Limited population structure but signals of recent selection in introduced African Fig Fly *(Zaprionus indianus*) in North America

Priscilla A. Erickson[1]*, Alyssa Bangerter[2], Ansleigh Gunter[1], Nikolaos T. Polizos[3], and Alan O. Bergland[2]

1. University of Richmond, Richmond, Virginia
2. University of Virginia, Charlottesville, Virginia
3. University of Miami, Coral Gables, Florida

**Corresponding author; perickso@richmond.edu*

Running title

Genomics of invasive *Zaprionus indianus*

Abstract

Invasive species have devastating consequences for human health, food security, and the environment. Many invasive species adapt to new ecological niches following invasion, but little is known about the early steps of adaptation. Here we examine population genomics of a recently introduced drosophilid in North America, the African Fig Fly, *Zaprionus indianus*. This species is likely intolerant of subfreezing temperatures and recolonizes temperate environments yearly. We generated a new chromosome-level genome assembly for *Z. indianus*. Using resequencing of over 200 North American individuals collected over four years in temperate Virginia, plus a single collection from subtropical Florida, we tested for signatures of recolonization, population structure, and adaptation within invasive populations. We show founding populations are sometimes small and contain close genetic relatives, yet temporal population structure and differentiation of populations is mostly absent across recurrent recolonization events. Although we find limited signals of genome-wide spatial or temporal population structure, we identify haplotypes on the X chromosome that are repeatedly differentiated between Virginia and Florida populations. These haplotypes show signatures of natural selection and are not found in African populations. We also find evidence for several large structural polymorphisms segregating within North America populations and show X chromosome evolution in invasive populations is strikingly different from the autosomes. These results show that despite limited population structure, populations may rapidly evolve genetic differences early in an invasion. Further uncovering how these genomic regions influence invasive potential and success in new environments will advance our understanding of how organisms evolve in changing environments.

Article Summary

Invasive species (organisms that have been moved outside their natural range by human activities) can cause problems for both humans and the environment. We studied the genomes

of over 200 individuals of a newly invasive fruit fly in North America, the African Fig Fly. We found genetic evidence that these recently introduced flies may be evolving in their new environments, which could make them stronger competitors and more likely to become pests.

Introduction

Understanding how species expand and adapt to new environments in an era of climate change and global commerce is central to controlling the spread of disease (Altizer et al. 2013; Hoberg and Brooks 2015), to maintaining crop security (Oerke 2006; Sutherst et al. 2011) and to preserving biodiversity (Bellard et al. 2012). Many organisms are moving to new, previously unoccupied ranges at rates that continue to accelerate (Ricciardi 2007; Seebens et al. 2015; Seebens et al. 2017; Platts et al. 2019; Sardain et al. 2019) due to changes in climate and habitat as well as anthropogenic introductions. Genetic adaptation to new environments may allow some vulnerable organisms to survive in new habitats but may also permit potentially harmful organisms to expand even further (Clements and Ditommaso 2011). The past two decades have produced a wealth of studies characterizing the genetic and genomic basis of adaptation in a variety of organisms, from experimental populations of microbes (Good et al. 2017; Nguyen Ba et al. 2019; Johnson et al. 2021) to natural populations of eukaryotes (Hancock et al. 2011; Jones et al. 2018; Barrett et al. 2019; Lovell et al. 2021; Schluter et al. 2021). Recent and ongoing invasions offer the opportunity to study rapid evolution and adaptation to new environments in nearly real-time (Koch et al. 2020; Pélissié et al. 2022; Parvizi et al. 2023; Soudi et al. 2023). Recently, genomics has helped trace the history and sources of many well-known invasions (Pélissié et al. 2022; Picq et al. 2023) and shown that genetic divergence and even local adaptation are common in invasive populations that have been established for decades or even centuries (Ma et al. 2020; Stuart et al. 2021; Li et al. 2023). However, much remains unknown about the genetic mechanisms that allow invasive organisms to colonize and thrive in new environments. A better understanding of adaptive pathways in invasion may assist in predicting the success of invasions and controlling their outcomes.

The African Fig Fly, *Zaprionus indianus*, serves as a unique model to study how invasion history and local environment influence patterns of genetic variation. The ongoing, recurrent invasion of *Z. indianus* in North America offers a premier opportunity to study the possibility of rapid genetic changes following invasion. The *Zaprionus* genus arose in Africa but Z. *indianus* was first described in India in 1970 (Gupta 1970), where it has adapted to a range of environments (da Mata et al. 2010). It is one of the most ecologically diverse drosophilids in Africa; its ability to utilize up to 80 different food sources (Yassin and David 2010) and its generation time of as few as ~13 days (Nava et al. 2007) likely fueled its spread around the world. In 1999, it was first detected in Brazil (Vilela 1999), where it subsequently spread and caused major damage to fig and berry crops as well as native fruit species (Leão and Tldon 2004; Oliveira et al. 2013; Roque et al. 2017; Zanuncio-Junior et al. 2018; Allori Stazzonelli et al. 2023). It was later found in Mexico and Central America in 2002-2003 (Markow et al. 2014) and eventually Florida in 2005 (Linde et al. 2006). In 2011-2012, its range expanded

northwards in eastern North America (Joshi et al. 2014; Timmeren and Isaacs 2014; Pfeiffer et al. 2019) and eventually reached as far north as Ontario (Renkema et al. 2013) and Minnesota (Holle et al. 2018). It has also recently been found in the Middle East, Europe, and Hawaii (Parchami-Araghi et al. 2015; Kremmer et al. 2017; Willbrand et al. 2018), suggesting that the invasion is ongoing. *Z. indianus* can damage fig and berry crops (Pfeiffer et al. 2019; Allori Stazzonelli et al. 2023), increasing concerns about its pest potential in its expanding range.

Despite its global success, *Z. indianus* males are sterile below 15 °C, making cold temperatures a limiting factor to their success (Araripe et al. 2004). Within the temperate environment of Virginia, the species exhibits strong seasonal fluctuations in abundance (Rakes et al. 2023). First detection in Virginia is usually in June or July, weeks after the appearance of other overwintering Drosophilids, and population sizes climb dramatically through the late summer and early fall, when it often dominates the drosophilid community in temperate orchards. Typically, the peak in early to mid-September is followed by a dip in abundance and then a second peak in October, suggesting a seasonal component to reproduction or fluctuations in factors influencing *Z. indianus'* relative fitness. However, despite its early post-colonization success, it does not appear to survive temperate winters; *Z. indianus* populations became undetectable in Virginia by early December (Rakes et al. 2023). In locations in Minnesota, Kansas, and the northeastern US, *Z. indianus* has been detected one year and then not the next, suggesting that the populations are not permanently established, but are extirpated by cold and re-introduced by stochastic dispersal processes (Holle et al. 2018; Gleason et al. 2019; Rakes et al. 2023). Therefore, *Z. indianus* likely repeatedly invades temperate environments and evolves for several generations in these new habitats, offering an opportunity to recurrently study the genetic impacts of invasion and post-colonization adaptation across multiple years of sampling.

Genetic studies *of Z. indianus* are limited but provide important context to understand its worldwide invasion. The invasion of North America likely resulted from separate founding events on the East and West coasts (Commar et al. 2012). Comeault *et al* (2020) showed that North American populations are genetically distinct from those from Africa. Invasive populations of *Z. indianus* have an approximately 30% reduction in genetic diversity relative to ancestral African populations (Comeault et al. 2020), though invasive populations of *Z. indianus* maintain levels of genetic diversity that are often higher than those of non-invasive congeners. Despite the loss of diversity, *Z. indianus* is extremely successful in temperate habitats (Rakes et al. 2023). Further studies demonstrated that genetically distinct populations from eastern and western Africa likely admixed prior to a single colonization of the Americas (Comeault et al. 2021). How the high degree of genetic diversity in invasive populations influences the potential for ongoing evolution in North America, which is in a critical early stage of invasion, remains understudied.

Here, we assembled and annotated a chromosome-level genome assembly for *Z. indianus* and used the newly improved genome to answer several questions with the whole genome sequences of over 200 North American flies collected from three locations over four

years. First, do recolonizing North American *Z. indianus* populations demonstrate spatial or temporal population structure and if so, do specific regions of the genome have an outsized contribution to population structure? Second, is the invasion and recolonization history recapitulated in population genetic data? And third, do temperate populations show signatures of selection relative to native and tropical invasive populations?

Materials and Methods

Hi-C based genome scaffolding

An inbred line was generated from flies originally captured from Carter Mountain Orchard, VA (37.9913° N, 78.4721° W) in 2018. Wild caught flies were reared in the lab for approximately one year prior to initiating isofemale lines. The offspring of the isofemale lines were propagated through 10 rounds of full-sib mating. The resulting lines were then passaged for approximately one additional year in the lab and the most vigorous remaining line ("24.2") was chosen for sequencing.

3rd instar larvae from a single inbred line were snap frozen in liquid nitrogen and sent to Dovetail corporation (now Cantata Bio, Scotts Valley, CA) for chromatin extraction, Hi-C sequencing and genome scaffolding. Briefly, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNAse I, chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30x sequence coverage.

The input *de novo* assembly was the *Z. indianus* "RCR04" PacBio assembly (assembly # ASM1890459v1) from Kim et al. (2021). This assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). Dovetail OmniC library sequences were aligned to the draft input assembly using *bwa* (Li and Durbin 2009). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. See Figure S1 for link density histogram of scaffolding data.

Annotation

Repeat families found in the genome assemblies of *Z. indianus* were identified de novo and classified using the software package *RepeatModeler* v. 2.0.1 (Flynn et al. 2020).

RepeatModeler depends on the programs *RECON* v. 1.08 (Bao and Eddy 2002) and *RepeatScout* v. 1.0.6 (Price et al. 2005) for the de novo identification of repeats within the genome. The custom repeat library obtained from *RepeatModeler* was used to discover, identify and mask the repeats in the assembly file using *RepeatMasker* v. 4.1.0 (Smit et al. 2015).

RNA sequencing was conducted on 3 replicates of 3rd instar larva and 3 replicates of mixed stage pupa that were snap frozen in liquid nitrogen. RNA extraction and sequencing was performed by GeneWiz (South Plainfield, NJ). New larval and pupal RNAseq reads were combined with adult RNA sequencing from Comeault et al. (2020) for annotation. Coding sequences from *D. grimshawi, D. melanogaster*, *D. pseudoobscura*, *D. virilis*, *Z. africanus*, *Z. indianus*, *Z. tsacasi* and *Z. tuberculatus* (Kim et al. 2021) were used to train the initial *ab initio* model for *Z. indianus* using the *AUGUSTUS* software v. 2.5.5 (Keller et al. 2011). Six rounds of prediction optimization were done with the software package provided by *AUGUSTUS*. The same coding sequences were also used to train a separate *ab initio* model for *Z. indianus* using *SNAP* (version 2006-07-28) (Korf 2004). RNAseq reads were mapped onto the genome using the *STAR* aligner software (version 2.7) (Dobin et al. 2013) and intron hints generated with the bam2hints tools within *AUGUSTUS*. *MAKER* v. 3.01.03 (Cantarel et al. 2008), SNAP and AUGUSTUS (with intron-exon boundary hints provided from RNAseq) were then used to predict for genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from *D. grimshawi, D. melanogaster, D. pseudoobscura, D. virilis, Z. africanus, Z. indianus, Z. tsacasi and Z. tuberculatus* to generate peptide evidence in the *MAKER* pipeline. Only genes that were predicted by both SNAP and AUGUSTUS were retained in the final gene sets. To help assess the quality of the gene prediction, AED scores were generated for each of the predicted genes as part of the *MAKER* pipeline. Genes were further characterized for their putative function by performing a BLAST search of the peptide sequences against the UniProt database. tRNA were predicted using the software *tRNAscan-SE* v. 2.05 (Chan and Lowe 2019). Transcriptome completeness was assessed with *BUSCO* v. 4.0.5 (Manni et al. 2021) using the eukaryota_odb10 list of 255 genes.

Wild fly collections

Flies were collected by aspiration and netting from Carter Mountain Orchard, VA (37.9913° N, 78.4721° W) in 2017-2020 and from Hanover Peach Orchard, VA (37.5694° N, 77.2660° W) in 2019-2020. Flies were sampled from Coral Gables, FL (25.7239° N, 80.2802° W) in June 2019 using traps baited with bananas, oranges, yeast, and red wine. Flies were frozen in 70% ethanol at -20°C (2017-2018) or dry at -80 °C (2019-2020) prior to sequencing. Collections performed in July and August were called “early season.” In 2019, the earliest collections were not made until September (typically when *Z. indianus* abundance peaks, Rakes *et al*. 2023), and were assigned “mid-season.” Collections from October and November were called “late season.” For some analyses, the mid-season collection and early collections were combined,

as they were the first collections available each year. See Table S1 for the number of individual flies sequenced from each location and timepoint.

