Genomic Diversity Illuminates the Environmental Adaptation of *Drosophila suzukii*

Running title: *Drosophila suzukii* Environmental Adaptation

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:

Siyuan Feng, siyuanfeng.bio@outlook.com; John E. Pool, jpool@wisc.edu

Mailing address: 425G Henry Mall, Madison, WI, 53705, USA

Abstract

Biological invasions carry substantial practical and scientific importance, and represent natural evolutionary experiments on contemporary timescales. Here, we investigated genomic diversity 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. Through a multifaceted analysis of this population genomic data, we increase our understanding of the *D. suzukii* genome, its diversity and its evolution, and we identify an appropriate genotype-environment association pipeline for our data set. Using this approach, we detect genetic signals of local adaptation associated with nine distinct environmental factors related to altitude, wind speed, precipitation, temperature, and human land use. We uncover unique functional signatures for each environmental variable, such as a prevalence of cuticular genes associated with annual precipitation. We also infer 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 to top genes associated with all nine environmental variables. Our findings therefore depict a finer-scale adaptive landscape underlying the rapid invasion success of this agronomically important species.

Keywords: environmental adaptation, genotype-environment association, *Drosophila suzukii*, invasion genomics

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 (Peng 1937; Kanzawa 1939; Tan et al. 1949). It first expanded to Hawaii in 1980, and within the past 15 years, it has invaded North America and Europe, and subsequently Réunion Island (Indian Ocean) and South America (Asplen et al. 2015), then northern and sub-Saharan Africa (Boughdad et al. 2021; Kwadha et al. 2021). *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

57 fruit crops (Knapp et al. 2021). It also exhibits remarkable genetic diversity and phenotypic
58 plasticity in behavior, morphology, and physiology, e.g., temperature and desiccation tolerance
59 (Little et al. 2020; Olazcuaga et al. 2020), which may notably facilitate its adaptation to different
60 climatic conditions and host plants (Gibert et al. 2019; Little et al. 2020). To obtain a
61 comprehensive evolutionary genetic understanding of the invasion success of *D. suzukii*, we need
62 to understand the genetic basis and ecological drivers of adaptive evolutionary changes that have
63 allowed this species to occupy diverse worldwide environments.

64 Multiple genetic studies have investigated the demographic history of invasive
65 populations of *D. suzukii*. Such studies have found greater levels of genetic structure between
66 than within continents (Adrion et al. 2014; Lewald et al. 2021), suggesting independent invasions
67 into Europe and North America. These inferences were supported by an approximate Bayesian
68 analysis of microsatellite data, which also indicated that some invading populations had multiple
69 genetic sources (Fraimout et al. 2017). While minor differences are suggested in the specific
70 admixture events that have occurred (Fraimout et al. 2017; Lewald et al. 2021), and inferences of
71 the geographic origins of invading populations are limited by incomplete sampling from the
72 species' Asian range, these studies suggest some emerging consensus about the invasion history of
73 *D. suzukii*.

74 In contrast, the genetic basis of environmental adaptation in *D. suzukii* is still largely
75 unexplored. Among the few relevant studies is that of Olazcuaga et al. (2020), which used
76 genomic sequencing data from 22 worldwide populations to search for SNPs with greater
77 frequency differences between Asian (China, Japan) and non-Asian populations than are
78 observed at most loci due to founder event bottlenecks, in hopes of identifying genetic variants
79 that may underlie the invasion success of introduced populations. A subsequent study examining

80 this same data set also found a small number of transposon insertions with strong frequency
81 differences between American/European and Asian populations (Mérel et al. 2021). Another
82 study identified F_{ST} outliers among Hawaiian *D. suzukii* populations (Koch et al. 2020). Apart
83 from these limited comparisons, the genetic changes that may have helped *D. suzukii* to adapt to
84 specific environmental changes are entirely unknown.

85 With the increasing availability of multi-population genomic resources, genotype-
86 environment association (GEA), also known as environmental association analysis (EAA), is
87 becoming a widely-used approach to understand the potential relationship between specific
88 environmental factors and adaptive genetic variation (Rellstab et al. 2015). GEA is also useful in
89 identifying subtle changes in allele frequencies that are difficult to detect with outlier tests based
90 on traditional population genomic approaches, especially when the number of studied
91 populations is relatively large, and there is high gene flow counteracting patterns of local
92 adaptation (Kawecki and Ebert 2004). The capability of GEA to identify adaptive genetic
93 changes and environmental drivers of local adaptation has been demonstrated with whole-
94 genome pool-seq data from *Drosophila melanogaster* (Bogaerts-Márquez et al. 2021). Therefore,
95 GEA could be helpful in understanding the genotype-environment relationships underlying the
96 invasion success of *D. suzukii*. Performing GEA on this model system could also facilitate insights
97 into environment-driven genetic changes during rapid biological invasions, as well as
98 mechanisms driving insect adaptation to rapid climate shifts *in situ*.

99 In the present study, we perform population genomic analyses on whole-genome pool-seq
100 data of 29 population samples from native and invasive ranges to investigate the environmental
101 adaptation of *D. suzukii*. We investigate the geographic pattern of genetic diversity and
102 population structure, in part to inform our choice of GEA methodology. We then test the

103 association between SNP frequencies and nine environmental variables across sampling
104 locations, identifying both specific and shared functional signatures of adaptation to these diverse
105 selective pressures.

106

107 **Results**

108

109 **Genomic Diversity and Population Structure of *D. suzukii***

110 To investigate the genetic diversity of *D. suzukii* and to examine the genetic input for
111 environmental association analysis, we summarized the genomic polymorphism of 29 *D. suzukii*
112 populations derived from Asia (n=8), Europe (n = 11), and North and South America (n = 10),
113 encompassing both newly reported and previously published samples (Figure 1A; Table S1;
114 Olazcuaga et al. 2020). Whole-genome sequences were obtained from 29 pooled samples
115 consisting of 50-212 female and male individuals. The depth of mapped reads after quality
116 control ranged from 23X to 66X among population samples, with an average of 45X (Table S1).

117 We first estimated nucleotide diversity (π_S) across synonymous SNPs to investigate the
118 effects of rapid invasions on neutral genetic diversity. The genome-wide π_S ranged from
119 averages of 0.045 (autosomes) and 0.025 (X chromosome) in introduced populations, to averages
120 of 0.051 (autosomes) and 0.034 (X chromosome) in Asian populations (Table S2). Previous
121 studies have hypothesized a relatively wide native range of *D. suzukii* in East and Southeast Asia
122 (Adrion et al. 2014; Fraimout et al. 2017), and the acute drop in π_S of introduced European and
123 American populations relative to that of the native Chinese and Japanese populations reflects
124 previously reported founder event bottlenecks (Figure 1B, diagonal; Figure S1, S2A). The

125 observed patterns were also recapitulated with SNPs at all types of sites (Figure S3). Lower
126 π_S and more contrasting among-population π_S differences in X chromosome reflect effectively
127 prolonged bottlenecks due to their lower effective population size (Pool & Nielsen, 2007; Figure
128 S1). We also observed a greater loss of rare alleles in the invasive populations as a typical
129 consequence of bottlenecks (Figure S4). These founder events also increased genetic
130 differentiation as measured by F_{ST}, especially between continents (Figure 1B; Figure S3), and
131 particularly for the X chromosome (Figure S2A).

132 Subsequent to our analysis, contamination in three of the published samples was
133 reported, involving reads from the closely-related *D. subpulchrella* and from *D. immigrans* (Gautier
134 2023), which can be estimated but not fully removed. The estimated level of *D. subpulchrella*-
135 sourced contamination for Tokyo, Japan (JP-Tok) was reduced from 4.47% to 3.58% after our
136 alignment and quality filtering pipeline (Table S3), which is slightly higher than the baseline noise
137 level of 1.14% to 2.78% from samples known to be pure *D. suzukii* (Gautier 2023). The published
138 DE-Jen sample's contamination from *D. immigrans* was reduced from 5.79% to 0.35%. Our
139 unpublished sample from the same area showed 0.28% *D. immigrans* contamination from mapped
140 reads, whereas the other unpublished samples showed no evidence for contamination (Table S3).
141 However, a higher proportion of reads from *D. subpulchrella* remained in the east Asian sample
142 CN-Nin (10.38%, versus 14.95% from raw reads). Hence, it is likely that the elevated π_S of CN-
143 Nin (and likewise its elevated D_{XY}values from the analysis described below) reflect an artefact of
144 this contamination.

145 We next analyzed genetic structure among the sampled populations, using the summary
146 statistics D_{XY} and F_{ST}, and via principal component analysis (PCA) of population allele
147 frequencies, particularly since the pattern of genetic structure present may influence the

148 performance of GEA methods (Rellstab et al. 2015). The top three principal components (PC 1-
149 3) explained 57.44% (autosomes, Figure 1C) and 72.35% (X chromosome, Figure S2B) of the
150 variance among populations. In both autosomes and X, three-dimensional PCA and matrices of
151 both D_{XY} and window F_{ST} recapitulated both continuous and hierarchical geographic structure,
152 which must be accounted for in GEA. These results together showed the expected clustering of
153 populations into four distinct ranges (East Asia, Hawaii, Americas, and Europe; Figure 1B, 1C;
154 Figure S2). Much of the observed population differentiation is most likely due to founder event
155 bottlenecks and admixture during worldwide expansion (Fraimout et al. 2017), whereas
156 migration following population establishment would need to be overwhelmingly high to have
157 significant impacts given the very brief time scale of the global invasion. The two more northerly
158 populations from the western US, Oregon (US-Sok) and central California (US-Wat), show an
159 affinity with the Hawaiian population, which aligns with the suggestion that these populations
160 received a genetic contribution from Hawaii in addition to East Asia (Fraimout et al. 2017), while
161 populations from southern California and the central and eastern US show less evidence of such
162 admixture.

