Genomic Diversity Illuminates the Species History and Environmental Adaptation of *Drosophila suzukii*

Siyuan Feng[1], Samuel P. DeGrey[2†], Christelle Guédot[2], Sean D. Schoville[2], John E. Pool[1*]

[1]Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA

[2] Department of Entomology, University of Wisconsin-Madison, Madison, WI, USA

[†]Present address: Kimberly Research and Extension Center, University of Idaho, Kimberly, ID, USA

[*]Corresponding authors - E-mail: siyuanfeng.bio@outlook.com, jpool@wisc.edu

Abstract

Biological invasions carry substantial practical and scientific importance, and represent natural evolutionary experiments on contemporary timescales. Here, we investigated genomic diversity, invasion history, and environmental adaptation of the crop pest *Drosophila suzukii* using whole-genome sequencing data and environmental metadata for 29 population samples from its native and invasive range. Our analysis of genome-wide variation revealed maximal diversity in a population from eastern China, with other populations largely containing a subset of the genetic diversity present in this sample, consistent with an ancestral or refugial species range encompassing this region of Asia. Our analyses of genomic diversity and population history recapitulated some previously suggested dynamics of genetic structure, invasion bottlenecks, and admixture events. However, we suggest that clearer inferences of the geographic origins of worldwide invasions will require denser geographic sampling of genetic variation, particularly in Asia. Using a genome-environment association approach, we detected genetic signals of local adaptation associated with nine distinct environmental factors related to altitude, wind speed, precipitation, temperature, and human land use. We detected some unique functional signatures for each environmental variable, such as a prevalence of cuticular genes associated with annual precipitation. We also inferred biological commonalities in the adaptation to diverse selective pressures, particularly in terms of the apparent contribution of nervous system evolution to enriched processes (ranging from neuron development to circadian behavior) and/or top genes associated with all nine environmental variables. Our findings provide insights into the invasion history of *D. suzukii* and depict a finer-scale adaptive landscape underlying this species' invasion success.

Introduction

One of the main goals of ecological and evolutionary genomics is to understand how organisms evolve in response to novel environments. Biological invasions, while often ecologically and economically damaging, represent unique opportunities to build our understanding of local adaptation, as natural experiments that expose introduced species to new biotic and abiotic factors on contemporary time scales (Lee 2002; Prentis et al. 2008; Colautti and Lau 2016). Invasive species can exhibit rapid phenotypic and genetic changes during the invasion process, driven by various evolutionary mechanisms such as selection, drift, mutation, and gene flow (Colautti and Lau 2016; Hodgins et al. 2018). These changes can result in the adaptive evolution of invasive populations to the novel environments they encounter (Colautti and Barrett 2013). Although there have been emerging studies on the evolutionary biology of invasive species in recent years, the source and nature of the genetic variation underlying such adaptation are still not well characterized (Reznick et al. 2019; Welles and Dlugosch 2019).

Drosophila suzukii, also known as spotted wing *Drosophila*, is a promising model for studying adaptive evolution during invasions. *Drosophila suzukii* is a highly polyphagous vinegar fly that originated from Asia. It first expanded to Hawaii in 1980, and has recently invaded North America and Europe, and subsequently Réunion Island (Indian Ocean) and South America since the late 2000s (Asplen et al. 2015). *Drosophila suzukii* differs from other *Drosophila* species in its unique ability to oviposit on both unripe and ripe fruits, using its serrated ovipositor to pierce the skin of soft-skinned fruits. This has allowed it to exploit a novel ecological niche and avoid competition with other vinegar flies that typically feed on overripe and rotting fruits (Cini et al. 2012; Atallah et al. 2014), causing severe economic losses to fruit crops (Knapp et al. 2021). It also exhibits remarkable phenotypic plasticity and genetic diversity, which may facilitate its

adaptation to different climatic conditions and host plants (Gibert et al. 2019; Little et al. 2020; Olazcuaga et al. 2020). To obtain a comprehensive evolutionary genetic understanding of the invasion history of *D. suzukii*, we need to understand both the demography of this species (including the history and structure of populations from Asia and introduced regions) and the genetic basis and ecological drivers of adaptive evolutionary changes that have allowed this species to occupy diverse worldwide environments.

Multiple genetic studies have investigated the demographic history of invasive populations of *D. suzukii*. Adrion et al. (2014) analyzed six X-linked gene fragments from populations in the continental US, Hawaii, Japan, and Spain. They detected differentiation signals between European, Asian, and US populations, but not among US populations, and suggested independent invasions into Europe and the continental US. Using data from 25 microsatellite markers, Fraimout et al. (2017) further characterized invasion bottlenecks and inferred more comprehensive worldwide invasion scenarios of *D. suzukii*, using an approximate Bayesian computation random forest (ABC-RF) approach. The resulting model, which was constrained by the observed chronology of the species' invasion, indicated that some invading populations had multiple genetic sources. A more recent population genomic study that focused on samples from the US, as well as single sites in Brazil, Ireland, Italy, Japan, and Korea, used autosomal single nucleotide polymorphisms (SNPs) to infer population structure and admixture (Lewald et al. 2021). Their model mostly agrees with the above microsatellite analysis, except for an admixture event from the western US back to Asia. Collectively, these studies suggest some emerging consensus about the invasion history of *D. suzukii*, along with some lingering uncertainty about admixture during this process. However, the limited study of Asian *D. suzukii*

77 samples has left the history of the species within this continent less clear, along with the precise
78 geographic origins of invading populations.

79 In contrast, the genetic basis of environmental adaptation in *D. suzukii* is still largely
80 unexplored. Among the few relevant studies is that of Olazcuaga et al. (2020), which used
81 genomic sequencing data from 22 worldwide populations to search for SNPs with greater
82 frequency differences between Asian (China, Japan) and non-Asian populations than are
83 observed at most loci due to founder event bottlenecks, in hopes of identifying genetic variants
84 that may underlie the invasion success of introduced populations. A subsequent study examining
85 this same data set also found a small number of transposon insertions with strong frequency
86 differences between American/European and Asian populations (Mérel et al. 2021). Another
87 study identified F_{ST} outliers among Hawaiian *D. suzukii* populations (Koch et al. 2020). Apart
88 from these limited comparisons, the genetic changes that may have helped *D. suzukii* to adapt to
89 diverse worldwide environments are entirely unknown.

90 With the increasing availability of genomic resources from non-model species, genotype-
91 environment association (GEA), also known as environmental association analysis (EAA), is
92 becoming a widely-used approach to understand the potential relationship between specific
93 environmental factors and adaptive genetic variation (Rellstab et al. 2015). GEA is also useful in
94 identifying subtle changes in allele frequencies that are difficult to detect with outlier tests based
95 on traditional population genomic approaches, especially when the number of studied
96 populations is relatively large, and there is high gene flow counteracting patterns of local
97 adaptation (Kawecki and Ebert 2004). The capability of GEA to identify adaptive genetic
98 changes and environmental drivers of local adaptation has been demonstrated with whole-
99 genome pool-seq data from *Drosophila melanogaster* (Bogaerts-Márquez et al. 2021). Therefore,

GEA could be helpful in understanding the genotype-environment relationships underlying the invasion success of *D. suzukii*.

In the present study, we perform population genomic analyses on whole-genome pool-seq data of 29 population samples from native and invasive ranges to investigate the genomic diversity, invasion history and environmental adaptation of *D. suzukii*. We investigate the geographic pattern of genetic diversity to obtain clues regarding the history of this species within Asia. We investigate population structure and perform demographic inference, assessing agreement with published models of invasion history. Finally, we test the association between SNP frequencies and nine environmental variables across sampling locations, identifying both specific and shared functional signatures of adaptation to these diverse selective pressures.

Results

Genomic Diversity Reaches its Maximum in Eastern China

We investigated the genomic polymorphism of 29 *D. suzukii* populations derived from Asia (n=8), Europe (n = 11), and North and South America (n = 10), and Asia (n = 8), encompassing both newly reported and previously published samples (Figure 1A; Table S1; Olazcuaga et al. 2020). Whole-genome sequences were obtained from 29 pooled samples consisting of 50-212 female and male individuals. The depth of mapped reads after quality control ranged from 23X to 66X among population samples, with an average of 45X (Table S1).

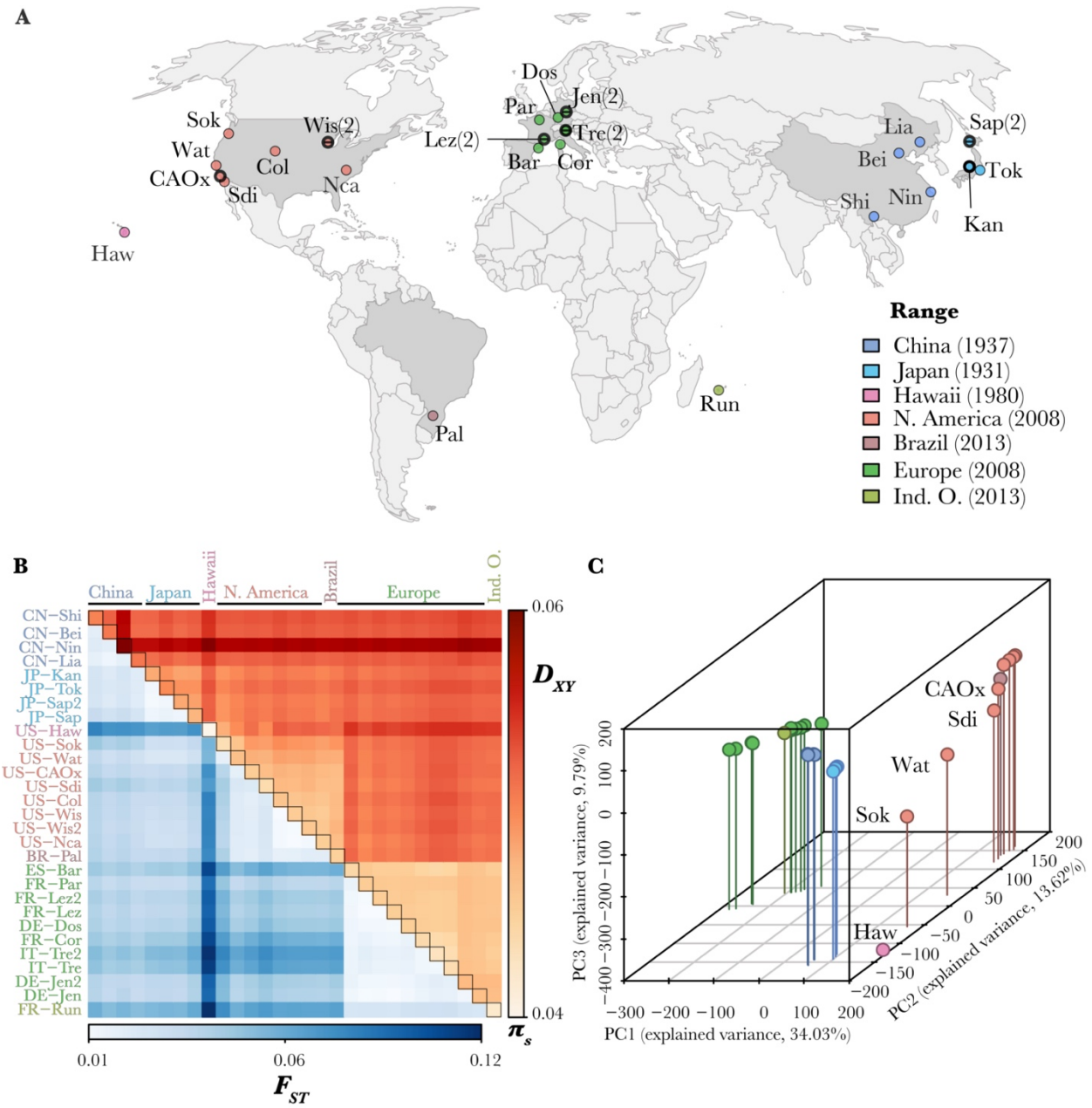
We first estimated nucleotide diversity (π_S) across synonymous SNPs to investigate the effects of rapid invasions on neutral genetic diversity. The genome-wide π_S ranged from a

minimum of 0.041 (autosomes) and 0.019 (X chromosome) in Hawaii, US (US-Haw) to a maximum of 0.059 (autosomes) and 0.045 (X chromosome) in Ningbo, China (CN-Nin). The acute drop in π_S of introduced European and American populations relative to that of the native Chinese and Japanese populations reflects previously reported founder event bottlenecks (Figure 1B, diagonal; Figure S1, S2). The observed patterns were also recapitulated with SNPs at all types of sites (Figure S3). Lower π_S and more contrasting among-population π_S differences in X chromosome reflect effectively prolonged bottlenecks due to their lower effective population size (Pool & Nielsen, 2007; Figure S1). We also observed a greater loss of rare alleles in the invasive populations as a typical consequence of bottlenecks (Figure S4). These founder events also increased genetic differentiation as measured by F_{ST}, especially between continents (Figure 1B; Figure S3), and particularly for the X chromosome (Figure S2).

Previous studies have hypothesized a relatively wide native range of *D. suzukii* in East and Southeast Asia (Adrion et al. 2014; Fraimout et al. 2017). We examined genetic variation within and among Asian populations to search for clues regarding the history of *D. suzukii* within its native continent. The Ningbo population in eastern China had the highest π_S – 15.6 % greater than any other population for the autosomes, and 30.6 % greater for the X chromosome (Figure 1B; Figure S1). Notably, D_{XY} values involving the Ningbo population did not meaningfully exceed the magnitude of Ningbo's π, indicating that all other investigated locations primarily have a subset of the genetic diversity present in Ningbo. These findings are consistent with the other sampled populations descending in relatively recent population genetic time from a Ningbo-like ancestor, and eastern China being within or closer to the ancestral range of the species.