Individual whole genome sequencing

The sex of each wild-caught fly was recorded, then DNA was extracted from individual flies using the DNAdvance kit (Beckman Coulter, Indianapolis, IN) in 96 well plates, including an additional RNAse treatment step. DNA concentration was measuring using the QuantIT kit (Invitrogen, Waltham, MA) and purified DNA was diluted to 1 ng/µL. Libraries were prepared from 1 ng of genomic DNA using a reduced-volume dual-barcoding Nextera (Illumina, San Diego, CA) protocol as previously described (Erickson et al. 2020). The libraries were quantified using the QuantIT kit and equimolar ratios of each individual DNA were combined for sequencing. The pooled library was size-selected for 500 bp fragments using a BluePippin gel cassette (Sage Sciences, Beverly, MA). The pooled libraries were sequenced in one Illumina NovaSeq 6000 lane using paired-end, 150 bp reads by Novogene (Sacramento, CA).

Existing raw reads from *Z. indianus* collections from North America, South America, and Africa (Comeault et al. 2020; Comeault et al. 2021) were downloaded from the SRA from BioProject number PRJNA604690. These samples were combined with the new sequence data and processed together with the same mapping and SNP-calling pipeline. Overlapping paired-end reads were merged with *BBMerge* v. 38.92 (Bushnell et al. 2017). Reads were mapped to the genome assembly described above using *bwa mem* v. 0.7.17 (Li and Durbin 2009). Bam files for merged and unmerged reads were combined, sorted and de-duplicated with *Picard* v. 2.26.2 (https://github.com/broadinstitute/picard).

We next used *Haplotype Caller* from *GATK* v. 4.2.0.0 (McKenna et al. 2010) to generate a gVCF for each individual. We built a GenomicsDBI database for each scaffold, then used this database to genotype each gVCF. We used *GATK's* hard filtering options to filter the raw SNPs based on previously published parameters (*--filter-expression "QD < 2.0 || FS > 60.0 || SOR > 3.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"*) (Comeault et al. 2020). We then removed SNPs within 20 bp of an indel from the output and removed all SNPs in regions identified by *RepeatMasker*. We analyzed several measures of individual and SNP quality using *VCFtools* v. 0.1.17 (Danecek et al. 2011). We removed 16 individuals with mean coverage < 7X or over 10% missing genotypes. Next, we filtered SNPs with mean depths <10 or > 50 across all samples. We removed individual genotypes supported by 6 or fewer reads or with more than 100 reads to produce a final VCF with 5,185,389 SNPs and 2,099,147 non-singleton SNPs. See Table S1 for the final number of individuals included in the analysis from each population. See Figure S2 for the average SNP depth per sampling time and location.

Sex chromosome and Muller element identification

samtools v. 1.12 (Li et al. 2009) was used to measure coverage and depth of mapped reads from individual sequencing. This analysis revealed that the five main scaffolds (all over 25 Mb

in length) had a mean depth of ~16X coverage in both males and females in our dataset, except for scaffold 3, which had ~16X coverage in females but ~8X coverage in males, suggesting it is the X chromosome (Figure S3). Some of the previously sequenced samples had no sex recorded, so we used the ratio of X chromosome reads (scaffold 3) to autosome (scaffolds 1, 2, 4 and 5) reads to assign sexes to those individuals. Individuals with a ratio greater than 0.8 were assigned female, and ratios less than 0.8 were assigned male (Figure S4). For two known-sex individuals, the sex recorded prior to sequencing did not match the sex based on coverage; for those two samples we used the coverage-based sex assignment for analyses. We used D-GENIES (Cabanettes and Klopp 2018) to create dot-plots comparing the *Z. indianus* and *D. melanogaster* genome (BDGP6.46, downloaded from ensemble.org) to confirm the sex chromosome identification and assign Muller elements to *Z. indianus* autosomes (Figure S5, Table S2). Five additional scaffolds had lengths over 1 Mb. Scaffold 8 is the dot chromosome (Muller element F) based on sequence comparison to *D. melanogaster* (Figure S5) and had similar coverage to the autosomes (Figure S3). Scaffolds 6,7,9, and 10 had reduced coverage (Figure S3) and contain mostly repetitive elements. Downstream SNP calling and population genetic analysis included the five large scaffolds (named chromosomes 1-5) and excluded all smaller scaffolds.

Related individuals

Preliminary exploration of population genetic data indicated that some individual samples may be close relatives. For downstream analyses, we used the–*king-cutoff 0.0625* argument in *Plink* v. 2.0 (Chang et al. 2015) to generate a list of unrelated individuals. This filtering removed 21 individuals from the dataset. To quantify relatedness between all individuals, we used the function *snpgdsibdKING* in *SNPRelate* v. 1.38.0 (Zheng et al. 2012) to determine the kinship coefficients and probability of zero identity by descent for pairs of individuals using autosomal SNPs. We used thresholds established in Thornton *et al.* (2012) to classify relatedness between individuals.

Population structure and F_{ST}

We conducted principal components analysis using the R package *SNPRelate* v. 1.38.0 (Zheng et al. 2012) in R v. 4.1.1 (R Core Team) using a vcf that excluded singleton SNPs. We LD pruned SNPs with a minor allele frequency of at least 0.05 using *SNPgdsLDpruning* with an LD threshold of 0.2 and then calculated principal components with *snpgdsPCA* using all four autosomes. For subsequent analyses, we repeated the LD pruning within subsets of the data (North America only, or Carter Mountain, VA only). We also calculated principal components using individual chromosomes; for the X chromosome, only females were used in the analysis. We used t-tests and one-way ANOVAs followed by Tukey post-hoc tests to compare PC values between sampling locations and time points.

We used *Plink* v. *1.9* (Purcell et al. 2007; Chang et al. 2015) to LD prune VCF files with parameters (*--indep-pairwise 1000 50 0.2*) and used *ADMIXTURE* v. 1.3.0 (Alexander and

Lange 2011) to evaluate population structure for each chromosome separately. For the X chromosome, only females were used. We tested up to *k*=10 genetic clusters and used cross-validation analysis to choose the optimal *k* for each chromosome separately.

We calculated F_{ST} between Florida samples and early season Virginia samples using the *snpgdsFST* function in *SNPRelate* for all SNPs with a minor allele frequency > 0.01. For the X chromosome, only females were used in F_{ST} calculations to ensure diploid genotypes. We used the same function to calculate genome-wide, pairwise F_{ST} between all Virginia collections using autosomal SNPs.

Testing for structural variants

We used *smoove* v. 0.2.6 (Pedersen et al. 2020) to identify and genotype insertions, deletions, and rearrangements in the paired-end sequencing data from all individuals as described in the documentation. We also used linkage disequilibrium (LD) of randomly sampled SNPs from each chromosome to visually inspect for linkage due to potential inversions. We generated a list of SNPs segregating in each focal population with no missing genotypes and randomly sampled 4,000 SNPs from each chromosome. We used the *snpgdsLDMat* function in *SNPRelate* to calculate LD between all pairs of SNPs. LD heatmaps were created with the *ggLD* package (https://github.com/mmkim1210/ggLD).

Estimation of historic population sizes

We used *smc*++ v. 1.15.4 (Terhorst et al. 2017) to estimate historic population sizes for several subpopulations of individuals using autosomal genotypes. We used individuals from each African location and used the earliest sampling available for each year and Virginia orchard. We used *vcf2smc* to prepare the input files for each autosome separately. We assigned each individual as the “distinguished individual” and ran the analysis using all possible combinations of distinguished individual as described in (Bemmels et al. 2021). We used 10-fold cross validation to estimate final model parameters with the option (*-cv –folds 10)*. We assumed a generation time of 0.08 years (~12 generations per year) based on Nava *et al*. (2007), which assumes year-round reproduction in tropical regions. We note that for Virginia populations experiencing temperate conditions in recent years, 12 generations per year is likely an overestimate due to the shortened breeding season.

Selection scan

We used *WhatsHap* v. 1.7 (Patterson et al. 2015) to perform read-based phasing of the full vcf including singletons. To polarize the vcf for the genome wide selection scan relative to the invasion, we reassigned the reference allele of the phased vcf as the allele that was most common across all African individuals sequenced in previous studies. We calculated allele frequencies using all African samples in *SNPRelate*, then used *vcf-info-annotator* (https://vatools.readthedocs.io/en/latest/index.html) to assign the “ancestral” allele in the INFO

column. Lastly we used *bcftools* v. 1.13 (Danecek et al. 2021) to make simplified vcfs containing only the GT and AA fields for each chromosome separately.

We used the R package *rehh* v. 3.2.2 (Gautier and Vitalis 2012) to conduct the selection scan using integrated haplotype homozygosity score. We split samples into four possible populations (Africa, Florida, all North America, Virginia only) and conducted the scans separately for each population using phased, polarized vcfs for each individual chromosome. We used the *haplo2hh*, *scan,* and *ihh2ihs* functions to implement the scan. For the X chromosome, we only used a single haplotype for each male in the dataset to avoid double counting haploid genotypes. Haplotypes under selection were visualized by plotting all SNPs with IHS > 5. LD between candidate SNPs was calculated in *SNPRelate*.

Genetic diversity statistics

Because we obtained variable sequencing coverage within and across populations (Figure S2) we used software designed for low coverage and missing data to analyze population genetic statistics in genomic windows. We used *pixy* v. 1.2.5 (Korunes and Samuk 2021) to calculate Pi, F_{ST} and D_{XY} in 5 kb windows. Samples were grouped by collection location and year or by collection location for different analyses. We used *ANGSD* v. 0.941 (Korneliussen et al. 2014) to calculate Tajima's D. We first calculated genotype likelihoods from the bam files using arguments *-doSaf* and *-GL*. We then calculated Tajima's D and theta using the folded site frequency spectrum across 5 kb windows with 5 kb steps as described in ANGSD documentation.

Data management and plotting

We used the R packages *foreach* (Microsoft and Weston 2017) and *data.table* (Dowle and Srinivasan 2019) for data management and manipulation and used *ggplot2* (Wickham 2016) for all plotting. The *ggpubfigs* (Steenwyk and Rokas 2021) and *viridis* (Garnier 2018) packages were used for color palettes.

Results and Discussion

Genome assembly and annotation

High quality genome assemblies and annotations are a critical component of tracking and controlling invasive species and understanding the potential evolution of invasive species in invaded ranges (Matheson and McGaughran 2022). We conducted Hi-C based scaffolding of a previously sequenced *Z. indianus* genome (Kim et al. 2021) to achieve a chromosome-level assembly. There were 1,014 scaffolds with an N50 of 26.6 Mb, an improvement from an N50 of 4.1-6.8 Mb in previous assemblies (Kim et al. 2021). The five main chromosomes (Figure S1, named in order of size from largest to smallest) varied in length from 25.7 to 32.3 Mb (total length of five main scaffolds = 146,062,119 bp), in agreement with *Z. indianus* karyotyping

(Gupta and Kumar 1987; Campos et al. 2007). Chromosome 3 was identified as the sex chromosome using sequencing coverage of known-sex individuals (Figure S3, S4) and sequence comparison dot-plots (Figure S5). See Table S2 for assignment of *Z. indianus* chromosomes to Muller elements based on alignment to the *D. melanogaster* genome.

The annotation using RNAseq from larvae, pupae, and adults predicted 13,162 transcripts and 13,075 proteins, with 93% of 255 benchmarking universal single copy orthologs (BUSCO) genes (Simão et al. 2015) identified as complete and an additional 1.2% of BUSCO genes identified as fragmented. This transcriptome-based completeness estimate is lower than the genome-based estimate of 99% complete (Kim et al. 2021) but is in line with other arthropod genomes (Feron and Waterhouse 2022). Within the 5 main scaffolds, 24.6% of sequences were repetitive; within the entire assembly including all smaller scaffolds, 41% were repetitive. The five main chromosomes contain 11,327 predicted mRNAs (87% of all predicted), including 99.5% of all complete BUSCO genes. This improved genome resource will be valuable for future evolutionary studies *of Z. indianus*, which is becoming an increasingly problematic pest in some regions of the world (Allori Stazzonelli et al. 2023).

Limited spatial or temporal population structure in North American Z. indianus

To study spatial and temporal patterns of genetic variation in the seasonally repeated invasion of *Z. indianus*, we resequenced ~220 individuals collected from two orchards in Virginia (Charlottesville and Richmond) from 2017-2020, as well as one population collected from Miami, Florida in 2019. Because temperate locations such as Virginia are thought to be recolonized by *Z. indianus* each year (Pfeiffer et al. 2019; Rakes et al. 2023), we sampled both early in the season (~July-August) and late in the season (~October-November) in each year to capture the founding event, population expansion, and potential adaptation to the temperate environment.