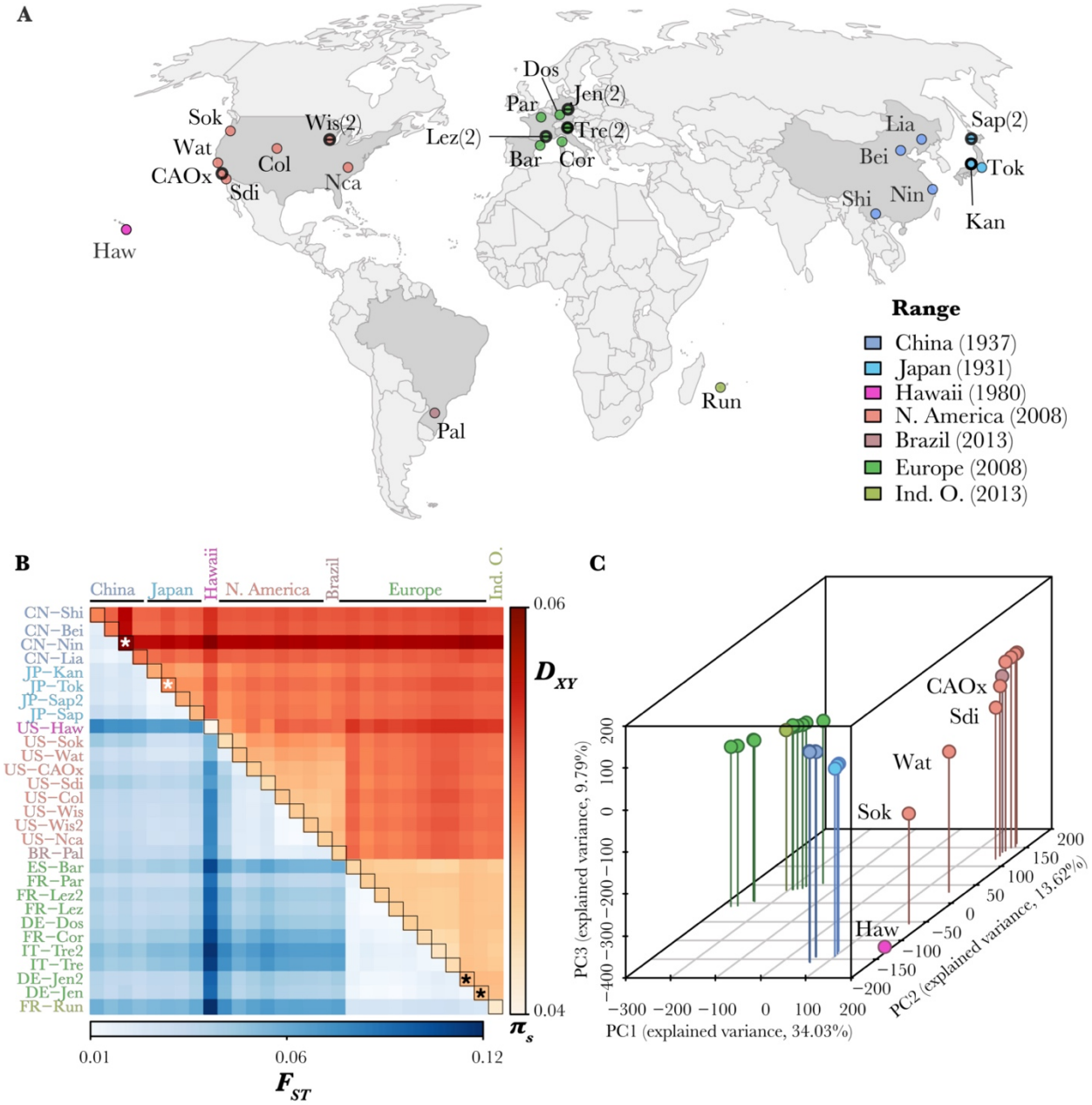
163 Our focus above on diversity and differentiation at synonymous sites was motivated by
164 the low selective constraint expected at these sites, as confirmed by our analysis of divergence
165 between *D. suzukii* and its relative *D. biarmipes* (Ometto et al. 2013; Suvorov et al. 2022; Figure
166 S5). We also note that compared to the divergence between *D. melanogaster* and *D. simulans*
167 (Begun et al. 2007; Lange and Pool 2018), *D. suzukii* and *D. biarmipes* show relatively higher ratios
168 intergenic (or intronic) divergence to non-synonymous divergence (Figure S5). Compared with *D.*
169 *melanogaster*, the genome of *D. suzukii* contains a notable expansion of repetitive sequences (Paris et
170 al. 2020), which could reflect a lower long-term effective population size (N_e) in the *D. suzukii*

171 lineage. However, the greater π of *D. suzukii* than *D. melanogaster* (Figure S1; Lack et al. 2016a;

172 Lewald et al. 2021) could instead indicate a greater N_e for *D. suzukii* within the past $4N_e$

173 generations.

174

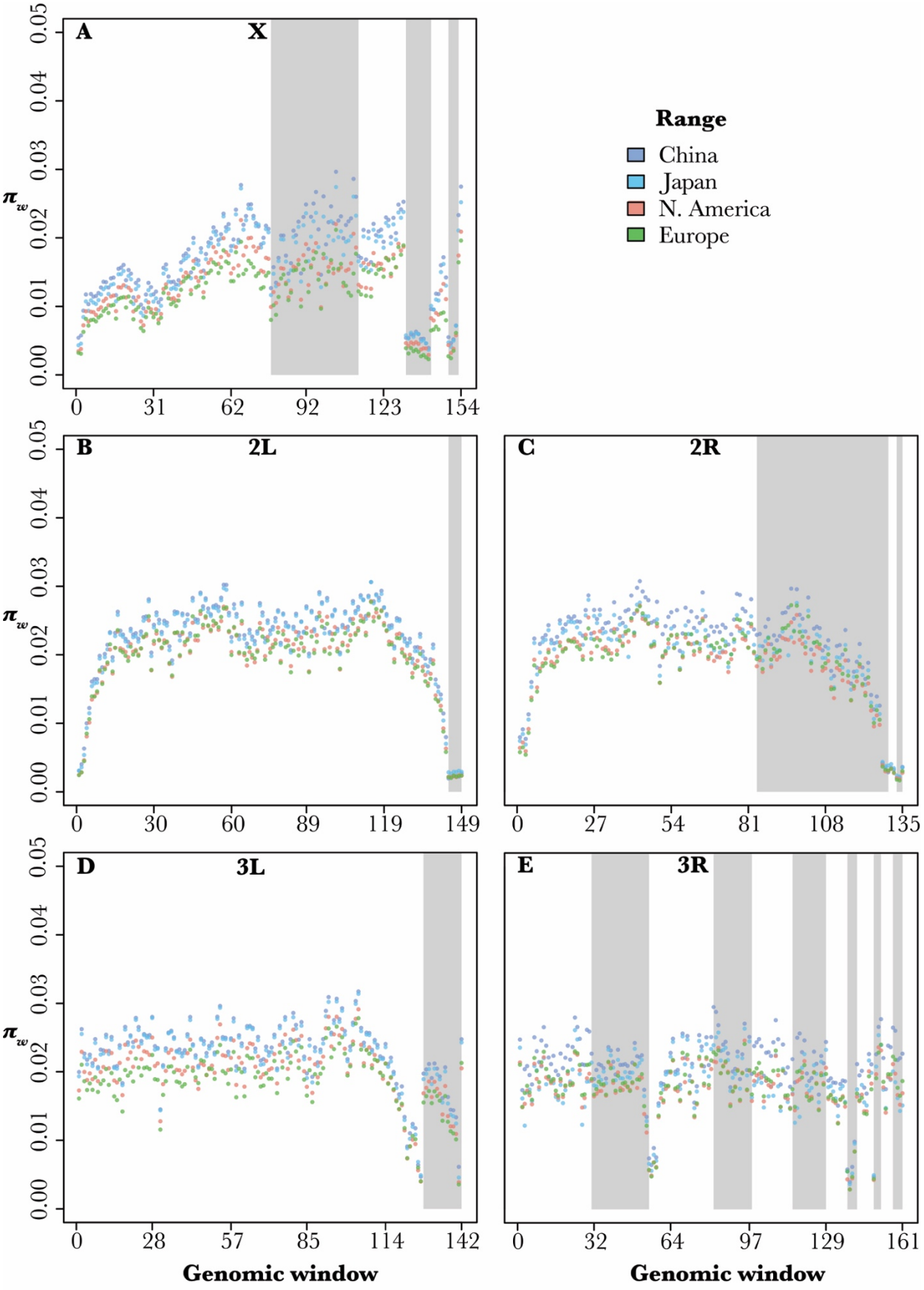


175 Figure 1. *Drosophila suzukii* populations show maximal diversity in eastern Asia and continent-level genetic structure.

176 (A) The geographic locations of the studied 29 natural populations are depicted as dots. In addition to the 22

177 populations sampled by Olazcuaga et al. (2020), populations newly sampled at independent locations are circled in
178 black. Populations newly sampled at nearby locations are circled and center-dashed in black, with the number of
179 total population samples in brackets. The year of the first recorded occurrence in each geographic range (colored
180 grey in the map) is given in brackets in the color legend. Further information about each sample is presented in
181 Table S1. (B) Population differentiation in allele frequencies (F_{ST}; lower triangle), between-population sequence
182 distances (D_{XY}; higher triangle), and within-population nucleotide diversity (π_S; diagonal) across autosomal
183 synonymous SNPs are displayed as a heatmap. Population names are colored by their geographic region. Asterisks
184 indicate samples contaminated by other *Drosophila* species, which may affect estimation of these statistics. (C)
185 Autosomal genetic structure is shown by three-dimensional principal components analysis (PCA) based on allele
186 frequencies of the two most frequent alleles across all populations. Each dot represents a population. Labeled are
187 Hawaii and western coastal US populations. See the X chromosome version of (B) and (C) in Figure S2.

