Footprints of worldwide adaptation in structured populations of D. melanogaster through the expanded DEST 2.0 genomic resource

Joaquin C. B. Nunez, Marta Coronado-Zamora, Mathieu Gautier, Martin Kapun, Sonja Steindl, Lino Ometto, Katja M. Hoedjes, Julia Beets, R. Axel W. Wiberg, Giovanni R. Mazzeo, et al.

Footprints of worldwide adaptation in structured populations of *D. 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 M. 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 Serga[5,13,‡], Cristina Amor-Jimenez[16,17], Sònia Casillas[16,17,‡], Alejandro Sanchez-Gracia[18,19,‡], Aleksandra Patenkovic[20,‡], Amanda Glaser-Schmitt[21,‡], Antonio Barbadilla[16,17,‡], Antonio J. Buendia-Ruiz[22], Astra Clelia Bertelli[7,6], 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 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], Florencia Camus[41,‡], Hervé Colinet[42,‡], Jan Hrcek[43,‡], Jesper G. 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,‡], Manolis Ladoukakis[52,‡], Maria Josefa Gomez-Julian[22], Maria 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[59,‡], Pau Carazo[60,‡], Pinar Kohlmeier[26], Priscilla A. Erickson[2,61], Renaud Vitalis[5], Roberto Torres[62,‡], Sara Guirao-Rico[18,19,‡], Sebastian E. Ramos-Onsins[63], Silvana Castillo[64], Tânia F. Paulo[65,‡], Venera Tyukmaeva[66,‡], Zahara Alonso[67], Vladimir Alatortsev[68,‡], Elena Pasyukova[68,‡], Dmitry Mukha[69,‡], Dmitri Petrov[70,71,°‡§], Paul Schmidt[72,°‡§], Thomas Flatt[38,°∧‡§], Alan O. Bergland[2,°∧‡§], Josefa Gonzalez[3,4,°∧‡§]

* = Equal Contribution (Co-First authors)

∧ = Equal Contribution (Co-Senior authors)

° = Corresponding Author

‡ = The European Drosophila Population Genomics Consortium (DrosEU)

§ = The Drosophila Real-Time Evolution Consortium (DrosRTEC)

Corresponding author emails:

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

---

1: Department of Biology, University of Vermont, Burlington, Vermont, USA
2: Department of Biology, University of Virginia, Charlottesville, Virginia, USA
3: Institut Botànic de Barcelona (IBB) CSIC-CMCNB. Catalonia, Spain
4: Institute of Evolutionary Biology, CSIC, UPF. Barcelona, Spain
5: CBGP, Univ Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, France
6: Natural History Museum, Vienna, Austria
7: Department of Biology and Biotechnology, University of Pavia, Italy
8: Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
9: Department of Zoology, Stockholm University, Stockholm, Sweden
10: Department of Biology, Johns Hopkins University, Baltimore, Maryland, USA
11: Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland, 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 Foresting, 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, Spain
17: Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de 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
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, the Netherlands
27: Neuroscience group, Genetics Institute, Universidad Nacional de Colombia, Bogotá, Colombia
28: Justus-von-Liebig-Schule, Baden-Württemberg, 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, 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: Pontificia Universidad Católica del 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, 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, Denmark
45: Biology Department, Lund University, Sweden
46: Instituto de Enseñanza Secundaria Benjamín Jarnés, Zaragoza, Spain
47: Instituto de Enseñanza Secundaria Azuer, Ciudad Real, Spain
48: Entomology Laboratory, Research Unit on Nuclear Techniques, 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, Greece
53: Faculty of Biology, University of Belgrade, Belgrade, Serbia
54: University of St. Andrews, Scotland, UK
55: Institut Jacques Monod, Paris, France
56: Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Lab, Berkeley, California
57: Center Agriculture Food Environment (C3A), University of Trento, Trento, Italy
58: Instituto de Enseñanza Secundaria Carpetania, Toledo, Spain
59: Laboratoire de Biométrie et Biologie Evolutive, CNRS, Université Claude Bernard Lyon 1, Villeurbanne, France
60: Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia
61: Department of Biology, University of Richmond, 138 UR Drive, Richmond, Virginia, USA
62: La Ciència Al Teu Món, Barcelona, Spain
63: Centre for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Barcelona, Spain
64: Instituto de Enseñanza Secundaria Jose de Mora, Granada, Spain
65: Instituto Gulbenkian de Ciência, Oeiras, Portugal
66: Institute of Infection, Veterinary, and Ecological Sciences, University of Liverpool, Liverpool, UK
67: Centro de Educación Infantil y Primaria Ramón y Cajal, Zaragoza, Spain
68: Institute of Molecular Genetics of Russian Academy of Sciences, Moscow, Russia
69: Vavilov Institute of General Genetics of Russian Academy of Sciences, Moscow, Russia
70: Department of Biology, Stanford University, Stanford, California, USA
71: CZ Biohub, San Francisco, California, USA
72: Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

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 (2009-2021), 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. We showed that patterns of secondary contact, originally characterized in North America, are replicated in South America and Australia. We also found 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 spatio-temporal patterns of genetic variation in this classic model system.

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 (e.g., 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, *D. simulans,* between 1.4 and 3.6 million years ago (Obbard et al. 2009; 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 modern fruit fly populations expanded out of Africa after the last glacial maximum ~10,000 ya (Kapopoulou et al. 2020), entering Asia around 3-4 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; Kapun et al. 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 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 (e.g., Ives 1945; Ives 1970; Machado et al. 2016; Kapun et al. 2021; Nunez et al. 2024). In temperate regions, *D. melanogaster* reaches its largest local population size during the peak of the growing season (e.g., late summer and early fall) and drastically decreases upon the onset of winter. These yearly boom-and-bust cycles are responsible for estimates of “local” population size that are orders of magnitude smaller than the “global” 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 2021). 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 clinal and/or seasonal adaptation (Lemeunier and Aulard 1992; Kapun et al. 2016; Kapun and Flatt 2019; Kapun et al. 2023; Nunez et al. 2024). Likewise, several studies have successfully linked clinally and/or seasonally varying polymorphisms in *D. melanogaster* to fitness-relevant phenotypes (Lemeunier and Aulard 1992; Schmidt et al. 2008; Cogni et al. 2014; Paaby et al. 2014; Kapun et al. 2016; Kapun et al. 2016; Durmaz et al. 2019; Kapun and Flatt 2019; Betancourt et al. 2021; Yu and Bergland 2022; Glaser-Schmitt et al. 2023; Kapun et al. 2023; Nunez et al. 2024). Populations of *D. melanogaster* can thus be thought of as powerful “natural laboratories” to study adaptation across spatial and temporal scales, 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 lines (Langley et al. 2012; Mackay et al. 2012; Lack et al. 2015; Lack et al. 2016), while others use sequencing of outbred individuals sequenced as a pool (i.e., Pool-Seq; Schlötterer et al. 2014), and the two data types can be difficult to reconcile. 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 (**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, as well as 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 highly 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 more than 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 we applied demographic inference tools (Jouganous et al. 2017) to infer the history of population subdivision in Europe (Kapun et al. 2020).

Here, we introduce the second release of the DEST resource (DEST 2.0), with substantial 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 new 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* (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 (v2.0) includes 530 high-quality samples as well as an additional 207 pools of varying quality (excluded from the analysis; see **Table S1**). In its totality, the 737 pooled libraries originated from multiple sources including both releases of the DEST dataset (i.e., v1.0 and v2.0), the *Drosophila* Genome Nexus (**DGN**; Lack et al. 2016; including one sample from *D. simulans*), as well as from previous publications (i.e., 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 (abbreviated as "**RD**") of 4X to 300X and from an effective haploid sample size ($\mathbf{n_e}$; the sample size accounting for pool size and pool-seq sampling effects) of 3.7 to 77.2 (**Fig. S1**; see **Text S1**; Kolaczkowski et al. 2011; Feder et al. 2012; Gautier et al. 2013). To ensure the highest possible quality of each sample, we calculated a battery of sequencing statistics including rate of PCR duplication, fraction of missing data, coverage, and number of private single nucleotide polymorphisms (SNPs) across the totality of the dataset (all 737 pools). In addition, we also estimated the pN/pS statistic (i.e., the ratio of the number of genome-wide non-synonymous polymorphisms to the number of genome-wide synonymous polymorphisms, as in Kapun et al. 2021; **Fig. S2**), and assessed non-*D. melanogaster* contamination through competitive mapping and k-mer approaches (Kapun et al. 2021, Gautier 2023; **Fig. S3**). 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 **Fig. S4**; see Materials and Methods: *Estimation of nucleotide diversity*). 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 on Data filtering procedures and recommendations for users see **Text S2**. 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 seven from South America; **Fig. 1A**) and across a time span of 12 years (2009-2021). In total, our 530 high-quality samples represent 164 localities, of these, 112 were sampled only in one year (68%), 18 were sampled across two 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.

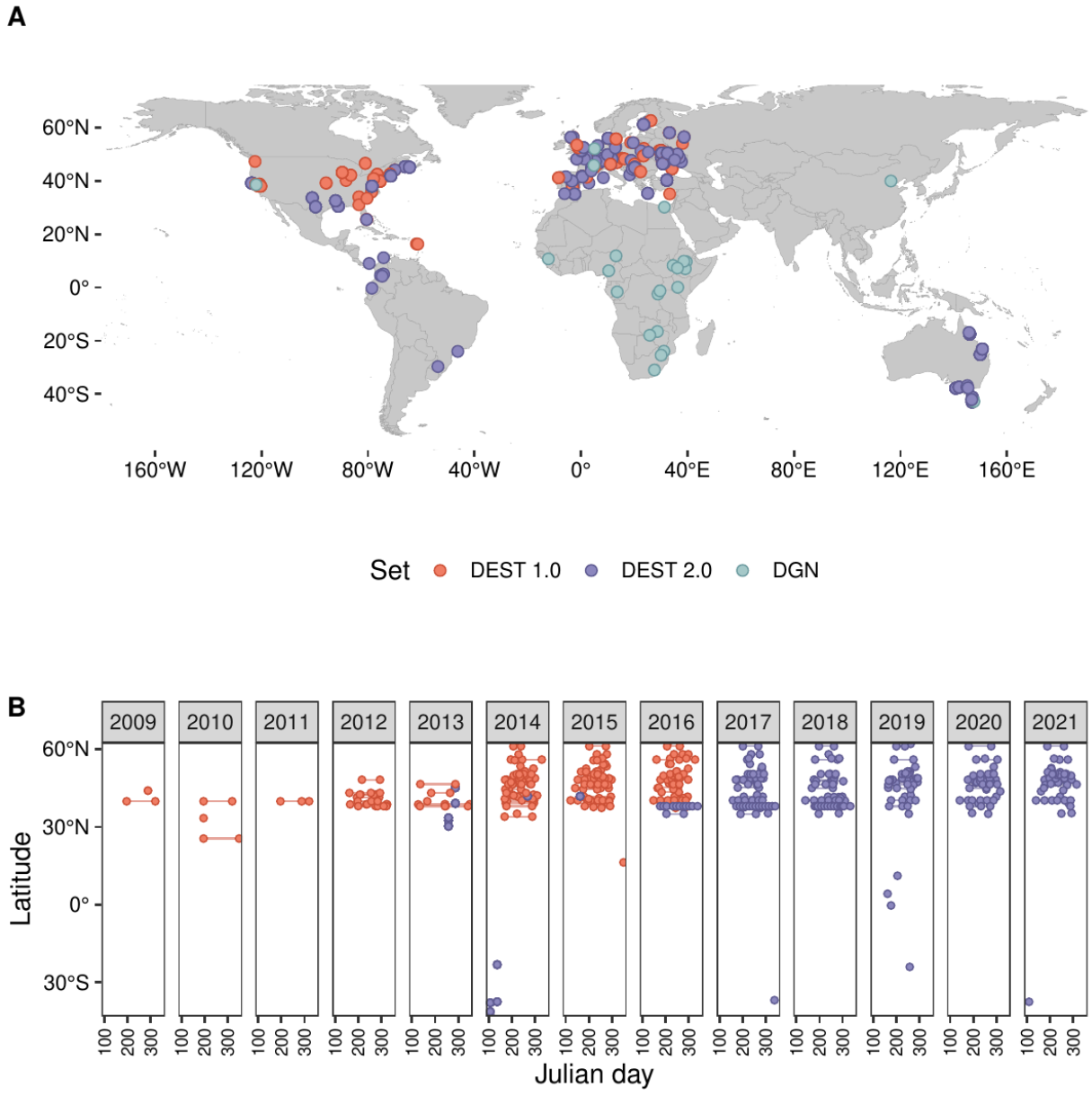


Figure 1. Spatial and temporal scales of DEST. (A) World map showing samples part of DEST 1.0 (Kapun et al. 2020), DEST 2.0 (this study), and the DGN (Lack et al. 2016). **(B)** Sampling density across a decade of sampling contained in the DEST dataset. The colors are consistent with panel A.

278

Table 1: SNP calling information for DEST 2.0 across major autosomes and chromosome X. SNPs inside the inversion are estimated of *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.