We were first interested in studying geographic and temporal variation in population structure in North American populations of *Z. indianus*. For this analysis, we incorporated previous sequencing data from the Western Hemisphere and Africa (Comeault et al. 2020). While previous studies have shown limited structure within North America (Comeault et al. 2020; Comeault et al. 2021), we wanted to test for structure using deeper sampling within introduced locations and with greater temporal resolution across the *Z. indianus* growing season (Rakes et al. 2023). As shown previously, in an autosome-wide principal component analysis, PC1 separated Western Hemisphere and African samples (Figure 1A; t-test: $t = 78.92$, $df = 36$, $p < 2 \times 10^{-16}$). However, with the increased sample size of North American flies relative to previous studies, PC2 separated North American samples into two clusters, explaining 8% of total variation. To focus on potential structure within invasive North American samples, we excluded the African samples and recalculated principal components. This analysis revealed little genome-wide differentiation of North American populations collected from different locations (Figure 1B; ANOVA $P > 0.05$ for PC1 and PC2), though the samples

did fall into three large groups based on PC1, which may be indicative of structural variation (Li and Ralph 2019); see below.

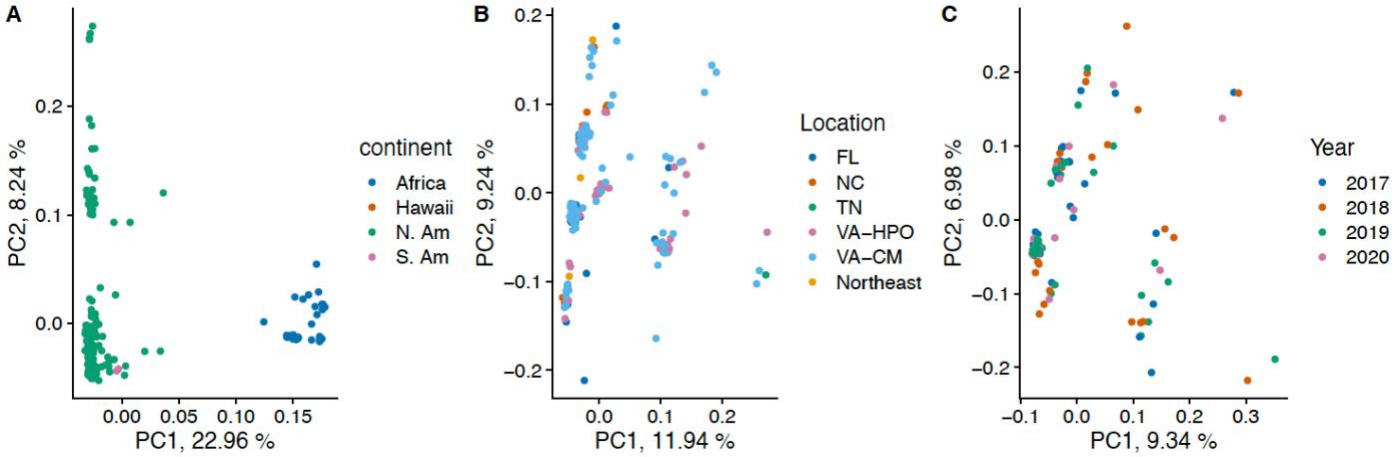

Figure 1: Principal component analysis of individual Z. indianus from this study and previous studies using autosomal SNPs. *(A) All unrelated individuals (n=247), color coded by continent/locale of collection. (B). All unrelated North American individuals (n=190), color coded by collection site; HPO and CM are two orchards in Virginia; Northeast refers to samples from NY, NJ, and PA. (C) All unrelated individuals from Carter Mountain, Virginia (n=110), color coded by year of collection. For each analysis, only the individuals shown in the plot were included in the PC calculation.*

North American samples clustered in single-chromosome PCA for chromosomes 1, 2 and 5, but these clusters generally did not correspond to sampling locations (Figure S6; PC2 separated the two Virginia orchards for chromosome 5; Tukey P = 0.01). Interestingly, PC1 separated Florida from both Virginia orchards for the X chromosome (Tukey P < 0.01 for each comparison), suggesting some degree of genetic differentiation on the X. Visual inspection of plots of PC3 and PC4 did not indicate additional geographic population structure (results not shown). The overall lack of genome- or chromosome-wide geographic population structure suggests that there is not a high degree of genetic differentiation between eastern North American populations spread over a latitudinal transect (~1600 km) encompassing distinct climates, but some localized patterns of population structure may exist on the X chromosome. Many invasive species evolve complex population structures in the invaded range due to a combination of bottlenecks, founder effects and rapid local adaptation (Koch et al. 2020; Atsawawaranunt et al. 2023; García-Escudero et al. 2023). On the other hand, some invasive species have more homogenous populations across widespread invaded ranges in eastern North America (Friedline et al. 2019; Barrett et al. 2023). A high rate of migration between orchards (occurring naturally or due to human-mediated transport) or large founding population sizes could result in a lack of geographic differentiation between populations.

We next hypothesized that founder effects during each recolonization event might lead to unique genetic compositions of temperate populations sampled in different years (Uller and

Leimu 2011). We calculated principal components using only samples collected from Carter Mountain, VA in 2017-2020. Surprisingly, in these samples, we saw no evidence of population structure between years across the genome (Figure 1C; ANOVA $P > 0.05$ for PC1 and PC2) or on individual chromosomes, except for chromosome 4, which showed subtle separation of some years (Figure S6; Tukey $P < 0.05$ for PC1: 2018 vs. 2019 and PC2: 2017 vs. 2019). These data suggest that the founding fly populations in Virginia are relatively homogeneous each year at a genome-wide scale. This result is consistent with the lack of spatial population structure and likewise could indicate large founding populations or ongoing migration. Alternatively, the Virginia population could be permanently established with little genetic differentiation year-to-year, though this possibility is not supported by field data (Rakes et al., 2023).

We used ADMIXTURE (Alexander and Lange 2011) to test for population structure using individuals from Africa, Florida, and the two focal Virginia orchards, calculating the most likely number of genetic clusters for each chromosome separately. Consistent with the PCA, the four autosomes each produced between two to four genetic groups, but there was no apparent geographic population structure, aside from African samples mostly belonging to different clusters from all North American samples for each chromosome (Figure 2). Notably, for chromosomes 1 and 2, many individuals showed ~50% ancestry assignment to different clusters, which could reflect genotypes for large structural rearrangements (see below). For the X chromosome, using females only, we identified structure within African samples as previously described (Comeault et al. 2020; Comeault et al. 2021) and a total of five genetic clusters within North American populations, including one of the African genetic groups which was found in Florida (Figure 2 third row; see orange grouping). X chromosomes have smaller effective population sizes in species with XY sex determination systems and often experience more extreme loss of genetic diversity upon population contraction (Ellegren 2009). The complex population structure seen on the X chromosome may be the result of this small population size or caused by selection on X-linked variants in different environments.

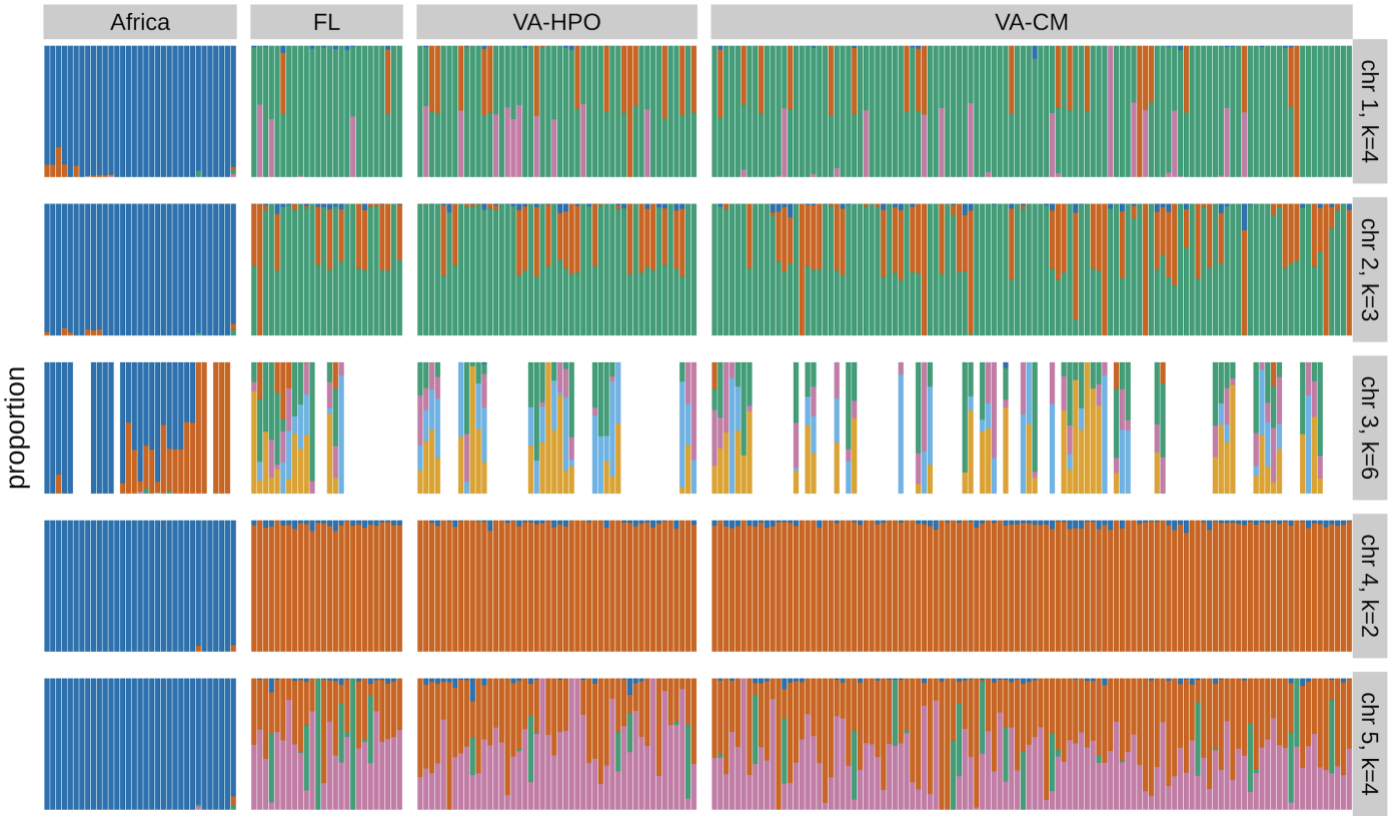

Figure 2: Admixture analysis of individual* Z. indianus *chromosomes from different locations. *Each column is an individual, and colors represent assignment to distinct genetic clusters. The most likely number of genetic clusters for each chromosome (k) was obtained with cross-validation analysis and is shown at right. For chromosome 3, the X chromosome, only female flies were used for admixture analysis, resulting in reduced sample size. FL=Miami, Florida, VA-HPO = Richmond, VA, VA-CM = Charlottesville, VA. African sequences represent five geographic locations and are taken from Comeault et al. (2020 & 2021).*

Structural polymorphism

The clustering of samples in the single-chromosome PCA (Figure S6), combined with many individuals showing ~50% assignment to genetic clusters (Figure 2), suggested that large structural variants may be segregating in *Z. indianus* (Li and Ralph 2019; Nowling et al. 2020). Analysis of paired-end sequencing data with *smoove* provided evidence of two large rearrangements on chromosome 1 located at 7.1 and 9.1 Mb; the genotypic combinations for these variants largely correlate with the clustering of samples in the PCA (Figure S7; PC1 correlation with variant at 7.1Mb: $P = 2 \times 10^{-9}$; PC1 correlation with variant at 9.1Mb: $P = 3 \times 10^{-5}$; PC2 correlation with variant at 9.1Mb: P = 0.001). Since chromosome 1 is the longest chromosome in our assembly, these rearrangements likely correspond to the complex *In(IV)EF* polymorphism, made up of two overlapping inversions (Ananina et al. 2007). *smoove* did not

identify large structural variants on chromosomes 2 or 5 whose genotypes correlated to the PCA clusters.

To look for evidence of structural variants via depressed recombination rates, we examined linkage disequilibrium (LD) from 4,000 randomly sampled SNPs on each chromosome. In North American samples, we discovered large blocks of LD spanning substantial portions of chromosomes 1, 2, 3, and 5 (Figure S8), potentially indicative of inversions (Fang et al. 2012; da Silva et al. 2019). However, there was no evidence of long-distance LD in these regions in the African samples (Figure S8). *smoove* did not identify inversions that corresponded to the sizes and locations of these linkage blocks. These results support the read-based evidence of a complex rearrangement on chromosome 1 (Figure S7) and suggest inversions on chromosomes 1, 2, 3, and 5 are segregating in North America but are relatively rare in Africa. Given the relative chromosome sizes in the genome assembly, the linkage blocks on chromosome 2 and 5 likely correspond to *In(V)B* and *In(II)*A, respectively (Ananina et al. 2007). The X chromosome has three described inversions in *Z. indianus* (Ananina et al. 2007), which may explain to the complex pattern of linkage observed in North American samples and the population structure observed for the X chromosome within North America (Figure 2). Major chromosomal polymorphisms are known to be important for local adaptation and phenotypic divergence in a wide variety of species (Joron et al. 2011; Küpper et al. 2016; Lee et al. 2016; Huang et al. 2020; Nunez et al. 2024), including inversions that facilitate invasive phenotypes (Galludo et al. 2018; Tepolt and Palumbi 2020; Tepolt et al. 2022; Ma et al. 2024). These inversions may have been present at low frequency in the bottlenecked population that founded *Z. indianus* populations in the Western Hemisphere, but then experienced subsequent selection in the invaded range. Alternatively, these polymorphisms may have arisen in a currently undescribed population and then been introduced to the Western Hemisphere.