188


A
X
B
2L
C
2R
D
3L
E
3R
Range
China
Japan
N. America
Europe
π_w
0.00
0.01
0.02
0.03
0.04
0.05
0
31
62
92
123
154
30
60
89
119
149
27
54
81
108
135
28
57
85
114
142
32
64
97
129
161
Genomic window
Genomic window

189

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

197

198 **Polymorphism and the Genomic Locations of *D. suzukii* Contigs**

199 If extensive genomic regions of low recombination are present in a genome-wide scan for
200 local adaptation, then because of the larger-scale influence of natural selection on linked sites in
201 such regions (Smith and Haigh 1974; Charlesworth et al. 1993), the precision of outlier
202 identification will be reduced (Lotterhos and Whitlock 2015; François et al. 2016). Although we
203 do not have recombination rate estimates for *D. suzukii*, we can begin to assess the genomic
204 abundance of low recombination regions through an examination of nucleotide diversity, in light
205 of its expected correlation with recombination rate (Begun and Aquadro 1992). Fortunately, we
206 observe that regions of low nucleotide diversity (which probably coincide with regions of low
207 recombination) cover a relatively small fraction of the genome (Figure 2). These patterns are
208 more similar to those in *D. simulans* than to *D. melanogaster* – which has broader centromeric
209 regions of low crossing over on the autosomes (Langley et al. 2012) – potentially suggesting a
210 relatively weaker suppression of crossing over in the centromere-proximal regions in *D. suzukii*.

211 Population genomic data may also provide clues regarding genome structure. The *D.*
212 *suzukii* reference assembly (Paris et al. 2020) is somewhat fragmented, with two or more contigs
213 assigned to each major chromosome arm, but with unknown orientation with respect to each

214 other. In other examined *D. melanogaster* group species, crossing-over rate and window nucleotide
215 diversity (π_w) are reduced in centromere- and telomere-proximal regions (*e.g.,* True et al. 1996).
216 When we examined the chromosomal distribution of π_w (Figure 2), we noticed that certain
217 arrangements of contigs would result in the expected patterns of reduced diversity at the ends of
218 each arm, and relatively smooth shifts in the diversity of large windows. For instance, the second
219 longest contig in chromosome arm 3L is likely to the left of the longest contig in a reverse
220 orientation, potentially with the smaller high diversity contig in between, (Figure 2D), although
221 either of the low diversity contig end could be centromeric or telomeric. Similarly in
222 chromosome arm 3R, although the assembly was more fragmented and the centromere- or
223 telomere-proximal regions consist of multiple contigs, the second and third longest contigs
224 appeared to contain the transitions between the high diversity mid-arm region and the
225 centromeric/telomeric zones (Figure 2E). Therefore, genome-wide polymorphisms could serve
226 as useful information to aid the ordering and orienting of contig-level genome assemblies
227 (Nakabayashi and Morishita 2020), although a combination of both systematic neighbor-
228 matching approaches and manual correction would be needed, and integration with other forms
229 of evidence (*e.g.,* comparative genomic or experimental data) would be preferable.

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

237 Methods). Based on this analysis, we assigned 170 contigs as autosomal or X-linked (Table S4).
238 Our classification of previously-assigned contigs was 96% consistent with past inferences, but we
239 corrected four previous assignments that were based on female-to-male read depth ratio, whereas
240 our method did not assign eight previously-assigned contigs.

241

242 **Selection of Robust Methodology and Distinct Environmental Variables for GEA**

243 The worldwide expansion of *D. suzukii* has exposed this species to selection pressures from
244 varying local environmental conditions (Olazcuaga et al. 2020; Mérel et al. 2021). To identify
245 environmental factors that have contributed to adaptive genetic differentiation at various levels
246 and loci under positive selection, we performed a whole-genome scan using GEA analysis
247 between environmental and genetic differentiation (see Materials and Methods). We focused on
248 the method BayeScEnv because it has been reported to produce fewer false positives in the
249 presence of hierarchical genetic structure such as the continental-scale patterns documented here
250 (de Villemereuil and Gaggiotti 2015). We excluded SNPs with a global average minor allele
251 frequency (MAF) below 5%, which should minimize the influence of the contamination
252 documented above. No alleles specific to contaminating species should meet that threshold, so
253 the only remaining effect should be a slight bias in allele frequency estimation toward ancestral
254 variants in contaminated samples at genuine *D. suzukii* SNPs. Given the low levels of
255 contamination present and the small number of samples affected, this effect may represent a
256 modest source of noise, but it should not lead to spurious adaptation signals.

257 Our selection of environmental variables for GEA started with a preliminary set of 26
258 candidate variables that are potentially relevant in the adaptation process of *D. suzukii* (Figure
259 3A), based on data availability and prior knowledge of *Drosophila* ecology (Kellermann et al. 2012;

260 Hamby et al. 2016; Bogaerts-Márquez et al. 2021). Among common strategies of performing
261 GEA with a large set of relevant environmental variables, univariate association with all
262 environmental variables could increase the number of statistical tests, thus increasing the
263 difficulty of controlling rates of false discovery. On the other hand, including multiple variables
264 might cause a multicollinearity issue (Rellstab et al. 2015). We opted to retain nine of the least
265 correlated environmental variables for univariate tests, representing altitude, wind speed, as well
266 as multiple aspects of temperature, precipitation, and human land usage for GEA analysis (Figure
267 3B). Although the temperature of the coldest quarter had a significant negative correlation with
268 wind speed, we kept both variables for GEA, as cold stress and wind-related factors are known to
269 be potential drivers of local adaptation in *Drosophila* (Bogaerts-Márquez et al. 2021). As indicated
270 by their coefficients of variation (*CVs*), the selected environmental variables had moderate
271 ($15\% \leq CV < 30\%$) to high ($CV \geq 30\%$) variability across our sampling locations (Table S4).

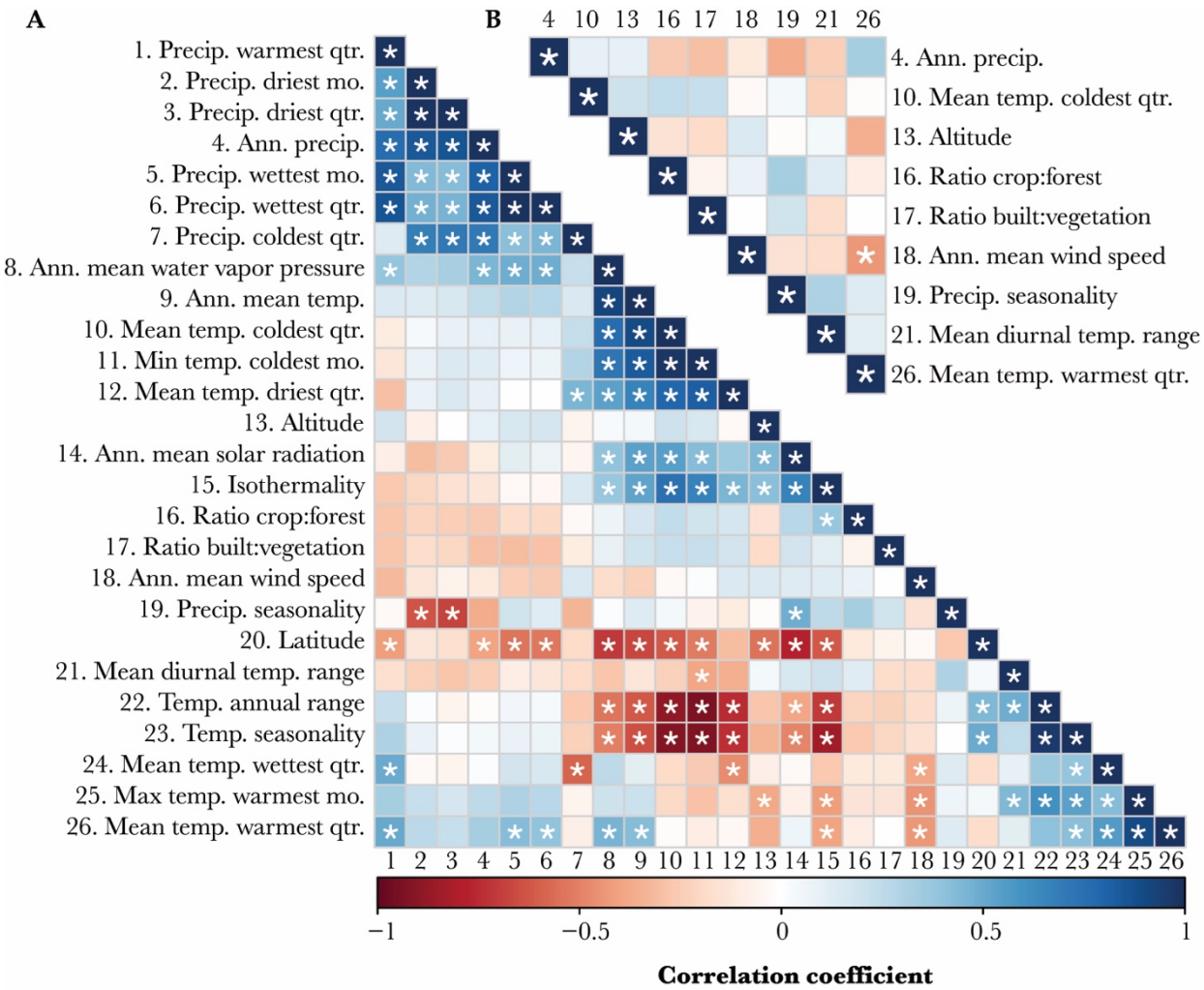


Figure 3. Identification of least-correlated environmental variables for genotype-environment association (GEA) analysis in *D. suzukii*. (A) Pairwise correlations among a preliminary set of 26 environmental variables that are potentially impactful on *D. suzukii*. (B) A final set of nine of the most 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 S4 for environmental values used to calculate correlation coefficients.