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
Non-synonymous	71,534	75,921	72,843	90,905	25,072
Proportion of missing data	0.0511	0.0507	0.0508	0.0493	0.0533

279

280

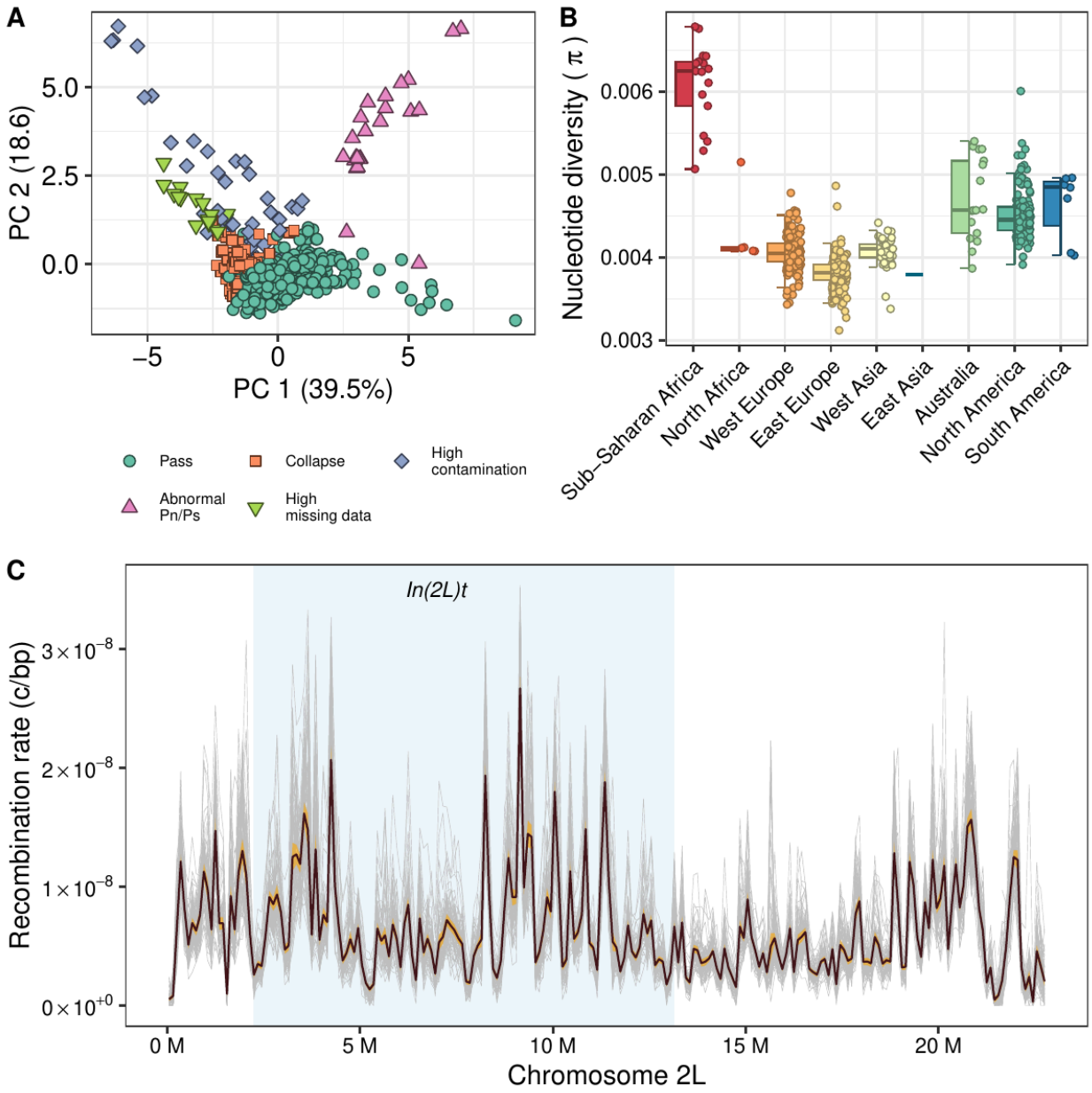


Figure 2. Patterns of filtering, genetic variation, and recombination in DEST 2.0. (A) Visualization of filtering information of samples using PCA. Each dot is a sample's QC metric and the color indicates the filtering decision (legend: Pass: samples that pass filter and are used in downstream analyses; Collapse: biological and/or technical replicates collapsed into a single representative sample; otherwise samples were excluded due to abnormal pN/pS levels of high levels of missing data or contamination). **(B)** Nucleotide diversity (π) calculated across continents (see *Estimation of nucleotide diversity* for details). **(C)** Recombination landscape of chromosome 2L in samples representative of the 75 *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 s.d.) are shown.

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; Tajima 1989) estimated with *npStat* (Ferretti et al. 2013). As previously observed (e.g., Begun and Aquadro 1993; Andolfatto 2001; Mackay et al. 2012; Kapun et al. 2021), we found that sub-Saharan African populations had 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 75 samples representative of the populations analyzed (see Materials and Methods: *Recombination landscape*) using the deep learning method *ReLERNN* (Adrion et al. 2020; see **Fig. 2C; Fig. S5**). Overall, recombination rate is highly heterogeneous among samples and , among chromosomes (two-way ANOVA, $F_{74,296}$ = 20.0, $P < 1.0\text{x}10^{-25}$, and $F_{4,296}$ = 1605.1, $P < 1.0\text{x}10^{-25}$, respectively; Tukey's HSD tests, all pairwise comparisons between chromosomes $P < 1.0\text{x}10^{-7}$, except for 3R *vs*. 2R, where P = 0.073). In most populations there is a statistically significant positive correlation between recombination rate and genetic diversity, consistent with recurrent genetic hitchhiking and background selection (Begun and Aquadro 1993; **Table S2**).

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 five out of the seven inversions analyzed (Wilcoxon test, $P < 0.01$; for inversions *In(2L)t*, *In(3L)P*, *In(3R)Payne*, *In(3R)C* and *In(3R)K*; **Table S3**). Recombination was also lower for those regions spanning the three inversions than for the rest of the chromosome (Wilcoxon test, $P < 0.01$; for inversions *In(2R)NS*, *In(3R)Payne* and *In(3R)K*; **Table S3**).

PCA analyses showed that populations belonging to the same geographic cluster share similar recombination landscapes (**Fig. S6**; see **Table S1** for metadata). The geographic clustering 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 **Fig. S6**).

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 (**Figs. 3A-B;** also see **Fig. S7**). 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 = 2x10^{-16} and r = -0.59; P = 2.2x10^{-16}, respectively) and longitude in Europe (r = -0.46; P = 2.2x10^{-16}; **Fig. 3C-D**). These patterns of population structure were consistent with previously published studies (Kapun et al. 2020; Kapun et al. 2021; Machado et al. 2021). 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. PC3 primarily resolved discrete European clusters and also suggests that North American, South American and Australian samples behave like admixed samples (Ma and Amos 2012).

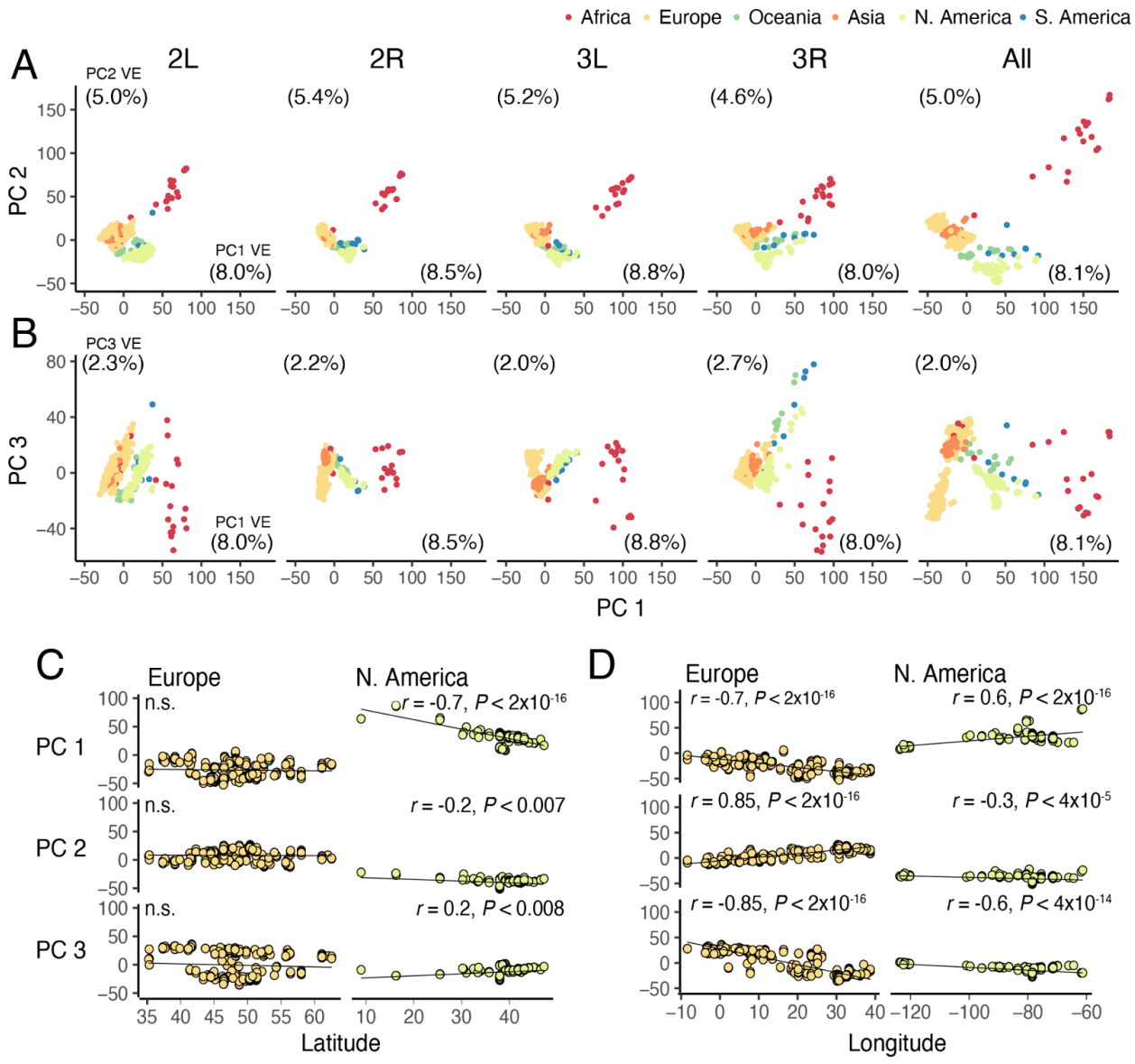


Figure 3. Principal component analysis and projections. (A) PCA projections showing PCs 1 and 2. Analyses were done for each chromosome arm and all arms combined. The proportion of variance explained (VE) is shown at the corners of each axis. (**B)**

PCA projections showing PCs 1 and 3. (**C)** Projections of PCs 1, 2, and 3 relative to latitude for Europe and North American pools. (**D)** Same as C but for longitude. Notice that, in this analysis, Asia refers primarily to samples from Turkey (which is located in Western Asia).

The patterns seen across chromosome-specific PCA were strongly correlated to that of the whole genome for both PCs 1 and 2 ($r_{2L\text{-}All}$ = ~0.97, $r_{2R\text{-}All}$ = ~0.98, $r_{3L\text{-}All}$ = ~0.97, $r_{3L\text{-}All}$ = ~0.96; note that all P are < $1.0\text{x}10^{-15}$). PC3 is peculiar in that the whole-genome results were similar only to those for chromosomes 2R ($r_{2R\text{-}All}$ = 0.95; P = $2.2\text{x}10^{-16}$) and 3L ($r_{3L\text{-}All}$ = -0.95; P-value = $2.2\text{x}10^{-16}$), but not for 2L ($r_{2L\text{-}All}$ = 0.18; P = $1.4\text{x}10^{-5}$) or 3R ($r_{3R\text{-}All}$ = 0.05; P = 0.17). This observation suggests that the signal captured by PC3 at 2L and 3R were strongly influenced by the frequencies of *In(2L)t* and *In(3R)Payne*, two large adaptive cosmopolitan inversion polymorphisms (e.g., 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; **Fig. S8**). Many inversions showed significant clinal patterns along latitude or longitude that were consistent across different continents (see **Table S4** for statistical details). Our results are in line with previous observations, in particular for *In(3R)Payne* (Lemeunier and Aulard 1992; Kapun et al. 2016; Kapun and Flatt 2019; Kapun et al. 2020; Kapun et al. 2023), which showed significant latitudinal clines in North America, Europe and along the Australian east coast. Notably, these patterns did not differ across sampling years in Europe and Australia, indicating temporal stability of the clines on these continents. 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 latent 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 dominated by 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 latent 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 dominated by a single cluster, were divided into three clusters: the Caribbean and most of South America (henceforth "Latin America"), a southeast U.S. coastal group (henceforth "Southeast"), and all other samples from the Americas (henceforth "mainland"; 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 4B**). We used model-based demographic inference with *moments* (Jouganous et al. 2017) to test the statistical support of these additional populations suggested by the k = 8 analysis while simultaneously estimating demographic parameters. Specifically, we fit simple, neutral population history models that we call "one-population," "split," "admixture," and "two-splits" (see **Fig. S9**; see description in the Materials and Methods: *Demographic inference with moments*) to subsets of the DEST 2.0 variant data consisting of the Southeast and mainland clusters, all samples from the Americas, and European samples (**Table S5**).