Population Structure and Invasion History of *D. suzukii*

Principal component analysis (PCA) of allele frequencies was next performed to offer insights into the population structure of *D. suzukii*. The top three principal components (PC 1-3) explained 57.44% (autosomes, Figure 1C) and 72.35% (X chromosome, Figure S2B) of the variance among populations. In both autosomes and X, three-dimensional PCA and matrices of both D_{XY} and window F_{ST} recapitulated both continuous and hierarchical geographic structure. These results together showed the expected clustering of populations into four distinct ranges (East Asia, Hawaii, Americas, and Europe; Figure 1B, 1C; Figure S2). Much of the observed population differentiation is most likely due to founder event bottlenecks and admixture during worldwide expansion (Fraimout et al. 2017), whereas migration following population establishment would need to be overwhelmingly high to have significant impacts given the very brief time scale of the global invasion. The two more northerly populations from the western US, Oregon (US-Sok) and central California (US-Wat), show an affinity with the Hawaiian population, which aligns with the suggestion that these populations received a genetic contribution from Hawaii in addition to East Asia (Fraimout et al. 2017), while populations from southern California and the central and eastern US show less evidence of such admixture. These findings are also reflected by the clustering patterns in F_{ST} -based neighbor-joining trees (Figure S5).



163 Figure 1. *Drosophila suzukii* populations show maximal diversity in eastern China and continent-level genetic
164 structure. (A) The geographic locations of the studied 29 natural populations are depicted as dots. In addition to the
165 22 populations sampled by Olazcuaga et al. (2020), populations newly sampled at independent locations are circled
166 in black. Populations newly sampled at nearby locations are circled and center-dashed in black, with the number of
167 total population samples in brackets. The year of the first recorded occurrence in each geographic range (colored
168 grey in the map) is given in brackets in the color legend. Further information about each sample is presented in
169 Table S1. (B) Population differentiation in allele frequencies (F_{ST}; lower triangle), between-population sequence

distances (D_{XY}; higher triangle), and within-population nucleotide diversity (π_S; diagonal) across autosomal synonymous SNPs are displayed as a heatmap. Population names are colored by their geographic region. (C) Autosomal genetic structure is shown by three-dimensional principal components analysis (PCA) based on allele frequencies of the two most frequent alleles across all populations. Each dot represents a population. Labeled are Hawaii and western coastal US populations. See the X chromosome version of (B) and (C) in Figure S2.

To provide a complementary understanding of the invasion history from genome-wide information, we used TreeMix (Pickrell and Pritchard 2012) to build maximum-likelihood trees with admixture graphs based on allele counts of SNPs from autosomes and X chromosome, separately (Figure 2; Figure S7). We rooted the tree along the branch leading to the Ningbo population (CN-Nin), since the above diversity analysis suggested that it could reflect the most basal lineage among the investigated populations. We also sequentially added a range of migration events (0 - 20) to trees to capture gene flow and parsimoniously improve the model fit, and found that tree models with 11 mixture events explain most of the variance in relatedness between populations (autosomes: 99.6%, X chromosome: 99.9%), while remaining interpretable for both autosomes and X chromosome, with greater numbers of migration events resulting in diminishing model improvements (Figure 2; Figure S6).

The topology of our TreeMix histories generally agreed with previous inferences (Adrion et al. 2014; Fraimout et al. 2017; Lewald et al. 2021) and with our above results, including the presence of clear genetic groupings from China, Japan, the United States, and Europe. The tree structure largely agreed between our autosomal and X chromosome trees, and showed general consistency across models where differing numbers of migration events were estimated. One exception concerned the differential placement of the American clade. The autosomal tree grouped the American and European lineages, which split from an Asian lineage closer to the

194 sampled Japan populations than to China. The X chromosome tree instead placed the
195 American group within the sampled Japan populations. Both results differ from the model of
196 Fraimout et al. (2017), which utilizes source populations from China for both continental
197 invasions, although each of our trees does include a specific migration event from China, and the
198 published model did entail ancestry from Japan to the mainland US via Hawaii.

199 The placement of the Hawaii population in both our autosomal and X chromosome trees
200 departed from the observed chronology of the species' expansion and from the model estimated
201 by Fraimout et al. (2017), which was conditioned on observed invasion dates. In both of our
202 trees, Hawaii was placed as an offshoot of the mainland US group, albeit with considerable drift
203 (Figure 2), even though *D. suzukii* was observed in Hawaii about 30 years sooner than on the
204 mainland. Although our trees do entail admixture from Hawaii into southern California
205 populations, the previously proposed history of Hawaii also contributing higher levels of
206 admixture into the Oregon (US-Sok) and central California (US-Wat) populations (versus Hawaii
207 arising from a population similar to those two) is easier to reconcile with observed invasion
208 history. These results highlight the challenges in estimating population history from genetic data
209 alone, from a finite number of populations, and in the presence of substantial drift and admixture
210 along some lineages.

211 The autosomal and X chromosome trees also differed in their placement of the
212 population from Réunion (an Indian Ocean island that is part of France). The autosomal tree
213 placed Réunion within the American clade, but with substantial bidirectional migration between
214 this population and European samples. Whereas, the X chromosome tree placed Réunion
215 alongside a population from France, which is more consistent with the European origin estimated
216 by Fraimout et al. (2017).

The autosomal tree also indicated an admixture event from western US back to Japan with an admixture proportion of 32% (Figure 2A), consistent with a previous inference of western US/Asia admixture using a larger number of US populations (Figure 4 of Lewald et al. 2021). Other instances of migration events from a more recent into an older population on a different continent were also suggested from Réunion to Europe (from the autosomal tree discussed above) and from Brazil to three separate French population samples based on the X chromosome tree (Figure 2B). Whereas some "migration events" might actually reflect independent invasions into the same continent, followed by secondary contact during the range expansion process, these "reverse migrations" cannot be explained in the same way. It is not clear that neutral migration between continents within just a few years should really be expected to displace any noticeable portion of an established population's gene pool. Alternatively, the precise accuracy of the TreeMix models may be limited by incomplete population sampling, or by violations of the neutral assumptions of the method, such as adaptive introgression, as was argued for gene flow back into the African ancestral range of *D. melanogaster* (Svedberg et al. 2021).

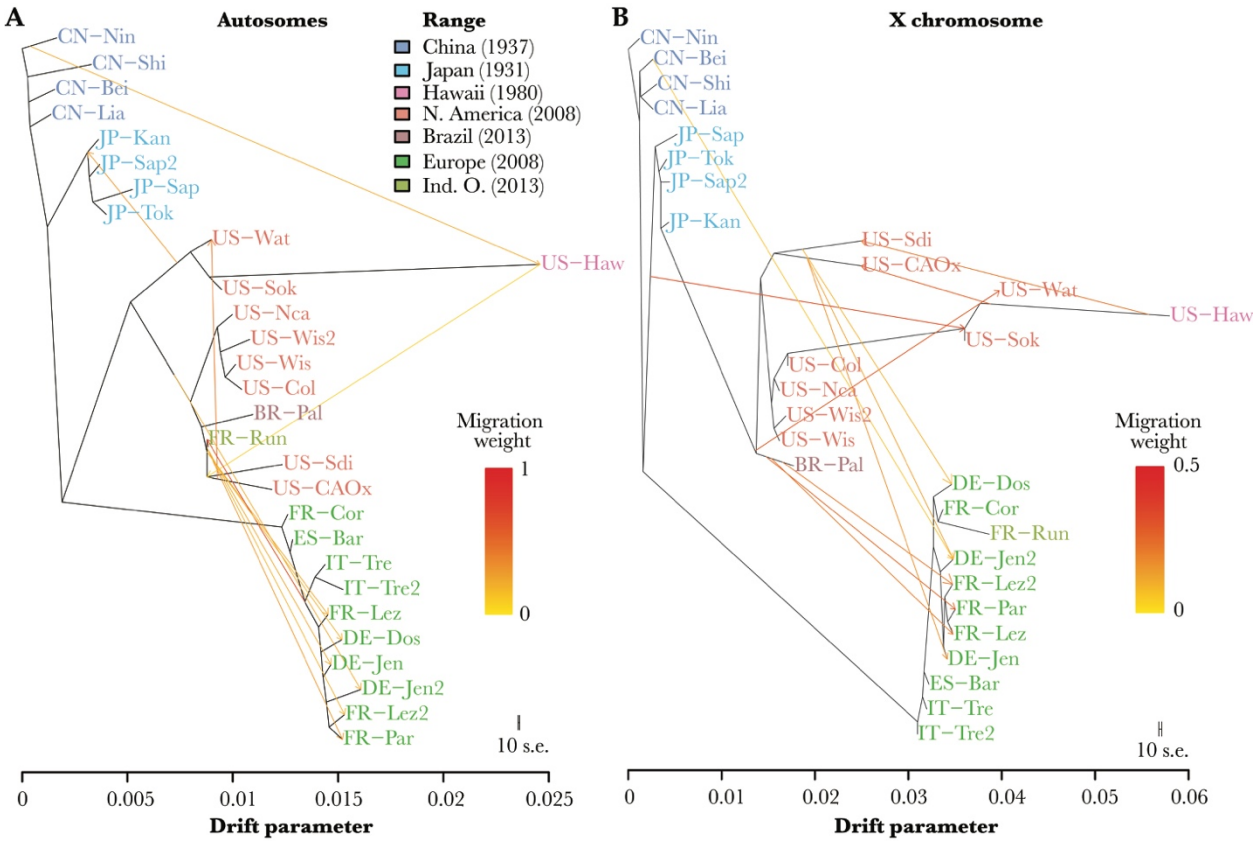
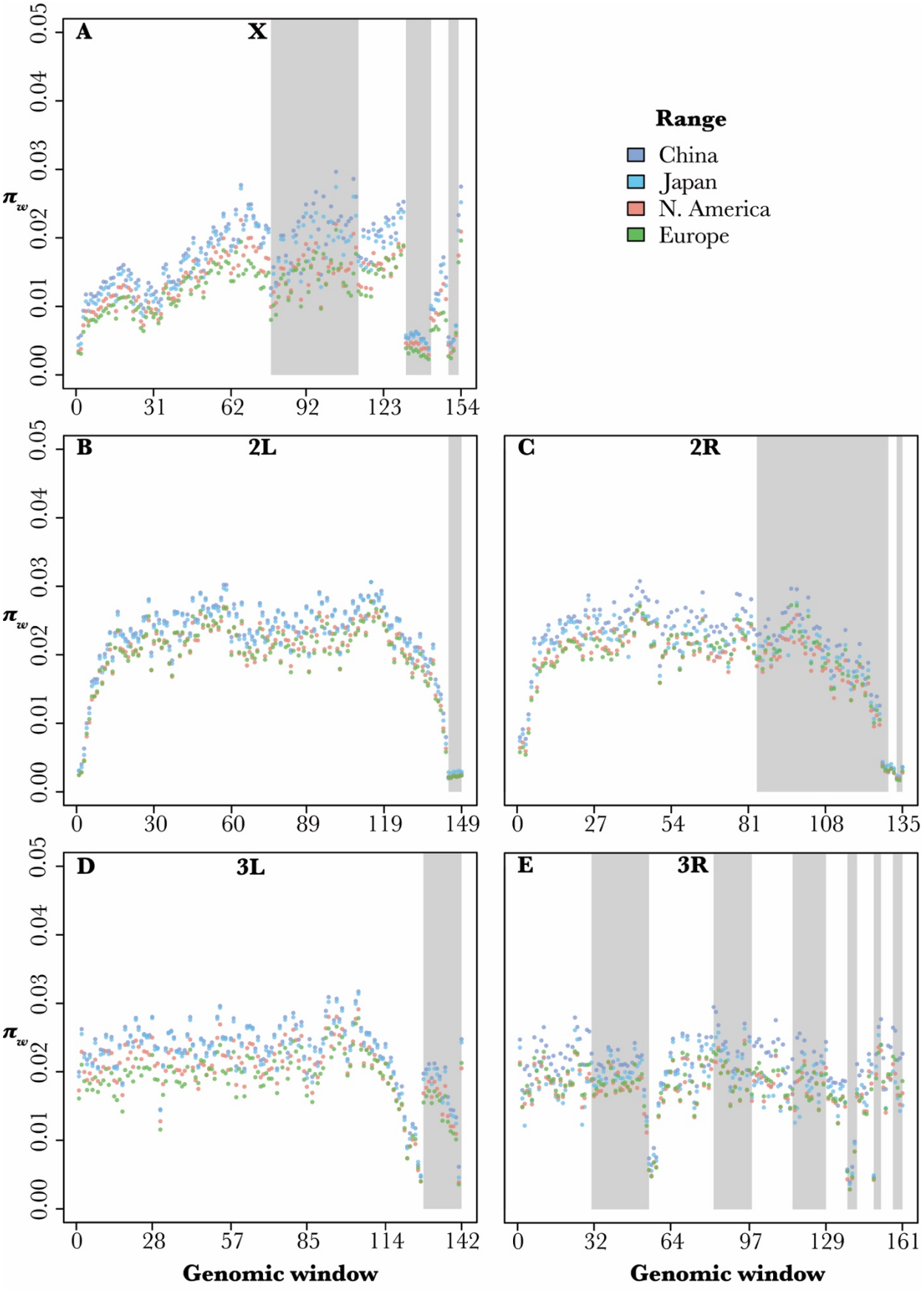


Figure 2. Population topology and migration events of *D. suzukii* suggested by TreeMix. Maximum likelihood graphs were built by allele counts at (A) autosomal and (B) X-chromosomal SNPs, allowing 11 migration events. Horizontal branch lengths are proportional to the amount of genetic drift that has occurred on the branch. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix, reflecting the precision of branch lengths estimates. Population names are colored by their geographic region. The year of the first recorded occurrence in each geographic range is given in brackets in the color legend. Migration arrows are colored according to their weight, which approximates admixture proportions. The residual fit from these graphs is presented in Figure S7.



