# New insights into past and future invasion of *Drosophila suzukii* using novel genomic resources and statistical methods to combine individual and pool sequencing data

Louise Camus[1,*], Nicolas Rode[1], Svitlana Serga[1,2], Anne Loiseau[1], Xiao Chen[3], Carole Iampietro[4], Marc Kenis[5], William Marande[6], Julián Mensch[7], Hugues Parinello[8], Marija Savić Veselinović[9], Sophie Valière[4], Jinping Zhang[10], Arnaud Estoup[1], Simon Boitard[1,†,*] and Mathieu Gautier[1,†,*]

[1] CBGP, Univ Montpellier, CIRAD, INRAE, IRD, Institut Agro Montpellier, Montpellier, France

[2] National Antarctic Scientific Center of Ukraine, Taras Shevchenko blv. 16, Kyiv, 01601, Ukraine

[3] College of Plant Protection, Yunnan Agricultural University, Kunming, 650201 China

[4] INRAE, GeT-PlaGe, Genotoul, 31326 Castanet-Tolosan, France

[5] CABI, Delémont, Switzerland

[6] French Plant Genomic Resource Center, CNRGV, INRAE, Castanet Tolosan, France

[7] Departamento de Ecología, Genética y Evolución. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires, Ciudad Universitaria, Pabellón II (C1428 EHA), Buenos Aires, Argentina

[8] MGX-Montpellier GenomiX, Univ Montpellier, CNRS, INSERM, Montpellier France

[9] University of Belgrade, Faculty of Biology, Studentski trg 16, 11000 Belgrade, Serbia

[10] MARA-CABI Joint Laboratory for Bio-safety, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China

[†]These authors are joint senior authors on this work

[*]Corresponding authors: louise.camus@inrae.fr, simon.boitard@inrae.fr and mathieu.gautier@inrae.fr

**Keywords**: *D. suzukii*, Biological invasions, Genomic Offset, Local Adaptation, GEA, BayPass

**Running title**: The genomics of past and future invasion in *D. suzukii*

# Abstract

Climate change and globalization are accelerating biological invasions, highlighting the urgent need to understand adaptation mechanisms in invaded environments to improve management strategies. Genomic data can provide insights into the adaptation of invasive species through Genotype-Environment Association (GEA) studies, identifying genes and biological processes associated with invasion success, and more globally to estimate genetic (mal)adaptation to new environments by calculating Genomic Offset (GO) statistics. In this study, we investigate genetic adaptation in the invasive crop pest *Drosophila suzukii* using novel genomic resources and statistical methods. We rely on a new chromosome-level genome assembly and a dataset representing 36 populations, combining both publicly available and newly generated pooled and individual sequencing data, which are analyzed using an enhanced version of the BayPass software, tailored for such hybrid datasets. We identify a limited number of genomic regions associated with invasion success, supporting the hypothesis of a polygenic architecture of the underlying adaptive traits. Using a GEA incorporating 28 environmental covariates, we further estimate GO between source environments and invaded areas to gain insight into the adaptive challenges faced by *D. suzukii* during past invasions. Reciprocally, we estimate GO between putative source environments and geographic areas that have not yet been invaded to predict regions at risk for potential future invasions. Finally, we used GO calculation to identify high risk regions from which pre-adapted populations could likely originate. Our results suggest that the adaptive challenge for *D. suzukii* populations to invade their current major invasion range was limited. We also identified uninvaded regions (in Africa, South America, and Australia) as being at high risk of future invasion. While further sampling and more extensive individual sequencing could refine these conclusions, our study provides important insights into *D. suzukii* adaptation and offers a generic and operational population genomics framework for studying and predicting biological invasions that can be applied to diverse species.

# Introduction

Climate change and globalization are accelerating biological invasions (Mainka and Howard, 2010; Hulme, 2017, 2021), underscoring the need to understand adaptation mechanisms in invaded environments and to predict future invasions for better management. Indeed, invasive species can have significant negative impacts on biodiversity (Mollot *et al.*, 2017), economy (Haubrock *et al.*, 2021; Bradshaw *et al.*, 2016), human health (Mazza *et al.*, 2014), and food security (Bruce, 2010; Paini *et al.*, 2016), with total reported costs recently estimated at more than one trillion US dollars over the past few decades (Diagne *et al.*, 2021). From an evolutionary perspective, the spread of invasive species raises questions about how populations are able to rapidly colonize new environments and the extent to which their genetic makeup is critical. One hypothesis suggests that invasive success may be due to various forms of pre-adaptation, enabled by standing genetic variation in the native range that allows exploitation of a wide range of environments, including anthropized ones (Estoup *et al.*, 2016; Hufbauer *et al.*, 2012). Alternatively, invasive success may be due to one or more *de novo* advantageous mutations (Colautti and Barrett, 2013), although this possibility can be considered less likely, primarily due to the short evolutionary time frame of biological invasions which reduces the likelihood of new mutations arising (Barrett and Schluter, 2008).

With the advent of next-generation genotyping and sequencing approaches, genome-environment association (GEA) studies have proven informative for unraveling the genetic mechanisms involved in local adaptation, including to invaded environments. Indeed, by identifying genomic regions whose variation covaries with environmental factors (Yang *et al.*, 2022; Hofmeister *et al.*, 2021; Ma *et al.*, 2020) or with phenotypic traits (Turner *et al.*, 2021; Olazcuaga *et al.*, 2020; Pfenninger *et al.*, 2021), they can provide insights into the biological processes that enable genetic adaptation. More generally, since GEA models summarize the relationship between the (adaptive) genomic composition of populations and their environment, they can be used to predict maladaptation to novel environments. Fitzpatrick and Keller (2015) introduced the genomic offset (GO) statistic to quantify the gap between the optimal genetic configuration for a new environment (as predicted by the GEA model) and the actual genetic makeup of a population, which can thus be interpreted as a potential maladaptation risk (Capblancq *et al.*, 2020). Several studies have confirmed the correlation between GO and average population fitness, using both computer simulations and common garden experiments (Gain *et al.*, 2023; Láruson *et al.*, 2020; Lotterhos, 2023; Rhoné *et al.*, 2020; Fitzpatrick *et al.*, 2021). Nevertheless, GO remains underutilized in invasion biology, although it shows promise as a valuable tool for predicting the success of invasive populations, as recently demonstrated in an extensive simulation study showing its correlation with the probability of establishment (Camus *et al.*, 2024). Further, the concept of GO, which implies that populations adapt to new environments through pre-existing variants, is consistent with the hypothesis of (genetic) pre-adaptation to invaded environments. Following this idea, calculating GO between source and invaded environments

can reveal which geographic areas have posed the greatest adaptive challenges to any native source population. Besides, GO can also be informative in predicting the future spread of an already or potentially invasive species by allowing the identification of areas posing the lowest adaptive challenges to native or already invasive populations.

Here, we propose to illustrate and test these ideas, and further develop methods regarding the genetic characterization of invasive species, using the *Drosophila suzukii* species as an empirical biological model. This species, also known as the spotted wing drosophila, provides a particularly vivid case study of a harmful biological invasion. Native to East Asia, it was first recorded in North America (California) and Southern Europe (Spain and Italy) in 2008 (Asplen *et al.*, 2015), with studies indicating independent invasion events on these continents (Fraimout *et al.*, 2017; Adrion *et al.*, 2014). By 2013, *D. suzukii* had spread to South America (Deprá *et al.*, 2014) and is now widespread throughout the Americas and Europe. The species has also been detected in various parts of Africa, such as Algeria (Aouari *et al.*, 2022), Morocco (Boughdad *et al.*, 2021), Kenya (Kwadha *et al.*, 2021), and most recently in South Africa (IPCC, 2024). Due to its ability to lay eggs in ripening fruit, *D. suzukii* has caused significant damage to crops, particularly berry crops (e.g. cherries or strawberries) in the various areas where it has invaded, making it a major agricultural pest (Walsh *et al.*, 2011).

To date, few studies have comprehensively investigated the adaptation of *D. suzukii* to invaded environments. To our knowledge, only one study by Feng *et al.* (2023) examined *D. suzukii* environmental adaptation using GEA, identifying genes associated with environmental variables such as temperature and precipitation. Another notable study by Olazcuaga *et al.* (2020) examined the broad genetic basis of *D. suzukii* invasion success and found evidence for only a few genes associated with population invasion status. Both studies were limited by their reliance on a limited number of pool-sequenced (Pool-Seq) population samples and a somewhat fragmented genome assembly derived from sequencing a strain originating from the North American invaded area (Paris *et al.*, 2020). As *D. suzukii* research expands, large-scale whole-genome sequencing (WGS) efforts that include individual (Ind-Seq) samples (Lewald *et al.*, 2021) in addition to the aforementioned Pool-Seq data have become publicly available; the heterogeneity of data types and origins however pose practical challenges for their combined analysis.

To deepen our understanding of the genetic basis of past *D. suzukii* invasion and to further identify regions at risk for future invasion, we present in this study the combined analyses of 27 pool-seq and 82 ind-Seq samples, together representing a total of 36 population samples. These data are primarily sourced from carefully curated publicly available datasets, with the addition of newly sequenced samples from four previously unsampled areas. To enable efficient integration of these heterogeneous data for GEA and other genome-scan approaches, we extended the Bayesian hierarchical models implemented in the BayPass software (Gautier, 2015) for the joint analysis of hybrid

data sets consisting here of read counts (from Pool-Seq experiments) and the likelihoods for all possible genotypes of the Ind-Seq individuals at the different markers. The newly introduced modeling of genotype likelihood (GL) in BayPass allows integrating over genotype uncertainties, and thus properly deals with low to mid-coverage Ind-Seq data. In addition, we present a new improved chromosome-level assembly for the genome of *D. suzukii*, derived from an isofemale strain originating from Japan and thus representative of the native range of the species. Based on these new datasets, genomic resources, and methodological developments, we first searched for key genes and their possible underlying biological processes that may have been involved in the adaptive process during invasion. To this end, and following Olazcuaga *et al.* (2020), we relied on the $C_2$ statistic to contrast the genomic composition of populations from the native and invaded areas, enhancing its power by exploiting linkage disequilibrium (LD) information using the local-score approach (Fariello *et al.*, 2017). Second, to further characterize the *D. suzukii* invasion and provide a global picture of its genetic underpinnings, we use GO statistics based on a GEA with 28 covariables characterizing the environment of the 36 studied populations to address three main purposes consisting of: i) identifying which parts of the invaded areas of Europe and North America may have posed the greatest adaptive challenge to *D. suzukii*; ii) predicting which parts of the geographic areas that have not yet been (fully) invaded (e.g. in South America and Africa) are at a higher risk of being invaded in the future; and iii) anticipating which region could host the highest-risk source population for a putative invasive *D. suzukii* population establishing in Australia where this pest is currently absent. These latter applications aims at illustrating how GO statistics may represent an additional potentially valuable tool to be considered in Integrated Pest Management (IPM) practice. Overall, by providing a comprehensive view of the adaptive history of *D. suzukii*, we seek to describe a generic and operational population genomics framework for studying and predicting biological invasions that can be applied to a wide range of species.

# Results

## A novel male chromosome-level genome assembly of a Japanese isofemale line

To date, all the available reference assemblies for the *D. suzukii* genome were obtained from isofemale lines originating from the European (Ometto *et al.*, 2013) or US (Chiu *et al.*, 2013; Paris *et al.*, 2020) invaded areas. These also only consisted of female assemblies, which remained suboptimal both in terms of contiguity and completeness. To circumvent these limitations, we generated a male chromosome level assembly for a Japanese isofemale line, i.e. originating from the native area, using PacBio HiFi and high-throughput chromosome conformation capture (HiC) sequencing data. The final curated contig assembly consisted of 75 contigs totaling 282,289,391 bp (N50=10.6 Mb

and N90=1.79Mb) and including 99.06% (98.48% complete and single) of the 3,285 BUSCO genes of the `diptera_odb10` dataset (Manni *et al.*, 2021). Based on the ratio of coverages observed in resequencing data of two male-only and female-only pools (see Material and Methods), eight and three contigs could be unambiguously assigned to the X (39.14 Mb in total) and Y (16.47 Mb) chromosomes, respectively. As detailed in Figure 1A, HiC data further allowed to generate a highly contiguous scaffolded assembly, hereafter referred to as *dsu_isojap1.0*, consisting of 54 scaffolds with N50=32.7 Mb and N90=2.97Mb (Table S1).

More specifically, seven out of the eight X–linked contigs were ordered into a single 32.72 Mb scaffold, the remaining eighth contig (6.316 Mb) was renamed *scf_Xc* as it was likely more centromeric (see below). For the Y chromosome, the two smallest contigs could not be scaffolded with the largest one (13.21 Mb), but were joined into an 3.338 Mb Y-linked scaffold (*scf_Y1*) based on the alignment of the Occludin-Related Y (*Ory*) *D. melanogaster* gene. Further, alignment of the *dsu_isojap1.0* assemblies with the chromosome-level reference assembly `Dbia2.1` of the closely related species *Drosophila biarmipes* (Figure 1B) showed that 5 large scaffolds of *dsu_isojap1.0* corresponded to the 5 autosome arms (2L, 2R, 3L, 3R and 4) of *D. biarmipes*; these scaffolds were thus named following *D. melanogaster* (and *D. biarmipes*) nomenclature as justified by cytogenetics studies (Drosopoulou *et al.*, 2019). Note that this comparison reflected the well known syntenic conservation of chromosomal arms (aka as Muller elements) across *Drosophila* species (e.g., Bhutkar *et al.*, 2008), since rearrangements among the two species were almost exclusively within these arms. It also suggested that the sequenced chromosome arms in *dsu_isojap1.0* are close to complete and highly accurate (see also Table 1 below). Moreover, the 3L and 3R arms could be scaffolded into a single large scaffold presumably including the whole centromeric region (see below).

Accordingly and as detailed in Table 1, a total of 3,197 genes of the `diptera_odb10` dataset were found complete and in single copy in both the *dsu_isojap1.0* and the *Dmel6* assembly for the *D. melanogaster* genome (Hoskins *et al.*, 2015). The overlap in BUSCO gene contents between the syntenic chromosome arms of the two species was almost complete with only a few discrepancies, which aligns with the very close but not perfect conservation of Muller elements among *Drosophila* species (Bhutkar *et al.*, 2008). For instance, a 99.4% overlap was found for the X chromosome: out of the 509 BUSCO genes mapping to the X chromosomes of either species, 506 were mapped to the X chromosome in both assemblies. Likewise, the overlap was equal to 99.9% and 100% for chromosomes 3 and 4 respectively. The overlap was a bit lower for chromosome arms 2L and 2R (95.8% and 96.0%) respectively, suggesting a less complete sequence. However, an additional large scaffold (40.07 Mb) containing only 19 BUSCO genes, 8 mapping to the 2L and 11 to the 2R *D. melanogaster* chromosome arms, likely corresponded to the (extended) centromeric regions of chromosome 2 (as confirmed by their repeat content, see below). This scaffold was thus named *2c*. Yet, out of the 1,212 (2.31%) BUSCO genes mapped to both *dsu_isojap1.0* and *Dmel6* assemblies

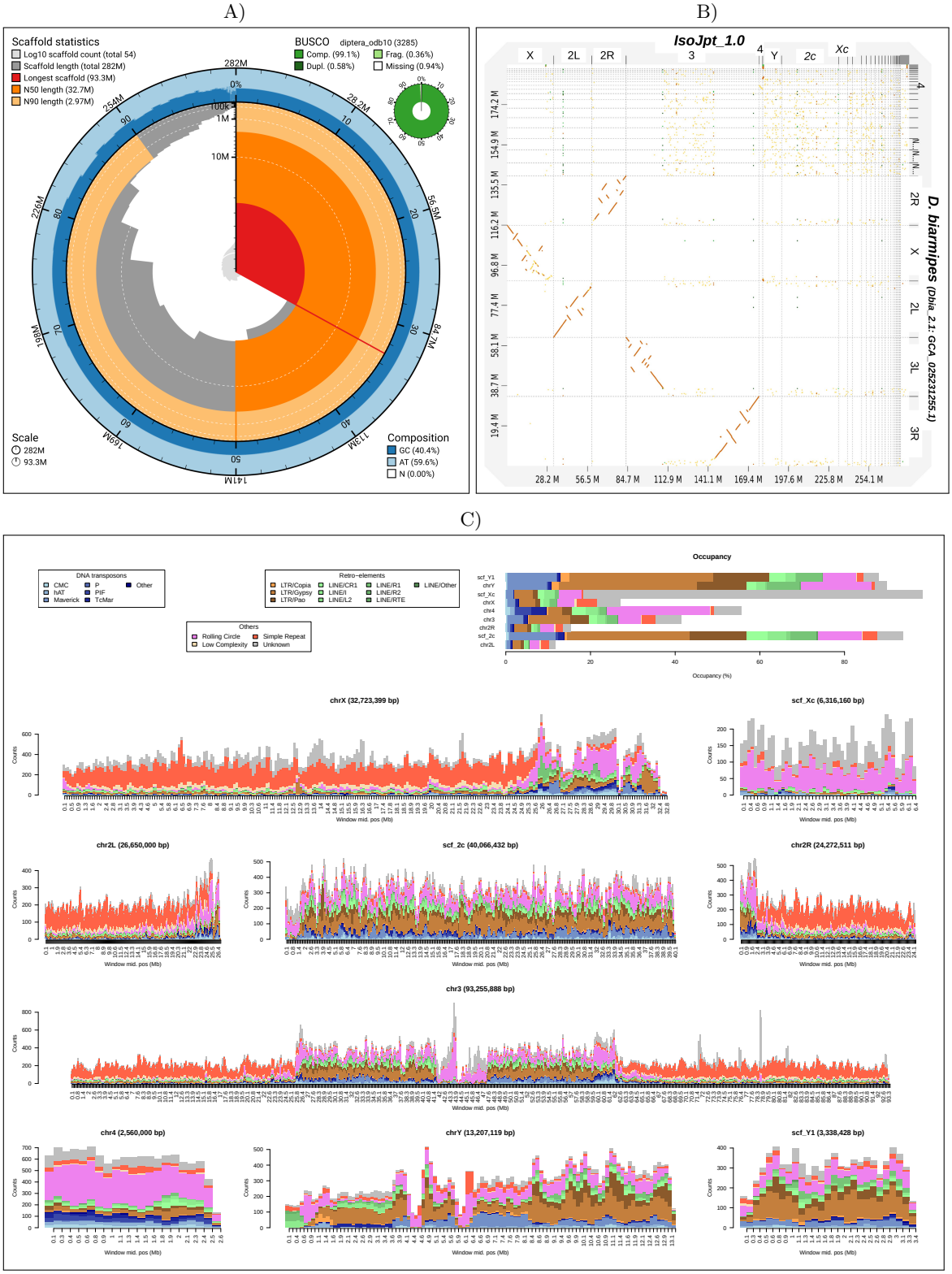


Figure 1: **Statistics of the novel *dsu_isojap1.0* male chromosome-level assembly of the *Drosophila suzukii* Japanese isofemale line.** A) Snailplot built using the `BlobToolKit v4.3.5` viewer (Challis *et al.*, 2020) based on the results from `BUSCO v5.4.7` for the `diptera_odb10` dataset that includes 3,285 genes (Manni *et al.*, 2021). B) Dotplot generated with `dgenies v1.5.0` (Cabanettes and Klopp, 2018) (default options) comparing the *dsu_isojap1.0* with the `Dbia2.1` reference assembly (GenBank ID: GCF_025231255.1) of the closely related species *Drosophila biarmipes*. C. Distribution of the number of copies of different families of repetitive elements over the assembly chromosomes per non overlapping 250 kb windows. The upper-right gives the overall occupancy per chromosome (in % of their length).