First, we fit the "one-population" and two-population "split" models to the Southeast and mainland clusters in North America to conclude that "one-population" better describes the region (Wilcoxon signed-rank test on model likelihoods, $P = 7.02 \times 10^{-7}$; **Fig. S10A**). 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. Thus, it is likely that the primary cause of the Southeast cluster in k = 8 analysis is the disproportionately dense sampling around Charlottesville, VA.

We then fit the "one-population" and "split" models to a population consisting of the Southeast and mainland clusters and the Latin America cluster, concluding again that "one-population" outperforms "split" (Wilcoxon signed-rank test on model likelihoods, $P = 6.90 \times 10^{-9}$; **Fig. S10B**). This result is complemented by the low F_{ST} = 0.062. This secondary result supports prior treatment of all flies of the Americas as a single cluster. This result does not contradict our findings of clines within the Americas, because the *demes*-type models employed rely on discretizing geography, and are thus largely blind to gradual changes with location.

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 are respectively 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). As in the Americas, we found support for the parsimonious two-population models that does not include the overlap zone as a discrete population (corrected Dunn's tests on model likelihoods, P = 3.3 x 10^{-7}; **Fig. S10C**). 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.

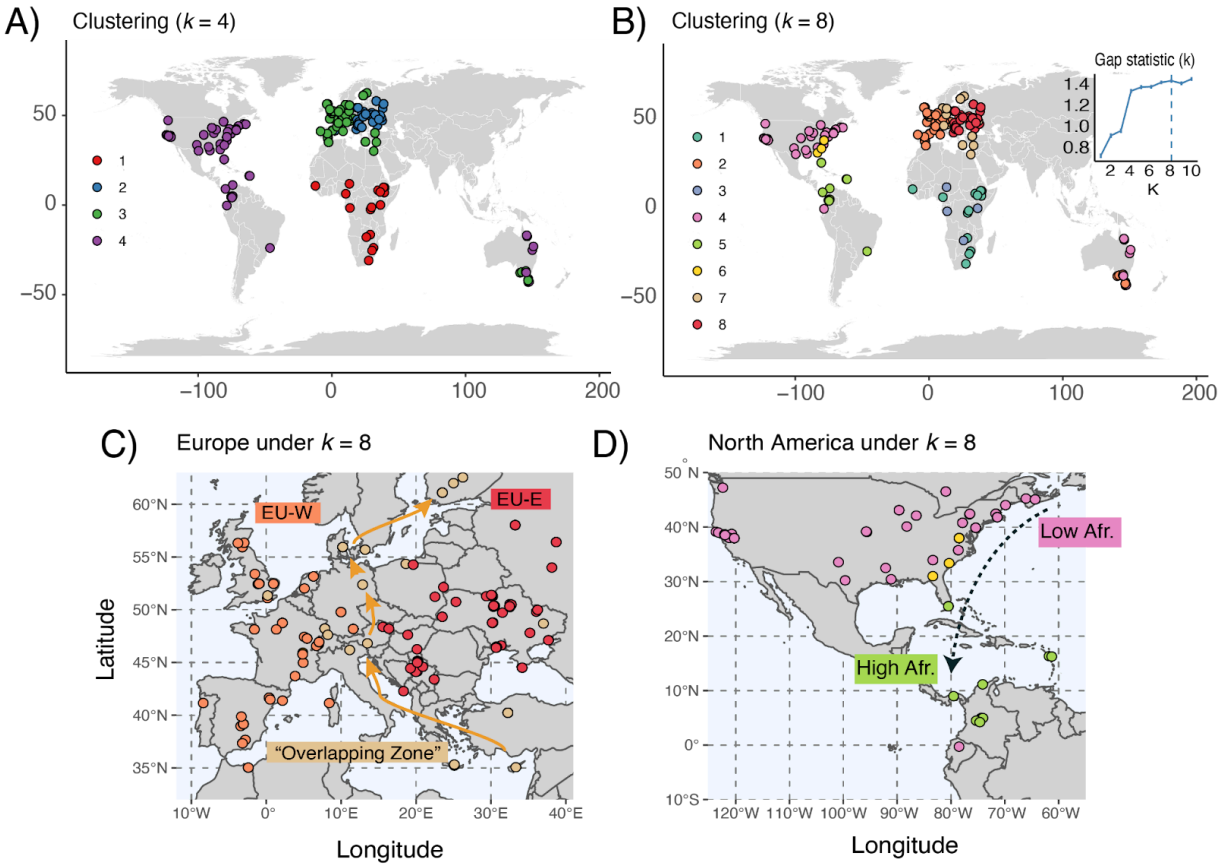


Figure 4: Spatial population structure and admixture in worldwide *Drosophila*. (A) Clustering map, based on PCA projections 1-3 built using k = 4 (as reported in DEST 1.0). **(B)** Same as A but with k=8 (the optimal number of clusters as defined by a heuristic Gap statistic search). **(C)** Zoom view 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).

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 (Bergland et al. 2016). To accomplish this, we first modeled the proportion of African and European admixture in the Americas and Australian pools as a linear combination of two "ancestral populations" from Europe and Africa (see **Dataset S1**). Our estimates of African admixture were consistent with previously published results (i.e., a positive, albeit non-significant, correlation between African admixture and latitude in Australia, $\beta_{\text{African anc.}}$ = 0.003, P = 0.162, see **Fig. 5A;** and a significant negative pattern in North America, $\beta_{\text{African anc.}}$ = -0.005, $P = 2.5\text{x}10^{-22}$, see **Fig. 5B**; Bergland et al. 2016). We calculated these estimates in the newly collected samples from South America and observed a trend of increasing African ancestry near the equator ($\beta_{\text{African anc.}}$ is 0.002, P = 0.139, **Fig. 5C**). We also estimated the relationship between levels of admixture and longitude in North America. Here, we identified a significant association between longitude and ancestry (LM; $\beta_{\text{African anc.}}$ = 0.0014, $P = 6.76\text{x}10^{-16}$). This was evidenced when levels of African ancestry were projected onto a map of North America (see **Fig. 5D**) revealing that westward samples (i.e., from the American midwest or California) have lower levels of African ancestry when compared to 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 **Fig. S11**).

We further explored patterns of admixture using a two-pronged approach. First, we calculated the f_3 statistic (Patterson et al. 2012; Gautier et al. 2022) using samples from North and South America as the targets of admixture and Europe and Africa as the "ancestral" populations. For African populations, we included samples from Cameroon, Egypt, Ethiopia, Morocco, Rwanda, South Africa, and Zambia. In total, we conducted 1,478,000 three-population comparisons (**Dataset S2**). Overall, all American populations displayed significant f_3 tests (i.e., had a Z-score < -1.65), which confirms pervasive admixture (**Table S6;** also **Fig. S11**); these results do not appear to be driven by differences in read depth ($r_{\text{signif } f3\text{-RD}}$ = -0.6, P = 0.10) or by the number of flies included in the pool or synthetic pool ($r_{\text{signif } f3\text{-Nflies}}$ = 0.2, P = 0.40).

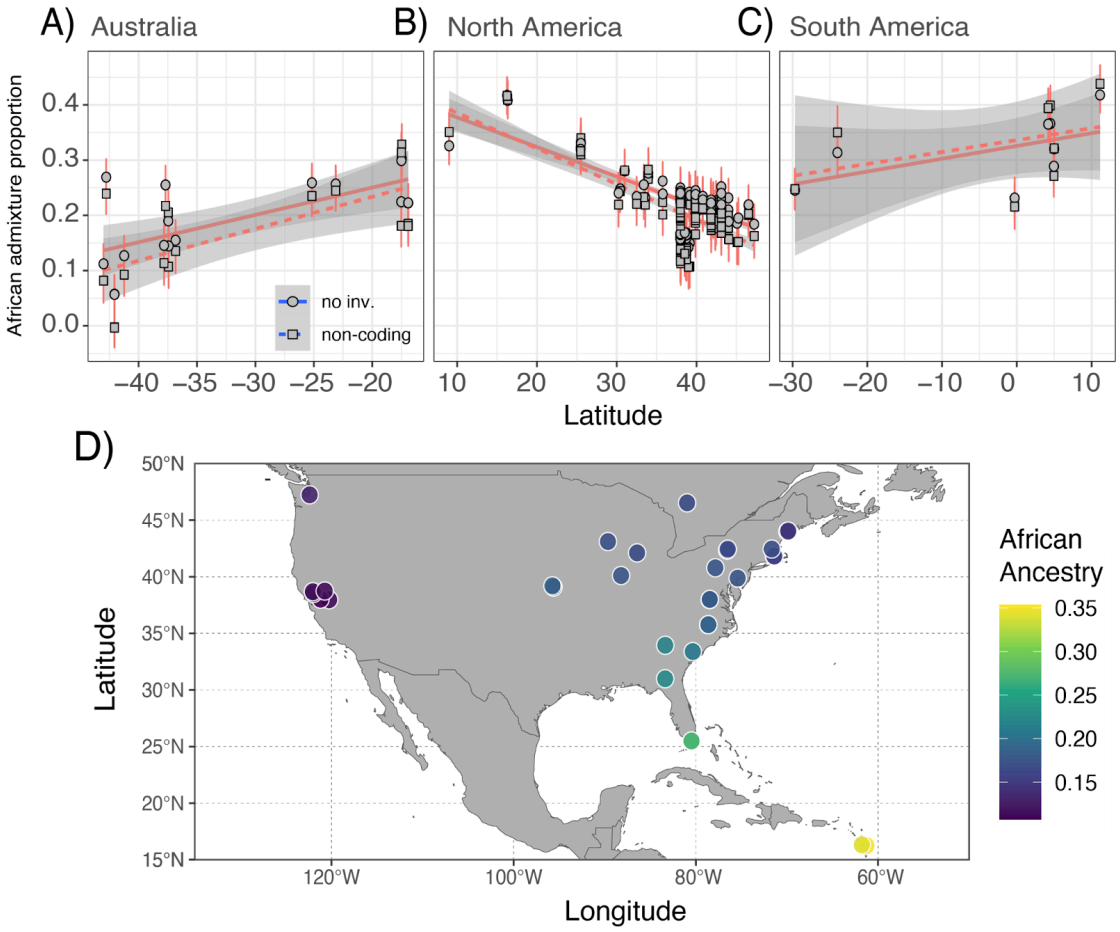


Figure 5: Patterns of admixture across the Americas and Australia. (A) Coefficients of linear admixture for Australia (excluding SNPs in inversions). **(B)** Same as A but for North America. **(C)** Same as A but for South America. **(D)** Map projection of levels of African ancestry in North American samples (note that the collapsed samples of Fournier-Level et al. 2019 were removed).

Lastly, we conducted a survey of genetic differentiation across the demographic clusters (see Materials and Methods: *Estimation of nucleotide diversity*). The overall differentiation was F_{ST} = 0.050 ± 0.001 for autosomes and nearly twice as high for the X chromosome (0.091 ± 0.004; **Fig. 6A, orange**). These results were robust to the removal of heterochromatin regions and low frequency alleles (MAF < 0.05; **Fig. S12**). To quantify the level of differentiation between population groups defined by their continental cluster (**Fig. 4A**), we further relied on a hierarchical F_{ST} model (Nei 1973), which consists of decomposing the total differentiation into an across-group (F_{GT}) and a within-group (i.e., a composite label of continent and cluster; F_{SG}) contributions, using unbiased estimators developed for Pool-Seq data (Gautier et al., *in prep.*). 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 (see above). 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} (i.e., for each pair of regions the underlying populations were analyzed under a hierarchical F_{ST} model with two groups), as shown on **Fig. 6B** (see results for k = 8 in **Fig. S13**). In general, all clusters involving Africa were consistently more differentiated than non-African groups. The highest level of differentiation was 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, reflecting a history of recent expansion and admixture between these two demes. 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, the latter being separated in clustering at k = 8 (**Figs. 4B-D**).