241

Figure 3. Chromosomal distribution of genetic polymorphism in *D. suzukii* informs the ordering and orientations of contigs, as well as levels of centromeric and telomeric repression. Window nucleotide diversity (π_w) values are displayed across (A) the X chromosome and major autosomal arms (B) 2L, (C) 2R, (D) 3L, (E) 3R. Chromosome 4 is not shown as it only contains 12 windows. Each window is a continuous genomic region that includes 125,000 analyzed sites. Each dot represents the average π_w across populations within their geographic range as colored. Only populations from major continental ranges are shown for chromosomal patterns to be clear. Within each chromosome arm, contigs are ordered by length. Separate contigs are indicated by grey or white shading.

Polymorphism, Divergence, and the Genomic Locations of *D. suzukii* Contigs

Genome-wide patterns of polymorphism and divergence are useful to reveal various aspects of the evolutionary history of a species. The *D. suzukii* reference assembly (Paris et al. 2020) is somewhat fragmented, with two or more contigs assigned to each major chromosome arm, but with unknown orientation with respect to each other. We note that in other examined *D. melanogaster* group species, crossing-over rate and window nucleotide diversity (π_w) are reduced in centromere- and telomere-proximal regions (*e.g.*, True et al. 1996). When we examined the chromosomal distribution of π_w (Figure 3), we noticed that certain arrangements of contigs would result in the expected patterns of reduced diversity at the ends of each arm, and relatively smooth shifts in the diversity of large windows. For instance, the second longest contig in chromosome arm 3L is likely to the left of the longest contig in a reverse orientation, potentially with the smaller high diversity contig in between, (Figure 3D), although either of the low diversity contig end could be centromeric or telomeric. Similarly in chromosome arm 3R, although the assembly was more fragmented and the centromere- or telomere-proximal regions consist of multiple contigs, the second and third longest contigs appeared to contain the transitions between the high diversity mid-arm region and the centromeric/telomeric zones (Figure 3E). Therefore,

genome-wide polymorphisms could serve as useful information to aid the ordering and orienting of contig-level genome assemblies (Nakabayashi and Morishita 2020), although a combination of both systematic neighbor-matching approaches and manual correction would be needed, and integration with other forms of evidence (*e.g.*, comparative genomic or experimental data) would be preferable.

Genomic polymorphism levels could also be indirectly influenced by centromeric and telomeric repression of crossing over, leading to a stronger influence of natural selection reducing linked variation (Maynard Smith & Haigh, 1974; Charlesworth et al. 1993). From the distribution of window-based nucleotide diversity (π_w) across chromosome arms, we noticed that centromere- and telomere-proximal regions of reduced diversity showed similarly narrow declines in π_w (Figure 3). These patterns are more similar to those in *D. simulans* than to *D. melanogaster* – which has broader centromeric regions of low crossing over (Langley et al. 2012). These results suggest a relatively weaker suppression of crossing over in the centromere-proximal regions in *D. suzukii*. It also appears that regions of low nucleotide diversity (which probably coincide with regions of low recombination) cover a relatively small fraction of the genome, which should limit the potentially biasing effects of natural selection on demographic inferences and facilitate the localization of selective sweeps (Schrider et al. 2016; Lange and Pool 2018).

We next leveraged our large pooled sequencing data set to improve inferences about which contigs map to the X chromosome. Out of a total of 546 contigs, 313 were previously assigned to autosomes and X chromosome through either direct mapping or comparing a female-to-male read depth ratio (Paris et al. 2020). We added to these annotations by implementing an approach based on correlations in sequencing depth of coverage across population samples that included varying numbers of females and males (see Materials and

289 Methods). Based on this analysis, we assigned 170 contigs as autosomal or X-linked. Our
290 classification of previously-assigned contigs was 96% consistent with past inferences, but we
291 corrected four previous assignments that were based on female-to-male read depth ratio, whereas
292 our method did not assign eight previously-assigned contigs.

293 To reveal levels of selective constraint, we estimated the divergence between *D. suzukii*
294 and its close relative *D. biarmipes* (Ometto et al. 2013; Suvorov et al. 2022; Figure S8). Compared
295 to the divergence of *D. melanogaster* from *D. simulans*, *D. suzukii* shows a higher genome-wide
296 divergence from *D. biarmipes*. The intronic and intergenic regions also have a higher divergence
297 relative to that of nonsynonymous sites, suggesting that a lower proportion of these non-coding
298 regions are constrained in *D. suzukii* and *D. biarmipes* than in *D. melanogaster* and *D. simulans* (Begun
299 et al. 2007; Lange and Pool 2018). Non-coding divergence is potentially inflated by repetitive
300 elements (i.e., repeatome), which represent about half of the genome expansion of *D. suzukii*
301 compared to the *D. melanogaster* genome (Paris et al. 2020). This expansion of the repeatome
302 could reflect a lower long-term effective population size (N_e) in the *D. suzukii* lineage. Given that
303 the repeatome is predominantly contributed by transposable elements (TEs), that hypothesis
304 aligns with the negative correlation between TE content and N_e found among 22 worldwide *D.*
305 *suzukii* populations also sampled in this study (Mérel et al. 2021). However, the greater π of *D.*
306 *suzukii* than *D. melanogaster* (Figure S1; Lack et al. 2016a; Lewald et al. 2021) could instead
307 indicate a greater N_e for *D. suzukii* within the past $4N_e$ generations.

308

309 **Environmental Adaptation Reveals the Genetic and Functional Basis of Invasion**
310 **Success**

311 The worldwide expansion of *D. suzukii* has exposed this species to selection pressures from
312 varying local environmental conditions (Olazcuaga et al. 2020; Mérel et al. 2021). To identify
313 environmental factors that have contributed to adaptive genetic differentiation at various levels
314 and loci under positive selection, we performed a whole-genome scan using GEA analysis
315 between environmental and genetic differentiation (see Materials and Methods). Our selection of
316 environmental variables for GEA started with a preliminary set of 26 candidate variables that are
317 potentially relevant in the adaptation process of *D. suzukii* (Figure 4A), based on data availability
318 and prior knowledge of *Drosophila* ecology (Kellermann et al. 2012; Hamby et al. 2016; Bogaerts-
319 Márquez et al. 2021). Among common strategies of performing GEA with a large set of relevant
320 environmental variables, univariate association with all environmental variables could increase
321 the number of statistical tests, thus increasing the difficulty of controlling rates of false discovery.
322 On the other hand, including multiple variables might cause a multicollinearity issue (Rellstab et
323 al. 2015). We opted to retain nine of the least correlated environmental variables for univariate
324 tests, representing altitude, wind speed, as well as multiple aspects of temperature, precipitation,
325 and human land usage for GEA analysis (Figure 4B). Although the temperature of the coldest
326 quarter had a significant negative correlation with wind speed, we kept both variables for GEA,
327 as cold stress and wind-related factors are known to be potential drivers of local adaptation in
328 *Drosophila* (Bogaerts-Márquez et al. 2021). As indicated by their coefficients of variation (*CVs*), the
329 selected environmental variables had moderate ($15\% \leq CV < 30\%$) to high ($CV \geq 30\%$)
330 variability across our sampling locations (Table S3).

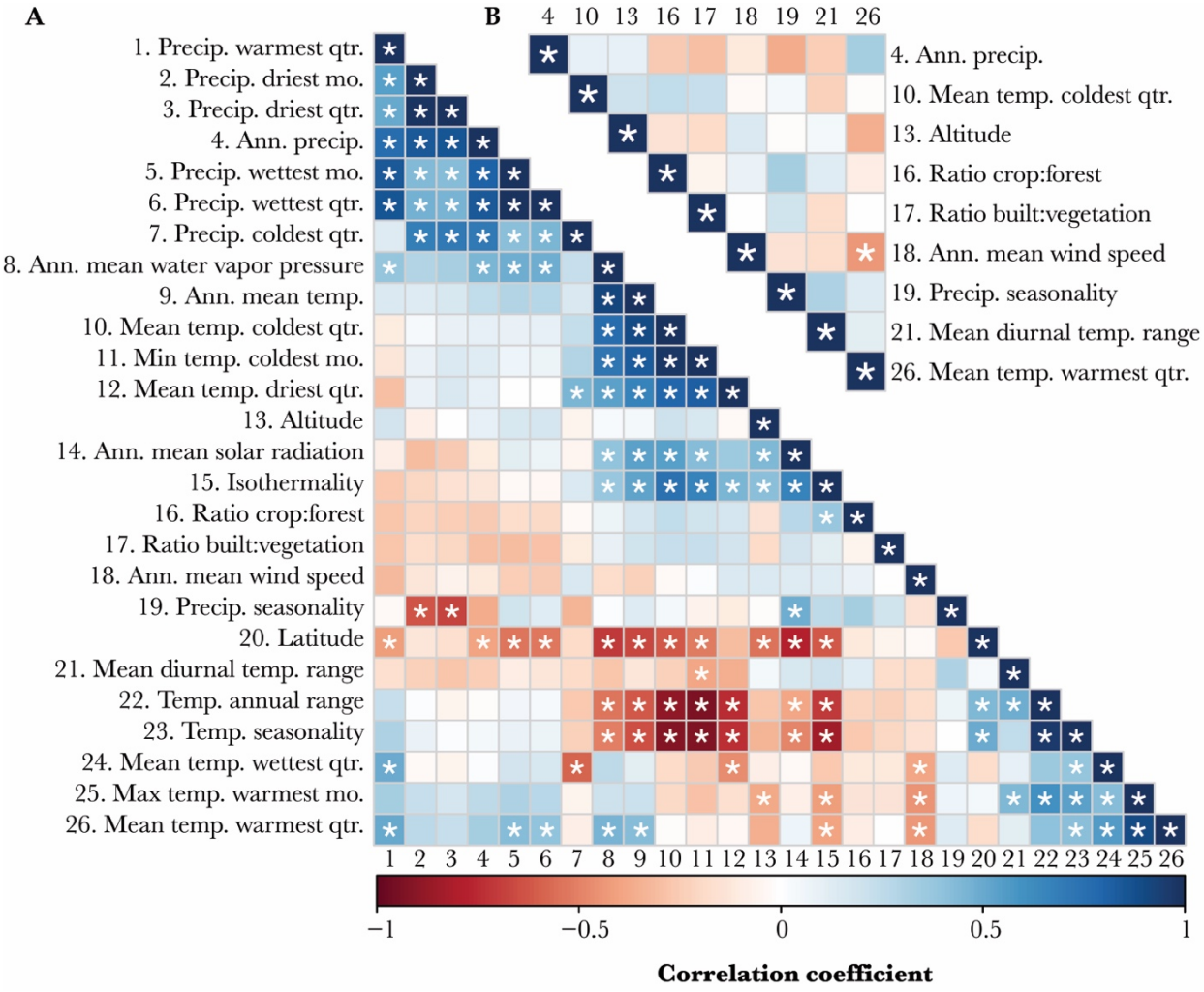


Figure 4. Identification of least-correlated environmental variables for genotype-environment association (GEA) analysis in *D. suzukii*. (A) Pairwise correlations among a preliminary set of 26 environmental variables that are potentially impactful on *D. suzukii*. (B) A final set of nine of the most relevant and least correlated environmental variables that were chosen for GEA analysis. The Pearson correlation coefficients are colored from -1 (perfect negative correlation) to 1 (perfect positive correlation). Significance correlations ($p < 0.05$) are indicated by asterisks. See Table S3 for environmental values used to calculate correlation coefficients.

From 5,752,156 genome-wide SNPs with a minor allele frequency (MAF) higher than 5%, we identified an average of 3,033 (SD = 823.4) unique candidate variants that were

significantly (genome-wide $q < 0.05$) associated with each of the nine candidate environmental variables (Table S4). These variants corresponded to an average of 3,345 overlapping or neighboring genes (Table S4), suggesting that selection pressures from the tested environmental variables (or correlated factors) have meaningfully contributed to adaptive genetic variation in *D. suzukii*, even on the brief time scale of its worldwide expansion. Among all tested environmental factors, mean temperature of the coldest quarter was associated with the greatest number of putatively adaptive variants (4,250 SNPs). Two precipitation-related variables, annual precipitation (4,141 SNPs) and precipitation seasonality (3,389 SNPs), have the next largest loci count. The ratio of built area to vegetation (i.e., crops and forests) and the ratio of crops to forests were associated with the fewest genetic variants (2,369 and 1,608 SNPs, respectively).

To reveal the potential genetic and functional basis of invasion success under multiple environmental challenges throughout the species' range, we first examined the functions of genes linked to the top 10 environment-associated loci for each variable as ranked by association q-value, and then by the g parameter estimating the sensitivity to environmental differentiation as a tie-breaker. We found many of these genes have known functions that could facilitate adaptation to the associated environmental factor (Table S5).

Among genes linked to altitude-associated loci, the second-ranked candidate *ab* is known to control wing size in *Drosophila* (Simoes da Silva et al. 2019). Interestingly, wing size was found to have increased in a highland Ethiopia *D. melanogaster* population, potentially assisting flight in thin, cool air (Lack et al. 2016b). Ranked next to *ab* is *Gbs-70E*, which plays roles in glycogen metabolism and the development of eggs inside the maternal ovary (Kerekes et al. 2014). Another top gene, the lysine demethylase *Kdm2*, is upregulated in response to hypoxia (Batie et al. 2017).

With wind speed, the top first candidate *Ttc30* is an essential gene in the biogenesis of sensory cilia, which are key to both chemosensory and mechanosensory functions in *Drosophila* (Avidor-Reiss et al. 2004; Avidor-Reiss and Leroux 2015). Another top candidate, *Arr2*, is involved in olfaction, hearing, and vision (Alloway and Dolph 1999; Elaine Merrill et al. 2005; Senthilan et al. 2012). In light of the relevance of wind for insect flight, we also noted that a third top candidate, *vn*, is a developmental gene named for its wing phenotype (Wang et al. 2000).