and located on the *D. melanogaster* 2L and 2R, 28 were mapped to other unscaffolded contigs of *dsu_isojap1.0*. In contrast, no BUSCO genes from *D. melanogaster* chromosomal arms other than 2L and 2R mapped to unscaffolded contigs of *dsu_isojap1.0*. Conversely, no BUSCO genes were found on the Y (and `Xc`) chromosomes of *dsu_isojap1.0*.

| *Dmel6* / *dsu_isojap1.0* | X (23.54 Mb) | 2L (23.51 Mb) | 2R (25.29 Mb) | 3L (28.11 Mb) | 3R (32.08 Mb) | 4 (1.348 Mb) | Total |
|---|---|---|---|---|---|---|---|
| X (32.72 Mb) | 506 | 1 | - | 1 | - | - | 508 |
| 2L (26.65 Mb) | - | 541 | 1 | - | - | - | 542 |
| 2R (24.72 Mb) | - | - | 622 | - | - | - | 622 |
| 3 (93.26 Mb) | 1 | - | - | 660 | 804 | - | 1465 |
| 4 (2.560 Mb) | - | - | - | - | - | 13 | 13 |
| Y (13.21 Mb) | - | - | - | - | - | - | - |
| *2c* (40.07 Mb) | - | 8 | 11 | - | - | - | 19 |
| *scf_Xc* (6.316 Mb) | - | - | - | - | - | - | - |
| *scf_Y1* (3.338 Mb) | - | - | - | - | - | - | - |
| other (n=12) | - | 14 | 14 | - | - | - | 28 |
| Total | 507 | 564 | 648 | 661 | 804 | 13 | 3197 |

Table 1: Comparison of BUSCO gene content between the *dsu_isojap1.0* assembly and the *Dmel6* assembly for the *D. melanogaster* genome (Hoskins *et al.*, 2015). In total, 3,197 genes from the `diptera_odb10` dataset that were found to be complete and in single copy in both assemblies. The table provides the number of these genes co-localizing on the chromosomes (or scaffolds) of the two assemblies.

We further assessed the repeat contents of the *dsu_isojap1.0* assembly using the curated and annotated TE database recently developed by Mérel *et al.* (2021) for the *D. suzukii* genome. In agreement with studies on the WT3_v2.0 assembly (Paris *et al.*, 2020; Mérel *et al.*, 2021, e.g.), we found that *dsu_isojap1.0* was characterized by a very high TE content (150.3 Mb i.e. 53.24% of the assembly) with a predominant occupancy of the long terminal repeat (LTR) order (20.9%) followed by LINEs (8.32%), DNA transposons (6.66%) and rolling circles (RC, 6.59%). As shown in Figure 1C, the repeat content varied among the assigned scaffolds. Repetitive sequences occupy almost all chrY (89.8%), scf_Y1 (87.6%), scf_2c (93.7%) and scf_Xc (98.5%) scaffolds and a high proportion of chr4 (53.7%). Conversely, the 2L (11.2%), 2R (14.8%) and chrX (25.7%) displayed the lowest proportion, the full chr3 scaffold an intermediary value of 40.4%. More interestingly, the count distribution of the different TE types over the scaffolds highlighted distinct chromosomal regions (Figure 1C). In particular, the center of chr3, one of the two extremities of 2L, X and 2R and the whole chr4, chrY, scf_Y1, scf_2c (93.7%) and scf_Xc scaffolds lacked simple repeats and were enriched in LTR (*Gypsy* and to a lesser extent *Copia* families) and RC. Aligning with Chang *et al.* (2019) who found islands of retroelements in *D. melanogaster* centromeres regions, this suggests that this observed TE signature specify centromeric regions in *D. suzukii* genome, although different TE families are involved. From a more practical point of view, such TE signature empirically allowed to orient the 2L, the 2R and the X scaffolds and support the scaffolding accuracy and chromosome-level nature of the *dsu_isojap1.0* assembly.

Overall, and as summarized in Table S1, the *dsu_isojap1.0* assembly represents a clear improvement

over the currently available assemblies (as of September 2024), both in terms of contiguity and completeness. The *dsu_isojap1.0* assembly was also found highly similar to the recently released `RU_1.0` chromosome level but partial (i.e., lacking ca. 100 Mb of mostly TE-rich and gene poor regions) assembly obtained from an US isofemale line, although three large-scale inversions could be identified on the X (from position 17.94 to 21.88 Mb), the 2R (from position 7.413 to 17.59 Mb) and 3R (from position 83.51 to 87.75 Mb) chromosomes (Figure S1). More precisely, 40 blocks of sequence similarity $> 1Mb$ totaling 171.5 Mb (i.e. 96.1% of the `RU_1.0` assembly) could be found with only 1.33% of nucleotide divergence. Likewise, when compared to the more fragmented `WT3-2.0` assembly, 35 blocks of sequence similarity $> 1Mb$ and totaling 153.6 Mb were found with a nucleotide divergence of 0.813%. Finally, 16,692 (96.0% of the 17,391 ones) and 16,018 (99.2% of the 16,148 ones) from the `WT3-2.0` and `RU_1.0` NCBI annotation files, respectively, could be found on *dsu_isojap1.0* using `liftoff` (Shumate and Salzberg, 2021).

## A curated genomic dataset combining Ind-Seq and Pool-Seq WGS representative of the worldwide genetic diversity of *D. suzukii*

As detailed in the Materials and Methods section (M&M), publicly available WGS data including both Ind-Seq (Lewald *et al.*, 2021) and Pool-Seq (Olazcuaga *et al.*, 2020; Feng *et al.*, 2023) samples, were carefully curated and combined to provide an optimal and balanced representation of the worldwide genetic diversity of *D. suzukii* (Tables S2 and S3). In addition, to further improve this representation, we newly sequenced i) 9 individuals previously collected in Chengdu (China) in 2014 (Table S3); ii) three pools of individuals sampled in Zajača (Serbia, n=100) in 2020, in Delemont (Switzerland, n=58) in 2014, and in Golondrinas (Argentina, n=100) in 2023, respectively (Table S2); iii) one pool of 100 females sampled in Jena (Germany) in 2016 to replace the pool from the same area included in Olazcuaga *et al.* (2020) that was found highly contaminated with *D. immigrans* individuals (Gautier, 2023). This resulted in a final dataset comprising genomic data for 36 populations representative of both the native area (n=9), and the two European (n=11) and American (n=16) invaded areas (Figure 2A, Table S2 and Table S3). The combined analysis of all Pool-Seq (n=27) and Ind-Seq (n=82 representing 9 populations) data allowed identifying and genotyping a total of 3,634,933 autosomal and 691,982 X-chromosome bi-allelic variants (mostly corresponding to SNPs, see M&M) on the new *dsu_isojap1.0* genome assembly.

To provide a global visualization of the genetic structuring of the populations, we first performed a random allele PCA on the whole sample and focusing on autosomal variants. As shown in Figure 2B and in agreement with previous studies (e.g. Fraimout *et al.*, 2017; Lewald *et al.*, 2021), the resulting first factorial plan of the random allele PCA grouped the 36 populations into four main clusters that could be associated with i) the native area; ii) the European invaded area; iii) populations from Northwest US and Hawaii; and iv) all other populations from the American invaded area (including

the newly generated AR-Gol from Argentina). A more detailed inspection showed that within the native area cluster (i.e., within Asia), Japanese populations tended to cluster together, as did Chinese populations.

Within the European cluster, which includes FR-Run from La Réunion Island in the Indian Ocean due to its recent European origin (Fraimout *et al.*, 2017), PC1 could only loosely be associated with a north-south geographical gradient. In fact, the northernmost IR-Dub (sampled in Ireland) was closer to the Mediterranean populations (ES-Bar, IT-Tre and FR-Cor), while PC1 (and PC2) tended to isolate the newly generated sample from Serbia (RS-Zaj), which may be related either to its easternmost geographic positions or to a more recent sampling date. Conversely, the positioning of the new CH-Del close to DE-Dos was in good agreement with its geographic origin. Finally, FR-Lez (from southern France) was more closely related to German (DE-Jen and DE-Dos) than to other Mediterranean populations, and FR-Par (sampled in Paris) remained relatively distant from all other populations. Overall, the positioning of populations from the European invaded area could not be clearly related to their geographic (or temporal) origin, which may indicate some degree of migration across the European continent, and perhaps some heterogeneous introgression events with propagules from the North-American continent (Fraimout *et al.*, 2017).

For the American invaded area, the observed differentiation between eastern and northwestern American populations is consistent with previous work (Lewald *et al.*, 2021). The proximity of US-Haw, US-Wa, and US-Sok, which form a distinct cluster, is consistent with evidence for gene flow from Hawaii to the northwestern America (Fraimout *et al.*, 2017; Gautier *et al.*, 2022). Five eastern US populations clustered together (US-Vir, US-Me, US-Ga, US-Nca, and US-Ny), along with three central US populations (US-Mi, US-Col, and US-Wis). South Californian and South American populations also clustered with this group, but also more closely with each other, in agreement with the previously inferred origin of the South American populations (Gautier *et al.*, 2022). Interestingly, US-Wat was positioned between the Northwestern and Eastern American clusters, consistent with its intermediate geographic origin.

As an alternative representation of the genetic structure over the whole dataset, we further estimated the scaled covariance matrix $\Omega$ of the population allele frequencies (of autosomal variants) with the new version of BayPass (v3.0; Gautier, 2015), which allows joint modeling of Ind-Seq and Pool-Seq data (see M&M). The correlation plot of the resulting $\Omega$ matrix showed a pattern similar to that observed with the random allele PCA, with a general structuring into the same four distinct groups identified above based on the random allele PCA (Figure 2C). More specifically and reassuringly, this $\Omega$ representation did not lead to a grouping of samples according to their sequencing type (Pool-Seq or Ind-Seq), suggesting that the statistical models implemented in the new BayPass version perform as expected on heterogeneous datasets.

Both the random allele PCA and the $\Omega$ matrix estimation were independently repeated with X–

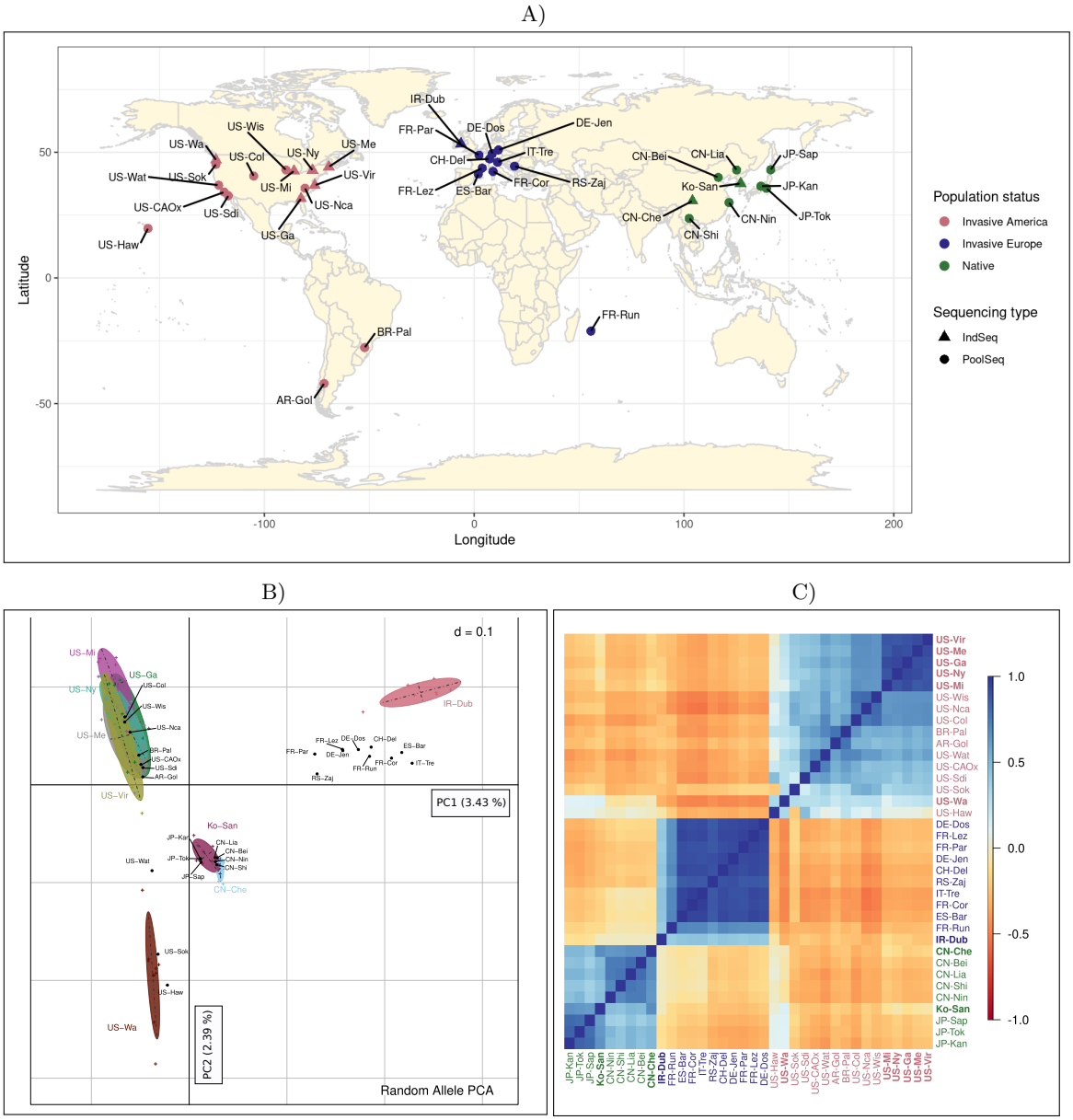


Figure 2: Description of the 36 studied *D. suzukii* population samples and structuring of genetic diversity on autosomal variants. A) Geographical location of the 36 *D. suzukii* populations. Populations from the native area are displayed in green, while invasive American and European populations are displayed in red and blue, respectively. B) Random allele PCA representation of the population samples. Pool-Seq samples are represented by black dots, while Ind-Seq samples are indicated by crosses colored according to their population of origin (with 95% ellipses summarizing the overall dispersion of the underlying individuals around their population center of gravity). C) Correlation plot of the scaled covariance matrix $\Omega$ estimated from the combined analysis of Pool-Seq and Ind-Seq data with the new version of BayPass software (v3.0; Gautier, 2015) on autosomal variants. The name of the populations represented by Ind-Seq samples are highlighted in bold.

linked variants. The results obtained were largely similar to those described above for autosomal variants (Figure S2). For the random allele PCA (Figure S2A), the only notable difference was that the two North and middle Californian populations did not cluster clearly with other American populations. Also, and as expected from higher overall differentiation on the X chromosome, the percentage of variance explained by the first two PC was about 40% higher (Figure S3). Regarding the estimated $\Omega$ matrix (Figure S2B), US-Wa and US-Haw were more differentiated and could not be clearly defined as part of the American cluster. They seemed to form a distinct cluster, closer to the native area. Another surprising result, regarding the $\Omega$ matrix for the X chromosome, was the correlation of allele frequencies between US-Sok and IR-Dub, positioning US-Sok closer to European populations. Despite these few differences, the overall clustering patterns remain consistent and robust, providing a clear description of population structure over the whole set of curated Pool-Seq and Ind-Seq samples.