Widespread Signals of Recent Adaptation to Diverse Environments

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

282 (genome-wide $q < 0.05$) associated with each of the nine candidate environmental variables
283 (Table S6). These variants corresponded to an average of 3,345 overlapping or neighboring
284 genes (Table S6), suggesting that selection pressures from the tested environmental variables (or
285 correlated factors) have meaningfully contributed to adaptive genetic variation in *D. suzukii*, even
286 on the brief time scale of its worldwide expansion. Among all tested environmental factors, mean
287 temperature of the coldest quarter was associated with the greatest number of putatively adaptive
288 variants (4,250 SNPs). Two precipitation-related variables, annual precipitation (4,141 SNPs)
289 and precipitation seasonality (3,389 SNPs), have the next largest loci count. The ratio of built
290 area to vegetation (i.e., crops and forests) and the ratio of crops to forests were associated with the
291 fewest genetic variants (2,369 and 1,608 SNPs, respectively).

292 To reveal the potential genetic and functional basis of invasion success under multiple
293 environmental challenges throughout the species' range, we first examined the functions of genes
294 linked to the top 10 environment-associated loci for each variable as ranked by association q-
295 value, and then by the g parameter estimating the sensitivity to environmental differentiation as
296 a tie-breaker. We found many of these genes have known functions that could facilitate
297 adaptation to the associated environmental factor (Table S7).

298 Among genes linked to altitude-associated loci, the second-ranked candidate *ab* is known
299 to control wing size in *Drosophila* (Simoes da Silva et al. 2019). Interestingly, wing size was found
300 to have increased in a highland Ethiopia *D. melanogaster* population, potentially assisting flight in
301 thin, cool air (Lack et al. 2016b). Ranked next to *ab* is *Gbs-70E*, which plays roles in glycogen
302 metabolism and the development of eggs inside the maternal ovary (Kerekes et al. 2014).
303 Another top gene, the lysine demethylase *Kdm2*, is upregulated in response to hypoxia (Batie et al.
304 2017).

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

311 There is also some evidence for precipitation-related local adaptation. The top gene *mmy*
312 associated with precipitation seasonality (i.e., annual range of precipitation) was shown to
313 regulate chitin synthesis and cuticle production. Since precipitation is correlated with desiccation
314 resistance across the *Drosophila* phylogeny (Kellermann et al. 2012), *D. suzukii* may have
315 developed adaptive strategies of modifying chitin biosynthesis under conditions of desiccation
316 (Rezende et al. 2008; Clark et al. 2009), which was also implied in seasonal plasticity of natural
317 *Drosophila* populations (Shearer et al. 2016; Horváth et al. 2023). In addition, the gene *osy*
318 (*CG33970*) contributes to the formation of the outer cuticle layer and is expressed more highly in
319 *D. suzukii* than in *D. melanogaster* (Wang et al. 2020). Furthermore, two of the top genes associated
320 with annual precipitation (*Abd-B* and *bab1*) regulate cuticle pigmentation (Rogers et al. 2013),
321 which may or may not correlate with desiccation tolerance in *Drosophila* species (Wang et al.
322 2021). We also note that although environmental fitness effects on these testes-expressed genes
323 are not known, the same SNP near *CG17944* and *nxf4* was among the highest-scoring variants for
324 both annual precipitation and precipitation seasonality (variables that have a non-significantly
325 negative correlation between them; Figure 3).

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

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

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

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

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

354

355 **Functional Commonalities in the Adaptation to Diverse Selective Pressures**

356 Next, we examined environment-specific adaptation on a more comprehensive basis
357 through a gene ontology (GO) enrichment analysis of the top 500 genes associated with each
358 environmental variable (Figure 4). As correlates of temperature in the coldest quarter, cAMP
359 metabolic process was the top enriched category, followed by two other related purine
360 metabolism groupings. We note that cAMP is important in circadian regulation (*e.g.*, Palacios-
361 Muñoz & Ewer, 2018), which is known to play an important role in *Drosophila* environmental
362 adaptation (*e.g.,* Helfrich-Förster et al. 2020), as also implicated by the presence of "entrainment
363 of circadian clock" on our top GO category list for altitude. More broadly, purine metabolism
364 was inferred as a strategy of cold acclimation in *D. suzukii* (Enriquez and Colinet 2019). For
365 diurnal temperature range, the top category was "regulation of growth", and we note that some
366 drosophilids have evolved to have larger body sizes in more challenging thermal environments
367 (Gilchrist and Partridge 1999; Calboli et al. 2003; Lack, Yassin, et al. 2016).

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

GO classes

- Biological process
- Molecular function
- Cellular component

	GO description	p-value	# genes
Altitude	nucleus localization	1.35e-03	7
	R7 cell fate commitment	1.89e-03	7
	R7 cell differentiation	2.44e-03	12
	sensory organ precursor cell fate determination	2.51e-03	6
	synaptic vesicle	2.67e-03	8
	intracellular protein-containing complex	2.76e-03	24
	endomembrane system	4.40e-03	53
	entrainment of circadian clock	4.99e-03	5
	translation initiation factor activity	5.85e-03	5
	centrosome duplication	6.85e-03	5
Ann. precip.	chitin metabolic process	9.70e-04	5
	transferase complex	1.84e-03	24
	I band	3.19e-03	8
	nucleoplasm	3.36e-03	24
	cellular protein-containing complex assembly	3.39e-03	23
	single fertilization	3.63e-03	5
	supramolecular fiber organization	3.84e-03	24
	cytoplasm	4.18e-03	145
	hydrolase activity, acting on glycosyl bonds	4.24e-03	8
	actin cytoskeleton organization	4.32e-03	27
Ann. mean wind speed	centrosome	2.40e-03	11
	ligand-gated sodium channel activity	2.61e-03	5
	cellular protein catabolic process	5.34e-03	19
	spindle	6.26e-03	9
	triglyceride lipase activity	7.07e-03	6
	nitrogen compound metabolic process	9.00e-03	167
	primary metabolic process	1.08e-02	175
	protein ubiquitination	1.19e-02	16
	organic substance metabolic process	1.26e-02	182
	actin cytoskeleton reorganization	1.39e-02	5
Mean temp. coldest qtr.	cAMP metabolic process	5.00e-04	6
	purine nucleotide metabolic process	1.35e-03	16
	oxidoreductase activity, acting on the aldehyde or oxo group of donors	6.18e-03	6
	sevenless signaling pathway	9.82e-03	5
	positive regulation of JNK cascade	1.03e-02	6
	positive regulation of cell projection organization	1.44e-02	8
	identical protein binding	1.54e-02	15
	cation transport	1.71e-02	32
	organic hydroxy compound metabolic process	2.35e-02	12
	locomotion	3.01e-02	60
Ratio crop to forest	cell-substrate junction	3.90e-04	9
	cell surface	1.15e-03	14
	extrinsic component of cytoplasmic side of plasma membrane	2.46e-03	5
	signaling receptor binding	5.99e-03	24
	positive regulation of filopodium assembly	7.48e-03	5
	cell-matrix adhesion	9.53e-03	5
	antimicrobial humoral response	1.05e-02	10
	cortical actin cytoskeleton organization	1.16e-02	6
	germ-band shortening	1.67e-02	5
	cell fate determination	1.85e-02	16
Precip. Seasonality	response to sucrose	2.45e-03	5
	signaling receptor activity	5.59e-03	48
	detection of chemical stimulus involved in sensory perception	8.82e-03	12
	exopeptidase activity	9.20e-03	9
	chitin binding	1.45e-02	8
	positive regulation of response to external stimulus	1.65e-02	11
	hindgut development	1.68e-02	10
	gastrulation involving germ band extension	1.81e-02	6
	positive regulation of response to biotic stimulus	2.01e-02	10
	histone binding	2.20e-02	6
Mean diurnal temp. range	regulation of growth	1.70e-04	33
	regulation of cellular component organization	1.74e-03	49
	cell growth	2.39e-03	18
	R3/R4 cell differentiation	2.52e-03	7
	dorsal/ventral pattern formation, imaginal disc	3.20e-03	11
	embryo development	3.22e-03	48
	post-embryonic animal organ development	3.92e-03	46
	regulation of developmental process	4.26e-03	56
	compound eye morphogenesis	4.81e-03	33
	generation of neurons	5.56e-03	79
Mean temp. warmest qtr.	extracellular matrix structural constituent	1.00e-03	7
	spindle pole	1.19e-03	6
	regulation of neurogenesis	1.64e-03	18
	centrosome	2.73e-03	11
	regulation of dendrite development	3.16e-03	9
	regulation of neuron projection development	3.26e-03	14
	cellular macromolecule biosynthetic process	3.73e-03	95
	sex differentiation	4.97e-03	17
	pattern specification process	6.56e-03	47
	brain development	7.22e-03	17
Ratio built to vegetation	appendage morphogenesis	3.88e-03	38
	adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	5.49e-03	6
	positive regulation of synaptic transmission	5.64e-03	5
	carbohydrate biosynthetic process	5.93e-03	6
	anterior Malpighian tubule development	5.97e-03	5
	wing disc morphogenesis	6.41e-03	32
	cleavage furrow	8.33e-03	5
	neuroblast differentiation	9.56e-03	11
	late endosome	1.02e-02	8
	carbohydrate derivative biosynthetic process	1.20e-02	17

Figure 4. 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

377 each panel (labelled on the left), with permutation p-values and the number of associated genes in each GO
378 category. Descriptions of GO categories are colored by their GO class (see legend at top right). Only GO
379 categories including more than five associated genes are listed here. For a full list of enriched GO categories, see
380 Table S8.