There is also some evidence for precipitation-related local adaptation. The top gene *mmy* associated with precipitation seasonality (i.e., annual range of precipitation) was shown to regulate chitin synthesis and cuticle production. Since precipitation is correlated with desiccation resistance across the *Drosophila* phylogeny (Kellermann et al. 2012), *D. suzukii* may have developed adaptive strategies of modifying chitin biosynthesis under conditions of desiccation (Rezende et al. 2008; Clark et al. 2009), which was also implied in seasonal plasticity of natural *Drosophila* populations (Shearer et al. 2016; Horváth et al. 2023). In addition, the gene *osy* (*CG33970*) contributes to the formation of the outer cuticle layer and is expressed more highly in *D. suzukii* than in *D. melanogaster* (Wang et al. 2020). Furthermore, two of the top genes associated with annual precipitation (*Abd-B* and *bab1*) regulate cuticle pigmentation (Rogers et al. 2013), which may or may not correlate with desiccation tolerance in *Drosophila* species (Wang et al. 2021). We also note that although environmental fitness effects on these testes-expressed genes are not known, the same SNP near *CG17944* and *nxf4* was among the highest-scoring variants for both annual precipitation and precipitation seasonality (variables that have a non-significantly negative correlation between them; Figure 4).

Another important environmental barrier to invasion success is temperature. For the mean temperature of the coldest quarter, a top gene was *Ac78C*, which has roles in circadian

regulation and taste (Ueno and Kidokoro 2008; Duvall and Taghert 2013). With the mean temperature of the warmest quarter, the top genes *crp*, *Mrtf*, and *Ubx* help control the development of trachea (Han et al. 2004; Guha and Kornberg 2005; Wong et al. 2015), which may be important in limiting water loss in hot environments (Gibbs et al. 2003).

For the ratio of built to vegetated area, a different variant near the cuticle-related gene *osy* (which was also indicated above for precipitation seasonality) was detected. Another top outlier was the nervous system gene *trv*, which is involved in thermosensitivity (Honjo et al. 2016). For the relative levels of crop and forest cover, the first-ranked variant was near *Mtk*, which encodes an antifungal and antibacterial peptide (Levashina et al. 1995), and we note that mushrooms (which are more available in forest) have been proposed as overwintering food sources for *D. suzukii* (Wallingford et al. 2018), and the evolution of immune genes has been found to differ strongly between mushroom-feeding and human commensal *Drosophila* species (Hill et al. 2019). With regard to the differential light environments entailed by forest versus farm habitats, we note that the next highest gene, *CadN2*, helps connect photoreceptor neurons to their targets (Prakash et al. 2005).

Beyond the top candidate genes that have related functions to specific types of environmental changes, we also found a wide range of nervous system genes associated with multiple environmental factors. For instance, among the top five altitude-associated loci, three have known functions in the nervous system of *Drosophila*, including the first-ranked gene *Cmpy*, which enables proper growth control at neuromuscular junctions (James and Broihier 2011), *ab*, which regulates dendritic complexity (Li et al. 2004; Sugimura et al. 2004), and *not*, which is essential for stabilizing synaptic homeostasis within glia (Wang et al. 2020). Such genes were also linked to top-10 loci associated with wind speed (*dpr6*), precipitation (*Msp300*, *Tusp*, *5-HT2A*),

410 temperature (*ATP6AP2*, *CG13579*, *D*), and land use variables (*Bsg*, *Mp*, *velo*). Different variants
411 associated with *scrt*, a regulator of neuronal cell fate, were among the top results for both mean
412 diurnal range and the ratio of crop to forest cover.

413 Next, we examined environment-specific adaptation on a more comprehensive basis
414 through a gene ontology (GO) enrichment analysis of the top 500 genes associated with each
415 environmental variable (Figure 5). As correlates of temperature in the coldest quarter, cAMP
416 metabolic process was the top enriched category, followed by two other related purine
417 metabolism groupings. We note that cAMP is important in circadian regulation (*e.g.*, Palacios-
418 Muñoz & Ewer, 2018), which is known to play an important role in *Drosophila* environmental
419 adaptation (*e.g.,* Helfrich-Förster et al. 2020), as also implicated by the presence of "entrainment
420 of circadian clock" on our top GO category list for altitude. More broadly, purine metabolism
421 was inferred as a strategy of cold acclimation in *D. suzukii* (Enriquez and Colinet 2019). For
422 diurnal temperature range, the top category was "regulation of growth", and we note that some
423 drosophilids have evolved to have larger body sizes in more challenging thermal environments
424 (Gilchrist and Partridge 1999; Calboli et al. 2003; Lack, Yassin, et al. 2016).

425 With precipitation, we identified "chitin metabolic process" as the top GO term
426 associated with annual precipitation, as well as "chitin-binding" with precipitation seasonality.
427 Together with the chitin synthesis genes we described above for precipitation, adaptation to the
428 overall intensity and seasonal variation of precipitation by modifying cuticular chitin may be
429 implied. For crop to forest ratio, the category "antimicrobial humoral response" included the top
430 gene *Mtk* listed above.

431

GO description | p-value | # genes

Figure 5. GO enrichment analysis of candidate genes from the gene-environment association analysis of *D. suzukii*. The top 10 GO categories enriched by the top 500 genes associated with each environmental variable are shown in each panel (labelled on the left), with permutation p-values and the number of associated genes in each GO

435 category. Descriptions of GO categories are colored by their GO class (see legend at top right). Only GO
436 categories including more than five associated genes are listed here. For a full list of enriched GO categories, see
437 Table S6.

438

439 As broader evidence for a shared (or biologically similar) underlying genetic basis of
440 adaptation to multiple environmental factors, we examined the overlap of the most significant
441 genes and most enriched GO categories between different environmental variables. We found
442 the top candidate genes to be mostly associated with both temperature and precipitation. The
443 gene sets showed relatively greater overlap among climatic variables (including altitude), whereas
444 the two land usage variables had less overlap with climatic variables or with each other (Figure
445 6A). Since the patterns of shared genes cannot be fully explained by correlations between
446 environmental values (Figure 4B), at least some of the genes may have been responding to
447 multiple selective pressures. The overall proportions of shared GO categories were lower than
448 those of shared genes, indicating that the shared genes do not necessarily lead shared functional
449 categories between environmental variables. Relatively higher GO term sharing was observed
450 between altitude and either wind speed or precipitation, and between diurnal temperature range
451 and temperature of the warmest quarter (Figure 6B). Based on the shared genes and GO terms
452 observed, it is possible that during the rapid range expansion of *D. suzukii*, pleiotropy may have
453 facilitated local adaptation to multiple selective pressures (Hämälä et al. 2020; Kinsler et al.
454 2020).

455 Consistent with our gene-based analyses of universal adaptive function, we found three of
456 the top shared biological processes clearly related to nervous system functions, including the
457 topmost "synaptic transmission, glutamatergic" (shared by altitude, wind speed, diurnal

458 temperature range, and temperature of the warmest quarter), "regulation of neurogenesis"
459 (altitude, diurnal temperature range, temperature of the warmest quarter, and ratio of built area
460 to vegetation), and "central nervous system development" (precipitation seasonality, diurnal
461 temperature range, and temperature of the warmest quarter) (Figure 6C). Further, as mentioned
462 above, "cAMP metabolic process" (shared by temperature of the coldest quarter, precipitation
463 seasonality and ratio of built area to vegetation) could entail neurologically modulated changes in
464 circadian behavior. Each of these traits was associated with at least three environmental
465 variables, which suggests a multifaceted role for nervous system evolution in facilitating the
466 invasion success of *D. suzukii* under multiple environmental challenges.

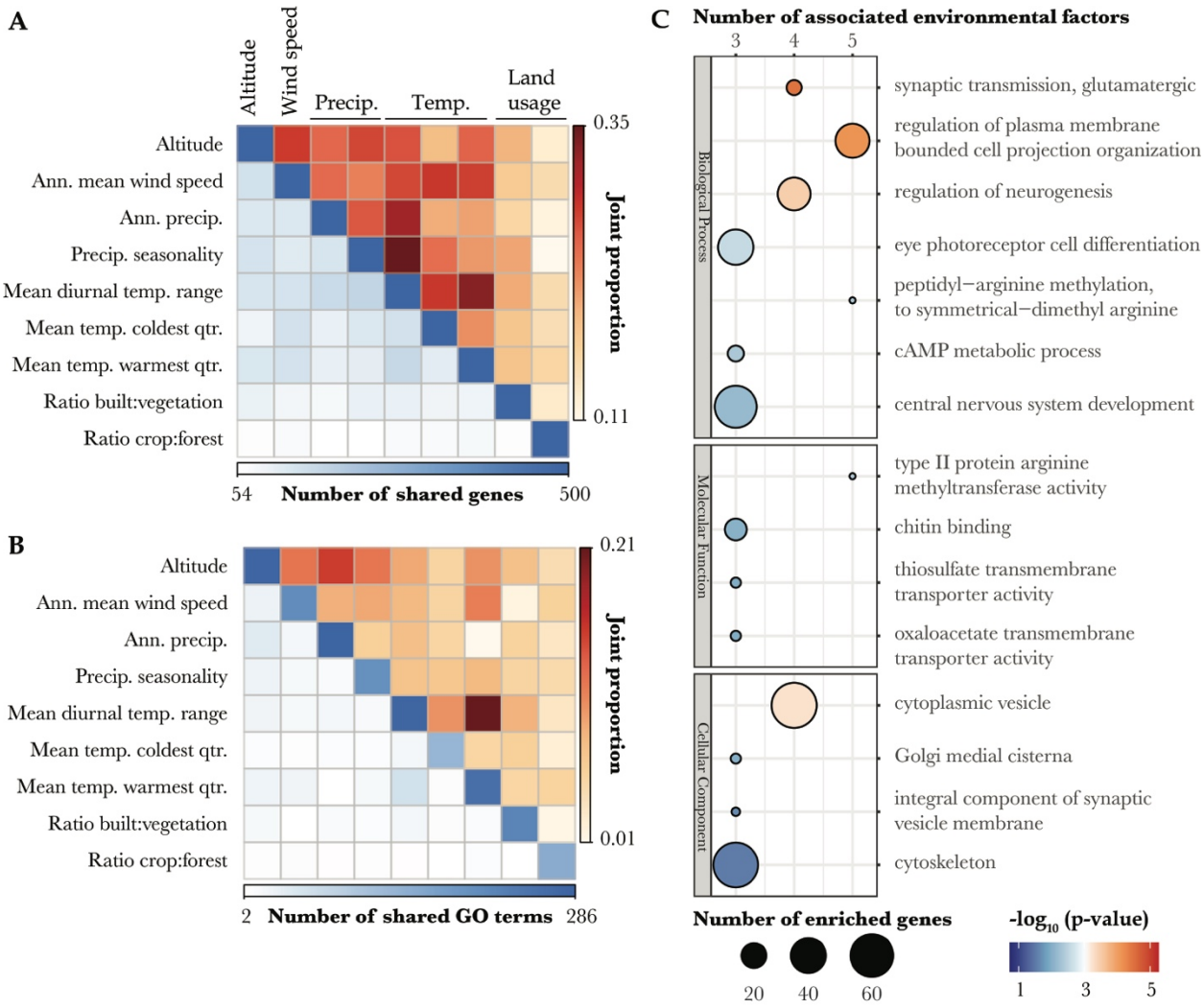


Figure 6. Overlapping genes and GO categories among environmental factors reveal the shared genetic and functional basis of environmental adaptation in *D. suzukii*. The numbers and proportions of shared (A) environment-associated genes and (B) enriched GO categories among environmental factors are shown in heatmaps. Here, joint proportion represents the fraction of the genes or GO terms associated with either of two environmental variables that are associated with both variables. (C) Top GO categories of each type are depicted as bubbles. Bubbles are colored by the negative logarithm of the combined p-value of enrichment across all environmental variables, and are scaled by the number of enriched genes. The number of environmental variables that enrich a given GO category is indicated by the top horizontal axis.

Discussion

We performed population genomic analyses of 29 population samples of *D. suzukii* to investigate the genomic diversity, invasion history, and environmental adaptation of this highly invasive species across its worldwide distribution. Our data shows a genetic grouping of these populations into four primary geographic regions: eastern Asia (containing the native range), Hawaii, the Americas, and Europe. We estimated the highest genetic diversity in eastern China (Ningbo), suggesting this region could be within or near the ancestral range of *D. suzukii*. We confirmed moderate founder event bottlenecks in introduced populations, and our analysis suggested that admixture during the expansion of *D. suzukii* populations might be more complex than previously estimated, but that further sampling in Asia is needed to address the precise origins of introduced populations. We then identified the environmental drivers for local adaptation of *D. suzukii*, including altitude, wind speed, precipitation, temperature, and human land usage, where temperature of the coldest quarter shows the clearest evidence of environmental adaptation. Our results suggested extensive local adaptation in response to specific environmental challenges. We also revealed appreciable sharing of genes and functional pathways underlying invasion success across multiple environmental pressures, which were most obvious with nervous system genes.

Insights and mysteries regarding the history of *D. suzukii*

Surprisingly, the relative levels of diversity of *D. suzukii* populations have been the subject of conflicting findings. Here, we detected the highest overall genetic diversity for *D. suzukii* in eastern Asia, especially in the eastern China Ningbo population (Figure 1B; Figure S1). Our results contrast with some prior findings, such as higher diversity in several North American populations relative to a Japanese population (Adrion et al. 2014), the species' highest diversity

being in north-central China (Fraimout et al. 2017), or alternatively in Japan (Gautier et al. 2022). These conflicts may reflect challenges in accurate estimation of genetic diversity, especially when the number of investigated loci is limited (Adrion et al. 2014; Fraimout et al. 2017), or when rare alleles are filtered by a relatively strict allele count threshold (Gautier et al. 2022), given that rare variants are especially likely to be lost during bottlenecks. In this study, we accounted for such biases by applying stringent filtering by sequencing read base quality and mapping quality, by including putative rare alleles in genome-wide analyses, and by focusing on synonymous diversity which has a relatively higher ratio of signal to noise (more true variation relative to sequencing errors).