Recolonization, bottlenecks and seasonal dynamics in Z. indianus

Invasive species typically experience a genetic bottleneck due to small founding population sizes (Barrett 2015; Estoup et al. 2016). We hypothesized that North American populations would show reduced effective population size (N_e) relative to African populations, and that Virginia populations would show a further, more recent reduction in N_e relative to Florida populations as the result of a secondary population bottleneck upon temperate recolonization. Our prediction was correct with respect to Africa vs North America: African populations show historical fluctuations but population sizes typically in the range of ~ 10^5-10^7 individuals. Interestingly, introduced populations in North America demonstrate population sizes that increased, decreased, then increased again in the past ~500 years. Comeault et al. (2021) suggested that introduced populations in the Americas are derived from a historically admixed population composed of both East African and West African flies, and the historic expansion of introduced populations might correspond to this admixture event. The subsequent drop in population size to 10^4-10^5 may then reflect a bottleneck following colonization of Brazil in the late 1990s (Yassin et al. 2008), followed by a rebound as introduced populations expanded.

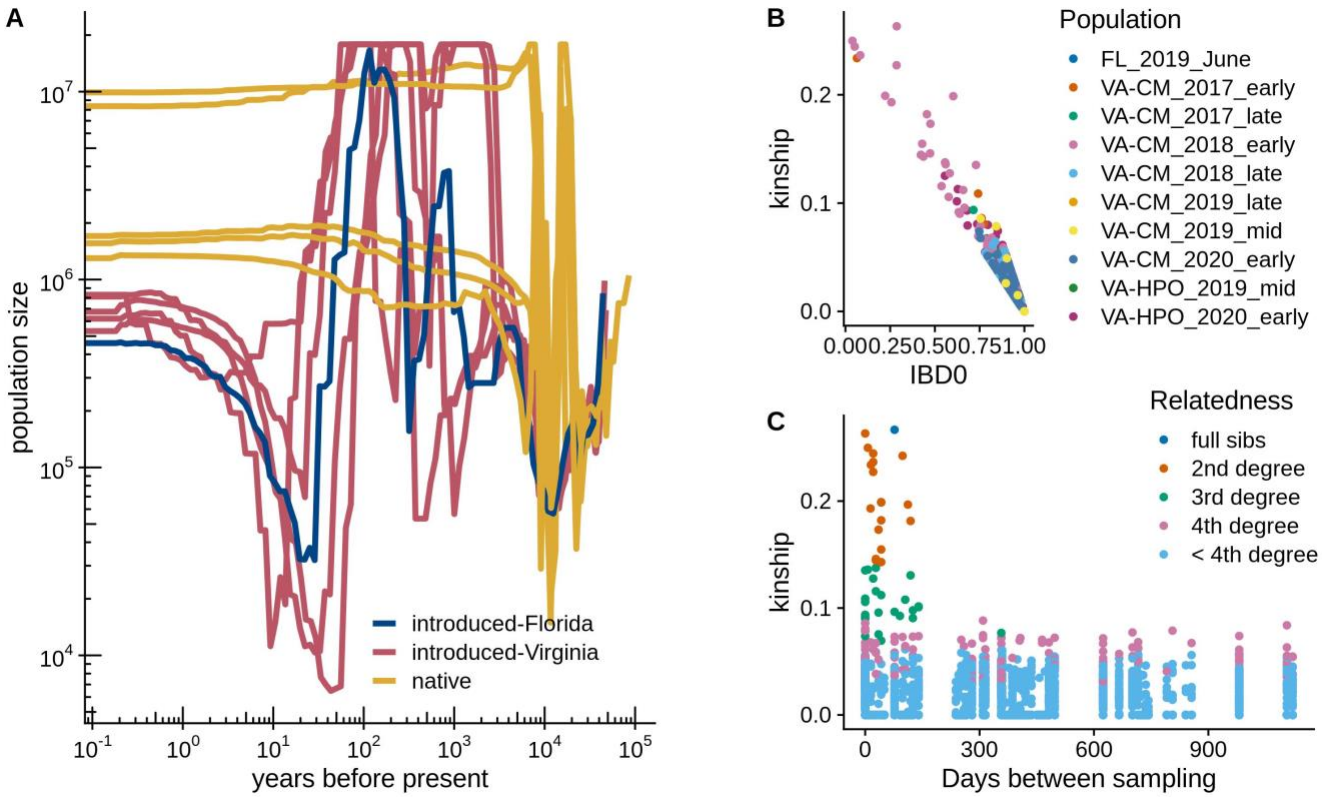

Figure 3: Demographic effects of bottlenecks in* Z. indianus *populations *A) Population history reconstruction with* smc++ *using autosomal genotypes. Introduced-Florida flies were collected in Miami in 2019. Introduced-Virginia flies were collected in the early-mid season (June-September) from two Virginia orchards in 2017-2020 (n=5 populations grouped by orchard and year). Native populations are distinct African populations (Kenya, Zambia, Senegal-Forest, Senegal-Desert, and Sao Tome [Comeault et al 2020]). B) Kinship and probability of zero identity by descent for pairs of individual flies from the same collection location and season within North America calculated with autosomal SNPs. C) Kinship coefficients for pairs of individual flies collected at Carter Mountain Orchard, Virginia, as a function of the number of days between sampling. Relatedness was assigned according to thresholds from (Thornton et al. 2012).*

Overall, the ancestral population sizes for Virginia and Florida were quite similar, and our prediction of reduced recent population sizes in Virginia relative to Florida was not well-supported. The minimum population sizes for Florida and Virginia (10^4-10^5) are larger than expected for a single small colonization event. Field data suggest founding populations in orchards are small and then rapidly expand (Rakes et al. 2023), suggesting that these large population sizes could be caused by ongoing gene flow from the source population after colonization, which is consistent with the lack of temporal population structure. Given our limited sample sizes and potential differences in the number of generations per year in

temperate and subtropical environments, detecting fine-scale differences in very recent population fluctuations may be beyond the detection ability of the software; *smc*++ becomes less accurate at timescales less than ~133 generations (Patton et al. 2019). Alternatively, the Virginia populations may be admixed populations reflecting individuals from multiple sources, producing larger effective population sizes than would otherwise be expected if recolonization occurs from a single source population undergoing a bottleneck. Admixture and gene flow ae important factors fueling genetic diversity and invasiveness in introduced species (McGaughran et al. 2024) and could potentially contribute to *Z. indianus'* local success following each recolonization event.

We additionally tested for bottlenecks by looking for inbreeding, which might be a product of small founding populations. Using two measures of genetic similarity, we discovered many pairs of related flies in our dataset (Figure 3B). Most dramatically, many flies collected in 2018 appeared to be close relatives (Figure S9). In collections from late July and early August 2018, 26 pairs of close relatives involving 13 individual flies were collected. Of those, 21 pairs of relatives were collected on different days, suggesting the relatedness was not solely a sampling artifact due to collecting closely related flies in the same microhabitat of the orchard. The effect of this apparent bottleneck was sometimes retained throughout the growing season, as a pair of full sibs was sampled 77 days apart in 2018, two pairs of second-degree relatives were sampled over 110 days apart in 2018, and two pairs of third-degree relatives were sampled 140 days apart in 2017 (Figure 3C). Given that *Z. indianus* are collected in small numbers early in the season (Rakes et al. 2023) and 2017 and 2018 had particularly early captures (Table S1), we suggest small founding population size followed by inbreeding could produce individuals sampled distantly in time that still show close genetic similarity. Alternatively, flies may live for a relatively long time or have slower generations in the wild, allowing us to capture close relatives separated by longer time periods. However, we note that the same pattern was not seen in every year of our collections, suggesting that colonization dynamics might differ dramatically from year to year, which is expected if recolonization occurs due to chance events each year.

The founder effect could generate temporal population structure by creating populations that were more similar within a year than between years, creating a positive relationship between F_{ST} and the elapsed time between collections (Bergland et al. 2014). We tested this prediction with samples collected from Carter Mountain, Virginia over four years, and there was no relationship between F_{ST} and the time between sampling (linear model, df=17, P=0.9, Figure S10). This lack of temporal differentiation is consistent with the PCA and the relatively large minimum population sizes previously described and could be produced by ongoing gene flow that eliminates any signal of a founder effect and inbreeding. This finding is distinct from trends observed in *D. melanogaster,* which experiences a strong overwintering bottleneck and shows temporal patterns of differentiation (Bergland et al. 2014; Nunez et al. 2024).

Repeated differentiation between Florida and Virginia populations

Despite the lack of genome-wide differentiation between different North American locales, we were interested in testing whether specific regions of the genome might differ between populations given environmental differences: Virginia has a temperate, seasonal climate with a relatively limited variety cultivated produce, and southern Florida is subtropical with an abundance and diversity of fruits throughout the year. Other factors such as diseases, insecticide use, and competing species may also differ widely between locales. In the absence of genome-wide population structure, genomic regions differentiated between these locations are candidates for local adaptation. We conducted a SNP-level F_{ST} analysis comparing all flies collected in Florida to those collected in the early season in Virginia over four years. We observed elevated F_{ST} throughout much of the X chromosome, with a pronounced peak at 690 kb (Figure 4A). This peak was observed when comparing the Florida collection to Virginia collections from both Charlottesville and Richmond across all four years of Virginia sampling both early and late in the season (Figure S11), suggesting that this differentiation is maintained through recurrent rounds of recolonization, potentially via local adaptation. Alternatively, this region could correspond to alleles that directly promote dispersion and/or invasion (Weinig et al. 2007) and are found at higher frequency in invaded populations. One limitation of our sampling strategy is that we have only a single year of sampling in Florida; additional data will be needed to determine whether the genetic composition of this population (and differentiation from Virginia) remains steady across multiple years. However, assuming that this result is not an artifact related to the Florida sample, an alternative possible explanation for the repeated differentiation seen between Florida and Virginia is that Virginia is recolonized by a source population that is genetically distinct from the southern Florida population we sampled here. Regardless, the finding implies localized genetic structure across a latitude gradient in North American.

