



Low levels of genetic differentiation with isolation by geography and environment in populations of *Drosophila melanogaster* from across China

Lei Yue¹ · Li-Jun Cao¹ · Jin-Cui Chen¹ · Ya-Jun Gong¹ · Yan-Hao Lin^{1,2} · Ary Anthony Hoffmann³ · Shu-Jun Wei¹

Received: 19 October 2020 / Revised: 16 February 2021 / Accepted: 17 February 2021 / Published online: 8 March 2021
© The Author(s), under exclusive licence to The Genetics Society 2021

Abstract

The fruit fly, *Drosophila melanogaster*, is a model species in evolutionary studies. However, population processes of this species in East Asia are poorly studied. Here we examined the population genetic structure of *D. melanogaster* across China. There were 14 mitochondrial haplotypes with 10 unique ones out of 23 known from around the globe. Pairwise F_{ST} values estimated from 15 novel microsatellites ranged from 0 to 0.11, with geographically isolated populations showing the highest level of genetic uniqueness. STRUCTURE analysis identified high levels of admixture at both the individual and population levels. Mantel tests indicated a strong association between genetic distance and geographical distance as well as environmental distance. Full redundancy analysis (RDA) showed that independent effects of environmental conditions and geography accounted for 62.10% and 31.58% of the total explained genetic variance, respectively. When geographic variables were constrained in a partial RDA analysis, the environmental variables bio2 (mean diurnal air temperature range), bio13 (precipitation of the wettest month), and bio15 (precipitation seasonality) were correlated with genetic distance. Our study suggests that demographic history, geographical isolation, and environmental factors have together shaped the population genetic structure of *D. melanogaster* after its introduction into China.

Introduction

Patterns of population genetic variation can be shaped by neutral and adaptive evolutionary processes and there is an increasing understanding of how demography, gene flow, and selection interact to drive extant genetic patterns across space

(Bradburd & Ralph, 2019; Lowe et al., 2017; Semenov et al., 2019). The fruit fly, *Drosophila melanogaster*, has been recognized as a useful model organism in understanding such interactions following many decades of research. The species has a cosmopolitan distribution and shows an ability to rapidly adapt to varying environments (David & Capy, 1988; Markow & O'Grady, 2005). Past biogeographical and systematic studies have confirmed that *D. melanogaster* was native to sub-Saharan Africa (Lachaise et al., 1988), and expanded its range from Middle East into the Eurasian continent roughly at 1800 years ago (Sprengelmeyer et al., 2020) and to America and Australia only about 100–200 years ago (Arguello et al., 2019; Stephan & Li, 2007). Current studies have shown that both historical demographic events and local adaptation as well as their interplay play crucial roles in shaping the population genetic structure of *D. melanogaster* (Arguello et al., 2016; Flatt, 2016; Keller, 2007; Laurent et al., 2011).

One factor influencing the population genetic structure of *D. melanogaster* is historical dispersal (Keller, 2007; Sprengelmeyer et al., 2020). North American and African populations of *D. melanogaster* have been affected by African and European migrations associated with human cargo transportation (Bergland et al., 2016; David & Capy,

Associate Editor Darren Obbard

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-021-00419-8>.

✉ Li-Jun Cao
gmatjhpl@163.com

✉ Shu-Jun Wei
shujun268@163.com

- ¹ Institute of Plant and Environmental Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China
- ² International Department of Beijing No. 80 High School, Beijing, China
- ³ Bio21 Institute, School of BioSciences, The University of Melbourne, Victoria, Australia

1988; Kao et al., 2015; Lachaise et al., 1988). The exchange of fruit around the world has provided an opportunity for ongoing gene flow among populations, contributing to genetic admixture worldwide in recent years as documented in African (Kauer et al., 2003; Pool et al., 2012), North American (Duchen et al., 2013), and Australian (Bergland et al., 2016) populations. Compared to rural populations, African urban populations exhibit substantial genetic admixture from European populations (Kauer et al., 2003; Pool et al., 2012); although the high levels of inferred non-African ancestry can be affected by adaptation, ongoing trade might also contribute (Kauer et al., 2003; Pool et al., 2012).

Another factor influencing the population genetic structure of *D. melanogaster* is natural selection (Flatt, 2016; Hoffmann et al., 2003; Sigh et al., 1982). Numerous studies have revealed clinal variation in phenotypes, chromosomal arrangements, and genotypes across environmental gradients for this species, implicating spatially varying selection (Flatt, 2016; Hoffmann & Weeks, 2007). DNA variation in genic targets of clines can be connected with adaptation to ecological factors like temperature, precipitation, and oxygen concentration (Parkash & Munjal, 2000; Schmidt et al., 2005; Weeks et al., 2002). To understand such clinal variation in *D. melanogaster*, the effect of natural selection should be disentangled from demographic events (Miller et al., 2020; Sexton et al., 2014; Smith et al., 2020).

The Eurasian continent has been proposed as one of the earliest areas colonized by *D. melanogaster* out of its native range (David & Capy, 1988). The East Asian populations of *D. melanogaster* may have some unique genetic characteristics compared to other populations (Schlötterer et al., 2006). However, there are very few studies on the population genetic structure of *D. melanogaster* across East Asia, with most previous studies concentrating on African, European, Australian, and American populations (Duchen et al., 2013; Kapun et al., 2018; Kauer et al., 2003; Stephan & Li, 2007). East Asia has a diversified geographical environment and a long history of human colonization. This area is one of the most important regions in the spread and diffusion of many species (Leipe et al., 2019; Molina et al., 2011; Xiang et al., 2018; Yu et al., 2018), especially those associated with human trade—which are likely to include *D. melanogaster* (Arguello et al., 2019; Keller, 2007). However, limited knowledge on the population genetic structure of *D. melanogaster* in East Asia has hindered an understanding of expansion and adaptation processes in this region and connections to other areas.

Here, we examined the genetic diversity and population structure of *D. melanogaster* populations collected across China based on the mitochondrial *cox1* gene and microsatellite loci. We hypothesized that (1) low population genetic differentiation exists among Chinese populations of *D. melanogaster* owing to its relatively short history of

introduction into China; and (2) genetic distance across Chinese populations are correlated with both geographical distance and environmental factors, given the rapid adaptive evolution of *D. melanogaster* and varied ecological environment across China, as already seen in other non-migratory invertebrates in this region (Cao et al., 2016b, 2019; Wei et al., 2015).

Materials and methods

Sample collection and DNA extraction

Specimens of *D. melanogaster* were collected from 16 locations in China through trapping by using rotten fruits (e.g., grapes, watermelon, and banana) in 2019 (Table 1 and Fig. 1). The sampled flies were first checked morphologically using an anatomical lens, followed by molecular identification by BLAST searching of the mitochondrial cytochrome oxidase subunit I (*cox1*) gene (see below) against the nucleotide database (nt) and in the BOLD system (<http://www.boldsystems.org>) to ensure the reliability of morphological identification. All samples were stored in 98% alcohol, and frozen at -80°C before DNA extraction. A DNeasy Blood and Tissue Kit (Qiagen, Germany) was used to extract total genomic DNA individually for 330 samples randomly selected from 16 populations.

Mitochondrial gene amplification and sequencing

To characterize the mitochondrial variation and validate correct identification of the specimens, a fragment of the *cox1* gene involving the DNA barcoding region of insects was sequenced using the universal primer pairs AF (5'-GGTCAACAAATCATAAAGATATTGG-3') and AR (5'-TAAACTTCAGGGTGACCAAAAATCA-3') (Folmer et al., 1994). Polymerase chain reaction (PCR) was conducted using the Mastercycler Pro system (Eppendorf, Germany) in a 10 μL volume consisting of 0.5 μL of template DNA (5–20 ng/ μL), 0.5 μL of LA Taq (Takara Biomedical, Japan, 5 U/ μL), 1 μL of LA PCR Buffer II (Mg²⁺), 0.16 μL of forward primer (10 mM), 0.16 μL of reverse primer (10 mM), and 7.68 μL of ddH₂O. The thermal profiles for DNA amplification were as follows: 4 min at 94 $^{\circ}\text{C}$; 35 cycles of 30 s at 94 $^{\circ}\text{C}$, 30 s at 52 $^{\circ}\text{C}$, and 60 s at 72 $^{\circ}\text{C}$, followed by a final 10-min extension at 72 $^{\circ}\text{C}$. Amplified products were purified and sequenced on an ABI 3730xl DNA Analyzer by TianYi HuiYuan Biotechnology Co. Ltd (Beijing, China).

