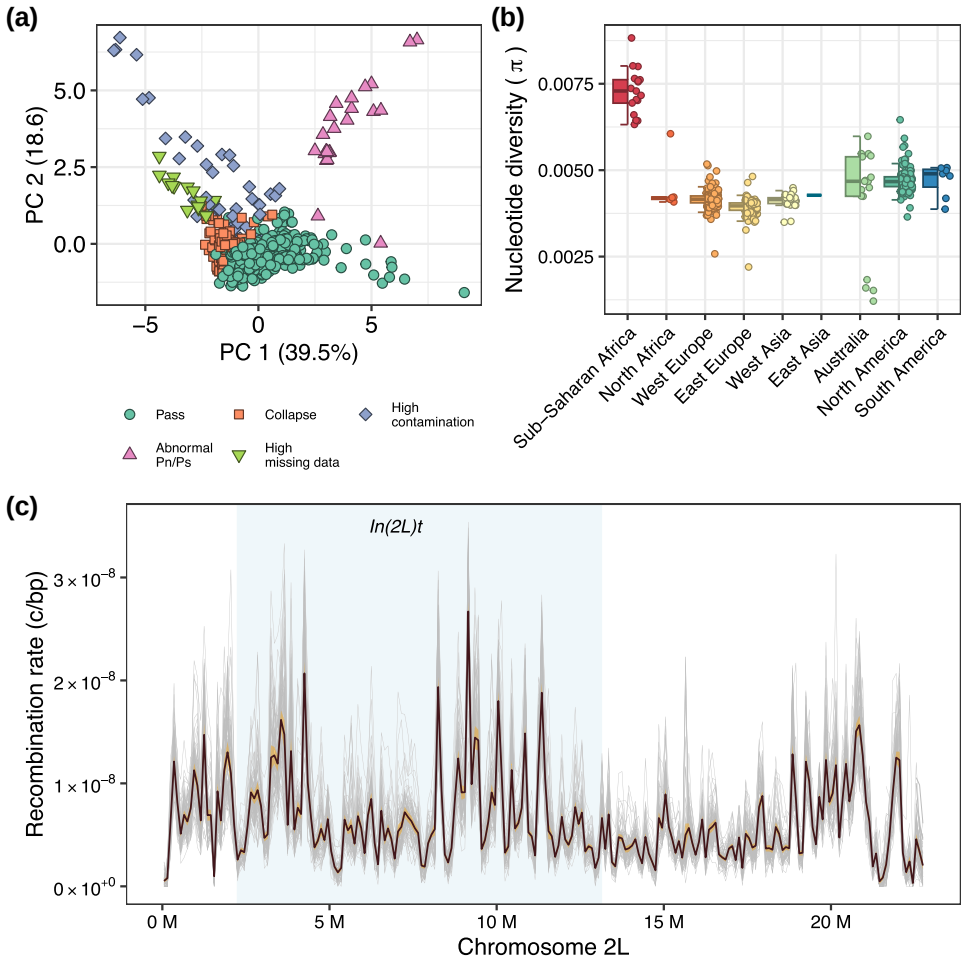

Fig. 2. Patterns of filtering, genetic variation, and recombination in DEST 2.0. a) Visualization of filtering information of samples using PCA. Each dot is a sample's QC metric, and the color indicates the filtering decision (legend: Pass: samples that pass the filter and are used in downstream analyses; 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 https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

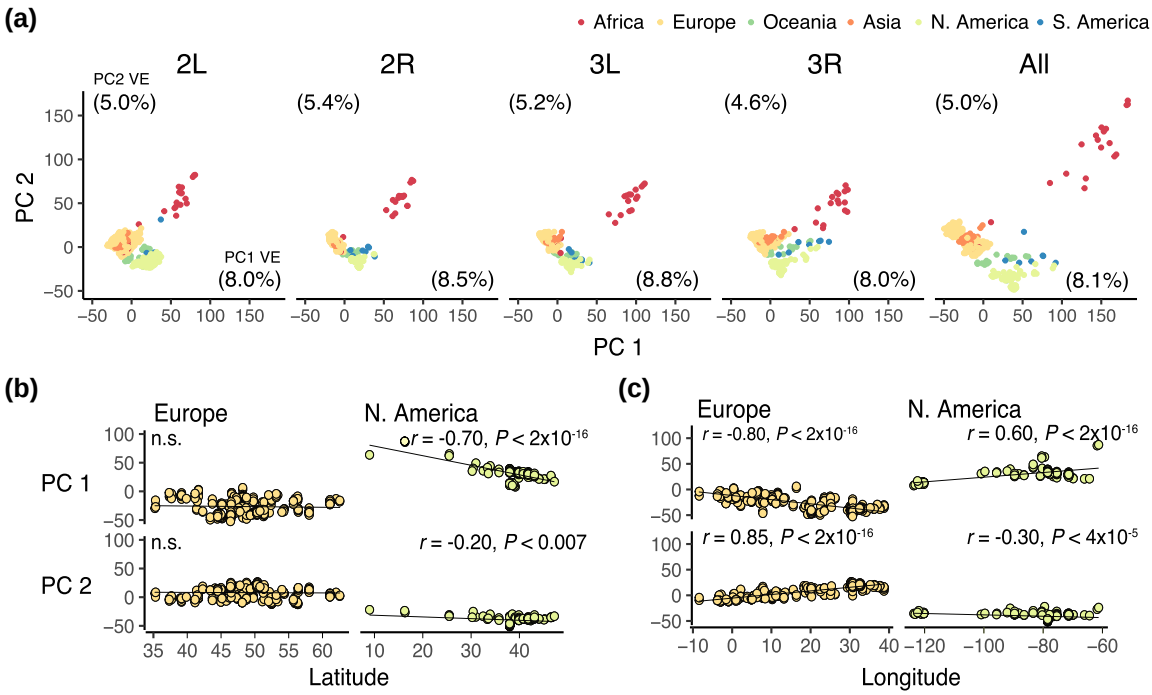

Fig. 3. PCA and projections. a) PCA projections showing PCs 1 and 2. Analyses were done for each chromosome arm and all arms combined. The proportion of variance explained (VE) is shown at the corners of each axis. b) Projections of PCs 1 and 2 relative to latitude for Europe and North American pools. c) Same as (b) but for longitude. Notice that, in this analysis, Asia refers primarily to samples from Turkey (which is located in Western Asia).

significant clinal patterns along latitude or longitude that were consistent across different continents (see supplementary table S5, Supplementary Material online for statistical details). Our results are in line with previous observations, in particular for *In(3R)Payne* (Lemeunier and Aulard 1992; Anderson et al. 2005; Kapun et al. 2016b, 2020, 2023; Kapun and Flatt 2019), which showed significant latitudinal clines in North America, Europe, and along the Australian east coast. Latitudinal clines were also significant for *In(2L)t* and *In(3R)Mo* in North America and Australia and for *In(2R)NS* and *In(3L)P* in North America, Australia, and Europe. Additionally, while overall not being very frequent, *In(2R)NS* exhibited a highly significant longitudinal cline across European populations.

Characterizing Population Structure in European and North American Populations

We applied *k*-means clustering analysis on the first three autosomal PCs to identify spatially defined clusters. First, with $k = 4$ clusters, we fully recapitulated the results of DEST 1.0 (Fig. 4a), with clusters composed of sub-Saharan African samples, the Americas, and two clusters in Europe [as in Kapun et al. 2021; "Europe West" (EU-W) and "Europe East" (EU-E)]. North African and West Asian samples clustered with EU-W. Australian samples were split between the clusters that contain Western Europe and the Americas. We also estimated population clusters using $k = 8$, which was estimated to be the optimal value based on the gap statistic (Tibshirani et al. 2001; Fig. 4b-inset). For $k = 8$, new hypotheses of structure emerged (Fig. 4b). In Europe, the previously known EU-W and EU-E clusters appeared, separated by a putative third cluster at the boundary between EU-E and EU-W (i.e. an "overlapping zone"; Fig. 4c). Newer populations (namely the Americas and Australia), previously grouped as a single cluster, were divided into three clusters: the Caribbean and most of South America, a southeast US coastal group, and all other samples from mainland North America (see green, yellow, and pink points, respectively, in Fig. 4b). Notably, samples from Australia do not show any new levels of clustering when $k = 8$, relative to $k = 4$. Instead, they retain their original cluster association, whereby samples from the south of the continent cluster with samples from EU-W, and those from the north cluster with North American populations (Fig. 4a and b).

While the gap statistic showed that $k = 8$ was the optimal number of clusters, the difference between $k = 4$ and $k = 8$ was marginal (see Fig. 4b-inset). To test whether the additional clusters provide novel biological insights, we assessed the support for each set of clusters using model-based demographic inference with *moments* (Jouganous et al. 2017). Specifically, we aimed to determine whether these clusters represent distinct population introductions into Europe (in the case of the overlap zone, potentially indicating an alternative out-of-Africa migration) or into the Americas (in the case of the Caribbean cluster), relative to the known demographic clusters at these sites. We fit neutral models of population history that we call "one-population," "split," "admixture," and "two-splits" (see supplementary fig. S9, Supplementary Material online; see description in the "Demographic Model Selection with Moments" section in Materials and Methods) to subsets of the DEST 2.0 variant data consisting of the clusters identified in $k = 8$ (supplementary table S6, Supplementary Material online). We used this framework primarily for model selection among the proposed demographic histories, rather than for estimating specific demographic parameters. We expect the analysis to remain robust across model selection, and note that the models we compared mainly differ in the topology of the graph summarizing population histories (see Discussion).

First, we fit the "one-population" and two-population "split" models to the Southeastern North America clusters (i.e. cluster 6 vs. cluster 4 in Fig. 4b) to conclude

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

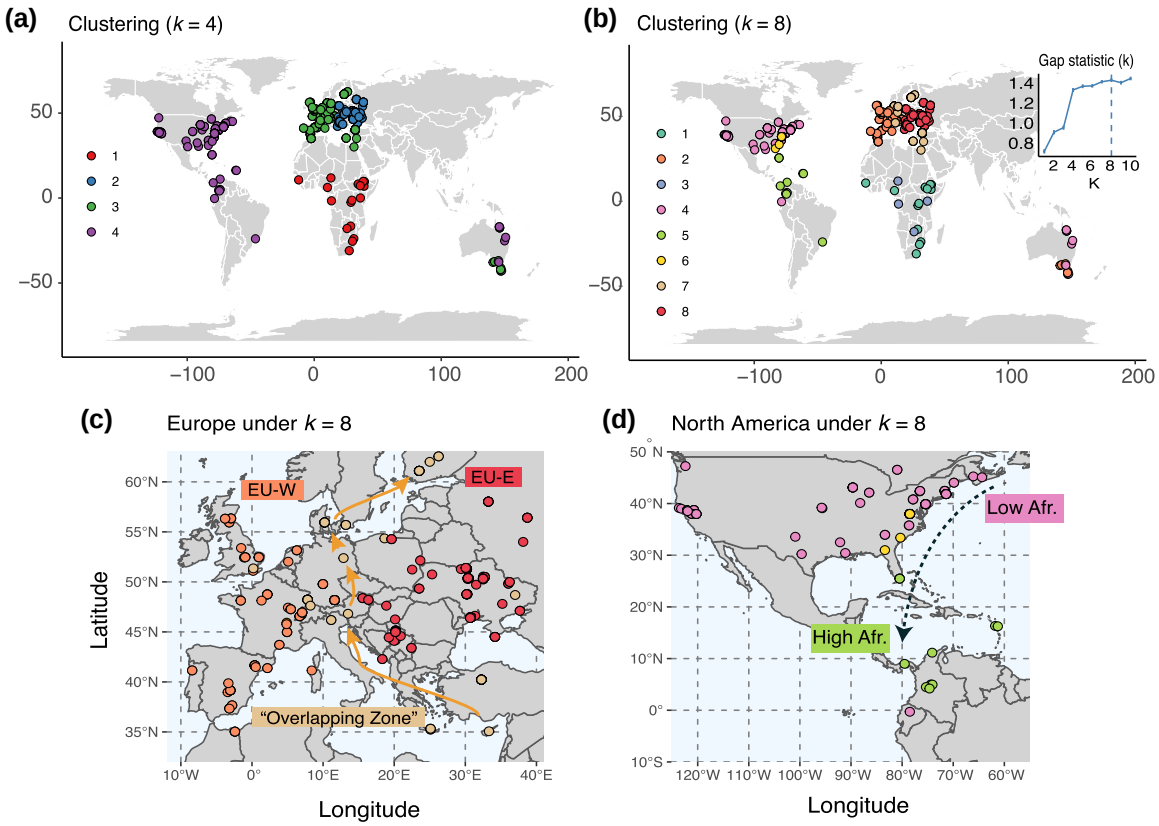

Fig. 4. Spatial population structure and admixture in worldwide *Drosophila*. a) Clustering map, based on PCA projections 1 to 3 built using $k=4$ (as reported in DEST 1.0). b) Same as (a) but with $k=8$ (the optimal number of clusters as defined by a heuristic Gap statistic search, as shown in the inset). c) Zoom view of $k=8$ into Europe to show the hypothetical overlap zone. d) Zoom view of $k=8$ into North America showing the hypothetical "Latin America" cluster (green) and Southeast cluster (yellow).

that "one-population" better describes the region (Wilcoxon signed-rank test on model likelihoods, $P=7.02\times10^{-7}$; supplementary fig. S10a, Supplementary Material online). This result, in which there is no strong evidence of historic divergence between the two clusters, along with low F_{ST} (0.034), supports the parsimony of clustering at $k=4$. Similarly, our analysis also showed that Caribbean populations (cluster 5 in Fig. 4b) are not a distinct cluster relative to the rest of the Americas (clusters 4 and 6 in Fig. 4b; supplementary fig. S10b, Supplementary Material online).

In Europe, we conducted model comparisons among a two-population "split" model, three variants of the three-population "admixture" model in which EU-W, the overlap region, and EU-E (clusters 2, 7, and 8, respectively, in Fig. 4b) are treated as the admixed population, and three variants of the three-population "two-splits" model in which EU-W, the overlap region, and EU-E are, respectively, treated as a sister group to the other two populations. We found support for the two-population models that do not include the overlap zone as a discrete population (Dunn's tests on model likelihoods, six corrected P-values $<3.3\times10^{-7}$; supplementary fig. S10c, Supplementary Material online). This result and the low three-way F_{ST} (0.036), indicate that only the EU-E and EU-W clusters are distinguished as discrete populations and that the overlap zone may simply be an active area of gene flow between EU-W and EU-E. Overall, these findings suggest that the optimal demographic partitioning of the data coincides with clustering at $k=4$, as reported in the original DEST release (Kapun et al. 2021).