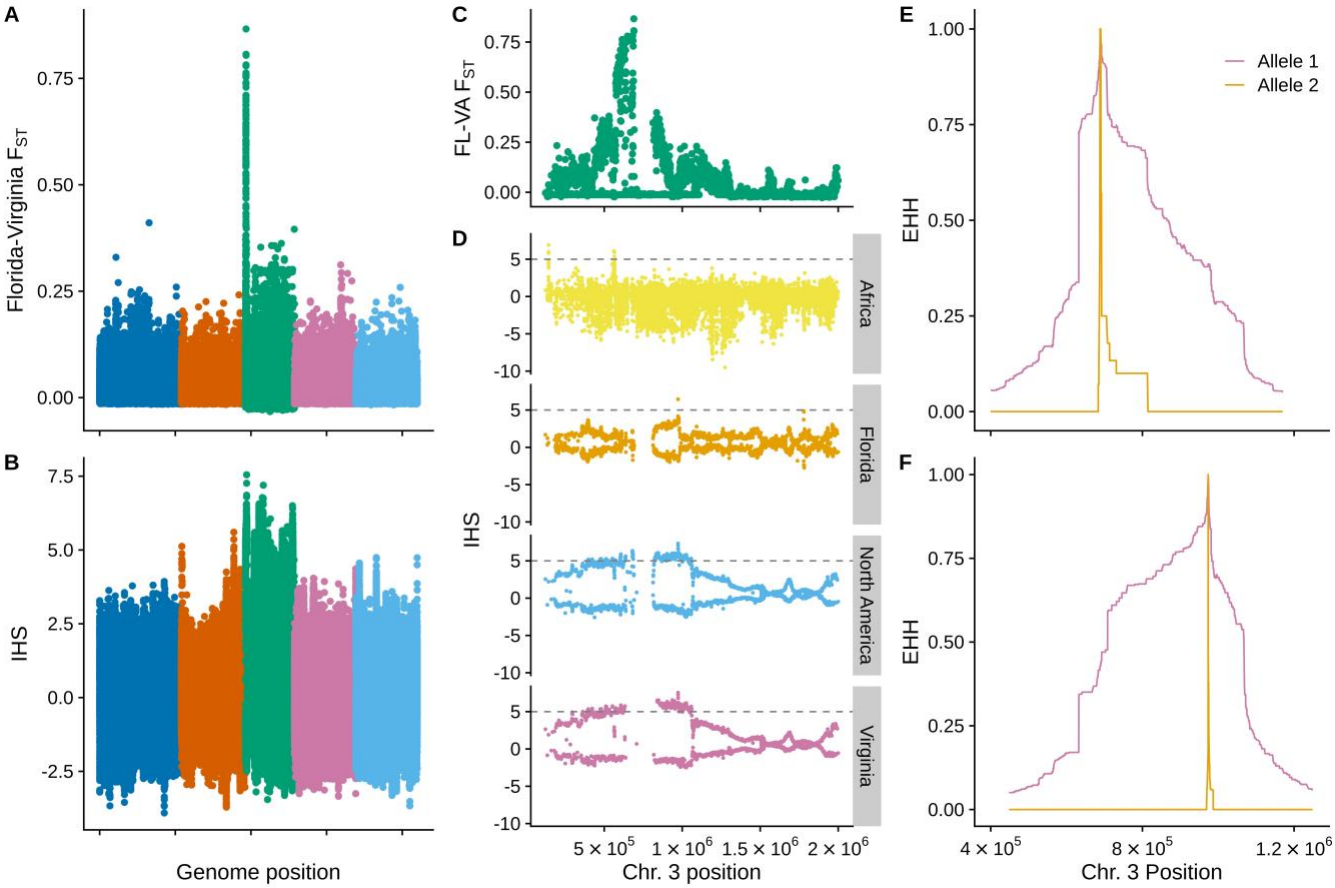

Figure 4: Signals of selection in temperate* Z. indianus *populations. *A) Genome-wide SNP-level F_{ST} comparing individual flies sampled in Florida (n=26) to all flies sampled in the early season in Virginia (n=123), color-coded by chromosome. Only females were used for the X chromosome (chromosome 3, green). B) Integrated haplotype homozygosity score (IHS) using all flies collected in Virginia. C) Zoomed in view of SNP-level FST between Florida and Virginia on chr 3: 0-2Mb. D) IHS for the X chromosome (chromosome 3:0-2 Mb) calculated separately for flies from Africa, Florida, all North America, and Virginia. Dashed line indicates IHS = 5 to facilitate comparisons between populations. E-F) Extended haplotype homozygosity for the two alleles of the SNP with highest F_{ST} (E; chr 3: 689841) and highest IHS (F; chr 3: 973443), calculated using all haplotypes from Virginia.*

Genomic signals of differentiation and selection

The elevated F_{ST} seen on the X chromosome raised the intriguing possibility that some genetic variation could potentially be under selection in temperate environments (Virginia) relative to subtropical Florida. We phased the paired-end sequencing data and calculated extended haplotype homozygosity (EHH) and IHS (integrated haplotype homozygosity score) using all Virginia individuals to look for long, shared haplotypes that can be signatures of selective sweeps (Sabeti et al. 2007). As in the F_{ST} analysis, we observed a region on chromosome 3

that stood out in this analysis with many SNPs with IHS > 5; this region overlapped with the F_{ST} peak (Figure 4, B-D). The peak F_{ST} SNP was approximately 300 kb away from the peak IHS SNP. We then repeated the IHS analysis using flies from Africa, Florida, and all North America (Virginia + Florida + Comeault (2020) locations) to determine whether this signature was unique to temperate populations. There was no signal of elevated IHS in African flies (Figure 4D, 1st row), suggesting this selective signature is unique to invasive populations. Further, this region showed a less substantial IHS peak when analyzing flies collected in Florida (Figure 4D, 2nd row) but was prominent when examining Virginia flies or all North American flies (Figure 4D, 3rd-4th rows), suggesting the signal of the selective sweep is primarily driven by individuals collected in temperate environments. Both the peak IHS SNP and the peak F_{ST} SNP showed evidence of long extended haplotypes characteristic of sweeps (Figure 4E-F). These results further support the possibility that this locus is advantageous to invasive potential or survival in temperate habitats.

We investigated this region of the genome by examining linkage disequilibrium (LD) and haplotype structure of the 400 SNPs with a Virginia IHS > 5 (Figure 5A). We discovered this region spanning ~700 kb has several large haplotype blocks in temperate North American samples (Figure 5C-D) and in Florida (Figure S12B), but these same haplotypes are not found in Africa (Figure 5C, Figure S12A), suggesting they are unique to introduced populations. In invasive copepods, haplotypes under selection in the invasive range are ancestral polymorphisms under balancing selection in the native range (Stern and Lee 2020). A similar situation was found for a balanced inversion polymorphism that fuels invasion in invasive crabs (Tepolt and Palumbi 2020; Tepolt et al. 2022). However, ancestral polymorphism selected in the invaded range does not appear to be the case in *Z. indianus*, as the haplotypes from North America were not found in any African flies. These novel haplotypes could be new mutations or derived due to hybridization/introgression from another species or divergent population; hybridization can be an important evolutionary force in invasive species (Ellstrand and Schierenbeck 2000; Fournier and Aron 2021). The *Zaprionus* genus shows signals of historic introgression, though *Z. indianus* was not directly implicated in a previous analysis (Suvorov et al. 2022). Therefore, two major haplotypes not found in Africa contribute to the differentiation of Florida and Virginia populations, though the source of these haplotypes remains to be determined. Though we focus on one genomic region, we note that most of the X chromosome shows elevated IHS scores (Figure 3B), and many SNPs on the X show F_{ST} > 0.25 between Virginia and Florida (Figure 3A). This observation is in line with the findings of Comeault et al 2021, who showed that many X-linked scaffolds showed signs of selection in invasive populations and is likely related to the presence of several inversions on this chromosome. We also note that our approach would not detect sweeps involving multiple alleles from standing variation (soft sweeps; (Messer and Petrov 2013; Garud et al. 2015)), which could be an important potential component of *Z. indianus* evolution given the high levels of genetic diversity found even in invasive populations (Avalos et al. 2017).

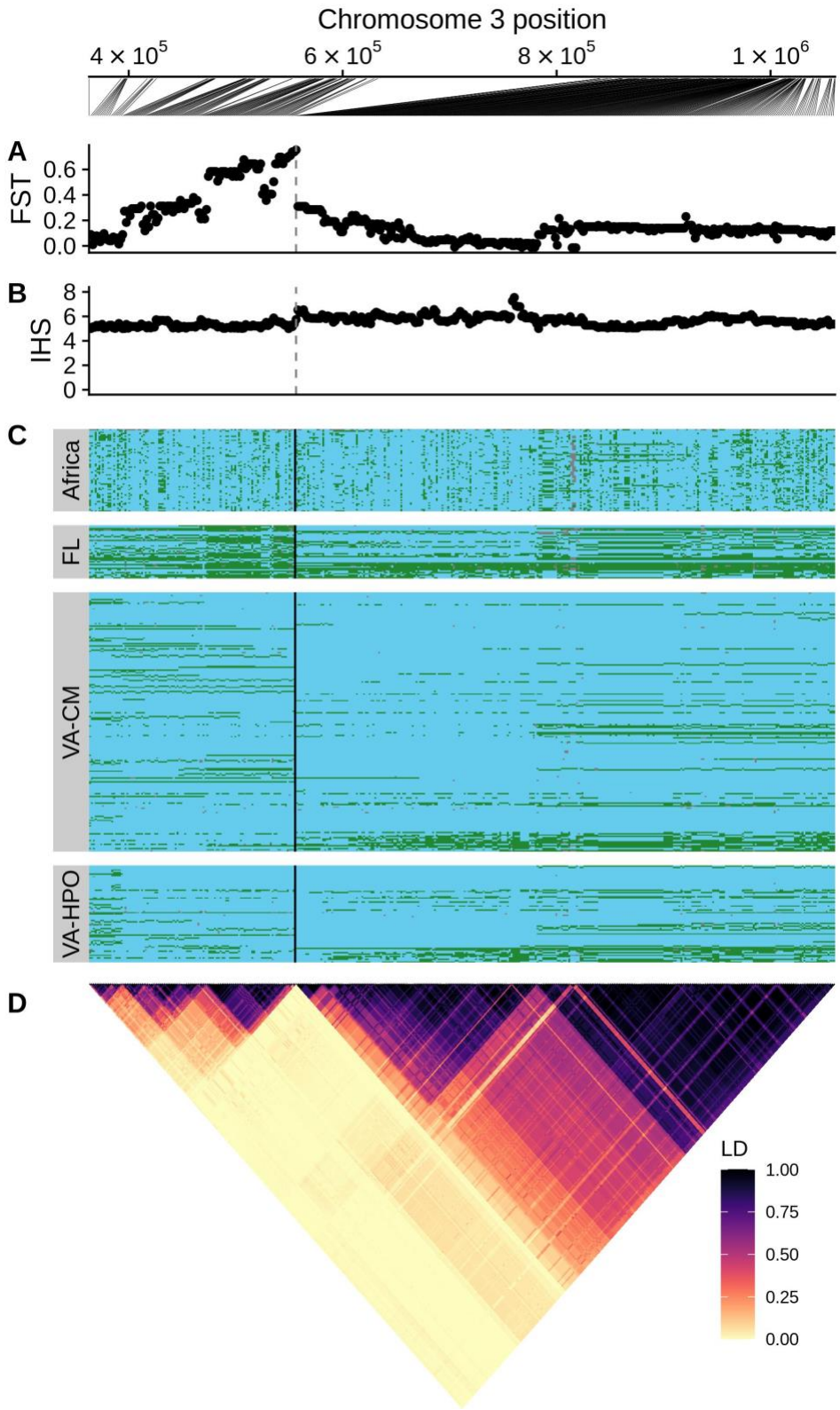

***Figure 5: Major haplotypes on the X chromosome with signals of selection and differentiation**. Only SNPs with IHS > 5 in Virginia (n=400) are shown in this figure for clarity; scale at top shows physical positions of SNPs, which are equally spaced in panels A-D. A) FST of individual SNPs comparing Florida and*

Virginia populations (see Figure 4). B) IHS for individual SNPs. C) Haplotypes: each horizontal row shows genotypes for a single haploid chromosome phased with read-backed phasing. Blue indicates the allele more common in African populations and green is the other allele. Missing genotypes are shown in gray. D) LD (R^2) for these SNPs in all North American flies, excluding Florida.

To explore population genetic signals around this highly divergent region of the X chromosome, we broadly grouped flies into three populations: Africa, Florida, and Virginia and calculated population genetic statistics in 5 kb non-overlapping windows for all females. This analysis confirmed two regions of relatively high F_{ST} between Florida and Virginia, though they are separated by a region of nearly zero differentiation within North America, as measured by F_{ST}, D_{xy}, and nucleotide diversity (Figure 6A-C, ~700-800kb). Virginia and Florida are both highly differentiated from Africa in this region, and it has negative Tajima's D in both Florida and Virginia (Figure 6D), potentially indicating recovery from a selective sweep in North America. The region of no divergence may represent a selective sweep of a haplotype that existed on two different genetic backgrounds that were subsequently favored in Florida and Virginia, producing a high degree of genetic differentiation in the surrounding sequences. The region with a potential sweep in North America contains ~6 genes, including the gene *yin/opt1*, which is important for absorption of dietary peptides in *D. melanogaster* (Roman et al. 1998). Allelic differences between African and invasive range flies in this gene could be involved in adaptation to new diets in new environments.

