GENETICS

From sub-Saharan Africa to China: Evolutionary history and adaptation of *Drosophila melanogaster* revealed by population genomics

Junhao Chen[1]†, Chenlu Liu[1]†, Weixuan Li[1]†, Wenxia Zhang[1], Yirong Wang[2], Andrew G. Clark[3]*, Jian Lu[1]*

***Drosophila melanogaster* is a widely used model organism for studying environmental adaptation. However, the genetic diversity of populations in Asia is poorly understood, leaving a notable gap in our knowledge of the global evolution and adaptation of this species. We sequenced genomes of 292 *D. melanogaster* strains from various ecological settings in China and analyzed them along with previously published genome sequences. We have identified six global genetic ancestry groups, despite the presence of widespread genetic admixture. The strains from China represent a unique ancestry group, although detectable differentiation exists among populations within China. We deciphered the global migration and demography of *D. melanogaster*, and identified widespread signals of adaptation, including genetic changes in response to insecticides. We validated the effects of insecticide resistance variants using population cage trials and deep sequencing. This work highlights the importance of population genomics in understanding the genetic underpinnings of adaptation, an effort that is particularly relevant given the deterioration of ecosystems.**

INTRODUCTION

The genetic mechanisms driving adaptation to diverse environments have been a central focus of evolutionary biology for decades. With increasing threats to biodiversity and habitat loss from human activities, understanding how organisms adapt to changing environments has become more critical than ever before. Population genomics has emerged as a powerful tool to identify the genetic variants responsible for adaptation to local conditions in humans (*1*) and a variety of other organisms [e.g., (*2*, *3*)]. This approach involves sequencing the genomes of many individuals within populations to identify the genetic variants associated with adaptive traits. However, identifying the causative loci underlying environmental adaptation is often complicated by population structure and demographic history, making it a challenge to distinguish between the effects of adaptation and other factors (*4*). Although functional and analytical tools in population genomics have improved considerably, establishing causal relationships between candidate loci and phenotypes remains a challenging task.

Drosophila melanogaster is a model organism that has been used extensively in the study of the genetic basis of environmental adaptation. This species originated in sub-Saharan Africa and has since colonized a wide variety of habitats as a human commensal, making it a valuable system for studying environmental adaptation despite its complex demographic history. Recent studies, many of which have been conducted at the population genomic level, have remarkably improved our understanding of the demographic history and adaptation of *D. melanogaster* (*5*–*16*). These studies have revealed the genetic architecture of various adaptive traits in natural populations, including starvation resistance (*14*, *15*), thermal tolerance (*14*, *15*, *17*), insecticide resistance (*13*, *16*), and pigmentation (*18*). Furthermore, the impact of seasonal and clinal variation in allele frequencies in response to changing environments has been established (*19*–*21*).

Despite these findings, previous studies have primarily focused on *D. melanogaster* in Europe, Africa, Oceania, and North America (*5*–*12*, *14*–*16*, *22*), leaving a notable gap in our knowledge of the evolution and adaptation of this species at the global level. East Asia, which boasts a rich resource of *Drosophila* species (*23*), is an area of particular interest for studying the genetic diversity and adaptation of *D. melanogaster*. Although early studies have identified phenotypic divergences between *D. melanogaster* in Asia and other continents (*24*) and conducted some genetic diversity analysis of certain loci or whole-genome sequencing of a limited number ($n = 15$) of *D. melanogaster* strains in China (*9*), these studies have provided limited or even conflicting information on the genetic diversity of *D. melanogaster* in China. Further investigations are, therefore, necessary to enhance our understanding of the genetic diversity and adaptation of *D. melanogaster* at the global scale.

China is renowned for its high levels of biodiversity compared to other countries at similar latitudes, owing to its diverse climate gradient and rich geodiversity. In addition, rapid increases in human population, urbanization, and widespread insecticide usage, among other anthropogenic factors, have led to considerable changes in biodiversity patterns (*25*). Here, we present a large-scale sequencing of 292 *D. melanogaster* strains collected from 52 geographic locations across China, including regions such as Tibet and Xinjiang that are both geographically and ecologically distinct from other areas. This sampling approach provides a unique opportunity to explore the genetic diversity and local adaptation of *D. melanogaster* across a range of ecological settings. By combining our data with previously published genome sequences from other continents (*5*–*9*, *14*–*16*, *20*, *26*–*31*), we can investigate the population history, gene flow, and local adaptation of *D. melanogaster* to diverse environments during its migration out of Africa. Our findings reveal

[1] State Key Laboratory of Protein and Plant Gene Research, Center for Bioinformatics, School of Life Sciences, Peking University, Beijing 100871, China. [2] College of Biology, Hunan University, Changsha 410082, China. [3] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA.
*Corresponding author. Email: ac347@cornell.edu (A.G.C.); luj@pku.edu.cn (J.L.)
†These authors contributed equally to this work.

widespread signals of adaptation that are either population-specific or shared across populations, including signals of adaptation to insecticides, a substantial environmental pressure that has been increasingly used to control agricultural pests. We further verified the effects of insecticide resistance variants that showed signals of adaptation. Our results shed light on the genetic mechanisms that underlie environmental adaptation and insecticide resistance in *D. melanogaster* and offer important insights into the impact of environmental changes on ecosystems.

RESULTS

Genetic variant discovery by whole-genome sequencing of 292 *D. melanogaster* strains in China

We sequenced the genomes of 292 wild-derived iso-female strains of *D. melanogaster* collected from 52 geographic locations representing different ecological settings in China between 2017 and 2020 (referred to as the CN population; Fig. 1A). Male adults of each strain were sequenced to a mean genome coverage of 26.2× (ranging from 21.9 to 30.3×) using Illumina NovaSeq 6000 technology (table S1). By integrating previously published genome sequences of *D. melanogaster* (*5*–*9*, *14*–*16*, *20*, *26*–*31*), we analyzed genome sequences in a total of 1356 *D. melanogaster* strains, representing 20 geographic populations from sub-Saharan Africa, northern Africa, Europe, North America, Oceania, and Asia (Fig. 1B, Table 1, and table S1). We used the previously described approach (*6*) to identify heterozygous blocks in a strain (table S2). Iso-female lines, characterized by passive inbreeding, demonstrated a propensity for longer heterozygous blocks (e.g., average total length for strains: CN 113.69 Mb; CAS 124.74 Mb; STO 124.76 Mb) compared to full-sib inbreeding strains (e.g., FR 67.20 Mb; RAL 25.28 Mb; EG 68.45 Mb). We excluded the strains sequenced using haploid embryos (i.e., ZI and RG populations) in the heterozygous block analysis because they are not expected to have genuine heterozygosity (*6*, *8*). We identified 10,488,937 biallelic single-nucleotide polymorphisms (SNPs) on autosomes and 2,383,127 biallelic SNPs on the X chromosome. Using *D. simulans* and *D. yakuba* as outgroups, we found that 87.7% of the SNPs had derived allele frequencies lower than 5%, and nonsynonymous SNPs showed even lower frequencies (fig. S1). We identified 3,019,941 autosomal and 360,412 X-linked SNPs in the 292 CN strains (table S3). The DFE-alpha (*32*) analysis revealed that when using fourfold degenerate sites as a neutral proxy, 66.9% and 71.8% of the nonsynonymous mutations were strongly deleterious ($-N_e s > 100$) in autosomes and the X chromosome, respectively (fig. S2).

The autosomal genetic diversity (π_A) in the CN population was 0.320%, comparable to that in the other 10 cosmopolitan populations (π_A ranging from 0.280% to 0.343%, with a median of 0.320%). Within CN, π_A values were similar for strains collected in Xinjiang (CN_XJ, $n = 25$, $\pi_A = 0.322\%$), on the Qinghai-Tibet Plateau (CN_QTP, $n = 63$, $\pi_A = 0.297\%$), and in other regions of China (CN_Other, $n = 204$, $\pi_A = 0.323\%$). The genetic diversity of the X chromosome (π_X) was also comparable between CN (0.186%) and other cosmopolitan populations (ranging from 0.150% to 0.215%, with a median of 0.187%). π_X was substantially lower than π_A in all the cosmopolitan populations (Table 1), even after adjusting for the difference in effective population size (fig. S3). As expected (*5*, *8*, *9*), the populations in sub-Saharan Africa showed substantially higher genetic diversities than CN and other cosmopolitan populations (Table 1), and these patterns were observed for different functional categories of variants (fig. S4 and table S4). Overall, this expanded dataset provides a substantial genetic resource for decoding the genetic diversity of *D. melanogaster*, enabling us to investigate both common and population-specific selective signals in various environmental settings worldwide.

Population structure of *D. melanogaster* at the global scale

To determine the genetic structure of worldwide populations of *D. melanogaster*, we performed principal components analysis (PCA) using SNPs in the putative neutral regions [including short introns (*33*) and fourfold degenerate sites of protein-coding regions]. Chromosomal inversions in the autosomes are a common occurrence within *D. melanogaster* populations (*34*). Using established methodologies (*35*), we identified these inversions in each individual strain (table S5 and fig. S5). The PCA involving all strains separates those with inversions from those without, distinctly evident on a global scale (fig. S6A), within cosmopolitan populations (fig. S6B), and even within CN populations (fig. S6C). These findings align with previous studies suggesting that inversions play an important role in shaping genetic diversity and can vary within and between populations (*34*).

Next, for each chromosome arm, we considered only the strains without inversions in the PCA. The PCAs of autosomes (Fig. 1C and fig. S7) and the X chromosome (Fig. 1D) yielded very similar results. At the global scale, the cosmopolitan populations grouped together and were well distinguished from sub-Saharan populations by the first principal component (PC1), which captured 6.6 to 7.6% of the variance on autosomes (Fig. 1C and fig. S7A) and 9.5% of the variance on the X chromosome (Fig. 1D). PC2 separated strains from sub-Sahara into three clusters, with strains from Ethiopia far from strains from southern Africa, and western and eastern Africa in between [the strains from western and eastern Africa were not well distinguishable on autosomes (Fig. 1C and fig. S7A) but clearly separated on the X chromosome (Fig. 1D)]. The PCAs split the cosmopolitan populations into three clusters (Fig. 1, C and D, and fig. S7B). PC1 separated strains from Asia (including CN and B) from the other populations, whereas PC2 separated the strains in Europe and northern Africa from those in North America and Oceania (Fig. 1, C and D, and fig. S7B). CN_QTP strains clustered together and were distinguishable from CN_Other or CN_XJ in both autosomes and the X chromosome (Fig. 1, C and D, and fig. S7C). The differentiation between CN_XJ and CN_Other strains varies across chromosomes, with no apparent separation on 2L, a modest level of differentiation on 2R, 3L, and 3R (Fig. 1C and fig. S7C), and pronounced differentiation on the X chromosome (Fig. 1D).

The PCA results were well supported by the results from ADMIXTURE (*36*), which quantitatively estimates the composition and admixture of genetic ancestries of each strain. The best-fitting ADMIXTURE results were obtained when the number of ancestry groups *K* was 6 to 8, dependent on the chromosome analyzed (Fig. 1, E and F, and fig. S8). Although the six ancestry groups identified on chromosome 2L ($K = 6$) were broadly present across other chromosomes, some chromosome-specific ancestry groups were observed, contributing a minor portion to the genetic makeup. To be more specific, for chromosome 2R ($K = 7$), the seventh ancestry group dominated in European (N, STO, CAS), T, EG, and EA strains. On chromosome 3L ($K = 8$), the seventh ancestry group was detected in

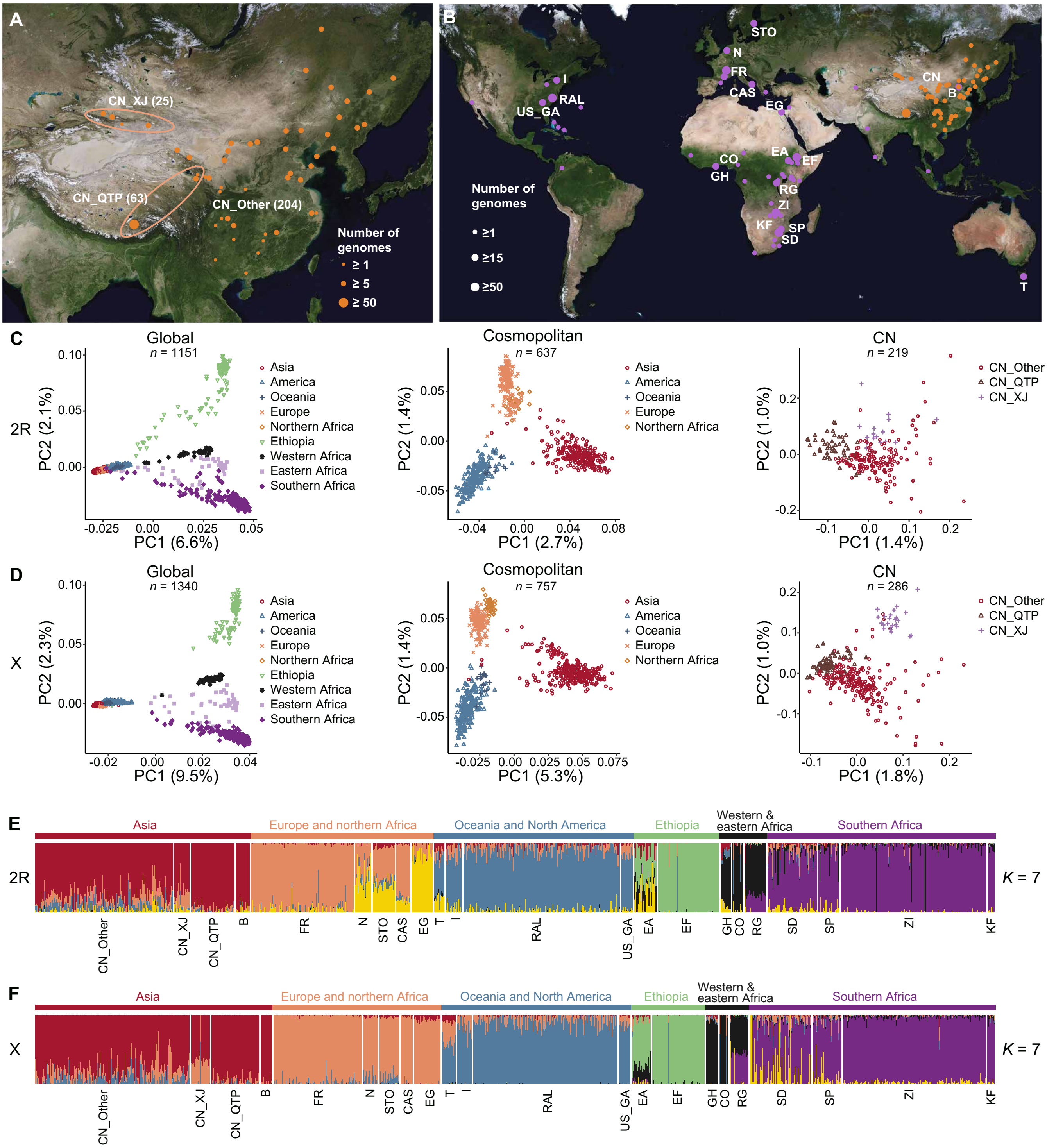


Fig. 1. Sample locations and population structure of *D. melanogaster*. (**A**) Geographic locations of 292 strains collected in China (CN), including three subpopulations: Xinjiang (CN_XJ), Qinghai-Tibet Plateau (CN_QTP), and other parts of China (CN_Other), with the sample size shown in parentheses. (**B**) Sample locations of 1356 genomes of *D. melanogaster* involved in this study, with the names of 20 representative populations provided. The orange dots represent the strains collected and sequenced in this study, while the purple dots represent the genomes previously published. (**C** and **D**) PCA based on neutral SNPs on inversion-free chromosome arm 2R (C) and the X chromosome (D) for global strains, cosmopolitan strains, and strains in China. Similar results were obtained for other autosomal arms (fig. S7). The number of samples (*n*) labeled on the graph represents the remaining sample count after one round of outlier removal in PCA analysis. The proportion of variance explained by each corresponding principal component is indicated in parentheses. (**E** and **F**) Admixture proportions inferred for each strain in these 20 populations, with subpopulations in CN shown separately. The results, based on neutral SNPs in inversion-free 2R (E) and X (F), are displayed. Strains are sorted by their geographic origin and represented by color-coded vertical columns that reflect the compositions of their ancestry. The optimal number of ancestries (*K*) was determined when the minimum CV error was reached (fig. S8A).

Table 1. Population IDs and diversity.

Population	Locality	Number of genomes	π_A (%)	π_X (%)	Reference
Asia					
CN	China	292	0.320	0.186	This study
CN_Other	Other parts of China	204	0.323	0.187	
CN_QTP	Qinghai-Tibet Plateau, China	63	0.297	0.168	
CN_XJ	Xinjiang, China	25	0.322	0.188	
B	Beijing, China	15	0.287	0.15	(*9*)
Europe and northern Africa					
FR	Lyon and Montpellier, France	117	0.295	0.181	(*7*, *8*, *26*)
STO	Stockholm, Sweden	26	0.296	0.194	(*16*)
N	Houten, Netherlands	19	0.291	0.176	(*9*)
CAS	Castellana Grotte, Italy	16	0.280	0.183	(*16*)
EG	Cairo, Egypt	35	0.343	0.182	(*6*, *8*)
Oceania and North America					
RAL	Raleigh, NC, United States	191	0.325	0.199	(*14*, *15*)
I	Ithaca, NY, United States	19	0.319	0.202	(*9*)
US_GA	Athens, GA, United States	15	0.338	0.215	(*26*)
T	Sorell, TAS, Australia	18	0.336	0.201	(*9*)
Ethiopia					
EF	Fiche, Ethiopia	69	0.342	0.324	(*6*, *8*)
EA	Gambella, Ethiopia	25	0.386	0.359	(*6*, *8*)
Western and eastern Africa					
GH	Accra, Ghana	15	0.390	0.369	(*26*)
CO	Oku, Cameroon	12	0.367	0.35	(*7*)
RG	Gikongoro, Rwanda	25	0.406	0.423	(*7*)
Southern Africa					
ZI	Siavonga, Zambia	189	0.459	0.455	(*6*, *7*)
SD	Dullstroom, South Africa	81	0.431	0.437	(*6*, *8*)
SP	Phalaborwa, South Africa	38	0.435	0.438	(*7*, *8*)
KF	Kafue, Zambia	10	0.393	0.334	(*27*)

European (N, STO), EG, EA, and GH strains, while the eighth ancestry group was detected in southern African strains (SD, SP, ZI, KF). Chromosomes 3R and X (both at $K = 7$) primarily displayed the seventh ancestry group within the SD and SP strains. Integrating the PCA and ADMIXTURE results, we proposed that the *D. melanogaster* populations examined here consist of six genetic ancestry groups: southern African (KF, ZI, SD, and SP), western and eastern African (RG, CO, and GH), Ethiopian (EA and EF), American and Australian (RAL, I, US_GA, and T), European and northern African (EG, FR, N, STO, and CAS), and Asian (CN and B). Although ADMIXTURE detected European and North American ancestries in the CN strains, the CN strains represent a unique genetic ancestry (Fig. 1, E and F; fig. S8; and table S6). Furthermore, ADMIXTURE analysis showed no differentiation within CN strains on 2L ($K = 1$). However, across other chromosomes, ADMIXTURE revealed two distinct ancestries: CN_Other and CN_XJ displayed a similar level of mixed ancestry, while CN_QTP showed relatively lower admixture (fig. S9). Overall, ADMIXTURE analysis aligns with PCA results, indicating the genetic distinctiveness of *D. melanogaster* in China from other populations. Within CN strains, CN_QTP is clearly distinguished from the other two populations, and a moderate differentiation exists between CN_XJ and CN_Other strains.