381

382 As broader evidence for a shared (or biologically similar) underlying genetic basis of
383 adaptation to multiple environmental factors, we examined the overlap of the most significant
384 genes and most enriched GO categories between different environmental variables. We found
385 the top candidate genes to be mostly associated with both temperature and precipitation. The
386 gene sets showed relatively greater overlap among climatic variables (including altitude), whereas
387 the two land usage variables had less overlap with climatic variables or with each other (Figure
388 5A). Since the patterns of shared genes cannot be fully explained by correlations between
389 environmental values (Figure 3B), at least some of the genes may have been responding to
390 multiple selective pressures. The overall proportions of shared GO categories were lower than
391 those of shared genes, indicating that the shared genes do not necessarily lead shared functional
392 categories between environmental variables. Relatively higher GO term sharing was observed
393 between altitude and either wind speed or precipitation, and between diurnal temperature range
394 and temperature of the warmest quarter (Figure 5B). Based on the shared genes and GO terms
395 observed, it is possible that during the rapid range expansion of *D. suzukii*, pleiotropy may have
396 facilitated local adaptation to multiple selective pressures (Hämälä et al. 2020; Kinsler et al.
397 2020).

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

401 temperature range, and temperature of the warmest quarter), "regulation of neurogenesis"
402 (altitude, diurnal temperature range, temperature of the warmest quarter, and ratio of built area
403 to vegetation), and "central nervous system development" (precipitation seasonality, diurnal
404 temperature range, and temperature of the warmest quarter) (Figure 5C). Further, as mentioned
405 above, "cAMP metabolic process" (shared by temperature of the coldest quarter, precipitation
406 seasonality and ratio of built area to vegetation) could entail neurologically modulated changes in
407 circadian behavior. Each of these traits was associated with at least three environmental
408 variables, which suggests a multifaceted role for nervous system evolution in facilitating the
409 invasion success of *D. suzukii* under multiple environmental challenges.

410

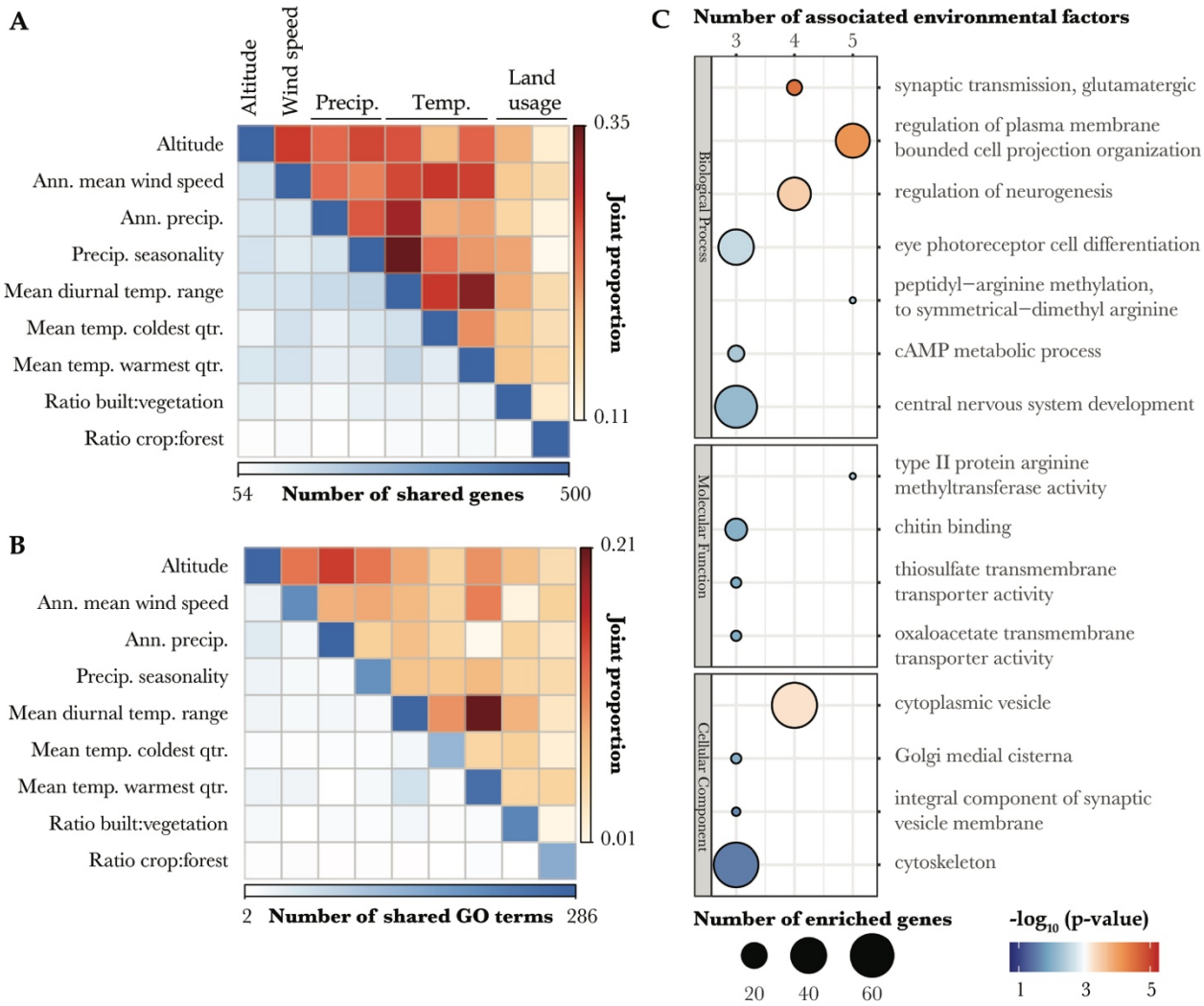


Figure 5. Overlapping genes and GO categories among environmental factors reveal the shared genetic and functional basis of environmental adaptation in *D. suzukii*. The numbers and proportions of shared (A) environment-associated genes and (B) enriched GO categories among environmental factors are shown 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.

421 **Discussion**

422 We performed population genomic analyses of 29 population samples of *D. suzukii* to
423 investigate the genomic diversity and environmental adaptation of this highly invasive species
424 across its worldwide distribution. Our data supported a genetic grouping of these populations
425 into four primary geographic regions: eastern Asia (containing the native range), Hawaii, the
426 Americas, and Europe. We also confirmed that all non-Asian populations have reduced
427 diversity, consistent with moderate founder event bottlenecks in introduced populations.

428 Our analyses also added to our knowledge of the *D. suzukii* genome and its evolution. We
429 used population genomic data to improve the classification of X-linked and autosomal contigs.
430 We determined that relatively few contigs showed strongly reduced nucleotide diversity, implying
431 that only a small fraction of the genome experiences minimal crossing-over. And we
432 documented the influence of an expanded repeatome on non-coding divergence.

433 The above analyses placed us in a more confident position to perform a robust analysis of
434 genotype-environment association (GEA). We selected nine distinct environmental variables,
435 including altitude, wind speed, precipitation, temperature, and human land usage. Our results
436 suggested extensive local adaptation in response to specific environmental challenges, along with
437 appreciable sharing of genes and functional pathways underlying invasion success across multiple
438 environmental pressures, which were most obvious with nervous system genes.

439

440 **Environmental Drivers of Adaptation in *D. suzukii***

441 Here, we presented a GEA analysis that investigated the most geographically and
442 genetically diverse set of *D. suzukii* populations and the most comprehensive set of environmental

443 factors to date, which enabled unprecedented power to capture even minor adaptive genetic
444 differentiation in response to distinct environmental challenges during the species' rapid
445 invasions. In addition to our identification of climatic factors including temperature,
446 precipitation-related variables and wind speed as the most frequently correlated with putatively
447 local adaptative variants (consistent with previous GEA analysis in *D. melanogaster*, *e.g.,* Bogaerts-
448 Márquez et al. 2021), we also for the first time identified large numbers of genome-wide variants
449 associated with altitude and human land usage-related variables (Table S6), which were not
450 included in most GEA studies despite their potential significance to local adaptation. In
451 particular, the detected associations with ratios of developed land to vegetation and of cropland
452 to forests highlights the ecological impacts of urbanization and agriculture on natural populations
453 of insects. Since the prior selection of environmental variables is critical for successful GEA
454 analyses, we also provided an instructive example for correlation-based selection to identify the
455 most relevant and least redundant environmental factors (Rellstab et al. 2015).

456

457 **Nervous System Evolution is Ubiquitous in Environmental Adaptation of**
458 ***Drosophila***

459 In *D. suzukii*, we found nervous system and related sensory and behavior annotations
460 associated with top genes for all nine environmental variables studied. Concordantly, we found
461 that GO categories related to the nervous system were among the most shared across
462 environmental variables (Figure 5C). In *D. melanogaster*, related GO categories like 'neuron
463 development', 'nervous system development' and 'eye development' were also enriched among
464 genes associated with environmental variation among natural populations within North America
465 or Europe, and across seasons within Europe (Bogaerts-Márquez et al. 2021). GO categories

466 associated with the nervous system have also shown evidence of positive selection in various
467 genome scans of *D. melanogaster* (Langley et al. 2012; Pool et al. 2012; Pool 2015), including a
468 study of parallel evolution in cold-adapted populations (Pool et al. 2017). Given the
469 morphological evidence of neuron-muscular junction evolution across the entire *Drosophila*
470 phylogeny (Campbell and Ganetzky 2012), we therefore propose a broad adaptive importance of
471 the nervous system in *Drosophila* species and potentially other insects. Such evolutionary
472 processes may have either maintained ancestral neural functions in novel challenging
473 environments, or created novel phenotypes that better fit the new optima arising from complex
474 combinations of environmental factors.