## Genome scan for association with invasive status

In order to investigate the genetic basis of *D. suzukii* invasive success, we employed the BayPass $C_2$ statistic (Olazcuaga *et al.*, 2020), which identifies overly differentiated markers between two population groups while accounting for population structure. We estimated the $C_2$ statistic contrasting native versus invasive worldwide populations ($C_2^{WW}$), as well as native versus invasive European ($C_2^{EU}$) or American ($C_2^{AM}$) populations, respectively. As mentioned above, BayPass analyses were conducted using our new version (v3.0) of the software (Gautier, 2015), which supports the joint analyses of both read count data (e.g., from Pool-Seq experiment) and Ind-Seq data consisting of the genotype likelihood (GL) for all possible genotypes of the Ind-Seq individuals at the different markers. We further employed the local-score approach (Fariello *et al.*, 2017), which accounts for linkage disequilibrium by cumulating the scores (here related to the $C_2$ statistic) of neighboring SNPs from a same genomic region, to identify significant windows associated with invasive status based on the three different contrasts. Cumulating SNPs with a $C_2$ p-value below 0.01 (i.e. $\xi = 2$ in the local score, see M&M), we could identify 20 significant windows for $C_2^{WW}$, 13 windows for $C_2^{EU}$, and 5 windows for $C_2^{AM}$ (Figure 3), predominantly on chromosomes 2R (n=14), 3R (n=12) and 3L (n=7), only one window being found on chromosomes 2L and 4 and three on chromosome X.

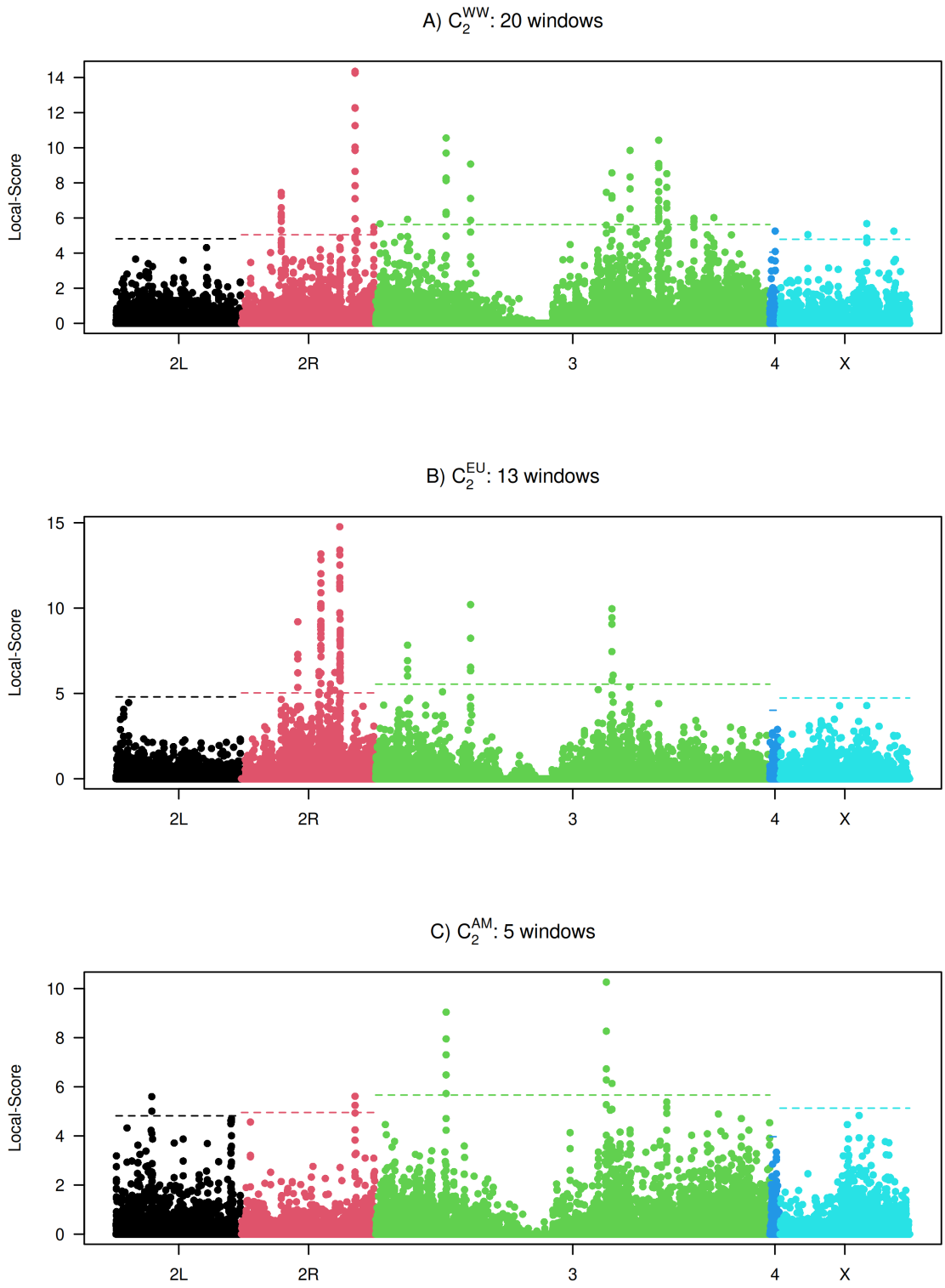


Figure 3: Manhattan plot of the Local-score based on $C_2$ p-values for the comparison between (from top to bottom) : native and invasive populations worldwide ($C_2^{WW}$), native and invasive European populations ($C_2^{EU}$), and native and invasive American populations ($C_2^{AM}$). The dashed lines indicate the chromosome-specific 1% significance threshold for the local score. No significant windows were found on the scf_2c and scf_Xc which are not plotted to improve readability. The number of significant local-score windows is indicated on each plot (see Table 2 for details).

Overall, the windows tended to be evenly distributed across the genome, with the notable exceptions

of the centromeric regions of chromosomes 2 (2C) and 3, both of which lacked significant signals, and conversely, chromosome 2R, which shows a concentration of significant windows for $C_2^{EU}$. As detailed in Table 2, the identified windows were small, ranging from 568 to 15,369 bp and including only 6 to 36 SNPs. Merging the windows that overlapped between different $C_2$ tests resulted in a final set of 32 candidate regions : 14 specific to $C_2^{WW}$, 10 specific to $C_2^{EU}$, 2 specific of $C_2^{AM}$, 3 in common between $C_2^{WW}$ and $C_2^{EU}$ and 3 in common between $C_2^{WW}$ and $C_2^{AM}$. Thus, no overlap was found between any of the 13 $C_2^{EU}$ and the 5 $C_2^{AM}$ windows. The regions that were detected only in the $C_2^{EU}$ test (10 regions) and only in the $C_2^{AM}$ test (2 regions) most likely correspond to continental specific signatures, otherwise they would also have been detected by the worldwide test. Indeed, $C_2^{WW}$ is expected to have more power due to the larger underlying sample size. However, our ability to assess the real convergence between adaptation signatures in Europe and North America is limited by the lower power of the $C_2^{EU}$ and $C_2^{AM}$ tests, which detected fewer significant regions than $C_2^{WW}$. For instance, the 14 $C_2^{WW}$–specific regions might result from the combination of signals that are consistent between continents but too weak to be detected in any of them. In agreement with this, several $C_2^{WW}$–specific regions included SNPs with high but non-significant $C_2^{EU}$ or $C_2^{AM}$ values (Figure S4). To investigate this pattern further, we repeated the local score analysis using a less stringent threshold, cumulating $C_2$ p-values below 0.1 (i.e. with $\xi = 1$) and found 8 significant regions shared by $C_2^{EU}$ and $C_2^{AM}$ (Table S4). Given the number and the size of the genomic regions detected by the two tests in this analysis, we would expect to find only 0.046 of such overlapping windows if the positions of these regions were chosen at random (see M&M). Thus, there is clear evidence that a few genomic regions have been selected convergently in the European and American continents, although these regions most likely represent a minor proportion of the genomic regions related to adaptation in the entire invaded area.

Gene annotation of the most significant windows (i.e., setting $\xi = 2$) highlighted 58 positional candidate genes (Table 2) among which 51 could be associated with a *D. melanogaster* ortholog. To investigate biological processes potentially related to invasive success, we performed a Gene Ontology enrichment analysis for Biological Process and Molecular Function on the list of these candidate genes (combining all three contrasts) but no significant enrichment signal could be detected.

| Chr. | Position in bp | Size in bp (Nmarkers) | $C_2$ | max. Lscore ($-\log_{10} \mathrm{P}_{C2}$) | Nb of genes: known *Dmel* orthologs |
|---|---|---|---|---|---|
| 2L | 7,392,604-7,393,484 | 880 (8) | $C_2^{AM}$ | 5.60 (5.07) | 1: CG15439 |
| | 6,776,116-6,779,675 | 3,559 (19) | $C_2^{WW}$ | 7.45 (5.03) | 1: Vps13 |
| | 9,873,930-9,876,158 | 2,228 (12) | $C_2^{EU}$ | 9.20 (4.60) | 1: Pcf11 |
| | 13,979,252-13,980,432 | 1,180 (14) | $C_2^{EU}$ | 6.27 (3.93) | 1: CG4554 |
| | 14,273,226-14,287,238 | 14,012 (36) | $C_2^{EU}$ | 13.2 (5.19) | 1: jbug |
| | 15,859,266-15,860,447 | 1,181 (10) | $C_2^{EU}$ | 5.55 (3.52) | 1: enok |
| | 15,868,534-15,869,278 | 744 (7) | $C_2^{EU}$ | 5.13 (5.11) | 1: gek |
| 2R | 16,494,070-16,499,464 | 5,394 (26) | $C_2^{EU}$ | 6.22 (4.65) | 5: CG8185-87-92, Vha36-1, eIF2B$\gamma$ |
| | 17,517,853-17,526,745 | 8,892 (29) | $C_2^{EU}$ | 14.8 (5.52) | 5: mim, CheB38b, ppk25 |
| | 17,580,698-17,596,067 | 15,369 (13) | $C_2^{EU}$ | 9.35 (5.64) | 5: Tsp42Ec-Ed, nclb, Taf5, Pex6 |
| | 17,596,280-17,602,111 | 5,831 (30) | $C_2^{EU}$ | 8.71 (4.29) | 1: Taf5 |
| | 20,371,155-20,383,674 | 12,519 (25) | $C_2^{WW}$ | 14.4 (5.41) | 1: CG15711 |
| | 20,371,155-20,376,124 | 4,969 (13) | $C_2^{AM}$ | 5.62 (4.53) | |
| | 20,804,560-20,805,667 | 1,107 (8) | $C_2^{WW}$ | 5.27 (7.11) | 1: mute |
| | 23,908,056-23,919,155 | 11,099 (10) | $C_2^{WW}$ | 5.49 (4.79) | 2: CG3829, zip |
| | 1,242,426-1,254,694 | 12,268 (11) | $C_2^{WW}$ | 5.67 (3.80) | 2: CG9205, bab1 |
| | 8,168,914-8,170,171 | 1,257 (9) | $C_2^{WW}$ | 5.93 (5.61) | 1: brv2 |
| | 8,168,914-8,170,197 | 1,283 (10) | $C_2^{EU}$ | 7.83 (6.04) | |
| 3(L) | 17,190,748-17,193,749 | 3,001 (13) | $C_2^{WW}$ | 10.6 (5.28) | 2: Pop2, CG6928 |
| | 17,190,748-17,193,662 | 2,914 (12) | $C_2^{AM}$ | 9.04 (4.60) | |
| | 22,093,311-22,106,186 | 12,875 (10) | $C_2^{WW}$ | 9.07 (5.79) | 3: CG7857, CG7272, CG7275 |
| | 22,093,311-22,108,254 | 14,943 (12) | $C_2^{EU}$ | 10.2 (6.27) | |
| | 68,326,018-68,330,780 | 4,762 (10) | $C_2^{WW}$ | 7.46 (4.80) | 1 |
| | 68,326,018-68,330,881 | 4,863 (13) | $C_2^{AM}$ | 10.3 (5.98) | |
| | 69,181,253-69,184,142 | 2,889 (12) | $C_2^{WW}$<br>$C_2^{EU}$ | 8.57 (6.87)<br>9.96 (7.05) | 3: BcDNA:RE41733, mdlc |
| | 69,194,278-69,206,486 | 12,208 (9) | $C_2^{AM}$ | 6.14 (5.58) | 2: CG31807 |
| | 69,404,501-69,409,533 | 5,032 (10) | $C_2^{EU}$ | 6.07 (4.75) | 1 |
| 3(R) | 70,603,616-70,606,546 | 2,930 (9) | $C_2^{WW}$ | 6.05 (6.41) | 2: Unc-115a, trbd |
| | 72,009,730-72,013,388 | 3,658 (15) | $C_2^{WW}$ | 9.85 (5.80) | 2: mthl11, cwo |
| | 75,619,062-75,627,177 | 8,115 (30) | $C_2^{WW}$ | 10.4 (5.86) | 4: CG43273, Kaz, RabX4, niki |
| | 76,781,158-76,785,234 | 4,076 (15) | $C_2^{WW}$ | 8.52 (5.32) | 2: CG3909 |
| | 81,357,185-81,360,108 | 2,923 (19) | $C_2^{WW}$ | 5.98 (3.31) | 1: Doa |
| | 84,087,208-84,088,556 | 1,348 (12) | $C_2^{WW}$ | 6.02 (4.29) | 1: Alg9 |
| 4 | 1,359,819-1,363,817 | 3,998 (9) | $C_2^{WW}$ | 5.25 (4.23) | 1 |
| | 8,823,167-8,825,627 | 2,460 (8) | $C_2^{WW}$ | 5.06 (4.68) | 1: sdk |
| X | 19,831,144-19,833,169 | 2,025 (9) | $C_2^{WW}$ | 5.68 (4.23) | 2: Flo2, Gld |
| | 25,685,679-25,686,247 | 568 (6) | $C_2^{WW}$ | 5.25 (7.25) | 1: CCKLR-17D1 |

Table 2: Description of the significant $C_2$ local-score windows surrounding a total of 58 candidate genes. The table gives for each window i) its position in the new *dsu_isojap1.0* assembly of the *D. suzukii* genome; ii) its size in pb and its number of markers; iii) the maximum Lindley score (Lscore) and the most significant $C_2$ Pvalues (in $-\log_{10}$ scale) observed in the window; and iv) the number of overlapping annotated genes and the names of the corresponding *D. melanogaster* orthologs when identified, separated by a comma.

Unsurprisingly, many of the detected genes are implicated in developmental pathways. Some others are related to specific functions that might favor the adaptation to environments, especially for those within genomic regions being found differentiated using the $C_2^{WW}$ contrast. For instance, *bab1* is involved in pigmentation in response to temperature (De Castro *et al.*, 2018), while *cwo* regulates circadian rhythm (Richier *et al.*, 2008; Chen *et al.*, 2014), two biological processes thought to be involved in the overwintering of the sister species *D. melanogaster* (Shearer *et al.*, 2016). The *nclb* gene may also play a role in response to environmental conditions, as it potentially contributes to

desiccation tolerance in natural *D. melanogaster* populations (Horváth *et al.*, 2023). Additionally, *CCKLR-17D1* gene is notable for its diverse roles, including adult locomotory behavior (Kiss *et al.*, 2013), stress-induced escape response in larvae (Chen *et al.*, 2012), and aggressive behavior in *D. melanogaster* male flies (Wu *et al.*, 2020; Agrawal *et al.*, 2020). Finally, the *Flo2* gene located on the X-chromosome has been identified as a region under positive selection (without clear function described) in European populations of *D. melanogaster* (Werzner *et al.*, 2013).

Interestingly, several other genes might be related to insecticides resistance. For instance, the *Vps13* gene, producing vacuolar protein sorting-associated protein 13, plays a role in the lysosomal degradation pathway in *D. melanogaster* and is likely implicated in the response to fluralaner, a common pesticide, in the pest insect *Spodoptera litura* (Jia *et al.*, 2020). In bumblebees, a gene producing an analog of Filamin-A protein, similarly to the *jbug* gene in *D. melanogaster*, and a glucose dehydrogenase [FAD, quinone]-like gene, similar to the *Gld* gene in *D. melanogaster*, are both implicated in the response to neonicotinoid insecticides, with the latter being strongly downregulated upon exposure (Erban *et al.*, 2019). In our study, the genes *Vps13* and *Gld* were located in genomic regions showing congruent local-score signals for all the three contrasts analyses (Figure S4, panels A and B).

Finally, several genes identified in our contrast analyses were found to have significant roles in lifespan or immunity related traits. For instance, the *jbug* gene is crucial for long-term memory and lifespan extension in *D. melanogaster* (Bolduc *et al.*, 2010; Landis *et al.*, 2003), while the *trbd* gene is involved in both lifespan and immune response (Fernando *et al.*, 2014). The *Vha36-1* gene also plays a vital role in immune defense against ingested bacteria (Vaibhvi *et al.*, 2022). Three of the genes mentioned above (*Vps13*,*cwo* and *trbd*) were found in regions shared by $C_2^{EU}$ and $C_2^{AM}$ analyses when using $\xi = 1$.

It is worth stressing that most of the genes mentioned above have been found in candidate regions where not any or very few other genes are located (Table 2). There is therefore good statistical evidence that these genes have been the drivers of the selection signal that has been detected in these regions.

## Insights into past invasions using Genomic Offset statistics

We performed a GEA analysis with the first ten PCs (explaining together >95% of the total variance) derived from 28 different environmental variables describing the environment of sampling sites of the 36 natural *D. suzukii* populations considered in this study. The GEA was based on the combined Pool-Seq and Ind-Seq samples using the new version of BayPass (v3.0; Gautier, 2015), and considering autosomal and X-chromosome datasets separately. This allowed us to estimate the effect of each of the ten environmental PCs on genomic variation, which corresponds to the PC regression coefficients on the scaled population allele frequencies of each variant. Such regression coefficients

have then be used to calculate the geometric Genomic Offset (gGO), which estimate the degree of (mal)adaption of a population adapted to a given environment when (re)located in a new one (Gain *et al.*, 2023; Camus *et al.*, 2024). Among the different GO estimators, we have chosen the gGO, recently introduced by (Gain *et al.*, 2023), due to its good and validated performance, in particular for the prediction of the probability of establishment of a population in a given (new) environment (Camus *et al.*, 2024).

