






Genomic Diversity Illuminates the Environmental Adaptation of *Drosophila suzukii*

Siyuan Feng ^{1,*}, Samuel P. DeGrey ^{2,3}, Christelle Guédot ², Sean D. Schoville ², John E. Pool ^{1,*}

¹Laboratory of Genetics, University of Wisconsin–Madison, Madison, WI, USA

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

³Present address: Kimberly Research and Extension Center, University of Idaho, Kimberly, ID, USA

*Corresponding authors: E-mails: jpool@wisc.edu (J.E.P.); siyuanfeng.bio@outlook.com (S.F.)

Accepted: August 21, 2024

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 dataset. 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 the 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.

Key words: environmental adaptation, genotype–environment association, *Drosophila suzukii*, invasion genomics.

Significance

While prior population genetic studies have examined the demographic history of *Drosophila suzukii*, the genetic changes underlying this important agricultural pest's very recent adaptation to diverse worldwide environments remain essentially unknown. We apply population genomic analyses on whole-genome data of 29 population samples across 4 continents and 2 islands, allowing us to gain an unprecedented view of the environmental adaptation of *D. suzukii*. We find, in spite of the recent timescale of the species' geographic expansion, variants at numerous genes show significant associations with altitude, wind speed, precipitation, temperature, and human land usage. We also find some processes—particularly those associated with the nervous system—have had broad adaptive importance with regard to different environmental gradients.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

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 timescales (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 Matsumura, 1931, also known as spotted wing drosophila (SWD), 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, followed by Réunion Island (Indian Ocean) and South America, and then North and sub-Saharan Africa (Hauser 2011; Calabria et al. 2012; Asplen et al. 2015; 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 fruit crops (Knapp et al. 2021). It also exhibits remarkable genetic diversity and phenotypic plasticity in behavior, morphology, and physiology, e.g. temperature and desiccation tolerance (Little et al. 2020; Olazuaga et al. 2020), which may facilitate its adaptation to different climatic conditions and host plants (Gibert et al. 2019; Little et al. 2020). To obtain a comprehensive evolutionary genetic understanding of the invasion success of *D. suzukii*, we need to understand the genetic basis and ecological drivers of adaptive evolutionary changes that have allowed this species to occupy diverse worldwide environments.

Multiple genetic studies have investigated the demographic history of invasive populations of *D. suzukii*. Such studies have found greater levels of genetic structure between than within continents (Adrion et al. 2014; Lewald et al. 2021), suggesting independent invasions into Europe

and North America. These inferences were supported by an approximate Bayesian analysis of microsatellite data, which also indicated that some invading populations had multiple genetic sources (Fraimout et al. 2017). While minor differences are suggested in the specific admixture events that have occurred (Fraimout et al. 2017; Lewald et al. 2021), and inferences of the geographic origins of invading populations are limited by incomplete sampling from the species' Asian range, these studies suggest some emerging consensus about the invasion history of *D. suzukii*, such as multiple invasions from Asia.

In contrast, the genetic basis of environmental adaptation in *D. suzukii* is still largely unexplored. Among the few relevant studies is that of Olazuaga et al. (2020). They used whole-genome sequencing data from 22 worldwide populations to search for single-nucleotide polymorphisms (SNPs) with greater frequency differences between Asian (China and Japan) and non-Asian populations than are observed at most loci due to founder event bottlenecks, in hopes of identifying genetic variants that may underlie the invasion success of introduced populations. A subsequent study examining this same dataset also found a small number of transposon insertions with strong frequency differences between American/European and Asian populations (Mével et al. 2021). A different study identified F_{ST} outliers among Hawaiian *D. suzukii* populations (Koch et al. 2020), whereas no prior study has incorporated environmental information into the population genomic study of adaptive evolution in *D. suzukii*, and therefore, the genetic changes that may have helped *D. suzukii* to adapt to specific environmental conditions during its global expansion are largely unknown.

With the increasing availability of multi-population genomic resources, genotype–environment association (GEA), also known as environmental association analysis (EAA), is becoming a widely used approach to understand the relationship between specific environmental factors and adaptive genetic variation (Rellstab et al. 2015). GEA is also useful in identifying subtle changes in allele frequencies that are difficult to detect with outlier tests based on traditional population genomic approaches, especially when the number of studied populations is relatively large, and there is high gene flow counteracting patterns of local adaptation (Kawecki and Ebert 2004). The capability of GEA to identify adaptive genetic changes and environmental drivers of local adaptation has been demonstrated with whole-genome pool-seq data from *Drosophila melanogaster* (Bogaerts-Márquez et al. 2021). Therefore, GEA could be helpful in understanding the GE relationships underlying local invasion success in *D. suzukii*.

In the present study, we perform population genomic analyses on whole-genome pool-seq data of 29 population samples, integrating both published and novel data (supplementary table S1, Supplementary Material online),

from native and invasive ranges to investigate the environmental adaptation of *D. suzukii*. We investigate the geographic pattern of genetic diversity and population structure, in part to inform our choice of GEA methodology. We then test the association between SNP frequencies and nine environmental variables across sampling locations, identifying both specific and shared functional signatures of adaptation to these diverse selective pressures.

Results

Genomic Diversity and Population Structure of *D. suzukii*

To investigate the genetic diversity of *D. suzukii* and to examine the genetic input for environmental association analysis, we summarized the genomic polymorphism of 29 *D. suzukii* populations derived from Asia ($n = 8$), Europe ($n = 11$), North ($n = 9$) and South ($n = 1$) America, and the Indian ($n = 1$) and Pacific ($n = 1$) Oceans. These collections encompass both newly reported and previously published samples (Fig. 1a; [supplementary table S1, Supplementary Material online](#); Olazcuaga et al. 2020). Whole-genome sequences were obtained from 29 pooled samples consisting of 50 to 212 female and male individuals (i.e. autosomal haploid sample size of 100 to 424). The depth of mapped reads after quality control ranged from 23 \times to 66 \times among population samples, with an average of 45 \times ([supplementary table S1, Supplementary Material online](#)).

We first estimated nucleotide diversity (π_S) across 1,031,687 (autosomes) and 198,846 (X chromosome) putative synonymous SNPs to investigate the effects of rapid invasions on neutral genetic diversity. The genome-wide π_S ranged from an average of 0.045 (autosomes, $\sigma = 0.0016$) and 0.025 (X chromosome, $\sigma = 0.0028$), in introduced populations, to an average of 0.051 (autosomes, $\sigma = 0.0036$) and 0.034 (X chromosome, $\sigma = 0.0042$) in native Asian populations ([supplementary table S2, Supplementary Material online](#)). Previous studies have hypothesized a relatively wide native range of *D. suzukii* in East and Southeast Asia (Adrion et al. 2014; Fraimout et al. 2017), and the significant drop in π_S of introduced European and American populations relative to that of the native Chinese and Japanese populations reflects previously reported founder event bottlenecks (Fig. 1b, diagonal; [supplementary figs. S1 and S2a and table S2, Supplementary Material online](#)). The patterns observed at synonymous sites were also recapitulated with SNPs at all types of sites ([supplementary fig. S3, Supplementary Material online](#)). Lower π_S and more contrasting among-population π_S differences in X chromosome reflect effectively prolonged bottlenecks due to their lower effective population size (Pool and Nielsen 2007; [supplementary fig. S1 and table S2, Supplementary Material online](#)). We also observed a greater loss of rare alleles in the introduced populations as a

typical consequence of bottlenecks, which is similarly more obvious for the X chromosome ([supplementary fig. S4, Supplementary Material online](#)). These founder events also increased genetic differentiation as measured by F_{ST} , especially between continents (Fig. 1b; [supplementary fig. S3, Supplementary Material online](#)), and particularly for the X chromosome ([supplementary fig. S2a, Supplementary Material online](#)).

Subsequent to our analysis, contamination in three of the published samples was reported, involving reads from the closely related *Drosophila subpulchrella* and from *Drosophila immigrans* (Gautier 2023), which can be estimated but not fully removed. The estimated level of *D. subpulchrella*-sourced contamination for Tokyo, Japan (JP-Tok), was reduced from 4.47% to 3.58% after our alignment and quality filtering pipeline ([supplementary table S3, Supplementary Material online](#)), which is slightly higher than the baseline noise level of 1.14% to 2.78% from samples known to be pure *D. suzukii* (Gautier 2023). The published Jena, Germany (DE-Jen), sample's contamination from *D. immigrans* was reduced from 5.79% to 0.35%. Our unpublished sample from the same area showed 0.28% *D. immigrans* contamination from mapped reads, whereas the other unpublished samples showed no evidence for contamination ([supplementary table S3, Supplementary Material online](#)). However, a higher proportion of reads from *D. subpulchrella* remained in the East Asian sample CN-Nin (10.38% vs. 14.95% from raw reads). Hence, it is likely that the elevated π_S of CN-Nin (and likewise its elevated D_{XY} values from the analysis described below) reflect an artifact of this contamination. By comparing results with and without contaminated samples, we noted that their inclusion neither changes the statistical significance of the diversity difference between native and introduced populations ([supplementary table S2, Supplementary Material online](#)) nor the patterns of population grouping in PCA ([supplementary fig. S5, Supplementary Material online](#)).

We next analyzed genetic structure among the sampled populations, using the summary statistics D_{XY} and F_{ST} and via principal component analysis (PCA) of population allele frequencies, particularly since the pattern of genetic structure present may influence the performance of GEA methods (Rellstab et al. 2015). The top three principal components (PCs 1 to 3) explained 57.44% (autosomes, Fig. 1c) and 72.35% (X chromosome; [supplementary fig. S2b, Supplementary Material online](#)) of the variance among populations. For both autosomes and the X, 3D PCA and matrices of both D_{XY} and window F_{ST} recapitulated both continuous and hierarchical geographic structure (Olazcuaga et al. 2020), which must be accounted for in GEA. These results together showed the expected clustering of populations into four distinct ranges (East Asia, Hawaii, Americas, and Europe; Fig. 1b and c; [supplementary fig. S2, Supplementary Material online](#)). Much of the observed population