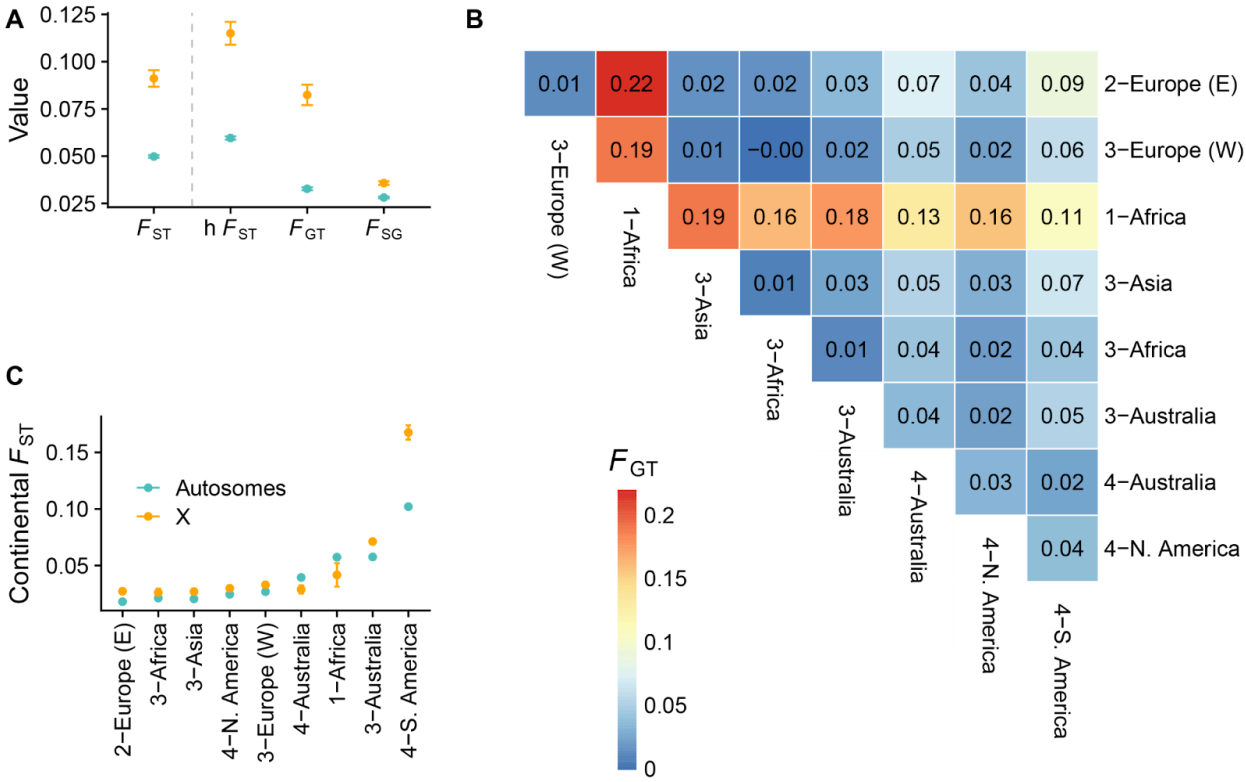


Figure 6: Genetic differentiation. (A) Values of the F_{ST} estimates over all DEST samples and their 95% CI (corresponding to ±1.96 s.e. estimated using block−jackknife with blocks of 50,000 consecutive SNPs). Note that the hF_{ST}, F_{GT} and F_{SG} statistics were estimated using the hierarchical F_{ST} model, over all DEST samples grouped according to the k = 4 clustering analysis and their 95% CI. Colors indicate autosomes (blue) and X chromosomes (orange). **(B)** Pairwise comparisons between cluster-continents (under k

491 = 4) results in a heatmap. In this plot, "1-Africa" refers to Sub-Saharan African populations, "3-Africa" refers to North Africa. The
492 clusters "Australia-3" and "Australia-4" represent samples with low and high levels of African admixture, respectively. **(C)** F_{ST}
493 estimates within clusters from the k = 4 analysis.
494
495 **Updated geographically informative markers improve predictive resolution of samples**
496 Our previous release of DEST generated a panel of geographically informative markers (GIMs).
497 The second release of our data gives us the unique opportunity to test the accuracy of our
498 previously published markers. To this end, we applied our previously DEST 1.0 GIMs to our new
499 data and we assessed the distance (d_{hav}; as great circle distance, see *Materials and Methods*)
500 between the predicted locality and the "real" locality as recorded in the metadata. Overall, both
501 DEST 1.0 models trained at the level of "city" and "region" (i.e., resolution at the level of state or
502 province), perform similarly well on the new data (r = 0.995, P = 2.2x10^{-16}; **Fig. 7A**). Next, we
503 aggregated the d_{hav} estimates at the level of continents (here we report only the results of the
504 region model). We did this to assess whether the quality of our predictions vary as a function of
505 continent. Overall, the best performance was observed in European samples (median resolution
506 of ~409 km to real location; **Fig. 7B**), followed by the North American samples, with a resolution
507 of 794 km. Unsurprisingly, the worst predictions from the DEST 1.0 markers occurred when
508 deployed on samples from South America and Australia, two locations that were not included in
509 the first release (**Fig. 7B**).
510 While our published markers performed well on samples from regions present in DEST
511 1.0, the addition of new regions to DEST required the generation of new GIMs. As such, we
512 trained a new demographic model (DEST-GIM 2.0) including the new samples reported in this
513 paper. Our new model was trained using the same workflow as DEST-GIM 1.0 (i.e., by retaining
514 40 PCs). Yet, the models differ in that DEST-GIM 2.0 was created by exclusively using
515 non-coding SNPs as well as loci outside genomic regions spanning major cosmopolitan
516 inversions. This new panel of GIMs is composed of 29,952 SNPs across all autosomes.
517 Performance assessment of the new model by the d_{hav} analysis shows that DEST-GIM 2.0
518 performs similarly to the 1.0 version for existing locales (e.g., Europe or North America; **Fig.**
519 **7B**), yet they provide improved prediction accuracy for new regions (**Fig. 7B** and **7C**).
520

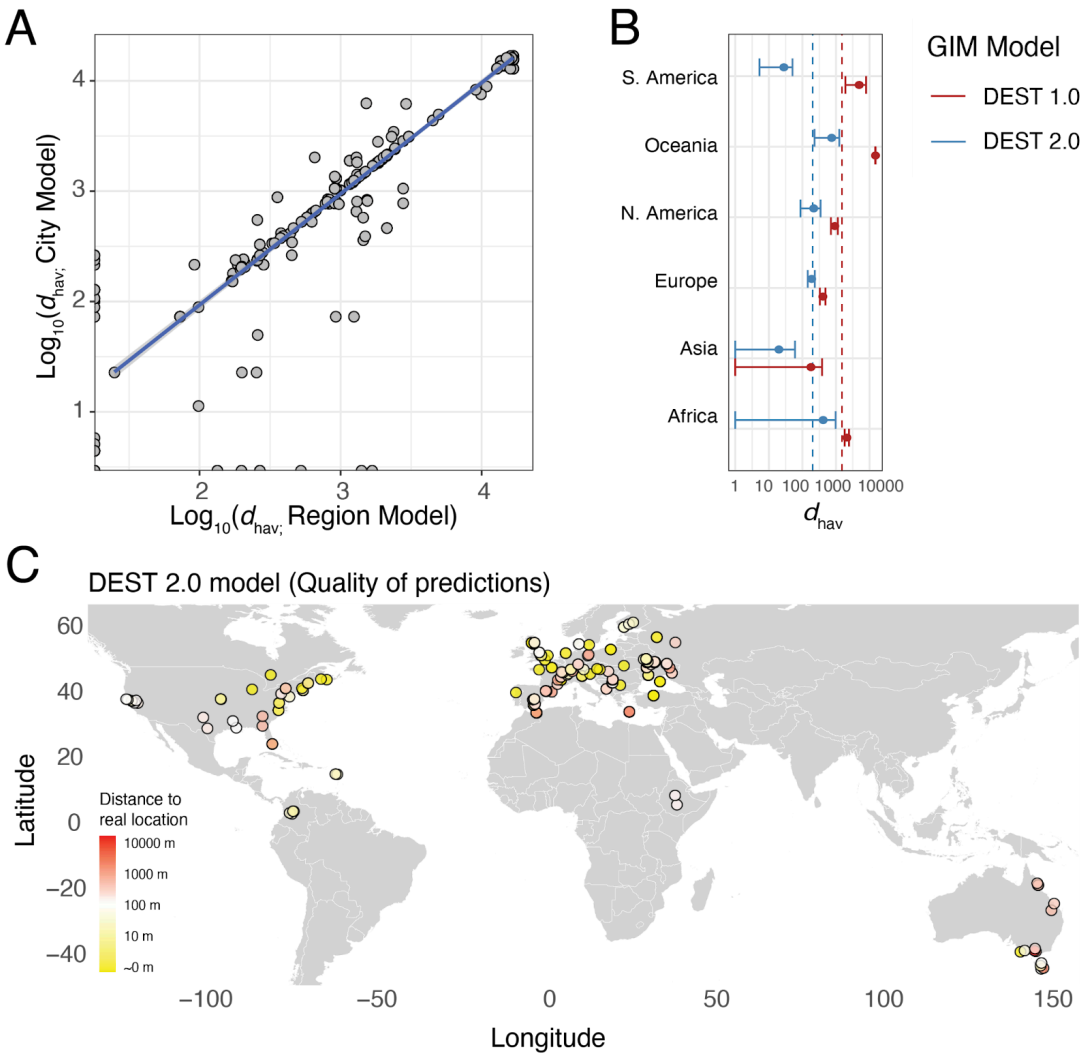


Figure 7: Geographically informative markers. (A) Bi-plot of d_{hav} from the 1.0 GIMs. City model (y-axis) and Region model (x-axis). **(B)** Mean and 95% confidence intervals (CIs) of d_{hav} for the 1.0 GIM and 2.0 GIM model (to improve readability the x-axis has been $\log_{10}$ transformed and CIs < 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 = 0-10 m), white are "adequate" predictions (10-100 m), and red are poor predictions (1000-10000 m).

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 post-overwintering (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 (1-2 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 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 one 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 hypothesis 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 **8E**). Samples below 50.3°N tend to have lower values of year-to-year F_{ST} as compared to 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 = 2.2×10^{-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-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**).

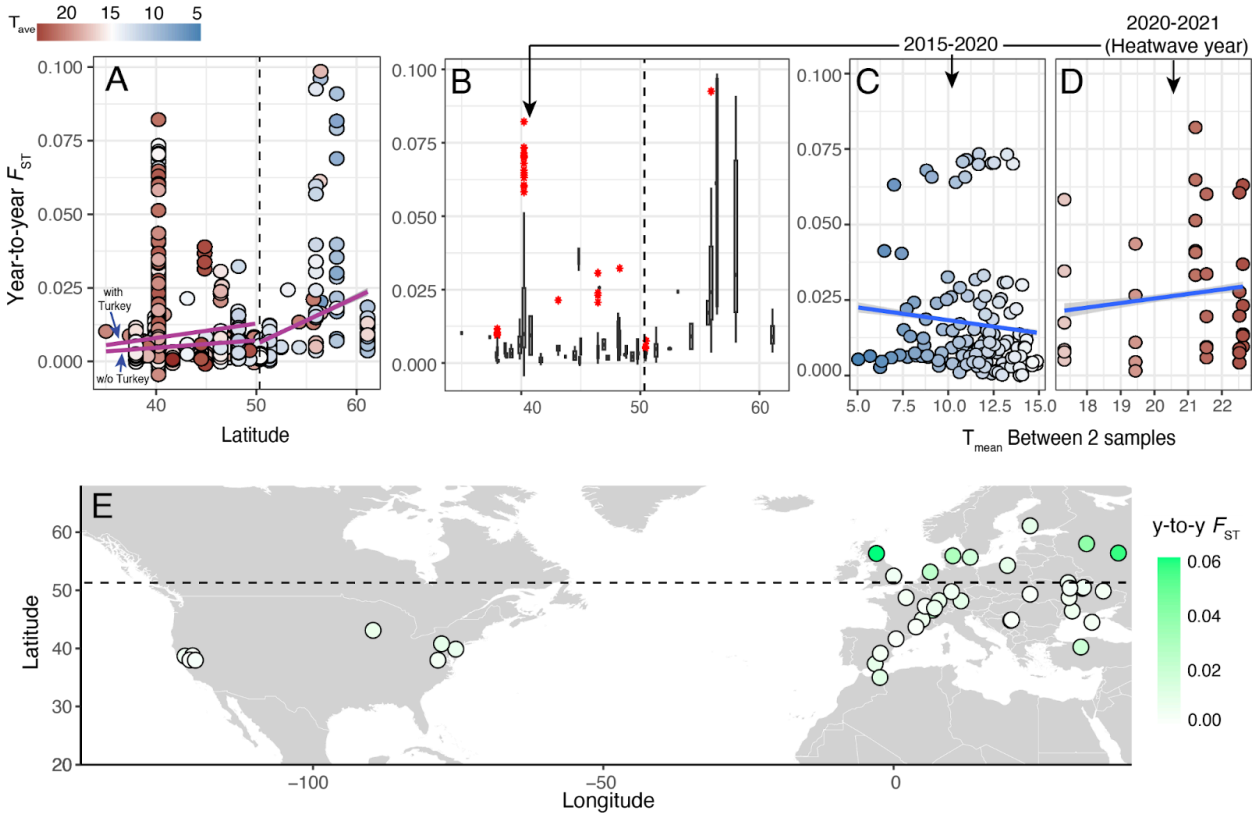
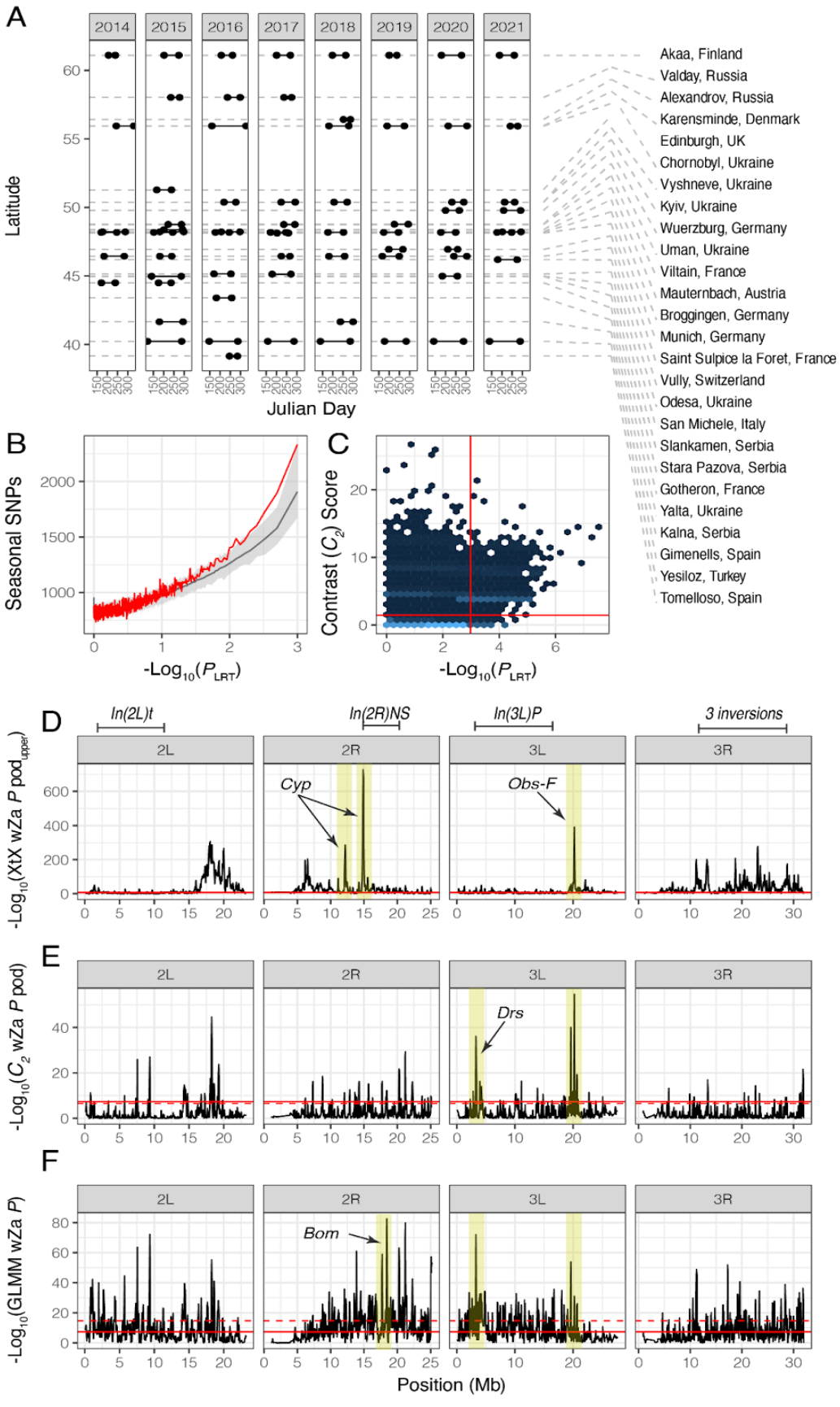