Next, we investigated the signals in the data that may have given rise to the clusters proposed by $k=8$. We focused our analyses on the role of African–European admixture in the samples, as this is a primary driver of standing genetic variation in recently expanded populations. To achieve this, we estimated admixture levels in the Americas and Australian populations using a two-pronged approach: a model-based method implemented with *moments*, and a linear modeling approach that has been previously applied in *Drosophila*, as described by Alkorta-Aranburu et al. (2012) and Bergland et al. (2016; see supplementary dataset S1a and b, Supplementary Material online). Results from both methods are highly correlated ($r=0.9$, $P<2\times10^{-16}$; Fig. 5a); however, the linear method consistently produces higher estimates of African ancestry relative to *moments* [concordance correlation $\rho=0.59$, 95% confidence interval (CI) = 0.53 to 0.66]. We investigated this systematic discordance among the inference methods using population genetic simulations with SLiM (Haller and Messer 2023; see "Population Genetics Simulations with SLiM" section in Materials and Methods). We simulated a stepping-stone model resulting from secondary contact and evaluated the performance of the two admixture-estimation methods. Both methods capture the overall clinal pattern of admixture; however, neither consistently recovers the true ancestry proportions (correlation with simulated ancestry: $r_{Linear}=0.79$, $r_{moments}=0.80$, both $P<2.0\times10^{-16}$; supplementary fig. S11a, Supplementary Material online). However, the two methods exhibit distinct error profiles. The *moments*-based approach is generally more precise, consistently overestimating ancestry

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

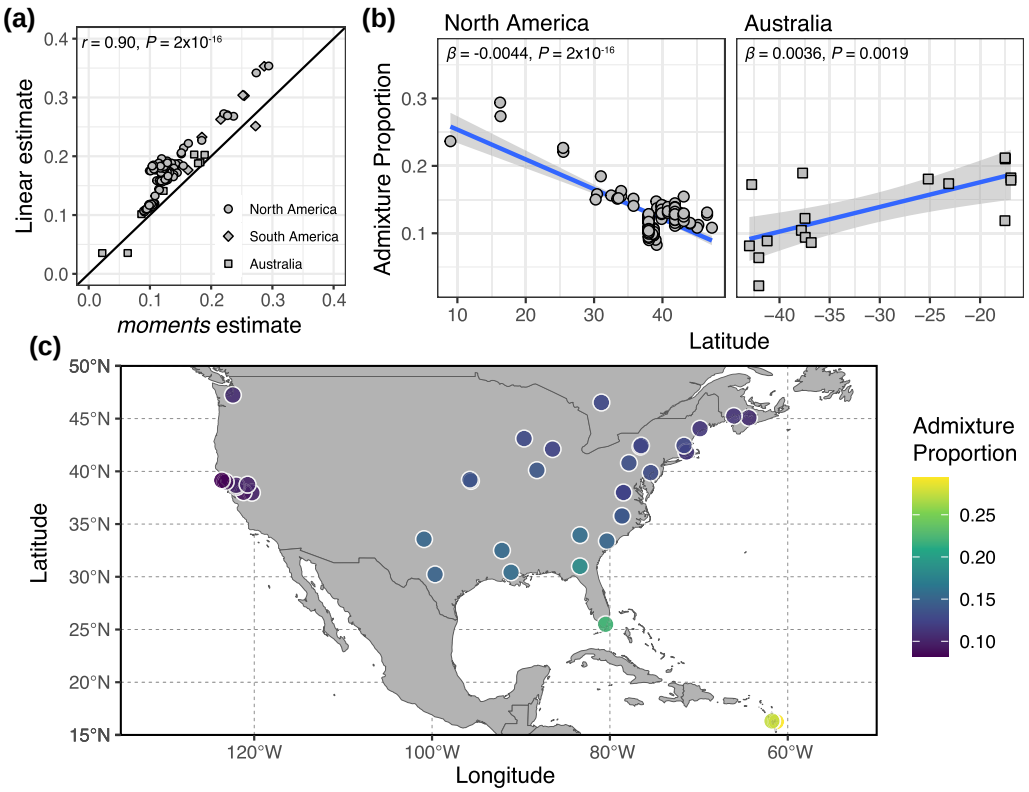

Fig. 5. Patterns of admixture across the Americas and Australia. a) Correlation between admixture estimates obtained using *moments* as well as with the linear modeling method. b) Coefficients of linear admixture for Australia and North America, inferred using *moments*. c) Map projection of levels of African ancestry in North American samples (*moments* estimate).

by only 2% to 10% on average. In contrast, the linear modeling approach shows greater variability, either overestimating or underestimating ancestry depending on simulation parameters, with errors ranging from 10% to 20% (supplementary fig. S11b, Supplementary Material online). We also evaluated the slope of the ancestry cline across the stepping-stone model ($\beta_{ancestry}$) for both methods. Consistent with the ancestry proportion results, both methods underestimated the steepness of the ancestry cline (supplementary fig. S11c, Supplementary Material online); however, the *moments*-based approach more closely approximated the true slope observed in the simulations (supplementary fig. S11d and e, Supplementary Material online).

Based on the simulation results, we conducted all subsequent analyses using the *moments*-based estimates. Overall, our estimates of African admixture were consistent with previously published findings (i.e. a significant negative pattern in North America, $\beta_{\text{African anc.}} = -0.0044$, $P < 2.2 \times 10^{-16}$, and a positive correlation between African admixture and latitude in Australia, $\beta_{\text{African anc.}} = 0.0036$, $P = 0.0019$, see Fig. 5b; Bergland et al. 2016; Corbett-Detig and Nielsen 2017; Coughlan et al. 2022). Analyses on South American samples revealed no significant correlation, likely due to limited sample size, yet showed considerable variation in admixture levels (9% to 28%). We also estimated the relationship between levels of admixture and longitude in North America. Here, we identified a significant association between longitude and ancestry ($\beta_{\text{African anc.}} = 0.0007$, $P = 1.96 \times 10^{-6}$). This was evidenced when levels of African ancestry were projected onto a map of North America (Fig. 5c), revealing that West Coast samples have lower levels of African ancestry when compared with samples in the eastern seaboard at comparable latitudes. These results suggest that, in North America, the patterns seen under $k = 8$ emerge due to the different levels of African admixture (Fig. 4d, also supplementary fig. S11f, Supplementary Material online).

Lastly, we conducted a survey of genetic differentiation across the demographic clusters (see "Estimation of Nucleotide Diversity" in Materials and Methods). The overall differentiation was $F_{ST} = 0.050 \pm 0.001$ for autosomes and nearly twice as high for the X chromosome (0.092 ± 0.004; Fig. 6a). These results were robust to the removal of heterochromatic regions and low-frequency alleles [minimum allele frequency (MAF) < 0.05; supplementary fig. S12, Supplementary Material online]. To quantify the level of differentiation between population groups defined by their continental cluster (Fig. 4a), we used a hierarchical F_{ST} (hF_{ST}) model (Nei 1973) that decomposes the total differentiation into across-group (F_{GT}) and within-group components (i.e. a composite label of continent and cluster; F_{SG}) contributions, using unbiased estimators developed for Pool-Seq data (Gautier et al. 2024). Note that here we refer to the overall differentiation under the hierarchical model as hF_{ST} [with $(1 - hF_{ST}) = (1 - F_{SG})(1 - F_{GT})$] to distinguish it from the standard F_{ST} defined under a model without population groups. As shown in Fig. 6a, F_{SG} was always lower than F_{GT}, demonstrating that there is less differentiation within than between most clusters. We evaluated the level of differentiation across all cluster-continent pairs by computing pairwise F_{GT}. For each pair of regions, the underlying populations were analyzed under a hierarchical F_{ST} model with two groups, as shown in Fig. 6b (see results for $k = 8$ in supplementary fig. S13, Supplementary Material online). In general, all clusters involving Africa were consistently more differentiated than non-African groups. The highest level of differentiation was

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

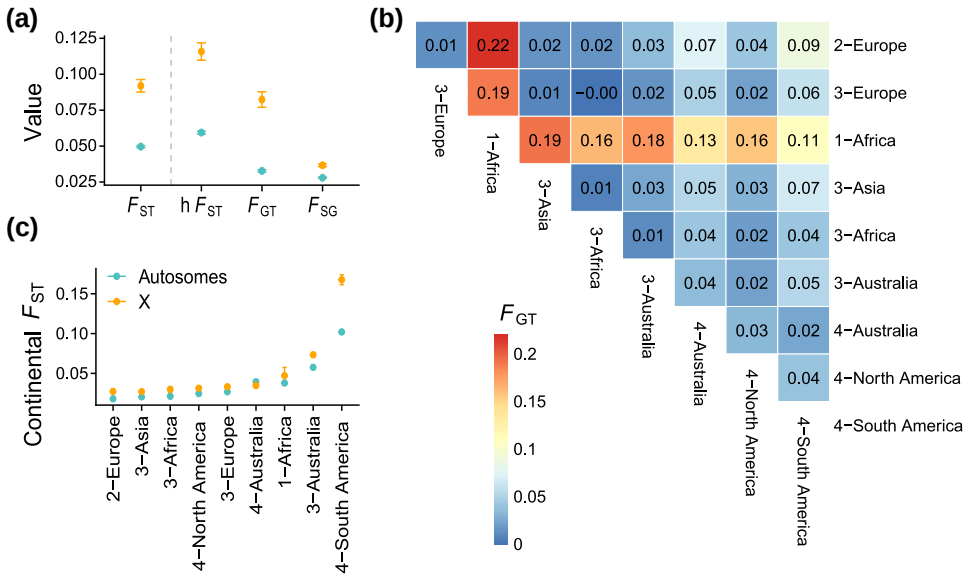

Fig. 6. Genetic differentiation analyses. a) Values of the F_{ST} estimates over all DEST samples and their 95% CI (corresponding to ±1.96 SE estimated using block–jackknife with blocks of 50,000 consecutive SNPs). Note that the hF_{ST}, F_{GT}, and F_{SG} statistics were estimated using the hierarchical F_{ST} model, over all DEST samples grouped according to the $k=4$ clustering analysis and their 95% CI. Colors indicate autosomes and X chromosomes. b) Pairwise comparisons between cluster-continents (under $k=4$) results in a heat map. In this plot, 1-Africa" refers to Sub-Saharan African populations, and "3-Africa" refers to North Africa. The clusters "Australia-3" and "Australia-4" represent samples with low and high levels of African admixture, respectively. c) F_{ST} estimates within clusters from the $k=4$ analysis. For (b) and (c), the population names represent the continent of provenance as well as the cluster number identified in Fig. 4a.

observed between Africa and EU-E ($F_{GT}=0.22$; Fig. 6b). Despite being located geographically between EU-W and EU-E, samples from the overlapping zone in Europe and Asia were more similar to EU-W than to EU-E (Fig. 6b). All populations in the Americas and Australia (i.e. "recent expansion" populations) were more similar to each other than to Africa or Europe. These findings align with previously published work (Bergland et al. 2016), indicating that flies in Australia and the Americas are subject to admixture from a recent secondary encounter between European and African populations. Finally, we estimated the differentiation (i.e. standard F_{ST}) within each cluster-continent level (Fig. 6c). Europe (cluster $2_{k=4}$) exhibited the lowest levels of differentiation, and South America (cluster $4_{k=4}$) the highest, which was essentially driven by a Brazilian and an Ecuadorian sample (Fig. 4c).

Updated Geographically Informative Markers Improve Predictive Resolution of Samples

Our previous release of DEST generated a panel of GIMs (Kapun et al. 2021). The second release of our data gives us the unique opportunity to test the accuracy of our previously published markers. To this end, we applied our previous DEST 1.0 GIMs to our new data, and we assessed the distance (d_{hav}; as great circle distance, see the "GIM Predictive Models" section in Materials and Methods) between the predicted locality and the "real" locality as recorded in the metadata. Overall, both DEST 1.0 models trained at the level of "city" and "region" (i.e. resolution at the level of state or province) perform similarly well on the new data ($r=0.995$, $P=2.2\times10^{-16}$; Fig. 7a). Next, we aggregated the d_{hav} estimates at the level of continents. We did this to assess whether the quality of our predictions varies as a function of continent. Overall, the best performance was observed in European samples (median resolution of ~409 km to real location; Fig. 7b), followed by the North American samples, with a resolution of 794 km. Unsurprisingly, the worst predictions from the DEST 1.0 markers occurred when deployed on samples from South America and Australia, two locations that were not included in the first release (Fig. 7b).

While our published markers performed well on samples from regions present in DEST 1.0, the addition of new regions to DEST required the generation of new GIMs. As such, we trained a new demographic model (DEST-GIM 2.0) including the new samples reported in this paper. Our new model was trained using the same workflow as DEST-GIM 1.0 (i.e. by retaining 40 PCs). Yet, the models differ in that DEST-GIM 2.0 was created by exclusively using noncoding SNPs as well as loci outside genomic regions spanning major cosmopolitan inversions. This new panel of GIMs is composed of 29,952 SNPs across all autosomes. Performance assessment of the new model by the d_{hav} analysis shows that DEST-GIM 2.0 performs similarly to the 1.0 version for existing locales (e.g. Europe or North America; Fig. 7b), yet they provide improved prediction accuracy for new regions (Fig. 7b and c).

Winter Severity Drives Year-to-Year Levels of Genetic Variation in Overwintering Populations

While much of demographic research in *D. melanogaster* has focused on spatial patterns of genetic variation, there is strong evidence that temporal demography, driven by yearly cycles of summer "booms" and winter "busts," can have strong and quantifiable effects on the frequency and levels of standing genetic variation in wild populations (Bergland et al. 2014; Nunez et al. 2024). For example, levels of postoverwintering (i.e. year-to-year) F_{ST} are generally higher than F_{ST} between samples collected within a growing season, even though overwintering F_{ST} captures a smaller number of generations (1 to 2 generations) than comparisons within a growing season (ca. 10 generations). This observation has led to the hypothesis that strong bottlenecks due to overwintering alter the genetic

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

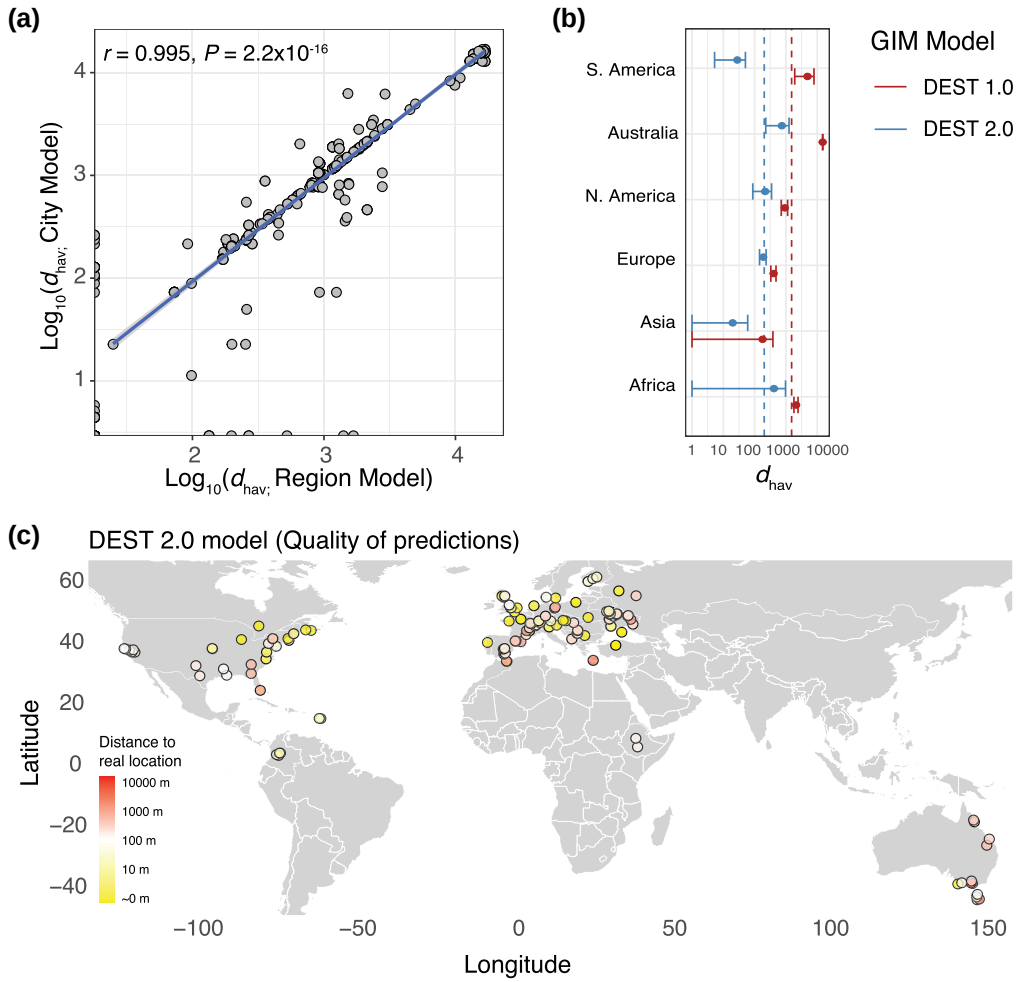

Fig. 7. Geographically informative markers. a) Bi-plot of d_{hav} from the DEST 1.0 GIMs. City model (*y*-axis) and Region model (*x*-axis). b) Mean and 95% CIs of d_{hav} for the DEST 1.0 and 2.0 GIM model (to improve readability, the *x*-axis has been $\log_{10}$ transformed, and CIs < 0 were set to 1; as 0 is logarithmically undefined). The mean distance to the true value is shown by dashed vertical lines (red for DEST 1.0, blue for DEST 2.0, models). c) Quality of predictions for the GIM DEST 2.0 model. The color indicates the average distance between the real d_{hav} of a sample and its predicted d_{hav}. Yellow are good predictions (accuracy = 0 to 10 m), white are "adequate" predictions (10 to 100 m), and red are poor predictions (1,000 to 10,000 m).

composition of fly populations, both due to changes in the amount of genetic drift (Nunez et al. 2024) and due to seasonally varying selection (Bergland et al. 2014; Machado et al. 2021; Behrman and Schmidt 2022; Johnson et al. 2023). A prediction of this hypothesis is that the strength and intensity of winter, an ecological driver of yearly population busts, should be correlated with the levels of overwintering F_{ST} from 1 year to the next. To test this prediction, we investigated patterns of temporal structure in worldwide DEST samples and asked whether latitude (a proxy for winter severity) is correlated with the levels of year-to-year F_{ST}.

For a given site, we assessed levels of F_{ST} between samples collected in two consecutive years (i.e. growing seasons) from the same locality. We implemented this analysis across 43 localities and estimated the relationship between mean year-to-year F_{ST} and latitude. We tested the prediction that higher latitude populations with stronger winter conditions exhibit higher levels of year-to-year F_{ST}. Indeed, we found a significant positive correlation between overwintering F_{ST} and latitude, yet the correlation is not monotonic. Using "broken-stick" regression (Muggeo 2003), we identified a change in the latitude–F_{ST} relationship at 50.3°N (Fig. 8a and e). Samples below 50.3°N tend to have lower values of year-to-year F_{ST} when compared with those above 50.3°N (Fig. 8b), and the magnitude of correlation between latitude and F_{ST} varies before and after this latitude mark (Fig. 8b; $r_{all} = 0.182$, $r_{>50\ lat} = 0.333$, $r_{<50\ lat} = 0.117$; all $P = 2.2 \times 10^{-16}$). These correlations are statistically significant and outperform 500 random permutations where latitude is shuffled.