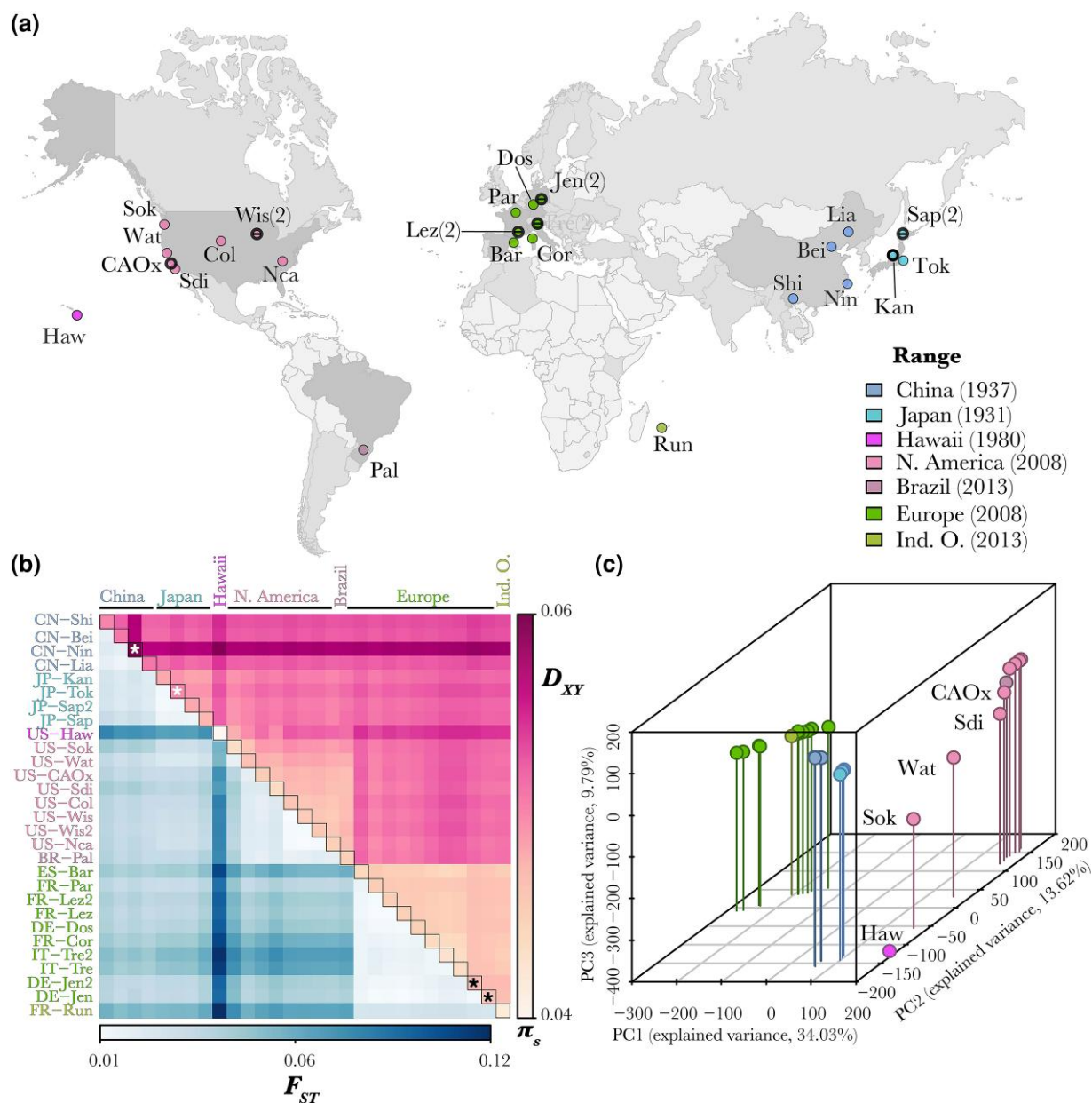


Fig. 1. *D. sukukii* populations show maximal diversity in Eastern Asia and continent-level genetic structure. a) The geographic locations of the studied 29 natural populations are depicted as dots. In addition to the 22 populations sampled by Olazcuaga et al. (2020), populations newly sampled at independent locations are circled in black. Populations newly sampled at nearby locations are circled and center-dashed in black, with the number of total population samples in brackets. The year of the first recorded occurrence in each geographic range (colored grey in the map) is given in brackets in the color legend. China (CN) and Japan (JP) are within the native range of *D. sukukii*. The gray shading indicates countries with samples represented in this study (darkest), those with documented occurrence of *D. sukukii* but not sampled in this study (medium), or those lacking occurrence records of *D. sukukii* (lightest) (Bächli 2016; Rossi Stacconi 2022). Further information about each sample is presented in [supplementary table S1, Supplementary Material](#) online. b) Population differentiation in allele frequencies (F_{ST} ; lower triangle), between-population sequence distances (D_{XY} ; higher triangle), and within-population nucleotide diversity (π_s ; diagonal) across autosomal synonymous SNPs are displayed as a heatmap. D_{XY} and π_s share the same color scale, since theoretically D_{XY} between genetically identical populations is π . Population names are colored by their geographic region. Asterisks indicate samples contaminated by other *Drosophila* species, which may affect estimation of these statistics. c) Autosomal genetic structure is shown by three-dimensional principal components analysis (PCA) based on allele frequencies of the two most frequent alleles across all populations. Each dot represents a population. Labeled are Hawaii and western coastal US populations, to illustrate potential admixture. See the X chromosome version of b) and c) in [supplementary fig. S2, Supplementary Material](#) online.

differentiation is most likely due to founder event bottlenecks and admixture during worldwide expansion (Framout et al. 2017), whereas migration among introduced populations following population establishment would need to be overwhelmingly high to have significant impacts given the very brief timescale of the global invasion (under 40 years). The two more northerly populations from the Western United States, Oregon (US-Sok), and Central California (US-Wat), are genetically closer than other populations to the Hawaiian population (Fig. 1c), which aligns with the suggestion that these populations received a genetic contribution from Hawaii in addition to East Asia (Framout et al. 2017), while populations from southern California and the Central and Eastern United States show less evidence of such admixture.

Our focus above on diversity and differentiation at synonymous sites was motivated by the low selective constraint expected at these sites, which may offer a closer estimate of neutral diversity and differentiation. This expectation was confirmed by our analysis of divergence between *D. sukukii* and its relative *Drosophila biarmipes* (supplementary fig. S6, Supplementary Material online; see Materials and Methods). We also note that *D. sukukii* and *D. biarmipes* show relatively higher ratios of intergenic (or intronic) divergence to nonsynonymous divergence ($\frac{Div_{intergenic}}{Div_{non-synonymous}} = 4.09$, $\frac{Div_{intronic}}{Div_{non-synonymous}} = 4.44$; supplementary fig. S6, Supplementary Material online), compared to those between *D. melanogaster* and *Drosophila simulans* ($\frac{Div_{intergenic}}{Div_{non-synonymous}} = 3.13$, $\frac{Div_{intronic}}{Div_{non-synonymous}} = 3.05$; Table S1 in Lange and Pool 2018). Compared with *D. melanogaster*, the genome of *D. sukukii* contains a notable expansion of repetitive sequences (Paris et al. 2020), which could reflect a lower long-term effective population size (N_e) in the *D. sukukii* lineage since its divergence from the *D. melanogaster* lineage (Lynch and Conery 2003). However, the greater π of *D. sukukii* than *D. melanogaster* (supplementary fig. S1, Supplementary Material online; Lack et al. 2016a; Lewald et al. 2021) could instead indicate a greater N_e for *D. sukukii* within the past four N_e generations.

Polymorphism and the Genomic Locations of *D. sukukii* Contigs

If extensive genomic regions of low recombination are present in a genome-wide scan for local adaptation, then because of the larger-scale influence of natural selection on linked sites in such regions (Smith and Haigh 1974; Charlesworth et al. 1993), the precision of outlier identification will be reduced (Lotterhos and Whitlock 2015; François et al. 2016). Although we do not have recombination rate estimates for *D. sukukii*, we can begin to assess the genomic abundance of low recombination regions through an examination of nucleotide diversity, in light of its expected correlation with recombination rate (Begun and Aquadro 1992). Fortunately, we observe that regions of low nucleotide

diversity (which probably coincide with regions of low recombination) cover a relatively small fraction of the genome (Fig. 2). These patterns are more similar to those in *D. simulans* than to *D. melanogaster*—which has broader centromeric regions of low crossing-over on the autosomes (Figure 15 in Langley et al. 2012)—potentially suggesting a relatively weaker suppression of crossing-over in the centromere-proximal regions in *D. sukukii*.

We next leveraged our large pooled sequencing dataset to improve inferences about which contigs map to the X chromosome, in part so that X-linked and autosomal contigs could be more accurately partitioned in our subsequent GEA. Out of a total of 546 contigs, 313 were previously assigned to autosomes and X chromosome through either direct mapping or comparing a female-to-male read depth ratio (Paris et al. 2020). We added to these annotations by implementing an approach based on correlations in sequencing depth of coverage across population samples that included varying numbers of females and males (see Materials and Methods). Based on this analysis, we assigned 170 contigs as autosomal or X-linked (supplementary table S4, Supplementary Material online). Our classification of previously assigned contigs was 96% consistent with past inferences, but we corrected four previous assignments that were based on female-to-male read depth ratio, whereas our method did not assign eight previously assigned contigs.

Selection of Robust Methodology and Distinct Environmental Variables for GEA

The worldwide expansion of *D. sukukii* has exposed this species to selection pressures from varying local environmental conditions (Olazcuaga et al. 2020; Mérel et al. 2021). To identify environmental factors that have contributed to adaptive genetic differentiation at various levels and loci under positive selection, while controlling for hierarchical genetic structure, we performed a whole-genome scan using GEA analysis between environmental and genetic differentiation using BayeScEnv (de Villemereuil and Gaggiotti 2015; Materials and Methods).

Our selection of environmental variables for GEA started with a preliminary set of 26 candidate variables that are potentially relevant in the adaptation process of *D. sukukii* (Fig. 3a; see detailed definitions of variables in supplementary table S5, Supplementary Material online). Although overlooked in many other GEA studies, we included human land use variables because land use has major impacts on insect ecology (Uhler et al. 2021; John et al. 2022; Harvey et al. 2023). Since univariate association with all environmental variables could increase the number of statistical tests, thus increasing the difficulty of controlling rates of false discovery, we opted to retain nine of the least correlated environmental variables for univariate tests, representing altitude, wind speed, and multiple aspects of