Microsatellite marker development and genotyping

We developed genome-wide microsatellites from the *D. melanogaster* genome (accession PGRM00000000). The

Table 1 Collection information for *Drosophila melanogaster* specimens used in this study.

Population code	Collection location	Longitude (°E), Latitude (°N)	Collection date (year/month)
XJKL	Korla, Xinjiang Province	86.18, 41.73	2019/10
YNDH	Dehong, Yunnan Province	98.59, 24.44	2019/10
SCPZ	Panzhihua, Sichuan Province	101.73, 26.59	2019/10
QHYN	Xining, Qinghai Province	101.78, 36.62	2019/10
YNKM	Kunming, Yunnan Province	102.85, 24.87	2019/11
SCCD	Chengdu, Sichuan Province	104.08, 30.66	2019/10
SXAK	Ankang, Shannxi Province	109.04, 32.69	2019/10
NMHH	Huhhot, Neimenggu Province	111.76, 40.85	2019/10
GDGZ	Guangzhou, Guangdong Province	113.27, 23.14	2019/10
HBWH	Wuhan, Hubei Province	114.31, 30.60	2019/09
JXSR	Shangrao, Jiangxi Province	117.95, 28.46	2019/10
SDLY	Linyi, Shandong Province	118.36, 35.11	2019/10
FJFZ	Fuzhou, Fujian Province	119.30, 26.08	2019/09
NMHB	Hulun Buir, Neimenggu Province	120.19, 50.25	2019/09
ZJHZ	Hangzhou, Zhejiang Province	120.22, 30.25	2019/09
LNHL	Huludao, Liaoning Province	120.84, 40.72	2019/09

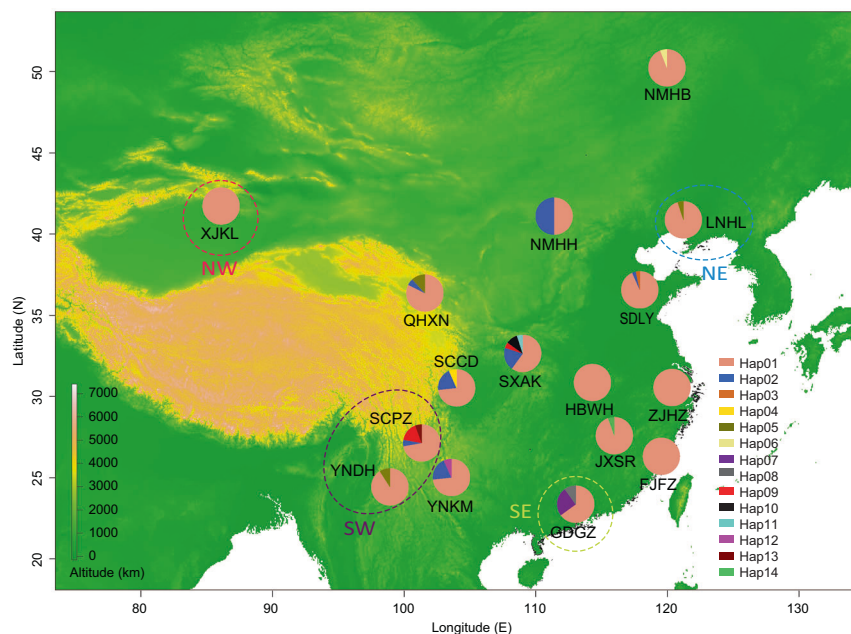


Fig. 1 Collection site and geographical distribution of the mitochondrial *coxI* haplotypes of *Drosophila melanogaster*. The haplotype frequency for each population is represented by colors in the pie charts. See Table 1 for population codes. Four of the five population

groups identified in population genetic structure analysis are marked by circles: NW northwest, SW southwest, SE southeast, NE northeast. The rest of the populations formed the CE (central) group.

QDD3 program (Meglecz et al., 2010) was used to extract microsatellites along with their flanking sequences (300 bp each) from the scaffolds, and to design the primers. According to previous reports, dinucleotide microsatellites are prone to polymerase slippage during the amplification process, which may lead to mistyping, so multi- (including tri-, tetra-, penta-, and hexa-) nucleotide loci were

preferentially selected in this study. The criteria and parameters of primer design for the isolated microsatellite markers followed Cao et al. (2016a). Next, the output primers were further filtered manually using more stringent criteria: (i) only perfect (without repeat region interruptions) loci were retained; (ii) the minimum distance between the 3' end of two primer pairs and their target region had to be

Table 2 Fifteen microsatellite markers developed to genotype *Drosophila melanogaster* in this study.

Locus	Chr.	Position	Repeat motif	Allele number	Size range (bp)	Forward primer (5'-3')	Reverse primer (5'-3')
S04	2L	2,932,573	AGC	6	136–149	CTTCGCCTTTCCAGCTTTGT	GTCGTAGTTGTGTCGCTGTC
S09	2R	23,181,060	AGC	10	135–175	GGGCTTCCTCATCCGGAATT	GCCACCTCGCTGATAATTG
S10	2R	12,896,467	AGCC	21	141–229	GTCCTTAGTCGAACCGCTCT	CACGAGCGACCATTGCATTT
S14	2R	13,082,536	AGC	9	159–194	TTGCTTTGTCTGATTGCGGA	CGAATCTGGCAACTGTGGAG
S17	2L	3,419,414	AGC	8	183–206	TCCGTTCCCATCTTCTCGG	GCTCCGGCTGCAACTTAATT
S18	3L	12,644,547	AGC	8	181–214	CAAGGCAGAGCATCACCATC	CAAACACTTCTGCGCTG
S28	3R	15,812,776	ACT	3	223–235	ATCCACGTCCCACCAGAAG	AAACGCACAACAACGCCATA
S39	2L	21,941,431	ATCCG	9	265–304	AAACGCATTTGCCACCTTT	GGCACCGAACATCAATCTCC
S41	2L	9,658,593	AGC	11	290–325	AGTTAACAAGTGACACGCC	ATCTCGAATCGCCTTGGCTA
S42	2L	16,449,207	ACC	6	287–305	TATTTAAGGGCATGGCGGGT	GCCACGAGGAATTGGACATC
S44	2R	11,206,856	AGC	8	312–337	GAGGTCGGTGAAAGGAGTGG	CGATCATCATTCTGCCGCAA
S47	2R	13,921,847	AGCC	11	307–357	GGTGCCTCTTAATGTGTCAG	ATGCATAAAGTCCGGCAAG
S48	3R	5,290,483	AGG	4	324–336	GCTTCCCTCTCCAATTCTGC	ATGGATGGGACTAACGGAA
S52	3L	13,589,683	AGC	7	350–367	GCCTTGCCAAGTTTCCAGTT	AAGCACACCATATCGCAGAC
S55	3L	8,663,943	ATC	7	343–359	GTTGTTGAAAGCGATTGCC	CTTGTTGACGCGGCATCTAA

Allele number and size range for each locus was calculated from 16 populations of *Drosophila melanogaster* in this study.

Chr. chromosome.

>10 bp; and (iii) primers using the “A” design strategy that do not have multiple microsatellites, nanosatellites, or homopolymers in the amplicon were retained. A total of 60 primer pairs were ultimately screened out, which were used for further isolation of polymorphic microsatellite loci.