475

476 **Cautions with Interpretations of Association Results and Future Directions**

477 While we have generated intriguing hypotheses about gene functions that may underlie
478 the environmental adaptation of *D. suzukii*, it is difficult to distinguish between correlated
479 environmental selective pressures that may have driven the detected associations, including not
480 only the 17 environmental factors that were excluded in the process of variable reduction, but
481 also correlated biotic or abiotic factors not represented in global databases. As an intrinsic
482 limitation of GEA analysis that cannot be accounted for by applying stricter thresholds,
483 associations observed with a particular environmental factor might stem from adaptation to other
484 covarying factors (Rellstab et al. 2015). For example, the two tracheal branching genes *crp* and
485 *Mrtf* (Han et al. 2004; Wong et al. 2015) associated with mean temperature of the warmest
486 quarter could represent adaptations to reduce water loss under conditions of elevated water
487 vapor pressure (Telonis-Scott et al. 2012), which is closely related to humidity and has a
488 significant positive correlation with mean temperature of the warmest quarter (Figure 3B).

489 Therefore, expanded characterization of the relationships between genotype, phenotype,
490 and fitness in this species is needed to further clarify the functional and phenotypic
491 interpretations associated with certain environmental factors and genes. Experimental
492 validations that leverage RNA interference (Boutros and Ahringer 2008) and/or transgenic
493 overexpression (Prelich 2012) to modify the expression of associated genes, and/or genome
494 editing techniques (Stern 2014; Turner 2014; Shalem et al. 2015) to target putatively adaptive
495 variants would also bring a more solid understanding about the invasive biology of this species in
496 distinct environments. Such functional studies could be complemented by population
497 experiments under controlled laboratory environments or field conditions (*e.g.,* Behrman et al.
498 2015; Rudman et al. 2022), in order to more clearly demonstrate the connections between
499 specific selective pressures and alleles or traits of interest.

500

501 **Broader Impacts and Significance**

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

511

512 **Materials and Methods**

513

514 **Fly Collection, DNA Preparation, and Pooled Sequencing**

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

524 Pool-sequenced reads of 22 additional *D. suzukii* population samples, including from
525 Europe, the Americas, and Asia, were obtained from public data provided by Olazcuaga et al.
526 (2020) at EBI's SRA (Figure 1; Table S1). Taken together, we formed a comprehensive dataset
527 of 29 populations sampled from native and invasive ranges of *D. suzukii*.

528

529 **Quality Control, Alignment, Contamination Analysis, and Variant Calling from**
530 **Pool-seq Data**

531 To maximize the quality of our analyzed data, we built a high-throughput assembly and
532 quality control pipeline *poolWGS2SNP* with optimized performance, stringent filtering,
533 compatibility with large numbers of genomic contigs, and customized functions to call high-

534 confidence single-nucleotide variants from pool-sequenced data in *D. suzukii* (Figure S6), in part
535 by utilizing resources from the DrosEU bioinformatics pipeline (Kapun et al. 2020).

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

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

551 We then checked sample contamination for both newly and previously reported pool-seq
552 data (Table S3), by estimating the proportion of pool-seq reads from different *Drosophila* species as
553 the proportion of reads assigned to species-discriminating *k-mers* (i.e., unique sequences of each
554 species' reference genome assembly) using the approached described by Gautier (2023). Although
555 the estimation should be reliable, to fully eliminate contaminated reads is not currently practical
556 because substantial proportion of reads can not be confidently assigned to any species (up to 42%

557 of total pool-seq reads), due to the sequence similarity among genomes within the *Drosophila*
558 family (Gautier 2023). Therefore, we focus on identifying samples with contamination and
559 interpreting results these samples with caution.

560 To call SNPs, we merged the quality-checked BAM files of all population samples into
561 one file using Samtools *mpileup*, only retaining alignments with mapping quality no less than 20
562 and sites with base quality no less than 20. Variant calling was then performed on the mpileup
563 file using the heuristic SNP caller *PoolSNP* (Kapun et al. 2020). We used a nominally low value
564 for the parameter miss-frac (0.001) to require for each population sample individually, that depth
565 of coverage at a given site be 12 or greater (min-cov = 12), and that this site not be in the top 1%
566 of sites genome-wide for depth of coverage (max-cov = 0.99; calculated separately for each
567 population and for autosomal and X-linked contigs), in order to filter sites subject to copy
568 number variation. In the initial data set used for analysis of genome-wide diversity, we avoided
569 potential biases from allele frequency filters by using min-count = 1 and min-freq = 0. We
570 termed the resulting high-quality sites as 'analyzed sites' for brevity.

571

572 **Identifying Autosomal and X-linked Contigs**

573 We chose to perform all population genomic analyses and whole-genome scans separately
574 for SNPs from autosomes and the X chromosome for the following reasons: 1) autosomal and X-
575 linked variants have different allelic sample sizes as samples were obtained from both male and
576 female flies; 2) autosomes and the X chromosome could reflect different demographic histories
577 and outcomes of natural selection, *e.g.*, the lower effective population size of the X chromosome
578 than autosomes could lead to a higher impact of bottlenecks and selection on genomic diversity;

579 3) unbalanced sex ratios and male-biased dispersal could further differentiate autosomal versus X
580 chromosome variation (Clemente et al. 2018; Olazcuaga et al. 2020).

581 Since the assembly of *D. suzukii* reference genome is still at the contig level, chromosomal
582 identities of each contig are needed to perform separate analyses. However, 497 contigs that
583 represent ~43% of the assembly length have not been unambiguously mapped onto chromosome
584 arms of the *D. melanogaster* dm6 genome assembly. Although 264 of the unplaced contigs had
585 been assigned to autosomes and the X chromosome based on a female-to-male read depth ratio,
586 233 contigs that represent ~5% of the genome remained unassigned due to the lack of statistical
587 power (Paris et al. 2020).

588 Given our interest in accurately analyzing a larger proportion of the euchromatic
589 genome, we identified ~70% of these 233 unassigned contigs as autosomal and X-linked based
590 on the correlation between the mean read depth of each contig (among population samples) and
591 that across unambiguously aligned autosomal or X-linked contigs. We chose Spearman's rank
592 correlation instead of the Pearson correlation, as the distribution of depth data failed the
593 assumption of bivariate normality. A contig that has mean depth significantly correlated with
594 that of either known autosomal or X-linked contigs was assigned to the chromosome with a
595 higher correlation coefficient. Our method completely confirmed all prior mapping-based
596 assignments and had a ~96% consistency with the previous assignment based on female-to-male
597 read depth ratios. Inconsistent assignments for four contigs were corrected according to our
598 method (Table S4). The eight previously-assigned contigs that could not be assigned using our
599 method, as well as other unassigned contigs using all methods (totaling ~2.7 mb), were excluded
600 from downstream GEA analyses, because the assignment information is needed for estimating
601 effective sample size that were used to correct allele count data as an input to GEA.

602

603 **Annotating Genomic Features and Estimating Divergence**

604 To explore genomic diversity at synonymous sites and selective constraint for other site
605 types in *D. suzukii*, we classified the reference genome into nine exclusive categories of site
606 degeneracy and function (Lange and Pool 2018), including: non-degenerate (i.e.,
607 nonsynonymous) sites, two-, three-, and four-fold degenerate (i.e., synonymous) sites, 3' and 5'
608 UTRs, RNA-coding genes, introns, and intergenic regions. From input files including the
609 eukaryotic codon table, the published genome sequence and GFF3 annotation obtained at NCBI
610 RefSeq, we generated a letter-coded annotation (in FASTA format) mirroring both strands of the
611 whole genome sequence of *D. suzukii* and a coordinate-based annotation (in BED format) that
612 combines adjacent sites of the same category into a single row. Degeneracy was determined
613 based on the standard codon table. 5' UTRs were defined as regions between the start of the
614 first exon and the start of the first coding sequence (CDS), while 3' UTRs were defined as regions
615 between the end of the last exon and the end of the last CDS. In cases of overlapping genes and
616 alternative splicing that raise annotation conflicts, we followed an annotation priority in the
617 category order listed above.

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

623

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

625 To compare genome-wide polymorphism among populations, we estimated nucleotide
626 diversity (π) across SNPs at four-fold degenerate sites (π_S) in addition to that at all categories of
627 sites (π_A), as π_s estimation is relatively less affected by sequencing errors than nucleotide diversity
628 estimated from other site categories (due to a higher ratio of real variation to errors). To
629 calculate π for each population sample, we adopted an unbiased estimator of nucleotide diversity
630 ($\hat{\theta}_{\Pi}$) based on heterozygosity (Π), which has been optimized for pool-seq data (Ferretti et al.
631 2013). Numerically,

632 $$\pi_S = \hat{\theta}_{\Pi} = \frac{n_c}{n_c-1}\frac{\Pi}{L} = \frac{n_c}{n_c-1}\frac{1}{L}\sum_l \frac{2}{n_r(l)(n_r(l)-1)} m_l(n_r(l) - m_l) \quad (1)$$

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

642 To examine patterns of polymorphism across chromosome arms, we also estimated
643 window nucleotide diversity (π_W) for all polymorphic sites. Each window was defined as a
644 continuous genomic region that includes 125,000 analyzed sites (Figure 2). Since chromosomal
645 identity was required in this analysis, we only took windows from 32 major contigs that contain

646 at least one full-size window and were unambiguously mappable to a chromosome arm of the *D.*
647 *melanogaster* dm6 genome assembly. Although such contigs only make up 57% of the *D. suzukii*
648 genome assembly, they contain a relatively larger proportion of all identified SNPs (72%), and
649 thus are still representative of genome-wide polymorphism.