A second finding of our year-to-year F_{ST} analysis was the discovery that several samples collected from Yesiloz, Turkey, are outliers (red dots in Fig. 8b) among samples below the 50.3 latitude mark (see Fig. 8a and b). This pattern was most apparent when considering samples between 2020 and 2021 (Fig. 8d) relative to comparisons at other years (Fig. 8c). This signal in Turkey appears to be associated with a historical heatwave and unusually warm winters in 2021 (see Discussion; Fig. 8d).

Footprints of Adaptive Differentiation to Insecticides in Europe

The broad sampling inherent to DEST allows us to test hypotheses about spatial adaptation in wild flies. We first took a heuristic approach where we extracted all regions of the genome with high across-cluster differentiation (i.e. $F_{GT} > 0.2$; see "Characterizing Population Structure in European and North

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

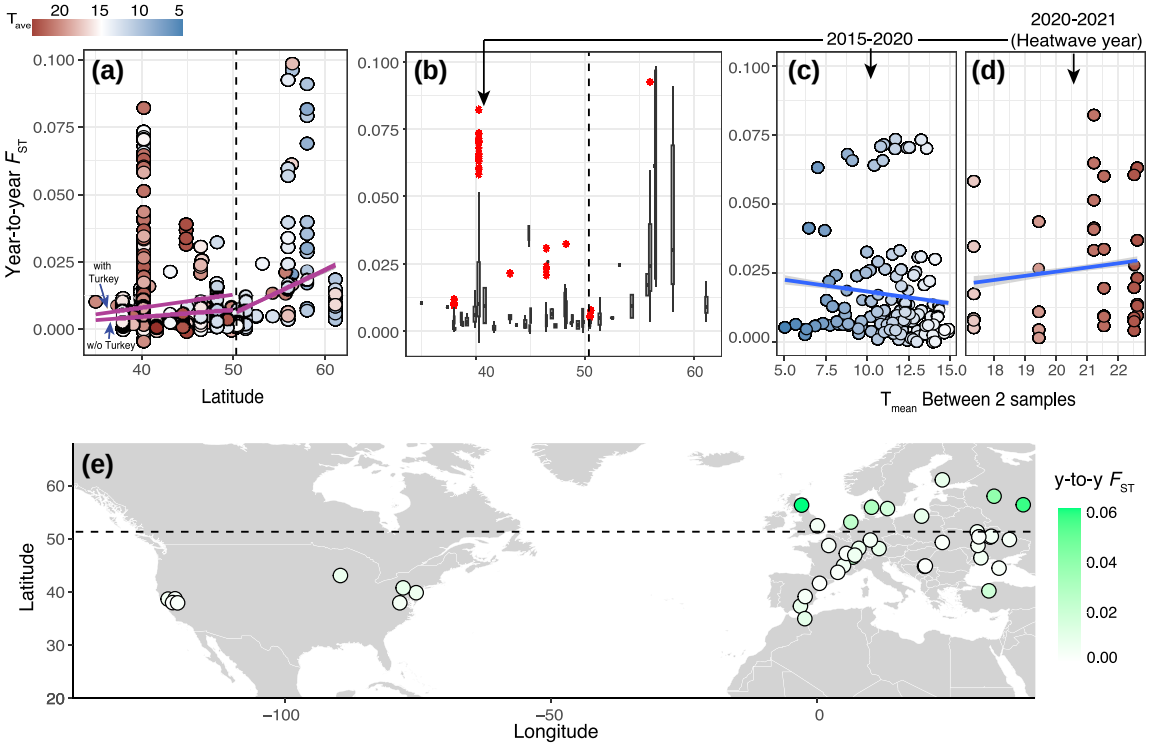

Fig. 8. Temporal genetic differentiation due to overwintering. a) F_{ST} values across DEST 2.0 samples as a function of latitude. Broken-stick regression and breakpoint are shown, for samples below latitude 50.3, the regression is shown with and without Turkey. The color indicates the average temperature in Celsius between the samples for which the F_{ST} was calculated. b) Distribution of year-to-year F_{ST} values across DEST 2.0 samples as a function of latitude, for comparisons spanning one winter only. Outliers (i.e. data above the 75th percentile) are shown in red. c) Distribution of temporal F_{ST} values as a function of the mean temperature in Turkey (Yesiloz) samples collected between 2015 and 2020 (logit transformed; correlation between F_{ST} and mean temperature; $r = 0.135$; $P = 4.60 \times 10^{-7}$). d) Same as (c) but for comparisons of 2020 and 2021, a historical heatwave year in Turkey and in Southern Europe (correlation between F_{ST} and mean temperature; $r = -0.100$; $P = 7.74 \times 10^{-13}$). e) Mean year-to-year F_{ST} overlaid over a world map of northern seasonal habitats.

American Populations" section above) and performed a gene-ontology enrichment analysis of genes located in these regions of high differentiation (Kofler and Schlötterer 2012). Overall, we found an enrichment of genes associated with environmental adaptation, such as responses to oxidative stress, metal ions, and pesticides (supplementary table S7, Supplementary Material online). One of the strongest signals of population differentiation was observed for the region surrounding the gene *Cyp6g1*, a cytochrome P450 (*Cyp*) gene (supplementary fig. S14, Supplementary Material online; a result also observed in DEST 1.0), a well-known gene involved in resistance to DDT and neonicotinoid insecticides (Le Goff and Hilliou 2017). This signal was particularly high when comparing North American and European samples. Elevated F_{GT} was also observed when comparing South American and North American samples, but not when comparing South American and European samples (supplementary fig. S14, Supplementary Material online). These signatures of differentiation suggest different adaptations likely driven by distinct environmental pressures and insecticide exposure levels in each continent.

To formally detect footprints of adaptive differentiation in our dataset, we applied the "Bayesian Population Association Analysis" framework, *BayPass* (Gautier 2015; Olazcuaga et al. 2020) to DEST samples from European localities (irrespective of sampling year or season; 138 samples in total; Fig. 9a) and relied on the estimated *XtX** statistic to identify overly differentiated SNPs. The analysis identified two regions in chromosome 2R as candidates of local adaptation (12,188,558 to 12,126,181 and 14,826,182 to 14,976,108; Fig. 9d). Both these regions harbor several *Cyp* genes. For example, the window at ~12 Mb contains *Cyp6g2*, and *Cyp6t3*, whereas the window at ~14 Mb contains *Cyp6a22*, *Cyp6a19*, *Cyp6a9*, *Cyp6a20*, *Cyp6a21*, *Cyp6a8*, and *Cyp317a1*. These genes are associated with hormonal metabolism as well as responses to insecticides (Danielson et al. 1995; Le Goff and Hilliou 2017). We performed gene-ontology enrichment analysis of genes within all *XtX** outlier regions and found an enrichment of terms such as "oxidation-reduction process," "cellular response to radiation," and "amide biosynthetic process," reflecting results from F_{GT} outlier regions above (supplementary table S8, Supplementary Material online).

Antimicrobial Peptides are Enriched Among Continent-Wide Targets of Seasonal Adaptation

We explored signals of seasonal evolution in DEST using paired spring-fall collections from Europe. In order to ensure that this test was not influenced by signals from previously analyzed data, we only used samples that were not included in previously published analyses (i.e. Bergland et al. 2014; Machado et al. 2021; Nunez et al. 2024; Fig. 9a). First, we ran the *BayPass* model, including both the Ω matrix as a demographic prior as well as categorical "spring" or "fall" labels (defined by the first and last sample collected in a locality within a year) in a contrast analysis. Under these conditions, *BayPass* outputs the C_2 statistic (Olazcuaga et al. 2020) that quantifies the degree of association of allele frequency with

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

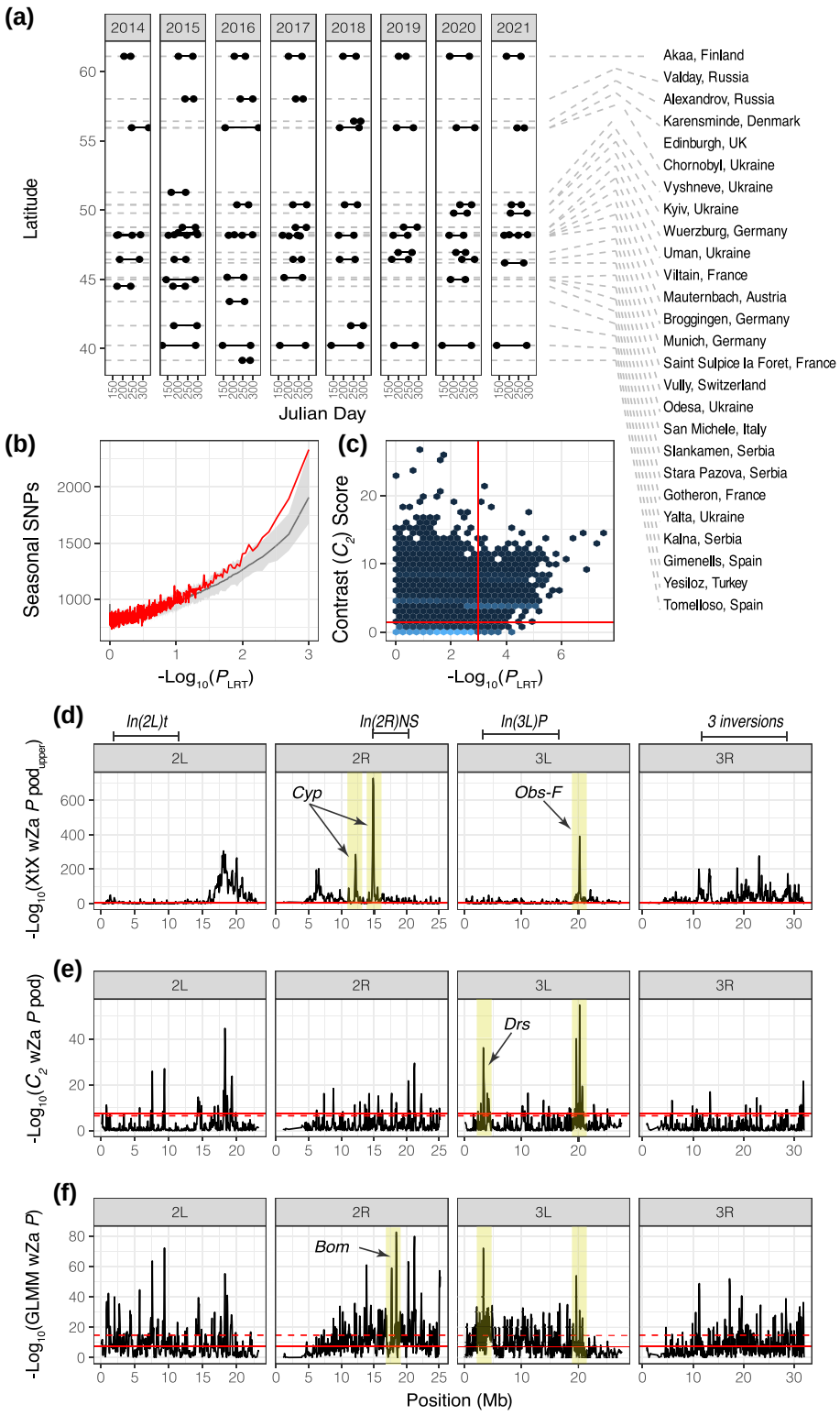

Fig. 9. Local and seasonal adaptation in *Drosophila*. a) Schematic of sampling for the local and seasonal analysis. In total, we used 138 samples collected in 26 European localities across an 8-year period. We selected localities where there was more than one sample per year and designated the first sample as "spring" and the last sample as "fall." There is no overlap between the samples used here and the samples used in seasonal analysis in Machado et al. (2021), Bergland et al. (2014), and Nunez et al. (2024). b) Results of the GLMM analysis. The permutations are shown in gray (95% CIs) and the real data in red. There are more SNPs with low seasonal *P*-values than expected by permutations. c) We performed the contrast analysis using *BayPass* 2.4. The contrast score (C_2 statistic) is the test statistic for the seasonal term and follows a χ^2 distribution with 1 degree of freedom. The *x*-axis is the $-\log_{10}$(*P*-value) from the GLMM. The red horizontal line represents the 99.9% significance threshold from the pseudo-observed data (POD) for ~10 M simulated sites. The red vertical line represents the 99.9% significance threshold from the permutations of the GLM analysis. d) Bayesian local adaptation scan. The plot shows the $\log_{10}$ transformed wZa *P*-value of the local adaptation (*XtX**) *BayPass* analysis. For d, e, and f, regions of interest are highlighted in yellow. Inversions are demarcated along the top of the figure. e) Bayesian seasonal adaptation scan. The plot shows the $\log_{10}$ transformed wZa *P*-value of the contrast (C_2) adaptation *BayPass* analysis. f) GLMM seasonal adaptation scan. The plot shows the $\log_{10}$ transformed wZa *P*-value of the LRT of base and seasonal models.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

season. We identified significant C_2 values using a simulation approach that is part of the *BayPass* workflow (see the "Scans for Adaptive Differentiation" section in Materials and Methods; supplementary dataset S2, Supplementary Material online). We observe that several regions across the *Drosophila* genome are enriched for signals of parallel seasonal evolution (Fig. 9e and f). A notable example appears in chromosome 3L (3,222,669 to 3,422,464), inside the region spanned by the inversion *In(3L)P*, where we observe the antimicrobial peptide *Drosomycin* (*Drs*) as well as several *Drs*-associated genes (i.e. *Drsl2*, *Drsl3*, *Drsl4*, *Drsl5*, and *Drsl6*). In view of previous observations of seasonal allele frequency oscillations in several immune genes, this result might imply functional shifts in immune tolerance and resistance across seasons in natural populations (Behrman et al. 2018). We performed gene-ontology enrichment analysis of all genes within C_2 outlier regions (supplementary table S9, Supplementary Material online). We found an enrichment of, among other terms, genes associated with "alcohol dehydrogenase (NAD) activity," including the gene *Adh* itself (supplementary table S10, Supplementary Material online).

We conducted an enrichment analysis comparing our C_2 SNPs (in the top 0.0001%) with loci reported in previous seasonal studies, done mostly in North American populations (i.e. FDR < 0.3 in Bergland et al. 2014; top 1% SNPs in Machado et al. 2021), to assess whether seasonal SNPs in Europe are also likely to be seasonal in North America. Our results indicate no significant enrichment of North American seasonal SNPs among our European C_2 SNPs (supplementary fig. S15, Supplementary Material online). Indeed, when compared with Pennsylvania data from Bergland et al. (2014), we observed a significant deficiency of these targets at both a global level ($P = 0.024$; supplementary fig. S15a, Supplementary Material online) and specifically on chromosome 3L ($P = 0.0055$).

Beyond the C_2 analysis, we implemented a generalized linear mixed model (GLMM) using the spring-fall seasonal labels, showing a global enrichment of seasonal SNPs relative to permutations (Fig. 9b). Comparing GLMM and *BayPass* results, we found a large number of SNPs exceeding the simulated 99.9% significance threshold for the C_2 statistic (Fig. 9c, vertical line), with the C_2 and GLMM models producing a similar set of candidate SNPs (Fig. 9c, red horizontal line). Likewise, a sliding window wZa analysis (Booker et al. 2024) of the GLMM results (window size of 100 kb, step size of 50 kb) identified the *Drs* region as a hotspot of seasonal adaptation (as in the C_2 analysis), and also revealed a second region of interest on chromosome 2R (18,376,129 to 18,475,992). This region contains several *Bomanin* genes (abbr. *Bom*; e.g. *BomBc1*, *BomT1*, *BomS1*, *BomBc2*, *BomS6*) known to play key roles in *Drosophila* antifungal responses (Xu et al. 2023). A region on 3L, near 20,172,964 to 20,271,926 bp, notable for harboring adjacent signal peaks across analyses of seasonal and local adaptation (see Fig. 9d to f; yellow band), contains *obstructor-F* (*obst-F*), a gene previously reported as a candidate of insecticide adaptation (Campo et al. 2013; Bogaerts-Márquez et al. 2021).

Discussion

A Unified Resource for Wild *Drosophila* Genomics

Drosophila melanogaster is a cosmopolitan species with resident populations across all human-inhabited continents that evolve adaptively in response to spatially and temporally varying selection in the wild (clinal patterns reviewed in Adrion et al. 2015; seasonal patterns reviewed in Johnson et al. 2023). To achieve a comprehensive understanding of the evolutionary patterns within this species, we need to create panels of variation sampled across wide geographical scales and densely across time. This is not a trivial undertaking for any single lab to achieve. The original impetus behind DEST was to generate a unified dataset and workflow that would capitalize on the collaborative efforts of labs and consortia around the world (Kapun et al. 2021). DEST 2.0 builds on the original release by adding twice as many new samples, significantly expanding the dataset.