F_{ST} analysis supporting the isolation by distance theory of *D. melanogaster*

As in humans (*37*), a positive correlation was detected in *D. melanogaster* between the pairwise within-continental geographic distance and F_{ST}

(the fixation index) (*38*) using SNPs in autosomes (linear regression, $r^2 = 0.43$, $P = 3.88 \times 10^{-8}$) or the X chromosome ($r^2 = 0.47$, $P = 4.21 \times 10^{-9}$; fig. S10). This finding supports the theory of isolation by distance (*39*, *40*). Although F_{ST} between the CN and other cosmopolitan populations was comparable to that between two sub-Saharan populations (fig. S11), D_{XY} (the absolute nucleotide divergence) (*41*) between CN and other cosmopolitan populations was much lower than that between two sub-Saharan populations (fig. S12). Since F_{ST} is sensitive to within-population diversity, while D_{XY} is less affected, these data support a unique origin of the cosmopolitan populations. Furthermore, both F_{ST} and D_{XY} analyses revealed modest differentiation among the three CN subpopulations (CN_QTP, CN_XJ, and CN_Other).

The phylogenetic relationships between *D. melanogaster* populations

To decipher the phylogenetic relationships of these populations, we used TreeMix (*42*) to construct maximum likelihood (ML) trees while accounting for gene flow. Through testing multiple migration edges (*m* ranging from 0 to 10), we determined that the optimal *m* value was 1 for autosomes (2R and 3L) and 2 for the X chromosome (figs. S13 and S14). When the KF population from Zambia (*27*) was used as the root, the phylogenetic trees revealed a distinct separation between sub-Saharan and cosmopolitan populations (Fig. 2A and figs. S14 and S15A). Specifically, in the phylogenetic tree based on X-linked neutral SNPs (Fig. 2A), populations from southern and eastern Africa (KF, ZI, SD, SP, and RG) were situated near the root

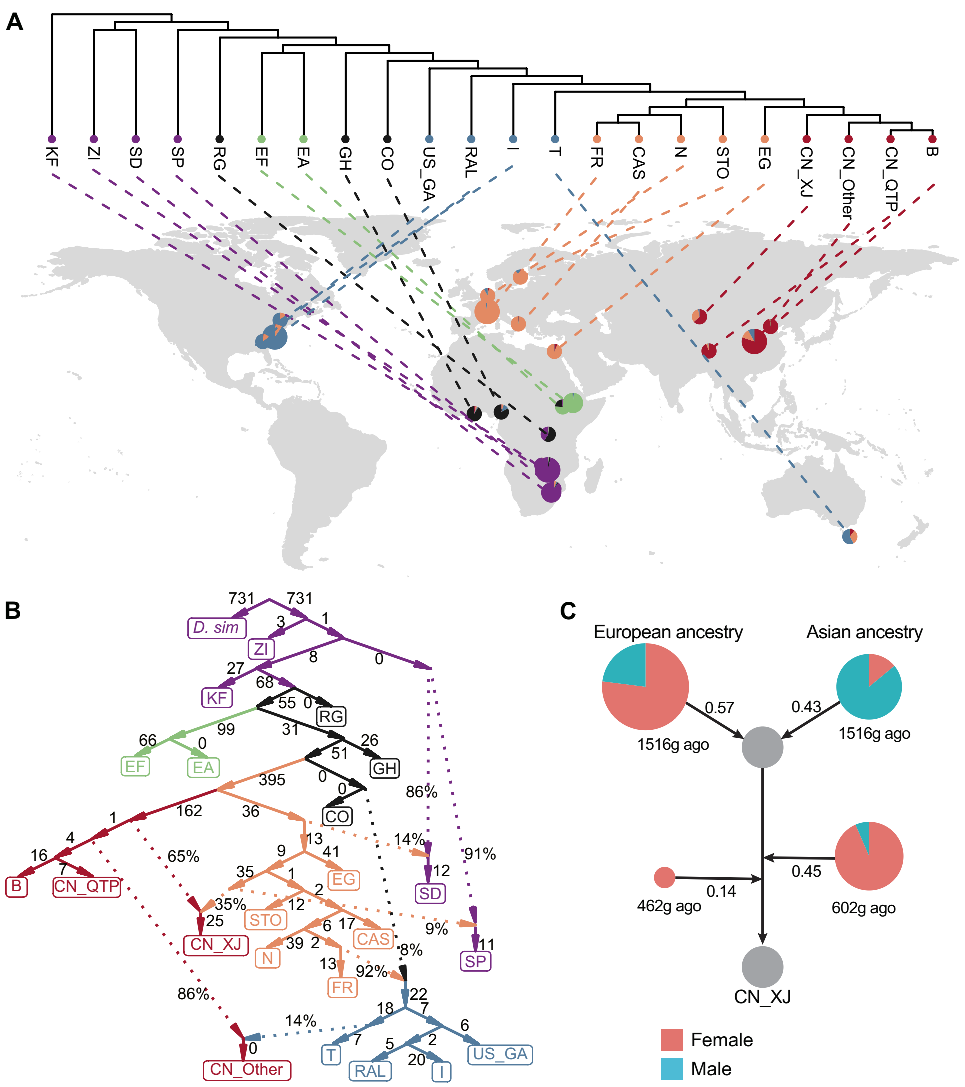


Fig. 2. Phylogeny and gene flow of worldwide *D. melanogaster* populations. (**A**) Maximum-likelihood tree for global populations inferred by TreeMix and projected onto the geographic map. The tree was inferred based on neutral SNPs of the X chromosome, assuming two migration edges. The six major ancestries were coded by different colors, and the average ancestry composition based on the X chromosome by ADMIXTURE for each population was plotted as a pie chart. (**B**) Admixture graph estimated by ADMIXTOOLS2 according to the likelihood score based on neutral SNPs of the X chromosome. Population split is represented by solid lines, and branch lengths are provided in units of 10 × *f*2 drift distance. The inferred admixture events are shown by dotted lines with proportions annotated. The results for the autosomes are shown in fig. S15. (**C**) Admixture history of CN_XJ inferred using MultiWaverX based on all inversion-free chromosomes. CN_QTP and FR were used to represent Asian and European ancestries, respectively. The proportion, time point (in generation), and sex ratio of each gene flow event are provided.

of the tree, followed by populations from Ethiopia (EF and EA) and western Africa (GH and CO). The cosmopolitan populations grouped together and were more closely related to the populations from western Africa than to other sub-Saharan populations. Strains from North America and Oceania (RAL, I, US_GA, and T) are more closely related to sub-Saharan populations. Populations in these two regions are known to be established through secondary contact of European strains and sub-Saharan strains (*21*, *43*, *44*). Strains in Europe (FR, N, CAS, and STO) and strains in China (CN_XJ, CN_QTP, CN_Other, and B) were clustered, respectively. Within the latter cluster, CN_QTP first clustered with B and then with CN_Other, and CN_XJ was more basal than the other three populations, presumably due to admixture with the strains in Europe (Fig. 2A). The phylogenetic tree based on autosomal sites was comparable to the X-linked tree except for a few topological deviations (Fig. 2A and figs. S14 and S15A). Notably, in the autosomal tree, CN_QTP first clustered with B and then with CN_XJ, and then with CN_Other. The difference between X-linked and autosomal topologies among the Chinese strains could potentially stem from differential gene flow across distinct chromosomes, as explored in more detail below.

Genetic admixture among populations

Admixture frequently occurred during *D. melanogaster* evolution (*6*, *7*, *20*, *21*, *43*, *44*). We used ADMIXTOOLS2 (*45*) to build an admixture graph of the populations and estimated the proportion of admixture among them (Fig. 2B and fig. S15B). Consistent with the TreeMix results, the best-fitting graphical models indicated that the strains in southern Africa represented the ancestral state, after which the flies moved to western Africa, eastern Africa, and Ethiopia, and then to northern Africa (Egypt), Europe, and Asia. Five introgression events were inferred from the X chromosome admixture graph. Specifically, two introgression events were found in southern African populations: (i) SD resulted from the admixture of southern African (86%) and European (14%) ancestries, and (ii) SP resulted from the admixture of southern African (91%) and European (9%) ancestries. One introgression event was found in strains in North America and Australia, which resulted from the admixture of western African (8%) and European (92%) ancestries, corroborating previous results (*20*, *21*, *44*). Another two events were inferred in strains in China: (i) CN_Other resulted from the admixture of Asian (86%) and American (14%) ancestries, and (ii) CN_XJ resulted from the admixture of Asian (65%) and European (35%) ancestries (Fig. 2B). These introgression events were corroborated by the admixture graph analysis of autosomes, with the exception of CN_XJ, possibly due to disparities in gene flow between autosomes and the X chromosome. Moreover, the autosomal analysis revealed that EA resulted from the admixture of Ethiopian (71%) and Asian (29%) ancestries (fig. S15B), providing additional evidence that introgression occurred from cosmopolitan populations back into sub-Saharan populations of *D. melanogaster* (*6*, *7*, *46*).

To capture more comprehensively the introgression events across these populations, we calculated Patterson's *D* (*47*) for the 1257 population trios that showed consistent TreeMix phylogenetic relationships for both the X chromosome and autosomes, using *D. simulans* as an outgroup. Under the criteria of $D > 0.05$, z score > 3, and false discovery rate (FDR) < 0.05, 141 trio tests exhibited gene flow in the X chromosome, and 64 trio tests indicated gene flow in the autosomes (table S7). Out of the 141 trio tests indicating gene flow in the X chromosome, 101 were in line with ADMIXTOOLS2 results. Specifically, 52 trios signaled gene flow between SD and cosmopolitan populations, five trios showed gene flow between the SP and cosmopolitan populations, and 29 trios indicated gene flow from Europe to populations in North America and Oceania. Moreover, 15 trio tests underscored gene flow for the X chromosome between CN_XJ and strains from Europe, northern Africa, or North America, corroborating the ADMIXTURE result (Fig. 1F). Within the 64 trio tests revealing gene flow in the autosomes, 27 were in agreement with ADMIXTOOLS2 outcomes (13 trios supported gene flow between SD and cosmopolitan populations, 10 trios supported gene flow between populations in Europe and recently colonized populations in North America and Oceania, and 4 trios supported gene flow between CN and EA). The remaining trio tests (40 for the X chromosome and 37 for autosomes) predominantly indicated gene flow between geographically adjacent populations within Africa (table S7).

Using Ancestry_HMM (*48*) and MultiwaverX (*49*), we further inferred three admixture events in the evolutionary history of CN_XJ (Fig. 2C). The first admixture wave, about 101 years ago, involved 43% Asian ancestry and 57% European ancestry, with female-dominant [female proportion: 77%, 95% confidence interval (CI): 73 to 81%] European contribution and male-dominant (male proportion: 86%, 95% CI: 80 to 91%) Asian contribution. The second wave, around 40 years ago, was female-dominant (female proportion: 93.5%, 95% CI: 87 to 99.3%) from Asian ancestry. The third wave, approximately 31 years ago, comprised an entirely female contribution from European ancestry. These findings suggest distinct autosomal and X-chromosomal admixture patterns in CN_XJ, possibly due to continuous female-biased introgression from European ancestry groups. Nonetheless, the possibility of selection favoring X-linked fragments introgressed from European strains cannot be ruled out. Our findings indicate that much like humans (*50*), Xinjiang serves as an east-west contact zone for *D. melanogaster*.

The global population migration and demography of *D. melanogaster*

The large-scale population genomic data enabled us to formulate a model of the possible range expansion and migration routes of *D. melanogaster*. The ancestral population of *D. melanogaster* evolved as a human commensal in the forests near Zambia (*27*). Among the populations we analyzed, the populations from southern Africa represented the ancient state, and the flies expanded to western and eastern Africa, Ethiopia, and then northern Africa (Egypt), followed by spreading into Europe and Asia (Fig. 3A). On the basis of the moments (*51*) analysis on neutral SNPs, we estimated that the ZI in sub-Saharan Africa diverged from the most recent common ancestor of the FR and CN about 9 (95% CI: 5.7 to 33.3) thousand years ago (ka) (Fig. 3, A and B), which is congruent with the previous estimation that the strains in Europe and sub-Saharan Africa separated 10 to 23 ka (*27*, *44*, *52*–*55*). We estimated that FR and CN diverged from the ancient cosmopolitan populations ~2.8 (95% CI: 0.36 to 9.85) to 4.4 (95% CI: 0.99 to 74.77) ka, and CN_QTP and CN_XJ diverged from CN_Other ~491 (95% CI: 34 to 893) and ~270 (95% CI: 71 to 1277) years ago, respectively (Fig. 3B, fig. S16, and table S8). The colonization of North America and Australia occurred very recently (Fig. 3A) (*20*, *54*, *56*).

We applied Stairway Plot 2 (*57*) to infer the historical effective population size (N_e) of a *D. melanogaster* population. On the basis of

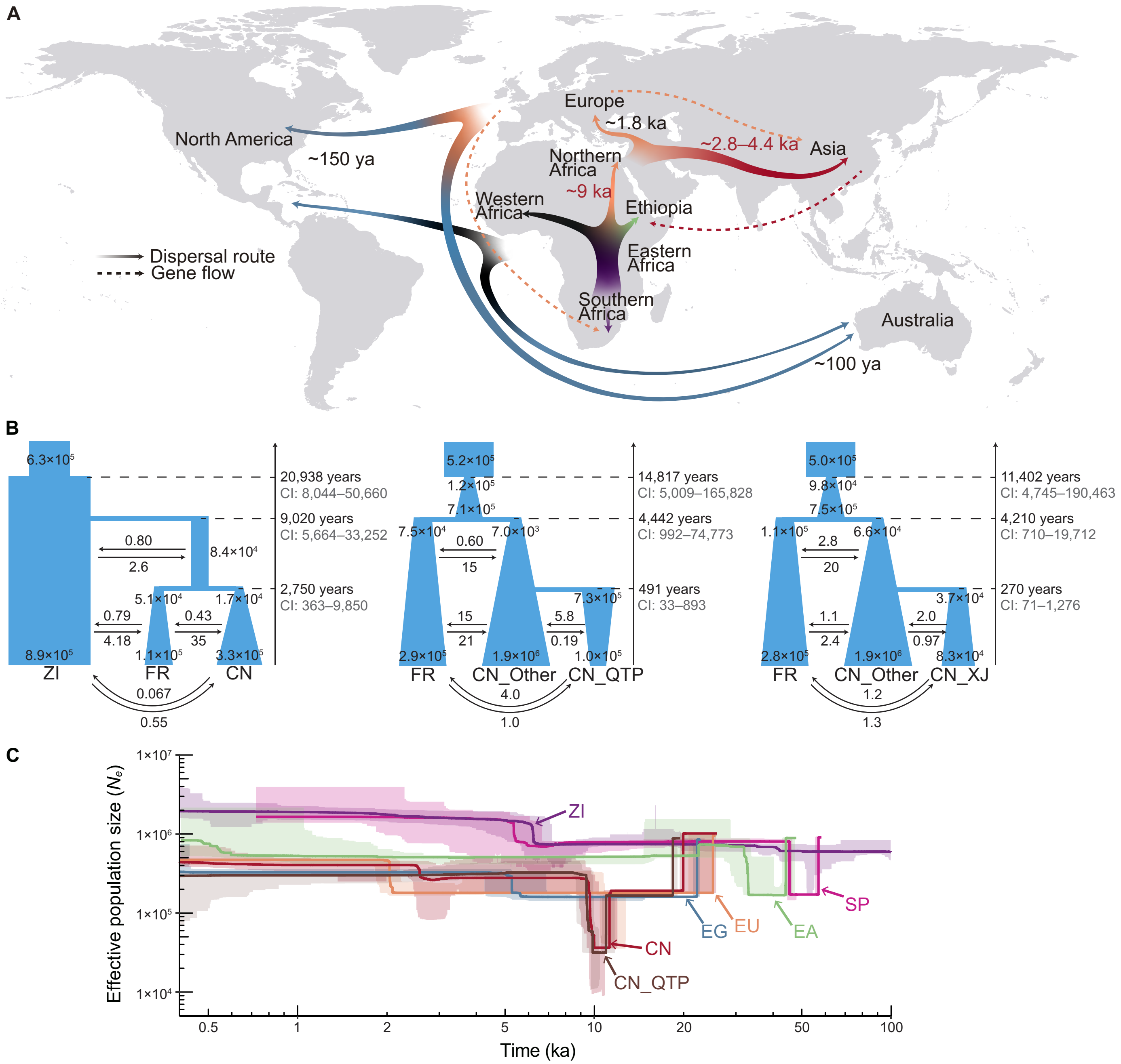


Fig. 3. Estimated evolutionary history of *D. melanogaster*. (**A**) Proposed range expansion and migration routes of *D. melanogaster*. Expansion routes are represented by solid arrows, and gene flow events are indicated as dashed lines. Time estimates were estimated from moments in this study (red) and previous studies (black). The out-of-Africa event occurred approximately 9 ka, followed by expansion to East Asia around 2.8 to 4.4 ka. *D. melanogaster* spread to Europe from the Middle East approximately 1.8 ka (*27*) and recently expanded to North America and Australia around 150 and 100 years ago, respectively (*56*, *85*). Introgressions from strains in Europe and Asia back into strains in southern Africa and Ethiopia are observed, and continuous gene flow occurs between strains in Europe and China. (**B**) Best-fitting evolutionary models involving the CN population inferred by moments based on neutral SNPs in inversion-free 2R and 3L. Parameters such as effective population size (N_e), split time, and migration strength ($2Nm$) are shown. (**C**) Inferred demographic history of seven representative populations using the unfolded SFS of neutral sites on inversion-free 2R, as analyzed by Stairway Plot 2. The EU population was created by merging CAS, FR, STO, and N. Historical effective population size changes from 0.4 to 100 ka are depicted. Solid lines represent the median of inferred population sizes, while light ribbons indicate the 95% CI. A mutation rate of 5.21×10^{-9} per site per generation and a generation time of 15 per year were used.

the analysis of chromosome 2R, the N_e of ZI grew dramatically at approximately 6.2 ka and has since climbed to 1.5×10^6. Ancient bottlenecks were detected in both SP and EA [45.4 to 56.9 (95% CI: 44.3 to 57.6) ka for the former and 32.4 to 44.1 (95% CI: 32.0 to 47.8) ka for the latter] (Fig. 3C). Nevertheless, caution is warranted when interpreting very ancient bottlenecks, as the Stairway Plot's resolution may be compromised for ancient histories, possibly resulting in artificial bottlenecks, especially in scenarios involving intricate ancient demographic history (*58*). The cosmopolitan populations [CN, EU (combination of CAS, FR, STO, and N), and EG] experienced a long bottleneck of 2 to 25 ka, which predated the divergence of the three populations and overlapped with the previous time estimate (10 to 26 ka) of the out-of-Africa event after the Last Glacial Maximum (*27*, *52*–*55*). The CN population showed a severe bottleneck [~9.4 to 11.2 (95% CI: 9.0 to 11.7) ka] during this long bottleneck (with N_e dropping to $\sim 4 \times 10^4$). These cosmopolitan populations expanded 2 to 5 ka, with extant N_e values of 4.4×10^5, 4.7×10^5, and 3.3×10^5 for CN, EU, and EG, respectively. The demographic history inferred from the CN_QTP strains resembled that of all CN strains, except that a population expansion around 2.5 ka was inferred using the entire CN dataset but not detected using the CN_QTP data alone (Fig. 3C). The analysis on chromosome 3L produced results akin to 2R, though with N_e and timing disparities (table S9 and fig. S17). For instance, on chromosome 3L, unlike 2R, the EU population experienced a bottleneck [~9.8 to 11.1 (95% CI: 9.3 to 11.8) ka], and CN_QTP underwent a recent population decline (~3.3 ka). These findings underscore the intricate demographic history across populations and chromosomes.