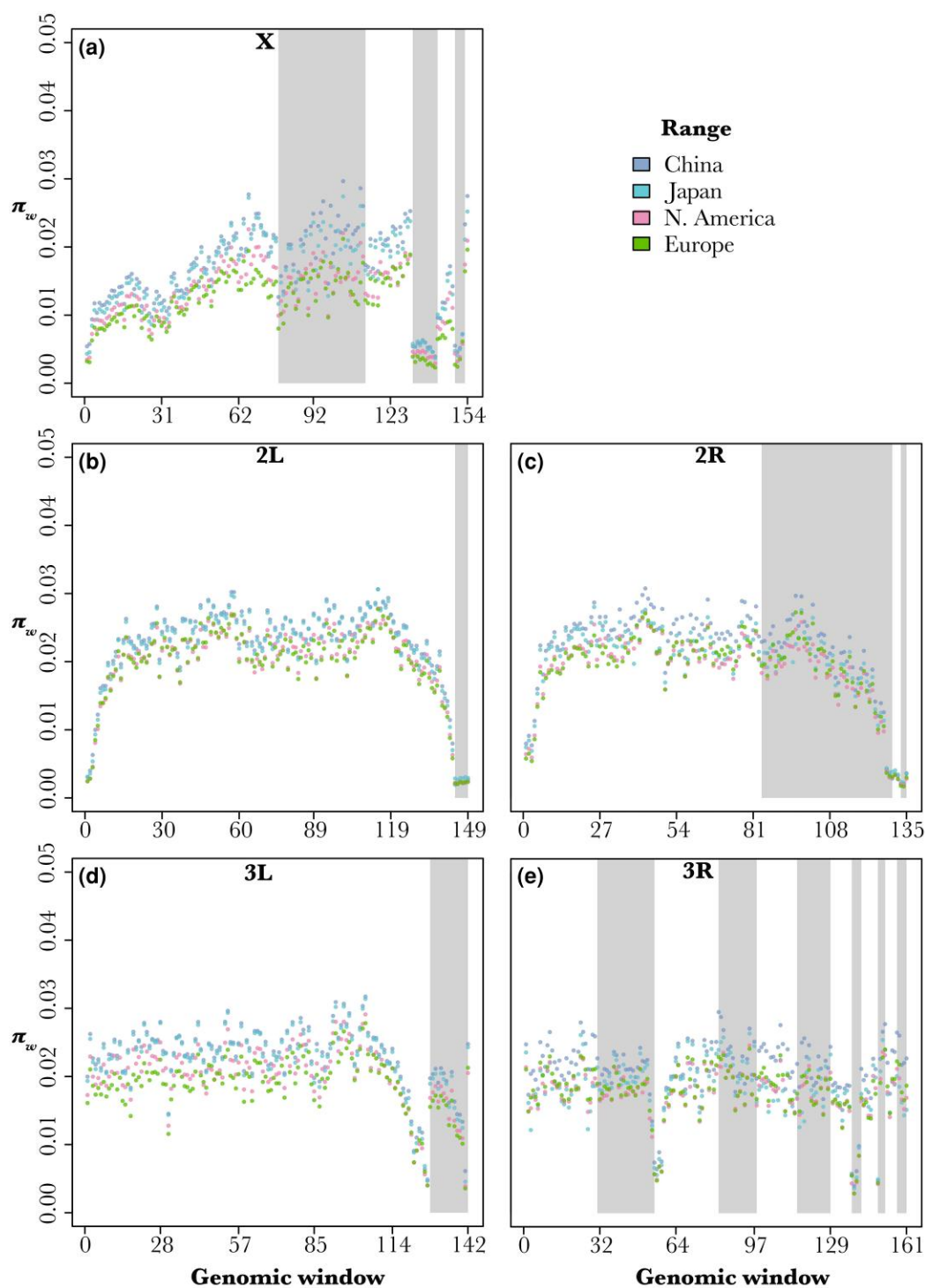


Fig. 2. Chromosomal distribution of genetic polymorphism in *D. suzukii* informs the ordering and orientations of contigs, as well as levels of centromeric and telomeric repression. Window nucleotide diversity (π_w) values are displayed across a) the X chromosome and major autosomal arms b) 2L, c) 2R, d) 3L, e) 3R. Chromosome 4 is not shown as it only contains 12 windows. Each window is a continuous genomic region that includes 125,000 analyzed sites. Each dot represents the average π_w across populations within their geographic range as colored. Only populations from major continental ranges are shown for chromosomal patterns to be clear. Within each chromosome arm, separate contigs are indicated by gray or white shading, and ordered by length. However, we note that certain arrangements of contigs would result in patterns of reduced π_w at the ends of each arm, as expected based on other examined *D. melanogaster* group species (e.g. True et al. 1996), and relatively smooth shifts in the diversity of large windows. Therefore, the landscape of genome-wide polymorphism could provide useful information to aid the ordering and orienting of contig-level genome assemblies like that of *D. suzukii*.

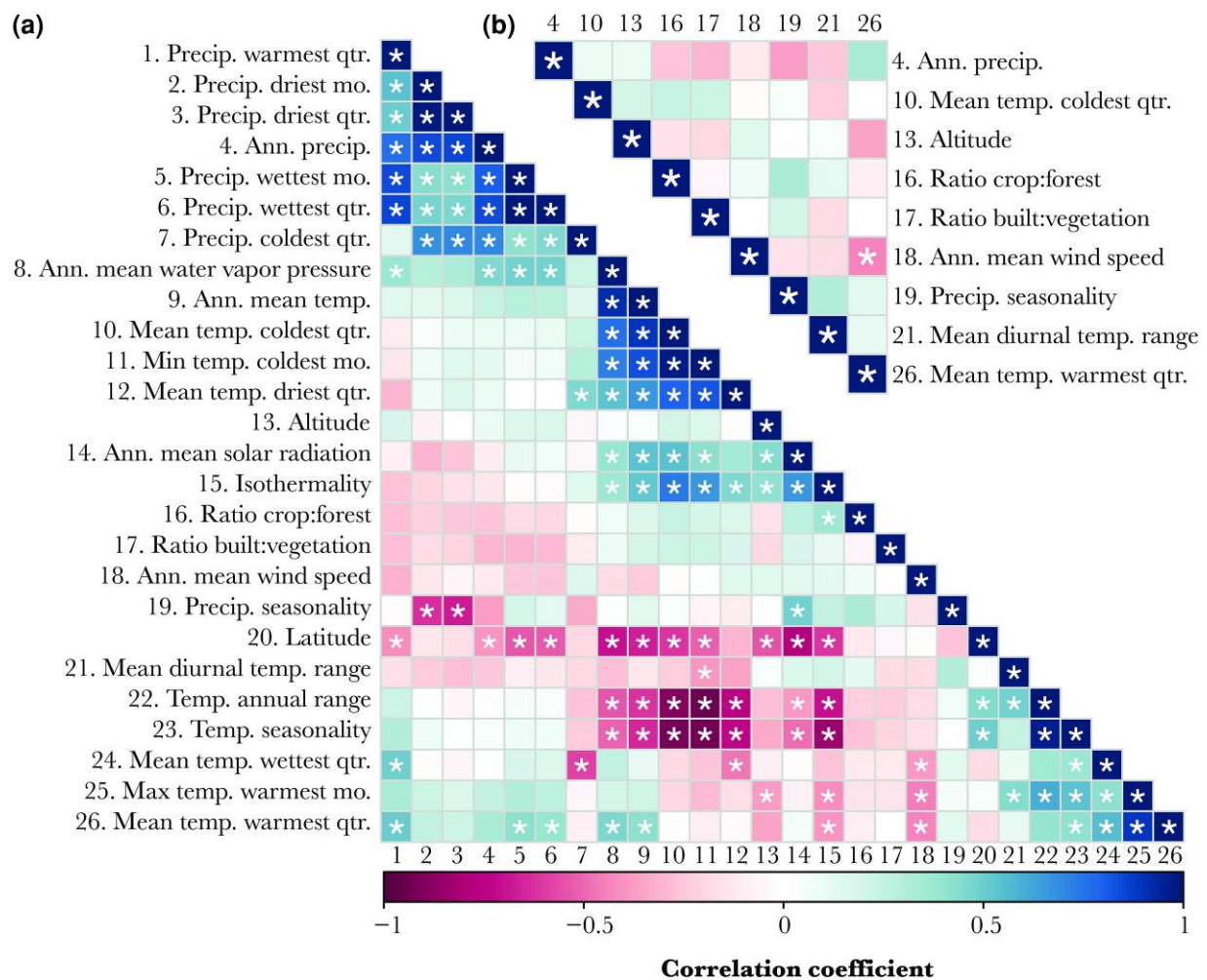


Fig. 3. Identification of least-correlated environmental variables for GEA analysis in *D. suzukii*. a) Pairwise correlations among a preliminary set of 26 environmental variables that are potentially impactful on *D. suzukii*. b) A final set of nine of the most relevant and least correlated environmental 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 [supplementary table S4, Supplementary Material](#) online for environmental values used to calculate correlation coefficients.

temperature, precipitation, and human land use for GEA analysis (Fig. 3b). Although the temperature of the coldest quarter had a significant negative correlation with wind speed, we kept both variables for GEA, as cold stress and wind-related factors are known to be potential drivers of local adaptation in *Drosophila* (Bogaerts-Márquez et al. 2021). As indicated by their coefficients of variation (CVs), the selected environmental variables had moderate ($15\% \leq CV < 30\%$)-to-high ($CV \geq 30\%$) variability across our sampling locations ([supplementary table S5, Supplementary Material](#) online). We further examined the environmental differentiation among those locations by performing PCA on standardized values of the full set of 26 candidate environmental variables ([supplementary fig. S7a, Supplementary Material](#) online), as well as the nine variables retained for GEA ([supplementary fig. S7b,](#)

[Supplementary Material](#) online). In both analyses, populations of different introduced/native status and continental origins were largely interspersed ([supplementary fig. S7a, Supplementary Material](#) online).

For genotype data, we excluded SNPs with a global average minor allele frequency (MAF) below 5%, which should minimize the influence of the contamination documented above. No alleles specific to contaminating species should meet that threshold, so the only remaining effect should be a slight bias in allele frequency estimation toward ancestral variants in contaminated samples at genuine *D. suzukii* SNPs. Given this consideration, as well as the lack of correlation between the contamination levels and environmental values ([supplementary table S3, Supplementary Material](#) online), contamination may represent a modest source of noise in the GEA, but it should not lead to spurious adaptation signals.

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 (genome-wide $q < 0.05$) associated with each of the nine candidate environmental variables and that have the lowest q within a 20 kb genomic interval (Materials and Methods; [supplementary table S6, Supplementary Material](#) online). These thinned sets of variants (which are the objects of all analyses described below) corresponded to an average of 3,346 overlapping or neighboring genes per environmental variable ([supplementary table S6, Supplementary Material](#) online), suggesting that selection pressures from the tested environmental variables (or correlated factors) have been associated with substantial adaptive genetic responses in *D. sukuzii*, even on the brief timescale of its worldwide expansion. Among all tested environmental factors, mean temperature of the coldest quarter was associated with the greatest number of putatively adaptive variants (4,250 SNPs). Two precipitation-related variables, annual precipitation (4,141 SNPs) and precipitation seasonality (3,389 SNPs), have the next largest loci count. The ratio of built area to vegetation (i.e. crops and forests) and the ratio of crops to forests were associated with the fewest genetic variants (2,369 and 1,608 SNPs, respectively).

Despite the varying numbers of thinned significant variants associated with each environmental variable, the enrichment of site types or genomic elements at these variants are similar across environmental variables ([supplementary table S7, Supplementary Material](#) online). Averaging across all environmental variables, the most enriched genomic element was RNA-coding genes, for which outliers were observed 42% more often than expected by chance, and positive enrichments were detected for all but one environmental variable. Perhaps surprisingly, not only 5' and 3' untranslated regions but also introns and intergenic regions were more enriched than sites from protein-coding exons, and <1% of outliers were nonsynonymous ([supplementary table S7, Supplementary Material](#) online), which may indicate a predominant role for regulatory rather than protein-coding changes in this species' environmental adaptation.

To reveal the potential genetic and functional basis of invasion success under multiple environmental challenges throughout the species' range, we examined the functions of genes linked to the top ten environment-associated loci for each variable as ranked by association q -value, and then by the g parameter estimating the sensitivity to environmental differentiation as a tiebreaker ([supplementary table S7, Supplementary Material](#) online). Many of the top variants implicating these genes had estimated q -values of 0, and all would remain significant if multiple testing corrections were extended on an experiment-wide basis across all nine

environmental variables. From an analysis of published literature, we found many of these genes have known functions that could facilitate adaptation to the associated environmental factor.

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

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

There is also some evidence for precipitation-related local adaptation. The top gene *mmy* associated with precipitation seasonality (i.e. the coefficient of variation) was shown to regulate chitin synthesis and cuticle production. Since precipitation is correlated with desiccation resistance across the *Drosophila* phylogeny (Kellermann et al. 2012), *D. sukuzii* may have developed adaptive strategies of modifying chitin biosynthesis under conditions of desiccation (Rezende et al. 2008; Clark et al. 2009), which was also implied in seasonal plasticity of natural *Drosophila* populations (Shearer et al. 2016; Horváth et al. 2023). In addition, the gene *osy* (*CG33970*) contributes to the formation of the outer cuticle layer and is expressed more highly in *D. sukuzii* than in *D. melanogaster* (Wang et al. 2020a). Furthermore, two of the top genes associated with annual precipitation (*Abd-B* and *bab1*) regulate cuticle pigmentation (Rogers et al. 2013), which may or may not correlate with desiccation tolerance in *Drosophila* species (Wang et al. 2021). Both genes were also found to be associated with precipitation in *D. melanogaster* and to be differentially expressed in response to desiccation stress (Bogaerts-Márquez et al. 2021; Horváth et al. 2023). We also note that although environmental fitness effects on these testes-expressed genes are not known, the same SNP near *CG17944* and *nx4* was among the highest-scoring variants for both annual precipitation and precipitation seasonality (variables that have a non-significantly negative correlation between them; Fig. 3).

Another important environmental barrier to invasion success is temperature. For the mean temperature of the coldest

quarter, a top gene was *Ac78C*, which has roles in circadian regulation and taste (Ueno and Kidokoro 2008; Duvall and Taghert 2013). With the mean temperature of the warmest quarter, the top genes *crp*, *Mrtf*, and *Ubx* help control the development of trachea (Han et al. 2004; Guha and Kornberg 2005; Wong et al. 2015), which may be important in limiting water loss in hot environments (Gibbs et al. 2003). *Ubx* was also found associated with temperature variables in *D. melanogaster* (Bogaerts-Márquez et al. 2021).