Overall, the incorporation of the aforementioned data into the dataset showcases the flexibility and capacity for growth of DEST, as a centralized and well-annotated repository of *Drosophila* genomics. Furthermore, the DEST 2.0 *Dockerized* pipeline now allows for pools generated using single-end (SE) sequencing approaches to be incorporated into its workflow, hence allowing for older pooled datasets to be included in DEST analyses. We plan to continue maintaining and updating the DEST workflow, with potential future expansions to explore other *Drosophila* species and additional data types. To keep pace with the influx of new genomic data, we have upgraded the DEST genome browser to the latest version of JBrowse, which has better scalability and performance when displaying large datasets (Diesh et al. 2023).

Heterogeneous Patterns of Recombination in DEST Samples

This release also includes genome-wide recombination rate estimations for 74 representative populations. In comparison with the findings of previous studies (Comeron et al. 2012; Adrion et al. 2020), our own estimates show a reduction of ~3-fold. This discrepancy may be attributed to the combination of our methodological approach and the nature of our data. The deep learning approach of *ReLERNN* (Adrion et al. 2020) is dependent on allele frequencies, and it is thus possible that levels of genetic polymorphism may affect the estimation of levels of recombination rate. In our analyses, we estimated allele frequencies on SNPs that were called with very conservative and stringent filtering methods. Furthermore, the polymorphism data were obtained from Pool-seq data from derived European and North American populations, which exhibit lower levels of genetic polymorphism (~2- to 3-fold; e.g. Ometto et al. 2005) than the ancestral African populations used in Adrion et al. (2020). Accordingly, there is a strong and significant correlation between the number of SNPs and the average recombination across the 74 populations (Spearman's $\rho = 0.821$, $S = 12{,}074$, $P < 1.0 \times 10^{-25}$; $R^2 = 0.667$). Notice that for this analysis, we estimated the population-scaled effective recombination rate (ρ), rather than the actual crossing-over rate (r, where, e.g. in autosomes, $\rho = 4N_e r$). A comparable finding was observed in the case of wild barley (Dreissig et al. 2019). It seems also probable, however, that our populations can indeed be characterized by heterogeneous levels of recombination, as has been reported by numerous studies in *Drosophila* (Hunter et al. 2016; Samuk et al. 2020; Wang et al. 2023).

New Insights into Ancestral and Recent Fly Phylogeography

The prior releases of DEST and similar datasets (Kapun et al. 2020, 2021; Machado et al. 2021) characterized patterns of population structure within North America and Europe.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

In this paper, we expand the repertoire of samples available for demographic inference and phylogeographic analysis and provide novel insight into colonization patterns in Australia and South America, as well as the genetic structure of Europe.

Our analyses provide new insights into the colonization history of the Americas and Australia. First, our analyses recapitulate published signals of a cline in African ancestry in North America (Kao et al. 2015; Bergland et al. 2016; Corbett-Detig and Nielsen 2017; Coughlan et al. 2022), with higher African ancestry in tropical versus temperate populations. These results support the hypothesis that African admixture is higher in equatorial populations, likely due to two separate introductions of *D. melanogaster* to the Americas: one from Europe and one from Africa. African ancestors likely entered through the Caribbean, with the earliest records of the species appearing in Cuba in 1862 (Sturtevant 1921). Introduction of the species, however, is documented in historical records starting in 1875 in New York. This suggests that African–European admixture in the eastern United States may have developed around 1876 (Lintner 1882, p. 217; Keller 2007).

Although the South American samples do not display a significant ancestry cline, the observed variation (9% to 28%) indicates the possibility that clinal patterns could emerge with broader sampling. Our results from Australia, on the other hand, revealed a reversed latitudinal trend compared with North America (Fig. 5b). This is consistent with previous work that has demonstrated clines in African/European ancestry across Australia (Bergland et al. 2016). Taken together, our results suggest that additional sampling is warranted in Australia and South America to further resolve these recent migrations.

Second, previous analysis of population structure in Europe (Kapun et al. 2020 , 2021) identified two phylogeographic clusters, referred to as Europe East" and "Europe West". Here, we identify a potential third cluster occupying the overlap zone between these clusters (in the $k = 8$ analysis; Fig. 4b). This cluster is notable since its placement closely mirrors the "suture zones" (Remington 1968) of other species, such as *Bombina* toads (Hofman et al. 2007), *Leuciscus cephalus* (Hewitt 2011), and *Mus musculus* (Ďureje et al. 2012). In our analyses, we tested whether this overlap zone is a zone of admixture between EU-E and EU-W or if it is a separate cluster, perhaps reflecting expansion of a Middle Eastern subpopulation. We show that the overlapping zone is not a distinct cluster (supplementary fig. S10c, Supplementary Material online) supporting $k = 4$. It is possible that the overlap zone is an artifact of the data that appears due to asymmetrical levels of migration between clusters, as reported previously (EU-W → EU-E as 0.209 flies/gen vs. EU-E → EU-W as 0.178 flies/gen; Kapun et al. 2021). These findings are further supported by our supplementary F_{ST} analyses that include the overlap zone [e.g. F_{ST} (EU-W vs. overlap) = 0.00; F_{ST} (EU-E vs. overlap) = 0.01]. Third, it is also possible that these patterns may arise from the action of a non-neutral force confounded with the complex demographic history of *D. melanogaster* in Europe. These hypotheses will be explored in deeper detail in future work.

To better explore these patterns of phylogeography, we implemented a variety of methods, including *F*-statistic comparisons, linear modeling of ancestry proportions, and model-based inference using the program *moments* (Jouganous et al. 2017). Each of these methods has its own strengths and weaknesses, especially in the context of Pool-Seq data. Our analysis with *moments* builds on our previous efforts to test demographic inference methods on Pool-Seq data (Kapun et al. 2021). However, these analyses should be interpreted with caution. Unlike most other methods in this paper, which directly or indirectly account for the inherent error structure of Pool-Seq datasets, our demographic inference approach using *moments* does not fully accommodate the unique biases of Pool-Seq. We highlight three aspects that influence our interpretation of the results:

First, as in DEST 1.0, we used the PoolSNP program to call SNPs in our dataset. Like many widely used Pool-Seq variant calling tools (Koboldt et al. 2009; Kofler et al. 2011), PoolSNP identifies polymorphisms by integrating evidence from multiple samples. This approach produces a high-quality SNP panel enriched for common mutations but comes at the cost of excluding rare and private mutations within pools (Kapun et al. 2021). While common SNPs are sufficient for most analyses presented here, the loss of rare variants can impact demographic inference—a challenge not unique to Pool-Seq (Gravel et al. 2011). Indeed, comparative studies have shown that rare variants are critical for detecting fine-scale population structure and recent demographic events, whereas common variants are more informative for older demographic events (O'Connor et al. 2015). Here, we aimed to test models that reflect relatively ancient events, such as out-of-Africa migrations, where common variants (i.e. of older origin) are more informative than rare variants (i.e. of very recent origin and likely geographically restricted). In other words, following a graph interpretation of population histories, we focused on the deeper part of the topology, for which common variants have already proven to be highly informative (Patterson et al. 2012).

Second, the current implementation of *moments* does not account for any source of noise in site frequency spectra. In the context of Pool-Seq, errors may arise from factors such as unequal DNA contributions from individuals, variation in sequencing error rates, and differences in effective coverage (Carvalho et al. 2023).

Third, we used forward-time population genetic simulations in SLiM to assess the performance of our admixture inference methods on discretized allele count data derived from estimated allele frequencies. While none of the methods perfectly recovered the true ancestry proportions, all were able to detect the general pattern of the admixture cline resulting from secondary contact. Notably, the method implemented in *moments* demonstrated higher precision and lower error rates compared with the linear modeling approach (supplementary fig. S11, Supplementary Material online). Furthermore, our admixture estimates aligned closely with findings from previously published studies (Bergland et al. 2016; Corbett-Detig and Nielsen 2017; Coughlan et al. 2022).

These results underscore the potential for integrating Pool-Seq data with model-based demographic inference methods to investigate complex population histories. Yet, a systematic evaluation of key parameters and known sources of variation—such as differences in effective coverage—is essential for improving the accuracy and reliability of future demographic inferences. We acknowledge these limitations and encourage cautious interpretation of our model-based results. Nonetheless, the *moments*-based analyses yielded valuable insights into population structure and admixture dynamics of the DEST 2.0 samples.

Inferring Targets of Adaptation Across Time and Space

The complex patterns of spatial population structure that we have described above are likely to alter the adaptive capacity

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

of fly populations. Indeed, a recent genomic analysis of the sibling species *D. simulans* across continents revealed that demographic ancestry, and not shared selection regime, is a better predictor for the genetic basis of local adaptation to thermal stressors (Otte et al. 2021). These results highlight that assessing footprints of adaptation requires robust controls for the complex demographic structure of species. We implemented the *BayPass* framework (Gautier 2015; Olazcuaga et al. 2020) to discover targets of spatially and temporally fluctuating selection across Europe. This framework is flexible, as it incorporates priors from population structure (via the Ω matrix) and, optionally, environmental variables (either as factors or covariates).

Our analyses of spatial adaptation reveal signatures of continent-wide differentiation around cytochrome P450 genes (e.g. *Cyp* genes) in 2R (Fig. 9). Follow-up analyses using estimates of across-group differentiation (F_{GT}) revealed that these genes are highly differentiated in comparisons between North American populations versus both European and South American populations (supplementary fig. S14, Supplementary Material online). Given that *Cyp* genes are important players in insect detoxification pathways and have been implicated in the evolution of insecticide resistance (Le Goff and Hilliou 2017), these findings suggest that flies have experienced continent-wide adaptation to different histories of land and pesticide use (see also Kapun et al. 2020). While further experimental validation is needed to disentangle the particular gene targets and drivers of selection, these data highlight the power of DEST to reveal the genetic bases of local adaptation to paralleled stressors.

We also explored patterns of temporal divergence in response to seasonality. Previous work has shown that seasonal adaptation, via adaptive tracking (Botero et al. 2015), is a ubiquitous and important evolutionary force affecting patterns of genetic variation across the genome of *Drosophila* (Bergland et al. 2014; Kapun et al. 2016a; Machado et al. 2021; Rudman et al. 2022; Bitter et al. 2024; Nunez et al. 2024). Here, we used the DEST 2.0 data to revisit footprints of seasonal adaptation across samples not used in previous analyses. Using this dataset, we tested the hypothesis that seasonal adaptive tracking is a general phenomenon of European temperate *Drosophila*. One challenge associated with testing this hypothesis is determining the appropriate covariate (e.g. temperature, humidity, and rainfall) and the timeframe of selection (e.g. 0 to 15, 0 to 30 d prior to collection) to use in the model. For example, Nunez et al. (2024) showed that, for the inversion *In(2L)t* in Virginia, the best seasonal model used the temperature 0 to 15 d prior to collection as a covariate. Yet, in Europe, Nunez et al (2024) showed that humidity 0 to 30 and 0 to 60 d prior to collection were the best models for EU-E and EU-W, respectively. Therefore, we used a contrast framework using the seasonal labels (i.e. "spring" and "fall") as comparison factors. This approach had been successfully used in the past by Bergland et al. (2014) and Machado et al. (2021) and allowed us to surmount the challenge of covariate selection.

We implemented a test of seasonality in a two-pronged approach using both the *BayPass* and the GLMM framework. Our results show multiple regions of interest across the genome, which are concordant across both BayPass and GLMM. For example, it highlights a region on 3L that encodes for *Drosomycin* and *Drosomycin-like* genes (Fig. 9e), canonical antifungal defense loci (Zhang and Zhu 2009), as a continent-wide hotspot of seasonal adaptation (Figs. 9c and f). These findings are noteworthy, as fungal communities are known to vary drastically across seasons, driven by changes in soil moisture, temperature, and carbon availability (Schadt et al. 2003). Furthermore, the analysis also reveals a region of interest on chromosome 2R containing *Bomanin* genes that are also associated with antifungal defense (Xu et al. 2023). Another gene of interest is *Obstructor-F*, a gene that has several functions and that has been associated with pesticide response (Campo et al. 2013).

Our gene-ontology enrichment analysis for targets of seasonality highlighted "alcohol dehydrogenase activity"—including the gene *Adh* itself—as being enriched among outlier regions. This is significant because patterns of genetic variation in *Adh* have long been recognized as classical examples of ecological adaptation (Kreitman 1983; Berry and Kreitman 1993). However, recent discussions have emphasized that the specific agents of selection acting on this gene remain unclear (Siddiq and Thornton 2019). We also assessed whether the seasonal SNPs observed in our C_2 analysis from Europe are enriched in seasonal datasets generated mostly from North American populations (Bergland et al. 2014; Machado et al. 2021). Our results showed no enrichment (see supplementary fig. S15, Supplementary Material online) between the datasets compared. In other words, these results suggest that the genetic basis of seasonality is different between continents. As mentioned above, this finding is consistent with previous studies positing that population ancestry is a more important predictor of adaptive genetic architecture than the existence of paralleled selection regimes (Otte et al. 2021).

Overall, our seasonal analyses reveal two major takeaways. First, they reveal that seasonal adaptive tracking is a detectable phenomenon across the temperate range of *D. melanogaster*. Yet, they also indicate that genetic ancestry may strongly influence the specific loci driving adaptation. And second, the data highlight a large role of pathogen response genes as major players in worldwide seasonality (Behrman et al. 2018). These findings suggest that follow-up studies of seasonality should take a more comprehensive approach to incorporate both abiotic (e.g. temperature) and biotic (e.g. pathogen) views of "seasonality." Further expansions of the DEST dataset will facilitate more granular exploration of adaptive tracking driven by spatially and temporally fluctuating selection.

The Impacts of Overwintering Demography on Genetic Variation

The results highlighted above showcase the power of DEST to examine fine-grained patterns of evolutionary change occurring within each population. Yet, seasonal adaptive tracking is not the only process at play in temperate habitats. As the seasons change, *Drosophila* populations expand and contract depending on resource availability (Atkinson and Shorrocks 1977). Indeed, the establishment and range limits of many insect species are tied to their ability to survive winter (Lawton et al. 2022). Previous work has suggested that local fly populations grow to their largest possible size during the summer months (with peaks in June and September; see Atkinson and Shorrocks 1977; Sanchez-Refusta et al. 1990; Gleason et al. 2019; Bangerter 2021) and drastically decrease in size following the onset of winter, when resources are scarce and reproduction is suppressed, leading flies to enter dormancy and overwinter until the next growing season. These seasonal

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

demographic cycles, called "boom-and-bust" demography, can result in yearly bottlenecks of up to ~97% in the "local" population (Nunez et al. 2024), and thus are likely to have fundamental consequences for standing genetic variation.

Consistent with studies on the impact of large overwintering bottlenecks, previous research on temperate *Drosophila* populations has revealed significant differences in genetic differentiation when comparing patterns of variation within a year and across years, a process likely driven by the effects of these cyclical winter contractions (Nunez et al. 2024). However, a key remaining question is whether these boom-and-bust dynamics correlate with winter severity: do harsher winters lead to greater year-to-year genetic differentiation than milder ones? We explored this question using year-to-year F_{ST} and tested the hypothesis that populations with harsher winters have, on average, larger levels of year-to-year F_{ST}. Our results support this hypothesis, revealing positive correlations between F_{ST} and latitude, particularly for samples collected at latitudes higher than 50.3°N (Fig. 8a and e). These patterns suggest that habitats with colder, harsher winters, typical of higher latitude habitats, impose stronger bottlenecks on overwintering flies relative to lower latitude habitats. Nevertheless, given that DEST presently includes only a limited number of populations with >5 years of sampling, a systematic comparison of these dynamics requires an expansion of sampling efforts. One notable exception to the pattern of both year-to-year and multiyear F_{ST} was found in the Turkish samples. There, populations in 2021 showed an unexpected positive correlation between F_{ST} and temperature (Fig. 8d; relative to patterns in previous years at the same site, Fig. 8c). These patterns may have arisen as a result of the harsh weather conditions of southern Europe in 2021. During that period, weather anomalies created unusually warm winters as well as the hottest and longest summer heat waves in the region's recent history (Lhotka and Kyselý 2022). These results, combined with the observed differences in F_{ST} levels both within and across multiple years at this site, suggest that extreme heat waves may also influence the standing genetic variation of flies in both tropical and temperate habitats. In this context, heat waves may have affected flies both directly, through physiological thermal challenges, and also indirectly by affecting their food sources.

Overall, our findings provide three major insights into the temporal structure of *D. melanogaster* populations. First, we showed that overwintering bottlenecks are associated with the severity of winter across habitats. Second, that there is a predictable relationship between the strength of winter and the genomic consequences of overwintering in fruit flies. And third, that temperate fly populations exhibit spatially stable genetic structure and thus accumulate divergence due to cyclical episodes of overwintering drift.

Future Directions

In conclusion, our findings not only highlight the power of DEST as a resource for fly biologists but also its promise and potential for growth. Indeed, as more temporal samples continue to be added, more detailed gene-environment association studies will undoubtedly shine a light on the drivers of selection across worldwide habitats. Our data may also be used in order to parameterize temporally and spatially explicit population genetic simulations, which, combined with climate change forecasting datasets, will help to model rapid evolutionary responses under various climate scenarios. Lastly, as our consortium continues to grow, we are working to include a variety of other *Drosophila* species into DEST. Such multispecies data will be pivotal to assess the evolutionary dynamics of adaptive tracking across the phylogeny.