Natural selection and environmental adaptation in global populations

To detect adaptation signals, we performed Fay and Wu's H (H_{FW}) (*59*), Ohana (*60*, *61*), and Population Branch Excess statistic (PBE) (*62*) analyses on all SNPs in the six populations (CN, FR, RAL, EF, SD, and ZI) with over 50 sequenced genomes. These methods detect positive selection events from distinct angles. H_{FW} detects excessive high-frequency derived alleles within one or multiple populations. While Ohana and PBE both assess allele frequency differentiation among populations, they differ in crucial aspects. Ohana accounts for population stratification and admixture, focusing on SNP-level selection within specific lineages. In contrast, PBE excels in detecting recent population-specific selection by comparing allele frequencies across populations.

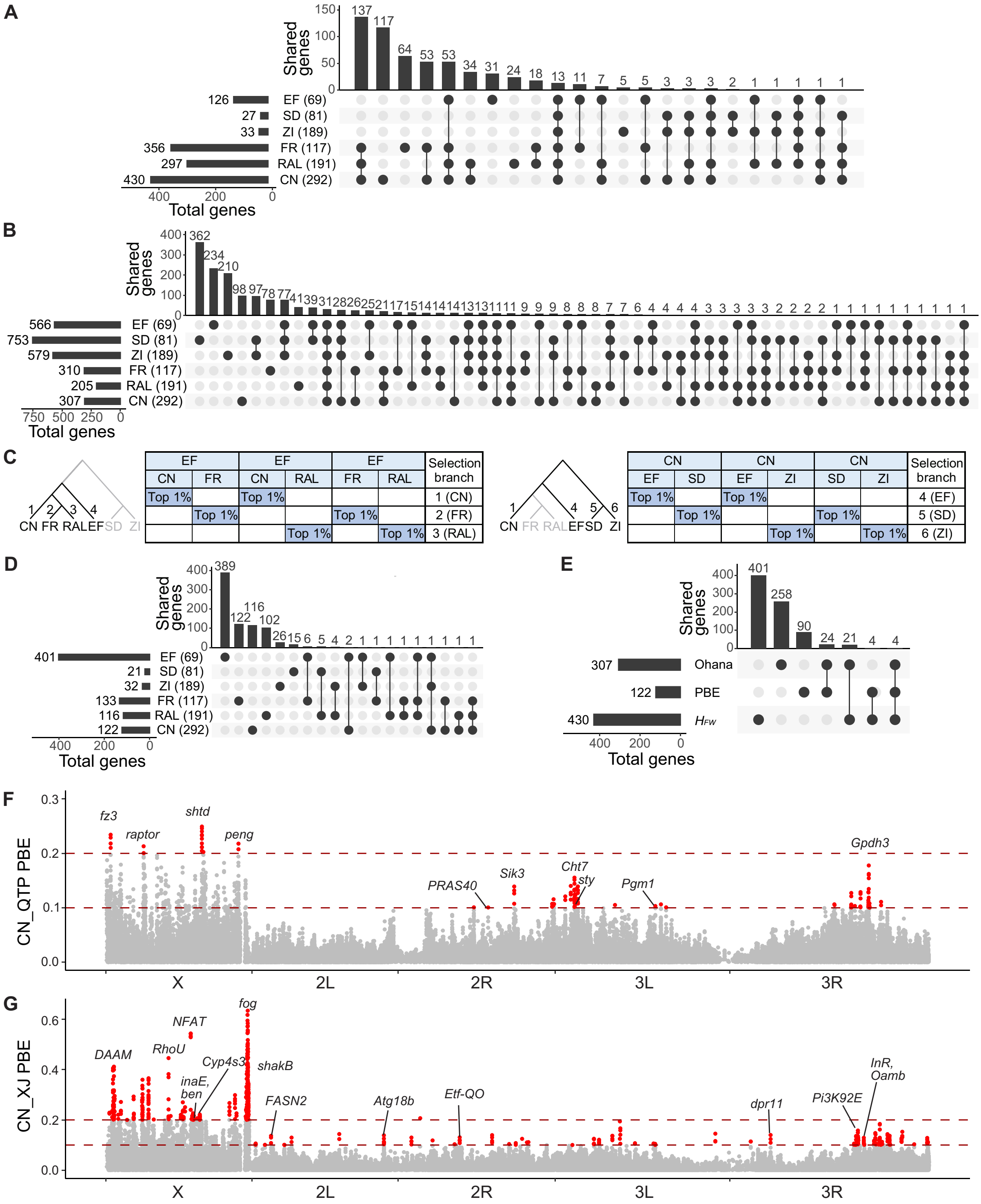
We divided the genome into 4-kb windows and subjected each window in every population to the H_{FW} test. Using a threshold of $H_{FW} < -0.4$ (typically below the lowest 1% of the lowest H_{FW} values; fig. S18), we identified 588 genes potentially under positive selection in at least one of the six populations (Fig. 4A and table S10). These genes displayed enrichments in pathways such as "insecticide catabolic process" and "response to DDT" (fig. S19). Among them, 550 genes were identified in at least one cosmopolitan population (430 in CN, 356 in FR, and 297 in RAL), 447 exclusively in the cosmopolitan populations, and 38 solely in the sub-Saharan populations (Fig. 4A).

We conducted Ohana analysis by assuming the global *D. melanogaster* consisting of seven ancestries ($K = 7$, as inferred from ADMIXTURE) (fig. S20). We identified 4654 X-linked and 24,250 autosomal SNPs potentially under positive selection by designating those within the top 1% of log-likelihood ratio scores (LLRSs). To attribute the selection signal to individual sites within a population, we required the selected allele's frequency to exceed 0.6 within a given ancestry and that this ancestry should have the highest proportion in the strains of that population. Using these criteria, we identified 1545 (RAL) to 8337 (SD) sites potentially under positive selection in a population (fig. S21). Generally, cosmopolitan populations (CN: 307, FR: 310, and RAL: 205) had fewer genes with at least two positively selected sites compared to sub-Saharan populations (EF: 566, SD: 753, and ZI: 579) (Fig. 4B). However, caution must be exercised when comparing the counts of positively selected genes between cosmopolitan and sub-Saharan populations. This is because Ohana may not accurately pinpoint the selection event on the genealogy (*60*), potentially leading to some positive selection in the ancestor of cosmopolitan populations being erroneously attributed to sub-Saharan populations. Overall, the Ohana analysis revealed limited overlap of positively selected sites (fig. S21) or genes (Fig. 4B) between populations. Genes under positive selection in at least one population were enriched in specific Gene Ontology (GO) terms such as "sensory perception of taste" or "detection of light stimulus involved in visual perception" (fig. S22), while those selected in a single population displayed enrichment in different GO terms, likely reflecting varied environmental selection pressures (fig. S23).

We conducted the PBE analysis based on the populations' phylogeny to identify selection signals exclusively along each terminal branch, using a previously described method (*63*). Specifically, we used the subtree ((CN, (FR, RAL)), EF) to identify branch-specific selection in CN, FR, or RAL (Fig. 4C). For instance, to identify selection signals in a 4-kb window exclusively present in the CN branch, we required the window to appear within the top 1% of highest PBE values in the CN branch for both ((CN, FR), EF) and ((CN, RAL), EF) trio tests. Furthermore, that window needed to lack selection signals (PBE within the top 1% highest PBE values) in both the FR and RAL branches during their trio tests. Similar tests were performed for the other two terminal branches within the subtree (Fig. 4C). Similarly, the subtree ((CN, EF), (SD, ZI)) was used to identify terminal branch–specific selection in populations ZI, SD, and EF (Fig. 4C).

The PBE analysis identified 796 genes potentially under positive selection in at least one of the six populations (Fig. 4D and table S10). These genes were enriched in pathways such as "sensory perception of touch" and "ribosomal small subunit assembly" (fig. S24). The analysis further revealed 122 potentially selected genes in CN, 133 in FR, and 116 in RAL. Among the sub-Saharan populations, EF displayed the most abundant selection signals (401 genes), likely due to its colonization of high-altitude regions. In contrast, branch-specific selection signals were sparse in SD (21 genes) and ZI (32 genes).

Given that these methods identified positive selection events spanning different time frames, there was limited gene overlap among them within a given population. For instance, in CN, a total of 802 genes displayed positive selection signals when combining results from all three approaches. Specifically, 401 genes were uniquely identified by H_{FW}, 258 genes solely by Ohana, and 90 genes exclusively by PBE. Only 49 genes were shared by two methods, and four genes (*Cyp6a17*, *shakB*, *Dora*, and *lncRNA:CR44997*) were identified by all three methods (Fig. 4E and table S10). Similar trends were observed in the FR (fig. S25), RAL (fig. S26), EF (fig. S27), SD (fig. S28), and ZI (fig. S29). Notably, GO analysis revealed significant enrichment of pathways such as insecticide catabolic process among the 802 genes in CN population ($P < 0.01$, fig. S30). For instance, *Ace* and



EF		EF		EF		Selection branch
CN	FR	CN	RAL	FR	RAL	
Top 1%		Top 1%				1 (CN)
	Top 1%			Top 1%		2 (FR)
			Top 1%		Top 1%	3 (RAL)

CN		CN		CN		Selection branch
EF	SD	EF	ZI	SD	ZI	
Top 1%		Top 1%				4 (EF)
	Top 1%			Top 1%		5 (SD)
			Top 1%		Top 1%	6 (ZI)

Fig. 4. Signals of positive selection detected by H_{FW}, Ohana, and PBE. (**A**) Overlap of positively selected genes detected by H_{FW} in the six major populations. The number of genomes used in the analysis for each population is indicated in parentheses. (**B**) Overlap of positively selected genes detected by Ohana in the six major populations. (**C**) Selection scan procedure and the phylogenetic tree used by PBE analysis. (**D**) Overlap of positively selected genes detected by PBE in the six major populations. (**E**) Overlap of positively selected genes detected by H_{FW}, Ohana, and PBE in CN. (**F** and **G**) Branch-specific positive selections in CN_QTP (F) and CN_XJ (G) detected by PBE. The PBE results are shown for each 4-kb sliding window with a step size of 1 kb in the trio (CN_QTP, CN_XJ, CN_Other). Windows with PBE > 0.2 on the X chromosome or PBE > 0.1 on the autosomes are colored in red. Genes located in outlier windows, potentially related to environmental adaptation and mentioned in the main text, are annotated.

Cyp6a23 were identified by both PBE and Ohana, *Cyp4d1* by both H_{FW} and Ohana, and *Cyp6a17* by all three methods (table S10). Moreover, *Cyp4d1* was identified by both H_{FW} and Ohana in FR (table S10). Overall, these findings suggest that diverse positive selection events have occurred in distinct populations during *D. melanogaster*'s evolution, underscoring the complexity of its history and environmental adaptation.

Population-specific selection among the populations in China

We conducted PBE analysis to identify population-specific positive selection within the three CN subpopulations (CN_QTP, CN_XJ, and CN_Other). Using arbitrary thresholds of PBE > 0.1 for autosomes and PBE > 0.2 for the X chromosome, we identified 146 putatively positively selected genes in CN_XJ, 54 genes in CN_QTP, and only 1 gene (*Myo81F*) in CN_Other (Fig. 4, F and G; fig. S31; and table S11).

In CN_XJ, the 146 genes under positive selection showed significant enrichment in the "response to oxidative stress" GO term ($P < 0.01$, fig. S32), including *Pi3K92E*, *InR*, *Oamb*, *inaE*, *ben*, and *Atg18b*. Furthermore, several identified genes were associated with stimulus responses, such as reactions to salt stress (*NFAT*, *dpr1*), light stimulus (*shakB*), chemical stimulus perception (*dpr11*, *dpr3*), and cellular response to sucrose stimulus (*FASN2*; fig. S33). Notably, a missense variant in *NFAT* (G585S) had a substantially higher frequency in CN_XJ than in other populations (fig. S34). These findings imply that the genes specifically selected in CN_XJ might be involved in adapting to environmental stresses.

In CN_QTP, the 54 genes potentially under positive selection were enriched in the GO term "carbohydrate metabolic process" (fig. S35), including genes like *Cht7*, *Gpdh3*, and *Pgm1*. Furthermore, three genes [*raptor*, *PRAS40*, and *Sik3* (fig. S36)] in the insulin signaling pathway showed positive selection specifically in CN_QTP. Notably, two missense mutations (S1493T and G945S) in *raptor* were more commonly in CN_QTP than in other populations (fig. S37). Additionally, the gene *sty*, associated with hypoxia response, also showed a selection signal in CN_QTP. Overall, these findings suggest that the genes uniquely selected in CN_QTP may play roles in metabolism and adaptation to hypoxia.

Moreover, we conducted Ohana analysis by assuming that the CN population consists of two ancestries ($K = 2$), as inferred from ADMIXTURE (fig. S38). With a threshold of LLRS > 20, we identified 239 genes potentially under positive selection. Among them, 11 genes overlap with those detected as positively selected by PBE (Table 2). However, due to the very recent differentiation among the three CN subpopulations and intricate gene flow between the CN_XJ and European strains, it is challenging to determine whether these events transpired before the divergence of the three populations or afterward.

Table 2. Genes with selection signals detected by both PBE and Ohana in CN_QTP and CN_XJ. Genes were ordered according to the highest PBE value of 4-kb windows overlapping with the gene.

Population	Gene	Description	Nonsynonymous SNPs	Synonymous SNPs	PBE
CN_QTP	*shtd*	Shattered, part of anaphase-promoting complex	0	42	0.249
CN_QTP	*CG11655*	Predicted to be involved in sodium-dependent organic anion transport	0	18	0.249
CN_QTP	*CG4666*	Uncharacterized protein, orthologous to human THEM6 (thioesterase superfamily member 6)	0	13	0.213
CN_QTP	*Tsp5D*	Tetraspanin 5D, predicted to be integral component of membrane	0	0	0.213
CN_QTP	*fz3*	Frizzled 3, involved in establishment or maintenance of cell polarity	0	0	0.210
CN_QTP	*raptor*	Part of TORC1 complex	2	13	0.201
CN_QTP	*Mipp2*	Multiple inositol polyphosphate phosphatase 2	0	2	0.200
CN_XJ	*lncRNA:CR44997*	Long noncoding RNA:CR44997	0	0	0.230
CN_XJ	*FASN2*	Fatty acid synthase 2	0	24	0.137
CN_XJ	*dpr3*	Defective proboscis extension response 3	0	1	0.101
CN_XJ	*Msp300*	Muscle-specific protein 300 kDa	0	2	0.101

Signals of natural selection on insecticide resistance–related genes

Insecticide resistance is commonly observed in *D. melanogaster* (*13*, *16*, *64*–*66*). We identified signals of positive selection in 26 genes or gene clusters that were associated with insecticide resistance (Fig. 5A). Notably, eight genes showed signatures of positive selection unique to a single population, including *Cyp28c1* in CN, *para* in FR, *Cyp310a1* in ZI, two genes (*Cyp4ac3, Cyp4c3*) in SD, and three genes (*mtt*, *Cyp313a1*, *Cyp313b1*) in EF. The *Cyp6a22-Cyp6a8* cluster exhibited positive selection signatures in both CN and RAL, and *GstE10* showed positive selection signatures in both CN and FR. Three gene regions (*Cyp318a1*, *Cyp308a1*, and *Cyp6v1*) showed signatures of positive selection only in all three cosmopolitan populations (CN, FR, and RAL). Four regions (*CHKov1*, *Cyp6t1*, *Mdr65*, and the *GstD10-GstD9* cluster) showed positive selection signals in all three cosmopolitan populations (CN, FR, and RAL), as well as the EF population. The *Cyp6g1-Cyp6t3* cluster showed a positive selection signal in cosmopolitan populations and SD. Two genes (*Ace* and *Cyp4d1*) showed positive selection signals in all six populations. We observed higher frequencies of three missense variants in *Ace* (T3I, G303A, and F368Y) within the CN populations compared to other populations, with similar variant frequencies across CN_QTP, CN_XJ, and CN_Other (fig. S39). In summary, we observed a similar count of insecticide resistance–related gene regions displaying positive selection in cosmopolitan populations (CN: 14, FR: 15, and RAL: 11) as compared to EF (14 regions), with these counts surpassing those in SD (8 regions) and ZI (6 regions).

Of note, the *Cyp6a22-Cyp6a8* cluster encompasses eight potential insecticide-resistant genes, including *Cyp6a17* and *Cyp6a23*, which have been linked to resistance against pyrethroid insecticides like deltamethrin and permethrin (*64*). Previous studies identified two structural variants of *Cyp6a17* in *D. melanogaster* (Fig. 5B and fig. S40). One variant is a *Cyp6a17 deletion* allele, which disrupts this gene (*67*). The second variant is a fusion allele, *Cyp6a17/23,* formed by a deletion spanning the 3′ portion of *Cyp6a17* and the 5′ portion of *Cyp6a23*, resulting in an in-frame fusion gene (*22*, *64*, *67*). We assessed the frequencies of these two structural variants in each of the six populations using next-generation sequencing (NGS) reads and validated them using Sanger sequencing whenever required. The *Cyp6a17 deletion* allele was primarily detected in SD, ZI, and RAL populations, with a frequency of 16.2%, 14.4%, and 19.3%, respectively. Besides the *Cyp6a17/23* fusion allele originally detected in the *Drosophila melanogaster* Genetic Reference Panel (DGRP) dataset (*64*) (denoted as "*fusion 1*"), we uncovered another *Cyp6a17/23* allele (called "*fusion 2*") with nine SNPs different from *fusion 1* (including five synonymous and four nonsynonymous SNPs; fig. S40). The frequency of the *Cyp6a17/23* alleles varied across populations, with *fusion 2* as the dominant allele in CN (~74.8%) and FR (~57.5%). Overall, the cosmopolitan populations exhibited higher frequencies of the *Cyp6a17/23* alleles, while the sub-Saharan populations had more intact alleles (Fig. 5B). The findings revealed the complicated evolutionary history of the *Cyp6a22-Cyp6a8* cluster.