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

Beyond genes that have related functions to specific types of environmental changes, we also found a wide range of nervous system genes associated with multiple environmental factors. For instance, among the top five altitude-associated loci, three have known functions in the nervous system of *Drosophila*, including the first-ranked gene *Cmpy*, which enables proper growth control at neuromuscular junctions (James and Broihier 2011), *ab*, which regulates dendritic complexity (Li et al. 2004; Sugimura et al. 2004), and *not*, which is essential for stabilizing synaptic homeostasis within glia (Wang et al. 2020b). Such genes were also linked to top 10 loci associated with wind speed (*dpr6*), precipitation (*Msp300*, *Tusp*, and *5-HT2A*), temperature (*ATP6AP2*, *CG13579*, and *D*), and land use variables (*Bsg*, *Mp*, and *velo*). Different variants associated with *scrt*, a regulator of neuronal cell fate, were among the top results for both mean diurnal range and the ratio of crop to forest cover.

Functional Commonalities in the Adaptation to Diverse Selective Pressures

Next, we examined environment-specific adaptation on a more comprehensive basis through a gene ontology (GO) enrichment analysis of the top 500 genes associated with each environmental variable (Fig. 4). As with the analysis of top genes mentioned above, all variants implicating the top 500 genes for each variable would remain significant if multiple testing correction was performed on an experiment-wide

basis, with the exception of some variants for the crop versus forest variable (supplementary table S7, Supplementary Material online). As correlates of temperature in the coldest quarter, cAMP metabolic process was the top enriched category, followed by two other related purine metabolism groupings. We note that cAMP is important in circadian regulation (e.g. Palacios-Muñoz and Ewer 2018), which is known to play an important role in *Drosophila* environmental adaptation (e.g. Helfrich-Förster et al. 2020), as also implicated by the presence of “entrainment of circadian clock” on our top GO category list for altitude. More broadly, purine metabolism was inferred as a strategy of cold acclimation in *D. suzukii* (Enriquez and Colinet 2019). For diurnal temperature range, the top category was “regulation of growth,” and we note that some drosophilids have evolved to have larger body sizes in more challenging thermal environments (Gilchrist and Partridge 1999; Calboli et al. 2003; Lack et al. 2016b).

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

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

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

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

Fig. 4. GO enrichment analysis of candidate genes from the gene-environment association analysis of *D. sukuzii*. The top 10 GO categories enriched by the top 500 genes associated with each environmental variable are shown in each panel (labelled on the left), with permutation *P*-values and the number of associated genes in each GO category. Descriptions of GO categories are colored by their GO class (see legend at top right). Only GO categories including more than five associated genes are listed here. For a full list of enriched GO categories, see [supplementary table S8, Supplementary Material](#) online.

temperature range, and temperature of the warmest quarter), “regulation of neurogenesis” (altitude, diurnal temperature range, temperature of the warmest quarter, and ratio of built area to vegetation), and “central nervous system development” (precipitation seasonality, diurnal

temperature range, and temperature of the warmest quarter) (Fig. 5c). Further, as mentioned above, “cAMP metabolic process” (shared by temperature of the coldest quarter, precipitation seasonality and ratio of built area to vegetation) could entail neurologically modulated changes

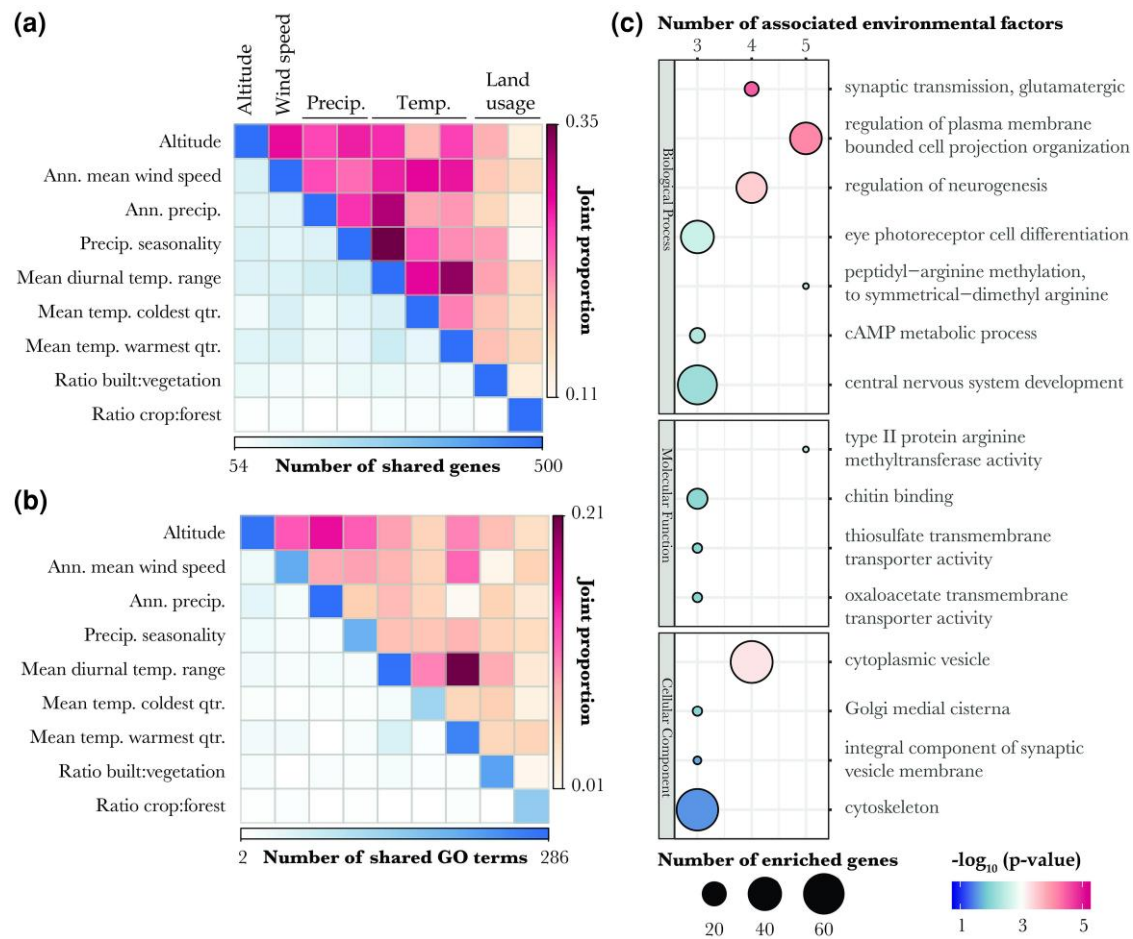


Fig. 5. Overlapping genes and GO categories among environmental factors reveal the shared genetic and functional basis of environmental adaptation in *D. sukukii*. The numbers and proportions of shared a) environment-associated genes and b) enriched GO categories among environmental factors are shown 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.

in circadian behavior. Each of these functional categories was enriched for at least three of the nine environmental variables. Thus, out of the 8,070 biological process categories analyzed, four out of the seven most shared enriched GO categories across environmental variables were connected to neurological function, potentially indicating a multifaceted role for nervous system evolution in facilitating local invasion success in *D. sukukii* under multiple environmental challenges.

Discussion

We performed population genomic analyses of 29 population samples of *D. sukukii* to investigate the genomic diversity and environmental adaptation of this highly invasive species across its worldwide distribution. Our data

supported a genetic grouping of these populations into four primary geographic regions: Eastern Asia (containing the native range), Hawaii, the Americas, and Europe. We also confirmed that all non-Asian populations have reduced diversity, consistent with moderate founder event bottlenecks in introduced populations.

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

The above analyses placed us in a more confident position to perform a robust analysis of GEA. We selected nine

distinct environmental variables, including altitude, wind speed, precipitation, temperature, and human land usage. Our results suggested extensive local adaptation in response to specific environmental challenges, along with appreciable sharing of genes and functional pathways underlying invasion success across multiple environmental pressures, which were most obvious with nervous system genes.

Environmental Drivers of Adaptation in *D. sukukii*

Here, we presented a GEA analysis that investigated the most geographically and genetically diverse set of *D. sukukii* populations and the most comprehensive set of environmental factors to date, which enabled unprecedented power to capture even minor adaptive genetic differentiation in response to distinct environmental challenges during the species' rapid invasions. While a previous study sought to identify invasion-related adaptive loci as those with allele frequency differences between native and introduced populations (Olazcuaga et al. 2020), this study is the first to explicitly dissect genetic associations with specific environmental factors in *D. sukukii*, independent of the invasive status of populations. In addition to our identification of climatic factors including temperature, precipitation-related variables, and wind speed as the most frequently correlated with putatively local adaptive variants (consistent with previous GEA analysis in *D. melanogaster*, e.g. Bogaerts-Márquez et al. 2021), we also for the first time identified large numbers of genome-wide variants associated with altitude and human land usage-related variables (supplementary table S6, Supplementary Material online), which were not included in most GEA studies despite their potential significance to local adaptation (Uhler et al. 2021; John et al. 2022; Harvey et al. 2023). In particular, the detected associations with ratios of developed land to vegetation and of cropland to forests highlights the ecological impacts of urbanization and agriculture on natural populations of insects. Since the selection of environmental variables is critical for successful GEA analyses, we also provided an instructive example for correlation-based selection to identify the most relevant and least redundant environmental factors (Rellstab et al. 2015).

Since we discovered widespread genetic signals of environmental adaptation in *D. sukukii*, it is natural to further ask to what extent introduced populations needed to adapt to novel environments, and the answer can depend on how different the introduced environments are from that of the source populations of specific invasions. We addressed this question of environmental differentiation by performing PCA on environmental values at the sampled locations. Since environmental variation is notable even within the native range (supplementary fig. S7, Supplementary Material online), and different invasive populations may have been founded from distantly distributed native populations (Framout et al.

2017), the environmental adaptation revealed from GEA is likely to be largely independent of the invasive status of *D. sukukii* populations and instead subject to distinct selective pressures posed by local and regional environmental challenges.

Our results, including the substantial numbers of environment-associated SNPs detected, raise the question of how much adaptation may have occurred since the introduction of *D. sukukii* populations to novel environments. The speed of adaptation in *D. sukukii* may be as accelerated by its rapid generation time (perhaps 13 generations per year in warm conditions; Tochen et al. 2014) and its high level of genetic diversity. Given that experimental evolution studies in other *Drosophila* species have detected adaptive phenotypic changes within just dozens of generations (e.g. Orozco-terWengel et al. 2012; Mallard et al. 2018), and that our collections of introduced populations were made between 5 and 36 years after local population establishment, some degree of local adaptation in these populations would be expected. That environmental adaptation might conceivably involve many thousands of variants worldwide may nevertheless seem surprising and could represent an intriguing topic for future simulation or theoretical studies regarding the number of potential targets of selection within a relatively brief interval.