Figure 8: Temporal genetic differentiation due to overwintering. (A) F_{ST} values across DEST 2.0 samples as a function of latitude. Broken-stick regression and breakpoint is shown, for samples below latitude 50.3 the regression is shown with and without Turkey. The color indicates the mean temperature 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 for samples between 2015 and 2020 (logit transformed; correlation between F_{ST} and mean temperature; r =0.135; P = 4.60x10^{-7}). **(D)** Same as B 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 = 7.74x10^{-13}). **(E)** Mean year-to-year F_{ST} overlaid over a world map of northern seasonal habitats.

Footprints of spatial adaptation 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 Results: *Population admixture and…*) 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 ion and pesticides (**Table S7**). One of the strongest signals of population differentiation was observed for the region surrounding the gene *Cyp6g1,* a cytochrome P450 (Cyp) gene (**Fig. S14**; a result also observed in DEST 1.0), a well-known gene involved in resistance to DDT and neonicotinoid

586 insecticides (Le Goff and Hilliou 2017). This signal was particularly high when comparing North
587 America and European samples. Elevated F_{GT} was also observed when comparing South
588 American and North American samples, but not when comparing South American and
589 European samples (**Fig. S14**). These signatures of differentiation suggest different adaptations
590 likely driven by distinct environmental pressures and insecticide exposure levels in each
591 continent. To formally detect footprints of adaptive differentiation in our dataset we applied the
592 "*Bayesian Population Association Analysis*" framework, *BayPass* (Gautier 2015; Olazcuaga et
593 al. 2022) to DEST samples from European localities (irrespective of sampling year or season;
594 138 samples in total; **Fig. 9A**) and relied on the estimated *XtX** statistic to identify overly
595 differentiated SNPs. The analysis identified two regions in chromosome 2R as candidates of
596 local adaptation (12,188,558-12,126,181 and 14,826,182-14,976,108; **Fig. 9D**). Both these
597 regions harbor several *Cyp* genes. For example, the window at ~12 Mb contains *Cyp6g2*, and
598 *Cyp6t3*, whereas the window at ~14 Mb contains *Cyp6a22, Cyp6a19, Cyp6a9, Cyp6a20,*
599 *Cyp6a21, Cyp6a8, and Cyp317a1*. These genes are associated with hormonal metabolism as
600 well as responses to insecticides (Danielson et al. 1995; Le Goff and Hilliou 2017). We
601 performed gene ontology enrichment analysis of genes within all *XtX** outlier regions and found
602 an enrichment of terms such as "oxidation-reduction process", "cellular response to radiation",
603 and "amide biosynthetic process", reflecting results from F_{GT} outlier regions above (**Table S8**).
604



605

Figure 9: Local and seasonal adaptation in *Drosophila*. (A) Schematic of sampling for the seasonal analysis. In total, we used 138 samples collected in 26 European localities across an 8 year period. We selected localities where there were more than one sample per year and designated the first sample as "spring" and last sample as "fall". There is no overlap between the samples used here and the samples used in seasonal analysis in Machado et al. (2020), Bergland et al. (2014), and Nunez et al. (2023). **(B)** GLMM seasonal adaptation scan. The plot shows the $\log_{10}$ transformed wZa *P*-value of the LRT of base and seasonal models. For A, B, and C, regions of interest are highlighted in yellow. Inversions are demarcated along the top of the figure. **(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 ~10M 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. **(E)** Bayesian seasonal adaptation scan. The plot shows the $\log_{10}$ transformed wZa *P*-value of the contrast (C_2) adaptation *BayPass* analysis. **(F)** Results of the GLMM analysis. The permutations are shown in gray (95% confidence intervals) and the real data in red. There are more SNPs with low seasonal p-values than expected by permutations.

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 that quantifies the degree of association of allele frequency with season. We identified significant C_2 values using a simulation approach that is part of the *BayPass* workflow (see Materials and Methods: S*cans for adaptive differentiation*; **Dataset S3**). We observe that several regions across the *Drosophila* genome are enriched for signals of parallel seasonal evolution (**Figs. 9D E, F**). A notable example appears in chromosome 3L (3,222,669-3,422,464), 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, Drsl6*). In view of previous observations of seasonal allele frequency oscillations in several immune genes, this result suggests 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 (**Table S9**). We found an enrichment of, among other terms, genes associated with "alcohol dehydrogenase (NAD) activity", including the gene *Adh* itself (**Table S10**).

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 (**Fig. S15**). Indeed, when compared to Pennsylvania data from Bergland et al. (2014), we observed a significant deficiency of these targets at both a global level (P = 0.024; **Fig. S15A**) 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**, red 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 (18,376,129-18,475,992). 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 20,172,964-20,271,926 bp, notable for harboring adjacent signal peaks across analyses of seasonal and local adaptation (see **Figs. 9D, 9E, 9F**; 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. 2020).

Discussion

A unified resource for wild *Drosophila* genomics

D. melanogaster is a cosmopolitan species with resident populations across all human-inhabited continents that evolves adaptively in response to spatially-varying and temporally-fluctuating selection in semi-natural settings and 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 expands data on the original release by adding twice as many new samples as the original release.

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 sequencing approaches to be incorporated into its workflow, hence allowing for older pooled data sets 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 75 representative populations. In comparison to the findings of previous studies (Comeron et al. 2012; Adrion et al. 2020) our own estimates show a reduction of approximately threefold. 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 (approximately two- to three-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 75 populations (Spearman's rho = 0.835, S = 11624, $P < 1.0\text{x}10^{-25}$; $R^2 = 0.672$). It is thus possible that our estimations can be approximated as a population-scaled effective recombination rate (ρ) rather than the actual crossing-over rate (r; where $\rho = 4N_e r$). 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* (e.g., 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; Kapun et al. 2021; Machado et al. 2021) characterized fine-grained levels of population structure within Europe, and dated their divergence at around ~1,000 ya. In this paper, we expanded the repertoire of samples available for demographic inference and phylogeographic analysis.

In the Americas and Australia, our data recapitulate published patterns of African admixture in North American fly populations (Kao et al. 2015; Bergland et al. 2016; Corbett-Detig and Nielsen 2017). Notably, in South America and Australia, while not significant, our results show a reversed trend with latitude, relative to North America (**Fig. 5A-C**). These support the general hypothesis of higher African admixture in equatorial populations relative to poleward ones, consistent with two separate introductions of *D. melanogaster* to the Americas. It is likely that the African ancestors entered the Americas through the Caribbean. In this region, the earliest record of *D. melanogaster* occurred in Cuba in 1862 (Sturtevant 1921), and it was first documented in Florida in 1894 (Keller 2007). While it is always important to consider that species distributions data may be incomplete, the entomological surveys conducted in the USA during the 1880s are extensive and they do not mention earlier records of the species under any of its old taxonomic names (i.e., *D. ampelophila or D. uvarum*; see Keller 2007). The origin and timing of European immigration is more complex. Notably, European entomological surveys only describe the presence of *D. melanogaster* as a “common” species in Central Europe (Sturtevant 1921), with reported sightings in German cities like Kiel or in Austrian towns in the 1830s (Keller 2007). Consistent with this chronology, the first recorded samples in North America come from New York in 1875 (Lintner 1882; Keller 2007). Thus, while African flies may have been in the Americas since the 1860s, it is possible that the African-European admixture cline in USA’s eastern seaboard originated later, during the late 1880s.

In Europe, the overlap zone we observed inside the continent (in the k = 8 analysis) 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. We reject this model and suggest that the overlap zone is a subpopulation of EU-W. These results are puzzling, and echo findings from our previous release (Kapun et al. 2021), whereby the levels of gene flow in this area appear to be asymmetric in favor of EU-W (e.g., as reported by Kapun et al. 2021, EU-W→EU-E as 0.209 flies/gen vs. EU-E→EU-W as 0.178 flies/gen). These findings are 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). As it stands, these patterns may indicate the action of a non-neutral force confounded with the complex demographic history of *D. melanogaster* in Europe, to be explored in future work.

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 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. 2022) 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 vs. both European and South American populations (**Fig. S14**). 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. 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. 2016; 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 worldwide temperate *Drosophila*. One challenge associated with testing this hypothesis is determining the appropriate covariate (e.g., temperature, humidity, rainfall) and the timeframe of

779 selection (e.g., 0-15, 0-30 days prior to collection) to use in the model. For example, Nunez et
780 al. (2024) showed that, in Virginia, the best seasonal model used the temperature 0-15 days
781 prior to collection as a covariate. Yet, in Europe, Humidity 0-30 and 0-60 prior to collection days
782 were the best models for EU-E and EU-W respectively. Therefore, we used a contrast
783 framework using the seasonal labels (i.e., “spring” and “fall”) as comparison factors. This
784 approach had been successfully used in the past by Bergland et al. (2014) and Machado et al.
785 (2021) and allowed us to surmount the challenge of covariate selection.

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

797 Our gene-ontology enrichment analysis for targets of seasonality highlighted “alcohol
798 dehydrogenase activity” —including the gene *Adh* itself— as being enriched among outlier
799 regions. This is significant because patterns of genetic variation in *Adh* have long been
800 recognized as classical examples of ecological adaptation (Kreitman 1983; Berry and Kreitman
801 1993). However, recent discussions have emphasized that the specific agents of selection
802 acting on this gene remain unclear, with some suggesting temperature-driven balancing
803 selection (Siddiq and Thornton 2019). We also assessed whether the seasonal SNPs observed
804 in our C_2 analysis from Europe are enriched in seasonal datasets generated mostly from North
805 American populations (Bergland et al. 2014; Machado et al. 2021). Our results showed no
806 enrichment (or under-enrichment; see **Fig. S22**) between the datasets compared. In other
807 words, these results suggest that the genetic basis of seasonality is different between
808 continents. This finding is consistent with previous studies positing that population ancestry is a
809 more important predictor of adaptive genetic architecture than the existence of paralleled
810 selection regimes (Otte et al. 2021).

811 Overall, our seasonal analyses reveal three major takeaways. First, they reveal that
812 seasonal adaptive tracking is a detectable phenomenon across the temperate range of *D.*

melanogaster. Yet, they also suggest that adaptive tracking may be driven by both natural and anthropogenic stressors, and that the specific loci that drive adaptation may be strongly shaped by genetic ancestry. 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." And third, our findings showcase an inherent strength of the *BayPass* model to successfully disentangle the dynamics of spatial and temporal adaptation in wild populations. 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 (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 diapause and overwinter until the next growing season. These seasonal 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.

One important question related to these boom-and-bust dynamics is whether populations that experience different severities of winter (harsher vs. milder) show elevated levels of year-to-year differentiation. 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 **8E**). 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. One notable exception to the pattern of year-to-year 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 at 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 extreme heat waves may have affected flies both directly, through physiological thermal challenges, and also indirectly by affecting their food sources.