Materials and Methods

Sample Mapping and SNP Discovery Using the DEST Mapping Pipeline

Samples were mapped to the *D. melanogaster* hologenome using the pipeline described in our first release (Kapun et al. 2021). This pipeline consists of a combination of genomic tools (fast-qc [v0.12.1], Cutadapt [v2.3] (Martin 2011), BBMap [v38.80] (Bushnell et al. 2017), BWA-mem [v0.7.15] (Li 2013), Picard [v3.1.1], SAMtools [v1.9] (Li et al. 2009)) in a Docker container. For our current release of DEST [v2.0], we have updated the Docker container to enable mapping of reads sequenced in both paired-end (PE) and SE configurations. This new version of the pipeline can be found in Dockerhub (https://hub.docker.com/) as destbio/dest_freeze2:latest. SNP calling was performed using the PoolSNP algorithm (Kapun et al. 2020). For SNP calling, we used the default parameters optimized in the first release of DEST (Kapun et al. 2021). Briefly, these parameters are: minimum allele count = 50, MAF = 0.001, minimum coverage (per pool) = 4, max-Coverage = 0.95, and missing Fraction (threshold) = 50%. The SNP calling step, as well as genome annotation with SNPEff (v5.2; Cingolani et al. 2012), was automated using SnakeMake (Mölder et al. 2021). We provide ready-to-use outputs of the DEST pipeline both in variant call format (VCF) as well as in genomic data structure format (Zheng et al. 2012). The entire DEST pipeline can be found on GitHub at https://github.com/DEST-bio/DESTv2.

Metadata for All DEST 2.0 Samples

Comprehensive metadata for all DEST 2.0 samples is included in supplementary table S1, Supplementary Material online, including collection information on sampling date and location. Flies from the previous release (DEST 1.0) were collected in a variety of methods, including aspirators, traps, and nets. New samples reported here as part of the DrosEU3 collection were sampled using standardized traps with a variety of baits (see supplementary table S1, Supplementary Material online). All newly acquired samples were collected in a coordinated manner and processed following the protocols outlined in Kapun et al. (2020). In brief, male flies were exclusively gathered from natural or seminatural habitats, such as orchards, vineyards, and compost piles. In Europe, collections primarily used baited traps with mashed banana or apples and live yeast, left at sampling sites for several days, or were obtained via sweep netting (Kapun et al. 2020). In North America, flies were collected using sweep nets, aspiration, or baited traps over natural substrates (Behrman et al. 2018; Machado et al. 2021; Nunez et al. 2024). All samples were preserved in 95% ethanol at −20 °C before DNA extraction. In addition, for the current release of DEST, we incorporated data from previously published studies (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). These data were added to DEST by processing the raw sequences using the Docker pipeline. These new samples include: 37 samples from Nunez et al. (2024), 16 samples from Fournier-Level et al. (2019), 2 samples from Hoffmann et al. (2002), 17 samples from Lange et al. (2022), 8 samples from Reinhardt et al. (2014), and 1 sample from Svetec et al.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

(2016). Samples from Fournier-Level et al. (2019) consist of multiple replicates from the same locality, each with low coverage. Accordingly, we collapsed all replicates from each site into a single "consolidated" library (see "Collapse" category in Fig. 2a), each with read depths (RDs) of ~60×.

Filtering Parameters

We filtered SNPs and samples using metrics and tools described in our first release (Kapun et al. 2021). In brief, we (i) calculated the levels of contamination by congenerics, (ii) levels of read duplication in the sequencing run, (iii) proportion of SNPs with missing allele frequency data, (iv) ratio of synonymous to nonsynonymous polymorphism (p_N/p_S), (v) nominal coverage, and (vi) the effective coverage. Levels of contamination by congenerics refer to the amount of non-*D. melanogaster* flies that were accidentally sequenced in pools.

We assessed contamination using a two-pronged approach. First, we assessed the levels of competitive mapping of reads to the genomes of *D. melanogaster* (RefSeq: GCF_000001215.4) and *D. simulans* (RefSeq: GCF_016746395.2). *Drosophila simulans* and *D. melanogaster* can be difficult to differentiate in the wild, and the wrong species may be sequenced by accident. The specifics of competitive mapping were discussed previously (Kapun et al. 2021; Machado et al. 2021). Our second approach uses a *k*-mer counting method that can be directly applied to raw read files and is flexible for multiple species that are represented or closely related to those represented in the target *k*-mer dictionary (Gautier 2023). Next, we generated in silico pools consisting of mixtures of panels of inbred *D. melanogaster* (Mackay et al. 2012) and *D. simulans* (Signor et al. 2018). We generated these in silico pools by varying the mixture levels of the two species. By analyzing these pools, we show that both the competitive mapping and the *k*-mer approach are accurate (supplementary fig. S3a, Supplementary Material online), with the competitive mapping approach slightly overestimating contamination (by 2.3% max) and the *k*-mer approach slightly underestimating contamination (by 6% max).

The levels of read duplication were extracted directly from the BAM files by mining the "mark_duplicates_report" output using a custom R script. Missing data was assessed by counting the number of sites reported as "NA" in a particular pool. The p_N/p_S statistic was calculated using the SNP annotations derived from SNPEff using a custom script (see GitHub). The nominal, genome-wide, RD is extracted directly from the BAM file using a custom script (see GitHub). Note that the per-site RD is a standard output of PoolSNP.

Masked gSYNC Files

Prior to SNP calling, we masked positions in each gSYNC file, which is a genome-wide extension of the SYNC file format (Kapun et al. 2021) for each sample based on minimum and maximum RD thresholds, as well as on proximity to putative indel polymorphisms as identified by GATK IndelRealigner [v3.8.1] (DePristo et al. 2011). In addition, we masked regions associated with repetitive elements identified as fragments of interrupted repeats by Repeat Masker (Smit et al. 1996; Jurka 2000), microsatellites and simple repeats identified by Tandem Repeat Finder (Benson 1999), repetitive windows identified by Window Masker and SDust (Morgulis et al. 2006), and transposable elements and other repetitive elements identified by Repeat Masker (all obtained from the UCSC Genome Browser), using the custom Python script MaskSYNC_snape_complete.py, as previously described in Kapun et al. (2021). Importantly, the positions of these masked sites are stored in BED file format, which allows accounting for masked sites both in mono- and polymorphic positions when calculating unbiased site-specific averages for population genetic statistics as described in the section "Estimation of Nucleotide Diversity" (see also Kapun et al. 2020).

Effective Read Depth

In addition to the nominal RD, multiple downstream analyses in this paper use the "effective RD" metric (n_e). This is a Pool-Seq-specific metric that corresponds to the number of individually genotyped chromosomes, after accounting for the double binomial sampling that occurs in Pool-Seq (Kolaczkowski et al. 2011; Feder et al. 2012; Gautier et al. 2013). An estimate of n_e for a Pool-Seq sample can be defined as

$$n_e = \frac{N\,c}{N + c - 1} \quad (1)$$

where N is the haploid sample size of the pool (i.e. number of pooled chromosomes) and c is the nominal RD at a given position or average across the genome (see supplementary text S1, Supplementary Material online for further details on the derivation of Equation (1) and for a more general formula applicable to collapsed Pool-Seq sample).

Recombination Landscape

We inferred the genome-wide recombination landscape for 74 of our samples using ReLERNN [v1.0.0] (Adrion et al. 2020). Samples were selected to cover the spatial distribution of the DEST 2.0 dataset, with a particular focus on Europe and North America. As ReLERNN has been shown to achieve modest accuracy when using allele frequencies derived from Pool-Seq data sequenced at low depth, we selected those samples with the highest possible coverage (mean sequencing depth = 68.8, SD = 35.8, min = 32, max = 234; supplementary table S1, Supplementary Material online). To further reduce any possible bias and to maximize the reliability of the allele frequency used by ReLERNN to estimate recombination, we used BCFtools (Danecek et al. 2021) to extract allele frequencies of all biallelic SNPs with a frequency >0.01 and RD >10. The resulting data was used to run ReLERNN. The parameters used in ReLERNN *simulate* module were as follows: assumed per-base mutation rate: --assumedMu 3.27×10^{-9}; assumed generation time (in years): --gentime 0.08; and upper rho/theta ratio --upperRhoThetaRatio 10. For the training module, we applied a MAF of 0.01 (–maf). For the prediction module, we considered windows with a minimum number of 50 sites (–minsites). Following the developers' recommendation, we let the program select the optimal size of the nonoverlapping windows on which per-base recombination rates were predicted. Analyses were run separately for autosomes and the X chromosome, to account for their different ploidy in the male-only pooled samples. To allow comparisons between samples, we estimated the average per-base recombination rates in larger 200 kb nonoverlapping sliding windows by combining the raw rates estimated in each ReLERNN-selected window weighted by the fraction of the overlap with the corresponding 200 kb sliding window. Using the same approach, we also calculated the recombination landscape using the raw data of (Comeron et al. 2012), which are

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

significantly correlated with our estimates for most of the populations (supplementary table S11, Supplementary Material online). We note a weak negative correlation between the coverage sequencing depth of a sample and the average CI of the recombination rate values measured across the genome (Spearman's $\rho = -0.330$, $S = 89{,}820$, $P = 0.0041$; $R^2 = 0.018$), consistent with a greater accuracy in the estimates of recombination in samples with high-sequencing coverage. Recombination rates are available in the genome browser.

Estimation of Nucleotide Diversity

We conducted population genetic analyses using *npStat* [v1.c] (Ferretti et al. 2013). Out of the 530 high-quality samples, we used a subset of 504 samples for which we also had the masked BAM files, which were necessary to compute the statistics. The remaining 26 samples do not have a masked BAM file, as they were incorporated from the DGN data. For those samples, diversity statistics were re-computed from the masked gsync files (see above) as described in DEST 1.0 data (Kapun et al. 2021). Standard nucleotide diversity statistics were first directly estimated from each BAM file, for nonoverlapping windows (10, 50, or 100 kb) over the whole genome, using the estimators for Pool-Seq data developed by Ferretti et al. (2013). Only positions covered by at least two reads and <250 reads with a minimum quality >20 were considered in the computations (*-mincov 2 -maxcov 250 -minqual 20* options). We further calculated genome-wide estimates for each sample (separating autosomes and the X-chromosome) as the median window estimates (excluding windows with <75% coverage) for window sizes of 10, 50, or 100 kb (i.e. as displayed in the genome browser). For the analysis of the X chromosome, the haploid sample size was set for each and every Pool-Seq sample to the number of flies included in the pool. Indeed, most of them consist of only males and for the few samples including females (i.e. samples included from: Reinhardt et al. 2014; Svetec et al. 2016; Fournier-Level et al. 2019; Lange et al. 2022), the estimates were very similar to those obtained when setting the haploid sample size to twice the number of flies. For autosomes, the haploid sample size was set to twice the number of flies, except for 13 DGN samples that consist of pools of haploid embryos, for which the pool haploid sample size was set to the number of flies.

Analyses of Chromosomal Inversions

Based on previously identified inversion-specific marker SNPs (Kapun et al. 2014), which are in tight linkage with the breakpoints of the common cosmopolitan inversions *In(2L)t*, *In(2R)NS*, *In(3L)P*, and *In(3R)Payne* and of the rare cosmopolitan inversions *In(3R)C*, *In(3R)K*, and *In(3R)Mo*, we estimated sample-specific inversion frequencies based on the median of the frequencies of inversion-specific alleles across SNP markers for a given inversion following the approach in Kapun et al. (2014). To test for associations between inversion frequencies and geographic variables, we partitioned the data by continent and analyzed each inversion separately. We fit general linear models including arcsine square-root transformed inversion frequencies as dependent variables, which accounts for the skewed variance distribution in binomial data when normality is assumed. We included latitude, longitude, and sampling year as independent variables and tested for the effect of the independent variables and all possible interactions with a likelihood ratio test. While we considered latitude and longitude as continuous numerical variables, we treated year as a categorical factor to account for the sparse sampling across years at most locations.

Principal Component Analysis

Global population structure analyses were done using the PCA algorithm implemented in the FactoMineR [v2.4] package (Lê et al. 2008). For these analyses, we included all available samples that passed the filter in DEST 2.0. We include all biallelic SNPs in autosomes provided they had <1.0% missing data and a mean allele frequency >1.0% (across all samples). We thinned the dataset by only selecting SNPs that were 500 bp apart from each other, reducing the dataset to 168,408 SNPs. Note that we ensured that this PCA was robust to variations in read coverage and haploid pool size by comparing the estimated PCs with those obtained with a random allele PCA, as implemented in *randomallele.pca*() from the R package *poolfstat* [v3.0.0] (Gautier et al. 2022, 2024; supplementary fig. S7, Supplementary Material online).

Admixture Estimates: Linear Modeling Method

We estimated the proportion of African and European admixture in North and South America, as well as Australian samples using a linear regression framework (Alkorta-Aranburu et al. 2012; Bergland et al. 2016). We modeled allele frequencies in each "admixed population" (i.e. North America, South America, and Australia) as a linear combination of the two "ancestral populations" (i.e. Europe and Africa) using an intercept-free linear model:

$$p_{i\text{–admix}} = \beta_1\,(\text{African ancestor}) + \beta_2\,(\text{European ancestor}_k) + \varepsilon \quad (2)$$

where $p_{i\text{–admix}}$ is a vector of allele frequencies composed of 5,000 randomly sampled SNPs across autosomes in the *i*th admixed sample, β_1 represents the proportion of African ancestry, and β_2 represents the proportion of European ancestry. The model is iterated over every *k*th sample from Europe and we used a sample from Zambia (sample ID = ZM_Sou_Sia_1_2010-07-16) to represent the African ancestor. We report the mean ancestry coefficients for each admix sample as the mean of β_1 for all iterations of European ancestors. For these admixture analyses, we omitted the "collapsed samples" from the (Fournier-Level et al. 2019) dataset. We performed this analysis by sampling SNPs across the entire genome, as well as inside chromosomal inversions, outside of inversions, and on noncoding mutations.

Admixture Estimates: *Moments* (Model-Based)

We also estimated the proportions of African and European admixture in North and South American populations, as well as in Australian samples, using the Python package *moments* [v1.2.2] (Jouganous et al. 2017). We employed a three-population model in which each pool from the Americas or Australia was modeled as a descendant of one African and one European source population. For this analysis, we used a pooled sample from Zambia (sample ID: ZM_Sou_Sia_1_2010-07-16) to represent the African ancestral population and a pooled sample from France (sample ID: FR_Ill_Sai_1_2017-09-16) as the European ancestral population.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Allele frequency data inferred from Pool-Seq were discretized into allele counts using a probabilistic approach implemented in the R package *genomalicious* [v0.7.11] (Thia 2024), with the rounded-down estimate of n_e as the expected number of individuals per pool. The resulting data were imported into Python (v3.12), and a folded allele frequency spectrum was constructed. We then performed 10 rounds of optimization for each population trio in *moments* to estimate admixture proportions. Additional details for this analysis are given in supplementary text S3, Supplementary Material online.

Population Genetics Simulations With SLiM

We used SLiM [v4.2.2] (Haller and Messer 2023) to simulate 999 stepping-stone populations resulting from the secondary encounter of two anciently diverged populations, mirroring the demographic history of *D. melanogaster*. We performed simulations using a non-Wright–Fisher model to generate a single chromosome of length 99,999 bp, with a mutation rate of 1.5×10^{-6} and a recombination rate of 1×10^{-8}.

The simulation consisted of three distinct epochs. In the first epoch, a single ancestral population (p0) of size $N = 5{,}000$ evolved neutrally for 7,999 generations, reaching an average genetic diversity of $\theta \approx 0.01$. At generation 8,000 (Epoch 2), a founder event initiated the formation of a new population (p1) when 0.2% of individuals from p0 colonized a novel habitat. This derived population had a smaller carrying capacity ($N = 2{,}000$) and experienced weak, asymmetric migration with the ancestral population ($m_{p0 \to p1} = 10^{-5}$ and $m_{p1 \to p0} = 10^{-4}$). This epoch continued until generation 14,999, by which time populations p0 and p1 had diverged to an F_{ST} of ~0.2. At generation 15,000 (Epoch 3), p0 and p1 seeded opposite ends of a linear stepping-stone cline (p2 and p8, respectively), with ~0.2% of individuals from each parental population initiating colonization. Prior to this event, genomes of individuals contributing to gene flow were tagged with an ancestry-informative mutation to enable downstream tracking of true ancestry proportions. Following colonization, the two populations expanded inward through six intermediate demes (p3 to p7) arranged in a one-dimensional stepping-stone model. Migration occurred only between neighboring demes, simulating a gradual secondary contact. The simulation proceeded until generation 16,900, at which point allele frequencies were estimated from populations p2 through p8. Prior to demographic inference, we discretized the allele frequencies as if they had been sampled from 25 individuals. We used this data to evaluate the accuracy and behavior of our admixture-estimation methods.