Soft sweeps on the insecticide resistance variants

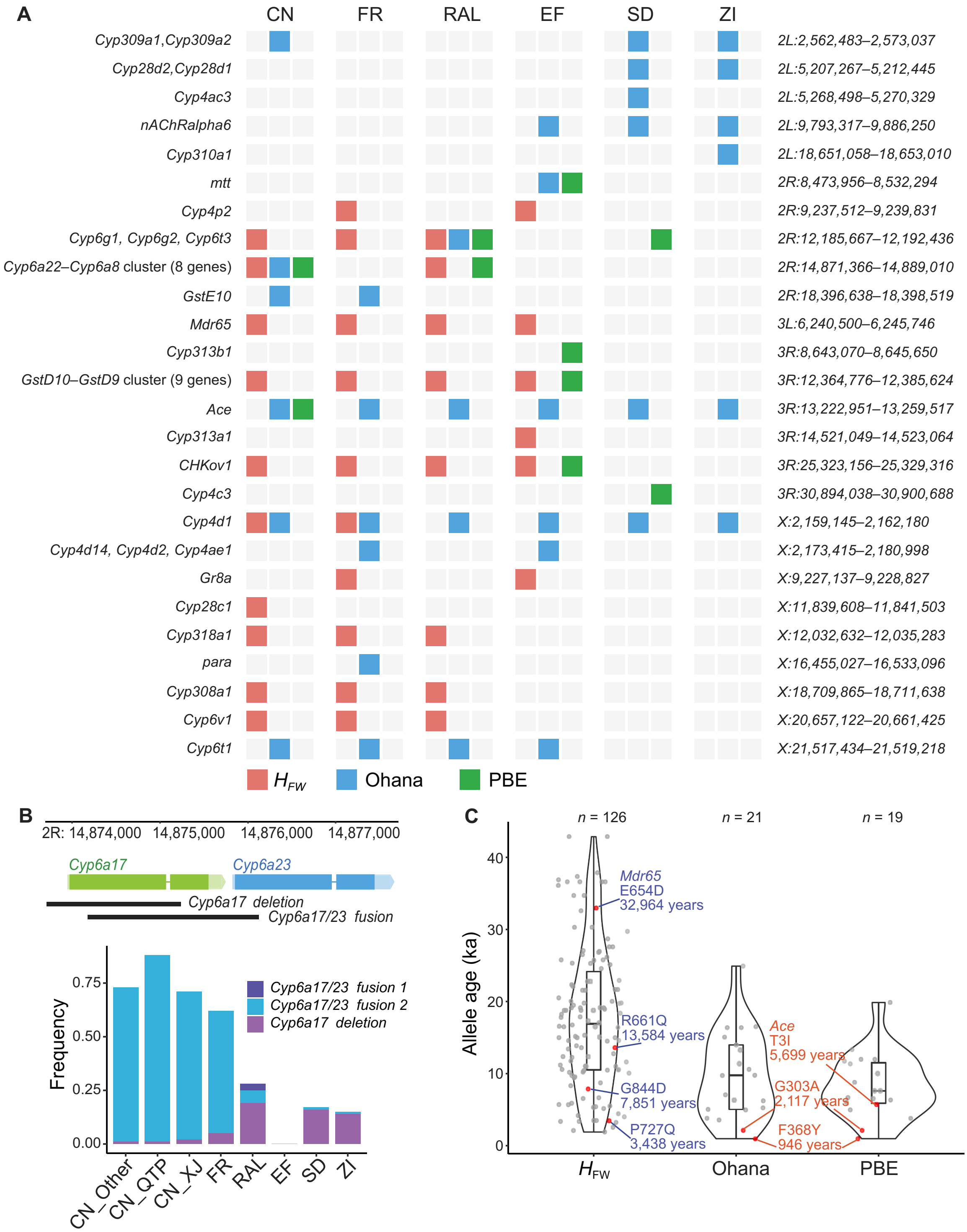
To estimate the age of insecticide resistance variants, we performed Relate (*68*) analysis, which simultaneously reconstructs the local tree near the target SNP and infers historical allele trajectories.

Considering 15 generations per year, the SNPs linked to insecticide resistance and displaying positive selection in the CN population (H_{FW}: 126 SNPs, Ohana: 21 SNPs, and PBE: 19 SNPs) exhibited ages spanning 946 to 42,872 years, with an average of 15,946 and a median of 14,581 years (Fig. 5C). Notably, two *Ace* missense variants (G303A and F368Y), known to affect insecticide resistance (*66*), exhibited notably higher frequencies in CN (0.70 and 0.58, respectively) compared to other populations (e.g., in FR, the corresponding frequencies were 0.16 and 0.07, respectively; fig. S39). G303A and F368Y were estimated to be 2117 and 946 years old, respectively. The allele age of another *Ace* missense variant (T3I, frequency: 0.92 in CN, 0.51 in FR) was approximately 5699 years. Four missense mutations (P727Q, G844D, R661Q, and E654D) in *Mdr65* arose approximately 3438 to 32,964 years ago. The 21 SNPs that showed signals of positive selection in the *Cyp6a22-Cyp6a8* gene cluster in the CN population were estimated to have arisen 4944 to 37,341 years ago. Overall, the insecticide resistance variants appear substantially older than the widespread adoption of synthetic pesticides (initiating around 1940), suggesting the role of soft sweeps on standing genetic variations in shaping the evolution of insecticide resistance (*65*, *69*).

Verifying insecticide resistance variants by combining population cage testing and deep sequencing

To verify the effects of the insecticide resistance variants and to identify other variants associated with insecticide resistance in natural strains in China, we pooled 2000 male adults from 200 strains collected from 20 geographic locations in China (10 males from each strain) into a population cage and treated the flies with lambda-cyhalothrin, one of the most commonly used pyrethroid insecticides. We sequenced the genomes of the flies that survived after 48 hours of treatment using the Illumina platform to a depth of ~100× for each replicate (101 and 36 individuals survived and were sequenced in the two biological replicates, respectively). In parallel, we pooled and sequenced the 200 strains (five males for each strain) to estimate the background frequency of each variant (two replicates were sequenced, each to a depth of ~400×). We compared the frequency of each variant in the insecticide-resistant versus the background sample to identify the variants that were associated with insecticide resistance. We tabulated the NGS reads that encompassed both the reference and alternative alleles in these samples for each SNP. Subsequently, we used the Cochran-Mantel-Haenszel (CMH) test to determine the likelihood of association between the variant and insecticide resistance. We selected a threshold of $P < 1 \times 10^{-9}$, which is more stringent than the common P value threshold of $P < 5 \times 10^{-8}$ used in human genetic genome-wide association studies (GWASs) (*70*), resulting in an approximate equivalent FDR of 1×10^{-7}. We identified 7202 candidate SNPs (out of 1,028,409 tested SNPs) under this criterion (fig. S41) and computed the allele frequency difference for each SNP between background samples and those resistant to lambda-cyhalothrin (Fig. 6A).

The associated SNPs were located in 2320 genes, and 164 of these genes overlapped with genes under positive selection in the CN population as detected by H_{FW}, Ohana, or PBE analysis (Fig. 6B and table S10). Notably, significantly associated SNPs were identified in five genes (*Cyp6a20*, *Cyp6a8*, *Cyp6a21*, *Cyp6a19*, and *Cyp6a9*) in the *Cyp6a22-Cyp6a8* cluster (Fig. 6A). Moreover, the *Cyp6a17/23 fusion 2* showed significantly higher frequencies in the insecticide-resistant samples than in the background samples (0.86 versus 0.75,



	CN	FR	RAL	EF	SD	ZI	
Cyp309a1,Cyp309a2	⬜🟦⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦⬜	⬜🟦⬜	2L:2,562,483–2,573,037
Cyp28d2,Cyp28d1	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦⬜	⬜🟦⬜	2L:5,207,267–5,212,445
Cyp4ac3	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦⬜	⬜⬜⬜	2L:5,268,498–5,270,329
nAChRalpha6	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	2L:9,793,317–9,886,250
Cyp310a1	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦⬜	2L:18,651,058–18,653,010
mtt	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜🟦🟩	⬜⬜⬜	⬜⬜⬜	2R:8,473,956–8,532,294
Cyp4p2	⬜⬜⬜	🟥⬜⬜	⬜⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	2R:9,237,512–9,239,831
Cyp6g1, Cyp6g2, Cyp6t3	🟥⬜⬜	🟥⬜⬜	🟥🟦🟩	⬜⬜⬜	⬜⬜🟩	⬜⬜⬜	2R:12,185,667–12,192,436
Cyp6a22–Cyp6a8 cluster (8 genes)	🟥🟦🟩	⬜⬜⬜	🟥⬜🟩	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	2R:14,871,366–14,889,010
GstE10	⬜🟦⬜	⬜🟦⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	2R:18,396,638–18,398,519
Mdr65	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	3L:6,240,500–6,245,746
Cyp313b1	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜🟩	⬜⬜⬜	⬜⬜⬜	3R:8,643,070–8,645,650
GstD10–GstD9 cluster (9 genes)	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	🟥⬜🟩	⬜⬜⬜	⬜⬜⬜	3R:12,364,776–12,385,624
Ace	⬜🟦🟩	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	3R:13,222,951–13,259,517
Cyp313a1	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	3R:14,521,049–14,523,064
CHKov1	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	🟥⬜🟩	⬜⬜⬜	⬜⬜⬜	3R:25,323,156–25,329,316
Cyp4c3	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜🟩	⬜⬜⬜	3R:30,894,038–30,900,688
Cyp4d1	🟥🟦⬜	🟥🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	X:2,159,145–2,162,180
Cyp4d14, Cyp4d2, Cyp4ae1	⬜⬜⬜	⬜🟦⬜	⬜⬜⬜	⬜🟦⬜	⬜⬜⬜	⬜⬜⬜	X:2,173,415–2,180,998
Gr8a	⬜⬜⬜	🟥⬜⬜	⬜⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	X:9,227,137–9,228,827
Cyp28c1	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	X:11,839,608–11,841,503
Cyp318a1	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	X:12,032,632–12,035,283
para	⬜⬜⬜	⬜🟦⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	X:16,455,027–16,533,096
Cyp308a1	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	X:18,709,865–18,711,638
Cyp6v1	🟥⬜⬜	🟥⬜⬜	🟥⬜⬜	⬜⬜⬜	⬜⬜⬜	⬜⬜⬜	X:20,657,122–20,661,425
Cyp6t1	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜🟦⬜	⬜⬜⬜	⬜⬜⬜	X:21,517,434–21,519,218

🟥 H_{FW} 🟦 Ohana 🟩 PBE

Fig. 5. Positive selection signals and allele age of genes related to insecticide resistance. (**A**) Selection signals detected by H_{FW}, Ohana, and PBE for insecticide resistance–related genes or gene clusters in six major populations. (**B**) Structural variations at *Cyp6a17* and *Cyp6a23*, including a *Cyp6a17* deletion and two *Cyp6a17/23* fusions (for details, see fig. S40). The frequency of deletion and fusion alleles varied among the six major populations, with *Cyp6a17/23 fusion 2* as the dominant allele in CN and FR. (**C**) Ages of insecticide resistance–related mutations potentially under positive selection. The age of each mutation was calculated as the mean of the lower and upper ends of the branch length estimated by Relate. The ages of three missense mutations in *Ace* (G303A, F368Y, T3I) and four missense mutations in *Mdr65* (P727Q, G844D, R661Q, E654D) are highlighted (red dots), as they are under positive selection in CN.

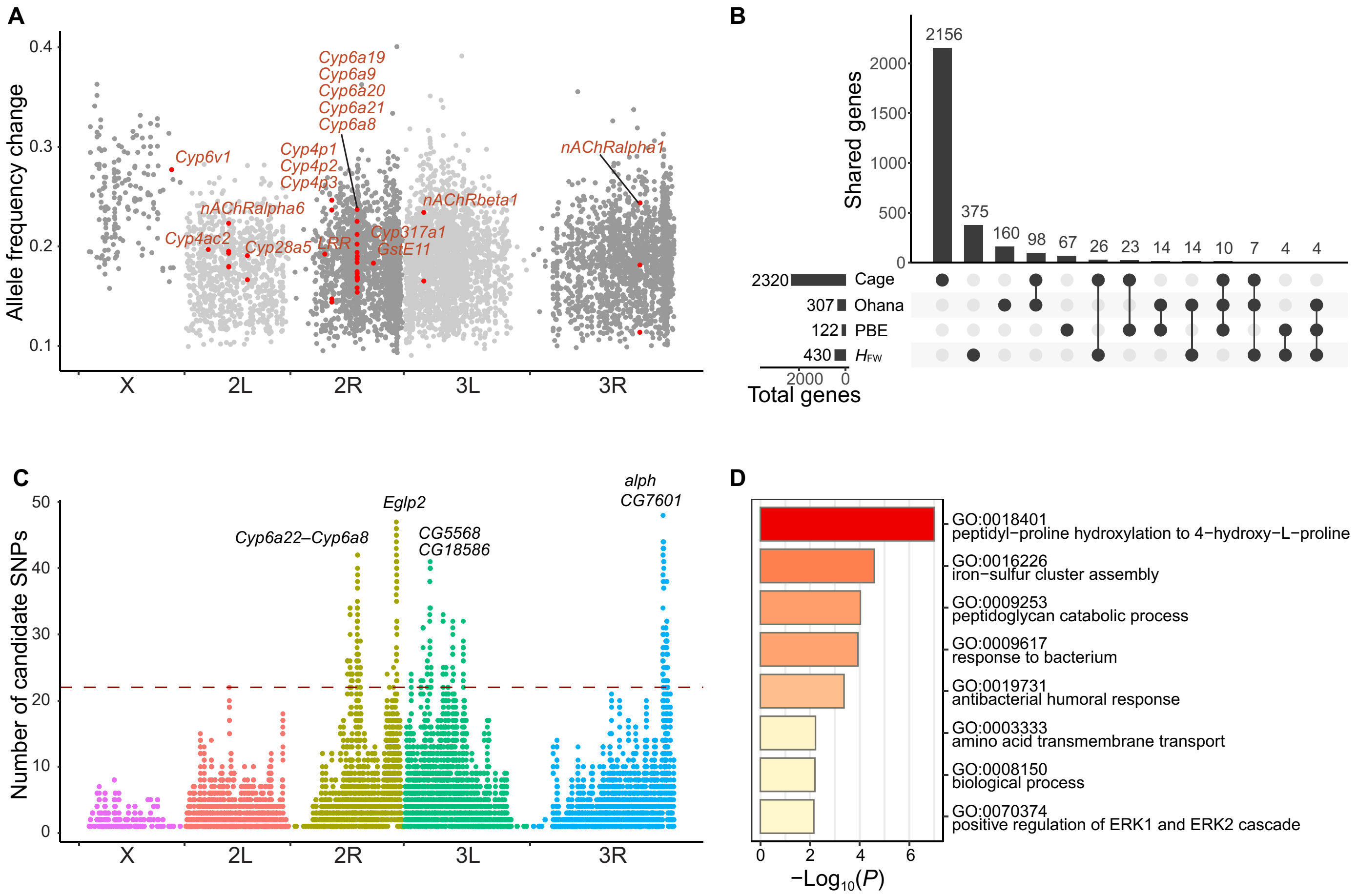


Fig. 6. Genes associated with lambda-cyhalothrin resistance in the population cage experiments. (**A**) Allele frequency difference of SNPs associated with lambda-cyhalothrin resistance (FDR < 1×10^{-7}) between the resistant and background samples. Genes previously known to be involved in insecticide resistance are highlighted as red dots. (**B**) Overlap of genes under positive selection and genes associated with lambda-cyhalothrin resistance. (**C**) Number of candidate SNPs associated with resistance to lambda-cyhalothrin in 50-kb sliding windows with a step of 5 kb. Windows with more than 22 candidate SNPs and *z* score > 6 are thought enriched with candidate SNPs. (**D**) Functional enrichment analysis of genes in windows enriched with candidate SNPs (*n* = 1234 SNPs). Gowinda was used for GO analysis, and GO BP terms with *P* < 0.01 and at least three candidate genes are shown.

$P = 2.76 \times 10^{-6}$; CMH test), providing further evidence that this fusion allele contributes to insecticide resistance (*64*). Thus, our association tests based on the population cage experiment and deep sequencing corroborated the population genetic analysis result that the selection signals in the *Cyp6a22-Cyp6a8* cluster were caused by the usage of the insecticide. The associated SNPs were also identified in other insecticide resistance genes, such as *Cyp4ac2*, *Cyp28a5*, *Cyp4p2*, *Cyp4p1*, *Cyp4p3*, *Cyp317a1*, *Cyp6v1*, *GstE11*, *LRR*, and nicotinic acetylcholine receptors such as *nAChRalpha1*, *nAChRalpha6*, and *nAChRbeta1* (Fig. 6A). These genes were previously reported to be associated with resistance to multiple insecticides (*64*, *71*).

To mitigate false positives stemming from linkage disequilibrium (LD) or other potential confounding factors, we focused on genetic regions enriched with candidate SNPs. We consolidated signals from adjacent SNPs and tallied the count of candidate SNPs within a sliding window of 50 kb, using a step size of 5 kb. We conducted 1000 permutations by shuffling the coordinates of candidate SNPs, thereby establishing the distribution of candidate SNP counts in 50-kb windows. According to this permutation analysis, windows with a *z* score larger than 6 and housing over 22 candidate SNPs were regarded as significantly enriched for these candidates (details in Materials and Methods and fig. S42). Through the merging of overlapping windows, we identified 25 genomic regions demonstrating an excess of SNPs compared to expectation (Fig. 6C). In total, these regions encompassed 1234 candidate SNPs located within 224 genes (table S12). Eleven of these candidate genes were found to overlap with genes undergoing positive selection in the CN population, as identified through H_{FW}, Ohana, or PBE analysis. These genes include *Egfr* (epidermal growth factor receptor), *Pcf11* (involved in regulation of RNA splicing), *Kank* (involved in negative regulation of actin filament polymerization), *CG7888* (involved in amino acid transmembrane transport), *CG30287* (involved in proteolysis), *CG33506* (integral component of membrane), *CG43693* (involved in amino acid transmembrane transport), *ABCB7* (involved in cellular iron ion homeostasis), *CG13917* (unknown function), *CG32264* (actin binding activity), and *aPKC* (involved in morphogenesis of an epithelium).

The region encompassing the *Cyp6a22-Cyp6a8* gene cluster (2R:14,790,001–15,040,000) emerged as one of the most enriched regions for candidate SNPs. Furthermore, our data revealed numerous previously unreported candidate genes within these regions (Table 3). For instance, the region 3R:30,466,050–30,648,551, which is enriched with candidate SNPs, featured 10 genes (*PH4alphaEFB*, *PH4alphaSG2*, *PH4alphaNE1*, *PH4alphaNE2*, *CG31021*, *PH4alphaPV*, *CG31013*, *CG31016*, *PH4alphaNE3*, and *CG34041*) that are enriched in the "peptidyl-proline hydroxylation to 4-hydroxy-L-proline" GO term (Fig. 6D), which may be involved in stress response (*72*). Another candidate region contains *CG5568* and *CG18586*, both of which are predicted to have long-chain fatty acid–CoA (coenzyme A) ligase activity. *CG5568* has four potential missense variants, while *CG18586* has two detected missense variants. Given that cuticle composition affects insecticide resistance by limiting insecticide penetration (*73*), these two genes might be linked to lambda-cyhalothrin resistance through the hydrocarbon biosynthetic pathway, potentially altering long-chain fatty acyl–CoA biosynthesis (*74*). Several other genes, such as *Eglp2* (involved in renal system processes), *shep* (associated with neuron remodeling and metamorphosis), *CG7601* (with nicotinamide adenine dinucleotide–retinol dehydrogenase activity), *alph* (involved in response to oxidative stress and paraquat), *CG9747* (related to lipid metabolic processes), and *CG15533* (participating in sphingomyelin catabolic processes), are within windows enriched for candidate SNPs. These genes could indirectly influence insect response to insecticides by affecting various biological processes, such as responding to reactive oxygen species (ROS) triggered by insecticides (*75*) or driving developmental changes in the nervous system, tracheal system, and muscles. Further research is required to elucidate their roles in insecticide resistance.