Beyond the eastern China population's maximal diversity, we found that no pairwise D_{XY} values meaningfully exceeded Ningbo π, implying that two Ningbo *D. suzukii* genomes are as different from each other as any pair of genomes worldwide (paralleling a result for a Zambia population of the human commensal *D. melanogaster*; Pool et al. 2012). Those findings are consistent with all sampled *D. suzukii* populations originating from a Ningbo-like ancestor in recent population genetic time. This inference could conceivably imply that *D. suzukii* only expanded across portions of its current eastern Asian range due to its relationship with human agriculture. Alternatively, in light of the apparently large effective population size of *D. suzukii*, the plausible time frame of divergence among Asian populations could extend back toward the last glacial maximum (LGM). We note that Ningbo is near the historical northern limit of temperate woodlands during the LGM (Ray and Adams 2001). If this region is indeed part of a limited overlap between the continental habitats of this temperate species during glacial and interglacial periods, then this stability of occupation could help explain the Ningbo population's

elevated genetic diversity, whereas other eastern Asian populations' somewhat lesser diversity might stem from mild founder event bottlenecks during post-glacial range expansions.

Further sampling and analysis of Asian *D. suzukii* populations is urgently needed to advance our understanding of the history of this species. Even the current distribution of *D. suzukii* in eastern Asia is not entirely clear. According to TaxoDros (https://www.taxodros.uzh.ch/), *D. suzukii* has been sampled from China, Japan, North and South Korea, Burma, Cambodia, Thailand, and Taiwan. It is unclear whether the lack of records from neighboring countries such as Laos and Vietnam, where the species is predicted to thrive (Santos et al. 2017), reflects a genuine absence or insufficient sampling effort. Understanding the history of *D. suzukii* in Asia will require the analysis of genomic diversity from dozens of population samples from across its Asian range, including from countries beyond China, Japan, and Korea, where all population genetic study to date has focused.

Geographically denser population genomic data from Asian *D. suzukii* would also facilitate clearer conclusions regarding the origins of lineages that invaded regions such as Hawaii, the mainland US, and Europe. Above, we noted some contrasts between our TreeMix results, which placed mainland US and Europe populations closer to sampled Japan populations than to China, and the model of Fraimout et al. (2017), which primarily drew upon China source populations for these invasions. However, our autosomal TreeMix results suggested that neither of these invasive lineages falls within the clades of sampled populations from either China or Japan (and our X chromosome analysis echoes this finding for Europe). Our results are compatible with the possible origins of continental *D. suzukii* invasions from one or more Asian populations that are somewhat genetically differentiated from any of those analyzed here. At

545 present, we suggest that the geographic origin of these economically significant introductions
546 remains an open question.

547 Our TreeMix analysis also suggested that worldwide patterns of admixture may be more
548 complex than suggested by previously inferred scenarios (Adrion et al. 2014; Fraimout et al.
549 2017; Gautier et al. 2022). However, caution is required to interpret admixture events that were
550 inferred solely from this analysis, and there may be several reasons why some inferences differ
551 between ours and prior analyses. First, as indicated above, we are still working with a very
552 incomplete panel of populations, especially from Asia, and methods may differ in terms of how
553 they respond to the absence of historically important populations such as the sources of the
554 invasive lineages. Second, whereas the study of Fraimout et al. (2017) restricted its consideration
555 to founder events compatible with the observed chronology of species expansion, our TreeMix
556 results represent an investigation of population history through genomic data alone, without
557 assuming that each worldwide invasion was detected rapidly upon its occurrence. In species with
558 complex population structure, multiple histories could lead to identical covariance matrices, thus
559 allowing for several different invasion histories compatible with the data (Pickrell and Pritchard
560 2012). Finally, as mentioned by Adrion et al. (2014) and Fraimout et al. (2017), the additional
561 migration events predicted in our analysis could indicate that genome-wide data is more sensitive
562 for discriminating among complex invasion scenarios and population structures than a limited
563 number of loci.

564 As indicated above, our TreeMix results also differed somewhat between autosomal and
565 X-linked loci. To some extent, demographic history is expected to differ between these loci,
566 including due to different levels of bottleneck severity at distinct effective population sizes, and
567 the unbalanced sex ratio of this species, especially when the multiple mating is present or sex

ratios depart from null expectations (Pool and Nielsen 2009; Tait et al. 2018; Weißinger et al. 2019; Manko et al. 2021). However, it is not clear that tree topology or gene flow presence should differ substantially on such grounds. In addition to concerns listed above such as model identifiability and missing populations, the consistency of autosomal versus X-linked inferences could also be influenced by natural selection, particularly if it influences a longer stretch of the genome with reduced recombination, such as a centromeric or telomeric region, or a polymorphic inversion.

A key limitation of our pool-seq data is that it cannot provide haplotype information. In light of the challenges in performing demographic inference from pooled sequencing data, there would be considerable benefit in high quality individual sequencing of multiple wild-caught flies from each of a broad array of populations (particularly in Asia), in order to facilitate quantitative inferences regarding population history. Such data, especially if combined with efficient model selection methods (*e.g.*, ABC-RF, (Pudlo et al. 2016; Fraimout et al. 2017), could provide a more confident and precise demographic model which could both inform species history and serve as null hypothesis for the detection of natural selection.

Environmental Drivers of Adaptation in *D. suzukii*

Here, we presented a GEA analysis that investigated the most geographically and genetically diverse set of *D. suzukii* populations and the most comprehensive set of environmental factors to date, which enabled unprecedented power to capture even minor adaptive genetic differentiation in response to distinct environmental challenges during the species' rapid invasions. In addition to our identification of climatic factors including temperature, precipitation-related variables and wind speed as the most frequently correlated with putatively

591 local adaptative variants (consistent with previous GEA analysis in *D. melanogaster*, *e.g.,* Bogaerts-
592 Márquez et al. 2021), we also for the first time identified large numbers of genome-wide variants
593 associated with altitude and human land usage-related variables (Table S4), which were not
594 included in most GEA studies despite their potential significance to local adaptation. In
595 particular, the detected associations with ratios of developed land to vegetation and of cropland
596 to forests highlights the ecological impacts of urbanization and agriculture on natural populations
597 of insects. Since the prior selection of environmental variables is critical for successful GEA
598 analyses, we also provided an instructive example for correlation-based selection to identify the
599 most relevant and least redundant environmental factors (Rellstab et al. 2015).

600

601 **Nervous System Evolution is Ubiquitous in Environmental Adaptation of**
602 ***Drosophila***

603 In *D. suzukii*, we found nervous system and related sensory and behavior annotations
604 associated with top genes for all nine environmental variables studied. Concordantly, we found
605 that GO categories related to the nervous system were among the most shared across
606 environmental variables (Figure 6C). In *D. melanogaster*, related GO categories like 'neuron
607 development', 'nervous system development' and 'eye development' were also enriched among
608 genes associated with environmental variation among natural populations within North America
609 or Europe, and across seasons within Europe (Bogaerts-Márquez et al. 2021). GO categories
610 associated with the nervous system have also shown evidence of positive selection in various
611 genome scans of *D. melanogaster* (Langley et al. 2012; Pool et al. 2012; Pool 2015), including a
612 study of parallel evolution in cold-adapted populations (Pool et al. 2017). Given the
613 morphological evidence of neuron-muscular junction evolution across the entire *Drosophila*

phylogeny (Campbell and Ganetzky 2012), we therefore propose a broad adaptive importance of the nervous system in *Drosophila* species and potentially other insects. Such evolutionary processes may have either maintained ancestral neural functions in novel challenging environments, or created novel phenotypes that better fit the new optima arising from complex combinations of environmental factors.

Cautions with Interpretations of Association Results and Future Directions

While we have generated intriguing hypotheses about gene functions that may underlie the environmental adaptation of *D. suzukii*, it is difficult to distinguish between correlated environmental selective pressures that may have driven the detected associations, including not only the 17 environmental factors that were excluded in the process of variable reduction, but also correlated biotic or abiotic factors not represented in global databases. As an intrinsic limitation of GEA analysis that cannot be accounted for by applying stricter thresholds, associations observed with a particular environmental factor might stem from adaptation to other covarying factors (Rellstab et al. 2015). For example, the two tracheal branching genes *crp* and *Mrtf* (Han et al. 2004; Wong et al. 2015) associated with mean temperature of the warmest quarter could represent adaptations to reduce water loss under conditions of elevated water vapor pressure (Telonis-Scott et al. 2012), which is closely related to humidity and has a significant positive correlation with mean temperature of the warmest quarter (Figure 4B).

Therefore, expanded characterization of the relationships between genotype, phenotype, and fitness in this species is needed to further clarify the functional and phenotypic interpretations associated with certain environmental factors and genes. Experimental validations that leverage RNA interference (Boutros and Ahringer 2008) and/or transgenic

637 overexpression (Prelich 2012) to modify the expression of associated genes, and/or genome
638 editing techniques (Stern 2014; Turner 2014; Shalem et al. 2015) to target putatively adaptive
639 variants would also bring a more solid understanding about the invasive biology of this species in
640 distinct environments. Such functional studies could be complemented by population
641 experiments under controlled laboratory environments or field conditions (*e.g.*, Behrman et al.
642 2015; Rudman et al. 2022), in order to more clearly demonstrate the connections between
643 specific selective pressures and alleles or traits of interest.

644

645 **Broader Impacts and Significance**

646 Our work integrates genetic and environmental data to improve the reconstruction of the
647 invasion genomics of a crop pest carrying significant economic costs (Knapp et al. 2021), which
648 will hopefully inspire future studies on developing diverse pest control methods given the
649 adaptive and neutral genetic differentiation among *D. suzukii* populations. Understanding the
650 extent of local adaptation and its potential environmental drivers will also help predict the spread
651 and future distributions of invasive species (Colautti and Lau 2016). More broadly, the enhanced
652 understanding of how organisms may adapt to geographical, climatic and artificial selective
653 pressures from this study is also of value in assessing the susceptibility of natural populations to
654 climate change (Kellermann et al. 2012) and human activities (Barange et al. 2010).

655 Our analysis of genome-wide diversity has uncovered novel patterns (such as the maximal
656 diversity of an eastern China population), while reinforcing other prior findings and adding new
657 perspectives on the invasive history of *D. suzukii*. As suggested above, our findings indicate a
658 strong need for the sequencing and analysis of individual genomes from a larger number of
659 worldwide populations, especially from additional regions of eastern and southern Asia. Such

efforts could lead to a vastly improved understanding of both the evolution of *D. suzukii* within its native continent and the geographic origins of worldwide invasions.

Materials and Methods

Fly Collection, DNA Preparation, and Pooled Sequencing

Fly samples from 29 populations were used, seven of which were sequenced for the present study. The fly samples sequenced in this study were collected from wild *D. suzukii* populations in two states of the USA, two provinces of Japan, and three European countries (Figure 1; Table S1). Pooled whole adult flies (n = 100 ~ 183) from each population (Table S1) were used for DNA extraction as previously described (Langley et al. 2011). Library preparations were conducted at the Next Generation Sequencing Core of University of Wisconsin Madison Biotechnology Center (https://dnaseq.biotech.wisc.edu), where pair-end (PE) reads at the length of 150bp were then generated for each of seven pooled DNA samples on an Illumina NovaSeq 6000.

Pool-sequenced reads of 22 additional *D. suzukii* population samples, including from Europe, the Americas, and Asia, were obtained from public data provided by Olazcuaga et al. (2020) at EBI's SRA (Figure 1; Table S1). Taken together, we formed a comprehensive dataset of 29 populations sampled from native and invasive ranges of *D. suzukii*.

Quality Control, Alignment, and Variant Calls from Pool-seq Data

To maximize the quality of our analyzed data, we built a high-throughput assembly and quality control pipeline *poolWGS2SNP* with optimized performance, stringent filtering, compatibility with large numbers of genomic contigs, and customized functions to call high-confidence single-nucleotide variants from pool-sequenced data in *D. suzukii* (Figure S9), in part by utilizing resources from the DrosEU bioinformatics pipeline (Kapun et al. 2020).

As an initial quality control of raw PE reads, adapters were removed, and the 3' end of reads with base quality < 20 were trimmed using *fastp* (Chen et al. 2018). Further trimming was performed using a self-developed python program *filter_PE_length_mem.py* (see Data Availability), where any pair of forward and reverse reads with less than a total of 150 bases with base quality (BQ) ≥ 20, as well as any individual reads with less than 25 bases with BQ ≥ 20 were discarded.

The trimmed and qualified reads were then mapped against the recently released near-chromosome level *D. suzukii* genome assembly Dsuz-WT3_v2.0 that covers autosomes and the X chromosome (Paris et al. 2020) using *bwa mem* (Li 2013). Reads with a mapping quality below 20 were then removed using *Samtools* (Li et al. 2009). We used Picard's *SortSam* to sort BAM files, and used Picard's *MarkDuplicates* to mark PCR duplicates to avoid false variant calls (http://broadinstitute.github.io/picard). Indel identification and realignment around indels were performed using GATK's *RealignerTargetCreator* and *IndelRealigner* (Auwera and O'Connor 2020). Finally, alignments in BAM format were checked for formatting errors using Picard's *ValidateSamFile*. Summary statistics for quality checking of BAM files were generated using *bamdst* (https://github.com/shiquan/bamdst).

To call SNPs, we merged the quality-checked BAM files of all population samples into one file using Samtools *mpileup*, only retaining alignments with mapping quality no less than 20 and sites with base quality no less than 20. Variant calling was then performed on the mpileup

file using the heuristic SNP caller *PoolSNP* (Kapun et al. 2020). We used a nominally low value for the parameter miss-frac (0.001) to require for each population sample individually, that depth of coverage at a given site be 12 or greater (min-cov = 12), and that this site not be in the top 1% of sites genome-wide for depth of coverage (max-cov = 0.99; calculated separately for each population and for autosomal and X-linked contigs), in order to filter sites subject to copy number variation. In the initial data set used for analysis of genome-wide diversity, we avoided potential biases from allele frequency filters by using min-count = 1 and min-freq = 0. We termed the resulting high-quality sites as 'analyzed sites' for brevity.