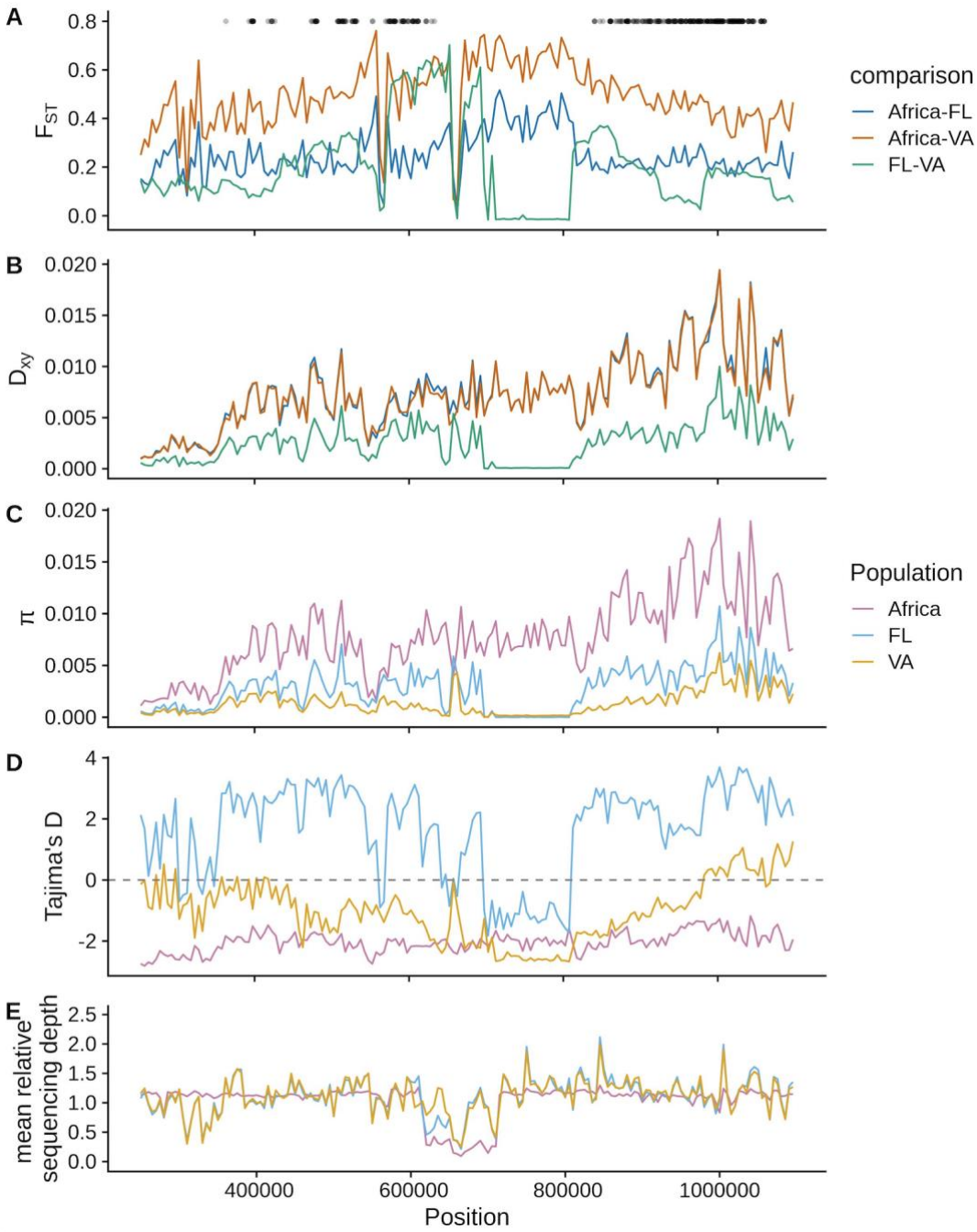

Figure 6: Population genetic statistics for the region surrounding a selected haplotypes on chromosome 3 (250-1,1000 kb). *All statistics were calculated for 5 kb, non-overlapping windows. Black points at the top indicate the locations of the 400 SNPs shown in Figure 5. A) F_{ST} comparing combinations of flies from Africa, Virginia (both focal orchards combined) and Florida. B) Absolute nucleotide divergence (D_{xy}) for the same comparisons. C) Nucleotide diversity (π) for each population. D) Tajima's D for the three populations. E) Average sequencing depth per window relative to the mean depth for the entire chromosome. Relative depths were averaged for all individuals in each population. See Figure S14 for whole-genome analysis of the same statistics.*

To confirm the patterns described above were not driven by genome assembly issues, we also examined the normalized depth of sequencing coverage relative to the chromosome average and discovered a region of variable coverage that overlaps the region of high F_{ST} between Florida and Virginia and is immediately adjacent to the region with zero divergence between Florida and Virginia (Figure 6E, Figure S13). This region (~600kb-700kb) has low coverage in Africa. In Florida and Virginia coverage varies from 0.5X -1X throughout the region. *smoove* identified paired-end evidence for a 52 kb duplication in this region, but genotype calls showed frequencies were similar in Virginia, Florida, and Africa. Combined with the sequencing depth data, these findings suggest copy number variation of these loci might contribute to the Florida-Virginia divergence, though long-read sequencing will likely be required to resolve the sequence variation. The region of elevated FST between Florida and Virginia and variable copy number contains several genes with neuronal and metabolic functions, offering exciting possibilities for future studies of the potential functional basis of this geographic divergence.

For comparison, we also examined the same five population signals genome-wide (Figure S14) and observed that the X chromosome is an outlier in many regards. Divergence between Africa and North American samples is greater on the X chromosome (Figure S14A). As previously described (Comeault et al. 2020; Comeault et al. 2021), the X chromosome has reduced genetic diversity relative to autosomes, especially in invaded populations (Figure S14B-C). Tajima's D is negative across the genome for African flies and mostly positive in North American autosomes, indicative of a strong bottleneck in North American flies. However, Tajima's D fluctuates between strongly positive and strongly negative in North American populations along the X chromosome (Figure S14D). This finding, combined with complex patterns of genetic ancestry on the X (Figure 2) and many regions with high haplotype homozygosity (Figure 4), suggest complex evolutionary dynamics on the X that warrant further investigation. These findings agree with Comeault et al. 2021, who found that regions under selection on the X chromosome typically showed higher divergence between invasive populations and African populations. Further global sampling and sequencing of X chromosomes with long reads to resolve inversion genotypes and CNVs may offer insight towards the role of X-linked genes in fueling the ongoing invasion of *Z. indianus*.

Conclusions

In addition to posing economic, health, and environmental threats, invasive species also serve as outstanding models for studying rapid evolution in new environments. Here we report an improved genome assembly and annotation for *Z. indianus*, an introduced drosophilid that is thought to repeatedly recolonize temperate environments each year and is a potential crop pest. We use it for a preliminary assessment of potential rapid evolution and genetic variation in the early stages of invasion. We show that recolonization is likely a stochastic process resulting in different evolutionary dynamics in different years, even within a single orchard. This finding demonstrates broad sampling is important for invasive species that are repeatedly

introduced or have multiple introduced populations that may undergo different evolutionary trajectories in different years or different locations. While some founding populations may be small, several population genetic patterns we observe could be explained by ongoing gene flow with the source population, or between temperate populations following recolonization, suggesting gene flow that spreads and maintains favorable alleles could be an important component in *Z. indianus*'s widespread success, as it is for many invasive species (Díez-del-Molino et al. 2013; Medley et al. 2015; Arredondo et al. 2018). Demographic simulations and additional whole genome data will be required to better describe the recent histories of and potential gene flow between invasive populations and to infer colonization routes within North America.

Though we find limited population structure across space or time in introduced North American populations, we find a region on the X chromosome that may have experienced a selective sweep in North America followed by separate sweeps in Virginia and Florida. Studying how genetic variation in this region of the genome influences survival in temperate environments will be an important direction of future research. We additionally find that the X chromosome has an unusually complex evolutionary history in *Z. indianus*. It may have several segregating inversions and CNVs, has strong signatures of selection, and shows regions of high divergence both between African and North American populations and within North America. Specifically, long-read sequencing strategies will be important to understand likely inversions both on the X and throughout the *Z. indianus* genome that are common in the invaded range. Large inversions can link together adaptive alleles and are often important drivers of evolution in rapidly changing environments (Thompson and Jiggins 2014), so these regions will be important to track over larger spatial and temporal scales in future studies.

These results underscore the complexity of genetic dynamics during invasions and the need for further studies to explore the adaptive potential and ecological impacts of *Z. indianus* in its invasive range. *Z. indianus* provides a unique system in which we can study independent invasion events across multiple years and locations. One limitation of our study is sample size for each year and location: our ability to estimate allele frequencies or detect subtle changes in allele frequencies across time or space is limited. Sampling strategies that incorporate more individuals, such as pooled sequencing (Bergland et al. 2014; Kapun et al. 2021; Machado et al. 2021; Nunez et al. 2024), will be required to detect these more subtle changes, if they occur, and to understand how they may contribute to rapid adaptation to new environments. The recurrent nature of *Z. indianus* colonization may also offer insight towards the predictability of rapid evolution of invasive species.

Data Availability

New individual sequencing data has been deposited in the SRA under project number # PRJNA991922. RNA sequencing from larval and pupal samples, and larval Hi-C data used for scaffolding are deposited under the same project number. The genome sequence has been deposited at DDBJ/ENA/GenBank under the accession JAUIZU000000000. The metadata for

all sequencing samples (including date and location of collection); the annotation information for transcripts, proteins and repeats; and VCFs of SNPs and structural variants have been deposited to Dryad: https://doi.org/10.5061/dryad.q2bvq83v3. All code to reproduce analyses has been deposited to Zenodo via Dryad. All code for analysis is also available at: https://github.com/ericksonp/Z.indianus_individual_sequencing/tree/main

Temporary reviewer dryad link: http://datadryad.org/stash/share/6td1mtLMrbgLL6IgyaEtpmqK4cigV0Vl2Hhqw9Aspvo

Acknowledgments

The authors acknowledge The University of Richmond's High Performance Computer (https://data.richmond.edu/About-HPC-at-UR/index.html) for providing computational resources that contributed to the results reported herein. We particularly thank George Flanagin for technical support. Preliminary analyses were conducted using resources provided by Research Computing at The University of Virginia (https://rc.virginia.edu). We thank the owners and managers of Carter Mountain Orchard and Hanover Peach Orchard for graciously allowing us to collect flies on their properties.

Funding

This work was funded by award #61-1673 from the Jane Coffin Childs Memorial Fund for Medical Research (to PAE), NIH NIGMS award # R15GM146208 (to PAE), NSF BIO-DEB (EP) award # 2145688 (to AOB), NIH NIGMS award # R35GM119686 (to AOB), and startup funds from the University of Richmond to PAE.

Author contributions

PAE, AB, AOB: conceptualization; PAE, AOB: funding; PAE, AG: investigation; PAE, AB, NP: resources; PAE: methodology, formal analysis, visualization, writing-original draft; PAE, AOB: writing-reviewing and editing

873

874 **References**

875

876 Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual
877 ancestry estimation. *BMC Bioinformatics* [Internet] 12:246. Available from:
878 https://doi.org/10.1186/1471-2105-12-246

879 Allori Stazzonelli E, Funes CF, Corral Gonzalez MN, Gibilisco SM, Kirschbaum DS. 2023.
880 Population fluctuation and infestation levels of Zaprionus indianus Gupta (Diptera:
881 Drosophilidae) in berry crops of northwestern Argentina | International Society for
882 Horticultural Science. *Acta Horticultura* [Internet]. Available from:
883 http://www.actahort.org/books/1381/1381_19.htm

884 Altizer S, Ostfeld RS, Johnson PTJ, Kutz S, Harvell CD. 2013. Climate Change and Infectious
885 Diseases: From Evidence to a Predictive Framework. *Science* 341:514–519.

886 Ananina G, Rohde C, David JR, Valente VLS, Klaczko LB. 2007. Inversion polymorphism and
887 a new polytene chromosome map of Zaprionus indianus Gupta (1970) (Diptera:
888 Drosophilidae). *Genetica* 131:117–125.

889 Araripe LO, Klaczko LB, Moreteau B, David JR. 2004. Male sterility thresholds in a tropical
890 cosmopolitan drosophilid, Zaprionus indianus. *Journal of Thermal Biology* [Internet]
891 29:73–80. Available from:
892 http://www.sciencedirect.com/science/article/pii/S0306456503000950

893 Arredondo TM, Marchini GL, Cruzan MB. 2018. Evidence for human-mediated range
894 expansion and gene flow in an invasive grass. *Proceedings of the Royal Society B:*
895 *Biological Sciences* [Internet] 285:20181125. Available from:
896 https://royalsocietypublishing.org/doi/full/10.1098/rspb.2018.1125

897 Atsawawaranunt K, Ewart KM, Major RE, Johnson RN, Santure AW, Whibley A. 2023. Tracing
898 the introduction of the invasive common myna using population genomics. *Heredity*
899 [Internet] 131:56–67. Available from: https://www.nature.com/articles/s41437-023-
900 00621-w

901 Avalos A, Pan H, Li C, Acevedo-Gonzalez JP, Rendon G, Fields CJ, Brown PJ, Giray T,
902 Robinson GE, Hudson ME, et al. 2017. A soft selective sweep during rapid evolution of
903 gentle behaviour in an Africanized honeybee. *Nat Commun* [Internet] 8:1550. Available
904 from: https://www.nature.com/articles/s41467-017-01800-0

905 Bao Z, Eddy SR. 2002. Automated De Novo Identification of Repeat Sequence Families in
906 Sequenced Genomes. *Genome Res.* [Internet] 12:1269–1276. Available from:
907 https://genome.cshlp.org/content/12/8/1269

908 Barrett CF, Corbett CW, Thixton-Nolan HL. 2023. A lack of population structure characterizes
909 the invasive Lonicera japonica in West Virginia and across eastern North America1,2.
910 *tbot* [Internet] 150:455–466. Available from: https://bioone.org/journals/the-journal-of-
911 the-torrey-botanical-society/volume-150/issue-3/TORREY-D-23-00007.1/A-lack-of-
912 population-structure-characterizes-the-invasive-Lonicera-japonica/10.3159/TORREY-D-
913 23-00007.1.full

914 Barrett RDH, Laurent S, Mallarino R, Pfeifer SP, Xu CCY, Foll M, Wakamatsu K, Duke-Cohan
915 JS, Jensen JD, Hoekstra HE. 2019. Linking a mutation to survival in wild mice. *Science*
916 363:499–504.