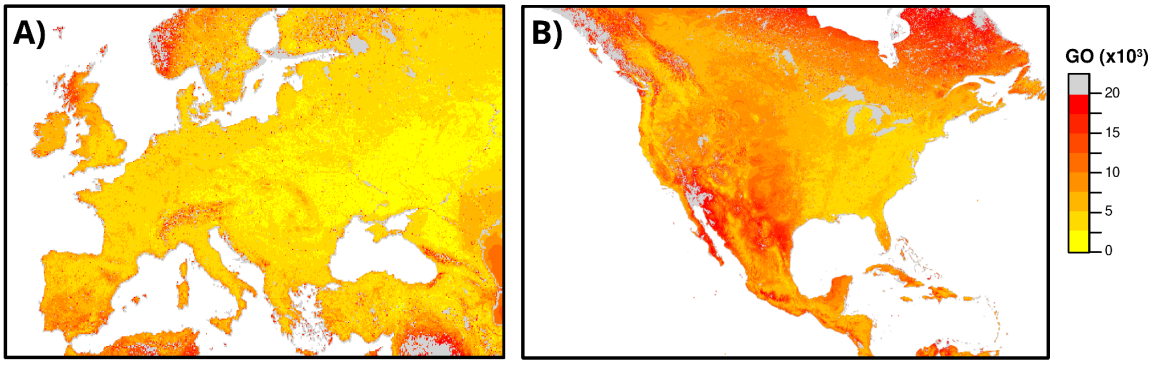


Figure 4: gGO values over European (A) and North American (B) invaded areas. For Europe, the reference environment corresponded to the one of the CN-Lia population, inferred as the closest proxy (see Figure S5). For North America, to account for the admixed origin of the North America invasive populations (Fraimout *et al.*, 2017; Gautier *et al.*, 2022), we computed weighted gGO values by combining the two GO maps (Figure S6) based on the environments of CN-Nin and US-Haw as references. The combined map used weights of 75% for CN-Nin and 25% for US-Haw, corresponding to their respective estimated admixture proportions in the primo-invading US population (Fraimout *et al.*, 2017). Lowest (highest) gGO values correspond to lowest (strongest) adaptive challenges. Gray pixels represent outlying gGO values (as defined in M&M).

To determine which parts of the invaded areas of Europe and North America posed the greatest adaptive challenge to *D. suzukii*, we computed GO maps over these two continents (Figure 4) by estimating, for each grid cell, the gGO values relative to a reference environment representative of the source populations. For the European continent (Figure 4A), for which a single original invasive source population has been previously inferred (Fraimout *et al.*, 2017), we relied on $F_3$ statistics (Patterson *et al.*, 2012; Gautier *et al.*, 2022) to identify CN-Lia as the closest source population proxy from our data (see Figure S5 for details). The environment of CN-Lia was then used as the reference environment to compute the GO map over Europe. The resulting estimated gGOs were generally very low, especially in the east (Figure 4A). Only a few areas, such as mountainous (e.g., the Alps) or cold regions in north-west Scotland and western Norway, showed high and sometimes extreme gGO values.

For North America, the determination of the reference source environment is more difficult due to the previously demonstrated admixed origin of North American invasive populations (Fraimout *et al.*, 2017; Gautier *et al.*, 2022). Extending the interpretation of GO as an environmental distance weighted by the adaptive genome composition (Gain *et al.*, 2023), we therefore chose to compute weighted gGO values combining the two GO maps (Figure S6) obtained when taking as reference

the environments of CN-Nin and US-Haw, the two main contributing source proxies (Fraimout *et al.*, 2017). The weights chosen were 75% (of gGO computed with CN-Nin) and 25% (of gGO computed with US-Haw), corresponding to the estimated admixture proportions for the primo-invading population in the US (Fraimout *et al.*, 2017). The resulting estimated gGOs were more heterogeneous across the North American subcontinent (Figure 4B) than in Europe, with high values found in the north (e.g., colder regions of northern Canada) and the south (e.g., arid regions of Mexico and east-southern USA). Only a few areas exhibited extreme gGO values, mostly in the norther most (cold) regions and the central part of the Sonoran Desert. The lowest gGO values were mostly found along some parts of the west coast and extreme south-eastern parts of the USA and in Central America. It should be noticed that our gGO weighting approach may only have had minor impact on our results since the GO maps computed separately with each of the environments associated with the two CN-Nin and US-Haw proxies were similar (Figure S6), although slightly higher values were found with CN-Nin in some areas, particularly in Central America and along a narrow coastal strip of the southeastern USA.

Taken together, these results suggest that the adaptive challenge for *D. suzukii* populations to invade two of their current major invasion areas (Europe and North America) was limited.

## Predicting future invasions using Genomic Offset statistics

We relied on gGO to predict which geographical regions in Africa, South America and Australia that have not been invaded to date are at high risk of invasion in the future. To do this, we followed a strategy similar to that described above, which consisted of calculating gGO over these continents with respect to a fixed source environment. For Africa and South America, we selected source environments from regions within each continent where *D. suzukii* populations have been found to be initially established and thus may be representative of actual source populations capable of spreading across these two continents. We hence selected Longonot Farm (Nakuru county, Kenya; Kwadha *et al.*, 2021) and Nova Veneza (Santa Catarina, Brazil; Deprá *et al.*, 2014) for Africa and South America, respectively. For Australia, where *D. suzukii* is not yet established, the environment associated with our Northwestern American sample US-Wa was used as a (putative) source reference based on trade statistics for fresh fruit exchanges (International Trade Center, 2024).

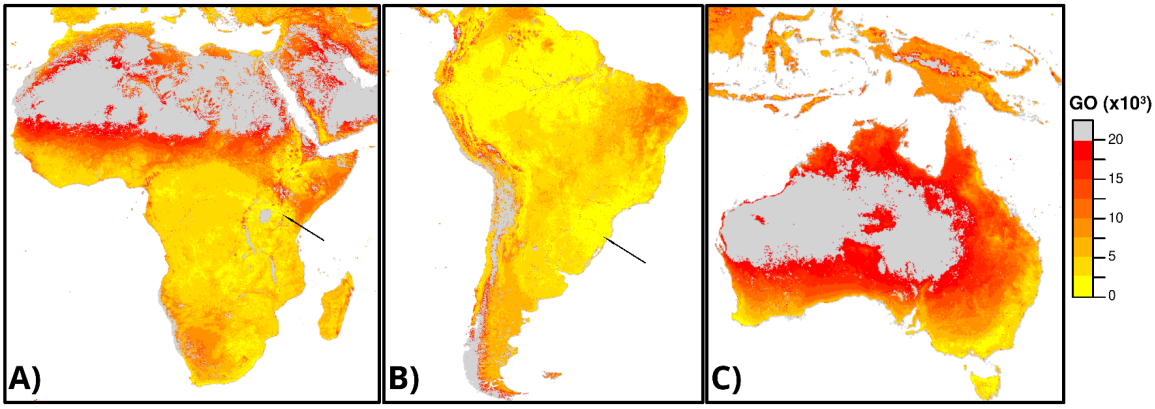


Figure 5: gGO estimates for the geographical regions in Africa (A), South America (B) and Australia (C) that have not been (fully) invaded to date. For Africa and South America, the black arrow indicates the (chosen) source reference environment where *D. suzukii* was first recorded and recently described as established in these continents (i.e. the environment of Longonot Farm - Nakuru county, Kenya - for Africa and of Nova Veneza - Santa Catarina, Brazil - for South America). For Australia, where *D. suzukii* has not yet been found as established, the reference environment associated with our Northwestern American sample US-Wa was used as a (putative) source reference based on trade statistics for fresh fruit exchanges (International Trade Center, 2024). Gray pixels represent outlying gGO values (as defined in M&M).

Estimated gGOs across Africa showed contrasting patterns, with arid regions (e.g., the Sahara and Namib deserts) showing predominantly outlying values and very high values at their borders, and other regions showing moderate values (e.g., in central Africa). The lowest gGOs values could be observed in various regions including in South Africa, the Great Lakes region, in central Cameroon or in the Madagascar highlands (Figure 5A). Interestingly, some of the lowest gGO values were located in the very northern part of the African continent, especially in northwestern Morocco where *D. suzukii* infestation has recently been reported in several farms growing small berries (Boughdad *et al.*, 2021). The north-eastern region of Algeria including pomegranate agroecosystem, where *D. suzukii* has been detected (Aouari *et al.*, 2022), also showed low gGO values but with a more heterogeneous pattern. If these results suggest an overall low risk for a widespread invasion of *D. suzukii* across the African continent, some of the regions with the lowest (or moderate) gGOs can be considered as having a higher risk of *D. suzukii* establishment and should be particularly monitored, as they may also coincide with important agricultural regions for fruit production.

In South America, the distribution of estimated gGO was less contrasted than in Africa, with most of the subcontinent showing low (e.g., the Amazon Basin, southeastern Brazil, or the eastern foothills of the Andes in northwestern Argentina and southern Bolivia) to moderate values. As expected, the high or extreme outlying values were mainly located in regions with extreme climates, such as the mountainous cordillera of the Andes, the west coast and southern parts of Patagonia, or the semi-arid region of the Sertão in northeastern Brazil (Figure 5B). It is worth noting that the Chubut region in central Patagonia (Argentina) where *D. suzukii* has only recently been detected and was considered as established (i.e., about ten years after the first observations of the species in Brazil) and from which our Argentinean AR-Gol sample has been collected, showed moderate to high gGO

values.

Finally, we found that the gGO values (with respect to the environment associated to the US-Wa sample) were very high over almost all Australia and extreme in desert areas (Figure 5C). The lowest gGO values were concentrated in the southern regions, particularly along the South East Coast, and in Tasmania. For areas free of *D. suzukii*, such as Australia, our gGO-based approach can be used to predict which regions of origin (i.e. source regions) are likely to pose the highest risk of providing the best pre-adapted *D. suzukii* populations to the environment of a new area. The aim is to use this information to improve the vigilance of IPM practices for fruit imports from these regions. As an illustration, the two largest exporters of fresh berries to Australia in 2023 (after New Zealand and Vietnam where *D. suzukii* is considered absent) are the USA, Thailand (International Trade Center, 2024). We hence estimated gGO over North America (Figure 6A) and Southeast Asia (Figure 6B) using Sydney Harbor as a reference target environment, an obvious potential sensitive introduction point of pest species in Australia.

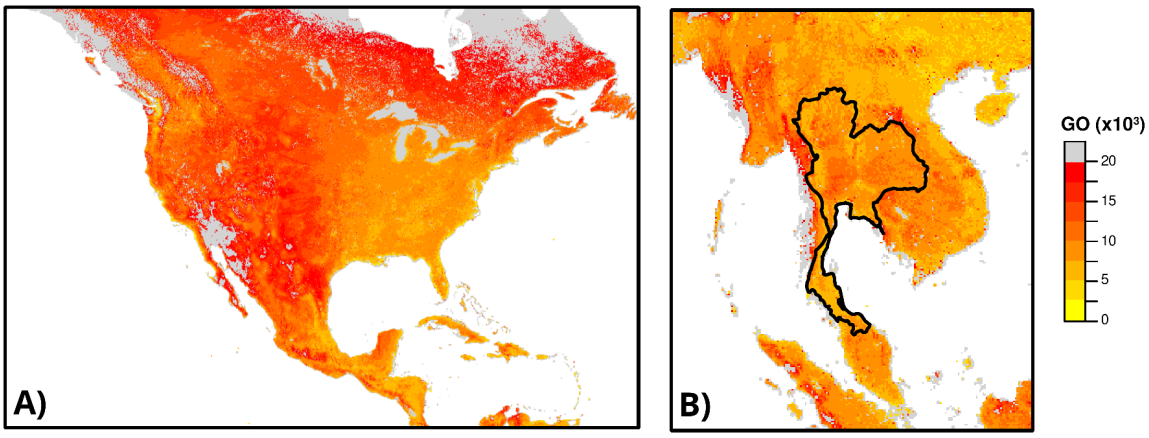


Figure 6: gGO map to assess the areas most at risk as sources of pre-adapted *D. suzukii* propagules with respect to Australia. gGO values were estimated over areas corresponding to that of the two largest exporters of fresh berries to Australia in 2023 (after New Zealand and Vietnam where *D. suzukii* is absent): North America (A) and Thailand (B, with national borders shown in black). For Australia, the reference environment is that of Sidney's harbor, an obvious potential sensitive introduction point of pest species in Australia. Gray pixels represent outlying gGO values (as defined in M&M).

When comparing the two gGO maps in Figure 6, we note that Thailand was characterized by overall lower gGO values than North America. More precisely, Thailand show consistently moderate gGO values, with no area of very high or extreme values, and only a small fraction of pixels with low gGO values. Interestingly, in North America, the regions with the lowest gGO values are mainly the east coast of the USA and Central America. These are not the major producers of *D. suzukii* host plant fruits imported into Australia.

# Discussion

The main objective of this study was to describe and empirically evaluate a general population genomics framework for assessing genetic adaptation during biological invasion. The biological model used as a guideline throughout this paper was *D. suzukii*, an invasive pest species originating from Asia that has recently and rapidly invaded the American and European continents. Following and extending the previous work by Olazcuaga *et al.* (2020), our strategy was first to identify key genomic regions associated with invasive success by contrasting genomic variation between populations from the native and invaded areas. Taking advantage of several recent studies, we were able to extend the worldwide representation of the original dataset by including several publicly available data sets consisting of both Ind-Seq (Lewald *et al.*, 2021) and Pool-Seq samples (Olazcuaga *et al.*, 2020; Feng *et al.*, 2023).These data were carefully curated (Gautier, 2023), along with newly generated data for nine individuals from the native area and four additional population pools.

Random allele PCA implemented in `poolfstat` (Gautier *et al.*, 2022) and originally proposed by Skoglund and Jakobsson (2011) to deal with ancient DNA sequencing data, has proven effective in accounting for the heterogeneity of the samples to provide a global picture of the structuring of genetic diversity. However, as it consists of sampling a single read in each (Pool-Seq or Ind-Seq) sample at each genomic position, the resulting loss of information would make such an approach suboptimal for detecting local genomic effects related to adaptive differentiation or association with environmental covariates. To the best of our knowledge, no population genomics method has been developed that fully exploits the information contained in heterogeneous genomic datasets. This is unfortunate, as such approaches would be extremely valuable as heterogeneous genomic datasets become increasingly available across a growing number of species due to the democratization of sequencing technologies used in a variety of contexts. Taking advantage of the versatility of Bayesian hierarchical modeling, we developed here an extension to the models implemented in the BayPass software (Gautier, 2015) that allows the joint analysis of allele count data of population samples (obtained from the genotypes of the corresponding sampled individuals) and/or read count data (e.g., PoolSeq) and/or individual genotype likelihoods (GLs) directly obtained from Ind-Seq data; see the manual of BayPass version 3.0 for more details.

Modeling GL in low to medium coverage Ind-Seq data (i.e., $< 15X$) has already been shown to be effective in properly accounting for and integrating over genotype uncertainty and is recommended in a variety of population genomics methods (e.g., Kim *et al.*, 2011; Skotte *et al.*, 2013; Korneliussen *et al.*, 2014; Druet and Gautier, 2017; Meisner and Albrechtsen, 2018). Note that GL modeling in BayPass also allows the genotype calling of individuals with a joint estimate of their population allele frequency, while accounting for relationships with other sampled populations. For populations characterized by high coverage Ind-Seq data (e.g., >20X), it may be more efficient and equally accurate in BayPass analyses to consider sample allele counts derived from individual genotypes, which

can then be reliably determined beforehand. Overall, these new model enhancements implemented in BayPass now allow a wide range of population genomics approaches to be performed on hybrid datasets, including estimation of population structure (via the $\Omega$ matrix), genome scans for adaptive differentiation and genome-wide association analyses with population-specific covariates (GEA) or based on contrast ($C_2$).

Based on our expanded combined data set, we searched for key genomic regions that may be associated with the invasive success of *D. suzukii* populations using the $C_2$ contrast statistic (Olazcuaga *et al.*, 2020). It is worth stressing here that we relied for this genome scan on a chromosome-level assembly for the *D. suzukii* genome that we had newly generated, leveraging the accuracy of PacBio HiFi reads and the structural insights from HiC data for scaffolding. Our assembly was obtained from males belonging to an isofemale line of Japanese origin established in the 1980's (Ohashi *et al.*, 1991), i.e., before the *D. suzukii* invasion outside Asia, in contrast to all current and publicly available assemblies obtained from females of isofemale lines representative of the European or North American invaded areas (e.g. Ometto *et al.*, 2013; Chiu *et al.*, 2013; Paris *et al.*, 2020). It can thus be considered as more representative of a *D. suzukii* "ancestral" genome for the native area (disregarding subsequent genetic changes that have occurred in the isofemale line) to identify *de novo* mutation in invasive populations, especially regarding structural variation (including TE content), which we did not examine here. Besides, this new assembly provided a critical improvement over available assemblies in terms of both contiguity and completeness (including the Y chromosome), and may thus represent a valuable reference for other genomic studies on *D. suzukii*. In particular, chromosome 3 could be completely scaffolded, whereas the 3R chromosome was poorly assembled in the previous reference assembly (Paris *et al.*, 2020) on which Olazcuaga *et al.* (2020) relied.