Identifying Autosomal and X-linked Contigs

We chose to perform all population genomic analyses and whole-genome scans separately for SNPs from autosomes and the X chromosome for the following reasons: 1) autosomal and X-linked variants have different allelic sample sizes as samples were obtained from both male and female flies; 2) autosomes and the X chromosome could reflect different demographic histories and outcomes of natural selection, *e.g.*, the lower effective population size of the X chromosome than autosomes could lead to a higher impact of bottlenecks and selection on genomic diversity; 3) unbalanced sex ratios and male-biased dispersal could further differentiate autosomal versus X chromosome variation (Clemente et al. 2018; Olazcuaga et al. 2020).

Since the assembly of *D. suzukii* reference genome is still at the contig level, chromosomal identities of each contig are needed to perform separate analyses. However, 497 contigs that represent ~43% of the assembly length have not been unambiguously mapped onto chromosome arms of the *D. melanogaster* dm6 genome assembly. Although 264 of the unplaced contigs had been assigned to autosomes and the X chromosome based on a female-to-male read depth ratio,

233 contigs that represent ~5% of the genome remained unassigned due to the lack of statistical power (Paris et al. 2020).

Given our interest in accurately analyzing a larger proportion of the euchromatic genome, we identified ~70% of these 233 unassigned contigs as autosomal and X-linked based on the correlation between the mean read depth of each contig (among population samples) and that across unambiguously aligned autosomal or X-linked contigs. We chose Spearman's rank correlation instead of the Pearson correlation, as the distribution of depth data failed the assumption of bivariate normality. A contig that has mean depth significantly correlated with that of either known autosomal or X-linked contigs was assigned to the chromosome with a higher correlation coefficient. Our method completely confirmed all prior mapping-based assignments and had a ~96% consistency with the previous assignment based on female-to-male read depth ratios. Inconsistent assignments for four contigs were corrected according to our method (Table S2). The eight previously-assigned contigs that could not be assigned using our method, as well as other unassigned contigs using all methods (totalling ~2.7 mb), were excluded from downstream GEA analyses, because the assignment information is needed for estimating effective sample size that were used to correct allele count data as an input to GEA.

Annotating Genomic Features and Estimating Divergence

To explore genomic diversity at synonymous sites and selective constraint for other site types in *D. suzukii*, we classified the reference genome into nine exclusive categories of site degeneracy and function (Lange and Pool 2018), including: non-degenerate (i.e., nonsynonymous) sites, two-, three-, and four-fold degenerate (i.e., synonymous) sites, 3' and 5' UTRs, RNA-coding genes, introns, and intergenic regions. From input files including the

eukaryotic codon table, the published genome sequence and GFF3 annotation obtained at NCBI RefSeq, we generated a letter-coded annotation (in FASTA format) mirroring both strands of the whole genome sequence of *D. suzukii* and a coordinate-based annotation (in BED format) that combines adjacent sites of the same category into a single row. Degeneracy was determined based on the standard codon table. 5' UTRs were defined as regions between the start of the first exon and the start of the first coding sequence (CDS), while 3' UTRs were defined as regions between the end of the last exon and the end of the last CDS. In cases of overlapping genes and alternative splicing that raise annotation conflicts, we followed an annotation priority in the category order listed above.

We then estimated the divergence between *D. suzukii* and its close relative *D. biarmipes* in each of these categories (Suvorov et al. 2022). We obtained results of multiple sequence alignment between the current reference genomes of *D. suzukii* and *D. biarmipes* (Paris et al. 2020). For each site category of *D. suzukii*, the unpolarized divergence was estimated as the number of substitutions over the total number of sites within aligned blocks of reference genome sequences.

Estimating Nucleotide Diversity, F_{ST} D_{XY}

To compare genome-wide polymorphism among populations, we estimated nucleotide diversity (π) across SNPs at four-fold degenerate sites (π_S) in addition to that at all categories of sites (π_A), as π_s estimation is relatively less affected by sequencing errors than nucleotide diversity estimated from other site categories (due to a higher ratio of real variation to errors). To calculate π for each population sample, we adopted an unbiased estimator of nucleotide diversity ($\hat{\theta}_{\Pi}$) based on heterozygosity (Π), which has been optimized for pool-seq data (Ferretti et al. 2013). Numerically,

$$\pi_S = \hat{\theta}_\Pi = \frac{n_c}{n_c-1}\frac{\Pi}{L} = \frac{n_c}{n_c-1}\frac{1}{L}\sum_l \frac{2}{n_r(l)(n_r(l)-1)} m_l(n_r(l)-m_l) \quad (1)$$

Here, L represents the total number of genome-wide analyzed sites. Of a given population sample, $n_r(l)$ represents the read depth of the top two alleles at the lth site (i.e., SNP) and m_l represents the minor allele count. n_c as a normalization factor represents the haploid sample size for either autosomes or X chromosome in a pool (Table S1). Strictly speaking, n_c as a normalization factor should represent equally contributing chromosomes in a pool. Nevertheless, for our data it is sufficient to use haploid sample size for either autosomes or X chromosome to approximate n_c in the above equation, as the estimation of $\hat{\theta}_\Pi$ is not substantially affected by the precise value of n_c when the number of individuals in the pool is large. The above formula is a simplified version for SNP data, based on equation 3 in Ferretti et al. (2013).

To examine patterns of polymorphism across chromosome arms, we also estimated window nucleotide diversity (π_W) for all polymorphic sites. Each window was defined as a continuous genomic region that includes 125,000 analyzed sites (Figure 3). Since chromosomal identity was required in this analysis, we only took windows from 32 major contigs that contain at least one full-size window and were unambiguously mappable to a chromosome arm of the *D. melanogaster* dm6 genome assembly. Although such contigs only make up 57% of the *D. suzukii* genome assembly, they contain a relatively larger proportion of all identified SNPs (72%), and thus are still representative of genome-wide polymorphism.

To estimate genome-wide pairwise F_{ST} between populations, we adopted an unbiased multi-loci estimator known as Reynolds' estimator of the coancestry coefficient, which accounts for unequal sample sizes among populations and is applicable for more than two alleles at a site (Reynolds et al. 1983). We first heuristically partitioned the genome into windows that exceeded

a cross-sample average accumulated heterozygosity threshold of 100. The genome-wide F_{ST} was then calculated as an average of window-F_{ST} $F_{ST}(w)$ weighted by the number of analyzed sites within each window ($L(w)$), where $F_{ST}(w)$ was calculated as a weighted average of single-site ratio estimators. Numerically,

$$F_{ST} = \frac{1}{L}\sum_w L(w)\, F_{ST}(w) \quad (2)$$

$$F_{ST}(w) = (\sum_l a_l)/\sum_l (a_l + b_l) \quad (3)$$

where

$$a_l = \frac{1}{2}\sum_u (\tilde{p}_{1lu} - \tilde{p}_{2lu})^2 - \frac{(n_{1l}+n_{2l})(n_{1l}\tilde{\alpha}_{1l}+n_{2l}\tilde{\alpha}_{2l})}{4n_{1l}n_{2l}(n_{1l}+n_{2l}-1)} \quad (4)$$

$$a_l + b_l = \frac{1}{2}\sum_u (\tilde{p}_{1lu} - \tilde{p}_{2lu})^2 + \frac{(4n_{1l}n_{2l}-n_{1l}-n_{2l})(n_{1l}\tilde{\alpha}_{1l}+n_{2l}\tilde{\alpha}_{2l})}{4n_{1l}n_{2l}(n_{1l}+n_{2l}-1)} \quad (5)$$

At the lth site in each population, $\tilde{p}_{1lu}$ and $\tilde{p}_{2lu}$ represent the frequency of the uth allele at the lth site; $\tilde{\alpha}_{1l}$ and $\tilde{\alpha}_{2l}$ represent the heterozygosity; n_{1l} and n_{2l} represent the sample size. Unlike the sequencing of individual genomes, pool-seq induces an uncertainty in the number of individual alleles actually sequenced at a locus (i.e., effective sample size), and this uncertainty decreases slowly even at high read depth (Ferretti et al. 2013). Since the sample size is an important parameter for F_{ST} estimation, we took measures to obtain an estimate of the effective sample size, n_{il}, at each given site (Ferretti et al. 2013). Numerically,

$$n_{il} = \sum_j^{n_c} j P_c(j \mid n_r, n_c) \quad (6)$$

where

$$P_c(j \mid n_r, n_c) = \frac{n_c! S(n_r, j)}{(n_c - j)! n_c^{n_r}} \quad (7)$$

Here, we explicitly estimated the probability of the number of j unique lineages sampled at a site given n_r sampled reads and n_c equally contributing chromosomes in a pool, where $S(n_r, j)$ are the Stirling numbers of the second kind, defined as the number of ways to partition n_r reads into j non-empty sets (Ferretti et al. 2013). We then estimated n_{il} as the expected number of lineages for each n_r and n_c. Ideally, n_c should be estimated as $2n_e$, where n_e is the effective pool size representing the number of diploid individuals contributing the same amount of reads to a pool (Gautier et al. 2013; Lange et al. 2022). Although we lack sample replicates to estimate n_e and therefore used haploid sample size for n_c as an approximation, the probability estimation is still reasonable given that our number of lineages for each pool is large (Table S1) (Ferretti et al. 2013).

Lastly, we estimated genome-wide pairwise D_{XY} as an absolute measure of population differentiation that is independent of levels of within-population diversity. It was calculated as pairwise differences per site between two populations, divided by L total analyzed sites (Nei 1987; Hahn 2018). Numerically,

$$D_{XY} = \frac{1}{L}\sum_l \sum_{ij} x_i y_j k_{ij} \quad (8)$$

where x_i and y_j represent frequencies of the *i*th allele from population *X* and the *j*th allele from population *Y*, and k_{ij} is either 1 or 0, depending on whether or not the alleles differ at the lth site. Calculations in this section were all implemented with Python and Shell scripts (see Data Availability).

Building Population Trees and Admixture Graphs

To infer the population structure assuming no migrations, we used MEGA X (Stecher et al. 2020) to build neighbor-joining population trees (Figure S5) for autosomes and X chromosome from Reynolds' distance, which is a transformation of pairwise F_{ST} : $D = -ln\,(1 - F_{ST})$ (Reynolds et al. 1983). Since the position of the tree root is not identifiable through this analysis, and the China-Ningbo population was identified through diversity and distance analyses as a potentially basal lineage among the sampled populations, we rooted the tree along the branch to Ningbo.

To further investigate population splits and mixtures, we used TreeMix (Pickrell and Pritchard 2012) to infer population trees using allele counts that were corrected based on effective sample size. The input SNPs were a subset of our total identified SNPs whose total minor allele count across all samples was no less than ten, as it would allow us to complete this computationally intensive analysis in a reasonable amount of time and avoid biases on tree construction from false positive SNPs. We combined SNPs into blocks of 54 (-k 54) to account for linkage disequilibrium across windows that have a median length of ~500 bp.

To determine the number of migration events (-m) that improved the model fit while avoiding overfitting and losing interpretability, we ran TreeMix on our actual SNP data with no migration and with -m from one to 20. Since the position of the root is only partially identifiable through TreeMix inference, we rooted the tree at China-Ningbo, as suggested by Pickrell & Pritchard (2012). We then estimated the fraction of the variance in relatedness between populations that is accounted for by each model (f). We found that f began to plateau for both the autosomal and X-linked models at m = 11. For the X chromosome model, we also confirmed that all migration edges were statistically significant ($p < 0.05$) based on calculating their p-values from the jackknife estimates of the migration weights and their standard error

(Pickrell and Pritchard 2012), whereas this analysis remained computationally intractable for the autosomal case. We interpreted the migration weights *w* as an admixture proportion, as it was correlated with ancestry fraction in previous simulations, although it tends to be a more conservative approximation for high admixture proportions (Pickrell and Pritchard 2012).

Preparing Environmental Data

To generate environmental data for GEA, we selected a preliminary set of 26 candidate environmental variables representing geographic, climatic and land cover-related factors (Figure 4A) that may be relevant in the adaptation process of *D. suzukii* based on prior knowledge (Kellermann et al. 2012; Bogaerts-Márquez et al. 2021). With R packages 'raster' (v. 3.5.2) and 'SpatialPoints' (v. 1.4.6), we retrieved environmental data of high spatial resolution (~100 km^2) in batch for the sampling locations of our 29 populations from online databases WorldClim (Fick and Hijmans 2017) and Esri 2020 Land Cover (Karra et al. 2021). Annual mean values of monthly climatic variables, including mean wind speed, solar radiation, and water vapor pressure, were derived by averaging across 12 months of data.

Due to the large number of statistical tests that would result from running GEA on all the environmental variables one by one, there is an increased difficulty in controlling rates of false discovery. Additionally, including multiple highly correlated variables in a model would lead to multicollinearity issues (Rellstab et al. 2015). To avoid these problems, we calculated a pairwise Pearson correlation matrix from values of environmental factors across sampled locations (Figure 4A; Table S3), and then selected a subset of nine least correlated environmental variables for one-by-one GEA analyses (Figure 4B). To avoid scale inconsistencies between estimated GEA statistics, the environmental differentiation of each population was calculated as the absolute

882 difference between the environmental value of that population and the average across all
883 populations, standardized by the standard deviation (de Villemereuil and Gaggiotti 2015). This
884 standardized differentiation was then input to GEA (Table S3).