DISCUSSION

Here, we sequenced 292 *D. melanogaster* strains from natural populations in China and examined them with more than 1000 genomes from previous studies. This extensive dataset enabled us to obtain a global perspective on the evolution and adaptation of *D. melanogaster*, and this work enhances our comprehension of how the population history and natural selection shape the genetic diversity of this model organism. Additionally, the incorporation of these strains and the genomic dataset presents prospects for further investigation into numerous fundamental questions in evolutionary biology, such as the genetic architecture of complex traits.

The influence of variant calling strategies on the detection of polymorphisms

In previous *Drosophila* population genomic studies, variant calling methods varied, including joint variant calling by JGIL (*14*) or GATK HaplotypeCaller and GenotypeGVCFs (*16*), as well as individual variant calling using GATK UnifiedGenotyper (*6*, *9*). These different approaches can introduce disparities in identified variant sites due to variations in sensitivity and specificity, potentially affecting downstream analyses. For instance, while individual calling might result in high false positives for singleton detection, joint calling could lead to a biased false-negative rate in singleton detection (*76*). Here, we combined our generated data with previously sequenced *D. melanogaster* genomes and used the joint calling method using freebayes software (*77*) for SNP calling. Heterozygosity was called with a minor allele threshold greater than 25% for each strain, aligning with the default setting of the GATK pipeline. The joint calling approach we used could potentially lead to the omission of certain singletons or low-frequency SNPs, and this omission might subsequently diminish growth signals during demographic inference and result in an underestimation of the proportion of deleterious variants in DFE-alpha. Nonetheless, it is important to note that despite these considerations, the joint calling method is generally recommended for variant calling (*78*) due to its capacity to distinguish reference homozygous sites from those with missing data.

Table 3. Examples of genes associated with resistance to lambda-cyhalothrin revealed by the population cage experiments. The gene ID, number of candidate SNPs, number of missense candidate SNPs, description, the maximum allele frequency change (df), and FDR of the most significantly associated SNPs in a gene in the population cage experiments were shown.

Gene/cluster	Total SNPs	Missense SNPs	Description	df	FDR
Cyp6a22-Cyp6a8	5	4	Insecticide metabolic process	0.23	6.55×10^{-18}
Eglp2	7	0	Water channel activity	0.25	5.73×10^{-21}
shep	12	0	Response to stimulus	0.23	3.11×10^{-15}
CG5568	9	4	Long-chain fatty acid metabolic process	0.35	1.36×10^{-13}
CG18586	9	2	Long-chain fatty acid metabolic process	0.31	3.56×10^{-18}
CG43897	5	0	Mesoderm development	0.25	1.65×10^{-15}
alph	10	0	Response to oxidative stress	0.28	6.30×10^{-17}
CG7601	5	1	Putative oxidoreductase	0.30	6.81×10^{-19}
CG9747	4	0	Lipid metabolic process	0.21	3.18×10^{-19}
CG15533	7	0	Sphingomyelin catabolic process	0.17	4.42×10^{-17}

Furthermore, this approach enhances sensitivity for low-frequency variants (excluding singletons) and improves false-positive filtering by capitalizing on information from all samples despite its computationally intensive nature.

Modest population stratification observed in the CN strains of *D. melanogaster*

Our analysis identified the presence of six distinct major genetic ancestry groups across global populations of *D. melanogaster*. Strains within China (CN and B) constituted a distinct group despite exhibiting detectable levels of admixture. While there was slight differentiation within the CN strains, this differentiation varied across different chromosomes, indicating a complex interplay of multiple forces shaping their genetic makeup. On the basis of the PCA and ADMIXTURE results, we categorized these strains into three subpopulations: CN_QTP, CN_XJ, and CN_Other. It is important to note that the strains in the CN_Other category come from diverse geographic locations. Two lines of evidence suggest the existence of population structure within CN_Other: first, strains in CN_Other exhibited varying proportions of ancestries (Fig. 1, E and F, and figs. S8 and S9); second, the first two principal components of each strain in CN_Other correlated with latitude or longitude, with a generally higher correlation observed with longitude (fig. S43). This observed structure within CN_Other might be attributed to multiple factors, such as uneven gene flow from Europe or North America, undetected inversions, and local adaptation. To assess this stratification within CN_Other, we compared the cumulative variance distribution from PCA across various datasets, including global strains, cosmopolitan strains, CN (comprising CN_Other, CN_QTP, and CN_XJ), CN_Other, and RAL datasets (fig. S44). While CN demonstrated a relatively steep cumulative distribution curve for PCA explained variance, indicating mild population structure within CN, the distribution of cumulative variance in CN_Other closely resembled that of the RAL dataset, which originates from a single location and lacks notable population structure (*14*).

To evaluate the possible effect of population structure on demographic inference, we performed comparative demographic analysis across the Lhasa population (CN_LS) and broader Chinese datasets (CN and CN_QTP) using Stairway Plot 2. The CN_LS data comprises 57 strains exclusively from Lhasa and minimizes the effect of population structure. Our analysis reveals consistent historical patterns, confirming a severe bottleneck across all datasets with minor differences in the timing (fig. S45). For instance, the analysis of SNPs in 2R suggests the 95% CIs for the bottleneck ranged from 9.0 to 11.7 ka for CN, 9.0 to 12.0 ka for CN_QTP, and 7.2 to 10.1 ka for CN_LS. A similar trend is observed for 3L, with the bottleneck ranging from 8.2 to 10.7 ka for CN, 7.5 to 10.4 ka for CN_QTP, and 6.4 to 8.8 ka for CN_LS. These overlapping timelines indicate no substantial statistical discrepancy among the datasets despite small variances, possibly due to minor population structure or the variability in sample sizes used in the analysis. Complementary moments analysis with the strains in Lhasa (CN_LS) aligns closely with the divergence times obtained for CN_QTP and the overall strains in China (CN), with the CN_LS and FR split time estimated at 2155 years ago (95% CI: 311 to 3791) (fig. S46A) and the CN_QTP and FR divergence at 2239 years ago (95% CI: 1140 to 7781) (fig. S46B), both consistent with the CN and FR split time (Fig. 3B). In summary, the analysis of CN_LS data shows that the population structure within *D. melanogaster* strains in China does not substantially bias our demographic inferences. Nonetheless, we recognize that population structure can influence demographic analysis, and additional sequencing data in future studies could help clarify any remaining uncertainties. Furthermore, pooling strains across different geological locations may reduce the detectability of positive selection, as suggested by the reduced number of selected genes in CN_Other versus CN_XJ or CN_QTP; however, the full impact of this aggregation on identifying signals of local adaptation across different environments warrants further investigation.

Temporal dynamics of the CN strains of *D. melanogaster* and the possible influencing factors

Our study fills the gap in the evolutionary history of *D. melanogaster* in East Asia. Our moments results showed that the ancestor of FR and CN split from sub-Saharan populations around 9 ka, and CN and FR diverged about 2.8 to 4.4 ka. This timeline broadly aligns with findings from another study that used fastsimcoal2 to infer a recent out-of-Africa expansion, with the estimated Saharan crossing occurring around 10 ka (*27*). We observed an extended duration (~6000 years) during which the ancestor of CN and FR persisted after branching from sub-Saharan populations, preceding the eventual split between CN and FR. This span coincides with the emergence and flourishing of agriculture and winemaking in the Middle East. It is plausible that the ancestral cosmopolitan populations remained in the Middle East before further dispersing into Europe and East Asia. Within the CN strains, our moments analysis indicated recent expansions (200 to 500 years ago) in the CN_XJ and CN_QTP subpopulations. This finding corresponds to previous evidence suggesting the ongoing expansion of *D. melanogaster* over the past several hundred years. For example, the colonization of North America and Australia occurred within the past 200 years (*56*), and the split among populations within Europe occurred several centuries ago (average divergence for western-eastern: 1013 years ago, western-western: 648 years ago, eastern-eastern: 294 years ago) (*11*). We offer a possible explanation for these recent expansions, considering the challenging cold and dry climate in Xinjiang and the Qinghai-Tibet Plateau, which affects both *D. melanogaster* and human populations. As *D. melanogaster* expansion is closely tied to human activities, initial colonization might have been hindered until specific conditions were met, including (i) increased human migration, (ii) growth in local human populations, and (iii) adaptation to the local environment.

Extensive gene flow between *D. melanogaster* populations revealed by multi-methods analysis

Our analysis, which used a variety of methods, reinforced the presence of extensive introgression between *D. melanogaster* populations. It is noteworthy that our TreeMix analysis might have potentially underestimated migration, as the chosen optimal *m* by OptM accounted for approximately 98.6% and 99.4% of the variation for autosomes and the X chromosome, respectively (fig. S13), slightly below the 99.8% anticipated from the original paper's simulations (*42*). Conversely, ADMIXTOOLS2 and Dsuite revealed more complex patterns of gene flow. These findings align well with previously reported secondary contact history of populations in North America and Australia, and cosmopolitan admixture in Southern Africa (SD and SP) from Europe (*6*, *7*). Our investigation highlights extensive admixture related to strains from Asia. For instance, cosmopolitan admixture identified in Ethiopia lowland (EA) likely originated from ancient Asian strains, potentially facilitated by maritime routes. A previous study reconstructed a two-pulse admixture model of EA using FR as an introgression donor (*46*), and further research using strains from Asia as the source of introgression will be needed.

We also observed an unexpected increase in X-linked gene flow from European strains to CN_XJ. This is surprising given that previous studies have typically reported lower levels of introgression on the X chromosome (*7*, *20*, *21*, *43*). The heightened X-linked admixture might arise from female-biased gene flow from European strains, although the precise mechanism remains elusive. Alternatively, this pattern could be influenced by selective processes favoring the incorporation of X-linked fragments from Europe into CN_XJ. Further research holds the potential to elucidate the connection between adaptation and introgression.

The potential influence of chromosomal inversions on gene flow

Chromosomal inversions are associated with reduced recombination and substantially affect polymorphism patterns and population stratification (fig. S6). In *D. melanogaster*, inversions are commonly observed, with their frequencies varying among different populations. Nevertheless, the extent to which these inversions affect gene flow patterns remains uncertain. When strains with autosomal inversions were included in Patterson's D calculations, we detected gene flow meeting the criteria (z score > 3 and FDR < 0.05) in 605 of 1044 trios within the autosomes. Conversely, excluding strains with inversions led to 463 of 1257 trios exhibiting gene flow. Moreover, applying an additional criterion ($D > 0.05$) resulted in a similar pattern (90 of 1044 with inversions versus 64 of 1257 without inversions). These data suggest that more gene flow signals were detected when considering strains with inversions. These observations do not align with the conventional belief that chromosomal inversions are associated with reduced introgression rates (*79*). Several explanations might account for this phenomenon: (i) gene flow involving chromosomal inversions might genuinely occur neutrally or under selection and (ii) the substantial LD introduced by ancient inversion polymorphisms might violate Patterson's D assumptions (*47*), potentially leading to misidentification of gene flow between populations with similar inversion frequencies. Here, we chose to exclude inversions when calculating Patterson's D to mitigate possible biases stemming from them. However, it is important to note that our approach might result in an underestimation of gene flow across populations. As a result, further research or methodological advancements are imperative to comprehensively elucidate the interplay between chromosomal inversions and gene flow among populations.

Evidence for faster-X evolution in *D. melanogaster* populations

The "faster-X evolution" theory (*80*) posits that both positive and purifying selection operate more efficiently on the X chromosome compared to autosomes in species with a heterogametic sex determination system, such as *Drosophila*. Supporting this theory, our DFE-alpha analysis indicated that approximately 66.9% and 71.8% of the polymorphic nonsynonymous mutations were strongly deleterious in autosomes and the X chromosome, respectively (fig. S2). This pattern supports the notion that purifying selection is more effective at removing harmful mutations from the X chromosome compared to the autosomes. Furthermore, our study revealed stronger adaptive signals on the X chromosome than on the autosomes. For instance, the H_{FW} tests showed that windows in the X chromosome overall exhibited lower H_{FW} values than autosomes across all populations (fig. S47). Additionally, the PBE values tended to be higher for the X chromosome compared to the autosomes in the cosmopolitan populations (fig. S48). These data overall support the faster-X evolution effect that positive selection is more efficient on the X chromosome than autosomes. We also found lower genetic diversities on the X chromosome than on autosomes in cosmopolitan populations, even after adjusting for differences in effective population size. This may indicate stronger selective sweeps on the X chromosome driven by adaptation to new environments. However, at this moment, we cannot exclude the possibility that this pattern was influenced by bottleneck effects following the migration out of Africa.

The effect of low-recombination regions on genetic adaptation of *D. melanogaster*

The Hill-Robertson effect posits that natural selection is less effective in genomic regions with low recombination rates (*81*). To investigate this theory using our data, we compared selective pressures in autosomal regions with high recombination (recombination rate > 0.5 cM/Mb; 10,813 genes) and low recombination (recombination rate < 0.5 cM/Mb; 3718 genes). To mitigate the faster-X effect, we only focused on the autosomes in this analysis. We found that the cosmopolitan populations generally exhibited more selection signals in high-recombination regions compared to low-recombination regions (fig. S49), supporting the Hill-Robertson effect. Specifically, the H_{FW} test revealed a significantly higher proportion of genes (1.48%, 160 of 10,813 genes) in high-recombination regions than in low-recombination regions (0.699%, 26 of 3718 genes) if all populations were considered ($P = 0.00018$, Fisher's exact test). This pattern was similarly observed in the H_{FW} test for CN strains (0.795%, 86 of 10,813 genes in high-recombination regions versus 0.350%, 13 of 3289 genes in low-recombination regions; $P = 0.0036$, Fisher's exact test). The Ohana analysis also indicated that LLRS scores tended to be higher in high-recombination regions compared to low-recombination regions (fig. S50), suggesting that positive selection signals were more abundant in high-recombination regions. Furthermore, the PBE analysis in CN strains revealed a significantly higher proportion of genes (0.916%, 99 of 10,813 genes) in high-recombination regions than in low-recombination regions (0.296%, 11 of 3718 genes; $P = 6.38 \times 10^{-5}$, Fisher's exact test). Additionally, high-recombination regions showed significantly higher genetic diversity (π) than low-recombination regions ($P < 0.01$, Wilcoxon test, fig. S51), likely due to the effects of hitchhiking or background selection (*82*). We observed that absolute nucleotide divergence (D_{XY}) was higher in high-recombination regions than in low-recombination regions (fig. S52), which might result from a faster accumulation of divergence in beneficial variants or a reduced number of neutral mutations removed by background selection in the high-recombination regions.

Notably, distinct patterns emerged between recombination and adaptation signals in sub-Saharan populations, with H_{FW} values tending to be lower in low-recombination regions than in high-recombination regions (fig. S49). This suggests possible differences in selective pressures between cosmopolitan and sub-Saharan populations. Cosmopolitan populations may face higher selective pressures as they adapt to new environments after migrating out of Africa, while diversity in sub-Saharan populations is likely influenced by selection effects at linked sites, especially background selection (*83*). In line with this notion, we found more abundant adaptive selection signals in cosmopolitan populations compared to sub-Saharan populations. In summary, these findings underscore the role of recombination in shaping genetic adaptation in cosmopolitan populations during their adaptation to new environments and provide insights into population-specific adaptive evolution through differences in selection signals between low- and high-recombination regions.

Effect of low-recombination regions on demographic analysis

By examining whether to incorporate low-recombination regions or not in our demographic analysis, we found that this factor has a minimal impact on our ability to infer population structure, gene flow patterns, and demographic history. Here, we primarily focused on autosomes for this analysis to avoid other factors that may confound the comparison on the X chromosome.

Regarding population structures, we observed that whether we included or excluded low-recombination regions, it did not notably alter the overall patterns of PCA (Fig. 1C and figs. S7 and S53) and ADMIXTURE (Fig. 1E and figs. S8 and S54). In both cases, the six major ancestral components remained consistent in ADMIXTURE, although the optimal value of K for 3L varied ($K = 8$ when including low-recombination regions; $K = 7$ when excluding them). In terms of tree topology and gene flow patterns, the results from the TreeMix analysis were consistent for autosomes, regardless of whether we included or excluded low-recombination regions (figs. S14A and S55). The ADMIXTOOLS2 analysis showed similar results when including or excluding low-recombination regions. The only difference is that when excluding low-recombination regions, the gene flow from North America into CN_Other was not detected, but another gene flow from cosmopolitan populations into GH was detected (figs. S15B and S56).

For demographic history, the moments results obtained from data with or without low-recombination regions revealed similar evolutionary models, albeit with slight differences in estimations of split time, effective population size, and migration strength. For example, the split time estimations between CN and FR were consistent between the analyses with or without low-recombination regions [2750 (95% CI: 363 to 9850) versus 2335 (95% CI: 266 to 3874) years ago]. Moreover, when analyzing the dataset without low-recombination regions, we estimated that CN_QTP and CN_XJ diverged from CN_Other approximately 244 (95% CI: 46 to 1095) and 181 (95% CI: 70 to 2095) years ago, respectively (fig. S57). These estimates did not notably differ from those based on the dataset that included low-recombination regions [~491 (95% CI: 34 to 893) years ago for CN_QTP and ~270 (95% CI: 71 to 1277) years ago for CN_XJ]. In the case of Stairway Plot 2, while the overall population size trends were very similar between datasets that included or excluded low-recombination regions (fig. S58), there were slight variations in the time estimations for population size changes. For instance, the most severe bottleneck in the CN population on 2R was inferred to be approximately 9.4 to 11.2 (95% CI: 9.0 to 11.7) ka using data that included low-recombination regions, which was similar to the estimate of ~10.8 to 12.9 (95% CI: 10.0 to 14.0) ka obtained when excluding these regions (fig. S58 and Fig. 3C). These minor differences in time estimations may be due to differences in genetic diversity and divergence across regions with differing recombination rates.