650 To estimate genome-wide pairwise F_{ST} between populations, we adopted an unbiased
651 multi-loci estimator known as Reynolds' estimator of the coancestry coefficient, which accounts
652 for unequal sample sizes among populations and is applicable for more than two alleles at a site
653 (Reynolds et al. 1983). We first heuristically partitioned the genome into windows that exceeded
654 a cross-sample average accumulated heterozygosity threshold of 100. The genome-wide F_{ST} was
655 then calculated as an average of window-F_{ST} $F_{ST}(w)$ weighted by the number of analyzed sites
656 within each window ($L(w)$), where $F_{ST}(w)$ was calculated as a weighted average of single-site
657 ratio estimators. Numerically,

658 $$F_{ST} = \frac{1}{L}\sum_{w} L(w)\, F_{ST}(w) \quad (2)$$

659 $$F_{ST}(w) = \left(\sum_{l} a_l\right) / \sum_{l}(a_l + b_l) \quad (3)$$

660 where

661 $$a_l = \frac{1}{2}\sum_{u} \left(\tilde{p}_{1lu} - \tilde{p}_{2lu}\right)^2 - \frac{(n_{1l}+n_{2l})(n_{1l}\tilde{\alpha}_{1l}+n_{2l}\tilde{\alpha}_{2l})}{4n_{1l}n_{2l}(n_{1l}+n_{2l}-1)} \quad (4)$$

662 $$a_l + b_l = \frac{1}{2}\sum_{u} \left(\tilde{p}_{1lu} - \tilde{p}_{2lu}\right)^2 + \frac{(4n_{1l}n_{2l}-n_{1l}-n_{2l})(n_{1l}\tilde{\alpha}_{1l}+n_{2l}\tilde{\alpha}_{2l})}{4n_{1l}n_{2l}(n_{1l}+n_{2l}-1)} \quad (5)$$

663 At the lth site in each population, $\tilde{p}_{1lu}$ and $\tilde{p}_{2lu}$ represent the frequency of the uth allele at the
664 lth site; $\tilde{\alpha}_{1l}$ and $\tilde{\alpha}_{2l}$ represent the heterozygosity; n_{1l} and n_{2l} represent the sample size. Unlike
665 the sequencing of individual genomes, pool-seq induces an uncertainty in the number of
666 individual alleles actually sequenced at a locus (i.e., effective sample size), and this uncertainty

667 decreases slowly even at high read depth (Ferretti et al. 2013). Since the sample size is an
668 important parameter for F_{ST} estimation, we took measures to obtain an estimate of the effective
669 sample size, n_{il}, at each given site (Ferretti et al. 2013). Numerically,

670 $$n_{il} = \sum_{j}^{n_c} j P_c(j \mid n_r, n_c) \quad (6)$$

671 where

672 $$P_c(j \mid n_r, n_c) = \frac{n_c! S(n_r, j)}{(n_c - j)! n_c^{n_r}} \quad (7)$$

673 Here, we explicitly estimated the probability of the number of j unique lineages sampled at a site
674 given n_r sampled reads and n_c equally contributing chromosomes in a pool, where $S(n_r, j)$ are
675 the Stirling numbers of the second kind, defined as the number of ways to partition n_r reads into
676 j non-empty sets (Ferretti et al. 2013). We then estimated n_{il} as the expected number of lineages
677 for each n_r and n_c. Ideally, n_c should be estimated as $2n_e$, where n_e is the effective pool size
678 representing the number of diploid individuals contributing the same amount of reads to a pool
679 (Gautier et al. 2013; Lange et al. 2022). Although we lack sample replicates to estimate n_e and
680 therefore used haploid sample size for n_c as an approximation, the probability estimation is still
681 reasonable given that our number of lineages for each pool is large (Table S1) (Ferretti et al.
682 2013).

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

687 $$D_{XY} = \frac{1}{L} \sum_{l} \sum_{ij} x_i y_j k_{ij} \quad (8)$$

688 where x_i and y_j represent frequencies of the *i*th allele from population *X* and the *j*th allele
689 from population *Y*, and k_{ij} is either 1 or 0, depending on whether or not the alleles differ at the
690 lth site.

691 Calculations in this section were all implemented with Python and Shell scripts (see Data
692 Availability).

693

694 **Preparing Environmental Data**

695 To generate environmental data for GEA, we selected a preliminary set of 26 candidate
696 environmental variables representing geographic, climatic and land cover-related factors (Figure
697 3A) that may be relevant in the adaptation process of *D. suzukii* based on prior knowledge
698 (Kellermann et al. 2012; Bogaerts-Márquez et al. 2021). With R packages 'raster' (v. 3.5.2) and
699 'SpatialPoints' (v. 1.4.6), we retrieved environmental data of high spatial resolution (~100 km^2)
700 in batch for the sampling locations of our 29 populations from online databases WorldClim (Fick
701 and Hijmans 2017) and Esri 2020 Land Cover (Karra et al. 2021). Annual mean values of
702 monthly climatic variables, including mean wind speed, solar radiation, and water vapor
703 pressure, were derived by averaging across 12 months of data.

704 Due to the large number of statistical tests that would result from running GEA on all the
705 environmental variables one by one, there is an increased difficulty in controlling rates of false
706 discovery. Additionally, including multiple highly correlated variables in a model would lead to
707 multicollinearity issues (Rellstab et al. 2015). To avoid these problems, we calculated a pairwise
708 Pearson correlation matrix from values of environmental factors across sampled locations (Figure
709 3A; Table S5), and then selected a subset of nine least correlated environmental variables for

710 one-by-one GEA analyses (Figure 3B). To avoid scale inconsistencies between estimated GEA
711 statistics, the environmental differentiation of each population was calculated as the absolute
712 difference between the environmental value of that population and the average across all
713 populations, standardized by the standard deviation (de Villemereuil and Gaggiotti 2015). This
714 standardized differentiation was then input to GEA (Table S5).

715

716 **Environmental Association Analyses**

717 To characterize the environmental adaptation of *D. suzukii*, we scanned the whole
718 genome for adaptive loci using the F_{ST}-based GEA method BayeScEnv (de Villemereuil and
719 Gaggiotti 2015). We chose this specific approach over other GEA methods mainly because it has
720 a lower false positive rate than other GEA approaches in the presence of hierarchical population
721 structure (de Villemereuil et al. 2014; de Villemereuil and Gaggiotti 2015). It also allows for
722 detecting patterns of allele frequency that are not linearly dependent on environmental factors
723 (Rellstab et al. 2015; de Villemereuil and Gaggiotti 2015).

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

732 To make this GEA analysis computationally feasible with our large SNP set, while still
733 analyzing all qualifying SNPs, we applied a split-run strategy: we subsampled SNPs across
734 concatenated sequences of contigs within the autosomes and the X chromosome separately, and
735 then ran subsamples with BayeScEnv in parallel. Since the null model of population structure is
736 estimated separately in each run, we subsampled non-adjacent SNPs at a fixed interval to limit
737 locus-specific biases in that estimation, where the length of the interval between jointly analyzed
738 SNPs was equal to the total number of subsamples. With a targeted subsample/interval size of
739 up to 10,000 SNPs, we divided the concatenated autosomal contigs into 490 subsamples (with
740 actual subsample sizes of 9982-9983 SNPs), and the concatenated X-linked contigs into 87
741 subsamples (with actual subsample sizes of 9893-9894 SNPs). Hence, the first autosomal
742 subsample contained SNP #1, SNP #491, and so on.

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

753

754 **Identifying Candidate Genes**

755 For each candidate SNP, the closest gene in each direction within a 200-exon flanking
756 region that overlapped with it was considered to be associated with that variant, in order to
757 encompass both potential coding and regulatory adaptation. To facilitate clear comparisons
758 among environmental variables with different numbers of significant variants, we focused on the
759 top 500 candidate genes that were linked to variants with the lowest significant q and highest g
760 within each environmental variable (Table S7).

761

762 **GO Enrichment and Semantic Clustering**

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

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

775

776 **Author contributions**

SF, JEP, SDS, and CG designed the research. SPD collected data for the project. SF analyzed the data. SF, JEP, and SDS wrote the paper. SDS, JEP, and CG obtained funding for the research.

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.

Funding

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).

Conflict of interest

All authors declare that they have no conflicts of interest.

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 (for WGS data processing and variant calling) and https://github.com/Sfeng666/Dsuz_popgen_GEA (for population genetics analyses and GEA).

Supplementary Material

Supplementary figures S1–S6 and tables S1–S7 are available at Evolution online.

808 **References**

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

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

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

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

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

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

824 Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, Zuker CS.
825 2004. Decoding Cilia Function: Defining specialized genes required for
826 compartmentalized cilia biogenesis. *Cell* 117:527–539.

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

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

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

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

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

840 Boughdad A, Haddi K, El Bouazzati A, Nassiri A, Tahiri A, El Anbri C, Eddaya T, Zaid A,
841 Biondi A. 2021. First record of the invasive spotted wing *Drosophila* infesting berry crops in
842 Africa. *J Pest Sci* 94:261–271.