The 60 primer pairs were initially chemically synthesized by Tsingke Biotechnology Co. Ltd (Beijing, China), with a universal primer (CAGGACCAGGCTACCGTG) added to the 5' end of the forward primers based on the method of Blacket et al. (2012). One individual *D. melanogaster* from each of the 16 populations was randomly selected to test the polymorphism level of target microsatellite sequences and the amplification efficiency of synthesized primers. The PCR amplification reaction was carried out in a 10 μ L volume consisting of 0.5 μ L of template DNA (5–20 ng/ μ L), 5 μ L of Master Mix (Promega, Madison, WI, USA), 0.08 μ L of PC tail modified forward primer (10 mM), 0.16 μ L of reverse primer (10 mM), 0.32 μ L of fluorescence-labeled PC tail (10 mM), and 3.94 μ L of ddH₂O. The thermal profiles for DNA amplification were as follows: 4 min at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 56 °C, and 45 s at 72 °C, followed by a final 10-min extension at 72 °C. The amplified PCR fragments were analyzed on an ABI 3730xl DNA Analyzer (Applied Biosystems) using the GeneScan 500 LIZ size standard (Applied Biosystems). Genotyping was conducted using GENEMAPPER 4.0 (Applied Biosystems, USA). Primer pairs that failed in the PCR amplification for more than two individuals, that were monomorphic in tested individuals, or that produced more than two peaks were discarded (Table 2).

Identification of relatives within populations

We identified relatives within each population to reduce the influence of close relatives on genetic diversity and population structure analysis. The software of SPAGeDi v 1.5 (Hardy & Vekemans, 2002) was used to calculate Loiselle's *K* based on microsatellite loci. Kinship among individuals was identified based on the *K* scores of pairwise individuals. The maximum *K* value among individual pairs with different mitochondrial *cox1* haplotypes was used as a standard to discriminate the related pairs. Individual pairs with *K* values larger than the standard were considered as relatives.

Genetic diversity analysis

For the mitochondrial *cox1* gene, sequencing results from both strands were assembled into a consensus sequence using the software SEQMAN in LASERGENE version 7.1.2 (DNASTAR, Inc., USA). Sequences from all individuals were aligned using CLUSTALW (Thompson et al., 1994) implemented in MEGA version 7 (Kumar et al., 2016). The number of polymorphic sites (*S*), haplotype diversity (*h*), and nucleotide diversity (π) were analyzed in DnaSP version 5.10 (Librado & Rozas, 2009).

For microsatellite loci, the genotyping results were corrected with the software MICRO-CHECKER (van Oosterhout et al., 2004). The number of alleles, observed heterozygosity (*H*_O), and expected heterozygosity (*H*_E) were calculated in the Microsoft Excel microsatellite toolkit version 3.1 (Park, 2001). Deviation from Hardy–Weinberg

equilibrium (HWE) at each locus/population combination, linkage disequilibrium (LD) among loci within each population, and pairwise mean population differentiation (F_{ST}) were examined using GENEPOP 4.2.1 (Rousset, 2008). We compared the total number of alleles (A_T), standardized number of alleles (A_S), and standardized expected heterozygosity (H_{ES}) among populations with different sample sizes using a rarefaction method in GENCLONE (Arnaud-Haond & Belkhir, 2007). Allelic richness (A_R) and allelic richness of private alleles (P_{AR}) were calculated with a rarefaction approach in HP-RARE version 1.1 (Kalinowski, 2005) on a minimum sample size of 12 diploid individuals. Finally, the inbreeding coefficient (F_{is}) within populations was estimated by FSTAT version 2.9.3 (Goudet, 2017).

Phylogenetic relationship and population genetic structure analysis

For mitochondrial DNA, a neighbor-joining phylogenetic tree was constructed by MEGA version 7 (Kumar et al., 2016) to examine phylogenetic relationships among mitochondrial *coxI* haplotypes of *D. melanogaster* from different sampling locations worldwide. Apart from sequences obtained in this study, we searched orthologous *coxI* gene sequences from NCBI nucleotide database and included those with sampling locations in phylogenetic analysis (Nunes et al., 2008a, b). A Kimura two-parameter distance model was used in the tree inference as suggested by analysis in PartitionFinder version 1.1.1 (Lanfear et al., 2012). We conducted a bootstrap test with 1000 replicates.

For microsatellite loci, a Bayesian cluster method implemented in STRUCTURE version 2.3.4 (Pritchard et al., 2000) and a discriminant analysis of principal components (DAPC) were used, respectively, to investigate genetic structure across the populations. In the STRUCTURE analysis, 30 replicates for each K (from 1 to 10) were run with 200,000 Markov Chain Monte Carlo iterations after a burn-in of 100,000 iterations. The optimal K value was obtained using the delta (K) method by submitting output results of STRUCTURE to the online software of STRUCTURE HARVESTER version 0.6.94 (Earl, 2012) (<http://taylor0.biology.ucla.edu/structureHarvester/>). CLUMPP version 1.1.2 (Jakobsson & Rosenberg, 2007) based on the greedy algorithm was used to combine the membership coefficient matrices (Q-matrices) of replicated runs for each K from STRUCTURE HARVESTER, and DISTRUCT version 1.1 (Rosenberg, 2004) was applied to prepare the visual representation. We ran a DAPC analysis in the R package ADEGENET 1.4-2 (Jombart et al., 2008) to identify genetic clusters among sampling individuals. DAPC analyses do not rely on any biological assumptions, providing a supplement to model-based STRUCTURE analysis.

Demographic history analysis

We used the approximate Bayesian computation, a model-based method, implemented in DIYABC version 2.1.0 (Cornuet et al., 2014), to infer the demographic history of *D. melanogaster* in China using the microsatellite data. This method allows a comparison of competing scenarios and incorporates unsampled populations in the analysis (Bertorelle et al., 2010). It has proven to be a powerful approach to test complex population genetic models in many species (Fraitout et al., 2017; Lombaert et al., 2014; Stone et al., 2017; Wei et al., 2015). In the analysis, we considered the relationship between the NW (XJKL) group, SW (SCPZ and YNDH) group, and CE (QHXN, YNKM, SCCD, SXAK, NMHH, HBWH, JXSR, SDLY, FJFZ, NMHL, and ZJHZ) group by comparing 18 competing scenarios involving three types of dispersal models (scenarios 1–6: any two groups were sequentially from the other group; scenarios 7–12: one group was an admixture of the other two; scenarios 13–18: the three groups dispersed in a stepping stone way) (Supplementary Fig. S1). We tested these scenarios using 22 datasets generated from one population of each group. Comparison of competing scenarios, estimation of posterior distributions of demographic parameters, model checking and evaluation of confidence in scenario choice, and accuracy of parameter estimation were performed following methods of previous studies (Cao et al., 2016b; Wei et al., 2015). Priors of each parameter involving in each scenario are shown in Supplementary Table S1.

Partitioning geographic and climate effects on genetic variation

To evaluate the effect of geographical separation and environmental factors on genetic differentiation of *D. melanogaster*, isolation by distance (IBD) and isolation by environment (IBE) analyses were performed. For IBD analysis, the correlation coefficient between pairwise population differentiation (calculated by $F_{ST}/(1 - F_{ST})$) and geographic distance was estimated using the R package *ade4*, with 10,000 permutations. For IBE analysis, 19 standard bioclimatic variables for each sampling location across China were downloaded from WorldClim database version 2 with a spatial resolution of 10 min, and other parameter settings were identical to those in the IBD analysis.

To examine the extent to which genetic distance based on microsatellite genotypes was explained by geographical and environmental factors and by their interaction, a multivariate redundancy analysis (RDA) was performed. As a constrained linear ordination method, RDA combines multiple linear regression and principal component analysis. The geographic information and climate data for 16

Table 3 Population genetic diversity of *Drosophila melanogaster* collected from China.