Soft sweep contributes to the adaptation to insecticide resistance

Here, we identified numerous signals of positive selection in insecticide-resistant genes. Despite *D. melanogaster* not being the primary target of insecticides, our findings emphasize the potential of fruit flies as ecological indicators for assessing the impact of chemical exposures. Furthermore, this adaptation to insecticides offers an excellent opportunity to explore the tempo and mechanisms of adaptation. The relative contributions of hard sweeps and soft sweeps to adaptation remain subjects of debate in evolutionary biology (*69*, *84*). In the hard sweep model, a new advantageous mutation arises and quickly spreads to fixation due to natural selection. In contrast, the soft selective sweep model involves selection acting on either standing genetic variation or recurrent de novo mutations, and adaptive alleles can be found on multiple haplotypes. Previous studies have uncovered soft sweeps on the *Ace* gene, primarily driven by the recurrent emergence of the I199V mutation in different populations (*65*, *69*). Our Relate analysis also supported the independent emergence of the I199V mutation in multiple phylogenetic lineages of *D. melanogaster*. Furthermore, we found the G303A and F368Y variants in *Ace*, which are associated with resistance to organophosphate insecticides (*66*), exhibited considerably higher frequencies in CN compared to other populations (fig. S39). Our Relate analysis, based on variation in the CN strains, estimated the ages of these mutations to be 2117 and 946 years old, respectively. These ages predate the use of insecticides (which began around 1940), suggesting adaptation through standing genetic variation. We also estimated the allele ages of other insecticide-associated SNPs in CN strains, and the results consistently suggest a scenario of soft selective sweep on standing genetic variation. Consequently, our study provides evidence that soft sweeps on standing genetic variation and recurrent de novo mutations have both played crucial roles in the rapid adaptation to insecticide resistance.

Evaluation of the association study by population cage

Our population cage experiment combined with insecticide resistance sorting and deep sequencing confirmed that variants in the *Cyp6a22-Cyp6a8* cluster and other insecticide resistance genes are associated with lambda-cyhalothrin resistance. Furthermore, we also identified several previously unreported candidate genes, which may contribute to insecticide resistance through multiple mechanisms, such as changes in epithelium, sensory organs, and development. Our method can be essentially used to dissect architectures of other phenotypes, since it is cost-effective and provides a homogeneous environment for different strains to be screened in the same environment and excludes the possible confounding effects on testing environments, which often confounded the GWAS tests by assaying individual strains. However, two caveats should be kept in mind with the population cage experimental approach.

First, as selection was done in one generation and no recombination occurred, SNPs linked with the causal SNPs may also show an association with the phenotype, generating false positives. To cope with the linkage, we combined the candidate SNPs in sliding windows with the assumption that background SNPs that are more adjacent to causal SNPs are more likely to be linked and identified as candidates, forming a cluster of SNPs surrounding the real signals. We tested windows of different lengths and found that 50-kb windows best revealed the enrichment of candidate SNPs, and no improvements were obtained by further increasing the window size (Fig. 6C and fig. S59). This approach effectively identified 25 regions with enriched association signals, which is consistent with the common architecture of insecticide resistance, where a few loci with large effects contribute to the phenotype. Second, population structure in strains from different geographic regions is another confounding factor. If local adaptation occurs in the phenotypes of interest, an association can be detected between the phenotypes and the local genetic components. In our study, the population structure is generally very weak in CN population, especially for autosomes (Fig. 1C and fig. S6C). The differentiation on the X chromosome will have very limited influence because the male sequencing strategy leads to lower depth and lower statistical power in detecting associations in the X chromosome, and most of the candidate SNPs are located in autosomes. Besides, the LD decays quickly (down to $r^2 < 0.2$) within 1 kb in the CN population (fig. S60), suggesting a well-mixed genetic background. Although this window-based approach effectively reduces false positives, identifying the true targets within these windows can still be difficult. Another noteworthy issue is that there might be some other traits associated with the lambda-cyhalothrin resistance, due to either the co-occurrence of selective pressures in nature or the experimental design of mixing individuals from structured populations. As a result, the candidate SNPs identified in the cage experiment may contribute to other phenotypes instead of lambda-cyhalothrin resistance. Further studies are required for functional validation of these candidate variants.

Overall, the population cage experiment, coupled with phenotype sorting and deep sequencing, is a promising method for identifying the candidate variants that are associated with insecticide resistance. We expected that it would improve the cage experiment with more replications and more individuals in pooled samples in future studies.

For most of the history of population genetics, *D. melanogaster* has been a widely used model organism for analyzing how evolutionary forces of mutation, drift, and selection shape patterns of variation in the genome. The perspective we want to emphasize is that this same species can also be used as a sensitive measure of the impact of environmental changes (*10*, *85*). This is particularly important given the current deterioration of ecosystems, mediated by both climate change and other human activities. Since *D. melanogaster* is a human commensal, it is anticipated that incorporating additional genome sequences of *D. melanogaster* sampled from other geographic locales will not only advance our knowledge of fundamental issues related to environmental adaptation but also provide insights into human evolution and migration.

In summary, our study offers insights into the evolutionary history and adaptive potential of *D. melanogaster* in China and globally. It demonstrates the power of population genomic analysis in deciphering the genetic underpinnings of environmental adaptation and gene introgression and provides a valuable resource for understanding the global evolution and adaptation of this important model organism.

MATERIALS AND METHODS

Sample collection

Flies were collected in orchards and parks from 52 cities throughout China in the summers of 2017 and 2020 (table S1). We used naturally rotten fruits or set up traps made of fruit (bananas, grapes, or other native fruits) and baker's yeast to attract flies and caught flies with a sweep net. Every female individual was transferred into a new vial and kept alone to establish an iso-female strain. For the 292 sequenced strains, 261 of these strains were collected in 2017 and maintained in vials with an N_e of 50 to 100 for 3 years (about 45 generations) before sequencing, while the other 31 strains (mel259 to mel292) were collected in 2020 and maintained in vials with an N_e of 50 to 100 for 6 months (about 7.5 generations) before sequencing. *D. melanogaster* was distinguished from *D. simulans* and other *Drosophila* species using the morphology of the male offspring of each iso-female strain. All strains were kept in vials containing standard cornmeal medium at 25°C with a 12:12-hour light/dark cycle.

Genome sequencing and variant calling

Genomic sequence data for the CN population were collected from 292 wild-derived iso-female strains from 52 cities (19 provinces) in China, including 25 strains from Xinjiang (CN_XJ), 63 strains from the Qinghai-Tibet Plateau (CN_QTP), and 204 strains from other parts of China (CN_Other). The genomic DNA for each strain was extracted from 20 to 25 males with a TIANamp Genomic DNA Kit (DP304). DNA library preparation and genomic sequencing were performed by Annoroad Gene Technology (Beijing, China) according to the DNA Sample Preparation Guide (Illumina, 15026486 Rev.C) and with the Illumina NovaSeq 6000 platform as paired-end fragments with a read length of 150 base pairs (bp). On average, 5 Gb of clean data for each strain was obtained, resulting in an average sequencing depth of 26.2×.

We also included 1147 genomes derived from global populations previously sequenced in multiple projects (*5*–*9*, *14*–*16*, *20*, *26*–*31*). To integrate these genomes, we downloaded the sequencing reads in Sequence Read Archive (SRA) format from the National Center for Biotechnology Information (NCBI)–SRA database and extracted fastq files using SRAtoolkit v2.11.0. Read quality control was performed by fastp v0.19.5 (*86*). A read was removed if it contained more than five undetermined (N) bases, was shorter than 50 bp, or had a quality score below 20 for more than 50% of its bases. Low-quality bases (mean quality below 20 in a 4-bp sliding window) were trimmed from both ends. After quality control, reads were mapped to the *D. melanogaster* reference genome (dmel_r6.24, FlyBase) by BWA-MEM v0.7.15 (*87*) using the default parameters. Optical duplicates were discarded by Picard v2.20.5. Variants were jointly called by freebayes v0.9.21 (*77*) with a minimum mapping quality of 20, a minimum base quality of 20, a haplotype length of 0, and no more than 4 best alleles (-m 20 -q 20 --haplotype-length 0 --use-best-n-alleles 4), while other parameters remained the default. Strains are considered heterozygous at an SNP if the fraction of minor allele is larger than 25%. This threshold is commonly applied in *Drosophila* population genomics studies (*6*, *16*). Complex mutations and multiple-nucleotide polymorphisms (MNPs) were decomposed by vcflib v1.0.0 (*88*).

For each strain, the genotype was considered missing if the depth of this site was below 4. Samples that showed a low sequencing depth or mapping rate were excluded from further analysis. Genomes with a genotype missing rate exceeding 20% or identified as PCA outliers in relation to all other genomes based on either autosomes or the X chromosome were also removed from further analysis. All 16 genomes from Winters, CA (*29*) were excluded due to their high missing rate or PCA outlier status (the PCA is described in detail below). Finally, the global cohort in our study consisted of 1356 genomes, including 14 populations with sample sizes ranging from

10 to 50 and 6 major populations with more than 50 genomes. For information on sequencing and sample origin, please see table S1. To obtain high-quality variants, the multisample VCF file was filtered to remove (i) variants with a quality below 20 and alternate allele counts below five, (ii) SNPs within 3 bp of indels or other complex mutations, (iii) SNPs falling in the repetitive or low-complexity regions annotated by RepeatMasker (*89*) or UCSC Genome Browser (http://genome.ucsc.edu/), and (iv) SNPs with a missing genotype rate higher than 10%. These filters are used for all analyses.

Inferring the functional consequences of mutations

Functional consequences of SNPs were annotated by SnpEff (*90*) using the BDGP6.86 database. We applied DFE-alpha v2.163 (*32*) to estimate the distribution of fitness effects (DFEs) for deleterious mutations occurring in different functional categories of SNPs in autosomes and the X chromosome separately in CN. A randomly sampled allele for each strain was extracted and used to generate the site frequency spectrum. The folded site frequency spectra were generated by the Spectrum.from_data_dict function in ∂a∂i v2.1.1 (*91*) for seven categories, including zerofold degenerate, 3′ untranslated region (UTR), 5′ UTR, intergenic, long intronic, short intronic, and fourfold degenerate SNPs. Using fourfold degenerate sites as the neutral reference, we ran a two-epoch model in DFE-alpha to estimate the DFEs for each category of SNPs, allowing variable mutation-effect sizes and shape parameters for the gamma distribution.

Heterozygous block identification

Heterozygous blocks were identified using the method outlined in the previous study (*6*). We computed the proportion of heterozygous sites within windows of 100-kb length and 5-kb steps for each genome, based on all SNPs. This proportion was then compared with the genetic diversity (π) within the corresponding population. The weighted average of π for each population was computed using pixy v1.2.4.beta1 (*92*) for all SNPs in the population. The process involved initiating heterozygosity blocks when the ratio of heterozygous proportion to diversity exceeded 0.2, and these blocks were extended in both directions until the ratio dropped below 0.05. The average total length of heterozygous blocks was calculated for each population, and the details of heterozygous block tracks for each strain were provided in table S2. It is important to note that when calculating allele frequencies, homozygous sites were counted twice for the respective allele, while heterozygous sites were counted once for each of their two different alleles. We excluded the strains sequenced using haploid embryos since they are not expected to have genuine heterozygosity (*6*, *8*).

Ancestral state inference

The ancestral state of each SNP was inferred by est-sfs (*93*) using two outgroup species, *D. simulans* and *D. yakuba*. Homologous sequences of *D. melanogaster*, *D. simulans* (dsim_r2.02, FlyBase), and *D. yakuba* (dyak3, UCSC Genome Browser) were obtained by LiftOver based on reciprocal-best chain files from UCSC Genome Browser. The probabilities of the inferred ancestral state were calculated using the R6 model in est-sfs. Only sites with ancestral state probabilities greater than 0.9 were used for further analysis.

Inversion genotyping

The frequencies of seven common inversion polymorphisms, *In(2L)t*, *In(2R)NS*, *In(3L)P*, *In(3R)C*, *In(3R)K*, *In(3R)Mo*, and *In(3R)Payne*, were characterized based on a previously published set of diagnostic SNP markers that were in tight LD with the corresponding inversions (*35*). For each strain, we inferred the genotype of each inversion by calculating the mean frequencies of inversion-specific SNP markers. Strains were regarded as heterozygous if the frequency was between 0.1 and 0.9, while strains with a frequency smaller than 0.1 or larger than 0.9 were considered homozygous for standard chromosomes or inverted chromosomes, respectively. The inversion genotypes of each CN strain and the inversion frequencies of global populations are given in table S5. The inversion genotypes of other populations can be found in the *Drosophila* Genome Nexus (*8*). The inversion genotype of *In(3L)Ok,* which was exclusively found within African populations, was also obtained from the *Drosophila* Genome Nexus (*8*).

Measuring genetic diversity and population differentiation

Genetic diversity (π) for each of the 22 populations listed in Table 1 was computed using pixy v1.2.4.beta1 (*92*). This calculation was based on all SNPs and all strains within each respective population. The calculation was based on 200-kb nonoverlapping windows. The weighted average of π was calculated as the total number of differences divided by the total number of comparisons. To compare π between autosomes and the X chromosome, the genetic diversity of all autosomes (2L and 2R, 3L and 3R, and 4) was combined as π_A, and π_X was corrected by multiplying by 4/3 (π_{XC}) to account for the difference in N_e. We also calculated π for different genomic functional categories, including 5′ UTR, 3′ UTR, missense variant, and synonymous variant, using a pipeline similar to that for all sites.

Genetic differentiation among populations was quantified using the fixation index (F_{ST}) and the absolute divergence (D_{XY}). The Weir and Cockerham F_{ST} and D_{XY} based on all SNPs were calculated by VCFtools (*94*) and pixy (*92*) for pairwise comparisons between the 22 populations. These calculations were performed separately for autosomes and the X chromosome. To test for the isolation by distance pattern, linear regression was carried out by using the weighted averages of F_{ST} or D_{XY} as the genetic distance and the great-circle distance as the measure of geographic distance. The longitude and latitude of Urumqi were used to represent CN_XJ, the longitude and latitude of Lhasa were used to represent CN_QTP, and the longitude and latitude of Beijing were used to represent CN_Other. To avoid the confounding effect of recent long-range dispersal along with human activities, the analysis was limited to comparisons within continents.

Population stratification

To analyze the population structure of *D. melanogaster,* we conducted PCA using the smartpca implemented in EIGENSOFT v7.2.1 (*95*). To minimize the influence of natural selection, we used biallelic SNPs with a minor allele frequency (MAF) greater than 0.01 in neutral regions, which included short introns (8 to 30 bp) (*33*) and fourfold degenerate sites within coding regions. PCA outliers in 1439 samples were identified using all neutral SNPs of all samples on autosomes or the X chromosome with two rounds of outlier detection. The final dataset for PCA comprised 185,521 SNPs (2L: 40,553, 2R: 38,789, 3L: 36,849, 3R: 41,555, X: 27,775). Given that inversions can affect genetic diversity across entire chromosome arms (*34*), we performed separate PCAs for each chromosome arm, considering both all genomes and genomes without inversions (inversion-free genomes). The analysis was conducted using default parameters, along with one round of outlier detection across three scales: (i) all 1356 strains, (ii) cosmopolitan strains, and (iii) only CN strains.

We used ADMIXTURE v1.3.0 (*36*) to estimate the genetic ancestry composition in each strain. This analysis was conducted across all strains or exclusively within the CN strains. The input SNP dataset was identical to that used for PCA, which included neutral loci and excluded chromosome arms with inversions. To determine the optimal number of ancestries (*K*), we performed the analysis 10 times, each time using different random seeds for *K* values ranging from 1 to 10. The selection of the best-fitting *K* was based on the lowest cross-validation (CV) error. The results of different runs were aggregated using pong v1.4.9 (*96*) with the parameter "--greedy -s.95." This approach combined redundant runs with a similarity threshold exceeding 95% for each *K*, or it used the greedy algorithm. The representative run for the best-fitting *K* was identified by pong and visualized.

Phylogeny and patterns of introgression

We used TreeMix v1.13 (*42*) to construct ML trees with consideration for gene flow, using allele frequencies across the 22 populations. This analysis was conducted separately for autosomes and the X chromosome, using neutral sites from chromosomal arms that were devoid of inversions. Specifically, for autosomes, we included only 2R and 3L, as 2L and 3R contained inversions in all GH strains, and 3R had inversions in all CO strains. The analysis was executed using blocks of 50 SNPs (-k 50), with KF as the root (-root KF), and incorporating migration events (*m*) spanning from 0 to 10. Each *m* value underwent 100 independent runs, and the optimal number of *m* was determined by evaluating the second-order rate of change in likelihood (Δm) using the R package OptM (*97*). The run with the highest likelihood and the optimal *m* value was visualized using the plotting_funcs.R script in TreeMix.

We further inferred the admixture graph of the 22 populations based on the *f*-statistics using the R package ADMIXTOOLS2 (*45*). The dataset analyzed here was the same as the one used in the TreeMix analysis, with the addition of *D. simulans* as the outgroup. We first applied the extract_f2 function to compute the f_2-statistics with a block size of 4000 bp (blgsize = 4000) for the 22 populations. Then, the find_graphs function was used to find the best-fitting graph topologies for the given set of *f*-statistics for autosomes and the X chromosome. During this process, a random topology was initially formed, with the number of admixture events ranging from 0 to 6 for different runs. The analysis for each number of admixture events was run 20 times. In each run, the topology underwent automatic optimization for up to 200 iterations, halting if no enhancement was observed after 20 iterations, assessed through the likelihood score. The optimal graph topology was determined by comparing the out-of-sample scores and the fitness of a graph using SNP blocks bootstrapped with the compare_fits function (nboot = 100). If a model with more admixture events demonstrated a lower score and a *P* value of less than 0.01, it was considered the better model.

Patterson's *D* statistic was used to analyze introgression patterns within three-population phylogenies (trios), with *D. simulans* used as the outgroup. All SNPs were used in the analysis, and chromosome arms containing inversions were excluded from the analysis. Trios that consistently displayed the same topologies on both autosomes and the X chromosome, as observed in both TreeMix and Dsuite (*98*), were retained for the calculation of *D* statistics. In total, 1257 trios were considered (details in table S7). Gene flow in a trio was considered significant if $D > 0.05$, *z* score > 3, and FDR < 0.05.

Modeling the introgression history of CN_XJ

To reconstruct the introgression history of CN_XJ, which is a mixture of European and Asian ancestries, we applied Ancestry_HMM (*48*) to infer the ancestral tracks of all inversion-free chromosome arms in CN_XJ. Inversion-free chromosome arms from CN_QTP and FR were used as reference populations for European and Asian ancestry, respectively. SNPs separated by at least 100 bp and with a minimum allele frequency difference of 0.1 between CN_QTP and FR were used to distinguish different ancestries by vcf2ahmm.py in Ancestry_HMM. The genetic map of these SNPs was generated using *D. melanogaster* Recombination Rate Calculator (RRC) v2.3 (*99*) based on Comeron *et al.* (*100*). Then, local ancestry inference for each CN_XJ strain was performed robustly by a one-pulse model using the parameter "-a 2 0.7 0.3" to specify the number of ancestral populations and overall ancestry proportions of 70% and 30% from Asia and Europe, respectively, and the parameter "-p 0 1000000 0.7 -p 1 -3000 0.3" to indicate 70% CN_QTP ancestry and 30% FR ancestry entering the population through a single pulse with an initial time estimate of 3000 generations. On the basis of the distribution of homozygous ancestral tracks from autosomes and the X chromosome, we used MultiWaverX (*49*) to infer the best-fitting model from four types of admixture models, including the multiple-wave model, the hybrid-isolation (HI) model, the gradual admixture (GA) model, and the continuous gene flow (CGF) model. We carried out this analysis with default parameters and 100 bootstrapping repeats to provide the support level of the models and the CI of the parameters, including the number of admixture events and the time, proportion, and sex ratio of each admixture event.