In practice, this more contiguous assembly enabled a better use of the local score approach of Fariello *et al.* (2017), which accounts for LD information by combining $C_2$ statistics at neighboring SNPs at a whole chromosome scale. This both increased the support and facilitated the delineation of genomic regions associated with the invasive status of populations. It also facilitated the comparison of the results obtained when contrasting native with European ($C_2^{EU}$), American ($C_2^{AM}$), or all ($C_2^{WW}$) invasive populations. In total, the local score approach led to the identification of 38 significant windows, which were all small (i.e., <15 kb), in agreement with the low level of SNP signal clustering already observed in Olazcuaga *et al.* (2020). Similarly, consistent with the substantially different subsets of SNPs found to be associated based on $C_2^{EU}$ and $C_2^{AM}$ in Olazcuaga *et al.* (2020), the corresponding continent-specific windows showed virtually no overlap. In addition, several highly significant windows (especially on chromosome 2R) were $C_2^{EU}$ specific and disappeared with $C_2^{WW}$. This may suggest a low degree of convergence between the adaptation processes observed in the two (almost) independently invaded areas by *D. suzukii* (i.e. Europe and America). However, it is important to note that the small number of detected windows and the reduced power of continent-specific

analyses, especially for $C_2^{AM}$, which may probably be also related to the larger genetic structure in this continent (Figure 2), greatly complicate the assessment of signal convergence.

Accordingly, the annotation of the genomic windows identified in the present study did not reveal any candidate genes in common with those identified by Olazcuaga *et al.* (2020). This discrepancy may be related to the increased power and robustness provided by our larger dataset and the local-score method, which led us to detect signals not captured by Olazcuaga *et al.* (2020) or, conversely, to discard genes from that previous study that were supported by only a few variants. In addition, we here relied on a much more accurate genome assembly along with an improved variant calling (and genotyping) procedure implemented in `FreeBayes` (Garrison and Marth, 2012), a haplotype caller that relies on local re-alignment of reads. In particular, this may explain why we could not confirm the strong association signal of the three neighboring SNPs upstream of the *cpo* gene, as these were actually located in a low complexity region of the 3R chromosome that was poorly assembled and encompassed several highly complex variants in the previous reference assembly in Paris *et al.* (2020). Although *cpo*, which seemed at first sight particularly attractive as a functional candidate due its role in diapause control (see the brief discussion on this point in Olazcuaga *et al.*, 2020) could not be validated by the present study, we identified some other genes potentially involved in overwintering (e.g., *bab1* and *cwo*) and resistance to extreme environments (e.g., *nclb*), which may represent valuable traits underlying invasion success. Additionally, *CCKLR-17D1* (controlling aggressive behavior and locomotion) and genes potentially related to pesticide resistance (e.g., *Vps13*) were found to be associated with invasive status. These genes may have contributed to invasive success by, for example, reducing intraspecific competition (Emiljanowicz *et al.*, 2017), improving larval response to stressful environments (Chen *et al.*, 2012), or increasing tolerance to pesticides to which *D. suzukii* is frequently exposed and can develop resistance rapidly according to laboratory assays (Haye *et al.*, 2016; Deans and Hutchison, 2022).

However, association analyses such as the one performed here based on the $C_2$ are expected to provide a very partial view of the genetic basis of the invasive success of a species, as the underlying traits are expected to be multiple and highly polygenic. Association analyses thus primarily aim at identifying a few important genes, such as the few candidates discussed above, that may deserve special interest for future validation studies to better characterize the traits they control.

To provide a global picture of the (possible) role of genetic adaptation in the invasive success of *D. suzukii*, we considered an alternative approach based on modeling the relationship between whole-genome genomic variation and environmental variables among populations, building on recent work on estimating population maladaptation using the so-called genomic-offset GO statistic (e.g. Capblancq *et al.*, 2020; Gain *et al.*, 2023). Indeed, under the assumptions that i) the considered environmental variables (at least for some of them) are among the main drivers of invasion success; and that ii) genomic information is available from a set of presumably locally adapted populations

representative of the species niches in both the native and the invaded area, then low (resp. high) GO values can be interpreted as a relative measure of (genetic) pre-adaptation (resp. mal-adaptation) of the populations. Accordingly, we recently showed in a simulation study that GO can be related to the probability of invasive population establishment and can be efficiently estimated from the GEA regression model implemented in BayPass (Camus *et al.*, 2024). From a practical point of view, our extension to the analysis of hybrid datasets allowed BayPass-based gGO estimation on our combined Pool-Seq and Ind-Seq data, which was not possible (or highly suboptimal) with other available methods.

Using the worldwide invasion of *D. suzukii* as a case study, we illustrated how GO-based analyses provide insights into the level of (genetic) adaptive challenge that new environments in the invaded areas may represent for a population adapted to a region of the native area, thereby allowing to better understand the past dynamics of biological invasion. Interestingly, despite evidence that some key genes were associated with adaptation in invaded areas, GO-based results tend to mitigate the overall contribution of genetic adaptation to the invasion success of *D. suzukii*. Indeed, gGO values across North America and Europe were generally low and homogeneous, particularly in Europe, suggesting that most of these invaded areas did not pose a significant adaptive challenge to propagules originating from the environments associated to the source populations. This finding is consistent with the rapid invasion of *D. suzukii*, suggesting that once it arrived on these continents, it faced minimal adaptive challenges because some forms of pre-adaptation to invaded areas existed in native source populations. While not necessarily contradicting the findings of Feng *et al.* (2023), who identified many candidate SNPs associated with environmental variables, our results suggest that the underlying genomic regions were likely not crucial in the invasion process. Our GO-based conclusions are also in line with a previous study using species distribution modeling on *D. suzukii*, which indicated the absence of obvious climatic niche shift between Asia and the first observation points in Europe and the US (Fraimout and Monnet, 2018).

Following a reverse logic, the GO maps we computed over Africa and South America illustrate how GO-based analyses can be used to identify the most vulnerable parts of areas that have not been, at least fully, invaded to date. According to our results, relatively low GO values were found over large parts of these continents (e.g., the Amazon basin in South America or the Great Lakes region in Africa) where *D. suzukii* is currently absent but could potentially establish rapidly. Some post-hoc surveillance of these regions would be required to assess the extent to which GO-based predictions are informative. Indeed, although GO values were found to be low in some of the recently invaded regions of this continent (e.g., South Africa), it remained more moderate in others (e.g., central Patagonia or Chile). In addition, the choice of the reference source environment from which the (pre-adapted) invading population would originate can be critical. For the sake of simplicity, we have assumed for both Africa and South America that the reference source environment is the one

where populations have been established for at least two consecutive years (and recently), ensuring that these regions can be considered as particularly vulnerable to *D. suzukii* invasion. Consideration of other source environments (e.g., Europe) may result in slightly to substantially different GO maps. For example, in South America, although the first record of *D. suzukii* dates back to 2013 (Deprá *et al.*, 2014), it was reported just one year later, in 2014, in northern Patagonia (Lavagnino *et al.*, 2018), approximately 500 km from the AR-Gol population studied here and sampled in 2023. Given the rapid expansion of the species in Patagonia, AR-Gol could also represent a suitable reference environment to compute the GO map of invasion across South America, leading to even more pessimistic predictions than those made based on the Nova Veneza (Brazil) environment in Figure 5B.

GO can also be used to identify the region of origin of a (locally adapted) population that would be the most (genetically) pre-adapted to a specific target area (see also Gautier *et al.*, 2024, for another recent application). Such information can be useful for IPM practice to increase vigilance on importing products from the highlighted high risk regions (i.e., those with low GO values and where *D. suzukii* is established). As illustrated for the non-invaded Australian continent in the present study, the computation of GO maps using a reference environment associated with the northwestern American US-Wa population, one of the major regions for fresh fruit imports, supports a limited overall risk of invasion from there. However, some densely populated areas of southern Australia that are likely targets of significant food trade still exhibited low gGO values, suggesting that such a US-Wa-like *D. suzukii* population could easily establish there and gradually spread throughout the country. While our GO analysis suggests that invasions from Thailand present a greater (adaptive) risk than those from the USA, implying that Australian management should also focus on monitoring imports from Thailand, it is worth noting that Australia already enforces strict control measures on most goods at risk of *D. suzukii* infestation. Previous research, which did not incorporate genomic data, similarly identified a high risk of *D. suzukii* invasion, particularly through passive human-mediated pathways (Maino *et al.*, 2021). As a result, our analysis serves more as an illustration of how GO-based methods could be applied in IPM rather than providing actionable insights for Australia, where rigorous biosecurity measures are already in place. In a more general perspective, the sequencing of newly established populations could serve as an empirical method to validate GO predictions in biological invasions. By comparing the most likely sources of invasion inferred without environmental data to those predicted through GO computation, we could assess the accuracy and relevance of GO-based predictions in real-world scenarios.

Overall, the GO-based approach appears promising for understanding and predicting *D. suzukii* colonization in different environments. From a biological perspective, GO estimates made sense globally, as regions with high estimated GO values corresponded to environmental characteristics known to be physiologically challenging for *D. suzukii*. In particular, high to extreme gGO values

were observed in cold regions, which are inhospitable to *D. suzukii* due to its "chill-susceptibility" (Kimura, 2004; Enriquez and Colinet, 2017; Ryan *et al.*, 2016; Winkler *et al.*, 2020), although *D. suzukii* can survive cold winters by using natural or anthropogenic shelters (Jakobs *et al.*, 2015; Stockton *et al.*, 2019a; Rossi-Stacconi *et al.*, 2016). Similarly, elevated gGO values have been found in arid and semi-arid desert regions, which are also unsuitable for *D. suzukii* due to its low tolerance to drought and high temperatures (Winkler *et al.*, 2020; Fanning *et al.*, 2019; Eben *et al.*, 2018). Nonetheless, it must be noted that predictions of potential future distributions based on Ecological Niche Models (ENM) show low to no suitability for some areas that were associated with moderate or low gGO values in our analyses, namely tropical region such as the Amazonian basin or Central Africa (Nair and Peterson, 2023; Ørsted and Ørsted, 2019). This discrepancy can be attributed to the inherent differences between the ENM-based and GO-based approaches, which do not rely on the same data to model the relationship between populations and their environments. On one hand, ENM can overestimate environmental suitability in areas with high densities of occurrence data and underestimate it in less sampled regions (Dubos *et al.*, 2022). In other words, the results of ENMs might underestimate environmental suitability in these tropical zones, overlooking the fact that the climatic characteristics of these regions are not necessarily limiting for the species, as indicated by gGO. On the other hand, the GO-based method focuses on a limited set of sequenced populations, which may limit its ability to capture the full ecological niche of the species.

Moreover, it is important to recognize that the biological interpretability of GO remains limited and has not yet been fully resolved (Camus *et al.*, 2024), making it difficult to rely on it to provide an absolute quantitative measure of invasion risk or adaptation difficulty. As a result, it is recommended to interpret the GO relatively, i.e., by ranking environments according to their susceptibility to invasion by source populations that are (genetically) adapted to a given reference region.

In addition, it is important to recognize that other factors not included in the models underlying GO calculations may be critical for successful invasion and must also be considered to explain or predict the establishment of invasive species. First, some pivotal environmental variables could have been disregarded in the GEA or their variation may be imperfectly captured with the analyzed PCs, which may have decisive impact on the GO predictive performance (Camus *et al.*, 2024). For instance, our GO calculation did not (directly) incorporate fruit-related variables other than land use due to data availability limitations at the considered continental scales. Given that fruit availability is an important component of *D. suzukii* fitness, its inclusion in future gGO analyses could greatly improve the accuracy of invasion predictions by accounting for additional factors that influence *D. suzukii* establishment success. Second, innate characteristics of the species, including phenotypic plasticity, can play an important role for invasion success and may be difficult (if not impossible) to include in GO computations. For instance, *D. suzukii* has several characteristics that may have facilitated its successful invasion such as its ability to i) tolerate cold environments through mecha-

nisms such as overwintering or diapause (Little *et al.*, 2020); ii) exploit a broad trophic niche with host plants spanning several families and many fruit species (Olazcuaga *et al.*, 2023); and iii) even subsist on a non-fruit diet (Poyet *et al.*, 2015; Stockton *et al.*, 2019b). Finally, contingent and extrinsic factors are obviously required for a species to successfully invade a new area. At the very least, enough individuals (of both sexes for sexual species) must first be able to reach it. In the case of *D. suzukii*, its widespread distribution is likely to be driven by human-mediated passive dispersal rather than active movement (Adrion *et al.*, 2014; Calabria *et al.*, 2012; Maino *et al.*, 2021). This form of assisted migration can significantly affect the dynamics of invasion, influencing which areas are invaded first and the spread of the species in a newly invaded area. Moreover, the number of introduced individuals (propagule size) and their arrival patterns (number and spatiotemporal distribution of introduction events), which together define the so-called "propagule pressure", have also been shown to be a crucial factor in the establishment of invasive populations (Simberloff, 2009; Estoup *et al.*, 2016). Indeed, high propagule pressure can buffer adaptive challenges, as shown in our recent simulation study where the performance of GO-based analyses in predicting the probability of establishment of invasive populations was found to be reduced for scenarios involving the single introduction of larger numbers of individuals (Camus *et al.*, 2024). For example, even though populations from Thailand may be the most pre-adapted to Australia, those from America may be more likely to become established in Australia due to greater fruit infestation and consequently greater propagule pressure.

Despite these limitations and the potential for further refinement, our findings provide valuable insights into the invasion success of *D. suzukii* and highlight the effectiveness of GO-based analyses in elucidating the factors underpinning the dynamic of past and future biological invasions. While further validation may be required, the proposed general framework, including our newly developed methodological tools, can be applied to various invasive species to help guiding management decisions.

# Materials and Methods

## Construction and annotation of a new *de novo* genome assembly

To generate a novel high-quality chromosome-level *de novo* assembly of the *Drosophila suzukii* genome, we relied on PacBio HiFi long reads in conjunction with HiC sequencing data for scaffolding. For HiFi sequencing, we first extracted high molecular weight (HMW) DNA using the QIAGEN Genomic-tips 500/G kit (Qiagen, MD, USA) from a pool of 30 flash-frozen males belonging to an isofemale line of Japanese origin established from females captured on Hachijo Island in 1978 (Ohashi *et al.*, 1991). A PacBio HiFi library was then constructed using SMRTbell Template Prep kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) according to PacBio recommendations

(SMRTbell express template prep kit 2.0 - PN: 100-938-900). The library was loaded on one SMRTcell 8M at an on-plate concentration of 90pM. Sequencing was performed on the Sequel II system at Gentyane Genomic platform (INRAE Clermont-Ferrand, France) with Sequel II Sequencing kit 3.0, a run movie time of 30 hours with an Adaptive Loading target (P1 + P2) at 0.75. After filtering and correcting the SMRTcell output, we obtained 743,798 reads totaling 14.7 Gb, with an N50 of 20,851 bp and a maximal length of 50.197 kb as assessed with LongQC v1.2.1 (Fukasawa *et al.*, 2020).

A whole-genome primary assembly was then built from the HiFi reads using `hifiasm` (v0.19.8; Cheng *et al.*, 2021) with the options `--primary` and `--n-hap 8` to account for residual genetic diversity among the pooled individuals included in the HiFi library. Based on preliminary tests, this latter value was indeed found to reduce the overall assembly sizes and duplicate rates estimated with BUSCO (Manni *et al.*, 2021, see below). The resulting primary assembly consisted of 167 contigs totaling 305.9 Mb (N50=6.18 Mb), which were further filtered by removing one (poorly assembled) mitochondrial contig and 25 contaminant contigs (i.e., non *D. suzukii*) identified following Gautier (2023) by querying a database constructed from the NCBI non-redundant nucleotides (nt) released in February 2020 using the program `Kraken2` (v2.1.2; Wood *et al.*, 2019). In particular, this included a 3.33 Mb contig assigned to *Acetobacter persici*, a 1.35 Mb contig assigned to *Lactobacillus fructivorans*, and a 93.0 kb contig assigned to *Acetobacter pasteurianus*. Finally, we completed the curation of the contig assembly by performing a self-alignment using `minimap2` (v2.20-r1061; Li, 2021) with the options `-xasm5 -DP --dual=no` to identify residual haplotigs. Note that all contigs $<$ 100 kb were discarded as most were found to be haplotigs or junk DNA. The final curated contig assembly consisted of 75 contigs totaling 282,289,391 bp (N50=10.6 Mb and N90=1.79Mb). Following Gautier *et al.* (2018), we were able to identify eight X-linked contigs (totaling 39.14 Mb) and three Y–linked contigs (totalling 16.74 Mb) based on the ratio of the relative mean read coverage of the two pools `RS-Zaj` and `DE-Jen` consisting exclusively of males and females, respectively (see below and Table S2). More precisely, the two corresponding Pool-Seq data set were filtered with `fastp` (v0.23.2; Chen *et al.*, 2018) and mapped to the contig assembly with `bwa-mem2` (v2.2.1; Vasimuddin *et al.*, 2019), both run with default options, and the coverages for each contig were estimated from the resulting *bam* files using `mosdepth` (v0.3.6; Pedersen and Quinlan, 2017) run with the option `-Q 20`.