Population	Mitochondrial DNA					Microsatellite loci						
	<i>N</i>	<i>H</i>	Hd	Pi	<i>S</i>	<i>A_R</i>	<i>P_{AR}</i>	<i>A_T</i>	<i>A_S</i>	<i>H_o</i>	<i>H_{ET}</i>	<i>F_{IS}</i>
XJKL	17	1	0.000	0.00000	0	3.70	0.04	59	55.49	0.526	0.535	0.018
YNDH	20	2	0.189	0.00036	1	3.72	0.02	64	57.79	0.562	0.535	-0.074
SCPZ	15	4	0.467	0.00098	3	3.76	0.05	60	56.45	0.507	0.510	0.006
QHYN	17	3	0.324	0.00065	2	4.21	0.10	72	63.11	0.553	0.556	0.005
YNKM	14	3	0.473	0.00094	2	4.32	0.07	69	67.35	0.574	0.602	0.031
SCCD	15	3	0.448	0.00090	2	4.40	0.14	81	75.59	0.499	0.575	0.076
SXAK	20	5	0.616	0.00356	10	4.28	0.13	74	64.23	0.620	0.596	-0.041
NMHH	14	2	0.538	0.00103	1	4.09	0.08	64	61.23	0.581	0.553	-0.052
GDGZ	12	3	0.545	0.00405	7	3.93	0.12	59	59.00	0.528	0.535	0.013
HBWH	17	1	0.000	0.00000	0	4.05	0.01	66	60.58	0.549	0.549	-0.001
JXSR	17	2	0.118	0.00022	1	4.05	0.15	67	60.77	0.573	0.541	-0.059
SDLY	29	3	0.135	0.00039	3	4.38	0.16	96	70.94	0.558	0.589	0.013
FJFZ	16	1	0.000	0.00000	0	4.01	0.07	69	63.48	0.544	0.556	-0.014
NMHB	16	2	0.125	0.00048	2	3.98	0.05	75	67.48	0.552	0.593	0.020
ZJHZ	16	1	0.000	0.00000	0	4.22	0.12	70	63.84	0.547	0.576	0.042
LNHL	19	2	0.105	0.00020	1	4.14	0.11	71	62.53	0.514	0.536	0.035

N number of individuals used in the analysis, *H* number of haplotypes, *Hd* haplotype diversity, *Pi* nucleotide diversity, *S* number of polymorphic sites, *A_R* average allelic richness for 12 individuals per population, *P_{AR}* private allelic richness for 12 individuals per population, *A_T* total number of alleles, *A_S* standardized total number of alleles (for 12 individuals), *H_o* observed heterozygosity, *H_{ET}* expected heterozygosity, *F_{IS}* inbreeding coefficient.

different sampling sites were used, as for the IBD and IBE analyses above. Genetic distance between populations (not individuals) was computed. Before RDA analysis, the matrices of pairwise geographic distances were transformed into principal components of neighborhood matrices (PCNM) with the *pcnm* function implemented in the R package *vegan* (<https://github.com/vegandevs/vegan>), and only the first half of positive eigenvectors was used as explanatory variables of genetic variance among populations. To avoid the high collinearity among explanatory variables, a standard value of variance inflation factor (VIF) was set in advance; any candidate variables with VIF value higher than 10 were excluded as described in Cao et al. (2019). This led to three geographic variables (PCNM1, PCNM2, and PCNM4) and five climate variables (bio2, mean diurnal air temperature range; bio3 isothermality, bio5, maximum temperature of warmest month; bio13, precipitation of the wettest month; and bio15, precipitation seasonality) being retained in the model. Both the full RDA (environmental variables and geographical components) and partial RDA (environmental variables or geographical components) analyses were performed using the *vegan* package in R. With a full RDA model, the degree of geography, environment, and their collinear proportion were analyzed, while the independent effect of the environment was estimated using the partial RDA model in which geographic effects were constrained.

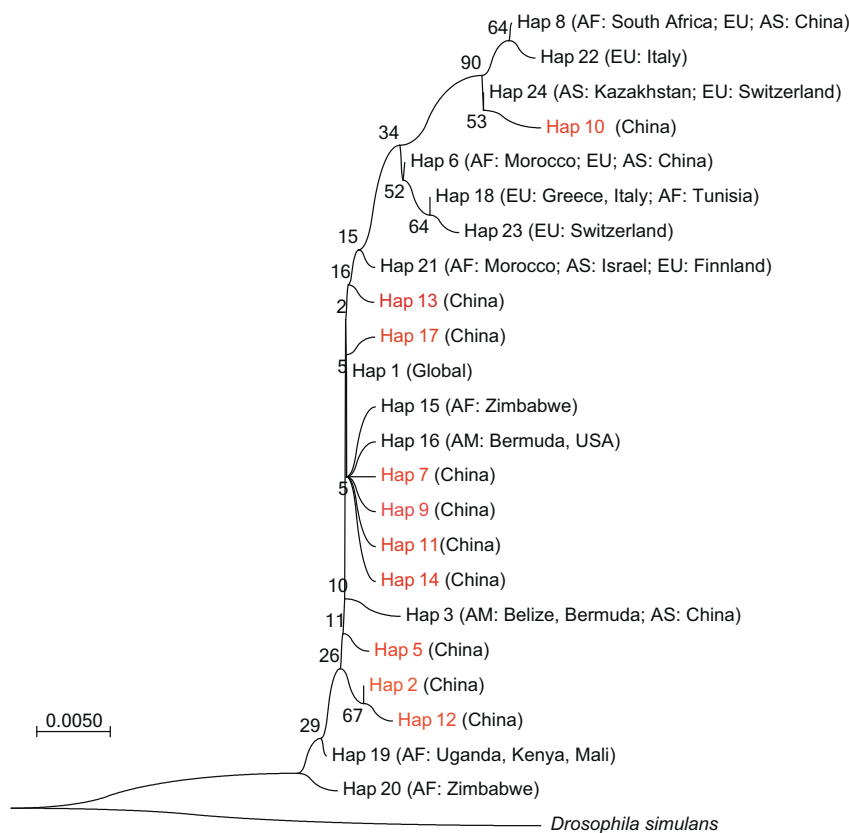
Results

Mitochondrial haplotype diversity and phylogeny

After removing one of the individuals from pairs of relatives (in total, 58 removed), there were 272 individuals available for analysis. Fourteen mitochondrial haplotypes were identified in these individuals from the 16 populations collected across China (Fig. 1). Hap01 was dominant in all the populations. Hap02 was mainly found in populations from western China, excepting XJKL and YNDH. In addition, populations from the western and southwest areas, except for XJKL and YNDH, possessed a higher number of haplotypes than most populations from the central and eastern areas. Populations from SXAK adjacent to the southwestern regions of China showed the highest level of mitochondrial DNA diversity, with the highest values for both haplotype diversity (Hd) and nucleotide diversity (Pi) (Table 3). NMHH and GDGZ populations also exhibited relatively high diversity, as indicated by Hd and Pi.

We related Chinese haplotypes in this study to cosmopolitan haplotypes by including *D. melanogaster cox1* haplotypes from previous reports into the analysis. The results showed that 14 of 23 currently identified mitochondrial haplotypes could be found in China, and 10 of them were unique to China (Fig. 2). The phylogenetic relationships among all haplotypes indicated that Hap12

Fig. 2 Neighbor-joining phylogenetic relationships among mitochondrial *cox1* haplotypes of *Drosophila melanogaster* worldwide. Hap1 to Hap 14 were found in this study, while Hap15 to Hap23 were reported in previous studies. The haplotypes unique to China are in red. *Drosophila simulans* was used as the outgroup.



(unique to YNKM) and Hap 2 (from NMHH, SDLY, SCCD, SXAK, YNKM, SCPZ, QHXN, and Hsinchu of Taiwan) in China were relatively ancestral, following the African haplotypes Hap19 and Hap 20.