Overall, our findings provide two 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.

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 multi-species 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 (2.0), we have updated the Docker container to enable mapping of reads sequenced in both paired-end (PE) and single-end (SE) configuration. 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). The SNP calling step as well as genome annotation with SNPEff (v5.2; Cingolani et al. 2012) were automated using SnakeMake (Mölder et al. 2021). We provide ready to use outputs of the DEST pipeline both in variant call format (VCF) format as well as in genomic data structure (GDS) format (Zheng et al. 2012). The entire DEST pipeline can be found on GitHub at: https://github.com/DEST-bio/DESTv2.

Previously published datasets added to DEST 2.0

We incorporated data from previously published studies (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), two samples from Hoffmann et al. (2002), 17 samples from Lange et al. (2022), eight samples from Reinhardt et al. (2014), and one sample from Svetec et al. (2016). Comprehensive metadata for these samples is included in **Table S1.** 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; orange squares in **Fig. 1C**), each with read depths of ~60X.

Filtering parameters

We filtered SNPs and samples using metrics and tools described in our first release (Kapun et al. 2021). In brief, we 1) calculated the levels of contamination by congenerics, 2) levels of read duplication in the sequencing run, 3) proportion of SNPs with missing allele frequency data, 4) ratio of synonymous to non-synonymous polymorphism (p_N/p_S), 5) nominal coverage, and 6) the effective coverage. Levels of contamination by congenerics refers to the amount of non-*D. melanogaster* flies accidentally sequenced in pools.

We assessed contamination using a two-pronged approach. First, we assessed levels of competitive mapping of reads to the genomes of *D. melanogaster* (RefSeq: GCF_000001215.4) and *D. simulans* (RefSeq: GCF_016746395.2). *D. 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 are discussed in the methods of the first release (Kapun 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. This approach is described in (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 (**Fig. S3A**), with the competitive mapping approach slightly over-estimating contamination (by 2.3% max) and the *k*-mer approach slightly under-estimating 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 custom script (see GitHub). The nominal, genome wide, read depth (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 read depth 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 position 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 below 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

$$n_e = \frac{NC}{N+C-1} \quad \text{(eq. 1)}$$

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 **Text S1** for further details on the derivation of eq. 1 and for a more general formula applicable to collapsed Pool-Seq sample).

Recombination landscape

We inferred the genome-wide recombination landscape for 75 of our samples using ReLERNN v1.0.0 (Adrion et al. 2020). The samples were selected to cover the entire spatial distribution of the DEST 2.0. sampling and based on the coverage sequencing depth (mean = 68.3, SD = 35.8, min. = 32, max. = 234), which was chosen to be as high as possible to maximize the reliability of the allele frequency used by ReLERNN to estimate recombination (**Table S1**). We used BCFtools (Danecek et al. 2021) to extract allele frequency of all biallelic SNPs with a frequency > 0.01 and read depth > 10. The resulting data was used to run ReLERNN. The parameters used in ReLERNN *simulate* module were as follow: 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 train 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 non-overlapping windows on which per-base recombination rates were predicted. To allow comparisons between samples, we estimated the average per-base recombination rates in larger 200 kb non-overlapping 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 significantly correlated with our estimates for most of the populations (**Table S11**). Recombination rates are available in the genome browser.

Estimation of nucleotide diversity

We conducted population genetic analyses using *npStat* (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 come from DEST 1.0 data (Kapun et al. 2021). Standard nucleotide diversity statistics were first directly estimated from each *bam* file, for non-overlapping windows (10 kb, 50 kb 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 less than 250 reads with a min quality > 20 were considered in the computations (*-mincov 2 -maxcov 250 -minqual 20* options) and windows with less than 9,000 remaining positions were discarded. We further calculated window-specific average estimates for each sample, using window sizes of 10k, 50k and 100k (i.e., window size that are displayed in the genome browser) using a custom Python script (BED2Window.py).

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 (PCA)

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 less than 1% missing data and a mean allele frequency greater than 1% (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* (v 2.3.0, Gautier et al., *in prep.*; **Fig. S7**).

Demographic inference with *moments*

We fit demographic models to subsets of the DEST 2.0 variant data with the Python package *moments* (Jouganous et al. 2017). We adapted *moments* code to construct site frequency spectra (SFSs) from autosomal SNPs from the Pool-Seq VCF file, subset to include only the pool with greatest effective sample size (n_e) from each locality in order to avoid geographic sampling bias. For simplicity, we normalized population-specific sample sizes to the average n_e of respective subsets of pools in consideration. For different subsets of the data, we constructed *demes*-type models (Gower et al. 2022) dubbed "one-population," "split," "two-splits," and "admixture" (see **Fig. S9**) in order to infer demographic parameters of global *Drosophila* populations while simultaneously performing likelihood-based model selection. A significant limitation of SFS-based demographic inference (e.g. Gutenkunst et al. 2009; Kamm et al. 2020) 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 with 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 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 re-calculate 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 "Southeast" population in order to compare two- and one-population models of the "mainland" region and then by collapsing the "Latin America"

population in order to compare two- and one-population models of the "Americas" region. Simulated validation of CLL as a powerful statistic for selection between models of different dimensions can be found at **Text S3**.

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 non-randomly 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. We then calculated 95% confidence intervals as being between the second-least and second-greatest values for each estimate among each set of 40 replicates. The hypothesis tests that we reported as being performed "on model likelihoods" in the Results section are comparisons of sets of 40 CLLs of model fits to jackknife replicates.

Linear admixture modeling and f_3 analysis

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, Australia) as a linear combination of the two "ancestral populations" (i.e., Europe and Africa) using an intercept-free linear model:

$$p_{i-admix} = \beta_1 (African\ Ancestor) + \beta_2 (European\ Ancestor_k) + \varepsilon \quad \text{(eq. 2)}$$

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 on the entire genome, as well as inside chromosomal

inversions, outside of inversions, and on non-coding mutations. In total we ran 1,313,070 comparisons (all available in **Dataset S2**).

We also assessed evidence of admixture using the f_3 statistic in the R package *poolfstat* (v2.3.0, Gautier et al., 2022). A significantly negative f_3 for a triplet configuration of the form f_3 (A;B,C) provides evidence for the target population A to originate from an admixture event between two source populations related to sampled populations B and C. We tested samples in the Americas and Australia to identify the most likely ancestral populations from Africa and Europe. For this analysis, we included 15 African populations (derived from seven countries: Cameroon, Egypt, Ethiopia, Morocco, Rwanda, South Africa, and Zambia) and all European samples as source population proxies. We used all populations in Australia and the Americas as targets of admixture.

Population differentiation

We analyzed patterns of population differentiation across samples and clusters using the R package *poolfstat* (v2.3.0, Gautier et al., *in prep.*). 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) autosomes, excluding heterochromatin; and iii) excluding heterochromatin and SNPs with MAF < 0.05. 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 the supplement, **Fig. S13**). 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), as follows:

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

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 (v2.3.0, Gautier et al., in prep). In addition to whole genome-estimates, window-wise hierarchical F_{ST} parameters were estimated across windows of 10 kb, 50 kb and 100 kb and are available in the DEST 2.0 browser.

GIM predictive models

GIMs 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 (SSE) 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 455 new 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 (MSE) 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 radius of 6378.137 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 (**Table S12**). First, we implemented the *BayPass* 2.4 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* 2.4, 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. (2022). We generated a null distribution of *XtX** statistics by simulating allele frequencies for ~9M 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 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 generalized linear mixed model (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 :

$$p_i = \alpha + X(year_{factor} : locality_{factor}) + \varepsilon \quad \text{(eq. 4)}$$

$$p_i = \alpha + \beta_1(season) + X(year_{factor} : locality_{factor}) + \varepsilon \quad \text{(eq. 5)}$$

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, and X is the random effect term coded as an interaction term between the year of collection and the locality where flies were collected, 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. 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 100000) on genes located in 10 kb windows of high differentiation (F_{GT} > 0.2; **Table S7**), -$\log_{10}$(wZa *p*-values) > 188.96 for the *XtX** statistic (**Table S8)**, and -$\log_{10}$(wZa *p*-values) > 3.65 for the C_2 statistic (**Table S9**), representing the 99.9th percentile from the simulated POD data (see above).

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 here among researchers to transport 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.

Author Contributions

All author contributions to this work are denoted in **Table S13**.

Acknowledgements

We 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). Nunez

acknowledges the Henderson-Harris fellowship program at the University of Vermont, also the Vermont Advanced Computing Center (VACC; URL: https://www.uvm.edu/vacc) for providing computational resources that contributed to this publication. Bergland 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. Coronado-Zamora and González 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). Gautier acknowledges the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. Obbard acknowledges Sue and Keith Obbard and Sandy Bayne for permission to collect flies on their land. Ansari 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. Serga acknowledges support from the PAUSE-ANR Ukraine Program. We also wish to thank Pavlo A. Kovalenko and Nadiia M. Pirko for their assistance with collecting flies in 2017-2021. Note: After 24 February 2022, no collaborative actions or exchanges have taken place within our project between Ukrainian and Russian scientists nor their institutions.

Funding

Nunez was supported by Start-up funds from the University of Vermont; Kapun was supported by the Horizon Europe project FAIRiCUBE (grant #101059238); Steindl was supported by the Horizon Europe project FAIRiCUBE (grant #101059238); Petrov was supported by the NIH 2R35GM11816506 (MIRA grant); Flatt was supported by the Swiss National Science Foundation (SNSF) grants 31003A-182262, 310030_219283, and FZEB-0-214654; Bergland was supported by the National Institutes of Health R35 GM119686, and National Science Foundation CAREER #2145688 grants. Gonzalez was supported by grant PID2020-115874GB-I00 funded by MICIU/AEI /10.13039/501100011033, MICIU/AEI /10.13039/501100011033, and by the European Commission NextGenerationEU/ PRTR, 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; Sánchez-Gracia was supported by the Ministerio de Ciencia e Innovación of Spain (MCIN/AEI/10.13039/501100011033; grant PID2020-113168GB-I00 to AS-G, and Comissió

Interdepartamental de Recerca I Innovació Tecnològica of Catalonia, Spain (2021SGR00279); Patenkovic was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-66/2024-03/ 200007; Barbadilla was supported by Ministerio de Ciencia e Innovación (PID2021-127107NB-I00), AGAUR Generalitat de Catalunya (SGR 00526); Schlötterer was supported by the Austrian Science Funds, FWF, 10.55776/P32935, 10.55776/P33734; Fricke was supported by the German Science Foundation (DFG, grant # FR2973/11-1); Obbard was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/T007516/1; Vela was supported by project QINV0196-IINV529010100 from the Pontificia Universidad Católica del Ecuador; Abbott was supported by VR-2015-04680, VR-2020-05412; Parsch was supported by the Deutsche Forschungsgemeinschaft (DFG) projects 255619725 and 503272152; Kankare was supported by the Academy of Finland project 322980; Guerreiro was supported by the Ministerio de Ciencia e Innovacion (PID2021-127107NB-I00), AGAUR Generalitat de Catalunya (SGR 00526); Veselinovic was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-65/2024-03/ 200178; Tanaskovic was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-66/2024-03/ 200007; Stamenkovic-Radak 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; Ritchie was supported by NERC, UK NE/V001566/1; Rera was supported by the Bettencourt Schueller Foundation long term partnership, this work was also partly supported by a CRI Core Research Fellowship; Jelić was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-65/2024-03/ 200178; Rakic was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (NITRA) grant no. 451-03-65/2024-03/ 200178; Erickson was supported by award #61-1673 from the Jane Coffin Childs Memorial Fund for Medical Research (www.jccfund.org); Ramos-Onsins 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); Casillas was supported by Ministerio de Ciencia e Innovación (PID2021-127107NB-I00); AGAUR Generalitat de Catalunya (SGR 00526); Hernandes was supported by Australian Research Council DP190102512; Kerdaffrec was supported by EMBO long-term fellowship ALT 248-02018;

Lawler was supported by Australian Research Council DP190102512; Colinet was supported by ANR Drothermal (ANR-20-CE02-011-01).

Data availability and the new DEST 2.0 web browser

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 (**Fig. S16**). 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 at our 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 sequences are available on the SRA (https://www.ncbi.nlm.nih.gov/sra) at PRJNA993612. Code is available in GitHub at: https://github.com/DEST-bio/DESTv2_data_paper. 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

Gautier M, Coronado-Zamora M and Vitalis R (2024). Estimating hierarchical *F*-statistics from Pool-Seq data.

Adrion JR, Galloway JG, Kern AD. 2020. Predicting the Landscape of Recombination Using Deep Learning.Wilke C, editor. *Mol. Biol. Evol.* 37:1790–1808.

Adrion JR, Hahn MW, Cooper BS. 2015. Revisiting classic clines in *Drosophila melanogaster* in the age of genomics. *Trends Genet.* 31:434–444.

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, Di Rienzo A. 2012. The Genetic Architecture of Adaptations to High Altitude in Ethiopia.Malik HS, editor. *PLoS Genet.* 8:e1003110.

Andolfatto P. 2001. Contrasting Patterns of X-Linked and Autosomal Nucleotide Variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* 18:279–290.