To further scaffold the contig assembly at the chromosome level, we incorporated HiC data, which provides long-range chromatin interaction information. The HiC library was prepared from a pool of three flash frozen males using the Arima-HiC kit according to the manufacturer instructions and paired-end (2x150bp) sequenced on an Illumina NovaSeq 6000 on the Get Plage platform (INRAE Toulouse, France). The obtained raw paired-end reads were filtered with `fastp` (v0.23.2; Chen *et al.*, 2018) with default options, resulting in a total of 573,783,218 read pairs with an estimated 18.25% of duplicate reads. The filtered paired-end reads were then mapped to the contig assembly using the `HiCstuff` pipeline program (v3.2.1; Baudry *et al.*, 2020; Matthey-Doret *et al.*, 2020) with the

options `--mapping=iterative`, `--read-len=150`, `--duplicates`, `--distance-law`, `--filter` and `-e "DpnII,HinfI"`. A total of 148,588,798 valid pairs were successfully identified and used to scaffold the assembly with `yahs` (v1.2; Zhou *et al.*, 2022) and default options, except `{file-type PA5` to read the input *pair* file.

The completeness of the different assemblies was evaluated using `BUSCO` for the `diptera_odb10` dataset containing 3,285 genes (v5.4.7; Manni *et al.*, 2021). The overall properties of the assemblies (BUSCO completeness and contiguity) were summarized with snail plots using the `BlobToolKit` viewer (v4.3.5; Challis *et al.*, 2020), and they were compared with other available assemblies, including for closely related species (*D. subpulchrella*, *D. biarmipes* and *D. melanogaster*) using dot plots generated with `dgenies` (v1.5.0; Cabanettes and Klopp, 2018). To further annotate the new assembly, we identified repetitive elements using `RepeatMasker` (v4.1.4; Smit *et al.*, 2023) run with options `-s -no_is -cutoff 200 -frag 20000`, the `wublast` search engine (option `-e wublast`) and using the curated and annotated TE database developed by (Mérel *et al.*, 2021) for the *D. suzukii* genome. Finally, the program `liftoff` (v1.6.3; Shumate and Salzberg, 2021) with default options was used to transfer the NCBI *Gnomon* annotation available from the NCBI website.

## *D. suzukii* population samples and whole genome sequencing data

Whole-genome sequencing data representative of 31 worldwide population samples, derived from either Pool-Seq (n=23 pools; Table S2) or Ind-Seq (n=73 individuals representative of 8 populations; Table S3) were obtained from the NCBI SRA repository (`https://www.ncbi.nlm.nih.gov/sra/`). These data are the result of a careful curation of publicly available WGS data from three recent studies (Olazcuaga *et al.*, 2020; Lewald *et al.*, 2021; Feng *et al.*, 2023). Following (Gautier, 2023), we removed all samples with a high level of contaminating sequences, except for the CN-Nin sample (collected in Ningbo, China). This latter sample indeed remained highly valuable as it represented a native populations closely related to one of the contributing sources of the America invaded area identified by (Fraimout *et al.*, 2017). Although this sample exhibited 15% *D. subpulchrella* contamination (i.e., 7 *D. subpulchrella* out of the 50 pooled individuals), we do not expect any strong bias introduced into our global analyses, since the *D. subpulchrella* specific variants may remain at very low frequency in the total analyzed data sets. Accordingly, the CN-Nin sample was found to be closely related to other samples from China in Figures 2B or 2C. We considered only two (JP-Kan and US-CAOx) of the seven Pool-Seq samples from Feng *et al.* (2023), the others being either already sequenced samples (n=3) or closely related geographically (n=2) to those from Olazcuaga *et al.* (2020). Finally, we selected a subset of the 235 individuals sequenced by Lewald *et al.* (2021) based on both geographic and preliminary genetic analyses to limit redundancy in the representation of US populations and to ensure a minimum sample size per population (from n=7 to n=10 individuals). Note that the samples representing the US states of Georgia (US-Ga, n=9),

New York (US-Ny, n=10), and Washington (US-Wa, n=10) each combined individuals collected simultaneously from several neighboring sampling sites located at a maximum of 27 km, 45 km, and 15 km apart, respectively (Table S3).

To further improve worldwide representation of the samples, we newly sequenced four DNA pools representative of one South-American (AR-Gol from Argentina, n=100) and three European (CH-Del from Switzerland, n=58; DE-Jen from Germany, n=100; and RS-Zaj from Serbia, n=100) populations, as well as nine native individuals collected in Chengdu (China) in 2014. Note that the DE-Jen individuals are from the same sampling used in (Olazcuaga *et al.*, 2020) and (Feng *et al.*, 2023), but with all non-*D. suzukii* individuals discarded (Gautier, 2023). DNA extraction was performed from whole bodies using the DNeasy Blood and Tissue Kit (Qiagen, USA) according to the manufacturer's recommendations. Libraries were constructed using the NEBNext Ultra II DNA Library prep kit for Illumina (New England Biolabs). All samples were paired-end sequenced (2x150 bp or 2x250 bp) on a NovaSeq6000 Illumina sequencer (Tables S2 and S3) on the Get Plage platform (INRAE, Toulouse). Some analyses (see below) also relied on paired-end sequencing data (2x125 bp) of a pool of 6 individuals from a *D. subpulchrella* isofemale line that had previously been generated (in 2018 but not yet published) on a HiSeq2500 sequencer at the MGX platform (CNRS, Montpellier).

## WGS data processing, SNP calling and genotyping

WGS data were first filtered using `fastp` (v0.23.2; Chen *et al.*, 2018) run with default options to remove contaminating adapter sequences and trim for poor quality bases (i.e., with a phred quality score <15). Read pairs with either one read containing more than 40% low-quality bases or more than 5 N bases were removed. The filtered reads were then mapped onto the new *dsu_isojap1.0* assembly, complemented with the complete mitochondrial sequence of *D. suzukii* (GenBank Accession ID: *KU588141*) and the complete sequence for the Wolbachia endosymbiont (wRI strain) of *D. simulans* (NCBI ID: *GCA_000022285*), a Wolbachia strain closely related to that of *D. suzukii* one (Turelli *et al.*, 2018), using the default options of `bwa-mem2` (v2.2.1; Vasimuddin *et al.*, 2019). PCR duplicates were further removed using the `view` and `markdup` programs of `SAMtools` (v1.14; Li *et al.*, 2009). The mean realized coverage was further estimated using `mosdepth` (v0.3.6; Pedersen and Quinlan, 2017) with the option *-Q 20* to ignore reads with low mapping quality. As detailed in Tables S2 and S3, over all autosomal contigs the coverages varied from 38.2 (D-sub pool) to 95.7 (JP-Kan) with a median of 54.5 for Pool-Seq data; and from 1.24 to 12.1 with a median of 6.23X for Ind-Seq data.

Variant calling was further performed using the Haplotype Caller implemented in `FreeBayes` (v1.3.6; Garrison and Marth, 2012), which does not require realignment around indels and base recalibration, jointly considering all 27 Pool-Seq and 82 Ind-Seq alignment files from *D. suzukii* samples, with the

following options *-K -C 1 -F 0.01 -G 5 –limit-coverage 500 -n 4 -m 30 -q 20*. For X-linked contigs, we also specify the ploidy of the 5 male individuals with the *-A* option. To improve computational efficiency, variant calling was performed in parallel on a computer grid for the 1,179 non-overlapping 250 kb windows covering the whole genome assembly. The resulting vcf files were reordered and concatenated per chromosome type using the `view` and `concat` programs of `bcftools` (v1.20; Danecek *et al.*, 2021), and filtered using a custom `awk` script to keep only bi-allelic variants (including the reference allele) with `QUAL>20`, `NS>100`, `100<DP<5000`, `SAF>0`, `SAR>0`, `RPR>1`, `RPL>1`, `EPP>0`, `SRP>0`, and a global MAF computed from the overall read counts $> 0.01$. This yielded 3,634,933 autosomal bi-allelic variants consisting of 3,506,159 SNPs (i.e., 96.5%), 28,797 MNPs (Multi-Nucleotide Polymorphisms), 56,893 indels and 43,084 so-called complex variants. Likewise, 691,982 X–linked bi-allelic variants were obtained, including 666,740 SNPs (i.e., 96.4%), 5,182 MNPs, 11,929 indels and 8,131 complex variants. To provide a global visualization of the genetic structuring of the populations, we performed a random allele PCA using the `randomallele.pca` function of the R package `poolfstat` (v2.2.0 Gautier *et al.*, 2022) on the entire *vcf* combining Pool-Seq and Ind-Seq data parsed with the `vcf2pooldata` function. Note that this is the equivalent to treating each Ind-Seq data as a pool (e.g., of haploid size 2), which has no effect on the random allele PCA since only a single read per sample (from either Pool-Seq or Ind-Seq) and per locus is randomly sampled. Finally, we estimated the read counts for the outgroup *D. subpulchrella* pool (D-sub) for all these identified variants by running `FreeBayes` (v1.3.6; Garrison and Marth, 2012) with the option *–target* to specify their position on the assembly and the options *–haplotype-length 0 -C 1 -F 0 -K –report-monomorphic -m 30 -q 20*.

## Genome scan analyses using combined Pool-Seq and Ind-Seq data

All genome-wide scans were performed using an upgraded version (3.0) of BayPass (Gautier, 2015; Olazcuaga *et al.*, 2020) available from `https://forgemia.inra.fr/mathieu.gautier/baypass_public` (last accessed Sep. 2024). This new version extends all the previously described models and approaches to the analysis of population genetic data consisting of sets of individual genotyping data coded as GL. This allowed noticeably to properly deal with low-to mid-coverage Ind-Seq data (e.g., <10-15X) by integrating over genotype uncertainty. Briefly, let each genotype be coded by its number of alternate allele $l$, i.e., for a diploid individual $l = 0$ if homozygote for the reference allele, $l = 1$ if heterozygote, or $l = 2$ if homozygote for the alternate allele. For haploid individuals (e.g., males at X–linked SNPs), only the two values $l = 0$ or $l = 1$ are observable. Further, let $GL_{ijkl} = \mathrm{P}\left(y_{ijk}^{(ind)} \mid g_{ijk} = l\right)$ denote the likelihood, computed by the genotype caller, of the observed Ind-Seq data for the $k$th individual from population $j$ at locus $i$ when its genotypes $g_{ijk} = l$. Following the same notations as in Gautier (2015) and assuming Hardy-Weinberg equilibrium, the (conditional) probability of the observed data may then simply

be defined as a function of the (reference) allele frequency $\alpha_{ij}$ at a bi-allelic locus $i$ in population $j$ and individual ploidy $p_{ijk}$ ($p_{ijk}$=1 or 2 if individual $k$ from population $j$ is haploid or diploid at SNP $i$, respectively), as $\mathrm{P}\left(y_{ijk}^{(ind)} \mid \alpha_{ij}\right) = \sum_{l=0}^{p_{ijk}} \binom{p_{ijk}}{l} \alpha_{ij}^{p_{ijk}-l} \left(1-\alpha_{ij}\right)^{l} GL_{ijkl}$. Finally, assuming that individuals are randomly sampled (i.e., conditional independence), the conditional probability of the observed data $Y_{ij}^{(ind)} = \left\{y_{ijk}^{(ind)}\right\}$ at SNP $i$ for the $n_j$ individuals from population $j$ is $f\left(Y_{ij}^{(ind)} \mid \alpha_{ij}\right) = \prod_{k=1}^{n_j} \mathrm{P}\left(y_{ijk}^{(ind)} \mid \alpha_{ij}\right)$, which may then be incorporated on the first level of the Bayesian hierarchical models described in Gautier (2015). Sampling of the $\alpha_{ij}$'s from the full posterior distribution was implemented within the BayPass Markov Chain Monte Carlo (MCMC) algorithm using a random walk Metropolis update similar to the alternate algorithm 1 described in Supplementary File S1 of Gautier (2015). In addition, capitalizing on the versatility of Bayesian hierarchical modeling, the new version 3.0 of BayPass was modified to allow for joint analysis of allele count and/or read count (e.g., PoolSeq) and/or GL data. More precisely, let $\boldsymbol{Y} = \left\{\left\{Y_{ij}^{(ac)}\right\}_{j\in\left[1,J^{(ac)}\right]}; \left\{Y_{ij'}^{(rc)}\right\}_{j'\in\left[1,J^{(rc)}\right]}; \left\{Y_{ij''}^{(ind)}\right\}_{j''\in\left[1,J^{(ind)}\right]}\right\}_{i\in[1,nsnp]}$ a genomic data set consisting of (sample) allele count $Y^{(ac)}$, read count (e.g., Pool-Seq) $Y^{(rc)}$ and individuals' GLs (e.g., from Ind-Seq) $Y^{(ind)}$ data for $J^{(ac)}$, $J^{(rc)}$ and $J^{(ind)}$ sampled populations, respectively. Assuming conditional independence of markers (i.e., not accounting for LD), given the underlying population allele frequencies $\boldsymbol{\alpha} = \left\{\{\alpha_{ij}\}_j; \{\alpha_{ij'}\}_{j'}; \{\alpha_{ij''}\}_{j''}\right\}_i$, the full likelihood of the combined data set $\boldsymbol{Y}$ becomes:

$$f\left(\boldsymbol{Y} \mid \boldsymbol{\alpha}\right) = \prod_{i=1}^{nsnp} \left( \prod_{j=1}^{J^{(ac)}} f\left(Y_{ij}^{(ac)} \mid \alpha_{ij}\right) \prod_{j'=1}^{J^{(rc)}} f\left(Y_{ij'}^{(rc)} \mid \alpha_{ij'}\right) \prod_{j''=1}^{J^{(ind)}} f\left(Y_{ij''}^{(ind)} \mid \alpha_{ij''}\right) \right)$$

, where $f\left(Y_{ij}^{(ac)} \mid \alpha_{ij}\right)$ and $f\left(Y_{ij}^{(ac)} \mid \alpha_{ij}\right)$ are the likelihoods previously defined in Gautier (2015) for allele count and read count data, respectively, and $f\left(Y_{ij''}^{(ind)} \mid \alpha_{ij''}\right)$ is the likelihood based on GL's defined above. Sampling of all the $\alpha_{ij}$'s from the full posterior distribution is then performed using MCMC as described in Gautier (2015) for the $\alpha_{ij}$'s associated with allele count or read count data, and as described above for the $\alpha_{ij}$'s associated with GL data (see the BayPass manual for additional details).

Following the BayPass manual recommendation (see also Gautier *et al.*, 2018), the complete dataset was first split into sub-data sets of ca. 106,500 markers each for autosomes and 86,000 markers each for the X-chromosome. The program was then run on each of these sub-data sets with the default option to both estimate the scaled covariance matrix of population allelic frequencies ($\Omega$), which allows accounting for the neutral population structure in the different analyses, and three $C_2$ statistics (Olazcuaga *et al.*, 2020) contrasting native versus i) worldwide ($C_2^{WW}$), ii) European ($C_2^{EU}$), or iii) American ($C_2^{AM}$) invasive populations. As observed in previous similar analyses (e.g., Gautier *et al.*, 2018; Olazcuaga *et al.*, 2020), we found a fairly high correlation of the estimates of $\Omega$ entries over sub-data sets (from the same chromosome type), and we thus only reported posterior mean

estimates obtained from a single randomly chosen sub-data set. Likewise, $C_2$ estimates were highly consistent across three independent analyses (i.e., with different seeds) of each sub-data set, and only results from the first analysis were further considered. To identify and delineate significant windows associated with invasive status based on the three different contrasts, i.e. with an excess of high $C_2$ values, we relied on the local-score approach described in Fariello *et al.* (2017) and (re)implemented in the R function `compute.local.scores` available in the BayPass software package (version 3.0). More precisely, for each contrast, individual SNP scores were computed as $-\log 10(p) - \xi$, where $p$ is the p-value associated with a given $C_2$ and $\xi$ determines the threshold of p-values that are considered to identify candidate windows. We set $\xi = 2$ (i.e., p-values$< 10^{-2}$ were cumulated and others were scored 0 in the underlying Lindley process). As recommended in the BayPass manual, to avoid penalizing the approach by the inclusion of non-informative low-polymorphic SNPs, SNPs with an extreme overall allele frequency, as estimated by parameter $\pi$ (i.e., $\pi < 0.2$ or $\pi > 0.8$) were discarded from the local-score computation.

To evaluate the (approximate) expected number of overlapping genomic windows between two independent tests, we considered all possible pairs $(R_1, R_2)$, where $R_i$ is a region detected by test number $i$, whose size in bp is denoted $l_i$. The two regions overlap when $R_2$ starts either within $R_1$, which occurs with probability $l_1/L$ where $L$ is the total genome size, or less than $l_2$ bp ahead of $R_1$, which occurs with probability $l_2/L$. The total number of expected overlaps was thus computed as the sum of $(l_1 + l_2)/L$ over all possible pairs. To avoid under-estimating this number due to the existence of large parts of the genome associated with a limited detection power, only chromosomes arms where signals had been actually detected were accounted when computing $L$.

All significant windows were further manually annotated on the *Jbrowse 2* genome browser (Diesh *et al.*, 2023), using the `gff` annotation file obtained for the *dsu_isojap1.0* new assembly by tansfering NCBI *Gnomon* annotation (see above). *D. melanogaster* gene orthologs of each of the NCBI ID of *D. suzukii* genes were obtained from the *OrthoDB* database (Kuznetsov *et al.*, 2023). Finally, to investigate biological processes, Gene Ontology enrichment analysis for Biological Process and Molecular Function was performed on the resulting list of candidate genes (from all three contrasts) using the R package *clusterProfiler* (v4.12.0; Wu *et al.*, 2021) with default options. Note that for this analysis, gene names were either obtained from the annotated NCBI *D. suzukii* genes, or the *Uniprot* database (The UniProt Consortium, 2023) for those found with *OrthoDB*.