Microsatellite marker development and genetic diversity

The QDD3 program provided 63,888 primer pairs, which included 50,412 tri-, 10,070 tetra-, 1651 penta-, and 1755 hexa-nucleotide microsatellites. Following our stringent filter, a total of 60 primer pairs flanking perfect microsatellites were retained. Among them, the number of primer pairs flanking tri-, tetra-, and penta-nucleotide microsatellites was 55, 4, and 1, respectively. In our initial test of these 60 primer pairs, 15 pairs generated polymorphic genotypes, 15 pairs failed to generate amplification in any individuals, and 30 pairs failed in more than two individuals. Finally, 15 microsatellite markers were retained for population-level genotyping (Table 2).

The genetic diversity in 16 populations of *D. melanogaster* from China was evaluated using the 15 developed microsatellite markers. Significant departures from HWE ($p < 0.05$) were detected in 18 of the 240 population-locus combinations after sequential Bonferroni correction.

However, multi-locus tests revealed that none of the 16 populations departed consistently from HWE. In addition, 65 of 1680 locus-locus pairs showed LD in at least one population ($p < 0.05$), whereas 2 (S04–S39 pair on the Chromosome 2L, and S10–S44 pair on the Chromosome 2R) of 105 locus pairs showed LD across all populations. Inversions could produce LD between distantly located loci on the same chromosome, but the two members had not been identified in the same inversion fragment for any LD pairs by referring to the website PopFly (<http://popfly.uab.cat/>) (Hervas et al., 2017). Perhaps new inversions exist in the chromosomes of *D. melanogaster* from China. Genetic diversity parameters (Table 3 and Supplementary Fig. S2) varied among populations, and most of the southwest populations except for YNDH showed relatively high values of diversity, including the average allelic richness (A_R) and the standardized total number of alleles (A_S), when compared to central and eastern populations.

There was no significant correlation in the diversity parameters between microsatellite loci and mitochondrial *cox1*. The association was significant for pairs of parameters estimated from the same type of genetic markers, such as H_d vs P_i ($r = 0.78$, $p < 0.001$), P_i vs S ($r = 0.93$, $p < 0.001$), A_R vs A_S ($r = 0.84$, $p < 0.001$), and A_R vs H_{ET} ($r = 0.74$, $p < 0.001$).

Table 4 Pairwise population differentiation (F_{ST} , lower triangle) and their statistical significance (p values from exact G test, upper triangle) among 16 populations of *Drosophila melanogaster* based on microsatellites.

Population	XJKL	YNDH	SCPZ	QHYN	YNKM	SCCD	SXAK	NMHH	GDGZ	HBWH	JXSR	SDLY	FJFZ	NMHB	ZJHZ	LNHL
XJKL																
YNDH	0.09	<0.01	<0.01	<0.01	<0.01	0.03	<0.01	<0.01	<0.01	<0.01	0.01	<0.01	0.01	<0.01	<0.01	<0.01
SCPZ	0.07	0.02		<0.01	0.22	0.70	0.02	0.02	<0.01	<0.01	0.09	0.10	0.02	<0.01	0.01	<0.01
QHYN	0.04	0.06	0.04		0.07	0.04	0.39	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
YNKM	0.05	0.02	<0.01	0.02		0.52	0.33	0.41	<0.01	<0.01	0.07	0.07	0.04	<0.01	0.09	<0.01
SCCD	0.04	0.01	<0.01	0.02	<0.01		0.48	0.21	<0.01	<0.01	0.45	0.78	0.19	<0.01	0.24	0.04
SXAK	0.05	0.03	0.02	<0.01	<0.01	0.01		0.37	<0.01	<0.01	0.01	<0.01	<0.01	0.15	0.04	<0.01
NMHH	0.08	0.04	0.02	0.04	<0.01	0.01	<0.01		<0.01	<0.01	0.02	0.06	<0.01	<0.01	<0.01	<0.01
GDGZ	0.11	0.05	0.05	0.07	0.05	0.05	0.05	0.07		<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
HBWH	0.05	0.03	0.02	0.03	0.02	0.02	0.02	0.04	0.03		<0.01	<0.01	<0.01	<0.01	0.02	<0.01
JXSR	0.07	0.02	<0.01	0.03	0.01	<0.01	0.01	0.02	0.04	0.02		0.22	<0.01	<0.01	0.04	<0.01
SDLY	0.05	0.02	<0.01	0.02	<0.01	-0.01	0.01	0.01	0.04	0.01	<0.01		0.04	<0.01	0.04	0.02
FJFZ	0.09	0.02	0.02	0.06	0.01	0.01	0.03	0.04	0.04	0.04	0.03	0.02		<0.01	<0.01	<0.01
NMHB	0.06	0.06	0.05	0.03	0.03	0.03	0.01	0.04	0.06	0.03	0.03	0.02	0.07		0.02	<0.01
ZJHZ	0.07	0.03	0.02	0.02	0.01	<0.01	0.01	0.03	0.07	0.03	0.01	<0.01	0.04	0.01		0.02
LNHL	0.06	0.06	0.03	0.03	0.02	0.02	0.04	0.06	0.06	0.04	0.03	0.01	0.04	0.04	0.01	

Population genetic structure

For the microsatellite loci, pairwise F_{ST} values ranged from 0 to 0.11, with an average F_{ST} value of 0.03 (Table 4 and Supplementary Table S2). Among them, XJKL, GDGZ, NMHB, and LNHL showed the highest level of genetic differentiation from the other populations. Contingency tests (based on likelihood ratios) on genotype frequencies across populations revealed that most pairwise F_{ST} estimates were significantly different ($p < 0.05$, Table 4).

STRUCTURE analysis indicated that almost all populations could not be well separated when K increased from 3 to 6, aside from XJKL when $K = 5$ and $K = 6$ (Fig. 3a). Some correlation was found between the proportion of some clusters and the longitudinal distribution of populations (the blue cluster when $K = 5$: $R^2 = 0.449$, $p < 0.001$; the yellow cluster when $K = 6$: $R^2 = 0.386$, $p < 0.001$). DAPC results showed a similar pattern of population differentiation as that revealed by the STRUCTURE analysis, with XJKL and LNHL populations forming outliers and the other populations clumping together (Fig. 3b).

Demographic history of *D. melanogaster* in China

In the DIYABC analysis, scenarios 8, 13, and 14 were supported with the highest posterior probabilities by 4, 5, and 13 out of 22 datasets, respectively (Supplementary Table S3). All of these three scenarios assumed that the NW group was ancestral of *D. melanogaster* in China. The posterior probability in datasets supporting scenarios 8, 13, and 14 ranged from 0.115 to 0.276, from 0.207 to 0.276, and from 0.170 to 0.285, respectively (Supplementary Table S3). The sum of posterior probabilities for all scenarios in each dataset that support the NW (scenarios 1, 2, 7, 8, 13, and 14), CE (scenarios 3, 4, 9, 10, 15, and 16), and SW (scenarios 5, 6, 11, 12, 17, and 18) origin ranged from 0.534 to 0.903, from 0.055 to 0.321, and from 0.032 to 0.182, respectively.

Geographic and climate effects on genetic variance

Mantel tests indicated an extremely significant association between genetic distance and geographical distance ($r = 0.5253$, $p = 0.0014$; Fig. 4a) and environmental distance ($r = 0.344$, $p = 0.009$; Fig. 4b) for all sampling individuals from China.