Demographic Model Selection With *Moments*

We fit demographic models to subsets of the DEST 2.0 variant data with the Python package *moments* [v1.2.2] (Jouganous et al. 2017). We used a combination of custom code as well as modified scripts adapted from *moments* code to construct site frequency spectra (SFSs) from autosomal SNPs from the Pool-Seq VCF file. First, we partition the DEST 2.0 VCF (including all autosomes) file as a function of the demographic clusters reported in the results. Each demographic cluster contains a number of localities sampled across space and time. To obtain a representative geographic sample from each cluster, we selected one pooled sample per locality. We chose the pooled sample with the highest n_e (Equation (1)) to represent a given locality. We then constructed a folded SFS for each demographic cluster by averaging the allele frequency of all constituent pools into 21 bins (i.e. discretizing the pool frequencies into counts of 20 diploid chromosomes). The SFS estimation process was repeated 40 times per demographic cluster using a jackknife approach whereby one sample was excluded at random. These jackknifed samples were used for demographic inference. Because the VCF files used to construct these SFS were generated using the PoolSNP caller, we expect that the SFS estimates will underestimate the number of rare alleles across clusters (see Discussion).

Using these data, we constructed *demes*-type models (Gower et al. 2022) dubbed "one-population," "split," "two-splits," and "admixture" (supplementary fig. S9, Supplementary Material online) in order to perform likelihood-based model selection of global *Drosophila* populations. A significant limitation of SFS-based demographic inference (Gutenkunst et al. 2009) is that model likelihoods are calculated from element-wise products of measures of deviations between data and model SFSs, thus making the likelihoods dependent on the number of elements of the SFS. This strategy inhibits comparison of models using classical approaches such as Akaike information criterion or Bayesian information criterion, since our models have different numbers of contemporary populations, whose corresponding SFSs have different numbers of dimensions (i.e. one dimension per population) and thus different numbers of elements. We overcome this limitation by introducing collapsed log-likelihood (CLL), in which direct comparison is enabled by "collapsing the additional populations of higher dimensional SFSs such that all SFSs to be compared have identical minimal shapes. For example, in order to compare three-population models of Europe that include the putative overlap zone (cluster 7 in Fig. 4b) to two-population models of Europe, we independently fit models, then "collapse" the data and model SFSs of the three-population models by summing over the axis representing the overlap zone in order to yield a 2D-SFS with the same shape as the SFSs in the two-population models, and then recalculate the log-likelihood of the collapsed data given the collapsed model SFS in order to achieve the CLL. This method was replicated by collapsing the populations from the Caribbean (cluster 5 in Fig. 4b) as well as from the southern United States (cluster 6 in Fig. 4b) in order to compare two- and one-population models relative to other American populations (cluster 4 in Fig. 4b). Simulated validation of CLL as a robust statistic for selection between models of different dimensions is summarized at supplementary text S4, Supplementary Material online.

Replicable fitting of each model necessitated thousands of replicate runs of *moments* inference through several rounds of manual adjustment of parameter space boundaries, optimization algorithms, and other optimization parameters. The general workflow for each model fit involved initially searching enormous parameter spaces (i.e. spanning orders of magnitude in each parameter's dimension) with the Nelder–Mead algorithm (Nelder and Mead 1965), then performing targeted searches with the BFGS algorithm (Fletcher 1987) until several runs were found to have nonrandomly converged to the same point in parameter space.

To validate model likelihoods and parameter estimates, we employed a jackknifing strategy, in which, for 40 replicates for each model fit to each region, we randomly removed one sample from each population. The hypothesis tests in the Results section compare sets of 40 CLLs from model fits to jackknife replicates.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Population Differentiation

We analyzed patterns of population differentiation across samples and clusters using the R package *poolfstat* [v3.0.0] (Gautier et al. 2022, 2024). This analysis was performed for 528 samples that passed quality filtering and for 9 clusters (clusters defined based on the spatial clustering using $k = 4$ and continent), thus excluding the *D. simulans* sample and "CN_Bei_Bei_1_1992-09-16," on three set of polymorphisms: (i) all chromosomes including heterochromatin; (ii) all chromosomes excluding heterochromatin; and (iii) all chromosomes excluding heterochromatin and SNPs with MAF < 0.05. In all analyses, we considered autosomes and X chromosomes separately to account for differences in pool size between male and female pools. For female-only pools, both autosomes and X chromosomes were counted as twice the number of flies in the pool. For male-only pools (or haploids), autosomes were counted as twice the number of flies, while X chromosomes equaled the number of flies in the pool. To examine pairwise population differentiation, the samples were grouped based on their spatial clusterings at $k = 4$ and $k = 8$ ($k = 8$ clustering results shown in supplementary fig. S13, Supplementary Material online). The *computeFST()* function was first used to estimate the global F_{ST} across all worldwide samples and also within each geographical cluster using the ANOVA method (Hivert et al. 2018).

To further quantify the impact of the structuring of the genetic diversity across continents, we used a hierarchical modeling of differentiation consisting of decomposing overall F_{ST} (here denoted as hF_{ST}) into an across-group (F_{GT}) and within-group (F_{SG}) contribution (Nei 1973; Gautier et al. 2024), as follows:

$$1 - hF_{ST} = (1 - F_{SG})(1 - F_{GT}) \quad (3)$$

with groups of population being defined a priori (e.g. according to their continent of origin and the clustering results as we did in the present study). We estimated these statistics using the unbiased estimator developed for Pool-Seq data implemented in the *computeFST*() function of *poolfstat* [v3.0.0] (Gautier et al. 2024). In addition to whole-genome estimates, window-wise hierarchical F_{ST} parameters were estimated across windows of 10, 50, and 100 kb and are available in the DEST 2.0 browser.

We also explored how recombination rates and gene density correlate with the levels of differentiation (as measured with F_{GT}). We used average recombination rates in 10 kb windows (Comeron et al. 2012; retrieved from Rech 2022). For gene density, we used gene counts in 10 kb windows. As previously reported (Keinan and Reich 2010; Nachman and Payseur 2012), we observed a general negative correlation between population differentiation and recombination rates across most pairwise comparisons, while no clear pattern with gene density (supplementary table S12, Supplementary Material online).

GIM Predictive Models

GIM analyses were conducted in the R package *adegenet* [v2.1.5] using discriminant analysis of the principal component (DAPC) framework (Jombart et al. 2010). While the original GIM set from DEST 1.0 consisted of 30,000 loci, here we use only 28,253 loci. This was done because some of the original markers were filtered out in the current DEST 2.0 panel. We used these markers to train the DAPC model using the sample's state/province as the grouping prior. We retained 30 PCs from the DEST 1.0 model for the state/province model. We retained PCs based on a leave-one-out analysis that minimized the sum of squared errors (SSEs) of the model. In addition, we also trained a second DEST-GIM 1.0 model using city labels (20 PCs were retained for this model; based on minimum SSE). We used 232 samples from DEST 1.0 to train the model and then predicted the provenance of all samples from DEST 2.0.

DAPC models were trained using a cross-validation routine where the data is subdivided into a training (90%) and a testing set (10%) across 30 replicates. For simplicity, we only explored the first 300 PCs across iterations. Parameters were optimized using the lowest mean square error statistic using the *xvalDapc* function in *adegenet*. Predictive GIM models were assessed by estimating the haversine distance (d_{hav}) between the predicted and expected latitude and longitude points. Haversine distances represent the lowest distance between two points across a spherical earth with a radius of 6,378.137 km using the R package geosphere [v.1.5-14] (Hijmans et al. 2022).

Temporal Genetic Structure and Latitudinal Analysis

We assessed levels of temporal structure across DEST by estimating F_{ST} between samples at the same locality collected a year apart from each other. These estimates of F_{ST} reflect differentiation resulting from the overwintering population "bust" across one winter. We call this summary statistic "year-to-year F_{ST}" as it captures levels of genetic variation for the population before and after a winter season. We correlated this data to latitude and performed a broken-stick regression analysis using the *segmented* [v.2.0-4] R package (Muggeo 2003).

Scans for Adaptive Differentiation

We tested for adaptive differentiation at ~908,543 SNPs that were polymorphic in a set of seasonally collected samples from across Europe (supplementary table S13, Supplementary Material online). First, we implemented the *BayPass* [v2.4] model for adaptive differentiation using the *XtX** test statistic (Olazcuaga et al. 2020) while controlling for population structure using a matrix of genetic relatedness (i.e. Ω matrix). We estimated the *XtX** for every autosomal SNP in the genome using five independent runs of *BayPass* and took the median value per SNP. We also generated a null distribution of *XtX** using the POD method outlined in Gautier (2015) and Olazcuaga et al. (2020). Briefly, we simulated allele frequencies for ~9 M SNPs, ten times the number of observed SNPs used in this analysis. We then generated empirical *P*-values for the observed *XtX** statistics by calculating the upper-tail probability of the observed data relative to the simulated POD data. We used the weighted Z analysis (wZa; Booker et al. 2024) to identify windows of signal enrichment across the genome. The wZa statistic combines the empirical *P*-values within a window for each test using Stouffer's method (Stouffer et al. 1949) weighted by average heterozygosity. We applied this approach in a sliding window approach with a window size of 100 kb and a step size of 50 kb.

Second, we ran the *BayPass* model including both the Ω matrix as a demographic prior as well as "spring" and "fall" labels as a proxy for seasonal selection pressures. We designated

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

the spring" sample as the first sample within a year, and the "fall" sample as the last sample within the year. Several samples from DEST 1.0 were characterized by the collectors as "spring" or "fall." For those samples, this label was used in the analysis. For more recent samples, including most sampled in DEST 2.0, samples are labeled as a function of date of collection. For such samples, we assigned seasonal labels by selecting the first and last sample collected in a locality within a year. For each SNP, we estimated the contrast statistics (C_2) with five independent runs of *BayPass* and took the median value. To generate a null distribution of C_2 statistics, we used the simulated SNP data described above, and ran *BayPass* five times. We took the median C_2 of the simulated data as our null distribution and calculated empirical *P*-values as described above. We performed a sliding window analysis of these empirical *P*-values using the wZa method.

Third, we implemented a GLMM approach that is similar to that applied previously by Machado et al. (2021). We modeled allele frequency at each SNP *i* using two models:

$$p_i = \alpha + X(\text{year}_{\text{factor}} : \text{locality}_{\text{factor}}) + \varepsilon \quad (4)$$

$$p_i = \alpha + \beta_1\,(\text{season}) + X(\text{year}_{\text{factor}} : \text{locality}_{\text{factor}}) + \varepsilon \quad (5)$$

Where p_i is the allele frequency at the *i*th locus, α is the intercept term, and β_1 is the term associated with season, X is the random effect term coded as an interaction term between the year of collection and the locality where flies were collected, and ϵ is the binomially distributed error. We assessed the statistical significance of the seasonal β_1 term using a likelihood ratio test between Equations (4) and (5). GLMMs were fitted using the *glmer* function of the *lme4* version [v1.1-35.5] package in R (Bates et al. 2015). We performed a permutation analysis following the methods outlined in Machado et al. (2021) by shuffling the seasonal labels 100 times and rerunning the GLMM analysis for each permutation. We conducted a sliding window analysis of the GLMM.

GO-Term Enrichment Analysis

We performed gene-ontology enrichment analysis using GOWINDA [v.1.12] (Kofler and Schlötterer 2012) in gene mode (with parameters: --min-genes 5 --min-significance 1 --simulations 100,000) on genes located in 10 kb windows of high differentiation ($F_{GT} > 0.2$; supplementary table S7, Supplementary Material online), $-\log_{10}$(wZa *P*-values) > 188.96 for the *XtX** statistic (supplementary table S8, Supplementary Material online), and $-\log_{10}$(wZa *P*-values) > 3.65 for the C_2 statistic (supplementary table S9, Supplementary Material online), representing the 99.9th percentile from the simulated POD data (see above).

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Ethics Statements

Fruit flies were collected either on public lands, where no permits are needed, or in private lands with explicit permission from the relevant stakeholders. To comply with the Nagoya protocol, material transfer agreements (MTAs) were secured and exchanged among researchers prior to transporting fly samples (for all new samples reported here) across borders. Permit MAE-DNB-CM-2015-0030, from the Environmental Ministry of Ecuador, was obtained by Vela to collect, export and perform molecular analysis on samples.

Acknowledgments

The authors are grateful to two anonymous reviewers for their helpful comments on our manuscript. The authors are indebted to all members of the DrosEU and DrosRTEC consortia for their support, collaboration, and for discussion over the years. DrosEU was funded by a Special Topic Networks (STN) grant from the European Society for Evolutionary Biology (ESEB). J.C.B.N. acknowledges the Henderson–Harris fellowship program at the University of Vermont, also the Vermont Advanced Computing Center (URL: https://www.uvm.edu/vacc) for providing computational resources that contributed to this publication. A.O.B. acknowledges Research Computing at the University of Virginia (URL: https://rc.virginia.edu) for providing computational resources and technical support that have contributed to the results reported within this publication. A.O.B. also acknowledges Kathryn Linehan from UVA's RC for support in updating the SNP calling pipeline that is part of DEST2.0. M.C.-Z. and J.G. acknowledge the Galician Supercomputing Center (CESGA), which provided access to its supercomputing infrastructure, the supercomputer FinisTerrae III and its permanent data storage system, funded by the Spanish Ministry of Science and Innovation, the Galician Government, and the European Regional Development Fund (ERDF). M.G. acknowledges the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. D.O. acknowledges Sue and Keith Obbard and Sandy Bayne for permission to collect flies on their land. M.H.A. acknowledges the Department of Evolution and Ecology at the University of Freiburg (Germany) for providing the necessary resources and support for sample preparations and DNA extractions. S.V.S. acknowledges support from the PAUSE-ANR Ukraine Program. The authors also thank Pavlo A. Kovalenko and Nadiia M. Pirko for their assistance with collecting flies in 2017 to 2021. After 2022 February 24, no collaborative actions or exchanges have taken place within our project between Ukrainian and Russian scientists nor their institutions.

Funding

J.C.B.N. was supported by start-up funds from the University of Vermont. M.Kap. was supported by the Horizon Europe project FAIRiCUBE (grant #101059238). S.S. was supported by the Horizon Europe project FAIRiCUBE (grant #101059238). D.A.P. was supported by the NIH 2R35GM11816506 (MIRA grant). T.F. was supported by the Swiss National Science Foundation (SNSF) grants 31003A-182262, 310030_219283, and FZEB-0-214654. A.O.B. was supported by the National Institutes of Health R35 GM119686 and National Science Foundation CAREER #2145688 grants. J.G. was supported by grant PID2020-115874GB-I00 funded by MICIU/AEI /10.13039/501100011033, grant PID2023-148838NB-I00 funded by MICIU/AEI/10.13039/501100011033 and FEDER/EU, and grant 2021 SGR 00417 funded by the Departament de Recerca i Universitats, Generalitat de Catalunya. A.S.-G. was supported by the Ministerio de Ciencia e Innovación of

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Spain (MCIN/AEI/10.13039/501100011033; grant PID2020-113168GB-I00) and Comissió Interdepartamental de Recerca I Innovació Tecnològica of Catalonia, Spain (2021SGR00279). A.P. and M.T. were supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA), grant no. 451-03-66/2024-03/200007. A.B., M.P.G.G., and S.Casi. were supported by Ministerio de Ciencia e Innovación (PID2021-127107NB-I00) and AGAUR Generalitat de Catalunya (SGR 00526). C.S. was supported by the Austrian Science Funds, FWF, 10.55776/P32935, 10.55776/P33734. C.Fr. was supported by the German Research Foundation (DFG, grant # FR2973/11-1). D.O. was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/T007516/1. D.V. was supported by project QINV0196-IINV529010100 from the Pontificia Universidad Católica del Ecuador; Abbott was supported by VR-2015-04680 and VR-2020-05412. J.P. was supported by the Deutsche Forschungsgemeinschaft (DFG) projects 255619725 and 503272152. M.Kan. was supported by the Academy of Finland project 322980. M.S.V., M.J., and M.Ra. were supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA), grant no. 451-03-65/2024-03/ 200178. M.S.-R. was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-47/2023-01/200178. M.G.R. was supported by Natural Environment Research Council (NERC), UK award NE/V001566/1. M.Re. was supported by the Bettencourt Schueller Foundation long-term partnership; this work was also partly supported by a CRI Core Research Fellowship. P.A.E. was supported by award #61-1673 from the Jane Coffin Childs Memorial Fund for Medical Research (www.jccfund.org). S.E.R.-O. was supported by PID2020-119255GB-I00 (MICINN, Spain), by the CERCA Programme/Generalitat de Catalunya and acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the Severo Ochoa Programme for Centres of Excellence in R&D 2016–2019 and 2020–2023 (SEV-2015-0533, CEX2019-000917) and the European Regional Development Fund (ERDF). N.H. and C.L. were supported by Australian Research Council DP190102512. E.K. was supported by the European Molecular Biology Organization (EMBO) long-term fellowship ALT 248-02018. H.C. was supported by the French National Research Agency (ANR) project Drothermal (ANR-20-CE02-011-01).

Author Contributions

All author contributions to this work are denoted in supplementary table S14, Supplementary Material online.