### Estimation of GO

To compute GO, we first performed a GEA analysis under the BayPass standard covariate model (Gautier, 2015) with option `-covmcmc` using the above estimated $\Omega$ matrix. This first step aimed at estimating the regression coefficients (environmental effect) of different environmental covariables on the (scaled) population allele frequencies to model the relationship between the genomic diversity

and environment variation of the studied populations. Different type of data were considered to characterize the environment. These included 19 commonly utilized bioclimatic variables, along with the mean monthly climate moisture index following Winkler *et al.* (2020) who showed the importance of humidity on fitness-related trait in *D. suzukii*, and mean monthly-near surface wind speed, which has been suggested as one of the main driver for adaptation in *D. melanogaster* European populations (Bogaerts-Márquez *et al.*, 2021). Data for these 21 covariables were sourced from the Chelsa database (Karger *et al.*, 2017), as averaged values over 1981 to 2010, and extracted using the R packages *dismo* (v1.3.5; Hijmans *et al.*, 2023) and *raster* (v3.5.15; Hijmans, 2022). In addition, we also included covariables related to land use (for year 2017) that were extracted from the ESRI Land Cover 2020 database (Karra *et al.*, 2021) and processed as percentages for seven of the nine class (clouds, crops, built area, trees, water, bareground and rangeland) with a resolution comparable to that of the Chelsa data. Note that the two other land use classes, ice/snow and flooded vegetation, were removed as their values were uninformative for the considered locations (all equal to 0 for ice/snow, and only one different of 0 for flooded vegetation). Finally, to address variable interdependence, which can reduce the GO predictive performance of establishment probability (Camus *et al.*, 2024), we performed a PCA on the 28 resulting covariables and retained the first 10 PCs (explaining $> 95\%$ of the variance) as environmental covariables in the GEA analysis.

Following Gain *et al.* (2023), and as previously described (Camus *et al.*, 2024; Gautier *et al.*, 2024), GO was computed in the form of gGO as:

$$\mathrm{gGO} = \frac{1}{n_{snp}} \left(\boldsymbol{e} - \boldsymbol{e}^{\star}\right)' \boldsymbol{B}' \boldsymbol{B} \left(\boldsymbol{e} - \boldsymbol{e}^{\star}\right)$$

where $n_{snp}$ is the number of genotyped markers; $\boldsymbol{e}$ and $\boldsymbol{e}^{\star}$ are the vectors of $m = 6$ environmental PCs for the two compared environments; and $\boldsymbol{B}$ is the $n_{snp} \times m$ matrix of the regression coefficients $\beta_{jk}$ of environmental variable $k$ (i.e. PC $k$) on population allele frequencies at SNP $j$ as estimated under a GEA linear model. Note that estimated regression coefficients for both autosomes and X-chromosome were considered for gGO computation, without any marker pre-selection as suggested by recent studies (Gain *et al.*, 2023; Lind and Lotterhos, 2024; Camus *et al.*, 2024), using a slightly modified version of the `compute_genetic_offset` function available in BayPass v3.0 software package. Finally, for graphical representation, the same scale was used for all gGO maps, defining the threshold for extreme gGO as the 90th percentile of the gGO estimated over North America, which roughly corresponds to the 94th percentile of the gGO computed over Europe (Figure 4), and thus seemed as an acceptable threshold to compare gGO maps on the same scale.

# Data availability

All data were deposited on the SRA repository (project PRJNA1032894). These included Pacbio HiFi raw reads (Run ID: SRR29552229) and the HiC library paired-end sequences (Run ID: SRR29552230) used for the de novo assembly of Japanese isofemale line, and all new Pool-Seq and Ind-Seq data (see accession details in Tables S2 and S3). The *dsu_isojap1.0* whole genome assembly has been deposited at DDBJ/ENA/GenBank under the accession JBFDSZ000000000. The version described in this paper is version JBFDSZ010000000.

# Acknowledgments

We wish to thank Toshiro Aigaki (TMU, Tokyo, Japan), Nicolas Gompel (Bonn University, Germany) and Benjamin Prud'homme (IBDML, Université Aix-Marseille, France) for providing and having maintained individuals from the Japanese isofemale line used for the genome assembly. We are also grateful to Benjamin Prud'homme for providing *D. subpulchrella* individuals, and Dr Marina Stamenković-Radak, Dr Mihailo Jelić, and Mina Rakić for their assistance in collecting the sample from Serbia. Mathieu Gautier, Simon Boitard and Arnaud Estoup have benefited from State aid managed by the Agence Nationale de la Recherche under the France 2030 programme, reference ANR-23-EXMA-0002 ("AgroStat" project). Svitlana Serga was supported by the PAUSE-ANR Ukraine Program. This work was performed in collaboration with the GeT core facility, Toulouse, France (GeT, `https://doi.org/10.15454/1.5572370921303193E12`). MGX and GeT core facility acknowledge financial support from France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We are also grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, `https://doi.org/10.15454/1.5572369328961167E12`) for providing computing resources. We finally would like to thank the DrosEU consortium for helpful discussion during stimulating meetings.

# Bibliography


Adrion, J. R., Kousathanas, A., Pascual, M., Burrack, H. J., Haddad, N. M., Bergland, A. O., Machado, H., Sackton, T. B., Schlenke, T. A., Watada, M., Wegmann, D., and Singh, N. D. 2014. Drosophila suzukii: The Genetic Footprint of a Recent, Worldwide Invasion. *Molecular Biology and Evolution*, 31(12): 3148–3163.

Agrawal, P., Kao, D., Chung, P., and Looger, L. L. 2020. The neuropeptide drosulfakinin regulates social isolation-induced aggression in drosophila. *Journal of Experimental Biology*, 223(2): jeb207407.

Aouari, I., Barech, G., and Khaldi, M. 2022. First record of the agricultural pest *Drosophila suzukii* (Matsumura, 1931) (Diptera: Drosophilidae) in Algeria. *EPPO Bulletin*, 52(2): 471–478.

Asplen, M. K., Anfora, G., Biondi, A., Choi, D. S., Chu, D., Daane, K. M., Gibert, P., Gutierrez, A. P., Hoelmer, K. A., Hutchison, W. D., Isaacs, R., Jiang, Z. L., Kárpáti, Z., Kimura, M. T., Pascual, M., Philips, C. R., Plantamp, C., Ponti, L., Vétek, G., Vogt, H., Walton, V. M., Yu, Y., Zappalà, L., and Desneux, N. 2015. Invasion biology of spotted wing Drosophila (Drosophila suzukii): a global perspective and future priorities. *Journal of Pest Science*, 88(3): 469–494.

Barrett, R. D. H. and Schluter, D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, 23(1): 38–44.

Baudry, L., Guiglielmoni, N., Marie-Nelly, H., Cormier, A., Marbouty, M., Avia, K., Mie, Y. L., Godfroy, O., Sterck, L., Cock, J. M., Zimmer, C., Coelho, S. M., and Koszul, R. 2020. instaGRAAL: chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffolder. *Genome Biol*, 21(1): 148.

Bhutkar, A., Schaeffer, S. W., Russo, S. M., Xu, M., Smith, T. F., and Gelbart, W. M. 2008. Chromosomal rearrangement inferred from comparisons of 12 Drosophila genomes. *Genetics*, 179(3): 1657–1680.

Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M., and González, J. 2021. Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of Drosophila melanogaster. *Molecular Ecology*, 30(4): 938–954.

Bolduc, F., Rosenfelt, C., *et al.* 2010. Fragile x mental retardation 1 and filamin a interact genetically in Drosophila long-term memory. *Frontiers in Neural Circuits*, 3: 751.

Boughdad, A., Haddi, K., El Bouazzati, A., Nassiri, A., Tahiri, A., El Anbri, C., Eddaya, T., Zaid, A., and Biondi, A. 2021. First record of the invasive spotted wing Drosophila infesting berry crops in Africa. *Journal of Pest Science*, 94(2): 261–271.

Bradshaw, C. J. A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., and Courchamp, F. 2016. Massive yet grossly underestimated global costs of invasive insects. *Nature Communications*, 7(1): 12986.

Bruce, T. J. A. 2010. Tackling the threat to food security caused by crop pests in the new millennium. *Food Security*, 2(2): 133–141.

Cabanettes, F. and Klopp, C. 2018. D-GENIES : Dot plot large GENomes in an interactive, efficient and simple way. *PeerJ Preprints*, 6: e26567v1.

Calabria, G., Máca, J., Bächli, G., Serra, L., and Pascual, M. 2012. First records of the potential pest species drosophila suzukii (diptera: Drosophilidae) in europe. *Journal of Applied entomology*, 136(1-2): 139–147.

Camus, L., Gautier, M., and Boitard, S. 2024. Predicting species invasiveness with genomic data: Is genomic offset related to establishment probability? *Evolutionary Applications*, 17(6): e13709.

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., and Keller, S. R. 2020. Genomic Prediction of (Mal)Adaptation Across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 51(1): 245–269.

Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. 2020. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3*, 10(4): 1361–1374.

Chang, C.-H., Chavan, A., Palladino, J., Wei, X., Martins, N. M. C., Santinello, B., Chen, C.-C., Erceg, J., Beliveau, B. J., Wu, C.-T., Larracuente, A. M., and Mellone, B. G. 2019. Islands of retroelements are major components of drosophila centromeres. *PLOS Biology*, 17(5): 1–40.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17): i884–i890.

Chen, W., Liu, Z., Li, T., Zhang, R., Xue, Y., Zhong, Y., Bai, W., Zhou, D., and Zhao, Z. 2014. Regulation of drosophila circadian rhythms by mirna let-7 is mediated by a regulatory cycle. *Nature communications*, 5(1): 1–13.

Chen, X., Peterson, J., Nachman, R. J., and Ganetzky, B. 2012. Drosulfakinin activates ccklr-17d1 and promotes larval locomotion and escape response in drosophila. *Fly*, 6(4): 290–297.

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2): 170–175.

Chiu, J. C., Jiang, X., Zhao, L., Hamm, C. A., Cridland, J. M., Saelao, P., Hamby, K. A., Lee, E. K., Kwok, R. S., Zhang, G., Zalom, F. G., Walton, V. M., and Begun, D. J. 2013. Genome of Drosophila suzukii, the spotted wing drosophila. *G3*, 3(12): 2257–71.

Colautti, R. I. and Barrett, S. C. H. 2013. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. *Science*, 342(6156): 364–366.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2): giab008.

De Castro, S., Peronnet, F., Gilles, J.-F., Mouchel-Vielh, E., and Gibert, J.-M. 2018. bric à brac (bab), a central player in the gene regulatory network that mediates thermal plasticity of pigmentation in drosophila melanogaster. *PLoS genetics*, 14(8): e1007573.

Deans, C. and Hutchison, W. D. 2022. Propensity for resistance development in the invasive berry pest, spotted-wing drosophila (drosophila suzukii), under laboratory selection. *Pest management science*, 78(12): 5203–5212.

Deprá, M., Poppe, J., Schmitz, H., De Toni, D., and Valente, V. 2014. The first records of the invasive pest Drosophila suzukii in the South American continent. *J Pest Sci*, 87.

Diagne, C., Leroy, B., Vaissière, A.-C., Gozlan, R. E., Roiz, D., Jarić, I., Salles, J.-M., Bradshaw, C. J. A., and Courchamp, F. 2021. High and rising economic costs of biological invasions worldwide. *Nature*, 592(7855): 571–576.

Diesh, C., Stevens, G. J., Xie, P., De Jesus Martinez, T., Hershberg, E. A., Leung, A., Guo, E., Dider, S., Zhang, J., Bridge, C., Hogue, G., Duncan, A., Morgan, M., Flores, T., Bimber, B. N., Haw, R., Cain, S., Buels, R. M., Stein, L. D., and Holmes, I. H. 2023. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology*, 24(1): 74.

Drosopoulou, E., Gariou-Papalexiou, A., Karamoustou, E., Gouvi, G., Augustinos, A. A., Bourtzis, K., and Zacharopoulou, A. 2019. The chromosomes of Drosophila suzukii (Diptera: Drosophilidae): detailed photographic polytene chromosomal maps and in situ hybridization data. *Molecular Genetics and Genomics*, 294(6): 1535–1546.

Druet, T. and Gautier, M. 2017. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20): 5820–5841.

Dubos, N., Préau, C., Lenormand, M., Papuga, G., Monsarrat, S., Denelle, P., Le Louarn, M., Heremans, S., May, R., Roche, P., *et al.* 2022. Assessing the effect of sample bias correction in species distribution models. *Ecological Indicators*, 145: 109487.

Eben, A., Reifenrath, M., Briem, F., Pink, S., and Vogt, H. 2018. Response of d rosophila suzukii (d iptera: D rosophilidae) to extreme heat and dryness. *Agricultural and Forest Entomology*, 20(1): 113–121.

Emiljanowicz, L. M., Hager, H. A., and Newman, J. A. 2017. Traits related to biological invasion: A note on the applicability of risk assessment tools across taxa. *NeoBiota*, 32: 31–64.

Enriquez, T. and Colinet, H. 2017. Basal tolerance to heat and cold exposure of the spotted wing drosophila, drosophila suzukii. *PeerJ*, 5: e3112.

Erban, T., Sopko, B., Talacko, P., Harant, K., Kadlikova, K., Halesova, T., Riddellova, K., and Pekas, A. 2019. Chronic exposure of bumblebees to neonicotinoid imidacloprid suppresses the entire mevalonate pathway and fatty acid synthesis. *Journal of proteomics*, 196: 69–80.

Estoup, A., Ravigné, V., Hufbauer, R., Vitalis, R., Gautier, M., and Facon, B. 2016. Is There a Genetic Paradox of Biological Invasion? *Annual Review of Ecology, Evolution, and Systematics*, 47(Volume 47, 2016): 51–72.

Fanning, P. D., Johnson, A. E., Luttinen, B. E., Espeland, E. M., Jahn, N. T., and Isaacs, R. 2019. Behavioral and physiological resistance to desiccation in spotted wing drosophila (diptera: Drosophilidae). *Environmental entomology*, 48(4): 792–798.

Fariello, M. I., Boitard, S., Mercier, S., Robelin, D., Faraut, T., Arnould, C., Recoquillay, J., Bouchez, O., Salin, G., Dehais, P., Gourichon, D., Leroux, S., Pitel, F., Leterrier, C., and San-Cristobal, M. 2017. Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: The local score approach. *Molecular Ecology*, 26(14): 3700–3714.

Feng, S., DeGrey, S. P., Guédot, C., Schoville, S. D., and Pool, J. E. 2023. Genomic Diversity Illuminates the Species History and Environmental Adaptation of Drosophila suzukii. *bioRxiv*, 2023.07.03.547576.

Fernando, M. D. A., Kounatidis, I., and Ligoxygakis, P. 2014. Loss of trabid, a new negative regulator of the drosophila immune-deficiency pathway at the level of tak1, reduces life span. *PLoS genetics*, 10(2): e1004117.

Fitzpatrick, M. C. and Keller, S. R. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18(1): 1–16.

Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., and Keller, S. R. 2021. Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources*, 21(8): 2749–2765.

Fraimout, A. and Monnet, A.-C. 2018. Accounting for intraspecific variation to quantify niche dynamics along the invasion routes of drosophila suzukii. *Biological Invasions*, 20(10): 2963–2979.

Fraimout, A., Debat, V., Fellous, S., Hufbauer, R. A., Foucaud, J., Pudlo, P., Marin, J.-M., Price, D. K., Cattel, J., Chen, X., Deprá, M., François Duyck, P., Guedot, C., Kenis, M., Kimura, M. T., Loeb, G., Loiseau, A., Martinez-Sañudo, I., Pascual, M., Polihronakis Richmond, M., Shearer, P., Singh, N., Tamura, K., Xuéreb, A., Zhang, J., and Estoup, A. 2017. Deciphering the Routes of invasion of Drosophila suzukii by Means of ABC Random Forest. *Molecular Biology and Evolution*, 34(4): 980–996.

Fukasawa, Y., Ermini, L., Wang, H., Carty, K., and Cheung, M.-S. 2020. LongQC: A Quality Control Tool for Third Generation Sequencing Long Read Data. *G3*, 10(4): 1193–1196.

Gain, C., Rhoné, B., Cubry, P., Salazar, I., Forbes, F., Vigouroux, Y., Jay, F., and François, O. 2023. A Quantitative Theory for Genomic Offset Statistics. *Molecular Biology and Evolution*, 40(6): msad140.

Garrison, E. and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*, 1207.3907.

Gautier, M. 2015. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, 201(4): 1555–1579.

Gautier, M. 2023. Efficient *k-mer* based curation of raw sequence data: application in *Drosophila suzukii*. *Peer Community Journal*, 3: e79.

Gautier, M., Yamaguchi, J., Foucaud, J., Loiseau, A., Ausset, A., Facon, B., Gschloessl, B., Lagnel, J., Loire, E., Parrinello, H., Severac, D., Lopez-Roques, C., Donnadieu, C., Manno, M., Berges, H., Gharbi, K., Lawson-Handley, L., Zang, L.-S., Vogel, H., Estoup, A., and Prud'homme, B. 2018. The Genomic Basis of Color Pattern Polymorphism in the Harlequin Ladybird. *Current Biology*, 28(20): 3296–3302.e7.

Gautier, M., Vitalis, R., Flori, L., and Estoup, A. 2022. f-Statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package poolfstat. *Molecular Ecology Resources*, 22(4): 1394–1416.

Gautier, M., Micol, T., Camus, L., Moazami-Goudarzi, K., Naves, M., Guéret, E., Engelen, S., Lemainque, A., Colas, F., Flori, L., and Druet, T. 2024. Genomic Reconstruction of the Successful Establishment of a Feralized Bovine Population on the Subantarctic Island of Amsterdam. *Molecular Biology and Evolution*, 41(7): msae121.