A full RDA model showed that only 9.1% of the total genetic variance could be explained by the effect of geography, the environment, and their interaction. For the explained genetic variance, independent effects of environmental conditions and geography accounted for 62.10% and 31.58%, respectively, while their collinear component accounted for 6.32%. When environmental

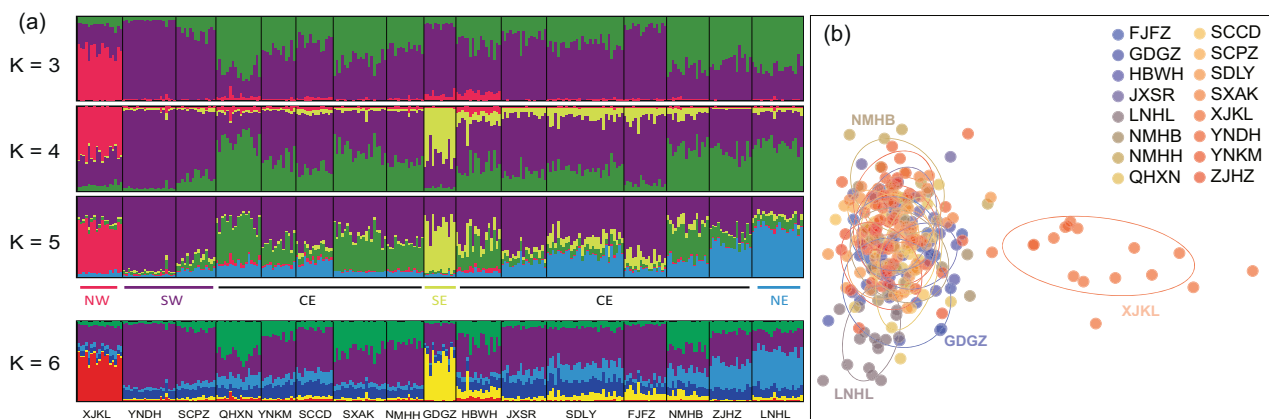


Fig. 3 Population genetic structure of 16 *Drosophila melanogaster* populations from China based on 15 microsatellite loci. **a** Genetic structure inferred from STRUCTURE and **b** DAPC analysis. The optimal number of clusters for the STRUCTURE analysis was five;

$K = 3, 4, 5,$ and 6 are presented (a) with populations arranged from west to east. DAPC analysis shows LNHL, XJKL, GDGZ, and NMHB to be outlier populations (b).

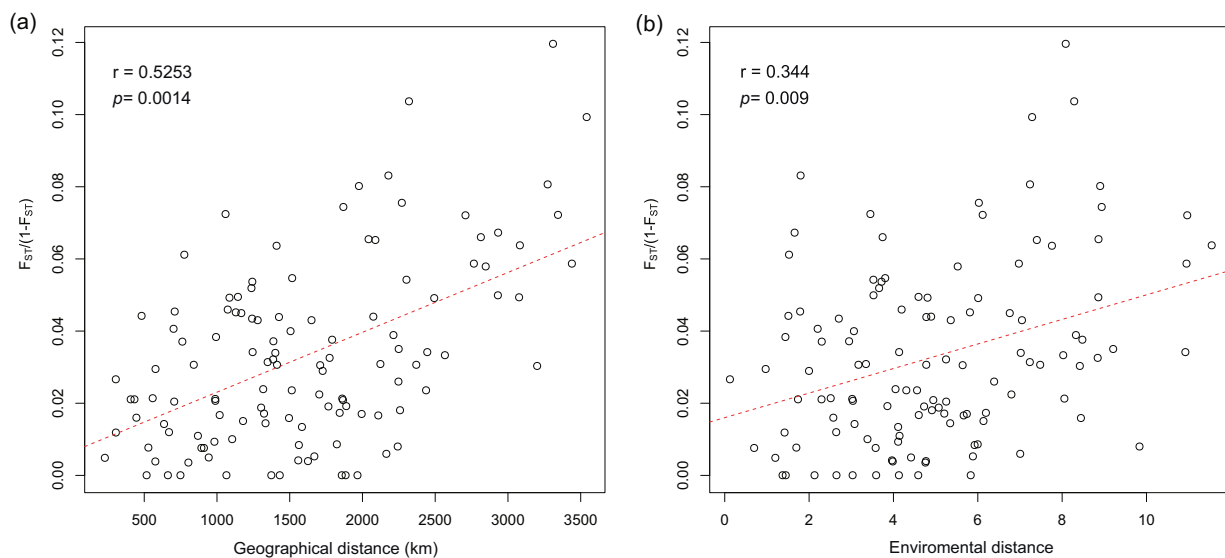


Fig. 4 Scatter plots of isolation by geographical distance and environment for all populations of *Drosophila melanogaster* in China. **a** Correlations between population genetic differentiation ($F_{ST}/(1-F_{ST})$), geographical distance, and **b** environmental distance

analyzed by Mantel tests Genetic differentiation was calculated based on 15 microsatellite loci. Correlation coefficients (r) and p values are given.

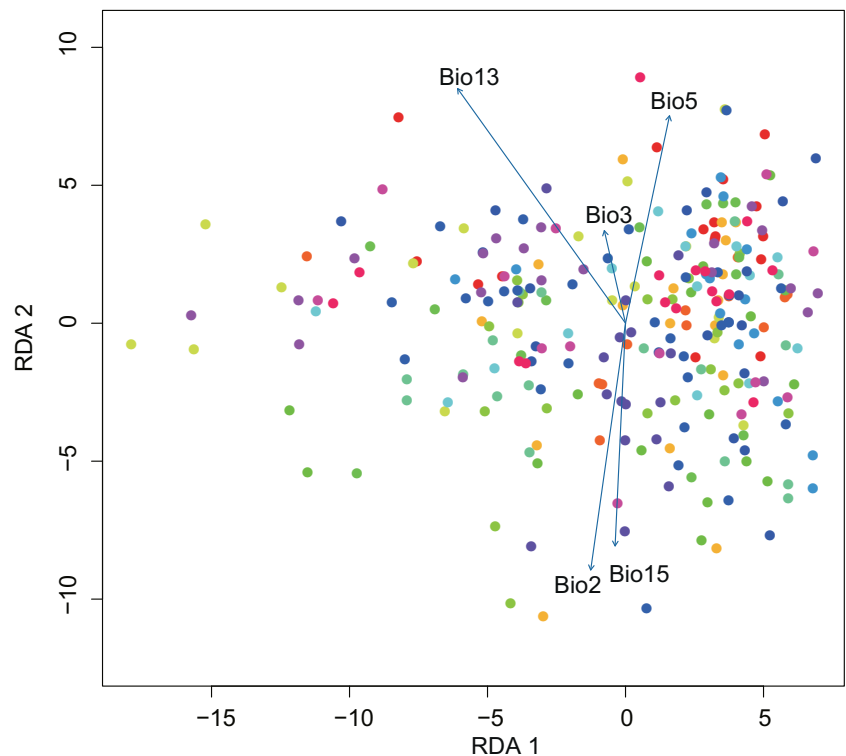
and geographic effects were considered simultaneously in the full RDA analysis, two climatic variables (bio3 (isothermality) and bio13 (precipitation of the wettest month)) and two geographic variables (PCNM3 and PCNM4) were correlated with genetic distance (Supplementary Fig. S3). When geographic variables were constrained in the RDA analysis, three environmental variables (bio2 (mean diurnal air temperature range), bio13 (precipitation of the wettest month), and bio15 (precipitation seasonality)) were correlated with genetic distance (Fig. 5).