917 Barrett SCH. 2015. Foundations of invasion genetics: the Baker and Stebbins legacy.
918 *Molecular Ecology* [Internet] 24:1927–1941. Available from:
919 https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.13014

920 Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F. 2012. Impacts of climate
921 change on the future of biodiversity. *Ecology Letters* 15:365–377.

922 Bemmels JB, Mikkelsen EK, Haddrath O, Colbourne RM, Robertson HA, Weir JT. 2021.
923 Demographic decline and lineage-specific adaptations characterize New Zealand kiwi.
924 *Proc Biol Sci* 288:20212362.

925 Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic Evidence of
926 Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS*
927 *Genet* [Internet] 10:e1004775. Available from:
928 http://dx.doi.org/10.1371/journal.pgen.1004775

929 Bushnell B, Rood J, Singer E. 2017. BBMerge – Accurate paired shotgun read merging via
930 overlap. *PLOS ONE* 12:e0185056.

931 Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient
932 and simple way. *PeerJ* [Internet] 6:e4958. Available from: https://peerj.com/articles/4958

933 Campos SRC, Rieger TT, Santos JF. 2007. Homology of polytene elements between
934 Drosophila and Zaprionus determined by in situ hybridization in Zaprionus indianus.
935 *Genet Mol Res* 6:262–276.

936 Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M.
937 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model
938 organism genomes. *Genome Res.* [Internet] 18:188–196. Available from:
939 https://genome.cshlp.org/content/18/1/188

940 Chan PP, Lowe TM. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences.
941 *Methods Mol Biol* [Internet] 1962:1–14. Available from:
942 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6768409/

943 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation
944 PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

945 Clements DR, Ditommaso A. 2011. Climate change and weed adaptation: can evolution of
946 invasive plants lead to greater range expansion than forecasted? *Weed Research*
947 51:227–240.

948 Comeault AA, Kautt AF, Matute DR. 2021. Genomic signatures of admixture and selection are
949 shared among populations of Zaprionus indianus across the western hemisphere.
950 *Molecular Ecology* 30:6193–6210.

951 Comeault AA, Wang J, Tittes S, Isbell K, Ingley S, Hurlbert AH, Matute DR. 2020. Genetic
952 Diversity and Thermal Performance in Invasive and Native Populations of African Fig
953 Flies. *Molecular Biology and Evolution* 37:1893–1906.

954 Commar LS, Galego LG da C, Ceron CR, Carareto CMA. 2012. Taxonomic and evolutionary
955 analysis of Zaprionus indianus and its colonization of Palearctic and Neotropical
956 regions. *Genet Mol Biol* 35:395–406.

957 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
958 Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics*
959 [Internet] 27:2156–2158. Available from: https://doi.org/10.1093/bioinformatics/btr330

960 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
961 McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools.
962 *Gigascience* 10:giab008.

963 Díez-del-Molino D, Carmona-Catot G, Araguas R-M, Vidal O, Sanz N, García-Berthou E,
964 García-Marín J-L. 2013. Gene Flow and Maintenance of Genetic Diversity in Invasive
965 Mosquitofish (Gambusia holbrooki). *PLOS ONE* [Internet] 8:e82501. Available from:
966 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0082501

967 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
968 Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]
969 29:15–21. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/

970 Dowle M, Srinivasan A. 2019. data.table: Extension of `data.frame`. Available from:
971 https://CRAN.R-project.org/package=data.table

972 Ellegren H. 2009. The different levels of genetic diversity in sex chromosomes and autosomes.
973 *Trends in Genetics* [Internet] 25:278–284. Available from:
974 https://www.sciencedirect.com/science/article/pii/S0168952509000900

975 Ellstrand NC, Schierenbeck KA. 2000. Hybridization as a stimulus for the evolution of
976 invasiveness in plants? *Proceedings of the National Academy of Sciences* [Internet]
977 97:7043–7050. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.97.13.7043

978 Erickson PA, Weller CA, Song DY, Bangerter AS, Schmidt P, Bergland AO. 2020. Unique
979 genetic signatures of local adaptation over space and time for diapause, an ecologically
980 relevant complex trait, in Drosophila melanogaster. *PLOS Genetics* 16:e1009110.

981 Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B. 2016. Is There a Genetic
982 Paradox of Biological Invasion? *https://doi.org/10.1146/annurev-ecolsys-121415-*
983 *032116* [Internet]. Available from:
984 https://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-121415-032116

985 Fang Z, Pyhäjärvi T, Weber AL, Dawe RK, Glaubitz JC, González J de JS, Ross-Ibarra C,
986 Doebley J, Morrell PL, Ross-Ibarra J. 2012. Megabase-Scale Inversion Polymorphism in
987 the Wild Ancestor of Maize. *Genetics* [Internet] 191:883–894. Available from:
988 https://doi.org/10.1534/genetics.112.138578

Feron R, Waterhouse RM. 2022. Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *GigaScience* [Internet] 11:giac006. Available from: https://doi.org/10.1093/gigascience/giac006

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* [Internet] 117:9451–9457. Available from: https://www.pnas.org/doi/full/10.1073/pnas.1921046117

Fournier D, Aron S. 2021. Hybridization and invasiveness in social insects — The good, the bad and the hybrid. *Current Opinion in Insect Science* [Internet] 46:1–9. Available from: https://www.sciencedirect.com/science/article/pii/S2214574521000018

Friedline CJ, Faske TM, Lind BM, Hobson EM, Parry D, Dyer RJ, Johnson DM, Thompson LM, Grayson KL, Eckert AJ. 2019. Evolutionary genomics of gypsy moth populations sampled along a latitudinal gradient. *Molecular Ecology* [Internet] 0. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15069

Galludo M, Canals J, Pineda-Cirera L, Esteve C, Rosselló M, Balanyà J, Arenas C, Mestres F. 2018. Climatic adaptation of chromosomal inversions in Drosophila subobscura. *Genetica* [Internet] 146:433–441. Available from: https://doi.org/10.1007/s10709-018-0035-x

García-Escudero CA, Tsigenopoulos CS, Manousaki T, Tsakogiannis A, Marbà N, Vizzini S, Duarte CM, Apostolaki ET. 2023. Population genomics unveils the century-old invasion of the Seagrass Halophila stipulacea in the Mediterranean Sea. *Mar Biol* [Internet] 171:40. Available from: https://doi.org/10.1007/s00227-023-04361-7

Garnier S. 2018. viridis: Default Color Maps from “matplotlib.” Available from: https://CRAN.R-project.org/package=viridis

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLOS Genetics* [Internet] 11:e1005004. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1005004

Gautier M, Vitalis R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* 28:1176–1177.

Gleason JM, Roy PR, Everman ER, Gleason TC, Morgan TJ. 2019. Phenology of Drosophila species across a temperate growing season and implications for behavior. *PLOS ONE* 14:e0216601.

Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. 2017. The dynamics of molecular evolution over 60,000 generations. *Nature* 551:45–50.

Gupta JP. 1970. Description of a new species of Phorticella zaprionus (Drosophilidae) from India. *Proceedings of the Indian National Science Academy* 36B:62–70.

Gupta JP, Kumar A. 1987. Cytogenetics of Zaprionus indianus Gupta (Diptera: Drosophilidae): Nucleolar organizer regions, mitotic and polytene chromosomes and inversion

polymorphism. *Genetica* [Internet] 74:19–25. Available from: https://doi.org/10.1007/BF00055090

Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux F, Bergelson J. 2011. Adaptation to Climate Across the Arabidopsis thaliana Genome. *Science* 334:83–86.

Hoberg EP, Brooks DR. 2015. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370:20130553.

Holle SG, Tran AK, Burkness EC, Ebbenga DN, Hutchison WD. 2018. First Detections of Zaprionus indianus (Diptera: Drosophilidae) in Minnesota. *ents* 54:99–102.

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology* [Internet] 29:2535–2549. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15428

Johnson MS, Gopalakrishnan S, Goyal J, Dillingham ME, Bakerlee CW, Humphrey PT, Jagdish T, Jerison ER, Kosheleva K, Lawrence KR, et al. 2021. Phenotypic and molecular evolution across 10,000 generations in laboratory budding yeast populations.Verstrepen KJ, Wittkopp PJ, Verstrepen KJ, Hodgins-Davis A, editors. *eLife* 10:e63910.

Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360:1355–1358.

Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* [Internet] 477:203–206. Available from: http://www.nature.com/nature/journal/v477/n7363/full/nature10341.html

Joshi NK, Biddinger DJ, Demchak K, Deppen A. 2014. First report of Zaprionus indianus (Diptera: Drosophilidae) in commercial fruits and vegetables in Pennsylvania. *J. Insect Sci.* 14:259.

Kapun M, Nunez JCB, Bogaerts-Márquez M, Murga-Moreno J, Paris M, Outten J, Coronado-Zamora M, Tern C, Rota-Stabelli O, Guerreiro MPG, et al. 2021. Drosophila Evolution over Space and Time (DEST) - A New Population Genomics Resource. *bioRxiv* [Internet]:2021.02.01.428994. Available from: https://www.biorxiv.org/content/10.1101/2021.02.01.428994v1

Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* [Internet] 27:757–763. Available from: https://doi.org/10.1093/bioinformatics/btr010

Kim BY, Wang JR, Miller DE, Barmina O, Delaney E, Thompson A, Comeault AA, Peede D, D'Agostino ER, Pelaez J, et al. 2021. Highly contiguous assemblies of 101 drosophilid

genomes.Coop G, Wittkopp PJ, Sackton TB, editors. *eLife* [Internet] 10:e66405. Available from: https://doi.org/10.7554/eLife.66405

Koch JB, Dupuis JR, Jardeleza M-K, Ouedraogo N, Geib SM, Follett PA, Price DK. 2020. Population genomic and phenotype diversity of invasive Drosophila suzukii in Hawai'i. *Biol Invasions* [Internet] 22:1753–1770. Available from: https://doi.org/10.1007/s10530-020-02217-5

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* [Internet] 5:59. Available from: https://doi.org/10.1186/1471-2105-5-59

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* [Internet] 15:356. Available from: https://doi.org/10.1186/s12859-014-0356-4

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* [Internet] 21:1359–1368. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13326

Kremmer L, David J, Borowiec N, Thaon M, Ris N, Poirie M, Gatti J-L. 2017. The African fig fly Zaprionus indianus: a new invasive pest in France? *Bulletin of Insectology* 70:57–62.

Küpper C, Stocks M, Risse JE, dos Remedios N, Farrell LL, McRae SB, Morgan TC, Karlionova N, Pinchuk P, Verkuil YI, et al. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet* [Internet] 48:79–83. Available from: http://www.nature.com/ng/journal/v48/n1/full/ng.3443.html

Leão BFD, Tldon R. 2004. Newly invading species exploiting native host-plants: the case of the African Zaprionus indianus (Gupta) in the Brazilian Cerrado (Diptera, Drosophilidae). *Annales de la Société entomologique de France (N.S.)* 40:285–290.

Lee YW, Fishman L, Kelly JK, Willis JH. 2016. A Segregating Inversion Generates Fitness Variation in Yellow Monkeyflower (Mimulus guttatus). *Genetics* [Internet] 202:1473–1484. Available from: http://www.genetics.org/content/202/4/1473

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* [Internet] 25:2078–2079. Available from: https://doi.org/10.1093/bioinformatics/btp352

Li H, Peng Y, Wang Y, Summerhays B, Shu X, Vasquez Y, Vansant H, Grenier C, Gonzalez N, Kansagra K, et al. 2023. Global patterns of genomic and phenotypic variation in the invasive harlequin ladybird. *BMC Biol* [Internet] 21:141. Available from: https://doi.org/10.1186/s12915-023-01638-7

Li H, Ralph P. 2019. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics* [Internet] 211:289–304. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6325702/

Linde K van der, Steck GJ, Hibbard K, Birdsley JS, Alonso LM, Houle D. 2006. FIRST RECORDS OF ZAPRIONUS INDIANUS (DIPTERA: DROSOPHILIDAE), A PEST SPECIES ON COMMERCIAL FRUITS FROM PANAMA AND THE UNITED STATES OF AMERICA. *flen* 89:402–404.

Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*:1–7.