Nervous System Evolution Is Ubiquitous in Environmental Adaptation of *Drosophila*

In *D. sukukii*, we found nervous system and related sensory and behavior annotations associated with top genes for all nine environmental variables studied. Concordantly, we found that GO categories related to the nervous system were among the most shared across environmental variables (Fig. 5c). In *D. melanogaster*, related GO categories like "neuron development," "nervous system development," and "eye development" were also enriched among genes associated with environmental variation among natural populations within North America or Europe and across seasons within Europe (Bogaerts-Márquez et al. 2021). GO categories associated with the nervous system have also shown evidence of positive selection in various genome scans of *D. melanogaster* (Langley et al. 2012; Pool et al. 2012; Pool 2015), including a study of parallel evolution in cold-adapted populations (Pool et al. 2017). Given the morphological evidence of neuron-muscular junction evolution across the entire *Drosophila* phylogeny (Campbell and Ganetzky 2012), we therefore propose a broad adaptive importance of the nervous system in *Drosophila* species and potentially other insects. Such evolutionary processes may have either maintained ancestral neural functions in novel challenging environments, or created novel phenotypes that better fit the new optima arising from complex combinations of environmental factors. Novel phenotypes conveyed by nervous system evolution could include behavioral traits that influence

the identification and selection of locally appropriate micro-environments and food sources.

Interpretations of Association Results and Future Directions

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

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

Broader Impacts and Significance

Our work integrates genetic and environmental data to improve the reconstruction of the invasion genomics of a crop pest carrying significant economic costs (Knapp et al. 2021), which will hopefully inspire future studies on developing diverse pest control methods given the adaptive and neutral genetic differentiation among *D. sukukii* populations. Understanding the extent of local adaptation and its potential environmental drivers will also help predict the spread and future distributions of invasive species (Colautti and Lau 2016). More broadly, the enhanced

understanding of how organisms may adapt to geographical, climatic, and artificial selective pressures from this study will also be of value in assessing the susceptibility of natural populations to climate change (Kellermann et al. 2012) and human activities (Barange et al. 2010).

Materials and Methods

Fly Collection, DNA Preparation, and Pooled Sequencing

Fly samples from 29 populations were used, 7 of which were sequenced for the present study. The fly samples sequenced in this study were collected from wild *D. sukukii* populations in two states of the USA, two provinces of Japan, and three European countries (Fig. 1; [supplementary table S1, Supplementary Material](#) online). Both previously and newly sequenced fly samples were collected in parallel within a 5-year span ([supplementary table S1, Supplementary Material](#) online). While five of the newly sequenced fly samples were collected from nearby locations of the previously sequenced ones, we kept all of them in our analyses, so that we could confirm the robustness of population genomic inferences and to slightly increase the power of GEA. Pooled whole adult flies ($n = 100$ to 183) from each population ([supplementary table S1, Supplementary Material](#) online) were used for DNA extraction as previously described (Langley et al. 2011). Library preparations were conducted at the Next Generation Sequencing Core of University of Wisconsin Madison Biotechnology Center (<https://dnaseq.biotech.wisc.edu>), where pair-end (PE) reads at the length of 150 bp were then generated for each of seven pooled DNA samples on an Illumina NovaSeq 6000.

Pool-sequenced reads of 22 additional *D. sukukii* population samples, including from Europe, the Americas, and Asia, were obtained from public data provided by Olazuaga et al. (2020) at EBI's SRA (Fig. 1; [supplementary table S1, Supplementary Material](#) online). Taken together, we formed a comprehensive dataset of 29 populations sampled from native and invasive ranges of *D. sukukii*.

Quality Control, Alignment, Contamination Analysis, and Variant Calling From Pool-seq Data

To maximize the quality of our analyzed data, we built a high-throughput assembly and quality control pipeline *poolWGS2SNP* with optimized performance, stringent filtering, compatibility with large numbers of genomic contigs, and customized functions to call high-confidence single-nucleotide variants from pool-sequenced data in *D. sukukii* ([supplementary fig. S8, Supplementary Material](#) online), in part by utilizing resources from the DrosEU bioinformatics pipeline (Kapun et al. 2020).

As an initial quality control of raw PE reads, adapters were removed, and the 3' end of reads with base quality < 20 were trimmed using *fastp* (Chen et al. 2018). Further trimming was

performed using a self-developed python program *filter_PE_length_mem.py* (see Data Availability), where any pair of forward and reverse reads with less than a total of 150 bases with base quality (BQ) ≥ 20 , as well as any individual reads with less than 25 bases with BQ ≥ 20 were discarded.

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

We then checked sample contamination for both newly and previously reported pool-seq data ([supplementary table S3, Supplementary Material](#) online), by estimating the proportion of pool-seq reads from different *Drosophila* species as the proportion of aligned reads assigned to species-discriminating *k-mers* (i.e. unique sequences of each species" reference genome assembly) using the approach described by Gautier (2023). Although the estimation should be reliable, fully eliminating contaminated reads is not currently practical because substantial proportion of reads cannot be confidently assigned to any species (up to 42% of total pool-seq reads), due to the sequence similarity among Drosophilid genomes (Gautier 2023). Therefore, we focus on identifying samples with contamination and interpreting results of these samples with caution. Comparisons between population genomic analyses with and without contaminated samples were also performed to evaluate their impacts on major conclusions ([supplementary table S2 and figs. S5 and S7, Supplementary Material](#) online).

To call SNPs, we merged the quality-checked BAM files of all population samples into one file using *Samtools mpileup*, only retaining alignments with mapping quality no less than 20 and sites with base quality no less than 20. As a default setting of *Samtools mpileup*, we only retained one base across any overlapping region between a pair of reads, so that the base count will not be artificially inflated. Variant calling was then performed on the *mpileup* file using the heuristic SNP caller *PoolSNP* (Kapun et al. 2020). We used a nominally low value for the parameter *miss-frac* (0.001) to require for each population sample individually, that depth of coverage at a given site be 12 or greater (*min-cov* = 12), and that this site not be in the top 1% of sites genome-wide for depth of coverage (*max-cov* = 0.99; calculated separately for each population and for autosomal and X-linked contigs), in order

to filter sites subject to copy number variation. In the initial dataset used for analysis of genome-wide diversity, we avoided potential biases from allele frequency filters by using *min-count* = 1 and *min-freq* = 0. We termed the resulting high-quality sites as "analyzed sites" for brevity.

Identifying Autosomal and X-Linked Contigs

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

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

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

Annotating Genomic Features and Estimating Divergence

To explore genomic diversity at synonymous sites and selective constraint for other site types in *D. sukuzii*, we classified the reference genome into nine exclusive categories of site degeneracy and function (Lange and Pool 2018), including non-degenerate (i.e. nonsynonymous) sites; 2-, 3-, and 4-fold degenerate (i.e. synonymous) sites; 3' and 5' untranslated regions (UTRs); RNA-coding genes; introns; and intergenic regions. From input files including the eukaryotic codon table, the published genome sequence and GFF3 annotation obtained at NCBI RefSeq, we generated a letter-coded annotation (in FASTA format) mirroring both strands of the whole-genome sequence of *D. sukuzii* and a coordinate-based annotation (in BED format) that combines adjacent sites of the same category into a single row. Degeneracy was determined based on the standard codon table. 5' UTRs were defined as regions between the start of the first exon and the start of the first coding sequence (CDS), while 3' UTRs were defined as regions between the end of the last exon and the end of the last CDS. In cases of overlapping genes and alternative splicing that raise annotation conflicts, we followed an annotation priority in the category order listed above.

We then estimated the divergence between *D. sukuzii* and its close relative *D. biarmipes* (Ometto et al. 2013; Suvorov et al. 2022) in each of these categories. We obtained results of multiple sequence alignment between the current reference genomes of *D. sukuzii* and *D. biarmipes* (Paris et al. 2020). For each site category of *D. sukuzii*, the unpolarized divergence was estimated as the number of substitutions over the total number of sites within aligned blocks of reference genome sequences.

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

To compare genome-wide polymorphism among populations, we estimated nucleotide diversity (π) across SNPs at 4-fold degenerate sites (π_s) in addition to that at all categories of sites (π_A), as π_s estimation is relatively less affected by sequencing errors than nucleotide diversity estimated from other site categories (due to a higher ratio of real variation to errors). To calculate π for each population sample, we adopted an unbiased estimator of nucleotide diversity ($\hat{\theta}_{\Pi}$) based on heterozygosity (H), which has been optimized for pool-seq data (Ferretti et al. 2013). Numerically,

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

Here, L represents the total number of genome-wide

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

To examine patterns of polymorphism across chromosome arms, we also estimated window nucleotide diversity (π_w) for all polymorphic sites. Each window was defined as a continuous genomic region that includes 125,000 analyzed sites (Fig. 2). Since chromosomal identity was required in this analysis, we only took windows from 32 major contigs that contain at least one full-size window and were unambiguously mappable to a chromosome arm of the *D. melanogaster* dm6 genome assembly. Although such contigs only make up 57% of the *D. sukuzii* genome assembly, they contain a relatively larger proportion of all identified SNPs (72%) and thus are still representative of genome-wide polymorphism.

To estimate genome-wide pairwise F_{ST} between populations, we adopted an unbiased multi-loci estimator known as Reynolds' estimator of the co-ancestry coefficient, which accounts for unequal sample sizes among populations and is applicable for more than two alleles at a site (Reynolds et al. 1983). Below, we detail our usage of this common F_{ST} estimator in a genomic context. We first heuristically partitioned the genome into windows that exceeded a cross-sample average accumulated heterozygosity threshold of 100. The window-specific value of the above estimator, denoted here as $F_{ST}(w)$, was calculated as a weighted average of single-site ratio estimators. Numerically,

$$F_{ST}(w) = \left(\sum_l a_l \right) / \sum_l (a_l + b_l) \quad (2)$$

where:

$$a_l = \frac{1}{2} \sum_u (\tilde{p}_{1lu} - \tilde{p}_{2lu})^2 - \frac{(n_{1l} + n_{2l})(n_{1l}\tilde{\alpha}_{1l} + n_{2l}\tilde{\alpha}_{2l})}{4n_{1l}n_{2l}(n_{1l} + n_{2l} - 1)} \quad (3)$$

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

following Reynolds et al. (1983). Above, at the l th site in each population, \tilde{p}_{1lu} and \tilde{p}_{2lu} represent the frequency of

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

$$n_{ij} = \sum_j^{n_c} j P_c(j|n_r, n_c) \quad (5)$$

where

$$P_c(j|n_r, n_c) = \frac{n_c! S(n_r, j)}{(n_c - j)! n_c^{n_r}} \quad (6)$$

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

The genome-wide value of the above F_{ST} estimator was then calculated as an average of window F_{ST} estimates, weighted by the number of analyzed sites within each window. We also estimated genome-wide pairwise D_{XY} as an absolute measure of population differentiation that is independent of levels of within-population diversity. It was calculated as pairwise differences per site between two populations, divided by L total analyzed sites (Nei 1987; Hahn 2018). Numerically,

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

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

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

Preparing Environmental Data

To generate environmental data for GEA, we selected a preliminary set of 26 candidate environmental variables representing geographic, climatic, and land cover-related factors (Fig. 3a) that may be relevant in the adaptation process of *D. sukuzii* based on prior knowledge (Kellermann et al. 2012; Bogaerts-Márquez et al. 2021). With R packages “raster” (v. 3.5.2) and “sp” (v. 1.4.6) (Bivand et al. 2013; Hijmans 2023), we retrieved environmental data of high spatial resolution (~ 100 km²) in batch for the sampling locations of our 29 populations from online databases WorldClim (Fick and Hijmans 2017) and Esri 2020 Land Cover (Karra et al. 2021). Annual mean values of monthly climatic variables, including mean wind speed, solar radiation, and water vapor pressure were derived by averaging across 12 months of data.

Due to the large number of statistical tests that would result from running GEA on all the environmental variables one by one, there is an increased difficulty in controlling rates of false discovery. Additionally, including multiple highly correlated variables in a model would lead to multicollinearity issues (Rellstab et al. 2015). To avoid these problems, we calculated a pairwise Pearson correlation matrix from values of environmental factors across sampled locations (Fig. 3a; supplementary table S5, Supplementary Material online), and then selected a subset of nine least correlated environmental variables for one-by-one GEA analyses (Fig. 3b). To avoid scale inconsistencies between estimated GEA statistics, the environmental differentiation of each population was calculated as the absolute difference between the environmental value of that population and the average across all populations, standardized by the standard deviation (de Villemereuil and Gaggiotti 2015). This standardized differentiation was then input to GEA (supplementary table S5, Supplementary Material online).