Discussion

Distribution of genetic diversity in *D. melanogaster* from China

The mitochondrial data in this study showed that Chinese populations had a large number of unique haplotypes, and some of them were previously undescribed (Fig. 2). The high differentiation between Asian *D. melanogaster* and populations from other parts of the world has been noted in previous studies (Hale & Singh, 1991; Nunes et al., 2008a; Schlötterer

Fig. 5 Partial RDA analysis on genetic variance explained environmental variables when geographical variables were constrained. Individuals from the same population are represented by dots with the same color. The blue vectors are the environmental predictors used in this study, and arrows indicate the strength of correlations between the environmental variable and genetic variance. Five climate variables were included in analysis: bio2 (mean diurnal air temperature range), bio3 (isothermality), bio5 (maximum temperature of warmest month), bio13 (precipitation of the wettest month), and bio15 (precipitation seasonality).



et al., 2006). A founder event (bottleneck) occurring in the early phase of colonization may be one reason for the high genetic divergence between Chinese populations and other populations. A “Far Eastern Race” has previously been theorized to explain differentiation in *D. melanogaster* between Asian and other populations, reflecting independent colonization events during the out-of-Africa habitat expansion of *D. melanogaster* into the Asian area (David & Capy, 1988). However, this theory has been questioned because of a lack of genomic evidence from Southeast Asian samples (Laurent et al., 2011). Further analyses that include multiple Asian and African populations from across these regions are needed to test these historical connections (Arguello et al., 2019).

In addition, both the mitochondrial variation and microsatellite data showed that, with a few exceptions, southwest China populations had relatively high levels of genetic polymorphism as measured by haplotype diversity (H_d), nucleotide diversity (p_i), average allelic richness (A_R), and standardized total number of alleles (A_S), when compared to most populations from eastern regions of China (Table 3). Further studies by including samples from and outside China may help to adequately examine this pattern of genetic diversity as well as its underlying evolutionary processes.

Population genetic structure of *D. melanogaster* in China

Pairwise F_{ST} results among the sampled populations revealed relatively low levels of population differentiation, with an

average value of 0.031 (Table 4). The low-level divergence among *D. melanogaster* populations within China might be attributed to a lack of time for genetic drift to occur in the absence of strong bottlenecks. Similarly, low levels of differentiation have also been documented within the European population and between populations on different continents that may have recent European ancestry (e.g. Australia, North America) (Nunes et al., 2008a, b; Kapun et al., 2020). In contrast, higher levels of differentiation have usually been found between populations from southeast Asia and Oceania (Schlötterer et al., 2006). Nevertheless, we were able to find some differentiation among populations. In particular, XJKL, GDGZ, NMHB,1 and LNHL, which were from the westernmost, southernmost, northernmost, and easternmost collection points in the current study, showed the highest genetic differentiation from other populations (Supplementary Table S2).

Demographic history of *D. melanogaster* in China

Our DIYABC analysis showed that scenarios of the northwestern origin of the *D. melanogaster* in China have the highest posterior probabilities. Although DIYABC analysis has proven to be a powerful approach in exploring the genetic relationship among the sampled or unsampled populations by comparing potential competing scenarios (Bertorelle et al., 2010), it should be acknowledged that limited coverage of sampling and a lack of potential source populations may bias the results. We found that the posterior probability for the best-fit scenario in each dataset was very low in our study.

This is our best estimate, but it provides weak support. Thus, our results cannot be used to reach a robust conclusion, especially in the absence of additional samples from the northwest and outside of China. Additional populations should be further sampled in an ensuing study to validate the dispersal history of *D. melanogaster* into and within China.

Ecological adaptation of *D. melanogaster* in China

Despite low F_{ST} s (Table 4), we did find a potential longitudinal pattern accompanied by admixture (blue cluster when $K = 5$, yellow cluster when $K = 6$ in Fig. 3a; Fig. 4b), and RDA analysis suggests that precipitation may have had an impact on patterns of genetic variation in Chinese populations of *D. melanogaster* (Fig. 5). Consistent with the RDA results, the XJKL and LNHL populations, which came from the driest and wettest regions among our sampling locations, had the highest levels of differentiation from other populations, as revealed by the DAPC and F_{ST} analyses. Adult *D. melanogaster* is sensitive to ambient humidity, and precipitation is usually regarded as one of the most important ecological factors limiting the distribution ranges of this species in nature (Gibbs et al., 1997; Hoffmann & Parsons, 1993). Laboratory selection in *D. melanogaster* can lead to rapid adaptation to desiccation stress (Telonis-Scott et al., 2016). Previously published work has shown genetically determined clinal variation in *D. melanogaster* along precipitation gradients (Hoffmann et al., 2001; Karan & Parkash, 1998; Parkash & Munjal, 2000). Our results may, therefore, partly reflect strong selection and adaptation in natural populations of *D. melanogaster* assuming some linkage between the microsatellite markers and traits under selection, which would be enhanced for microsatellite markers located within *D. melanogaster* inversion polymorphisms that respond rapidly to clinal selection (Kennington et al., 2006). Furthermore, an association between environmental factors and genetic structure over and above geographical distance could reflect patterns of genetic exchange driven by environmental conditions, which in themselves might reflect trade, given that regions with similar climates should support similar types of fruit production. Given the overall rapid decay of LD in parts of the *D. melanogaster* genome, these patterns need to be further investigated with high-density markers such as SNPs. Sampling will need to be robust because F_{ST} values were low and because the factors we considered here only accounted for around 9% of the genetic variance among populations.

Conclusion

In this study, we considered the population genetic structure of *D. melanogaster* across China. We found low levels of

genetic differentiation, high levels of admixture, and isolation by geographical and environmental factors in this species. China covers a range of climatic conditions within a complex topography where there has been a long history of human colonization. This context provides an opportunity for further research with *D. melanogaster* populations, such as in linking putative adaptive polymorphisms to climatic selection (Hoffmann et al., 2003) and in understanding the role of inversions in local adaptation by testing for parallel patterns across different continents (Kapun et al., 2018).

Data availability

Mitochondrial genes and microsatellite data generated in this study were deposited to the Dryad database: <https://doi.org/10.5061/dryad.37pvmcvgv>.

Acknowledgements We thank De-Qiang Pu, Qian Li, Xiang-Zhao Yue, Kai Liu, Jia-Ying Zhu, Li-Na Sun, and Jin-Yu Li for their help on the collection of the samples. This research was funded by the National Natural Science Foundation of China (32070464), the Joint Laboratory of Pest Control Research Between China and Australia (Beijing Municipal Science & Technology Commission, 201100008320013), Beijing Postdoctoral Research Foundation (2020-22-108), and Beijing Key Laboratory of Environmental Friendly Management on Pests of North China Fruits (BZ0432).

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Arguello JR, Laurent S, Clark AG (2019) Demographic history of the human commensal *Drosophila melanogaster*. *Genome Biol Evolution* 11(3):844–854
- Arguello JR, Cardoso-Moreira M, Grenier JK, Gottipati S, Clark AG, Benton R (2016) Extensive local adaptation within the chemosensory system following *Drosophila melanogaster*'s global expansion. *Nat Rev Commun* 7:11855
- Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes* 7(1):15–17
- Bergland AO, Tobler R, González J, Schmidt P, Petrov D (2016) Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. *Mol Ecol* 25(5):1157–1174
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol* 19(13):2609–2625
- Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Mol Ecol Resour* 12(3):456–463
- Bradburd GS, Ralph PL (2019) Spatial population genetics: it's about time. *Annu Rev Ecol, Evolution, Syst* 50(1):427–449