Demographic inference with moments

It is worth noting that variations in time estimations can arise from differences in mutation rate and generation time parameters. In all demographic analyses, we used a mutation rate of 5.21×10^{-9} per site per generation, which was derived from mutation accumulation experiments (*101*) and consistent with previous studies (*27*). Additionally, we adopted a generation time of 15 per year based on a previous study (*43*).

We used the Python package moments (*51*) to infer the evolutionary history of *D. melanogaster* in China. Demographic models were fitted for three trios independently, including ((CN, FR), ZI), ((CN_Other, CN_QTP), FR), and ((CN_Other, CN_XJ), FR). For ((CN, FR), ZI), we compared six demographic models: (1) symmetric migration and constant population sizes (OOA_SyM); (2) asymmetrical migration and constant population sizes (OOA_AsyM); (3) symmetric migration, constant population sizes for the sub-Saharan population, and population growth in cosmopolitan populations (OOA_BG_SyM); (4) similar to model 3 but allowing for asymmetrical migration (OOA_BG_AsyM); (5) symmetric migration and population growth in all populations (OOA_BG2_SyM); and (6) similar to model 5 but allowing for asymmetrical migration (OOA_BG2_AsyM) (fig. S16, models 1 to 6). For ((CN_Other, CN_QTP), FR) and ((CN_Other, CN_XJ), FR), we compared another six demographic models: (7) symmetric migration and constant population sizes after the divergence of the populations in Europe and Asia with a bottleneck in the ancestor (COS_SyM), (8) similar to model 7 but allowing for asymmetrical migration (COS_AsyM), (9) similar to model 7 but allowing for population size changes (COS_BG_SyM), (10) similar to model 9 but allowing for asymmetrical migration (COS_BG_AsyM), (11) similar to model 9 but allowing for population growth before the divergence of the populations in Europe and Asia (COS_BG2_SyM), and (12) similar to model 11 but allowing for asymmetrical migration (COS_BG2_AsyM) (fig. S16, models 7 to 12). We fit these models with moments_pipeline (*102*) using the three-dimensional joint site frequency spectrum (SFS), which was generated using neutral SNPs from 2R and 3L. To reduce the computational cost, the sample sizes of ZI, FR, CN, CN_Other, and CN_QTP were projected down to 20. According to this pipeline, we performed four rounds of model optimization (replicates = 100, 100, 100, 100; maxiter = 100, 100, 100, 100; fold = 3, 2, 2, 1) with the Nelder-Mead method (optimize_log_fmin) and estimated the log-likelihood by a multinomial approach. The replicate with the highest log-likelihood for each model was used to calculate Akaike weights. The model with the highest Akaike weight was selected as the best-fitting model. The summary of model choice and detailed parameters outputted for each model are available in table S8. We generated 100 simulated datasets with msms (*103*) according to the best-fitting models and re-estimated the demographic parameters for each of the simulated datasets. The mutation and recombination rates were set to 5.21×10^{-9} per site per generation (*101*) and 1 cM/Mb, respectively, in the simulation. The CIs were calculated using the 2.5th and 97.5th percentiles for all parameters estimated from the simulated data.

Stairway Plot 2

We inferred demographic history with the unfolded SFS by Stairway Plot 2 (*57*) to investigate the changes in population size of three sub-Saharan populations (EA, SP, and ZI) and three cosmopolitan populations (CN, EU, and EG), while EU was merged from CAS, FR, STO, and N to obtain a larger sample size. CN_QTP was also used to be compared with CN. Neutral SNPs from 2R and 3L were used in the analysis separately because inversion frequencies were low on these two arms, while we further excluded genomes with inversions [*In(2R)Ns*, *In(3L)P*, and *In(3L)Ok*] for a specific chromosome arm in consideration of distinct histories of inversions. The ancestral state was inferred as described above. We randomly sampled one allele from each sequenced genome to avoid the effect of inbreeding and generated the unfolded SFS for each population using ∂a∂i v2.1.1 (*91*) with a sample size chosen by the project function to maximize the number of SNPs. The mutation rate was set to 5.21×10^{-9} per site per generation according to the mutation accumulation lines (*101*), and a generation time of 15 per year was applied according to research on the strains in North America (*43*).

Detecting selection signals

Three methods, Fay and Wu's *H* (*59*), Ohana (*60*, *61*), and PBE (*62*), were used to detect selection signals in the six populations with a sample size larger than 50 (three sub-Saharan populations: EF, SD, and ZI and three cosmopolitan populations: CN, FR, and RAL).

The Fay and Wu's *H* (H_{FW}) tests were conducted using all SNPs containing ancestral allele information and all individuals from the six populations. Initially, the multisample VCF file was transformed into the diploid hapmap format using tassel v5.2.40-3 (*104*), with inferred ancestral states added through est-sfs (*93*). Using this modified file, we computed H_{FW} values for nonoverlapping 4-kb windows for each of the six populations with Variscan v2.0 (*105*). Windows containing fewer than 50 segregating sites were excluded. Normalization of H_{FW} was performed based on the number of segregating sites within each 4-kb window. Genes potentially under selection were identified using the following criteria: (i) situated in a 4-kb window with H_{FW} value less than −0.4 and (ii) containing the SNP with a derived allele frequency larger than 0.2 in the gene region.

The Ohana suite (*60*, *61*) was used following its standard pipeline. The selection scan for autosomes and the X chromosome was performed separately. To determine the admixture proportions required for the selection scan, we used neutral SNPs across all genomes. We further refined the dataset by excluding variants and samples with a missing rate exceeding 0.1. Subsequently, SNPs within 50-kb windows that displayed an LD with an r^2 value exceeding 0.2 were removed. Ultimately, the analysis was conducted on datasets consisting of 94,834 SNPs for autosomes and 19,475 SNPs for the X chromosome. Given the ADMIXTURE results indicating an optimal *K* value of 6 to 8 for distinct chromosome arms ($K = 6$ for 2L, $K = 8$ for 3L, and $K = 7$ for 2R, 3R, and X), we proceeded with $K = 7$ for the structure analysis. This choice aimed to effectively distinguish CN, FR, RAL, EF, and the two southern African populations (SD and ZI). The structure analysis involved 20 replication runs, with optimization iterations set to 30 (-mi 30). The run that yielded the highest log-likelihood was used for subsequent analyses. Utilizing the outcomes of the structure analysis, we used all SNPs with a minor allele frequency (MAF) greater than 0.01 to generate admixture-corrected allele frequencies for subsequent selection scans. To identify potential outliers subject to positive selection, we applied a threshold based on the top 1% of LLRSs. For the SNPs identified as under-selection, we assigned them to a specific ancestry if the f-pop value for that ancestry exceeded 0.6. Subsequently, the determined ancestry was linked to a particular population if the average proportion of that ancestry was the highest in the strains of that population (visualized in fig. S20). Genes with at least two selected sites were identified as positively selected genes, and candidate SNPs in these genes were used for GO enrichment. A similar analysis was done for strains in CN with the $K = 2$. A more stringent threshold of LLRS > 20 was used to identify SNPs under positive selection. Because of the very recent differentiation, the SNPs under selection were not assigned to a specific subpopulation in CN.

To compute the PBE statistic, we initiated by calculating the Weir and Cockerham F_{ST} using VCFtools within 4-kb windows with a step size of 1 kb. Subsequently, the PBE was computed using the formula outlined in previous studies (*62*). Windows containing 10 or fewer segregating sites were excluded. The identification of selection signals was executed on each terminal branch as per the phylogenetic tree topology (*63*). The subtree (CN, (FR, RAL), EF) was used to identify terminal branch–specific selection in populations CN, FR, and RAL. Briefly, a 4-kb region that resides within the top 1% of PBE values for both the ((CN, FR), EF) and ((CN, RAL), EF) trios but not for the ((FR, CN), EF), ((RAL, CN), EF), ((FR, RAL), EF), or ((RAL, FR), EF) trios (with the focal population as the first in the trio) would be recognized as specifically selected in CN (Fig. 4A). The second subtree, encompassing ((CN, EF), (SD, ZI)), was used to detect terminal branch–specific selection in populations ZI, SD, and EF, following a similar methodology. We carried out PBE scans on the trio (CN_QTP, CN_XJ, CN_Other) to detect population-specific selection occurring in the three subpopulations. To enhance the robustness of our results and minimize false-positive signals, we identified the putatively positive selected regions in each focal subpopulation by requiring PBE > 0.1 for autosomes and PBE > 0.2 for the X chromosome. In these regions putatively under positive selection, SNPs that exhibited high differentiation ($F_{ST} > 0.1$) between the focal population and any other population within the subtree were annotated, narrowing down the potential genes under selection.

To compare the effect of recombination on adaptation, we compared the selection signals in both low- and high-recombination regions. High-recombination regions were defined as regions with a recombination rate of >0.5 cM/Mb (*100*).

Insecticide-resistant genes

The list of genes associated with insecticide resistance was obtained from FlyBase (www.flybase.org) using the GO terms "insecticide metabolic process," "insecticide catabolic process," and "response to insecticide." Additionally, we conducted a literature search to identify genes linked to insecticide resistance. Notable examples include *Ace* (*66*), *CHKov1*, *para*, the Delta and Epsilon groups of glutathione *S*-transferases (GSTs), members of the cytochrome P450 gene family categorized as involved in "the breakdown of synthetic insecticides" by UniProtKB, and genes encoding subunits of the nicotinic acetylcholine receptor.

Estimate the frequency of structural variations at *Cyp6a17* and *Cyp6a23*

We used Pindel v0.2.5b9 (*106*) to identify the breakpoints of the deletions at *Cyp6a17* and *Cyp6a23* with the options set to "-x 6 -d 20." Deletion calls supported by at least three reads were accepted. We confirmed the breakpoints by polymerase chain reaction (PCR) and Sanger sequencing in CN strains and manually constructed sequences of four alleles for genotyping and frequency estimation, including the intact allele, the *Cyp6a17 deletion* allele, *Cyp6a17/23 fusion 1*, and *Cyp6a17/23 fusion 2* (fig. S40). To infer the genotype of each strain, we first calculated the ratio of the average depth within the deletion and the average depth of the whole chromosome arm. Strains with a ratio greater than 0.8 are regarded as homozygous for the intact allele, while strains with a ratio between 0.3 and 0.8 were considered to be heterozygous for deleted alleles and the intact allele. Strains with a ratio less than 0.3 were considered homozygous for deleted alleles. To distinguish different deleted alleles (e.g., the *Cyp6a17 deletion* allele, the *Cyp6a17/23 fusion 1*, and the *Cyp6a17/23 fusion 2*), reads were mapped to the 200-bp flanking region of the breakpoint by BLAT (*107*). A certain allele is present in a strain if three or more reads span across the breakpoint and are 100% identical to the allele-specific sequence. Strains were excluded from frequency estimation if three or more alleles were detected. For pool-seq samples, the frequency of *Cyp6a17/23 fusion 2* was calculated using the formula $n/[n + (x + y)/2]$, where n, x, and y are the number of reads specifically mapped to *Cyp6a17/23 fusion 2*, *Cyp6a17*, and *Cyp6a23*, respectively.

Estimate the allele age

For SNPs that showed signals of positive selection in CN and were located in genes related to insecticide resistance, we used a genealogy-based approach in Relate v1.1.8 (*68*) to estimate their allele ages. The genetic map was obtained as described above. The phased vcf files required by Relate were obtained by Beagle v5.3 (*108*) with a window size of 0.5 cM and a step size of 0.4 cM based on all SNPs of all genomes. During phasing, the number of model states is twice the number of individuals. To exclude the influence of inbreeding, we randomly selected one phased chromosome from each strain of CN and reconstructed pseudodiploid genomes. The ancestral and derived states of each SNP were obtained as described above. Then, Relate was conducted for all SNPs with the standard pipeline (https://myersgroup.github.io/relate). Detailed information for each SNP was extracted from the mut file. The age of a mutation was equal to the mean of age_begin and age_end, assuming 15 generations per year.

Population cage association study

We combined phenotype sorting by population cage and pool sequencing to identify SNPs associated with insecticide resistance. In this experiment, 200 strains from 20 geographic locations in China were used to represent the natural population. Parents for each strain were transferred into a new vial containing standard cornmeal medium and allowed to lay eggs for 24 hours. For the offspring raised under the low-density condition, 10 male adults of each strain (2000 individuals in total) were collected at the age of 3 to 5 days after eclosion and pooled into a population cage with a food supply to acclimate for 1 day. Then, an emulsion of lambda-cyhalothrin (200 μl/liter) was added to the surface of the food. The food with lambda-cyhalothrin was replaced every 24 hours. The individuals that survived for 48 hours were pooled and sequenced using the Illumina platform to a depth of ~100×. The experiments were performed for two replicates, and the resistant samples were labeled insA (101 individuals) and insB (36 individuals). We also sequenced two background samples (BgA and BgB) to a depth of ~400×, each containing 5 males of each strain (1000 males in total).

The sequencing reads were aligned competitively to the reference genomes of both *D. melanogaster* (dmel_r6.24) and *D. simulans* (dsim_r2.02) using BWA-MEM. To prevent potential contamination, reads mapped to the *D. simulans* reference genome were discarded. Variants were called using freebayes with the parameters "--haplotype-length 0 -m 20 -q 20 -F 0.01 -C 1 --pooled-continuous." We filtered out polymorphic sites with depths lower than 50. Our focus remained on common SNPs, which displayed an average MAF exceeding 5% in the background samples. We used the CMH test to uncover associations between lambda-cyhalothrin resistance and allele counts across the four samples. This process was followed by applying the Benjamini-Hochberg method to ascertain adjusted *P* values (FDR) for each variant. In accordance with common practice in human GWAS (*70*), we adopted a threshold of $P < 1 \times 10^{-9}$ (approximate equivalent to FDR $< 1 \times 10^{-7}$) as the cutoff. We also analyzed adjacent signals by calculating the count of candidate SNPs within sliding 50-kb windows, with a 5-kb step. To assess the enrichment of candidate SNPs along the genome, we randomly chose 7202 SNPs (the observed number of associated SNPs) from the total 1,028,409 SNPs detected in the population-cage sequencing to serve as candidate associates. We then calculated the number of associated SNPs in a sliding 50-kb window with a step size of 5 kb. We repeated the permutation process 1000 times. For each window i, we calculate the mean (E_i) and standard deviation (sd_i) of the candidate SNPs across the 1000 replicates. We then calculated the z score (Z_i) for the actual observed candidate SNPs in that window using the formula $Z_i = (O_i - E_i)/sd_i$, where O_i represents the count of observed associated SNPs in that window. Only windows with a z score exceeding 6, which represent approximately 4% (640 of 15,515) of all analyzed windows, were considered significant. This distribution is illustrated in fig. S42A. Furthermore, we summarized the number of candidate SNPs in each sliding window and found that, in the permutations, the maximum count of candidate SNPs within any window is 22 (fig. S42B). We then identified candidate windows fulfilling both criteria: a *z* score over 6 and more than 22 candidate SNPs. We found 244 windows that met these conditions for the observed associated SNPs. In contrast, no permutations met these stringent criteria (fig. S42C), indicating a significant enrichment with a *P* value less than 0.001. Overlapping windows were merged to ensure a coherent analysis of SNP enrichment. Our analysis ensures that our findings of SNP associations are not due to random chance but represent a statistically significant enrichment. Furthermore, we investigated the influence of LD on the outcomes by estimating LD r^2 in CN using PopLDdecay v3.41 (*109*) for SNPs with a MAF above 0.05.

Functional enrichment analysis

We conducted functional enrichment analyses using Gowinda (*110*). For the analysis of genes under natural selection, we used a set of candidate SNPs that met the aforementioned filtering criteria as input. As a background dataset, we used 12,873,602 SNPs that exhibit polymorphism across the six focal populations (CN, FR, RAL, ZI, SD, and EF). In the context of analyzing genes identified through the lambda-cyhalothrin resistance experiment, we used all candidate SNPs situated in windows containing more than 22 SNPs (totaling 1234 SNPs) as input. As the background dataset, we used 1,028,409 SNPs with a MAF exceeding 0.05 and a depth greater than 50.

GO terms and their associated genes were sourced from FlyBase (http://ftp.flybase.net/releases/current/precomputed_files/go/gene_association.fb.gz, accessed on 6 June 2023). The functional enrichment analysis primarily focused on the Gene Ontology biological process (GO BP) category. We configured Gowinda to execute 10,000,000 simulations, with the parameters "gene-definition" and "mode" set to "gene" (--gene-definition gene --mode gene).

Statistical analyses

All statistical analyses were performed using R (www.r-project.org).