Haubrock, P. J., Turbelin, A. J., Cuthbert, R. N., Novoa, A., Taylor, N. G., Angulo, E., Ballesteros-Mejia, L., Bodey, T. W., Capinha, C., Diagne, C., Essl, F., Golivets, M., Kirichenko, N., Kourantidou, M., Leroy, B., Renault, D., Verbrugge, L., and Courchamp, F. 2021. Economic costs of invasive alien species across Europe. *NeoBiota*, 67: 153–190.

Haye, T., Girod, P., Cuthbertson, A., Wang, X., Daane, K., Hoelmer, K., Baroffio, C., Zhang, J., and Desneux, N. 2016. Current swd ipm tactics and their practical implementation in fruit crops across different regions around the world. *Journal of Pest Science*, 89: 643–651.

Hijmans, R. J. 2022. *raster: Geographic Data Analysis and Modeling*. R package version 3.5-15.

Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J. 2023. *dismo: Species Distribution Modeling*. R package version 1.3-14.

Hofmeister, N. R., Werner, S. J., and Lovette, I. J. 2021. Environmental correlates of genetic variation in the invasive European starling in North America. *Molecular Ecology*, 30(5): 1251–1263.

Horváth, V., Guirao-Rico, S., Salces-Ortiz, J., Rech, G. E., Green, L., Aprea, E., Rodeghiero, M., Anfora, G., and González, J. 2023. Gene expression differences consistent with water loss reduction underlie desiccation tolerance of natural drosophila populations. *BMC biology*, 21(1): 35.

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., Booth, B. W., Pfeiffer, B. D., George, R. A., Svirskas, R., Krzywinski, M., Schein, J., Accardo, M. C., Damia, E., Messina, G., Méndez-Lago, M., de Pablos, B., Demakova, O. V., Andreyeva, E. N., Boldyreva, L. V., Marra, M., Carvalho, A. B., Dimitri, P., Villasante, A., Zhimulev, I. F., Rubin, G. M., Karpen, G. H., and Celniker, S. E. 2015. The release 6 reference sequence of the drosophila melanogaster genome. *Genome Research*, 25(3): 445–58.

Hufbauer, R. A., Facon, B., Ravigné, V., Turgeon, J., Foucaud, J., Lee, C. E., Rey, O., and Estoup, A. 2012. Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications*, 5(1): 89–101.

Hulme, P. E. 2017. Climate change and biological invasions: evidence, expectations, and response options. *Biological Reviews of the Cambridge Philosophical Society*, 92(3): 1297–1313.

Hulme, P. E. 2021. Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth*, 4(5): 666–679.

International Trade Center 2024. Trade Map. Accessed: 2024-07-22.

IPCC 2024. International Plant Protection Convention country report: Notification of the detection of Drosophila suzukii, the Spotted Wing Drosophila in the Republic of South Africa. Technical Report ZAF-58/2, National Plant Protection Organization of South Africa.

Jakobs, R., Gariepy, T. D., and Sinclair, B. J. 2015. Adult plasticity of cold tolerance in a continental-temperate population of drosophila suzukii. *Journal of Insect Physiology*, 79: 1–9.

Jia, Z.-Q., Liu, D., Peng, Y.-C., Han, Z.-J., Zhao, C.-Q., and Tang, T. 2020. Identification of transcriptome and fluralaner responsive genes in the common cutworm spodoptera litura fabricius, based on rna-seq. *BMC genomics*, 21: 1–11.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., and Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1): 170122.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., and Brumby, S. P. 2021. Global land use/land cover with Sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4704–4707.

Kim, S. Y., Lohmueller, K. E., Albrechtsen, A., Li, Y., Korneliussen, T., Tian, G., Grarup, N., Jiang, T., Andersen, G., Witte, D., Jorgensen, T., Hansen, T., Pedersen, O., Wang, J., and Nielsen, R. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics*, 12(1): 231.

Kimura, M. T. 2004. Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia*, 140: 442–449.

Kiss, B., Szlanka, T., Zvara, Á., Žurovec, M., Sery, M., Kakaš, Š., Ramasz, B., Hegedűs, Z., Lukacsovich, T., Puskás, L., *et al.* 2013. Selective elimination/rnai silencing of fmrf-related peptides and their receptors decreases the locomotor activity in drosophila melanogaster. *General and Comparative Endocrinology*, 191: 137–145.

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. 2014. Angsd: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1): 356.

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., and Zdobnov, E. M. 2023. Orthodb v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, 51(D1): D445–D451.

Kwadha, C. A., Okwaro, L. A., Kleman, I., Rehermann, G., Revadi, S., Ndlela, S., Khamis, F. M., Nderitu, P. W., Kasina, M., George, M. K., Kithusi, G. G., Mohamed, S. A., Lattorff, H. M. G., and Becher, P. G. 2021. Detection of the spotted wing drosophila, Drosophila suzukii, in continental sub-Saharan Africa. *Journal of Pest Science*, 94(2): 251–259.

Landis, G. N., Bhole, D., and Tower, J. 2003. A search for doxycycline-dependent mutations that increase drosophila melanogaster life span identifies the vhasfd, sugar baby, filamin, fwd and cctl genes. *Genome Biology*, 4: 1–14.

Láruson, A. J., Yeaman, S., and Lotterhos, K. E. 2020. The Importance of Genetic Redundancy in Evolution. *Trends in Ecology & Evolution*, 35(9): 809–822.

Lavagnino, N. J., CICHÓN, L. I., DE LA VEGA, G. J., GARRIDO, S. A., FANARA, J. J., *et al.* 2018. New records of the invasive pest drosophila suzukii (matsumura)(diptera: Drosophilidae) in the south american continent. *Revista de la Sociedad Entomológica Argentina*, 77(1): 27.

Lewald, K. M., Abrieux, A., Wilson, D. A., Lee, Y., Conner, W. R., Andreazza, F., Beers, E. H., Burrack, H. J., Daane, K. M., Diepenbrock, L., Drummond, F. A., Fanning, P. D., Gaffney, M. T., Hesler, S. P., Ioriatti, C., Isaacs, R., Little, B. A., Loeb, G. M., Miller, B., Nava, D. E., Rendon, D., Sial, A. A., Bezerra da Silva, C. S., Stockton, D. G., Van Timmeren, S., Wallingford, A., Walton, V. M., Wang, X., Zhao, B., Zalom, F. G., and Chiu, J. C. 2021. Population genomics of Drosophila suzukii reveal longitudinal population structure and signals of migrations in and out of the continental United States. *G3*, 11(12): jkab343.

Li, H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37(23): 4572–4574.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16): 2078–9.

Lind, B. M. and Lotterhos, K. E. 2024. The limits of predicting maladaptation to future environments with genomic data. *bioRxiv*, 2024.01.30.577973.

Little, C. M., Chapman, T. W., and Hillier, N. K. 2020. Plasticity is key to success of drosophila suzukii (diptera: Drosophilidae) invasion. *Journal of Insect Science*, 20(3): 5.

Lotterhos, K. E. 2023. The paradox of adaptive trait clines with nonclinal patterns in the underlying genes. *Proceedings of the National Academy of Sciences*, 120(12): e2220313120.

Ma, L., Cao, L.-J., Hoffmann, A. A., Gong, Y.-J., Chen, J.-C., Chen, H.-S., Wang, X.-B., Zeng, A.-P., Wei, S.-J., and Zhou, Z.-S. 2020. Rapid and strong population genetic differentiation and genomic signatures of climatic adaptation in an invasive mealybug. *Diversity and Distributions*, 26(5): 610–622.

Mainka, S. A. and Howard, G. W. 2010. Climate change and invasive species: double jeopardy. *Integrative Zoology*, 5(2): 102–111.

Maino, J. L., Schouten, R., and Umina, P. 2021. Predicting the global invasion of drosophila suzukii to improve australian biosecurity preparedness. *Journal of Applied Ecology*, 58(4): 789–800.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, 38(10): 4647–4654.

Matthey-Doret, C., Baudry, L., Bignaud, A., Cournac, A., Remi-Montagne, Guiglielmoni, N., Foutel-Rodier, T., and Scolari, V. F. 2020. hicstuff: Simple library/pipeline to generate and handle hi-c data.

Mazza, G., Tricarico, E., Genovesi, P., and Gherardi, F. 2014. Biological invaders are threats to human health: an overview. *Ethology Ecology & Evolution*.

Meisner, J. and Albrechtsen, A. 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, 210(2): 719–731.

Mérel, V., Gibert, P., Buch, I., Rodriguez Rada, V., Estoup, A., Gautier, M., Fablet, M., Boulesteix, M., and Vieira, C. 2021. The Worldwide Invasion of Drosophila suzukii Is Accompanied by a Large Increase of Transposable Element Load and a Small Number of Putatively Adaptive Insertions. *Molecular Biology and Evolution*, 38(10): 4252–4267.

Mollot, G., Pantel, J. H., and Romanuk, T. N. 2017. Chapter Two - The Effects of Invasive Species on the Decline in Species Richness: A Global Meta-Analysis. In D. A. Bohan, A. J. Dumbrell, and F. Massol, editors, *Advances in Ecological Research*, volume 56 of *Networks of Invasion: A Synthesis of Concepts*, pages 61–83. Academic Press.

Nair, R. R. and Peterson, A. T. 2023. Mapping the global distribution of invasive pest drosophila suzukii and parasitoid leptopilina japonica: implications for biological control. *PeerJ*, 11: e15222.

Ohashi, Y. Y., Haino-Fukushima, K., and Fuyama, Y. 1991. Purification and charachterization of an ovulation stimulating substance from the male accessory glands of drosophila suzukii. *Insect Biochemistry*, 21(4): 413–419.

Olazcuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., Diepenbrock, L. M., Kenis, M., Zhang, J., Chen, X., Borowiec, N., Facon, B., Vogt, H., Price, D. K., Vogel, H., Prud'homme, B., Estoup, A., and Gautier, M. 2020. A Whole-Genome Scan for Association with Invasion Success in the Fruit Fly Drosophila suzukii Using Contrasts of Allele Frequencies Corrected for Population Structure. *Molecular Biology and Evolution*, 37(8): 2369–2385.

Olazcuaga, L., Baltenweck, R., Leménager, N., Maia-Grondard, A., Claudel, P., Hugueney, P., and Foucaud, J. 2023. Metabolic consequences of various fruit-based diets in a generalist insect species. *eLife*, 12: e84370.

Ometto, L., Cestaro, A., Ramasamy, S., Grassi, A., Revadi, S., Siozios, S., Moretto, M., Fontana, P., Varotto, C., Pisani, D., Dekker, T., Wrobel, N., Viola, R., Pertot, I., Cavalieri, D., Blaxter, M., Anfora, G., and Rota-Stabelli, O. 2013. Linking Genomics and Ecology to Investigate the Complex Evolution of an Invasive Drosophila Pest. *Genome Biology and Evolution*, 5(4): 745–757.

Ørsted, I. V. and Ørsted, M. 2019. Species distribution models of the spotted wing drosophila (drosophila suzukii, diptera: Drosophilidae) in its native and invasive range reveal an ecological niche shift. *Journal of applied ecology*, 56(2): 423–435.

Paini, D. R., Sheppard, A. W., Cook, D. C., De Barro, P. J., Worner, S. P., and Thomas, M. B. 2016. Global threat to agriculture from invasive species. *Proceedings of the National Academy of Sciences*, 113(27): 7575–7579.

Paris, M., Boyer, R., Jaenichen, R., Wolf, J., Karageorgi, M., Green, J., Cagnon, M., Parinello, H., Estoup, A., Gautier, M., Gompel, N., and Prud'homme, B. 2020. Near-chromosome level genome assembly of the fruit pest Drosophila suzukii using long-read sequencing. *Scientific Reports*, 10(1): 11227.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. 2012. Ancient Admixture in Human History. *Genetics*, 192(3): 1065–1093.

Pedersen, B. S. and Quinlan, A. R. 2017. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, 34(5): 867–868.

Pfenninger, M., Reuss, F., Kiebler, A., Schönnenbeck, P., Caliendo, C., Gerber, S., Cocchiararo, B., Reuter, S., Blüthgen, N., Mody, K., Mishra, B., Bálint, M., Thines, M., and Feldmeyer, B. 2021. Genomic basis for drought resistance in European beech forests threatened by climate change. *eLife*, 10: e65532.

Poyet, M., Le Roux, V., Gibert, P., Meirland, A., Prevost, G., Eslin, P., and Chabrerie, O. 2015. The wide potential trophic niche of the asiatic fruit fly drosophila suzukii: the key of its invasion success in temperate europe? *PloS one*, 10(11): e0142785.

Rhoné, B., Defrance, D., Berthouly-Salazar, C., Mariac, C., Cubry, P., Couderc, M., Dequincey, A., Assoumanne, A., Kane, N. A., Sultan, B., Barnaud, A., and Vigouroux, Y. 2020. Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nature Communications*, 11(1): 5274.

Richier, B., Michard-Vanhée, C., Lamouroux, A., Papin, C., and Rouyer, F. 2008. The clockwork orange drosophila protein functions as both an activator and a repressor of clock gene expression. *Journal of biological rhythms*, 23(2): 103–116.

Rossi-Stacconi, M. V., Kaur, R., Mazzoni, V., Ometto, L., Grassi, A., Gottardello, A., Rota-Stabelli, O., and Anfora, G. 2016. Multiple lines of evidence for reproductive winter diapause in the invasive pest drosophila suzukii: useful clues for control strategies. *Journal of pest science*, 89: 689–700.

Ryan, G. D., Emiljanowicz, L., Wilkinson, F., Kornya, M., and Newman, J. A. 2016. Thermal tolerances of the spotted-wing drosophila drosophila suzukii (diptera: Drosophilidae). *Journal of Economic Entomology*, 109(2): 746–752.

Shearer, P. W., West, J. D., Walton, V. M., Brown, P. H., Svetec, N., and Chiu, J. C. 2016. Seasonal cues induce phenotypic plasticity of drosophila suzukii to enhance winter survival. *BMC ecology*, 16: 1–18.

Shumate, A. and Salzberg, S. L. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics*, 37(12): 1639–1643.

Simberloff, D. 2009. The role of propagule pressure in biological invasions. *Annual review of ecology, evolution, and systematics*, 40(1): 81–102.

Skoglund, P. and Jakobsson, M. 2011. Archaic human ancestry in east asia. *Proceedings of the National Academy of Sciences*, 108(45): 18301–18306.

Skotte, L., Korneliussen, T. S., and Albrechtsen, A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3): 693–702.

Smit, A., Hubley, R., and Green, P. 2023. *RepeatMasker Open-4.0*. Institute for Systems Biology.

Stockton, D., Wallingford, A., Rendon, D., Fanning, P., Green, C. K., Diepenbrock, L., Ballman, E., Walton, V. M., Isaacs, R., Leach, H., *et al.* 2019a. Interactions between biotic and abiotic factors affect survival in overwintering drosophila suzukii (diptera: Drosophilidae). *Environmental Entomology*, 48(2): 454–464.

Stockton, D. G., Brown, R., and Loeb, G. M. 2019b. Not berry hungry? discovering the hidden food sources of a small fruit specialist, drosophila suzukii. *Ecological Entomology*, 44(6): 810–822.

The UniProt Consortium 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1): D523–D531.

Turelli, M., Cooper, B. S., Richardson, K. M., Ginsberg, P. S., Peckenpaugh, B., Antelope, C. X., Kim, K. J., May, M. R., Abrieux, A., Wilson, D. A., Bronski, M. J., Moore, B. R., Gao, J.-J., Eisen, M. B., Chiu, J. C., Conner, W. R., and Hoffmann, A. A. 2018. Rapid global spread of wri-like wolbachia across multiple drosophila. *Current Biology*, 28(6): 963–971.e8.

Turner, K. G., Ostevik, K. L., Grassa, C. J., and Rieseberg, L. H. 2021. Genomic Analyses of Phenotypic Differences Between Native and Invasive Populations of Diffuse Knapweed (Centaurea diffusa). *Frontiers in Ecology and Evolution*, 8.

Vaibhvi, V., Künzel, S., and Roeder, T. 2022. Hemocytes and fat body cells, the only professional immune cell types in drosophila, show strikingly different responses to systemic infections. *Frontiers in immunology*, 13: 1040510.

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. 2019. Efficient architecture-aware acceleration of bwa-mem for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 314–324.

Walsh, D. B., Bolda, M. P., Goodhue, R. E., Dreves, A. J., Lee, J., Bruck, D. J., Walton, V. M., O'Neal, S. D., and Zalom, F. G. 2011. Drosophila suzukii (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. *Journal of Integrated Pest Management*, 2(1): G1–G7.

Werzner, A., Pavlidis, P., Ometto, L., Stephan, W., and Laurent, S. 2013. Selective sweep in the flotillin-2 region of european drosophila melanogaster. *PLoS One*, 8(2): e56629.

Winkler, A., Jung, J., Kleinhenz, B., and Racca, P. 2020. A review on temperature and humidity effects on Drosophila suzukii population dynamics. *Agricultural and Forest Entomology*, 22(3): 179–192.

Wood, D. E., Lu, J., and Langmead, B. 2019. Improved metagenomic analysis with kraken 2. *Genome Biology*, 20(1): 257.

Wu, F., Deng, B., Xiao, N., Wang, T., Li, Y., Wang, R., Shi, K., Luo, D.-G., Rao, Y., and Zhou, C. 2020. A neuropeptide regulates fighting behavior in drosophila melanogaster. *Elife*, 9: e54229.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., and Yu, G. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3): 100141.

Yang, F., Crossley, M. S., Schrader, L., Dubovskiy, I. M., Wei, S.-J., and Zhang, R. 2022. Polygenic adaptation contributes to the invasive success of the Colorado potato beetle. *Molecular Ecology*, 31(21): 5568–5580.

Zhou, C., McCarthy, S. A., and Durbin, R. 2022. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, 39(1): btac808.