Arguello JR, Laurent S, Clark AG. 2019. Demographic History of the Human Commensal *Drosophila melanogaster*.Gaut B, editor. *Genome Biol. Evol.* 11:844–854.

Atkinson W, Shorrocks B. 1977. Breeding Site Specificity in the Domestic Species of *Drosophila*. *Oecologia* [Internet] 29. Available from: https://www.jstor.org/stable/4215461

Bangerter A. 2021. Dense seasonal sampling of an orchard population uncovers population turnover, adaptive tracking, and structure in multiple *Drosophila* species. Available from: https://libraetd.lib.virginia.edu/public_view/2801ph17g

Begun DJ, Aquadro CF. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* 365:548–550.

Behrman EL, Howick VM, Kapun M, Staubach F, Bergland AO, Petrov DA, Lazzaro BP, Schmidt PS. 2018. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proc. Biol. Sci.* 285.

Behrman EL, Schmidt P. 2022. How predictable is rapid evolution? Evolutionary Biology Available from: http://biorxiv.org/lookup/doi/10.1101/2022.10.27.514123

Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS. 2015. Seasonal variation in life history traits in two *Drosophila* species. *J. Evol. Biol.* 28:1691–1704.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. *PLoS Genet.* 10:e1004775.

1342 Bergland AO, Tobler R, González J, Schmidt P, Petrov D. 2016. Secondary contact and local
1343 adaptation contribute to genome-wide patterns of clinal variation in *Drosophila*
1344 *melanogaster*. *Mol. Ecol.* 25:1157–1174.
1345 Berry A, Kreitman M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in
1346 *Drosophila melanogaster* on the east coast of North America. *Genetics* 134:869–893.
1347 Betancourt NJ, Rajpurohit S, Durmaz E, Fabian DK, Kapun M, Flatt T, Schmidt P. 2021. Allelic
1348 polymorphism at *foxo* contributes to local adaptation in *Drosophila melanogaster*. *Mol.*
1349 *Ecol.* 30:2817–2830.
1350 Bitter MC, Berardi S, Oken H, Huynh A, Lappo E, Schmidt P, Petrov DA. 2024. Continuously
1351 fluctuating selection reveals fine granularity of adaptation. *Nature* [Internet]. Available
1352 from: https://www.nature.com/articles/s41586-024-07834-x
1353 Boettiger C. 2015. An introduction to Docker for reproducible research. *ACM SIGOPS Oper.*
1354 *Syst. Rev.* 49:71–79.
1355 Bogaerts-Márquez M, Guirao-Rico S, Gautier M, González J. 2020. Temperature, rainfall and
1356 wind variables underlie environmental adaptation in natural populations of *Drosophila*
1357 *melanogaster*. *Mol. Ecol.*
1358 Booker TR, Yeaman S, Whiting JR, Whitlock MC. 2024. The WZA: A window-based method for
1359 characterizing genotype–environment associations. *Mol. Ecol. Resour.* 24:e13768.
1360 Botero CA, Weissing FJ, Wright J, Rubenstein DR. 2015. Evolutionary tipping points in the
1361 capacity to adapt to environmental change. *Proc. Natl. Acad. Sci.* 112:184–189.
1362 Buri P. 1956. Gene frequency in small populations of mutant *Drosophila*. *Evolution* 10:367–402.
1363 Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging via
1364 overlap.Biggs PJ, editor. *PLOS ONE* 12:e0185056.
1365 Campo D, Lehmann K, Fjeldsted C, Souaiaia T, Kao J, Nuzhdin SV. 2013. Whole-genome
1366 sequencing of two North American *Drosophila melanogaster* populations reveals
1367 genetic differentiation and positive selection. *Mol. Ecol.* 22:5084–5097.
1368 Capy P, David JR, Allemand R, Carton Y, Febvay G, Kermarec A. 1986. Genetic analysis of
1369 *Drosophila melanogaster* in the French West Indies and comparison with populations
1370 from other parts of the world. *Genetica* 69:167–176.
1371 Caracristi G. 2003. Genetic Differentiation Between American and European *Drosophila*
1372 *melanogaster* Populations Could Be Attributed to Admixture of African Alleles. *Mol. Biol.*
1373 *Evol.* 20:792–799.
1374 Casillas S, Barbadilla A. 2017. Molecular Population Genetics. *Genetics* 205:1003–1035.
1375 Chen J, Liu C, Li W, Zhang W, Wang Y, Clark AG, Lu J. 2024. From sub-Saharan Africa to

China: Evolutionary history and adaptation of *Drosophila melanogaster* revealed by population genomics. *Sci. Adv.* 10:eadh3425.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. 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)* 6:80–92.

Cogni R, Kuczynski C, Koury S, Lavington E, Behrman EL, O'Brien KR, Schmidt PS, Eanes WF. 2014. The intensity of selection acting on the couch potato gene-spatial-temporal variation in a diapause cline: spatial-temporal variation in diapause cline. *Evolution* 68:538–548.

Comeron JM, Ratnappan R, Bailin S. 2012. The Many Landscapes of Recombination in *Drosophila melanogaster*. Petrov DA, editor. *PLoS Genet.* 8:e1002905.

Corbett-Detig R, Nielsen R. 2017. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy.Kang HM, editor. *PLOS Genet.* 13:e1006529.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.

Danielson PB, Letman JA, Fogleman JC. 1995. Alkaloid metabolism by cytochrome P-450 enzymes in *Drosophila melanogaster*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 110:683–688.

David J, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet.* 4:106–111.

De Jong G, Bochdanovits Z. 2003. Latitudinal clines in *Drosophila melanogaster*: Body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *J. Genet.* 82:207–223.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43:491–498.

Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, Guo E, Dider S, Zhang J, Bridge C, et al. 2023. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol.* 24:74.

Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM, the FlyBase Consortium. 2015. FlyBase: introduction of the *Drosophila*

melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* 43:D690–D697.

Dreissig S, Mascher M, Heckmann S. 2019. Variation in Recombination Rate Is Shaped by Domestication and Environmental Conditions in Barley.Purugganan M, editor. *Mol. Biol. Evol.* 36:2029–2039.

Duchen P, Živković D, Hutter S, Stephan W, Laurent S. 2013. Demographic Inference Reveals African and European Admixture in the North American *Drosophila melanogaster* Population. *Genetics* 193:291–301.

Ďureje Ľ, Macholán M, Baird SJE, Piálek J. 2012. The mouse hybrid zone in Central Europe: from morphology to molecules. *Folia Zool.* 61:308–318.

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

Erickson PA, Weller CA, Song DY, Bangerter AS, Schmidt P, Bergland AO. 2020. Unique genetic signatures of local adaptation over space and time for diapause, an ecologically relevant complex trait, in *Drosophila melanogaster*. *PLoS Genet.* 16:e1009110.

Fabian DK, Kapun M, Nolte V, Kofler R, Schmidt PS, Schlötterer C, Flatt T. 2012. Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Mol. Ecol.* 21:4748–4769.

Feder AF, Petrov DA, Bergland AO. 2012. LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data.Wu R, editor. *PLoS ONE* 7:e48588.

Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. 2013. Population genomics from pool sequencing. *Mol. Ecol.* 22:5561–5576.

Flatt T. 2020. Life-History Evolution and the Genetics of Fitness Components in *Drosophila melanogaster*. *Genetics* 214:3–48.

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

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

Gautier M. 2015. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* 201:1555–1579.

Gautier M. 2023. Efficient k-mer based curation of raw sequence data: application in *Drosophila suzukii*. *Peer Community J.* 3:e79.

1444 Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P,
1445 Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from
1446 next-generation sequencing data: pool-versus individual-based genotyping. *Mol. Ecol.*
1447 22:3766–3779.
1448 Gautier M, Vitalis R, Flori L, Estoup A. 2022. *f*-Statistics estimation and admixture graph
1449 construction with Pool-Seq or allele count data using the R package *poolfstat*. *Mol. Ecol.*
1450 *Resour.* 22:1394–1416.
1451 Glaser-Schmitt A, Ramnarine TJS, Parsch J. 2023. Rapid evolutionary change, constraints and
1452 the maintenance of polymorphism in natural populations of *Drosophila melanogaster*.
1453 *Mol. Ecol.*:mec.17024.
1454 Gleason JM, Roy PR, Everman ER, Gleason TC, Morgan TJ. 2019. Phenology of *Drosophila*
1455 species across a temperate growing season and implications for behavior. Desneux N,
1456 editor. *PLOS ONE* 14:e0216601.
1457 Gower G, Ragsdale AP, Bisschop G, Gutenkunst RN, Hartfield M, Noskova E, Schiffels S,
1458 Struck TJ, Kelleher J, Thornton KR. 2022. Demes: a standard format for demographic
1459 models.Coop G, editor. *Genetics* 222:iyac131.
1460 Grenier JK, Arguello JR, Moreira MC, Gottipati S, Mohammed J, Hackett SR, Boughton R,
1461 Greenberg AJ, Clark AG. 2015. Global diversity lines-a five-continent reference panel of
1462 sequenced Drosophila melanogaster strains. *G3 Bethesda Md* 5:593–603.
1463 Guirao-Rico S, González J. 2021. Benchmarking the performance of Pool-seq SNP callers
1464 using simulated and real sequencing data. *Mol. Ecol. Resour.* 21:1216–1229.
1465 Günther T, Coop G. 2013. Robust Identification of Local Adaptation from Allele Frequencies.
1466 *Genetics* 195:205–220.
1467 Hales KG, Korey CA, Larracuente AM, Roberts DM. 2015. Genetics on the Fly: A Primer on the
1468 *Drosophila* Model System. *Genetics* 201:815–842.
1469 Haudry A, Laurent S, Kapun M. 2020. Population Genomics on the Fly: Recent Advances in
1470 *Drosophila*. In: Dutheil JY, editor. Statistical Population Genomics. Vol. 2090. Methods in
1471 Molecular Biology. New York, NY: Springer US. p. 357–396. Available from:
1472 https://link.springer.com/10.1007/978-1-0716-0199-0_15
1473 Hewitt GM. 2011. Quaternary phylogeography: the roots of hybrid zones. *Genetica*
1474 139:617–638.
1475 Hijmans RJ, Karney C, Williams E, Vennes C. 2022. Package ‘geosphere.’ Available from:
1476 10.32614/CRAN.package.geosphere
1477 Hivert V, Leblois R, Petit EJ, Gautier M, Vitalis R. 2018. Measuring Genetic Differentiation from

Pool-seq Data. *Genetics* 210:315–330.

Hoffmann AA, Anderson A, Hallas R. 2002. Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecol. Lett.* 5:614–618.

Hoffmann AA, Weeks AR. 2007. 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* 129:133.

Hofman S, Spolsky C, Uzzell T, Cogălniceanu D, Babik W, Szymura JM. 2007. Phylogeography of the fire-bellied toads *Bombina*: independent Pleistocene histories inferred from mitochondrial genomes. *Mol. Ecol.* 16:2301–2316.

Hunter CM, Huang W, Mackay TFC, Singh ND. 2016. The Genetic Architecture of Natural Variation in Recombination Rate in *Drosophila melanogaster*. Sekelsky J, editor. *PLOS Genet.* 12:e1005951.

Ives PT. 1945. The genetic structure of American populations of *Drosophila melanogaster*. *Genetics* 30:167–196.

Ives PT. 1970. Further genetic studies of the south amherst population of *Drosophila melanogaster*. *Evol. Int. J. Org. Evol.* 24:507–518.

Izquierdo JI. 1991. How does *Drosophila melanogaster* overwinter? *Entomol. Exp. Appl.* 59:51–58.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. 2005. Distinguishing Between Selective Sweeps and Demography Using DNA Polymorphism Data. *Genetics* 170:1401–1410.

Johnson OL, Tobler R, Schmidt JM, Huber CD. 2023. Fluctuating selection and the determinants of genetic variation. *Trends Genet.* 39:491–504.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 11:94.

Jouganous J, Long W, Ragsdale AP, Gravel S. 2017. Inferring the Joint Demographic History of Multiple Populations: Beyond the Diffusion Approximation. *Genetics* 206:1549–1567.

Jurka J. 2000. Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* 16:418–420.

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

Kapopoulou A, Kapun M, Pieper B, Pavlidis P, Wilches R, Duchen P, Stephan W, Laurent S. 2020. Demographic analyses of a new sample of haploid genomes from a Swedish

population of *Drosophila melanogaster*. *Sci. Rep.* 10:22415.

Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-Stabelli O, Kankare M, Bogaerts-Márquez M, et al. 2020. Genomic Analysis of European *Drosophila melanogaster* Populations Reveals Longitudinal Structure, Continent-Wide Selection, and Previously Unknown DNA Viruses.Falush D, editor. *Mol. Biol. Evol.* 37:2661–2678.

Kapun Martin, Fabian DK, Goudet J, Flatt T. 2016. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. *Mol. Biol. Evol.* 33:1317–1336.

Kapun M, Flatt T. 2019. The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Mol. Ecol.* 28:1263–1282.

Kapun M, Mitchell ED, Kawecki TJ, Schmidt P, Flatt T. 2023. An Ancestral Balanced Inversion Polymorphism Confers Global Adaptation.Rogers R, editor. *Mol. Biol. Evol.* 40: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. 2021. *Drosophila* Evolution over Space and Time (DEST): A New Population Genomics Resource.Nielsen R, editor. *Mol. Biol. Evol.* 38:5782–5805.