885

886 **Environmental Association Analyses**

887 To characterize the environmental adaptation of *D. suzukii*, we scanned the whole
888 genome for adaptive loci using the F_{ST}-based GEA method BayesScEnv (de Villemereuil and
889 Gaggiotti 2015). We chose this specific approach over other GEA methods mainly because it has
890 a lower false positive rate than other GEA approaches in the presence of hierarchical population
891 structure. It also allows for detecting patterns of allele frequency that are not linearly dependent
892 on environmental factors (Rellstab et al. 2015; de Villemereuil and Gaggiotti 2015).

893 For each environmental variable, the association analyses tested the relationship between
894 environmental and genetic differentiation among populations, for 5,752,156 genome-wide SNPs
895 with a MAF higher than 5%. To control for false positives, we chose stringent model parameters
896 expected to yield extremely conservative results, setting the prior probability of non-neutral
897 models as 0.02 (-pr_jump 0.02) and the prior probability of the competing environment-
898 unrelated locus-specific model as 0.9 (-pr_pref 0.9). These parameters correspond to assumptions
899 that genetic differentiation reflects the action of natural selection in just 2% of the genome, and
900 the focal environmental variable is only expected to be involved at 10% of the non-neutral loci.

901 To make this GEA analysis computationally feasible with our large SNP set, while still
902 analyzing all qualifying SNPs, we applied a split-run strategy: we subsampled SNPs across
903 concatenated sequences of contigs within the autosomes and the X chromosome separately, and

then ran subsamples with BayesScEnv in parallel. Since the null model of population structure is estimated separately in each run, we subsampled non-adjacent SNPs at a fixed interval to limit locus-specific biases in that estimation, where the length of the interval between jointly analyzed SNPs was equal to the total number of subsamples. With a targeted subsample/interval size of up to 10,000 SNPs, we divided the concatenated autosomal contigs into 490 subsamples (with actual subsample sizes of 9982-9983 SNPs), and the concatenated X-linked contigs into 87 subsamples (with actual subsample sizes of 9893-9894 SNPs). Hence, the first autosomal subsample contained SNP #1, SNP #491, and so on.

Convergence of each run was confirmed with the R package 'CODA'. Individual runs were then merged across autosomes and X chromosome to calculate the genome-wide q-value (q) of locally-estimated posterior error probability (PEP) across all sites, where we targeted a false discovery rate (FDR) of 5% by setting the q threshold at 0.05 (Storey 2003; Muller et al. 2006). For downstream analyses, to remove redundancy due to linkage disequilibrium, we pared down closely linked candidate sites by maintaining the site with the lowest q within each 20 kb genomic window. To assess the relative levels of support for associations between SNPs and a given environmental variable, we ranked all candidate loci first by q and then by the estimated g parameter as a tie-breaker, which measures the sensitivity of a locus to environmental differentiation.

Identifying Candidate Genes

For each candidate SNP, the closest gene in each direction within a 200-exon flanking region that overlapped with it was considered to be associated with that variant, in order to encompass both potential coding and regulatory adaptation. To facilitate clear comparisons

among environmental variables with different numbers of significant variants, we focused on the top 500 candidate genes that were linked to variants with the lowest significant q and highest g within each environmental variable (Table S5).

GO Enrichment and Semantic Clustering

Gene ontology (GO) enrichment of the top 500 candidate genes associated with candidate SNPs was performed via genomic permutation of outlier SNP positions (100,000,000 replicates), which accounts for the variability of gene length and the clustering of functionally related genes, as described in previous work (Pool et al. 2017). For each GO category, a p-value indicated the proportion of permutation replicates in which an equal or greater number of genes was implicated.

We then prioritized the most informative and significant GO terms and removed redundant terms that potentially share similar groups of genes by clustering GO terms based on their semantic similarity and ranking representative terms of each cluster by their p-value (Reijnders and Waterhouse 2021). For GO terms that were shared among associations with multiple environmental variables, a combined p-value was calculated from the p-values of independent enrichment tests using Fisher's method (Fisher 1938).

Supplementary Material

Supplementary figures S1–S9 and tables S1–S6 are available at Molecular Biology and Evolution online.

Data Availability

All sequence data generated for this project are available from the NIH Short Read Archive under project PRJNA973110, with specific sample information given in Table S1, Supplementary Material online. All computational scripts created for this study have been uploaded to https://github.com/Sfeng666/poolWGS2SNP and https://github.com/Sfeng666/Dsuz_popgen_GEA (last accessed May 27, 2023).

Acknowledgments

We thank members of the Pool lab for helpful comments on this manuscript. We also thank Masahito Kimura, Samantha Tochen, and Carandale Farms for assistance with fly collection. We also thank Mathilde Paris for providing a multiple alignment file across *Drosophila* species and updated genomic annotations, as well as Martin Kapun for assistance with SNP calling, and Pierre de Villemereuil for helping with our GEA analyses. The UW-Madison Center for High Throughput Computing provided computational assistance and resources for this work. This work was funded by USDA Hatch grant WIS01900 (to SDS, CG, and JEP) and by National Institutes of Health grant R35 GM13630 (to JEP).

References

Adrion JR, Kousathanas A, Pascual M, Burrack HJ, Haddad NM, Bergland AO, Machado H, Sackton TB, Schlenke TA, Watada M, et al. 2014. *Drosophila suzukii*: the genetic footprint of a recent, worldwide invasion. *Mol Biol Evol* 31:3148–3163.

Alloway PG, Dolph PJ. 1999. A role for the light-dependent phosphorylation of visual arrestin. *Proc Natl Acad Sci USA* 96:6072–6077.

Asplen MK, Anfora G, Biondi A, Choi D-S, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD, et al. 2015. Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. *J Pest Sci* 88:469–494.

Atallah J, Teixeira L, Salazar R, Zaragoza G, Kopp A. 2014. The making of a pest: the evolution of a fruit-penetrating ovipositor in *Drosophila suzukii* and related species. *Proc Roy Soc B* 281:20132840.

Auwera GAV der, O'Connor BD. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media.

Avidor-Reiss T, Leroux MR. 2015. Shared and distinct mechanisms of compartmentalized and cytosolic ciliogenesis. *Curr Biol* 25:R1143–R1150.

Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, Zuker CS. 2004. Decoding Cilia Function: Defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 117:527–539.

Barange M, Cheung WWL, Merino G, Perry RI. 2010. Modelling the potential impacts of climate change and human activities on the sustainability of marine resources. *Curr Opin Environ Sustain* 2:326–333.

Batie M, Druker J, D'Ignazio L, Rocha S. 2017. KDM2 family members are regulated by HIF-1 in hypoxia. *Cells* 6:8.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLOS Biol* 5:e310.

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

Bogaerts-Márquez M, Guirao-Rico S, Gautier M, González J. 2021. Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of *Drosophila melanogaster*. *Mol Ecol* 30:938–954.

Boutros M, Ahringer J. 2008. The art and design of genetic screens: RNA interference. *Nat Rev Genet* 9:554–566.

1001 Calboli FCF, Gilchrist GW, Partridge L. 2003. Different cell size and cell number contribution in
1002 two newly established and one ancient body size cline of *Drosophila subobscura*. *Evolution*
1003 57:566–573.

1004 Campbell M, Ganetzky B. 2012. Extensive morphological divergence and rapid evolution of the
1005 larval neuromuscular junction in *Drosophila*. *Proc Natl Acad Sci USA* 109:E648–E655.

1006 Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on
1007 neutral molecular variation. *Genetics*. 34:1289-1303.

1008 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
1009 *Bioinformatics* 34:i884–i890.

1010 Cini A, Ioriatti C, Anfora G. 2012. A review of the invasion of *Drosophila suzukii* in Europe and a
1011 draft research agenda for integrated pest management. *Bull Insectology* 65:149–160.

1012 Clark MS, Thorne MA, Purać J, Burns G, Hillyard G, Popović ŽD, Grubor-Lajšić G, Worland
1013 MR. 2009. Surviving the cold: molecular analyses of insect cryoprotective dehydration in
1014 the Arctic springtail *Megaphorura arctica* (Tullberg). *BMC Genomics* 10:328.

1015 Clemente F, Gautier M, Vitalis R. 2018. Inferring sex-specific demographic history from SNP
1016 data. *PLOS Genet* 14:e1007191.

1017 Colautti RI, Barrett SCH. 2013. Rapid adaptation to climate facilitates range expansion of an
1018 invasive plant. *Science* 342:364–366.

1019 Colautti RI, Lau JA. 2016. Contemporary evolution during invasion. In: Invasion Genetics. John
1020 Wiley & Sons, Ltd. p. 101–121.

1021 Duvall LB, Taghert PH. 2013. E and M circadian pacemaker neurons use different PDF
1022 receptor signalosome components in drosophila. *J Biol Rhythms* 28:239–248.

1023 Elaine Merrill C, Sherertz TM, Walker WB, Zwiebel LJ. 2005. Odorant-specific requirements
1024 for arrestin function in *Drosophila* olfaction. *J Neurobiol* 63:15–28.

1025 Enriquez T, Colinet H. 2019. Cold acclimation triggers lipidomic and metabolic adjustments in
1026 the spotted wing drosophila *Drosophila suzukii* (Matsumara). *Am J Physiol Regul Integr Comp*
1027 *Physiol* 316:R751–R763.

1028 Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. 2013. Population genomics from pool
1029 sequencing. *Mol Ecol* 22:5561–5576.

1030 Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global
1031 land areas. *Int J Climatol* 37:4302–4315.

1032 Fisher RA. 1938. Statistical methods for research workers. Oliver and Boyd, Edinburgh.

1033 Fraimout A, Debat V, Fellous S, Hufbauer RA, Foucaud J, Pudlo P, Marin J-M, Price DK,
1034 Cattel J, Chen X, et al. 2017. Deciphering the routes of invasion of *Drosophila suzukii* by
1035 means of ABC random forest. *Mol Biol Evol* 34:980–996.

1036 Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P,
1037 Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from next-
1038 generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol* 22:3766–
1039 3779.

1040 Gautier M, Vitalis R, Flori L, Estoup A. 2022. f-Statistics estimation and admixture graph
1041 construction with Pool-Seq or allele count data using the R package poolfstat. *Mol Ecol*
1042 *Resour* 22:1394–1416.

1043 Gibbs AG, Fukuzato F, Matzkin LM. 2003. Evolution of water conservation mechanisms in
1044 Drosophila. *J Exp Biol* 206:1183–1192.

1045 Gibert P, Debat V, Ghalambor CK. 2019. Phenotypic plasticity, global change, and the speed of
1046 adaptive evolution. *Curr Op Insect Sci* 35:34–40.

1047 Gilchrist AS, Partridge L. 1999. A comparison of the genetic basis of wing size divergence in
1048 three parallel body size clines of *Drosophila melanogaster*. *Genetics* 153:1775–1787.

1049 Guha A, Kornberg TB. 2005. Tracheal branch repopulation precedes induction of the
1050 Drosophila dorsal air sac primordium. *Devel Biol* 287:192–200.

1051 Hahn MW. 2018. Molecular population genetics. Sunderland, MA: Sinauer Associates.

1052 Hämälä T, Gorton AJ, Moeller DA, Tiffin P. 2020. Pleiotropy facilitates local adaptation to
1053 distant optima in common ragweed (Ambrosia artemisiifolia). *PLOS Genet* 16:e1008707.

1054 Hamby KA, E. Bellamy D, Chiu JC, Lee JC, Walton VM, Wiman NG, York RM, Biondi A.
1055 2016. Biotic and abiotic factors impacting development, behavior, phenology, and
1056 reproductive biology of *Drosophila suzukii*. *J Pest Sci* 89:605–619.

1057 Han Z, Li X, Wu J, Olson EN. 2004. A myocardin-related transcription factor regulates activity
1058 of serum response factor in *Drosophila*. *Proc Natl Acad Sci USA* 101:12567–12572.

1059 Helfrich-Förster C, Bertolini E, Menegazzi P. 2020. Flies as models for circadian clock
1060 adaptation to environmental challenges. *Eur J NeuroSci* 51:166–181.

1061 Hill T, Koseva BS, Unckless RL. 2019. The genome of *Drosophila innubila* reveals lineage-specific
1062 patterns of selection in immune genes. *Mol Biol Evol* 36:1405–1417.

1063 Hodgins KA, Bock DG, Rieseberg LH. 2018. Trait evolution in invasive species. In: Annual
1064 Plant Reviews online. New York: John Wiley & Sons, Ltd. p. 459–496.

1065 Honjo K, Mauthner SE, Wang Y, Skene JHP, Tracey WD. 2016. Nociceptor-enriched genes
1066 required for normal thermal nociception. *Cell Rep* 16:295–303.

1067 Horváth V, Guirao-Rico S, Salces-Ortiz J, Rech GE, Green L, Aprea E, Rodeghiero M, Anfora
1068 G, González J. 2023. Gene expression differences consistent with water loss reduction
1069 underlie desiccation tolerance of natural *Drosophila* populations. *BMC Biol* 21:35.

1070 James RE, Broihier HT. 2011. Crimpy inhibits the BMP homolog Gbb in motoneurons to
1071 enable proper growth control at the *Drosophila* neuromuscular junction. *Development*
1072 138:3273–3286.

1073 Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-
1074 Stabelli O, Kankare M, Bogaerts-Márquez M, et al. 2020. Genomic analysis of European
1075 *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection,
1076 and previously unknown DNA viruses. *Mol Biol Evol* 37:2661–2678.

1077 Karra K, Kontgis C, Statman-Weil Z, Mazzariello JC, Mathis M, Brumby SP. 2021. Global
1078 land use / land cover with Sentinel 2 and deep learning. In: 2021 IEEE International
1079 Geoscience and Remote Sensing Symposium IGARSS. p. 4704–4707.

1080 Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. *Ecol Let* 7:1225–1241.

1081 Kellermann V, Loeschcke V, Hoffmann AA, Kristensen TN, Fløjgaard C, David JR, Svenning J,
1082 Overgaard J. 2012. Phylogenetic constraints in key functional traits behind species'
1083 climate niches: patterns of desiccation and cold resistance across 95 *Drosophila* species.
1084 *Evolution* 66:3377–3389.