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

845 Calboli FCF, Gilchrist GW, Partridge L. 2003. Different cell size and cell number contribution in
846 two newly established and one ancient body size cline of *Drosophila subobscura*. *Evolution*
847 57:566–573.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

880 François O, Martins H, Caye K, Schoville SD. 2016. Controlling false discoveries in genome
881 scans for selection. *Mol Ecol* 25:454–469.

882 Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P,
883 Kerdelhué C, Estoup A. 2013. Estimation of population allele frequencies from next-
884 generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol* 22:3766–
885 3779.

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

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

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

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

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

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

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

900 Hamby KA, E. Bellamy D, Chiu JC, Lee JC, Walton VM, Wiman NG, York RM, Biondi A.
901 2016. Biotic and abiotic factors impacting development, behavior, phenology, and
902 reproductive biology of *Drosophila suzukii*. *J Pest Sci* 89:605–619.

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

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

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

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

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

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

916 James RE, Broihier HT. 2011. Crimpy inhibits the BMP homolog Gbb in motoneurons to
917 enable proper growth control at the *Drosophila* neuromuscular junction. *Development*
918 138:3273–3286.

919 Kanzawa, T. 1939. Studies on *Drosophila suzukii* mats. Yamanashi Agricultural Experimental
920 Station, Kofu, Japan.

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

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

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

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

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

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

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

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

943 Kwadha CA, Okwaro LA, Kleman I, Rehermann G, Revadi S, Ndlela S, Khamis FM, Nderitu
944 PW, Kasina M, George MK, et al. 2021. Detection of the spotted wing drosophila,
945 *Drosophila suzukii*, in continental sub-Saharan Africa. *J Pest Sci* 94:251–259.

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

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

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

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

955 Langley CH, Crepeau M, Cardeno C, Corbett-Detig R, Stevens K. 2011. Circumventing
956 Heterozygosity: Sequencing the Amplified Genome of a Single Haploid *Drosophila*
957 *melanogaster* Embryo. *Genetics* 188:239–246.

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

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

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

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

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

971 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
972 1000 genome project data processing subgroup. 2009. The Sequence Alignment/Map
973 format and SAMtools. *Bioinformatics* 25:2078–2079.

974 Li W, Wang F, Menut L, Gao F-B. 2004. BTB/POZ-zinc finger protein abrupt suppresses
975 dendritic branching in a neuronal subtype-specific and dosage-dependent manner. *Neuron*
976 43:823–834.

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

979 Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local
980 adaptation depends on sampling design and statistical method. *Mol Ecol* 24:1031–1046.

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

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

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

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

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

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

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

999 Paris M, Boyer R, Jaenichen R, Wolf J, Karageorgi M, Green J, Cagnon M, Parinello H, Estoup
1000 A, Gautier M, et al. 2020. Near-chromosome level genome assembly of the fruit pest
1001 *Drosophila suzukii* using long-read sequencing. *Sci Rep* 10:11227.

1002 Peng FT. 1937. On some species of *Drosophila* from China. *Annot zool jap* 16:20–27.

1003 Pool JE. 2015. The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D.*
1004 *melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol Biol*
1005 *Evol* 32:3236–3251.

1006 Pool JE, Braun DT, Lack JB. 2017. Parallel evolution of cold tolerance within *Drosophila*
1007 *melanogaster*. *Mol Biol Evol* 34:349–360.

1008 Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P,
1009 Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population genomics of sub-Saharan
1010 *Drosophila melanogaster*: African diversity and non-African admixture. *PLOS Genetics*
1011 8:e1003080.

1012 Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. *Evolution*
1013 61:3001–3006.

1014 Prakash S, Caldwell JC, Eberl DF, Clandinin TR. 2005. *Drosophila* N-cadherin mediates an
1015 attractive interaction between photoreceptor axons and their targets. *Nat Neurosci* 8:443–
1016 450.

1017 Prelich G. 2012. Gene overexpression: uses, mechanisms, and interpretation. *Genetics* 190:841–
1018 854.

1019 Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ. 2008. Adaptive evolution in
1020 invasive species. *Trends Plant Sci* 13:288–294.

1021 Reijnders MJMF, Waterhouse RM. 2021. Summary visualizations of Gene Ontology terms with
1022 GO-Figure! *Front Bioinform* 1:6.

1023 Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to
1024 environmental association analysis in landscape genomics. *Mol Ecol* 24:4348–4370.

1025 Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a
1026 short-term genetic distance. *Genetics* 105:767–779.

1027 Rezende GL, Martins AJ, Gentile C, Farnesi LC, Pelajo-Machado M, Peixoto AA, Valle D.
1028 2008. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized
1029 Serosal Cuticle. *BMC Devel Biol* 8:82.

1030 Reznick DN, Losos J, Travis J. 2019. From low to high gear: there has been a paradigm shift in
1031 our understanding of evolution. *Ecol Let* 22:233–244.

1032 Rogers WA, Salomone JR, Tacy DJ, Camino EM, Davis KA, Rebeiz M, Williams TM. 2013.
1033 Recurrent modification of a conserved cis-regulatory element underlies fruit fly
1034 pigmentation diversity. *PLOS Genet* 9:e1003740.

1035 Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T,
1036 Petrov DA, Schmidt P. 2022. Direct observation of adaptive tracking on ecological time
1037 scales in *Drosophila*. *Science* 375:eabj7484.

1038 Senthilan PR, Piepenbrock D, Ovezmyradov G, Nadrowski B, Bechstedt S, Pauls S, Winkler M,
1039 Möbius W, Howard J, Göpfert MC. 2012. *Drosophila* auditory organ genes and genetic
1040 hearing defects. *Cell* 150:1042–1054.

1041 Shalem O, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR–
1042 Cas9. *Nat Rev Genet* 16:299–311.

1043 Shearer PW, West JD, Walton VM, Brown PH, Svetec N, Chiu JC. 2016. Seasonal cues induce
1044 phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. *BMC Ecol* 16:11.

1045 Simoes da Silva CJ, Sospedra I, Aparicio R, Busturia A. 2019. The microRNA-306/abrupt
1046 regulatory axis controls wing and haltere growth in *Drosophila*. *Mech Dev* 158:103555.

1047 Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* 23:23–35.

1048 Stern DL. 2014. Identification of loci that cause phenotypic variation in diverse species with the
1049 reciprocal hemizygosity test. *Trends Genet* 30:547–554.

1050 Storey JD. 2003. The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann*
1051 *Stat* 31:2013–2035.

1052 Sugimura K, Satoh D, Estes P, Crews S, Uemura T. 2004. Development of Morphological
1053 Diversity of Dendrites in *Drosophila* by the BTB-Zinc Finger Protein Abrupt. *Neuron*
1054 43:809–822.

1055 Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ,
1056 Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny
1057 of 155 *Drosophila* genomes. *Curr Biol* 32:111-123.e5.

1058 Tan CC, Hsu TC, Sheng TC. 1949. Known *Drosophila* species in China with descriptions of
1059 twelve new species. *Univ Texas Publ* 4929:196–206.

1060 Telonis-Scott M, Gane M, DeGaris S, Sgrò CM, Hoffmann AA. 2012. High resolution mapping
1061 of candidate alleles for desiccation resistance in *Drosophila melanogaster* under selection. *Mol*
1062 *Biol Evol* 29:1335–1351.

1063 True JR, Mercer JM, Laurie CC. 1996. Differences in crossover frequency and distribution
1064 among three sibling species of *Drosophila*. *Genetics* 142:507–523.

1065 Turner TL. 2014. Fine-mapping natural alleles: quantitative complementation to the rescue. *Mol*
1066 *Ecol* 23:2377–2382.

1067 Ueno K, Kidokoro Y. 2008. Adenylyl cyclase encoded by AC78C participates in sugar
1068 perception in *Drosophila melanogaster*. *Eur J Neurosci* 28:1956–1966.

1069 de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. 2014. Genome scan methods
1070 against more complex models: when and how much should we trust them? *Mol Ecol*
1071 23:2006–2019.

1072 de Villemereuil P, Gaggiotti OE. 2015. A new F_{ST}-based method to uncover local adaptation
1073 using environmental variables. *Methods in Ecology and Evolution* 6:1248–1258.

1074 Wallingford AK, Rice KB, Leskey TC, Loeb GM. 2018. Overwintering behavior of *drosophila*
1075 *suzukii*, and potential springtime diets for egg maturation. *Environ Entomol* 47:1266–1273.

1076 Wang S-H, Simcox A, Campbell G. 2000. Dual role for *Drosophila* epidermal growth factor
1077 receptor signaling in early wing disc development. *Genes Dev* 14:2271–2276.

1078 Wang T, Morency DT, Harris N, Davis GW. 2020. Epigenetic signaling in glia controls
1079 presynaptic homeostatic plasticity. *Neuron* 105:491-505.e3.

1080 Wang Y, Farine J-P, Yang Y, Yang J, Tang W, Gehring N, Ferveur J-F, Moussian B. 2020.
1081 Transcriptional control of quality differences in the lipid-based cuticle barrier in *Drosophila*
1082 *suzukii* and *Drosophila melanogaster*. *Front Genet* 11:887.

1083 Wang Y, Ferveur J-F, Moussian B. 2021. Eco-genetics of desiccation resistance in *Drosophila*. *Biol*
1084 *Rev* 96:1421–1440.

1085 Welles SR, Dlugosch KM. 2019. Population genomics of colonization and invasion. In: Rajora
1086 OP, editor. Population genomics: concepts, approaches and applications. *Population*
1087 *Genomics. Cham: Springer International Publishing.* p. 655–683.

1088 Wong MM-K, Liu M-F, Chiu SK. 2015. Cropped, *Drosophila* transcription factor AP-4, controls
1089 tracheal terminal branching and cell growth. *BMC Devel Biol* 15:20.