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

Kapun M, Van Schalkwyk H, McAllister B, Flatt T, Schlötterer C. 2014. Inference of chromosomal inversion dynamics from Pool- Seq data in natural and laboratory populations of *Drosophila melanogaster* . *Mol. Ecol.* 23:1813–1827.

Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. *Curr. Biol.* 17:R77–R81.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26:2204–2207.

Kofler R, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28:2084–2085.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of Drosophila melanogaster. *Genetics* 187:245–260.

Köster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* 28:2520–2522.

Kreitman M. 1983. Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila*

melanogaster. *Nature* 304:412–417.

Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol.*:159–225.

Lachaise D, Silvain J-F. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster–D. simulans* palaeogeographic riddle. *Genetica* 120:17–39.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The *Drosophila* Genome Nexus: A Population Genomic Resource of 623 *Drosophila melanogaster* Genomes, Including 197 from a Single Ancestral Range Population. *Genetics* 199:1229–1241.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A Thousand Fly Genomes: An Expanded *Drosophila* Genome Nexus. *Mol. Biol. Evol.* 33:3308–3313.

Lange JD, Bastide H, Lack JB, Pool JE. 2022. A Population Genomic Assessment of Three Decades of Evolution in a Natural *Drosophila* Population.Rogers R, editor. *Mol. Biol. Evol.* 39:msab368.

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, et al. 2012. Genomic Variation in Natural Populations of *Drosophila melanogaster*. *Genetics* 192:533–598.

Lawton D, Huseth AS, Kennedy GG, Morey AC, Hutchison WD, Reisig DD, Dorman SJ, Dillard D, Venette RC, Groves RL, et al. 2022. Pest population dynamics are related to a continental overwintering gradient. *Proc. Natl. Acad. Sci.* 119:e2203230119.

Le Goff G, Hilliou F. 2017. Resistance evolution in *Drosophila*: the case of *CYP6G1* . *Pest Manag. Sci.* 73:493–499.

Lê S, Josse J, Husson F. 2008. **FactoMineR**: An *R* Package for Multivariate Analysis. *J. Stat. Softw.* [Internet] 25. Available from: http://www.jstatsoft.org/v25/i01/

Lemeunier F, Aulard S. 1992. Inversion polymorphism in *Drosophila melanogaster*. In: *Drosophila* Inversion Polymorphism. In C. B. Krimbas, & J. R. Powell (Eds.). Boca Raton, FL: CRC Press. p. 339–405.

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

Lhotka O, Kyselý J. 2022. The 2021 European Heat Wave in the Context of Past Major Heat Waves. *Earth Space Sci.* 9:e2022EA002567.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from: https://arxiv.org/abs/1303.3997

1580 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
1581 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map
1582 format and SAMtools. *Bioinformatics* 25:2078–2079.
1583 Li H, Stephan W. 2006. Inferring the Demographic History and Rate of Adaptive Substitution in
1584 Drosophila.Przeworski M, editor. *PLoS Genet.* 2:e166.
1585 Lintner JA. 1882. First Annual Report on the Injurious and Other Insects of the State of New
1586 York. Albany, New York: Weed, Parsons and Co.
1587 Ma J, Amos CI. 2012. Principal Components Analysis of Population Admixture.You M, editor.
1588 *PLoS ONE* 7:e40115.
1589 Machado HE, Bergland AO, O'Brien KR, Behrman EL, Schmidt PS, Petrov DA. 2016.
1590 Comparative population genomics of latitudinal variation in *Drosophila simulans* and
1591 *Drosophila melanogaster*. *Mol. Ecol.* 25:723–740.
1592 Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González
1593 J, Karasov TL, et al. 2021. Broad geographic sampling reveals the shared basis and
1594 environmental correlates of seasonal adaptation in *Drosophila*. Nordborg M, Wittkopp
1595 PJ, Nordborg M, editors. *eLife* 10:e67577.
1596 Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y,
1597 Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* Genetic Reference
1598 Panel. *Nature* 482:173–178.
1599 Mansourian S, Enjin A, Jirle EV, Ramesh V, Rehermann G, Becher PG, Pool JE, Stensmyr MC.
1600 2018. Wild African *Drosophila melanogaster* Are Seasonal Specialists on Marula Fruit.
1601 *Curr. Biol.* 28:3960-3968.e3.
1602 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1603 *EMBnet.journal* 17:10.
1604 Mateo L, Rech GE, González J. 2018. Genome-wide patterns of local adaptation in Western
1605 European *Drosophila melanogaster* natural populations. *Sci. Rep.* 8:16143.
1606 McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*.
1607 *Nature* 351:652–654.
1608 Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S,
1609 Twardziok SO, Kanitz A, et al. 2021. Sustainable data analysis with Snakemake.
1610 *F1000Research* 10:33.
1611 Morgulis A, Gertz EM, Schaffer AA, Agarwala R. 2006. WindowMasker: window-based masker
1612 for sequenced genomes. *Bioinformatics* 22:134–141.
1613 Muggeo VMR. 2003. Estimating regression models with unknown break-points. *Stat. Med.*

22:3055–3071.

Nei M. 1973. Analysis of Gene Diversity in Subdivided Populations. *Proc. Natl. Acad. Sci.* 70:3321–3323.

Nelder JA, Mead R. 1965. A Simplex Method for Function Minimization. *Comput. J.* 7:308–313.

Nunez JCB, Lenhart BA, Bangerter A, Murray CS, Mazzeo GR, Yu Y, Nystrom TL, Tern C, Erickson PA, Bergland AO. 2024. A cosmopolitan inversion facilitates seasonal adaptation in overwintering *Drosophila*. Ralph P, editor. *GENETICS* 226:iyad207.

Obbard DJ, Maclennan J, Kim K-W, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating Divergence Dates and Substitution Rates in the *Drosophila* Phylogeny. *Mol. Biol. Evol.* 29:3459–3473.

Obbard DJ, Welch JJ, Kim K-W, Jiggins FM. 2009. Quantifying Adaptive Evolution in the *Drosophila* Immune System.Begun DJ, editor. *PLoS Genet.* 5:e1000698.

Olazcuaga L, Foucaud J, Deschamps C, Loiseau A, Claret J-L, Vedovato R, Guilhot R, Sévely C, Gautier M, Hufbauer RA, et al. 2022. Rapid and transient evolution of local adaptation to seasonal host fruits in an invasive pest fly. *Evol. Lett.* 6:490–505.

Ometto L, Glinka S, De Lorenzo D, Stephan W. 2005. Inferring the Effects of Demography and Selection on *Drosophila melanogaster* Populations from a Chromosome-Wide Scan of DNA Variation. *Mol. Biol. Evol.* 22:2119–2130.

Otte KA, Nolte V, Mallard F, Schlötterer C. 2021. The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime. *Genome Biol.* 22:211.

Paaby AB, Bergland AO, Behrman EL, Schmidt PS. 2014. A highly pleiotropic amino acid polymorphism in the *Drosophila* insulin receptor contributes to life-history adaptation. *Evolution* 68:3395–3409.

Parsons PA. 1975. The Comparative Evolutionary Biology of the Sibling Species, *Drosophila melanogaster* and *D. Simulans*. *Q. Rev. Biol.* 50:151–169.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* 192:1065–1093.

Pavlidis P, Jensen JD, Stephan W. 2010. Searching for Footprints of Positive Selection in Whole-Genome SNP Data From Nonequilibrium Populations. *Genetics* 185:907–922.

Powell JR. 1997. Progress and prospects in evolutionary biology: the *Drosophila model.*

Rajpurohit S, Gefen E, Bergland AO, Petrov DA, Gibbs AG, Schmidt PS. 2018. Spatiotemporal dynamics and genome-wide association analysis of desiccation tolerance in *Drosophila melanogaster*. *Mol. Ecol.* 27:3525–3540.

1648 Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD. 2014. Parallel Geographic
1649 Variation in *Drosophila melanogaster*. *Genetics* 197:361–373.
1650 Remington CL. 1968. Suture-Zones of Hybrid Interaction Between Recently Joined Biotas. In:
1651 Dobzhansky T, Hecht MK, Steere WC, editors. Evolutionary Biology. Boston, MA:
1652 Springer US. p. 321–428. Available from:
1653 http://link.springer.com/10.1007/978-1-4684-8094-8_8
1654 Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T, Petrov
1655 DA, Schmidt P. 2022. Direct observation of adaptive tracking on ecological time scales in
1656 *Drosophila*. *Science* 375:eabj7484.
1657 Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. 2020. Natural Selection Shapes Variation in
1658 Genome-wide Recombination Rate in *Drosophila pseudoobscura*. *Curr. Biol.*
1659 30:1517-1528.e6.
1660 Sanchez-Refusta F, Santiago E, Rubio J. 1990. Seasonal fluctuations of cosmopolitan inversion
1661 frequencies in a natural population of *Drosophila melanogaster*. *Genet. Sel. Evol.*
1662 22:47–56.
1663 Schadt CW, Martin AP, Lipson DA, Schmidt SK. 2003. Seasonal Dynamics of Previously
1664 Unknown Fungal Lineages in Tundra Soils. *Science* 301:1359–1361.
1665 Schlötterer C, Tobler R, Kofler R, Nolte V. 2014. Sequencing pools of individuals — mining
1666 genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* 15:749–763.
1667 Schmidt PS, Conde DR. 2006. Environmental heterogeneity and the maintenance of genetic
1668 variation for reproductive diapause in *Drosophila melanogaster*. *Evol. Int. J. Org. Evol.*
1669 60:1602–1611.
1670 Schmidt PS, Zhu C-T, Das J, Batavia M, Yang L, Eanes WF. 2008. An amino acid polymorphism
1671 in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila*
1672 *melanogaster*. *Proc. Natl. Acad. Sci.* 105:16207–16211.
1673 Serga SV, Maistrenko OM, Rozhok AI, Mousseau TA, Kozeretska IA. 2015. Colonization of a
1674 temperate-zone region by the fruit fly *Drosophila simulans* (Diptera: Drosophilidae). *Can.*
1675 *J. Zool.* 93:799–804.
1676 Siddiq MA, Thornton JW. 2019. Fitness effects but no temperature-mediated balancing selection
1677 at the polymorphic *Adh* gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.*
1678 116:21634–21640.
1679 Signor SA, New FN, Nuzhdin S. 2018. A large panel of *Drosophila simulans* reveals an
1680 abundance of common variants. *Genome Biol. Evol.* 10:189–206.
1681 Smit A, Hubley R, Green P. 1996. RepeatMasker Open-3.0. Available from:

1682 http://www.repeatmasker.org.
1683 Sparks A. 2018. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and
1684 Climatology Data Client for R. *J. Open Source Softw.* 3:1035.
1685 Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV, Stensmyr MC,
1686 Pool JE. 2020. Recurrent Collection of *Drosophila melanogaster* from Wild African
1687 Environments and Genomic Insights into Species History. *Mol. Biol. Evol.* 37:627–638.
1688 Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams Jr RM. 1949. The American soldier:
1689 Adjustment during army life.(studies in social psychology in world war ii), vol. 1.
1690 Sturtevant AH. 1921. The North American species of *Drosophila*. Carnegie institution of
1691 Washington
1692 Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ,
1693 Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny
1694 of 155 *Drosophila* genomes. *Curr. Biol.* 32:111-123.e5.
1695 Svetec N, Cridland JM, Zhao L, Begun DJ. 2016. The Adaptive Significance of Natural Genetic
1696 Variation in the DNA Damage Response of *Drosophila melanogaster*. Presgraves DC,
1697 editor. *PLOS Genet.* 12:e1005869.
1698 Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*
1699 105:437–460.
1700 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA
1701 polymorphism. *Genetics* 123:585–595.
1702 Teshima KM, Coop G, Przeworski M. 2006. How reliable are empirical genomic scans for
1703 selective sweeps? *Genome Res.* 16:702–712.
1704 Thornton KR, Jensen JD. 2007. Controlling the False-Positive Rate in Multilocus Genome
1705 Scans for Selection. *Genetics* 175:737–750.
1706 Tibshirani R, Walther G, Hastie T. 2001. Estimating the Number of Clusters in a Data Set Via the
1707 Gap Statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 63:411–423.
1708 Wang Y, McNeil P, Abdulazeez R, Pascual M, Johnston SE, Keightley PD, Obbard DJ. 2023.
1709 Variation in mutation, recombination, and transposition rates in *Drosophila melanogaster*
1710 and *Drosophila simulans*. *Genome Res.* 33:587–598.
1711 Xu R, Lou Y, Tidu A, Bulet P, Heinekamp T, Martin F, Brakhage A, Li Z, Liégeois S, Ferrandon D.
1712 2023. The Toll pathway mediates *Drosophila* resilience to *Aspergillus* mycotoxins
1713 through specific Bomanins. *EMBO Rep.* 24:e56036.
1714 Yu Y, Bergland AO. 2022. Distinct signals of clinal and seasonal allele frequency change at
1715 eQTLs in *Drosophila melanogaster*. *Evolution* 76:2758–2768.

1716 Zhang Z, Zhu S. 2009. Drosomycin, an essential component of antifungal defence in
1717 *Drosophila*. *Insect Mol. Biol.* 18:549–556.
1718 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance
1719 computing toolset for relatedness and principal component analysis of SNP data.
1720 *Bioinformatics* 28:3326–3328.