Data Availability

The DEST 2.0 browser is built on the latest version of JBrowse 2 (Diesh et al. 2023), an enhanced successor to JBrowse 1, which powered the original DEST 1.0 browser (Kapun et al. 2021). JBrowse 2.0 offers improved performance through a modern software architecture that supports parallel rendering of tracks and allows for the visualization of new data types, such as VCF files. Similar to the first DEST browser, it features a user-friendly data selector that facilitates the selection of the multiple population genetic metrics and statistics compiled for the DEST 2.0 release (supplementary fig. S16, Supplementary Material online). Additionally, the browser provides a portal for downloading allelic information and precomputed population genetics statistics in multiple formats, along with a usage tutorial featuring worked examples. Bulk downloads of all compiled tracks are available in BigWig format (Kent et al. 2010), and Pool-Seq files (in VCF format) can be accessed through a dedicated data directory. All data, tools, and supporting resources for the DEST 2.0 release, including reference tracks from FlyBase (v.6.12; Dos Santos et al. 2015), are freely available through the website (https://dest.bio). The browser operates on an Apache server running CentOS 7.2 Linux x64, powered by 16 Intel Xeon 2.4 GHz processors and 32 GB of RAM. All new sequences reported here for the first time are available on the SRA (https://www.ncbi.nlm.nih.gov/sra) at PRJNA993612 and PRJNA1263695. Code is available in GitHub at: https://github.com/DEST-bio/DESTv2_data_paper and https://github.com/DEST-bio/DESTv2. All outputs from the DEST 2.0 pipeline can be found at https://dest.bio. Supplementary datasets can be found in Zenodo at https://doi.org/10.5281/zenodo.13731977.

References

Adrion JR, Galloway JG, Kern AD. Predicting the landscape of recombination using deep learning. *Mol Biol Evol*. 2020:37(6):1790–1808. https://doi.org/10.1093/molbev/msaa038.

Adrion JR, Hahn MW, Cooper BS. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends Genet*. 2015:31(8):434–444. https://doi.org/10.1016/j.tig.2015.05.006.

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. The genetic architecture of adaptations to high altitude in Ethiopia. *PLoS Genet*. 2012:8(12):e1003110. https://doi.org/10.1371/journal.pgen.1003110.

Anderson AR, Hoffmann AA, Mckechnie SW, Umina PA, Weeks AR. The latitudinal cline in the *In(3R)Payne* inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Mol Ecol*. 2005:14(3):851–858. https://doi.org/10.1111/j.1365-294X.2005.02445.x.

Andolfatto P. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol Biol Evol*. 2001:18(3):279–290. https://doi.org/10.1093/oxfordjournals.molbev.a003804.

Arguello JR, Laurent S, Clark AG. Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biol Evol*. 2019:11(3):844–854. https://doi.org/10.1093/gbe/evz022.

Atkinson W, Shorrocks B. Breeding site specificity in the domestic species of *Drosophila*. *Oecologia*. 1977:29(3):223–232. https://doi.org/10.1007/BF00345697.

Bangerter A. Dense seasonal sampling of an orchard population uncovers population turnover, adaptive tracking, and structure in multiple Drosophila species. 2021. [accessed 2023 Nov 15]. https://libraetd.lib.virginia.edu/public_view/2801ph17g.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015:67(1):1–48. https://doi.org/10.18637/jss.v067.i01.

Begun DJ, Aquadro CF. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature*. 1993:365(6446):548–550. https://doi.org/10.1038/365548a0.

Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, Lazzaro BP, Schmidt PS. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc Biol Sci*. 2018:285(1870):20172599. https://doi.org/10.1098/rspb.2017.2599.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Behrman EL, Schmidt P. How predictable is rapid evolution? bioRxiv 514123. https://doi.org/10.1101/2022.10.27.514123, 28 October 2022, preprint: not peer reviewed.
Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS. Seasonal variation in life history traits in two *Drosophila* species. *J Evol Biol*. 2015:28(9):1691–1704. https://doi.org/10.1111/jeb.12690.
Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999:27(2):573–580. https://doi.org/10.1093/nar/27.2.573.
Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet*. 2014:10(11):e1004775. https://doi.org/10.1371/journal.pgen.1004775.
Bergland AO, Tobler R, González J, Schmidt P, Petrov D. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol Ecol*. 2016:25(5):1157–1174. https://doi.org/10.1111/mec.13455.
Berry A, Kreitman M. Molecular analysis of an allozyme cline: alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. *Genetics*. 1993:134(3):869–893. https://doi.org/10.1093/genetics/134.3.869.
Betancourt NJ, Rajpurohit S, Durmaz E, Fabian DK, Kapun M, Flatt T, Schmidt P. Allelic polymorphism at *foxo* contributes to local adaptation in *Drosophila melanogaster*. *Mol Ecol*. 2021:30(12):2817–2830. https://doi.org/10.1111/mec.15939.
Bitter MC, Berardi S, Oken H, Huynh A, Lappo E, Schmidt P, Petrov DA. Continuously fluctuating selection reveals fine granularity of adaptation. *Nature*. 2024:634(8033):389 396. https://doi.org/10.1038/s41586-024-07834-x.
Boettiger C. An introduction to Docker for reproducible research. *ACM SIGOPS Oper Syst Rev*. 2015:49(1):71–79. https://doi.org/10.1145/2723872.2723882.
Bogaerts-Márquez M, Guirao-Rico S, Gautier M, González J. Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of *Drosophila melanogaster*. *Mol Ecol*. 2021:30(4):938–954. https://doi.org/10.1111/mec.15783.
Booker TR, Yeaman S, Whiting JR, Whitlock MC. The WZA: a window-based method for characterizing genotype–environment associations. *Mol Ecol Resour*. 2024:24(2):e13768. https://doi.org/10.1111/1755-0998.13768.
Botero CA, Weissing FJ, Wright J, Rubenstein DR. Evolutionary tipping points in the capacity to adapt to environmental change. *Proc Natl Acad Sci U S A*. 2015:112(1):184–189. https://doi.org/10.1073/pnas.1408589111.
Buri P. Gene frequency in small populations of mutant *Drosophila*. *Evolution*. 1956:10(4):367–402. https://doi.org/10.2307/2406998.
Bushnell B, Rood J, Singer E. BBMerge—accurate paired shotgun read merging via overlap. *PLoS One*. 2017:12(10):e0185056. https://doi.org/10.1371/journal.pone.0185056.
Campo D, Lehmann K, Fjeldsted C, Souaiaia T, Kao J, Nuzhdin SV. Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Mol Ecol*. 2013:22(20):5084–5097. https://doi.org/10.1111/mec.12468.
Capy P, David JR, Allemand R, Carton Y, Febvay G, Kermarec A. Genetic analysis of *Drosophila melanogaster* in the French West Indies and comparison with populations from other parts of the world. *Genetica*. 1986:69(3):167–176. https://doi.org/10.1007/BF00133519.
Caracristi G, Schlötterer C. Genetic differentiation between American and European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol*. 2003:20(5):792–799. https://doi.org/10.1093/molbev/msg091.
Carvalho J, Morales HE, Faria R, Butlin RK, Sousa VC. Integrating pool-seq uncertainties into demographic inference. *Mol Ecol Resour*. 2023:23(7):1737–1755. https://doi.org/10.1111/1755-0998.13834.
Casillas S, Barbadilla A. Molecular population genetics. *Genetics*. 2017:205(3):1003–1035. https://doi.org/10.1534/genetics.116.196493.
Chen J, Liu C, Li W, Zhang W, Wang Y, Clark AG, Lu J. From sub-Saharan Africa to China: evolutionary history and adaptation of *Drosophila melanogaster* revealed by population genomics. *Sci Adv*. 2024:10(16):eadh3425. https://doi.org/10.1126/sciadv.adh3425.
Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. *Fly (Austin)*. 2012:6(2):80–92. https://doi.org/10.4161/fly.19695.
Cogni R, Kuczynski C, Koury S, Lavington E, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF. The intensity of selection acting on the *couch potato* gene-spatial-temporal variation in a diapause cline: spatial-temporal variation in diapause cline. *Evolution*. 2014:68(2):538–548. https://doi.org/10.1111/evo.12291.
Comeron JM, Ratnappan R, Bailin S. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet*. 2012:8(10):e1002905. https://doi.org/10.1371/journal.pgen.1002905.
Corbett-Detig R, Nielsen R. A hidden Markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLoS Genet*. 2017:13(1):e1006529. https://doi.org/10.1371/journal.pgen.1006529.
Coughlan JM, Dagilis AJ, Serrato-Capuchina A, Elias H, Peede D, Isbell K, Castillo DM, Cooper BS, Matute DR. Patterns of population structure and introgression among recently differentiated *Drosophila melanogaster* populations. *Mol Biol Evol*. 2022:39(11):msac223. https://doi.org/10.1093/molbev/msac223.
Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al*. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008.
Danielson PB, Letman JA, Fogleman JC. Alkaloid metabolism by cytochrome P-450 enzymes in Drosophila melanogaster. *Comp Biochem Physiol B Biochem Mol Biol*. 1995:110(4):683–688. https://doi.org/10.1016/0305-0491(94)00214-F.
David J, Capy P. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet*. 1988:4(4):106–111. https://doi.org/10.1016/0168-9525(88)90098-4.
De Jong G, Bochdanovits Z. Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *J Genet*. 2003:82(3):207–223. https://doi.org/10.1007/BF02715819.
DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, *et al*. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011:43(5):491–498. https://doi.org/10.1038/ng.806.
Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, *et al*. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol*. 2023:24(1):74. https://doi.org/10.1186/s13059-023-02914-z.
Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM; The FlyBase Consortium. FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res*. 2015:43(Database issue):D690–D697. https://doi.org/10.1093/nar/gku1099.
Dreissig S, Mascher M, Heckmann S. Variation in recombination rate is shaped by domestication and environmental conditions in barley. *Mol Biol Evol*. 2019:36(9):2029–2039. https://doi.org/10.1093/molbev/msz141.
Duchen P, Živković D, Hutter S, Stephan W, Laurent S. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics*. 2013:193(1):291–301. https://doi.org/10.1534/genetics.112.145912.
Ďureje Ľ, Macholán M, Baird SJE, Piálek J. The mouse hybrid zone in Central Europe: from morphology to molecules. *Folia Zool*. 2012:61(3-4):308–318. https://doi.org/10.25225/fozo.v61.i3.a13.2012.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Durmaz E, Rajpurohit S, Betancourt N, Fabian DK, Kapun M, Schmidt P, Flatt T. A clinal polymorphism in the insulin signaling transcription factor *foxo* contributes to life-history adaptation in *Drosophila*. *Evolution*. 2019:73(9):1774–1792. https://doi.org/10.1111/evo.13759.

Erickson PA, Weller CA, Song DY, Bangerter AS, Schmidt P, Bergland AO. Unique genetic signatures of local adaptation over space and time for diapause, an ecologically relevant complex trait, in Drosophila melanogaster. *PLoS Genet*. 2020:16(11):e1009110. https://doi.org/10.1371/journal.pgen.1009110.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol Ecol*. 2012:21(19):4748–4769. https://doi.org/10.1111/j.1365-294X.2012.05731.x.

Feder AF, Petrov DA, Bergland AO. LDx: estimation of linkage disequilibrium from high-throughput pooled resequencing data. *PLoS One*. 2012:7(11):e48588. https://doi.org/10.1371/journal.pone.0048588.

Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. Population genomics from pool sequencing. *Mol Ecol*. 2013:22(22):5561–5576. https://doi.org/10.1111/mec.12522.

Flatt T. Life-history evolution and the genetics of fitness components in *Drosophila melanogaster*. *Genetics*. 2020:214(1):3–48. https://doi.org/10.1534/genetics.119.300160.

Fletcher R. *Practical methods of optimization*. 2nd ed. Chichester; New York: Wiley; 1987.

Fournier-Level A, Good RT, Wilcox SA, Rane RV, Schiffer M, Chen W, Battlay P, Perry T, Batterham P, Hoffmann AA, *et al*. The spread of resistance to imidacloprid is restricted by thermotolerance in natural populations of *Drosophila melanogaster*. *Nat Ecol Evol*. 2019:3(4):647–656. https://doi.org/10.1038/s41559-019-0837-y.

Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*. 2015:201(4):1555–1579. https://doi.org/10.1534/genetics.115.181453.

Gautier M. Efficient k-mer based curation of raw sequence data: application in Drosophila suzukii. *Peer Community J*. 2023:3:e79. https://doi.org/10.24072/pcjournal.309.

Gautier M, Coronado-Zamora M, Vitalis R. Estimating hierarchical F –statistics from pool–seq data. bioRxiv 624688. https://doi.org/10.1101/2024.11.22.624688, 22 November 2024, preprint: not peer reviewed.

Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol*. 2013:22(14):3766–3779. https://doi.org/10.1111/mec.12360.

Gautier M, Vitalis R, Flori L, Estoup A. *F*-statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package *poolfstat*. *Mol Ecol Resour*. 2022:22(4):1394–1416. https://doi.org/10.1111/1755-0998.13557.

Glaser-Schmitt A, Ramnarine TJS, Parsch J. Rapid evolutionary change, constraints and the maintenance of polymorphism in natural populations of *Drosophila melanogaster*. *Mol Ecol*. 2024:33(10):e17024. https://doi.org/10.1111/mec.17024.

Gleason JM, Roy PR, Everman ER, Gleason TC, Morgan TJ. Phenology of *Drosophila* species across a temperate growing season and implications for behavior. *PLoS One*. 2019:14(5):e0216601. https://doi.org/10.1371/journal.pone.0216601.

Gower G, Ragsdale AP, Bisschop G, Gutenkunst RN, Hartfield M, Noskova E, Schiffels S, Struck TJ, Kelleher J, Thornton KR. Demes: a standard format for demographic models. *Genetics*. 2022:222(3):iyac131. https://doi.org/10.1093/genetics/iyac131.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA; The 1000 Genomes Project; Bustamante CD, *et al*. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A*. 2011:108(29):11983–11988. https://doi.org/10.1073/pnas.1019276108.

Guirao-Rico S, González J. Evolutionary insights from large scale resequencing datasets in Drosophila melanogaster. *Current Opinion in Insect Science*. 2019:31:70–76. https://doi.org/10.1016/j.cois.2018.11.002.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD, McVean G. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet*. 2009:5(10):e1000695. https://doi.org/10.1371/journal.pgen.1000695.

Hales KG, Korey CA, Larracuente AM, Roberts DM. Genetics on the fly: a primer on the *Drosophila* model system. *Genetics*. 2015:201(3):815–842. https://doi.org/10.1534/genetics.115.183392.

Haller BC, Messer PW. SLiM 4: multispecies eco-evolutionary modeling. *Am Nat*. 2023:201(5):E127–E139. https://doi.org/10.1086/723601.

Haudry A, Laurent S, Kapun M. Population genomics on the fly: recent advances in *Drosophila*. *Methods Mol Biol*. 2020:2090:357–396. https://doi.org/10.1007/978-1-0716-0199-0_15.

Hewitt GM. Quaternary phylogeography: the roots of hybrid zones. *Genetica*. 2011:139(5):617–638. https://doi.org/10.1007/s10709-011-9547-3.

Hijmans RJ, Karney C, Williams E, Vennes C. 2022. Package 'geosphere'. https://doi.org/10.32614/CRAN.package.geosphere.

Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. Measuring genetic differentiation from pool-seq data. *Genetics*. 2018:210(1):315–330. https://doi.org/10.1534/genetics.118.300900.

Hoffmann AA, Anderson A, Hallas R. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecol Lett*. 2002:5(5):614–618. https://doi.org/10.1046/j.1461-0248.2002.00367.x.

Hoffmann AA, Weeks AR. Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. *Genetica*. 2007:129(2):133. https://doi.org/10.1007/s10709-006-9010-z.

Hofman S, Spolsky C, Uzzell T, Cogălniceanu D, Babik W, Szymura JM. Phylogeography of the fire-bellied toads *Bombina*: independent pleistocene histories inferred from mitochondrial genomes. *Mol Ecol*. 2007:16(11):2301–2316. https://doi.org/10.1111/j.1365-294X.2007.03309.x.

Hunter CM, Huang W, Mackay TFC, Singh ND. The genetic architecture of natural variation in recombination rate in *Drosophila melanogaster*. *PLoS Genet*. 2016:12(4):e1005951. https://doi.org/10.1371/journal.pgen.1005951.

Ives PT. The genetic structure of American populations of *Drosophila melanogaster*. *Genetics*. 1945:30(2):167–196. https://doi.org/10.1093/genetics/30.2.167.

Ives PT. Further genetic studies of the South Amherst population of *Drosophila melanogaster*. *Evol. Int. J. Org. Evol*. 1970:24(3):507–518. https://doi.org/10.2307/2406830.

Izquierdo JI. How does *Drosophila melanogaster* overwinter? *Entomol Exp Appl*. 1991:59(1):51–58. https://doi.org/10.1111/j.1570-7458.1991.tb01485.x.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*. 2005:170(3):1401–1410. https://doi.org/10.1534/genetics.104.038224.