1085 Kerekes É, Kókai E, Páldy FS, Dombrádi V. 2014. Functional analysis of the glycogen binding
1086 subunit CG9238/Gbs-70E of protein phosphatase 1 in *Drosophila melanogaster*. *Insect*
1087 *Biochem Mol Biol* 49:70–79.

1088 Kinsler G, Geiler-Samerotte K, Petrov DA. 2020. Fitness variation across subtle environmental
1089 perturbations reveals local modularity and global pleiotropy of adaptation. *eLife* 9:e61271.

1090 Knapp L, Mazzi D, Finger R. 2021. The economic impact of *Drosophila suzukii*: perceived costs
1091 and revenue losses of Swiss cherry, plum and grape growers. *Pest Manag Sci* 77:978–1000.

1092 Koch JB, Dupuis JR, Jardeleza M-K, Ouedraogo N, Geib SM, Follett PA, Price DK. 2020.
1093 Population genomic and phenotype diversity of invasive *Drosophila suzukii* in Hawai'i. *Biol*
1094 *Invasions* 22:1753–1770.

1095 Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes: an
1096 expanded *Drosophila* Genome Nexus. *Mol Biol Evol* 33:3308–3313.

1097 Lack JB, Yassin A, Sprengelmeyer QD, Johanning EJ, David JR, Pool JE. 2016. Life history
1098 evolution and cellular mechanisms associated with increased size in high-altitude
1099 *Drosophila*. *Ecol Evol* 6:5893–5906.

1100 Lange JD, Bastide H, Lack JB, Pool JE. 2022. A population genomic assessment of three decades
1101 of evolution in a natural *Drosophila* population. *Mol Biol Evol* 39:msab368.

1102 Lange JD, Pool JE. 2018. Impacts of Recurrent Hitchhiking on Divergence and Demographic
1103 Inference in *Drosophila*. *Genome Biol Evol* 10:1882–1891.

1104 Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C,
1105 Corbett-Detig RB, Kolaczkowski B, et al. 2012. Genomic variation in natural populations
1106 of *Drosophila melanogaster*. *Genetics* 192:533–598.

1107 Lee CE. 2002. Evolutionary genetics of invasive species. *Trends Ecol Evol* 17:386–391.

1108 Levashina EA, Ohresser S, Bulet P, Reichhart J-M, Hetru C, Hoffmann JA. 1995.
1109 Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with
1110 antibacterial and antifungal properties. *Eur J Biochem* 233:694–700.

1111 Lewald KM, Abrieux A, Wilson DA, Lee Y, Conner WR, Andreazza F, Beers EH, Burrack HJ,
1112 Daane KM, Diepenbrock L, et al. 2021. Population genomics of *Drosophila suzukii* reveal
1113 longitudinal population structure and signals of migrations in and out of the continental
1114 United States. *G3* 11:jkab343.

1115 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM,
1116 http://arxiv.org/abs/1303.3997, last accessed May 31, 2023.

1117 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
1118 1000 genome project data processing subgroup. 2009. The Sequence Alignment/Map
1119 format and SAMtools. *Bioinformatics* 25:2078–2079.

1120 Li W, Wang F, Menut L, Gao F-B. 2004. BTB/POZ-zinc finger protein abrupt suppresses
1121 dendritic branching in a neuronal subtype-specific and dosage-dependent manner. *Neuron*
1122 43:823–834.

1123 Little CM, Chapman TW, Hillier NK. 2020. Plasticity is key to success of *Drosophila suzukii*
1124 (Diptera: Drosophilidae) invasion. *J Insect Sci* 20:5.

1125 Manko P, de Figueroa JMT, Oboňa J. 2021. Responses of the pest *Drosophila suzukii* population
1126 to environmental variables in southern Spain (Granada, Andalusia) – results of a pilot
1127 study. *Acta Mus Sil Sci Nat* 70:65–74.

1128 Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. 23:23-
1129 35.

1130 Mérel V, Gibert P, Buch I, Rodriguez Rada V, Estoup A, Gautier M, Fablet M, Boulesteix M,
1131 Vieira C. 2021. The worldwide invasion of *Drosophila suzukii* is accompanied by a large
1132 increase of transposable element load and a small number of putatively adaptive
1133 insertions. *Mol Biol Evol* 38:4252–4267.

1134 Muller P, Parmigiani G, Rice K. 2006. FDR and Bayesian multiple comparisons rules. *JHU*
1135 *Biostatistics* 115.

1136 Nakabayashi R, Morishita S. 2020. HiC-Hiker: a probabilistic model to determine contig
1137 orientation in chromosome-length scaffolds with Hi-C. *Bioinformatics* 36:3966–3974.

1138 Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

1139 Olazcuaga L, Loiseau A, Parrinello H, Paris M, Fraimout A, Guedot C, Diepenbrock LM, Kenis
1140 M, Zhang J, Chen X, et al. 2020. A whole-genome scan for association with invasion
1141 success in the fruit fly *Drosophila suzukii* using contrasts of allele frequencies corrected for
1142 population structure. *Mol Biol Evol* 37:2369–2385.

1143 Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P,
1144 Varotto C, Pisani D, et al. 2013. Linking genomics and ecology to investigate the
1145 complex evolution of an invasive *Drosophila* pest. *Genome Biol Evol* 5:745–757.

1146 Palacios-Muñoz A, Ewer J. 2018. Calcium and cAMP directly modulate the speed of the
1147 Drosophila circadian clock. *PLOS Genet* 14:e1007433.

1148 Paris M, Boyer R, Jaenichen R, Wolf J, Karageorgi M, Green J, Cagnon M, Parinello H, Estoup
1149 A, Gautier M, et al. 2020. Near-chromosome level genome assembly of the fruit pest
1150 *Drosophila suzukii* using long-read sequencing. *Sci Rep* 10:11227.

1151 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide
1152 allele frequency data. *PLOS Genet* 8:e1002967.

1153 Pool JE. 2015. The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D.*
1154 *melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol Biol*
1155 *Evol* 32:3236–3251.

1156 Pool JE, Braun DT, Lack JB. 2017. Parallel evolution of cold tolerance within *Drosophila*
1157 *melanogaster*. *Mol Biol Evol* 34:349–360.

1158 Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P,
1159 Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population genomics of sub-Saharan
1160 *Drosophila melanogaster*: African diversity and non-African admixture. *PLOS Genetics*
1161 8:e1003080.

1162 Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. *Evolution*
1163 61:3001–3006.

1164 Pool JE, Nielsen R. 2009. Inference of historical changes in migration rate from the lengths of
1165 migrant tracts. *Genetics* 181:711–719.

1166 Prakash S, Caldwell JC, Eberl DF, Clandinin TR. 2005. *Drosophila* N-cadherin mediates an
1167 attractive interaction between photoreceptor axons and their targets. *Nat Neurosci* 8:443–
1168 450.

1169 Prelich G. 2012. Gene overexpression: uses, mechanisms, and interpretation. *Genetics* 190:841–
1170 854.

1171 Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. 2008. Adaptive evolution in
1172 invasive species. *Trends Plant Sci* 13:288–294.

1173 Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP. 2016. Reliable ABC model
1174 choice via random forests. *Bioinformatics* 32:859–866.

1175 Ray N, Adams JM. 2001. A GIS-based vegetation map of the world at the last glacial maximum
1176 (25,000-15,000 BP). *Internet Archaeol* 11.

1177 Reijnders MJMF, Waterhouse RM. 2021. Summary visualizations of Gene Ontology terms with
1178 GO-Figure! *Front Bioinform* 1:6.

1179 Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to
1180 environmental association analysis in landscape genomics. *Mol Ecol* 24:4348–4370.

1181 Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a
1182 short-term genetic distance. *Genetics* 105:767–779.

1183 Rezende GL, Martins AJ, Gentile C, Farnesi LC, Pelajo-Machado M, Peixoto AA, Valle D.
1184 2008. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized
1185 Serosal Cuticle. *BMC Devel Biol* 8:82.

1186 Reznick DN, Losos J, Travis J. 2019. From low to high gear: there has been a paradigm shift in
1187 our understanding of evolution. *Ecol Let* 22:233–244.

1188 Rogers WA, Salomone JR, Tacy DJ, Camino EM, Davis KA, Rebeiz M, Williams TM. 2013.
1189 Recurrent modification of a conserved cis-regulatory element underlies fruit fly
1190 pigmentation diversity. *PLOS Genet* 9:e1003740.

1191 Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T,
1192 Petrov DA, Schmidt P. 2022. Direct observation of adaptive tracking on ecological time
1193 scales in *Drosophila*. *Science* 375:eabj7484.

1194 Santos LA dos, Mendes MF, Krüger AP, Blauth ML, Gottschalk MS, Garcia FRM. 2017.
1195 Global potential distribution of *Drosophila suzukii* (Diptera, Drosophilidae). *PLOS ONE*
1196 12:e0174318.

1197 Schrider DR, Shanku AG, Kern AD. 2016. Effects of linked selective sweeps on demographic
1198 inference and model selection. *Genetics* 204:1207–1223.

1199 Senthilan PR, Piepenbrock D, Ovezmyradov G, Nadrowski B, Bechstedt S, Pauls S, Winkler M,
1200 Möbius W, Howard J, Göpfert MC. 2012. *Drosophila* auditory organ genes and genetic
1201 hearing defects. *Cell* 150:1042–1054.

1202 Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR–
1203 Cas9. *Nat Rev Genet* 16:299–311.

1204 Shearer PW, West JD, Walton VM, Brown PH, Svetec N, Chiu JC. 2016. Seasonal cues induce
1205 phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. *BMC Ecol* 16:11.

1206 Simoes da Silva CJ, Sospedra I, Aparicio R, Busturia A. 2019. The microRNA-306/abrupt
1207 regulatory axis controls wing and haltere growth in *Drosophila*. *Mech Dev* 158:103555.

1208 Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for
1209 macOS. *Mol Biol Evol* 37:1237–1239.

1210 Stern DL. 2014. Identification of loci that cause phenotypic variation in diverse species with the
1211 reciprocal hemizygosity test. *Trends Genet* 30:547–554.

1212 Storey JD. 2003. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann*
1213 *Stat* 31:2013–2035.

1214 Sugimura K, Satoh D, Estes P, Crews S, Uemura T. 2004. Development of Morphological
1215 Diversity of Dendrites in *Drosophila* by the BTB-Zinc Finger Protein Abrupt. *Neuron*
1216 43:809–822.

1217 Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ,
1218 Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny
1219 of 155 *Drosophila* genomes. *Curr Biol* 32:111-123.e5.

1220 Svedberg J, Shchur V, Reinman S, Nielsen R, Corbett-Detig R. 2021. Inferring adaptive
1221 introgression using hidden Markov models. *Mol Biol Evol* 38:2152–2165.

1222 Tait G, Grassi A, Pfab F, Crava CM, Dalton DT, Magarey R, Ometto L, Vezzulli S, Rossi-
1223 Stacconi MV, Gottardello A, et al. 2018. Large-scale spatial dynamics of *Drosophila suzukii*
1224 in Trentino, Italy. *J Pest Sci* 91:1213–1224.

1225 Telonis-Scott M, Gane M, DeGaris S, Sgrò CM, Hoffmann AA. 2012. High resolution mapping
1226 of candidate alleles for desiccation resistance in *Drosophila melanogaster* under selection. *Mol*
1227 *Biol Evol* 29:1335–1351.

1228 True JR, Mercer JM, Laurie CC. 1996. Differences in crossover frequency and distribution
1229 among three sibling species of *Drosophila*. *Genetics* 142:507–523.

1230 Turner TL. 2014. Fine-mapping natural alleles: quantitative complementation to the rescue. *Mol*
1231 *Ecol* 23:2377–2382.

1232 Ueno K, Kidokoro Y. 2008. Adenylyl cyclase encoded by AC78C participates in sugar
1233 perception in *Drosophila melanogaster*. *Eur J Neurosci* 28:1956–1966.

1234 de Villemereuil P, Gaggiotti OE. 2015. A new F_{ST}-based method to uncover local adaptation
1235 using environmental variables. *Methods in Ecology and Evolution* 6:1248–1258.

1236 Wallingford AK, Rice KB, Leskey TC, Loeb GM. 2018. Overwintering behavior of *drosophila*
1237 *suzukii*, and potential springtime diets for egg maturation. *Environ Entomol* 47:1266–1273.

1238 Wang S-H, Simcox A, Campbell G. 2000. Dual role for *Drosophila* epidermal growth factor
1239 receptor signaling in early wing disc development. *Genes Dev* 14:2271–2276.

1240 Wang T, Morency DT, Harris N, Davis GW. 2020. Epigenetic signaling in glia controls
1241 presynaptic homeostatic plasticity. *Neuron* 105:491-505.e3.

1242 Wang Y, Farine J-P, Yang Y, Yang J, Tang W, Gehring N, Ferveur J-F, Moussian B. 2020.
1243 Transcriptional control of quality differences in the lipid-based cuticle barrier in *Drosophila*
1244 *suzukii* and *Drosophila melanogaster*. *Front Genet* 11:887.

1245 Wang Y, Ferveur J-F, Moussian B. 2021. Eco-genetics of desiccation resistance in *Drosophila*. *Biol*
1246 *Rev* 96:1421–1440.

1247 Weißinger L, Schrieber K, Breuer M, Müller C. 2019. Influences of blackberry margins on
1248 population dynamics of *Drosophila suzukii* and grape infestation in adjacent vineyards. *J*
1249 *Appl Entomol* 143:802–812.

1250 Welles SR, Dlugosch KM. 2019. Population genomics of colonization and invasion. In: Rajora
1251 OP, editor. Population genomics: concepts, approaches and applications. Population
1252 Genomics. Cham: Springer International Publishing. p. 655–683.

1253 Wong MM-K, Liu M-F, Chiu SK. 2015. Cropped, *Drosophila* transcription factor AP-4, controls
1254 tracheal terminal branching and cell growth. *BMC Devel Biol* 15:20.