Environmental Association Analyses

To characterize the environmental adaptation of *D. sukuzii*, we scanned the whole genome for adaptive loci using the F_{ST} -based GEA method BayeScEnv (de Villemereuil and Gaggiotti 2015). We chose this specific approach over other GEA methods mainly because it allows for detecting patterns of allele frequency that are not linearly dependent on environmental factors (Rellstab et al. 2015; de Villemereuil and Gaggiotti 2015). It has also been reported to have a low false positive rate compared to other GEA approaches in the presence of hierarchical population structure such as the continental-scale patterns documented here (de Villemereuil et al. 2014; de Villemereuil and Gaggiotti 2015; Gautier 2015).

For each environmental variable, the association analyses tested the relationship between environmental and genetic differentiation among populations, for 5,752,156

genome-wide SNPs with a MAF higher than 5%. To adapt BayeScEnv to pool-seq data (e.g. Wiberg et al. 2021), we corrected the input allele count data based on the effective sample size estimated from Equation (5). To control for false positives, we chose stringent model parameters expected to yield extremely conservative results, setting the prior probability of non-neutral models as 0.02 (-pr_jump 0.02) and the prior probability of the competing environment-unrelated locus-specific model as 0.9 (-pr_pref 0.9). These parameters correspond to assumptions that genetic differentiation reflects the action of natural selection in just 2% of the genome, and the focal environmental variable is only expected to be involved at 10% of the non-neutral loci.

To make this GEA analysis computationally feasible with our large SNP set, while still analyzing all qualifying SNPs, we applied a split-run strategy: we subsampled SNPs across concatenated sequences of contigs within the autosomes and the X chromosome separately, and then ran subsamples with BayeScEnv in parallel. Since the null model of population structure is estimated separately in each run, we subsampled non-adjacent SNPs at a fixed interval to limit locus-specific biases in that estimation, where the length of the interval between jointly analyzed SNPs was equal to the total number of subsamples. With a targeted subsample/interval size of up to 10,000 SNPs, we divided the concatenated autosomal contigs into 490 subsamples (with actual subsample sizes of 9,982 to 9,983 SNPs), and the concatenated X-linked contigs into 87 subsamples (with actual subsample sizes of 9,893 to 9,894 SNPs). Hence, the first autosomal subsample contained SNP #1, SNP #491, and so on.

Convergence of each run was confirmed with the R package “CODA” (Plummer et al. 2006). Individual runs were then merged across autosomes and X chromosome to calculate the genome-wide q -value (q) of locally estimated posterior error probability (PEP) across all sites, where we targeted a false discovery rate (FDR) of 5% by setting the q threshold at 0.05 (Storey 2003; Muller et al. 2006). As with most GEA studies, these results reflect a separate correction for multiple testing for each environmental variable.

For downstream analyses, to remove redundancy due to linkage disequilibrium, we obtained a set of “thinned outliers” for each environmental variable, paring down closely linked outlier sites by only maintaining the site with the lowest q when they occurred within 20 kb of each other. To assess the relative levels of support for associations between SNPs and a given environmental variable, we ranked all candidate loci first by q and then by the estimated g parameter as a tiebreaker, which measures the sensitivity of a locus to environmental differentiation.

Identifying Candidate Genes

For each candidate SNP, the closest gene in each direction within a 200-exon flanking region that overlapped with the

SNP was considered to be associated with that variant, in order to encompass both potential coding and regulatory adaptation. To facilitate clear comparisons among environmental variables with different numbers of significant variants, we focused on the top 500 candidate genes that were linked to variants with the lowest significant q and highest g within each environmental variable (Supplementary table S7, Supplementary Material online).

GO Enrichment and Semantic Clustering

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

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

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Arnaud Estoup, 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, Martin Kapun for assistance with SNP calling, Pierre de Villemereuil for helping with our GEA analyses, and members of the Pool lab for helpful comments on this manuscript. The UW-Madison Center for High Throughput Computing provided computational assistance and resources for this work. This work was supported by the United States Department of Agriculture (USDA) Hatch program (grant WIS02005 to S.D.S., C.G., and J.E.P.); and by the National Institute for General Medical Sciences (NIGMS) at the National Institutes of Health (NIH) (grant number R35 GM13630 to J.E.P.).

Conflict of Interest

The authors have no conflicts of interest to declare.

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 [supplementary 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).

Literature Cited

- Adrion JR, Kousathanas A, Pascual M, Burrack HJ, Haddad NM, Bergland AO, Machado H, Sackton TB, Schlenke TA, Watada M, et al. *Drosophila suzukii*: the genetic footprint of a recent, worldwide invasion. *Mol Biol Evol.* 2014;31(12):3148–3163. <https://doi.org/10.1093/molbev/msu246>.
- Alloway PG, Dolph PJ. A role for the light-dependent phosphorylation of visual arrestin. *Proc Natl Acad Sci U S A.* 1999;96(11):6072–6077. <https://doi.org/10.1073/pnas.96.11.6072>.
- Asplen MK, Anfora G, Biondi A, Choi D-S, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD, et al. Invasion biology of spotted wing *Drosophila* (*Drosophila suzukii*): a global perspective and future priorities. *J Pest Sci.* 2015;88(3):469–494. <https://doi.org/10.1007/s10340-015-0681-z>.
- Atallah J, Teixeira L, Salazar R, Zaragoza G, Kopp A. The making of a pest: the evolution of a fruit-penetrating ovipositor in *Drosophila suzukii* and related species. *Proc Biol Sci.* 2014;281(1781):20132840. <https://doi.org/10.1098/rspb.2013.2840>.
- Avidor-Reiss T, Leroux MR. Shared and distinct mechanisms of compartmentalized and cytosolic ciliogenesis. *Curr Biol.* 2015;25(23):R1143–R1150. <https://doi.org/10.1016/j.cub.2015.11.001>.
- Avidor-Reiss T, Maer AM, Koundakjian E, Polyanovsky A, Keil T, Subramaniam S, Zuker CS. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell.* 2004;117(4):527–539. [https://doi.org/10.1016/S0092-8674\(04\)00412-X](https://doi.org/10.1016/S0092-8674(04)00412-X).
- Bächli G. TaxoDros—The Database on Taxonomy of Drosophilidae; 2016. [Accessed 2024 September 12]. <http://www.taxodros.uzh.ch/>.
- Barange M, Cheung WWL, Merino G, Perry RI. Modelling the potential impacts of climate change and human activities on the sustainability of marine resources. *Curr Opin Environ Sustain.* 2010;2(5-6):326–333. <https://doi.org/10.1016/j.cosust.2010.10.002>.
- Batie M, Druker J, D'Ignazio L, Rocha S. KDM2 family members are regulated by HIF-1 in hypoxia. *Cells.* 2017;6(1):8. <https://doi.org/10.3390/cells6010008>.
- Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature.* 1992;356(6369):519–520. <https://doi.org/10.1038/356519a0>.
- Behrman EL, Watson SS, O'Brien KR, Heschel MS, Schmidt PS. Seasonal variation in life history traits in two *Drosophila* species. *J Evol Biol.* 2015;28(9):1691–1704. <https://doi.org/10.1111/jeb.12690>.
- Bivand R, Pebesma EJ, Gómez-Rubio V. Applied spatial data analysis with R. 2nd ed. New York Heidelberg Dordrecht London: Springer; 2013.
- Bogaerts-Márquez M, Guirao-Rico S, Gautier M, González J. Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of *Drosophila melanogaster*. *Mol Ecol.* 2021;30(4):938–954. <https://doi.org/10.1111/mec.15783>.
- Boughdad A, Haddi K, El Bouazzati A, Nassiri A, Tahiri A, El Anbri C, Eddaya T, Zaid A, Biondi A. First record of the invasive spotted wing *Drosophila* infesting berry crops in Africa. *J Pest Sci.* 2021;94(2):261–271. <https://doi.org/10.1007/s10340-020-01280-0>.
- Boutros M, Ahringer J. The art and design of genetic screens: RNA interference. *Nat Rev Genet.* 2008;9(7):554–566. <https://doi.org/10.1038/nrg2364>.
- Calabria G, Mâca J, Bächli G, Serra L, Pascual M. First records of the potential pest species *Drosophila suzukii* (Diptera: Drosophilidae) in Europe. *J Appl Entomol.* 2012;136(1-2):139–147. <https://doi.org/10.1111/j.1439-0418.2010.01583.x>.
- Calboli FCF, Gilchrist GW, Partridge L. Different cell size and cell number contribution in two newly established and one ancient body size cline of *Drosophila subobscura*. *Evolution.* 2003;57(3):566–573. <https://doi.org/10.1111/j.0014-3820.2003.tb01548.x>.
- Campbell M, Ganetzky B. Extensive morphological divergence and rapid evolution of the larval neuromuscular junction in *Drosophila*. *Proc Natl Acad Sci U S A.* 2012;109(11):E648–E655. <https://doi.org/10.1073/pnas.1201176109>.
- Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 1993;34(4):1289–1303. <https://doi.org/10.1093/genetics/134.4.1289>.
- Chen S, Zhou Y, Chen Y, Gu J. Fastq: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Cini A, Ioriatti C, Anfora G. A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. *Bull Insectol.* 2012;65(1):149–160. <http://www.bulletinofinsectology.org/pdfarticles/vol65-2012-149-160cini.pdf>.
- Clark MS, Thorne MA, Purać J, Burns G, Hillyard G, Popović ZD, Grubor-Lajsić G, Worland MR. Surviving the cold: molecular analyses of insect cryoprotective dehydration in the Arctic springtail *Megaphorura arctica* (Tullberg). *BMC Genomics.* 2009;10:328. <https://doi.org/10.1186/1471-2164-10-328>.
- Clemente F, Gautier M, Vitalis R. Inferring sex-specific demographic history from SNP data. *PLoS Genet.* 2018;14(1):e1007191. <https://doi.org/10.1371/journal.pgen.1007191>.
- Colautti RI, Barrett SCH. Rapid adaptation to climate facilitates range expansion of an invasive plant. *Science.* 2013;342(6156):364–366. <https://doi.org/10.1126/science.1242121>.
- Colautti RI, Lau JA. Contemporary evolution during invasion. *Invasion genetics.* Hoboken, New Jersey, U.S.: John Wiley & Sons, Ltd; 2016. p. 101–121.
- de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE. Genome scan methods against more complex models: when and how much should we trust them? *Mol Ecol.* 2014;23(8):2006–2019. <https://doi.org/10.1111/mec.12705>.
- de Villemereuil P, Gaggiotti OE. A new F_{ST} -based method to uncover local adaptation using environmental variables. *Methods Ecol Evol.* 2015;6(11):1248–1258. <https://doi.org/10.1111/2041-210X.12418>.
- Duvall LB, Taghert PH. E and M circadian pacemaker neurons use different PDF receptor signalosome components in *Drosophila*. *J Biol Rhythms.* 2013;28(4):239–248. <https://doi.org/10.1177/0748730413497179>.
- Elaine Merrill C, Sherertz TM, Walker WB, Zwiebel LJ. Odorant-specific requirements for arrestin function in *Drosophila* olfaction. *J Neurobiol.* 2005;63(1):15–28. <https://doi.org/10.1002/neu.20113>.
- Enriquez T, Colinet H. Cold acclimation triggers lipidomic and metabolic adjustments in the spotted wing drosophila *Drosophila suzukii* (Matsumara). *Am J Physiol Regul Integr Comp Physiol.* 2019;316(6):R751–R763. <https://doi.org/10.1152/ajpregu.00370.2018>.