Johnson OL, Tobler R, Schmidt JM, Huber CD. Fluctuating selection and the determinants of genetic variation. *Trends Genet*. 2023:39(6):491–504. https://doi.org/10.1016/j.tig.2023.02.004.

Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010:11(1):94. https://doi.org/10.1186/1471-2156-11-94.

Jouganous J, Long W, Ragsdale AP, Gravel S. Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*. 2017:206(3):1549–1567. https://doi.org/10.1534/genetics.117.200493.

Jurka J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet*. 2000:16(9):418–420. https://doi.org/10.1016/S0168-9525(00)02093-X.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. *Mol Ecol*. 2015:24(7):1499–1509. https://doi.org/10.1111/mec.13137.

Kapopoulou A, Kapun M, Pieper B, Pavlidis P, Wilches R, Duchen P, Stephan W, Laurent S. Demographic analyses of a new sample of haploid genomes from a Swedish population of *Drosophila melanogaster*. *Sci Rep*. 2020:10(1):22415. https://doi.org/10.1038/s41598-020-79720-1.

Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-Stabelli O, Kankare M, Bogaerts-Márquez M, *et al*. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol Biol Evol*. 2020:37(9):2661–2678. https://doi.org/10.1093/molbev/msaa120.

Kapun M, Fabian DK, Goudet J, Flatt T. Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Mol Biol Evol*. 2016a:33(5):1317–1336. https://doi.org/10.1093/molbev/msw016.

Kapun M, Flatt T. The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol Ecol*. 2019:28(6):1263–1282. https://doi.org/10.1111/mec.14871.

Kapun M, Mitchell ED, Kawecki TJ, Schmidt P, Flatt T. An ancestral balanced inversion polymorphism confers global adaptation. *Mol Biol Evol*. 2023:40(6):msad118. https://doi.org/10.1093/molbev/msad118.

Kapun M, Nunez JCB, Bogaerts-Márquez M, Murga-Moreno J, Paris M, Outten J, Coronado-Zamora M, Tern C, Rota-Stabelli O, Guerreiro MPG, *et al*. *Drosophila* evolution over space and time (DEST): a new population genomics resource. *Mol Biol Evol*. 2021:38(12):5782–5805. https://doi.org/10.1093/molbev/msab259.

Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T. Parallel effects of the inversion *In(3R)Payne* on body size across the North American and Australian clines in *Drosophila melanogaster*. *J Evol Biol*. 2016b:29(5):1059–1072. https://doi.org/10.1111/jeb.12847.

Kapun M, Van Schalkwyk H, McAllister B, Flatt T, Schlötterer C. Inference of chromosomal inversion dynamics from pool-seq data in natural and laboratory populations of *Drosophila melanogaster*. *Mol Ecol*. 2014:23(7):1813–1827. https://doi.org/10.1111/mec.12594.

Keinan A, Reich D. Human population differentiation is strongly correlated with local recombination rate. *PLoS Genet*. 2010:6(3):e1000886. https://doi.org/10.1371/journal.pgen.1000886.

Keller A. *Drosophila melanogaster*'s history as a human commensal. *Curr Biol*. 2007:17(3):R77–R81. https://doi.org/10.1016/j.cub.2006.12.031.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*. 2010:26(17):2204–2207. https://doi.org/10.1093/bioinformatics/btq351.

Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009:25(17):2283–2285. https://doi.org/10.1093/bioinformatics/btp373.

Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. Popoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*. 2011:6(1):e15925. https://doi.org/10.1371/journal.pone.0015925.

Kofler R, Schlötterer C. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*. 2012:28(15):2084–2085. https://doi.org/10.1093/bioinformatics/bts315.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*. 2011:187(1):245–260. https://doi.org/10.1534/genetics.110.123059.

Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012:28(19):2520–2522. https://doi.org/10.1093/bioinformatics/bts480.

Kreitman M. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*. 1983:304(5925):412–417. https://doi.org/10.1038/304412a0.

Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol Biol*.. 1988:22:159–225. 10.1007/978-1-4613-0931-4_4.

Lachaise D, Silvain J-F. How two afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*–*D. simulans* palaeogeographic riddle. *Genetica*. 2004:120(1–3):17–39. https://doi.org/10.1023/B:GENE.0000017627.27537.ef.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*. 2015:199(4):1229–1241. https://doi.org/10.1534/genetics.115.174664.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. A thousand fly genomes: an expanded *Drosophila* genome nexus. *Mol Biol Evol*. 2016:33(12):3308–3313. https://doi.org/10.1093/molbev/msw195.

Lange JD, Bastide H, Lack JB, Pool JE. A population genomic assessment of three decades of evolution in a natural *Drosophila* population. *Mol Biol Evol*. 2022:39(2):msab368. https://doi.org/10.1093/molbev/msab368.

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, *et al*. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*. 2012:192(2):533–598. https://doi.org/10.1534/genetics.112.142018.

Lawton D, Huseth AS, Kennedy GG, Morey AC, Hutchison WD, Reisig DD, Dorman SJ, Dillard D, Venette RC, Groves RL, *et al*. Pest population dynamics are related to a continental overwintering gradient. *Proc Natl Acad Sci U S A*. 2022:119(37):e2203230119. https://doi.org/10.1073/pnas.2203230119.

Lê S, Josse J, Husson F. FactoMineR: an *R* package for multivariate analysis. *J Stat Softw*. 2008:25(1):1–18. https://doi.org/10.18637/jss.v025.i01.

Le Goff G, Hilliou F. Resistance evolution in *Drosophila*: the case of *CYP6G1*. *Pest Manag Sci*. 2017:73(3):493–499. https://doi.org/10.1002/ps.4470.

Lemeunier F, Aulard S. Inversion polymorphism in *Drosophila melanogaster*. In: Krimbas CB, Powell JR, editors. *Drosophila Inversion polymorphism*. Boca Raton (FL): CRC Press; 1992. p. 339–405.

Lewontin RC. *The genetic basis of evolutionary change*. New York: Columbia University Press; 1974.

Lhotka O, Kyselý J. The 2021 European heat wave in the context of past major heat waves. *Earth Space Sci*. 2022:9(11):e2022EA002567. https://doi.org/10.1029/2022EA002567.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv, arXiv:1303.3997, preprint: not peer reviewed.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in Drosophila. *PLoS Genet*. 2006:2(10):e166. https://doi.org/10.1371/journal.pgen.0020166.

Lintner JA. *First annual report on the injurious and other insects of the state of New York*. Albany (NY): Weed, Parsons and Co; 1882.

Machado HE, Bergland AO, O'Brien KR, Behrman EL, Schmidt PS, Petrov DA. Comparative population genomics of latitudinal variation in *Drosophila simulans* and *Drosophila melanogaster*. *Mol Ecol*. 2016:25(3):723–740. https://doi.org/10.1111/mec.13446.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González J, Karasov TL, *et al.* Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila*. *Elife*. 2021:10:e67577. https://doi.org/10.7554/eLife.67577.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, *et al.* The *Drosophila* melanogaster genetic reference panel. *Nature*. 2012:482(7384):173–178. https://doi.org/10.1038/nature10811.

Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC. Wild African *Drosophila melanogaster* are seasonal specialists on Marula fruit. *Curr Biol*. 2018:28(24):3960–-3968.e3. https://doi.org/10.1016/j.cub.2018.10.033.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011:17(1):10. https://doi.org/10.14806/ej.17.1.200.

Mateo L, Rech GE, González J. Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. *Sci Rep*. 2018:8(1):16143. https://doi.org/10.1038/s41598-018-34267-0.

McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. 1991:351(6328):652–654. https://doi.org/10.1038/351652a0.

Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, *et al.* Sustainable data analysis with snakemake. *F1000Res*. 2021:10:33. https://doi.org/10.12688/f1000research.29032.2.

Morgulis A, Gertz EM, Schaffer AA, Agarwala R. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*. 2006:22(2):134–141. https://doi.org/10.1093/bioinformatics/bti774.

Muggeo VMR. Estimating regression models with unknown breakpoints. *Stat Med*. 2003:22(19):3055–3071. https://doi.org/10.1002/sim.1545.

Nachman MW, Payseur BA. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos Trans R Soc B Biol Sci*. 2012:367(1587):409–421. https://doi.org/10.1098/rstb.2011.0249.

Nei M. Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A*. 1973:70(12):3321–3323. https://doi.org/10.1073/pnas.70.12.3321.

Nelder JA, Mead R. A simplex method for function minimization. *Comput J*. 1965:7(4):308–313. https://doi.org/10.1093/comjnl/7.4.308.

Nunez JCB, Lenhart BA, Bangerter A, Murray CS, Mazzeo GR, Yu Y, Nystrom TL, Tern C, Erickson PA, Bergland AO. A cosmopolitan inversion facilitates seasonal adaptation in overwintering *Drosophila*. Ralph P, editor. *Genetics*. 2024:226(2):iyad207. https://doi.org/10.1093/genetics/iyad207.

Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol Biol Evol*. 2012:29(11):3459–3473. https://doi.org/10.1093/molbev/mss150.

O'Connor TD, Fu W, Mychaleckyj JC, Logsdon B, Auer P, Carlson CS, Leal SM, Smith JD, Rieder MJ, Bamshad MJ, *et al.* Rare variation facilitates inferences of fine-scale population structure in humans. *Mol Biol Evol*. 2015:32(3):653–660. https://doi.org/10.1093/molbev/msu326.

Olazcuaga L, Loiseau A, Parrinello H, Paris M, Fraimout A, Guedot C, Diepenbrock LM, Kenis M, Zhang J, Chen X, *et al.* A whole-genome scan for association with invasion success in the fruit fly *Drosophila suzukii* using contrasts of allele frequencies corrected for population structure. *Mol Biol Evol*. 2020:37(8):2369–2385. https://doi.org/10.1093/molbev/msaa098.

Ometto L, Glinka S, De Lorenzo D, Stephan W. Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol Biol Evol*. 2005:22(10):2119–2130. https://doi.org/10.1093/molbev/msi207.

Otte KA, Nolte V, Mallard F, Schlötterer C. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biol*. 2021:22(1):211. https://doi.org/10.1186/s13059-021-02425-9.

Paaby AB, Bergland AO, Behrman EL, Schmidt PS. A highly pleiotropic amino acid polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation: ADAPTIVE POLYMORPHISM AT *InR*. *Evolution*. 2014:68(12):3395–3409. https://doi.org/10.1111/evo.12546.

Parsons PA. The comparative evolutionary biology of the sibling species, *Drosophila melanogaster* and *D. Simulans*. *Q Rev Biol*. 1975:50(2):151–169. https://doi.org/10.1086/408437.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. *Genetics*. 2012:192(3):1065–1093. https://doi.org/10.1534/genetics.112.145037.

Pavlidis P, Jensen JD, Stephan W. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*. 2010:185(3):907–922. https://doi.org/10.1534/genetics.110.116459.

Pitchers W, Pool JE, Dworkin I. Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many?: altitude and wing morphology in *Drosophila melanogaster*. *Evolution*. 2013:67(2):438–452. https://doi.org/10.1111/j.1558-5646.2012.01774.x.

Powell JR. *Progress and prospects in evolutionary biology: the Drosophila model*. Oxford University Press; 1997. 10.1093/oso/9780195076912.001.0001.

Rajpurohit S, Gefen E, Bergland AO, Petrov DA, Gibbs AG, Schmidt PS. Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Mol Ecol*. 2018:27(17):3525–3540. https://doi.org/10.1111/mec.14814.

Rech GE, Radío S, Guirao-Rico S, *et al.* Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in Drosophila. *Nat Commun*. 2022:13(1). 10.1038/s41467-022-29518-8.

Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD. Parallel geographic variation in *Drosophila melanogaster*. *Genetics*. 2014:197(1):361–373. https://doi.org/10.1534/genetics.114.161463.

Remington CL. Suture-zones of hybrid interaction between recently joined biotas. In: Dobzhansky T, Hecht MK, Steere WC, editors. *Evolutionary biology*. Boston (MA): Springer US; 1968. p. 321–428.

Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T, Petrov DA, Schmidt P. Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science*. 2022:375(6586):eabj7484. https://doi.org/10.1126/science.abj7484.

Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. *Curr Biol*. 2020:30(8):1517–1528.e6. https://doi.org/10.1016/j.cub.2020.03.053.

Sanchez-Refusta F, Santiago E, Rubio J. Seasonal fluctuations of cosmopolitan inversion frequencies in a natural population of *Drosophila melanogaster*. *Genet Sel Evol*. 1990:22(1):47–56. https://doi.org/10.1186/1297-9686-22-1-47.

Schadt CW, Martin AP, Lipson DA, Schmidt SK. Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science*. 2003:301(5638):1359–1361. https://doi.org/10.1126/science.1086940.

Schlötterer C, Tobler R, Kofler R, Nolte V. Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat Rev Genet*. 2014:15(11):749–763. https://doi.org/10.1038/nrg3803.

Schmidt PS, Conde DR. Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in *Drosophila melanogaster*. *Evolution*. 2006:60(8):1602–1611. https://doi.org/10.1111/j.0014-3820.2006.tb00505.x.

Schmidt PS, Zhu C-T, Das J, Batavia M, Yang L, Eanes WF. An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2008:105(42):16207–16211. https://doi.org/10.1073/pnas.0805485105.

Serga SV, Maistrenko OM, Rozhok AI, Mousseau TA, Kozeretska IA. Colonization of a temperate-zone region by the fruit fly

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025

Drosophila simulans (Diptera: Drosophilidae). *Can J Zool*. 2015:93(10):799–804. https://doi.org/10.1139/cjz-2015-0018.
Siddiq MA, Thornton JW. Fitness effects but no temperature-mediated balancing selection at the polymorphic *Adh* gene of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2019:116(43):21634–21640. https://doi.org/10.1073/pnas.1909216116.
Signor SA, New FN, Nuzhdin S. A large panel of *Drosophila simulans* reveals an abundance of common variants. *Genome Biol Evol*. 2018:10(1):189–206. https://doi.org/10.1093/gbe/evx262.
Smit A, Hubley R, Green P. 1996. RepeatMasker Open-3.0. [accessed 2022 Dec 12]. http://www.repeatmasker.org.
Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC, Pool JE. Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Mol Biol Evol*. 2020:37(3):627–638. https://doi.org/10.1093/molbev/msz271.
Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM Jr. *The American soldier: adjustment during army life. (studies in social psychology in world war ii), vol. 1*. Princeton (NJ): Princeton Univ. Press; 1949.
Sturtevant AH. *The North American species of Drosophila*. Washington (DC): Carnegie institution of Washington; 1921.
Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, *et al*. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol*. 2022:32(1):111–123.e5. https://doi.org/10.1016/j.cub.2021.10.052.
Svetec N, Cridland JM, Zhao L, Begun DJ. The adaptive significance of natural genetic variation in the DNA damage response of *Drosophila melanogaster*. *PLoS Genet*. 2016:12(3):e1005869. https://doi.org/10.1371/journal.pgen.1005869.
Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics*. 1983:105(2):437–460. https://doi.org/10.1093/genetics/105.2.437.
Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989:123(3):585–595. https://doi.org/10.1093/genetics/123.3.585.
Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? *Genome Res*. 2006:16(6):702–712. https://doi.org/10.1101/gr.5105206.
Thia JA. 2024. *genomalicious*: serving up a smorgasbord of R functions for performing and teaching population genomic analyses. bioRxiv 667337. https://doi.org/10.1101/667337, 22 April 2024, preprint: not peer reviewed.
Thornton KR, Jensen JD. Controlling the false-positive rate in multilocus genome scans for selection. *Genetics*. 2007:175(2):737–750. https://doi.org/10.1534/genetics.106.064642.
Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol*. 2001:63(2):411–423. https://doi.org/10.1111/1467-9868.00293.
Wang Y, McNeil P, Abdulazeez R, Pascual M, Johnston SE, Keightley PD, Obbard DJ. Variation in mutation, recombination, and transposition rates in *Drosophila melanogaster* and *Drosophila simulans*. *Genome Res*. 2023:33(4):587–598. https://doi.org/10.1101/gr.277383.122.
Xu R, Lou Y, Tidu A, Bulet P, Heinekamp T, Martin F, Brakhage A, Li Z, Liégeois S, Ferrandon D. The toll pathway mediates *Drosophila* resilience to *Aspergillus* mycotoxins through specific Bomanins. *EMBO Rep*. 2023:24(1):e56036. https://doi.org/10.15252/embr.202256036.
Yu Y, Bergland AO. Distinct signals of clinal and seasonal allele frequency change at eQTLs in *Drosophila melanogaster*. *Evolution*. 2022:76(11):2758–2768. https://doi.org/10.1111/evo.14617.
Zhang Z, Zhu S. Drosomycin, an essential component of antifungal defence in *Drosophila*. *Insect Mol Biol*. 2009:18(5):549–556. https://doi.org/10.1111/j.1365-2583.2009.00907.x.
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012:28(24):3326–3328. https://doi.org/10.1093/bioinformatics/bts606.

Downloaded from https://academic.oup.com/mbe/article/42/8/msaf132/8237493 by University of Virginia Law Library user on 19 August 2025