- Cao LJ, Wei SJ, Hoffmann AA, Wen JB, Chen M (2016b) Rapid genetic structuring of populations of the invasive fall webworm in relation to spatial expansion and control campaigns. *Diversity Distrib* 22(12):1276–1287
- Cao LJ, Li ZM, Wang ZH, Zhu L, Gong YJ, Chen M et al. (2016a) Bulk development and stringent selection of microsatellite markers in the western flower thrips *Frankliniella occidentalis*. *Sci Rep* 6:26512
- Cao LJ, Gao YF, Gong YJ, Chen JC, Chen M, Hoffmann A et al. (2019) Population analysis reveals genetic structure of an invasive agricultural thrips pest related to invasion of greenhouses and suitable climatic space. *Evolut Appl* 12(10):1868–1880
- Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R et al. (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30(8):1187–1189
- David JR, Capi P (1988) Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* 4(4):106–111
- Duchen P, Živković D, Hutter S, Stephan W, Laurent S (2013) Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* 193(1):291–301
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4(2):359–361
- Flatt T (2016) Genomics of clinal variation in *Drosophila*: disentangling the interactions of selection and demography. *Mol Ecol* 25(5):1023–1026
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 3(5):294–299
- Fraimout A, Debat V, Fellous S, Hufbauer RA, Foucaud J, Pudlo P et al. (2017) Deciphering the Routes of invasion of *Drosophila suzukii* by Means of ABC Random Forest. *Mol Biol Evolution* 34(4):980–996
- Gibbs AG, Chippindale AK, Rose MR (1997) Physiological mechanisms of evolved desiccation resistance in *Drosophila melanogaster*. *J Exp Biol* 200(12):1821–1832
- Goudet J (2017) FSTAT (Version 1.2): a computer program to calculate F-statistics. *J Heredity* 86(6):485–486
- Hale LR, Singh RS (1991) A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history vs. selection in the genetic structure of geographic populations. *Genetics* 129(1):103–117
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2(4):618–620
- Hervas S, Sanz E, Casillas S, Pool JE, Barbadilla A (2017) PopFly: the *Drosophila population* genomics browser. *Bioinformatics* 33(17):2779–2780
- Hoffmann AA, Parsons P (1993) Selection for adult desiccation resistance in *Drosophila melanogaster*: fitness components, larval resistance and stress correlations. *Biol J Linn Soc* 48(1):43–54
- Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion: a critical look at the temperate-tropical clines in *Drosophila melanogaster* from eastern Australia. *Genetica* 129(2):133
- Hoffmann AA, Sørensen JG, Loeschcke V (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J Therm Biol* 28(3):175–216
- Hoffmann AA, Hallas R, Sinclair C, Mitrovski P (2001) Levels of variation in stress resistance in *Drosophila* among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold resistance, and associated traits. *Evolution* 55(8):1621–1630
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801–1806
- Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers *Bioinformatics* 24:1403–1405
- Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* 5(1):187–189
- Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D (2015) Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from the south-eastern United States and Caribbean Islands. *Mol Ecol* 24(7):1499–1509
- Kapun M, Barrón MG, Staubach F, Vieira J, Obbard DJ, Goubert C, et al. (2018). Genomic analysis of European *Drosophila* populations reveals longitudinal structure and continent-wide selection. *bioRxiv*: 313759
- Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J et al. (2020) Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol Biol Evol* 37(9): msaa120
- Karan D, Parkash R (1998) Desiccation tolerance and starvation resistance exhibit opposite latitudinal clines in Indian geographical populations of *Drosophila kikkawai*. *Ecol Entomol* 23(4):391–396
- Kauer M, Dieringer D, Schlotterer C (2003) Nonneutral admixture of immigrant genotypes in African *Drosophila melanogaster* populations from Zimbabwe. *Mol Biol Evolution* 20(8):1329–1337
- Keller A (2007) *Drosophila melanogaster's* history as a human commensal. *Curr Biol* 17(3):77–81
- Kennington WJ, Partridge L, Hoffmann AA (2006) Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion In (3R) Payne in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* 172(3):1655–1663
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evolution* 33(7):1870–1874
- Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup *Evolutionary Biology*. Springer, pp 159–225
- Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evolution* 29(6):1695–1701
- Laurent SJ, Werzner A, Excoffier L, Stephan W (2011) Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol Biol Evolution* 28(7):2041–2051
- Leipe C, Long T, Sergusheva E, Wagner M, Tarasov P (2019) Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Sci Adv* 5(9):eaax6225
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452
- Lombaert E, Guillemaud T, Lundgren J, Koch R, Facon B, Grez A et al. (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird *Harmonia axyridis*. *Mol Ecol* 23(24):5979–5997

- Lowe WH, Kovach RP, Allendorf FW (2017) Population genetics and demography unite ecology and evolution. *Trends Ecol Evolution* 32(2):141–152
- Markow TA, O'Grady P (2005) *Drosophila*: a guide to species identification and use. Elsevier
- Meglecz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N et al. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26(3):403–404
- Miller AD, Coleman MA, Clark J, Cook R, Naga Z, Doblin MA, et al. (2020). Local thermal adaptation and limited gene flow constrain future climate responses of a marine ecosystem engineer. *Evolutionary Applications* eva.12909
- Molina J, Sikora M, Garud N, Flowers JM, Purugganan MD (2011) Molecular evidence for a single evolutionary origin of domesticated rice. *J Agric Biotechnol* 108(20):8351–8356
- Nunes MD, Nolte V, Schlotterer C (2008b) Nonrandom *Wolbachia* infection status of *Drosophila melanogaster* strains with different mtDNA haplotypes. *Mol Biol Evolution* 25(11):2493–2498
- Nunes MD, Neumeier H, Schlotterer C (2008a) Contrasting patterns of natural variation in global *Drosophila melanogaster* populations. *Mol Ecol* 17(20):4470–4479
- van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4(3):535–538
- Park S (2001). Animal Genomics Laboratory, UCD, Ireland.
- Parkash R, Munjal AK (2000) Evidence of independent climatic selection for desiccation and starvation tolerance in Indian tropical populations of *Drosophila melanogaster*. *Evolut Ecol Res* 2(5):685–699
- Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW et al. (2012) Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet* 8(12):e1003080
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes* 4(1):137–138
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 8(1):103–106
- Schlötterer C, Neumeier H, Sousa C, Nolte V (2006) Highly structured Asian *Drosophila melanogaster* populations: A new tool for hitchhiking mapping? *Genetics* 172(1):287–292
- Schmidt PS, Matzkin L, Ippolito M, Eanes WF (2005) Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. *Evolution* 59(8):1721–1732
- Semenov GA, Safran RJ, Smith CCR, Turbek SP, Mullen SP, Flaxman SM (2019) Unifying theoretical and empirical perspectives on genomic differentiation. *Trends Ecol Evolution* 34(11):987–995
- Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68(1):1–15
- Sigh RS, Hickey DA, David J (1982) Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. *Genetics* 101(2):235–256
- Smith AL, Hodkinson TR, Villellas J, Catford JA, Csergo AM, Blomberg SP et al. (2020) Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proc Natl Acad Sci USA* 117(8):4218–4227
- Sprengelmeyer QD, Mansourian S, Lange JD, Matute DR, Cooper BS, Jirle EV et al. (2020) Recurrent collection of *Drosophila melanogaster* from wild African environments and genomic insights into species history. *Mol Biol Evolution* 37(3):627–638
- Stephan W, Li H (2007) The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity (Edinb)* 98(2):65–68
- Stone GN, White SC, Csóka G, Melika G, Mutun S, Péntzes Z et al. (2017) Tournament ABC analysis of the western Palaearctic population history of an oak gall wasp, *Synergus umbraculus*. *Mol Ecol* 26(23):6685–6703
- Telonis-Scott M, Sgrò CM, Hoffmann AA, Griffin PC (2016) Cross-study comparison reveals common genomic, network, and functional signatures of desiccation resistance in *Drosophila melanogaster*. *Mol Biol Evolution* 33(4):1053–1067
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673–4680
- Weeks AR, McKechnie SW, Hoffmann AA (2002) Dissecting adaptive clinal variation: markers, inversions and size/stress associations in *Drosophila melanogaster* from a central field population. *Ecol Lett* 5(6):756–763
- Wei SJ, Cao LJ, Gong YJ, Shi BC, Wang S, Zhang F et al. (2015) Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. *Mol Ecol* 24(16):4094–4111
- Xiang H, Liu X, Li M, Yn Zhu, Wang L, Cui Y et al. (2018) The evolutionary road from wild moth to domestic silkworm. *Nat Ecol Evolution* 2(8):1268–1279
- Yu Y, Fu J, Xu Y, Zhang J, Ren F, Zhao H et al. (2018) Genome re-sequencing reveals the evolutionary history of peach fruit edibility. *Nat Commun* 9(1):5404