- Ferretti L, Ramos-Onsins SE, Pérez-Enciso M. Population genomics from pool sequencing. *Mol Ecol*. 2013;22(22):5561–5576. <https://doi.org/10.1111/mec.12522>.
- Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol*. 2017;37(12):4302–4315. <https://doi.org/10.1002/joc.5086>.
- Fisher RA. Statistical methods for research workers. Edinburgh: Oliver and Boyd; 1938.
- Fraimout A, Debat V, Fellous S, Hufbauer RA, Foucaud J, Pudlo P, Marin J-M, Price DK, Cattel J, Chen X, et al. Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. *Mol Biol Evol*. 2017;34(4):980–996. <https://doi.org/10.1093/molbev/msx050>.
- François O, Martins H, Caye K, Schoville SD. Controlling false discoveries in genome scans for selection. *Mol Ecol*. 2016;25(2):454–469. <https://doi.org/10.1111/mec.13513>.
- Gautier M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*. 2015;201(4):1555–1579. <https://doi.org/10.1534/genetics.115.181453>.
- Gautier M. Efficient k-mer based curation of raw sequence data: application in *Drosophila suzukii*. *Peer Community J*. 2023;3. <https://doi.org/10.24072/pcjournal.309>.
- Gautier M, Foucaud J, Gharbi K, Cézard T, Galan M, Loiseau A, Thomson M, Pudlo P, Kerdelhué C, Estoup A. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol*. 2013;22(14):3766–3779. <https://doi.org/10.1111/mec.12360>.
- Gibbs AG, Fukuzato F, Matzkin LM. Evolution of water conservation mechanisms in *Drosophila*. *J Exp Biol*. 2003;206(7):1183–1192. <https://doi.org/10.1242/jeb.00233>.
- Gibert P, Debat V, Ghalambor CK. Phenotypic plasticity, global change, and the speed of adaptive evolution. *Curr Opin Insect Sci*. 2019;35:34–40. <https://doi.org/10.1016/j.cois.2019.06.007>.
- Gilchrist AS, Partridge L. A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drosophila melanogaster*. *Genetics*. 1999;153(4):1775–1787. <https://doi.org/10.1093/genetics/153.4.1775>.
- Guha A, Kornberg TB. Tracheal branch repopulation precedes induction of the *Drosophila* dorsal air sac primordium. *Dev Biol*. 2005;287(1):192–200. <https://doi.org/10.1016/j.ydbio.2005.09.005>.
- Hahn MW. Molecular population genetics. Sunderland (MA): Sinauer Associates; 2018.
- Hämälä T, Gorton AJ, Moeller DA, Tiffin P. Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*). *PLoS Genet*. 2020;16(3):e1008707. <https://doi.org/10.1371/journal.pgen.1008707>.
- Han Z, Li X, Wu J, Olson EN. A myocardin-related transcription factor regulates activity of serum response factor in *Drosophila*. *Proc Natl Acad Sci U S A*. 2004;101(34):12567–12572. <https://doi.org/10.1073/pnas.0405085101>.
- Harvey JA, Tougeron K, Gols R, Heinen R, Abarca M, Abram PK, Basset Y, Berg M, Boggs C, Brodeur J, et al. Scientists' warning on climate change and insects. *Ecol Monogr*. 2023;93(1):e1553. <https://doi.org/10.1002/ecm.1553>.
- Hauser M. A historic account of the invasion of *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae) in the continental United States, with remarks on their identification. *Pest Manag Sci*. 2011;67(11):1352–1357. <https://doi.org/10.1002/ps.2265>.
- Helfrich-Förster C, Bertolini E, Menegazzi P. Flies as models for circadian clock adaptation to environmental challenges. *Eur J Neurosci*. 2020;51(1):166–181. <https://doi.org/10.1111/ejn.14180>.
- Hijmans R. Raster: geographic data analysis and modeling. R package version 3.5.2; 2023. [Accessed 2024 September 12]. <https://rspatial.org/raster>.
- Hill T, Koseva BS, Unckless RL. The genome of *Drosophila innubila* reveals lineage-specific patterns of selection in immune genes. *Mol Biol Evol*. 2019;36(7):1405–1417. <https://doi.org/10.1093/molbev/msz059>.
- Hodgins KA, Bock DG, Rieseberg LH. Trait evolution in invasive species. In: Annual plant reviews online. New York: John Wiley & Sons, Ltd; 2018. p. 459–496.
- Honjo K, Mauthner SE, Wang Y, Skene JHP, Tracey WD. Nociceptor-enriched genes required for normal thermal nociception. *Cell Rep*. 2016;16(2):295–303. <https://doi.org/10.1016/j.celrep.2016.06.003>.
- Horváth V, Guirao-Rico S, Salces-Ortiz J, Rech GE, Green L, Aprea E, Rodeghiero M, Anfora G, González J. Gene expression differences consistent with water loss reduction underlie desiccation tolerance of natural *Drosophila* populations. *BMC Biol*. 2023;21(1):35. <https://doi.org/10.1186/s12915-023-01530-4>.
- James RE, Broihier HT. Crimpy inhibits the BMP homolog Gbb in motoneurons to enable proper growth control at the *Drosophila* neuromuscular junction. *Development*. 2011;138(15):3273–3286. <https://doi.org/10.1242/dev.066142>.
- John AO, Sylvester AA, Kehinde AO, Michael AA. Land use impacts on diversity and abundance of insect species. In: Vegetation dynamics, changing ecosystems and human responsibility. Rijeka: IntechOpen; 2022.
- Kanzawa T. Studies on *Drosophila suzukii* mats. Kofu, Japan: Yamanashi Agricultural Experimental Station; 1939.
- Kapun M, Barrón MG, Staubach F, Obbard DJ, Wiberg RAW, Vieira J, Goubert C, Rota-Stabelli O, Kankare M, Bogaerts-Márquez M, et al. Genomic analysis of European *Drosophila melanogaster* populations reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses. *Mol Biol Evol*. 2020;37(9):2661–2678. <https://doi.org/10.1093/molbev/msaa120>.
- Karra K, Kontgis C, Statman-Weil Z, Mazzariello JC, Mathis M, Brumbyp SP. Global land use/land cover with Sentinel 2 and deep learning. In: IEEE international geoscience and remote sensing symposium IGARSS. Brussels, Belgium: IEEE; 2021. p. 4704–4707.
- Kawecki TJ, Ebert D. Conceptual issues in local adaptation. *Ecol Lett*. 2004;7(12):1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>.
- Kellermann V, Loeschcke V, Hoffmann AA, Kristensen TN, Fløjgaard C, David JR, Svenning J-C, Overgaard J. Phylogenetic constraints in key functional traits behind species' climate niches: patterns of desiccation and cold resistance across 95 *Drosophila* species. *Evolution*. 2012;66(11):3377–3389. <https://doi.org/10.1111/j.1558-5646.2012.01685.x>.
- Kerekes É, Kókai E, Páldy FS, Dombrádi V. Functional analysis of the glycogen binding subunit CG9238/Gbs-70E of protein phosphatase 1 in *Drosophila melanogaster*. *Insect Biochem Mol Biol*. 2014;49:70–79. <https://doi.org/10.1016/j.ibmb.2014.04.002>.
- Kinsler G, Geiler-Samerotte K, Petrov DA. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*. 2020;9:e61271. <https://doi.org/10.7554/eLife.61271>.
- Knapp L, Mazzi D, Finger R. The economic impact of *Drosophila suzukii*: perceived costs and revenue losses of Swiss cherry, plum and grape growers. *Pest Manag Sci*. 2021;77(2):978–1000. <https://doi.org/10.1002/ps.6110>.
- Koch JB, Dupuis JR, Jardeleza M-K, Ouedraogo N, Geib SM, Follett PA, Price DK. Population genomic and phenotype diversity of invasive *Drosophila suzukii* in Hawai'i. *Biol Invasions*. 2020;22(5):1753–1770. <https://doi.org/10.1007/s10530-020-02217-5>.
- Kwadha CA, Okwaro LA, Kleman I, Rehmann G, Revadi S, Ndlela S, Khamis FM, Nderitu PW, Kasina M, George MK, et al. Detection of