Supplementary Materials

This PDF file includes:
Figs. S1 to S60
Legends for tables S1 to S12

Other Supplementary Material for this manuscript includes the following:
Tables S1 to S12

REFERENCES AND NOTES

1. S. Fan, M. E. Hansen, Y. Lo, S. A. Tishkoff, Going global by adapting local: A review of recent human adaptation. *Science* **354**, 54–59 (2016).
2. S. Lamichhaney, J. Berglund, M. S. Almen, K. Maqbool, M. Grabherr, A. Martinez-Barrio, M. Promerova, C. J. Rubin, C. Wang, N. Zamani, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
3. The 1001 Genomes Consortium, 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* **166**, 481–491 (2016).
4. B. Charlesworth, D. Charlesworth, N. H. Barton, The effects of genetic and geographic structure on neutral variation. *Annu. Rev. Ecol. Evol. Syst.* **34**, 99–125 (2003).
5. C. H. Langley, K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider, J. E. Pool, S. A. Langley, C. Suarez, R. B. Corbett-Detig, B. Kolaczkowski, S. Fang, P. M. Nista, A. K. Holloway, A. D. Kern, C. N. Dewey, Y. S. Song, M. W. Hahn, D. J. Begun, Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* **192**, 533–598 (2012).
6. J. B. Lack, C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig, K. A. Stevens, C. H. Langley, J. E. Pool, 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 (2015).
7. J. E. Pool, R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, M. W. Crepeau, P. Duchen, J. J. Emerson, P. Saelao, D. J. Begun, C. H. Langley, Population Genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLOS Genet.* **8**, e1003080 (2012).
8. J. B. Lack, J. D. Lange, A. D. Tang, R. B. Corbett-Detig, J. E. Pool, A thousand fly genomes: An expanded *Drosophila* Genome Nexus. *Mol. Biol. Evol.* **33**, 3308–3313 (2016).
9. J. K. Grenier, J. R. Arguello, M. C. Moreira, S. Gottipati, J. Mohammed, S. R. Hackett, R. Boughton, A. J. Greenberg, A. G. Clark, Global diversity lines—A five-continent reference panel of sequenced *Drosophila melanogaster* strains. *G3* **5**, 593–603 (2015).
10. H. E. Machado, A. O. Bergland, R. Taylor, S. Tilk, E. Behrman, K. Dyer, D. K. Fabian, T. Flatt, J. Gonzalez, T. L. Karasov, B. Kim, I. Kozeretska, B. P. Lazzaro, T. J. Merritt, J. E. Pool, K. O'Brien, S. Rajpurohit, P. R. Roy, S. W. Schaeffer, S. Serga, P. Schmidt, D. A. Petrov, Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in *Drosophila*. *eLife* **10**, e67577 (2021).
11. M. Kapun, J. C. B. Nunez, M. Bogaerts-Marquez, J. Murga-Moreno, M. Paris, J. Outten, M. Coronado-Zamora, C. Tern, O. Rota-Stabelli, M. P. G. Guerreiro, S. Casillas, D. J. Orengo, E. Puerma, M. Kankare, L. Ometto, V. Loeschcke, B. S. Onder, J. K. Abbott, S. W. Schaeffer, S. Rajpurohit, E. L. Behrman, M. F. Schou, T. J. S. Merritt, B. P. Lazzaro, A. Glaser-Schmitt, E. Argyridou, F. Staubach, Y. Wang, E. Tauber, S. V. Serga, D. K. Fabian, K. A. Dyer, C. W. Wheat, J. Parsch, S. Grath, M. S. Veselinovic, M. Stamenkovic-Radak, M. Jelic, A. J. Buendia-Ruiz, M. J. Gomez-Julian, M. L. Espinosa-Jimenez, F. D. Gallardo-Jimenez, A. Patenkovic, K. Eric, M. Tanaskovic, A. Ullastres, L. Guio, M. Merenciano, S. Guirao-Rico, V. Horvath, D. J. Obbard, E. Pasyukova, V. E. Alatortsev, C. P. Vieira, J. Vieira, J. R. Torres, I. Kozeretska, O. M. Maistrenko, C. Montchamp-Moreau, D. V. Mukha, H. E. Machado, K. Lamb, T. Paulo, L. Yusuf, A. Barbadilla, D. Petrov, P. Schmidt, J. Gonzalez, T. Flatt, A. O. Bergland, *Drosophila* evolution over space and time (DEST): A new population genomics resource. *Mol. Biol. Evol.* **38**, 5782–5805 (2021).
12. M. Kapun, M. G. Barron, F. Staubach, D. J. Obbard, R. A. W. Wiberg, J. Vieira, C. Goubert, O. Rota-Stabelli, M. Kankare, M. Bogaerts-Marquez, A. Haudry, L. Waidele, I. Kozeretska, E. G. Pasyukova, V. Loeschcke, M. Pascual, C. P. Vieira, S. Serga, C. Montchamp-Moreau, J. Abbott, P. Gibert, D. Porcelli, N. Posnien, A. Sanchez-Gracia, S. Grath, E. Sucena, A. O. Bergland, M. P. G. Guerreiro, B. S. Onder, E. Argyridou, L. Guio, M. F. Schou, B. Deplancke, C. Vieira, M. G. Ritchie, B. J. Zwaan, E. Tauber, D. J. Orengo, E. Puerma, M. Aguade, P. Schmidt, J. Parsch, A. J. Betancourt, T. Flatt, J. Gonzalez, Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol. Biol. Evol.* **37**, 2661–2678 (2020).
13. N. R. Garud, P. W. Messer, D. A. Petrov, Detection of hard and soft selective sweeps from *Drosophila melanogaster* population genomic data. *PLOS Genet.* **17**, e1009373 (2021).
14. T. F. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L. L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y. Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, R. A. Gibbs, The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
15. W. Huang, A. Massouras, Y. Inoue, J. Peiffer, M. Ramia, A. M. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. F. Lyman, M. M. Magwire, K. Blankenburg, M. A. Carbone, K. Chang, L. L. Ellis, S. Fernandez, Y. Han, G. Highnam, C. E. Hjelmen, J. R. Jack, M. Javaid, J. Jayaseelan, D. Kalra, S. Lee, L. Lewis, M. Munidasa, F. Ongeri, S. Patel, L. Perales, A. Perez, L. Pu, S. M. Rollmann, R. Ruth, N. Saada, C. Warner, A. Williams, Y. Q. Wu, A. Yamamoto, Y. Zhang, Y. Zhu, R. R. Anholt, J. O. Korbel, D. Mittelman, D. M. Muzny, R. A. Gibbs, A. Barbadilla, J. S. Johnston, E. A. Stone, S. Richards, B. Deplancke, T. F. Mackay, Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* **24**, 1193–1208 (2014).
16. L. Mateo, G. E. Rech, J. González, Genome-wide patterns of local adaptation in Western European *Drosophila melanogaster* natural populations. *Sci. Rep.* **8**, 16143 (2018).
17. Y. Huang, J. B. Lack, G. T. Hoppel, J. E. Pool, Parallel and population-specific gene regulatory evolution in cold-adapted fly populations. *Genetics* **218**, iyab077 (2021).
18. H. Bastide, J. D. Lange, J. B. Lack, A. Yassin, J. E. Pool, A variable genetic architecture of melanic evolution in *Drosophila melanogaster*. *Genetics* **204**, 1307–1319 (2016).
19. S. M. Rudman, S. I. Greenblum, S. Rajpurohit, N. J. Betancourt, J. Hanna, S. Tilk, T. Yokoyama, D. A. Petrov, P. Schmidt, Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science* **375**, eabj7484 (2022).

20. J. Y. Kao, A. Zubair, M. P. Salomon, S. V. Nuzhdin, D. Campo, 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 (2015).
21. A. O. Bergland, R. Tobler, J. González, P. Schmidt, D. Petrov, Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol. Ecol.* **25**, 1157–1174 (2016).
22. B. Kolaczkowski, A. D. Kern, A. K. Holloway, D. J. Begun, Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics* **187**, 245–260 (2011).
23. F. G. Liu, S. C. Tsaur, H. T. Huang, Biogeography of *Drosophila* (Diptera: Drosophilidae) in East and Southeast Asia. *J. Insect Sci.* **15**, 69 (2015).
24. J. David, C. Bocquet, E. Pla, New results on the genetic characteristics of the Far East race of *Drosophila melanogaster*. *Genet. Res.* **28**, 253–260 (1976).
25. X. Mi, G. Feng, Y. Hu, J. Zhang, L. Chen, R. T. Corlett, A. C. Hughes, S. Pimm, B. Schmid, S. Shi, J. C. Svenning, K. Ma, The global significance of biodiversity science in China: An overview. *Natl. Sci. Rev.* **8**, nwab032 (2021).
26. C. M. Bergman, P. R. Haddrill, Strain-specific and pooled genome sequences for populations of *Drosophila melanogaster* from three continents. *F1000Res.* **4**, 31 (2015).
27. Q. D. Sprengelmeyer, S. Mansourian, J. D. Lange, D. R. Matute, B. S. Cooper, E. V. Jirle, M. C. Stensmyr, J. E. Pool, Recurrent Collection of *Drosophila melanogaster* from Wild African Environments and Genomic Insights into Species History. *Mol. Biol. Evol.* **37**, 627–638 (2020).
28. E. G. King, S. J. Macdonald, A. D. Long, Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics* **191**, 935–949 (2012).
29. D. Campo, K. Lehmann, C. Fjeldsted, T. Souaiaia, J. Kao, S. V. Nuzhdin, Whole-genome sequencing of two North American *Drosophila melanogaster* populations reveals genetic differentiation and positive selection. *Mol. Ecol.* **22**, 5084–5097 (2013).
30. K. Singhal, R. Khanna, S. Mohanty, Is *Drosophila*-microbe association species-specific or region specific? A study undertaken involving six Indian *Drosophila* species. *World J. Microbiol. Biotechnol.* **33**, 103 (2017).
31. T. Lian, D. Li, X. Tan, T. Che, Z. Xu, X. Fan, N. Wu, L. Zhang, U. Gaur, B. Sun, M. Yang, Genetic diversity and natural selection in wild fruit flies revealed by whole-genome resequencing. *Genomics* **110**, 304–309 (2018).
32. P. D. Keightley, A. Eyre-Walker, Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* **177**, 2251–2261 (2007).
33. J. Parsch, S. Novozhilov, S. S. Saminadin-Peter, K. M. Wong, P. Andolfatto, On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol. Biol. Evol.* **27**, 1226–1234 (2010).
34. R. B. Corbett-Detig, D. L. Hartl, Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1003056 (2012).
35. M. Kapun, H. van Schalkwyk, B. McAllister, T. Flatt, C. Schlotterer, Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. *Mol. Ecol.* **23**, 1813–1827 (2014).
36. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
37. J. Novembre, T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, C. D. Bustamante, Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
38. S. Wright, The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).
39. S. Wright, Isolation by distance. *Genetics* **28**, 114–138 (1943).
40. Y. Ishida, Sewall Wright and Gustave Malécot on isolation by distance. *Philos Sci.* **76**, 784–796 (2009).
41. T. E. Cruickshank, M. W. Hahn, Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
42. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012).
43. J. E. Pool, The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol. Biol. Evol.* **32**, 3236–3251 (2015).
44. P. Duchen, D. Zivkovic, S. Hutter, W. Stephan, S. Laurent, Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* **193**, 291–301 (2013).
45. R. Maier, P. Flegontov, O. Flegontova, P. Changmai, D. Reich, On the limits of fitting complex models of population history to genetic data. bioRix 2022.05.08.491072 [Preprint] (2022). https://doi.org/10.1101/2022.05.08.491072.
46. P. Medina, B. Thornlow, R. Nielsen, R. Corbett-Detig, Estimating the timing of multiple admixture pulses during local ancestry inference. *Genetics* **210**, 1089–1107 (2018).
47. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prufer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hober, B. Hoffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Z. Kucan, I. Gusic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Paabo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
48. R. Corbett-Detig, R. Nielsen, A hidden markov model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. *PLOS Genet.* **13**, e1006529 (2017).
49. R. Zhang, X. Ni, K. Yuan, Y. Pan, S. Xu, MultiWaverX: Modeling latent sex-biased admixture history. *Brief. Bioinform.* **23**, bbac179 (2022).
50. Y. Pan, C. Zhang, Y. Lu, Z. Ning, D. Lu, Y. Gao, X. Zhao, Y. Yang, Y. Guan, D. Mamatyusupu, S. Xu, Genomic diversity and post-admixture adaptation in the Uyghurs. *Natl. Sci. Rev.* **9**, nwab124 (2022).
51. J. Jouganous, W. Long, A. P. Ragsdale, S. Gravel, Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics* **206**, 1549–1567 (2017).
52. H. Li, W. Stephan, Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLOS Genet.* **2**, e166 (2006).
53. S. J. Laurent, A. Werzner, L. Excoffier, W. Stephan, Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol. Biol. Evol.* **28**, 2041–2051 (2011).
54. J. R. Arguello, S. Laurent, A. G. Clark, Demographic History of the Human Commensal *Drosophila melanogaster*. *Genome Biol. Evol.* **11**, 844–854 (2019).
55. K. Thornton, P. Andolfatto, Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**, 1607–1619 (2006).
56. A. Keller, *Drosophila melanogaster*'s history as a human commensal. *Curr. Biol.* **17**, R77–R81 (2007).
57. X. Liu, Y. X. Fu, *Stairway Plot 2*: Demographic history inference with folded SNP frequency spectra. *Genome Biol.* **21**, 280 (2020).
58. X. Liu, Y. X. Fu, Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).
59. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
60. J. Y. Cheng, A. J. Stern, F. Racimo, R. Nielsen, Detecting selection in multiple populations by modeling ancestral admixture components. *Mol. Biol. Evol.* **39**, msab294 (2022).
61. J. Y. Cheng, T. Mailund, R. Nielsen, Fast admixture analysis and population tree estimation for SNP and NGS data. *Bioinformatics* **33**, 2148–2155 (2017).
62. A. Yassin, V. Debat, H. Bastide, N. Gidaszewski, J. R. David, J. E. Pool, Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4771–4776 (2016).
63. D. Wu, J. Dou, X. Chai, C. Bellis, A. Wilm, C. C. Shih, W. W. J. Soon, N. Bertin, C. B. Lin, C. C. Khor, M. DeGiorgio, S. Cheng, L. Bao, N. Karnani, W. Y. K. Hwang, S. Davila, P. Tan, A. Shabbir, A. Moh, E. K. Tan, J. N. Foo, L. L. Goh, K. P. Leong, R. S. Y. Foo, C. S. P. Lam, A. M. Richards, C. Y. Cheng, T. Aung, T. Y. Wong, H. H. Ng, J. Liu, C. Wang, Large-scale whole-genome sequencing of three diverse Asian populations in Singapore. *Cell* **179**, 736–749.e15 (2019).
64. P. Battlay, P. B. Leblanc, L. Green, N. R. Garud, J. M. Schmidt, A. Fournier-Level, C. Robin, Structural variants and selective sweep foci contribute to insecticide resistance in the *Drosophila* Genetic Reference Panel. *G3* **8**, 3489–3497 (2018).
65. T. Karasov, P. W. Messer, D. A. Petrov, Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLOS Genet.* **6**, e1000924 (2010).
66. A. Mutero, M. Pralavorio, J. M. Bride, D. Fournier, Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 5922–5926 (1994).
67. R. T. Good, L. Gramzow, P. Battlay, T. Sztal, P. Batterham, C. Robin, The molecular evolution of *Cytochrome P450* genes within and between *Drosophila* species. *Genome Biol. Evol.* **6**, 1118–1134 (2014).
68. L. Speidel, M. Forest, S. Shi, S. R. Myers, A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
69. N. R. Garud, P. W. Messer, E. O. Buzbas, D. A. Petrov, Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genet.* **11**, e1005004 (2015).
70. E. Uffelmann, Q. Q. Huang, N. S. Munung, J. de Vries, Y. Okada, A. R. Martin, H. C. Martin, T. Lappalainen, D. Posthuma, Genome-wide association studies. *Nat. Rev. Method. Prim.* **1**, 59 (2021).
71. K. M. Seong, B. S. Coates, B. R. Pittendrigh, Cytochrome P450s *Cyp4p1* and *Cyp4p2* associated with the DDT tolerance in the *Drosophila melanogaster* strain 91-R. *Pest. Biochem. Physiol.* **159**, 136–143 (2019).
72. T. M. Bach, H. Takagi, Properties, metabolisms, and applications of (L)-proline analogues. *Appl. Microbiol. Biotechnol.* **97**, 6623–6634 (2013).
73. V. Balabanidou, L. Grigoraki, J. Vontas, Insect cuticle: A critical determinant of insecticide resistance. *Curr. Opin. Insect Sci.* **27**, 68–74 (2018).

74. L. Grigoraki, X. Grau-Bové, H. Carrington Yates, G. J. Lycett, H. Ranson, Isolation and transcriptomic analysis of *Anopheles gambiae* oenocytes enables the delineation of hydrocarbon biosynthesis. *eLife* **9**, e58019 (2020).
75. B. Dong, X. Y. Liu, B. Li, M. Y. Li, S. G. Li, S. Liu, A heat shock protein protects against oxidative stress induced by lambda-cyhalothrin in the green peach aphid *Myzus persicae*. *Pest. Biochem. Physiol.* **181**, 104995 (2022).
76. S. Q. Le, R. Durbin, SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Res.* **21**, 952–960 (2011).
77. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN] (2012).
78. D. C. Koboldt, Best practices for variant calling in clinical sequencing. *Genome Med.* **12**, 91 (2020).
79. A. A. Hoffmann, L. H. Rieseberg, Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annu. Rev. Ecol. Evol. Syst.* **39**, 21–42 (2008).
80. B. Charlesworth, J. A. Coyne, N. H. Barton, The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
81. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
82. H. Becher, B. C. Jackson, B. Charlesworth, Patterns of genetic variability in genomic regions with low rates of recombination. *Curr. Biol.* **30**, 94–100.e3 (2020).
83. B. Charlesworth, J. L. Campos, B. C. Jackson, Faster-X evolution: Theory and evidence from *Drosophila*. *Mol. Ecol.* **27**, 3753–3771 (2018).
84. J. D. Jensen, On the unfounded enthusiasm for soft selective sweeps. *Nat. Commun.* **5**, 5281 (2014).
85. A. A. Hoffmann, A. R. Weeks, 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**, 133–147 (2007).
86. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
87. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN] (2013).
88. E. Garrison, Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, P. Prins, Vcflib and tools for processing the VCF variant call format. bioRxiv 2021.05.21.445151v1 [Preprint] (2021). https://www.biorxiv.org/content/10.1101/2021.05.21.445151v1.
89. J. Jurka, Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).
90. P. Cingolani, A. Platts, L. W. Le, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
91. R. N. Gutenkunst, R. D. Hernandez, S. H. Williamson, C. D. Bustamante, Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genet.* **5**, e1000695 (2009).
92. K. L. Korunes, K. Samuk, pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol. Ecol. Resour.* **21**, 1359–1368 (2021).
93. P. D. Keightley, B. C. Jackson, Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics* **209**, 897–906 (2018).
94. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, G., The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
95. N. Patterson, A. L. Price, D. Reich, Population structure and eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
96. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, Pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics* **32**, 2817–2823 (2016).
97. R. R. Fitak, OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* **6**, bpab017 (2021).
98. M. Malinsky, M. Matschiner, H. Svardal, Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
99. A. S. Fiston-Lavier, N. D. Singh, M. Lipatov, D. A. Petrov, *Drosophila melanogaster* recombination rate calculator. *Gene* **463**, 18–20 (2010).
100. J. M. Comeron, R. Ratnappan, S. Bailin, The many landscapes of recombination in *Drosophila melanogaster*. *PLOS Genet.* **8**, e1002905 (2012).
101. W. Huang, R. F. Lyman, R. A. Lyman, M. A. Carbone, S. T. Harbison, M. M. Magwire, T. F. Mackay, Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *eLife* **5**, e14625 (2016).
102. A. D. Leaché, D. M. Portik, D. Rivera, M.-O. Rödel, J. Penner, V. Gvoždík, E. Greenbaum, G. F. M. Jongsma, C. Ofori-Boateng, M. Burger, E. A. Eniang, R. C. Bell, M. K. Fujita, Exploring rain forest diversification using demographic model testing in the African foam-nest *treefrogChiromantis rufescens*. *J. Biogeogr.* **46**, 2706–2721 (2019).
103. G. Ewing, J. Hermisson, MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**, 2064–2065 (2010).
104. P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, E. S. Buckler, TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
105. A. J. Vilella, A. Blanco-Garcia, S. Hutter, J. Rozas, VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* **21**, 2791–2793 (2005).
106. K. Ye, M. H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
107. W. J. Kent, BLAT—The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
108. B. L. Browning, X. Tian, Y. Zhou, S. R. Browning, Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).
109. C. Zhang, S. S. Dong, J. Y. Xu, W. M. He, T. L. Yang, PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
110. R. Kofler, C. Schlotterer, Gowinda: Unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* **28**, 2084–2085 (2012).

Acknowledgments: We thank the National Center for Protein Sciences at Peking University for technical assistance. Some of the analyses were performed on the High-Performance Computing Platform of the Center for Life Sciences. **Funding:** This work was supported by the Ministry of Science and Technology of the People's Republic of China (grant 2022YFE0132000) awarded to J.L., the National Natural Science Foundation of China (grant 32070597) awarded to J.L., and the Natural Science Foundation of Beijing (grant 5212006) awarded to J.L. **Author contributions:** Conceptualization: J.L. and A.G.C. Methodology: W.Z., Y.W., J.C., C.L., and W.L. Investigation: J.C., C.L., and W.L. Visualization: J.C., C.L., and W.L. Supervision: J.L. and A.G.C. Writing—original draft: J.C., C.L., and W.L. Writing—review and editing: J.L., A.G.C., J.C., C.L., and W.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Sequence read data are available from NIH Short Read Archive under accession PRJNA924845.

Submitted 4 March 2023
Accepted 13 March 2024
Published 17 April 2024
10.1126/sciadv.adh3425