- the spotted wing drosophila, *Drosophila suzukii*, in continental sub-Saharan Africa. *J Pest Sci.* 2021;94(2):251–259. <https://doi.org/10.1007/s10340-021-01330-1>.
- Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. A thousand fly genomes: an expanded *Drosophila* genome nexus. *Mol Biol Evol.* 2016a;33(12):3308–3313. <https://doi.org/10.1093/molbev/msw195>.
- Lack JB, Yassin A, Sprengelmeyer QD, Johanning EJ, David JR, Pool JE. Life history evolution and cellular mechanisms associated with increased size in high-altitude *Drosophila*. *Ecol Evol.* 2016b;6(16):5893–5906. <https://doi.org/10.1002/cece.3.2327>.
- Lange JD, Bastide H, Lack JB, Pool JE. A population genomic assessment of three decades of evolution in a natural *Drosophila* population. *Mol Biol Evol.* 2022;39(2):msab368. <https://doi.org/10.1093/molbev/msab368>.
- Lange JD, Pool JE. Impacts of recurrent hitchhiking on divergence and demographic inference in *Drosophila*. *Genome Biol Evol.* 2018;10(8):1882–1891. <https://doi.org/10.1093/gbe/evy142>.
- Langley CH, Crepeau M, Cardeno C, Corbett-Detig R, Stevens K. Circumventing heterozygosity: sequencing the amplified genome of a single haploid *Drosophila melanogaster* embryo. *Genetics.* 2011;188(2):239–246. <https://doi.org/10.1534/genetics.111.127530>.
- Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczowski B, et al. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics.* 2012;192(2):533–598. <https://doi.org/10.1534/genetics.112.142018>.
- Lee CE. Evolutionary genetics of invasive species. *Trends Ecol Evol.* 2002;17(8):386–391. [https://doi.org/10.1016/S0169-5347\(02\)02554-5](https://doi.org/10.1016/S0169-5347(02)02554-5).
- Levashina EA, Ohresser S, Bulet P, Reichhart J-M, Hetru C, Hoffmann JA. Metchnikowin, a novel immune-inducible proline-rich peptide from *Drosophila* with antibacterial and antifungal properties. *Eur J Biochem.* 1995;233(2):694–700. https://doi.org/10.1111/j.1432-1033.1995.694_2.x.
- Lewald KM, Abrieux A, Wilson DA, Lee Y, Conner WR, Andrezza F, Beers EH, Burrack HJ, Daane KM, Diepenbrock L, et al. Population genomics of *Drosophila suzukii* reveal longitudinal population structure and signals of migrations in and out of the continental United States. *G3 (Bethesda).* 2021;11(12):jkab343. <https://doi.org/10.1093/g3journal/jkab343>.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, arXiv, <http://arxiv.org/abs/1303.3997>, preprint: not peer reviewed.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Li W, Wang F, Menut L, Gao F-B. BTB/POZ-zinc finger protein abrupt suppresses dendritic branching in a neuronal subtype-specific and dosage-dependent manner. *Neuron.* 2004;43(6):823–834. <https://doi.org/10.1016/j.neuron.2004.08.040>.
- Little CM, Chapman TW, Hillier NK. Plasticity is key to success of *Drosophila suzukii* (Diptera: Drosophilidae) invasion. *J Insect Sci.* 2020;20(3):5. <https://doi.org/10.1093/jisesa/ieaa034>.
- Lotterhos KE, Whitlock MC. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol.* 2015;24(5):1031–1046. <https://doi.org/10.1111/mec.13100>.
- Lynch M, Conery JS. The origins of genome complexity. *Science.* 2003;302(5649):1401–1404. <https://doi.org/10.1126/science.1089370>.
- Mallard F, Nolte V, Tobler R, Kapun M, Schlötterer C. A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. *Genome Biol.* 2018;19(1):119. <https://doi.org/10.1186/s13059-018-1503-4>.
- Mérel V, Gibert P, Buch I, Rodriguez RV, Estoup A, Gautier M, Fablet M, Boulesteix M, Vieira C. The worldwide invasion of *Drosophila suzukii* is accompanied by a large increase of transposable element load and a small number of putatively adaptive insertions. *Mol Biol Evol.* 2021;38(10):4252–4267. <https://doi.org/10.1093/molbev/msab155>.
- Muller P, Parmigiani G, Rice K. FDR and Bayesian multiple comparisons rules. *JHU Biostatistics.* 2006:115. <https://doi.org/10.1093/oso/9780199214655.003.0014>.
- Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.
- Olazuaga L, Loiseau A, Parrinello H, Paris M, Fraimout A, Guedot C, Diepenbrock LM, Kenis M, Zhang J, Chen X, et al. A whole-genome scan for association with invasion success in the fruit fly *Drosophila suzukii* using contrasts of allele frequencies corrected for population structure. *Mol Biol Evol.* 2020;37(8):2369–2385. <https://doi.org/10.1093/molbev/msaa098>.
- Ometto L, Cestaro A, Ramasamy S, Grassi A, Revadi S, Siozios S, Moretto M, Fontana P, Varotto C, Pisani D, et al. Linking genomics and ecology to investigate the complex evolution of an invasive *Drosophila* pest. *Genome Biol Evol.* 2013;5(4):745–757. <https://doi.org/10.1093/gbe/evt034>.
- Orozco-terWengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlötterer C. Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol Ecol.* 2012;21(20):4931–4941. <https://doi.org/10.1111/j.1365-294X.2012.05673.x>.
- Palacios-Muñoz A, Ewer J. Calcium and cAMP directly modulate the speed of the *Drosophila* circadian clock. *PLoS Genet.* 2018;14(6):e1007433. <https://doi.org/10.1371/journal.pgen.1007433>.
- Paris M, Boyer R, Jaenichen R, Wolf J, Karageorgi M, Green J, Cagnon M, Parrinello H, Estoup A, Gautier M, et al. Near-chromosome level genome assembly of the fruit pest *Drosophila suzukii* using long-read sequencing. *Sci Rep.* 2020;10(1):11227. <https://doi.org/10.1038/s41598-020-67373-z>.
- Peng FT. On some species of *Drosophila* from China. *Annot Zool Jap.* 1937;16:20–27.
- Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News.* 2006;6(1):7–11.
- Pool JE. The mosaic ancestry of the *Drosophila* genetic reference panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. *Mol Biol Evol.* 2015;32(12):3236–3251. <https://doi.org/10.1093/molbev/msv194>.
- Pool JE, Braun DT, Lack JB. Parallel evolution of cold tolerance within *Drosophila melanogaster*. *Mol Biol Evol.* 2017;34(2):349–360. <https://doi.org/10.1093/molbev/msw232>.
- Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchon P, Emerson JJ, Saelao P, Begun DJ, et al. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet.* 2012;8(12):e1003080. <https://doi.org/10.1371/journal.pgen.1003080>.
- Pool JE, Nielsen R. Population size changes reshape genomic patterns of diversity. *Evolution.* 2007;61(12):3001–3006. <https://doi.org/10.1111/j.1558-5646.2007.00238.x>.
- Prakash S, Caldwell JC, Eberl DF, Clandinin TR. *Drosophila* N-cadherin mediates an attractive interaction between photoreceptor axons and their targets. *Nat Neurosci.* 2005;8(4):443–450. <https://doi.org/10.1038/nn1415>.
- Prelig G. Gene overexpression: uses, mechanisms, and interpretation. *Genetics.* 2012;190(3):841–854. <https://doi.org/10.1534/genetics.111.136911>.

- Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ. Adaptive evolution in invasive species. *Trends Plant Sci.* 2008;13(6):288–294. <https://doi.org/10.1016/j.tplants.2008.03.004>.
- Reijnders MJMF, Waterhouse RM. Summary visualizations of gene ontology terms with GO-figure! *Front Bioinform.* 2021;1:6. <https://doi.org/10.3389/fbinf.2021.638255>.
- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. A practical guide to environmental association analysis in landscape genomics. *Mol Ecol.* 2015;24(17):4348–4370. <https://doi.org/10.1111/mec.13322>.
- Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics.* 1983;105(3):767–779. <https://doi.org/10.1093/genetics/105.3.767>.
- Rezende GL, Martins AJ, Gentile C, Farnesi LC, Pelajo-Machado M, Peixoto AA, Valle D. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. *BMC Dev Biol.* 2008;8:82. <https://doi.org/10.1186/1471-213X-8-82>.
- Reznick DN, Losos J, Travis J. From low to high gear: there has been a paradigm shift in our understanding of evolution. *Ecol Lett.* 2019;22(2):233–244. <https://doi.org/10.1111/ele.13189>.
- Rogers WA, Salomone JR, Tacy DJ, Camino EM, Davis KA, Rebeiz M, Williams TM. Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genet.* 2013;9(8):e1003740. <https://doi.org/10.1371/journal.pgen.1003740>.
- Rossi Stacconi V. *Drosophila sukukii* (spotted wing drosophila). CABI Compendium; 2022. p. 109283.
- Rudman SM, Greenblum SI, Rajpurohit S, Betancourt NJ, Hanna J, Tilk S, Yokoyama T, Petrov DA, Schmidt P. Direct observation of adaptive tracking on ecological time scales in *Drosophila*. *Science.* 2022;375(6586):eabj7484. <https://doi.org/10.1126/science.abj7484>.
- Senthilan PR, Piepenbrock D, Ovezmyradov G, Nadrowski B, Bechstedt S, Pauls S, Winkler M, Möbius W, Howard J, Göpfert MC. *Drosophila* auditory organ genes and genetic hearing defects. *Cell.* 2012;150(5):1042–1054. <https://doi.org/10.1016/j.cell.2012.06.043>.
- Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR–Cas9. *Nat Rev Genet.* 2015;16(5):299–311. <https://doi.org/10.1038/nrg3899>.
- Shearer PW, West JD, Walton VM, Brown PH, Svetec N, Chiu JC. Seasonal cues induce phenotypic plasticity of *Drosophila sukukii* to enhance winter survival. *BMC Ecol.* 2016;16(1):11. <https://doi.org/10.1186/s12898-016-0070-3>.
- Simoës da Silva CJ, Sospedra I, Aparicio R, Busturia A. The microRNA-306/abrupt regulatory axis controls wing and haltere growth in *Drosophila*. *Mech Dev.* 2019;158:103555. <https://doi.org/10.1016/j.mod.2019.103555>.
- Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974;23(1):23–35. <https://doi.org/10.1017/S0016672300014634>.
- Stern DL. Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizyosity test. *Trends Genet.* 2014;30(12):547–554. <https://doi.org/10.1016/j.tig.2014.09.006>.
- Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann Stat.* 2003;31(6):2013–2035. <https://doi.org/10.1214/aos/1074290335>.
- Sugimura K, Satoh D, Estes P, Crews S, Uemura T. Development of morphological diversity of dendrites in *Drosophila* by the BTB-zinc finger protein abrupt. *Neuron.* 2004;43(6):809–822. <https://doi.org/10.1016/j.neuron.2004.08.016>.
- Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Curr Biol.* 2022;32(1):111–123.e5. <https://doi.org/10.1016/j.cub.2021.10.052>.
- Tan CC, Hsu TC, Sheng TC. Known *Drosophila* species in China with descriptions of twelve new species. *Univ Texas Publ.* 1949;4929:196–206.
- Telonis-Scott M, Gane M, DeGaris S, Sgrò CM, Hoffmann AA. High resolution mapping of candidate alleles for desiccation resistance in *Drosophila melanogaster* under selection. *Mol Biol Evol.* 2012;29(5):1335–1351. <https://doi.org/10.1093/molbev/msr294>.
- Tochen S, Dalton DT, Wiman N, Hamm C, Shearer PW, Walton VM. Temperature-related development and population parameters for *Drosophila sukukii* (Diptera: Drosophilidae) on cherry and blueberry. *Environ Entomol.* 2014;43(2):501–510. <https://doi.org/10.1603/EN13200>.
- True JR, Mercer JM, Laurie CC. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics.* 1996;142(2):507–523. <https://doi.org/10.1093/genetics/142.2.507>.
- Turner TL. Fine-mapping natural alleles: quantitative complementation to the rescue. *Mol Ecol.* 2014;23(10):2377–2382. <https://doi.org/10.1111/mec.12719>.
- Ueno K, Kidokoro Y. Adenylyl cyclase encoded by AC78C participates in sugar perception in *Drosophila melanogaster*. *Eur J Neurosci.* 2008;28(10):1956–1966. <https://doi.org/10.1111/j.1460-9568.2008.06507.x>.
- Uhrer J, Redlich S, Zhang J, Hothorn T, Tobisch C, Ewald J, Thorn S, Seibold S, Mitesser O, Morinière J, et al. Relationship of insect biomass and richness with land use along a climate gradient. *Nat Commun.* 2021;12(1):5946. <https://doi.org/10.1038/s41467-021-26181-3>.
- Van der Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol, CA: O'Reilly Media; 2020.
- Wallingford AK, Rice KB, Leskey TC, Loeb GM. Overwintering behavior of *Drosophila sukukii*, and potential springtime diets for egg maturation. *Environ Entomol.* 2018;47(5):1266–1273. <https://doi.org/10.1093/ee/nvy115>.
- Wang Y, Farine J-P, Yang Y, Tang W, Gehring N, Ferveur J-F, Moussian B. Transcriptional control of quality differences in the lipid-based cuticle barrier in *Drosophila sukukii* and *Drosophila melanogaster*. *Front Genet.* 2020b;11:887. <https://doi.org/10.3389/fgene.2020.00887>.
- Wang Y, Ferveur J-F, Moussian B. Eco-genetics of desiccation resistance in *Drosophila*. *Biol Rev Camb Philos Soc.* 2021;96(4):1421–1440. <https://doi.org/10.1111/brv.12709>.
- Wang T, Morency DT, Harris N, Davis GW. Epigenetic signaling in glia controls presynaptic homeostatic plasticity. *Neuron.* 2020a;105(3):491–505.e3. <https://doi.org/10.1016/j.neuron.2019.10.041>.
- Wang S-H, Simcox A, Campbell G. Dual role for *Drosophila* epidermal growth factor receptor signaling in early wing disc development. *Genes Dev.* 2000;14(18):2271–2276. <https://doi.org/10.1101/gad.827000>.
- Welles SR, Dlugosch KM. Population genomics of colonization and invasion. In: Rajora OP, editor. Population genomics: concepts, approaches and applications. Population genomics. Cham: Springer International Publishing; 2019. p. 655–683.
- Wiberg RAW, Tyukmaeva V, Hoikkala A, Ritchie MG, Kankare M. Cold adaptation drives population genomic divergence in the ecological specialist, *Drosophila montana*. *Mol Ecol.* 2021;30(15):3783–3796. <https://doi.org/10.1111/mec.16003>.
- Wong MM-K, Liu M-F, Chiu SK. Cropped, *Drosophila* transcription factor AP-4, controls tracheal terminal branching and cell growth. *BMC Devel Biol.* 2015;15(1):20. <https://doi.org/10.1186/s12861-015-0069-6>.

Associate editor: Shu-Dan Yeh