Ma L, Cao L-J, Hoffmann AA, Gong Y-J, Chen J-C, Chen H-S, Wang X-B, Zeng A-P, Wei S-J, Zhou Z-S. 2020. Rapid and strong population genetic differentiation and genomic signatures of climatic adaptation in an invasive mealybug. *Diversity and Distributions* [Internet] 26:610–622. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.13053

Ma L-J, Cao L-J, Chen J-C, Tang M-Q, Song W, Yang F-Y, Shen X-J, Ren Y-J, Yang Q, Li H, et al. 2024. Rapid and Repeated Climate Adaptation Involving Chromosome Inversions following Invasion of an Insect. *Molecular Biology and Evolution* [Internet] 41:msae044. Available from: https://doi.org/10.1093/molbev/msae044

Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K, Fabian DK, Flatt T, González J, Karasov TL, et al. 2021. Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in Drosophila.Nordborg M, Wittkopp PJ, Nordborg M, editors. *eLife* [Internet] 10:e67577. Available from: https://doi.org/10.7554/eLife.67577

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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* [Internet] 38:4647–4654. Available from: https://doi.org/10.1093/molbev/msab199

Markow TA, Hanna G, Riesgo-Escovar JR, Tellez-Garcia AA, Richmond MP, Nazario-Yepiz NO, Laclette MRL, Carpinteyro-Ponce J, Pfeiler E. 2014. Population genetics and recent colonization history of the invasive drosophilid Zaprionus indianus in Mexico and Central America. *Biol Invasions* 16:2427–2434.

da Mata RA, Tidon R, Côrtes LG, De Marco P, Diniz-Filho JAF. 2010. Invasive and flexible: niche shift in the drosophilid Zaprionus indianus (Insecta, Diptera). *Biol Invasions* 12:1231–1241.

Matheson P, McGaughran A. 2022. Genomic data is missing for many highly invasive species, restricting our preparedness for escalating incursion rates. *Sci Rep* [Internet] 12:13987. Available from: https://www.nature.com/articles/s41598-022-17937-y

McGaughran A, Dhami MK, Parvizi E, Vaughan AL, Gleeson DM, Hodgins KA, Rollins LA, Tepolt CK, Turner KG, Atsawawaranunt K, et al. 2024. Genomic Tools in Biological Invasions: Current State and Future Frontiers. *Genome Biology and Evolution* [Internet] 16:evad230. Available from: https://doi.org/10.1093/gbe/evad230

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.

Medley KA, Jenkins DG, Hoffman EA. 2015. Human-aided and natural dispersal drive gene flow across the range of an invasive mosquito. *Molecular Ecology* [Internet] 24:284–295. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.12925

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* [Internet] 28:659–669. Available from: https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(13)00207-3

Microsoft, Weston S. 2017. foreach: Provides Foreach Looping Construct for R. Available from: https://CRAN.R-project.org/package=foreach

Nava DE, Nascimento AM, Stein CP, Haddad ML, Bento JMS, Parra JRP. 2007. Biology, thermal requirements, and estimation of the number of generations of Zaprionus indianus (Diptera: Drosopholidae) for the main fig producing regions of Brazil. *flen* 90:495–501.

Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, Levy SF, Desai MM. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 575:494–499.

Nowling RJ, Manke KR, Emrich SJ. 2020. Detecting inversions with PCA in the presence of population structure. *PLOS ONE* [Internet] 15:e0240429. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240429

Nunez JCB, Lenhart BA, Bangerter A, Murray CS, Mazzeo GR, Yu Y, Nystrom TL, Tern C, Erickson PA, Bergland AO. 2024. A cosmopolitan inversion facilitates seasonal adaptation in overwintering Drosophila. *Genetics* [Internet] 226:iyad207. Available from: https://doi.org/10.1093/genetics/iyad207

Oerke E-C. 2006. Crop losses to pests. *The Journal of Agricultural Science* 144:31–43.

Oliveira CM, Auad AM, Mendes SM, Frizzas MR. 2013. Economic impact of exotic insect pests in Brazilian agriculture. *Journal of Applied Entomology* 137:1–15.

Parchami-Araghi M, Gilasian E, Keyhanian A. 2015. Olive infestation with Zaprionus indianus Gupta (Dip.: Drosophilidae) in northern Iran: a new host record and threat to world olive production. *Drosophila Information Service* 98:60–61.

Parvizi E, Dhami MK, Yan J, McGaughran A. 2023. Population genomic insights into invasion success in a polyphagous agricultural pest, Halyomorpha halys. *Molecular Ecology* [Internet] 32:138–151. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16740

Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, Schönhuth A. 2015. WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J Comput Biol* 22:498–509.

Patton AH, Margres MJ, Stahlke AR, Hendricks S, Lewallen K, Hamede RK, Ruiz-Aravena M, Ryder O, McCallum HI, Jones ME, et al. 2019. Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. *Molecular Biology and Evolution* [Internet] 36:2906–2921. Available from: https://doi.org/10.1093/molbev/msz191

Pedersen BS, Layer R, Quinlan AR. 2020. smoove: structural-variant calling and genotyping with existing tools.

Pélissié B, Chen YH, Cohen ZP, Crossley MS, Hawthorne DJ, Izzo V, Schoville SD. 2022. Genome Resequencing Reveals Rapid, Repeated Evolution in the Colorado Potato Beetle. *Molecular Biology and Evolution* [Internet] 39:msac016. Available from: https://doi.org/10.1093/molbev/msac016

Pfeiffer DG, Shrader ME, Wahls JCE, Willbrand BN, Sandum I, van der Linde K, Laub CA, Mays RS, Day ER. 2019. African Fig Fly (Diptera: Drosophilidae): Biology, Expansion of Geographic Range, and Its Potential Status as a Soft Fruit Pest. *J Integr Pest Manag* [Internet] 10. Available from: https://academic.oup.com/jipm/article/10/1/20/5514212

Picq S, Wu Y, Martemyanov VV, Pouliot E, Pfister SE, Hamelin R, Cusson M. 2023. Range-wide population genomics of the spongy moth, Lymantria dispar (Erebidae): Implications for biosurveillance, subspecies classification and phylogeography of a destructive moth. *Evolutionary Applications* [Internet] 16:638–656. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.13522

Platts PJ, Mason SC, Palmer G, Hill JK, Oliver TH, Powney GD, Fox R, Thomas CD. 2019. Habitat availability explains variation in climate-driven range shifts across multiple taxonomic groups. *Scientific Reports* 9:15039.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* [Internet] 21:i351–i358. Available from: https://doi.org/10.1093/bioinformatics/bti1018

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81:559–575.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res* [Internet] 26:342–350. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4772016/

R Core Team. R: A language and environment for statistical computing. Available from: http://www.R-project.org/

Rakes LM, Delamont M, Cole C, Yates JA, Blevins LJ, Hassan FN, Bergland AO, Erickson PA. 2023. A small survey of introduced Zaprionus indianus (Diptera: Drosophilidae) in orchards of the eastern United States. *Journal of Insect Science* [Internet] 23:21. Available from: https://doi.org/10.1093/jisesa/iead092

Renkema JM, Miller M, Fraser H, Légaré J-P, Hallett RH. 2013. First records of *Zaprionus indianus* Gupta (Diptera: Drosophilidae) from commercial fruit fields in Ontario and Quebec, Canada. *The Journal of the Entomological Society of Ontario* [Internet] 144. Available from: https://journal.lib.uoguelph.ca/index.php/eso/article/view/3745

Ricciardi A. 2007. Are modern biological invasions an unprecedented form of global change? *Conserv Biol* 21:329–336.

Roman G, Meller V, Wu KH, Davis RL. 1998. The opt1 gene ofDrosophila melanogaster encodes a proton-dependent dipeptide transporter. *American Journal of Physiology-Cell Physiology* [Internet] 275:C857–C869. Available from: https://journals.physiology.org/doi/full/10.1152/ajpcell.1998.275.3.C857

Roque F, Matavelli C, Lopes PHS, Machida WS, Von Zuben CJ, Tidon R. 2017. Brazilian Fig Plantations Are Dominated by Widely Distributed Drosophilid Species (Diptera: Drosophilidae). *Annals of the Entomological Society of America* 110:521–527.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, et al. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* [Internet] 449:913–918. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2687721/

Sardain A, Sardain E, Leung B. 2019. Global forecasts of shipping traffic and biological invasions to 2050. *Nature Sustainability* 2:274–282.

Schluter D, Marchinko KB, Arnegard ME, Zhang H, Brady SD, Jones FC, Bell MA, Kingsley DM. 2021. Fitness maps to a large-effect locus in introduced stickleback populations. *PNAS* [Internet] 118. Available from: https://www.pnas.org/content/118/3/e1914889118

Seebens H, Blackburn TM, Dyer EE, Genovesi P, Hulme PE, Jeschke JM, Pagad S, Pyšek P, Winter M, Arianoutsou M, et al. 2017. No saturation in the accumulation of alien species worldwide. *Nature Communications* 8:14435.

Seebens H, Essl F, Dawson W, Fuentes N, Moser D, Pergl J, Pyšek P, Kleunen M van, Weber E, Winter M, et al. 2015. Global trade will accelerate plant invasions in emerging economies under climate change. *Global Change Biology* 21:4128–4140.

da Silva VH, Laine VN, Bosse M, Spurgin LG, Derks MFL, van Oers K, Dibbits B, Slate J, Crooijmans RPMA, Visser ME, et al. 2019. The Genomic Complexity of a Large Inversion in Great Tits. *Genome Biology and Evolution* [Internet] 11:1870–1881. Available from: https://doi.org/10.1093/gbe/evz106

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* [Internet] 31:3210–3212. Available from: https://doi.org/10.1093/bioinformatics/btv351

Smit A, Hubley R, Green P. 2015. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org

Soudi S, Crepeau M, Collier TC, Lee Y, Cornel AJ, Lanzaro GC. 2023. Genomic signatures of local adaptation in recent invasive Aedes aegypti populations in California. *BMC Genomics* [Internet] 24:311. Available from: https://doi.org/10.1186/s12864-023-09402-5

Steenwyk JL, Rokas A. 2021. ggpubfigs: Colorblind-Friendly Color Palettes and ggplot2 Graphic System Extensions for Publication-Quality Scientific Figures. *Microbiology Resource Announcements* [Internet] 10:10.1128/mra.00871-21. Available from: https://journals.asm.org/doi/10.1128/MRA.00871-21

Stern DB, Lee CE. 2020. Evolutionary origins of genomic adaptations in an invasive copepod. *Nat Ecol Evol* [Internet] 4:1084–1094. Available from: https://www.nature.com/articles/s41559-020-1201-y

Stuart KC, Cardilini APA, Cassey P, Richardson MF, Sherwin WB, Rollins LA, Sherman CDH. 2021. Signatures of selection in a recent invasion reveal adaptive divergence in a highly vagile invasive species. *Molecular Ecology* [Internet] 30:1419–1434. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15601

Sutherst RW, Constable F, Finlay KJ, Harrington R, Luck J, Zalucki MP. 2011. Adapting to crop pest and pathogen risks under a changing climate. *WIREs Climate Change* 2:220–237.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell PJ, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. *Current Biology* [Internet] 32:111-123.e5. Available from: https://www.sciencedirect.com/science/article/pii/S0960982221014962

Tepolt CK, Grosholz ED, de Rivera CE, Ruiz GM. 2022. Balanced polymorphism fuels rapid selection in an invasive crab despite high gene flow and low genetic diversity. *Molecular Ecology* [Internet] 31:55–69. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.16143

Tepolt CK, Palumbi SR. 2020. Rapid Adaptation to Temperature via a Potential Genomic Island of Divergence in the Invasive Green Crab, Carcinus maenas. *Front. Ecol. Evol.* [Internet] 8. Available from: https://www.frontiersin.org/articles/10.3389/fevo.2020.580701

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole-genomes. *Nat Genet* [Internet] 49:303–309. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5470542/

Thompson MJ, Jiggins CD. 2014. Supergenes and their role in evolution. *Heredity* [Internet] 113:1–8. Available from: http://www.nature.com/hdy/journal/v113/n1/full/hdy201420a.html

Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. 2012. Estimating kinship in admixed populations. *Am J Hum Genet* 91:122–138.

Timmeren SV, Isaacs R. 2014. Drosophila suzukii in Michigan vineyards, and the first report of Zaprionus indianus from this region. *Journal of Applied Entomology* 138:519–527.

Uller T, Leimu R. 2011. Founder events predict changes in genetic diversity during human-mediated range expansions. *Global Change Biology* [Internet] 17:3478–3485. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2486.2011.02509.x

Vilela C. 1999. Is Zaprionus indianus Gupta, 1970 (Diptera, Drosophilidae) currently colonizing the Neotropical region? *Drosophila Information Service* 82:37–39.

Weinig C, Brock MT, Dechaine JA, Welch SM. 2007. Resolving the genetic basis of invasiveness and predicting invasions. *Genetica* [Internet] 129:205–216. Available from: https://doi.org/10.1007/s10709-006-9015-7

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis.

Willbrand B, Pfeiffer D, Leblanc L, Yassin A. 2018. First Report of African Fig Fly, Zaprionus indianus Gupta (Diptera: Drosophilidae), on the Island of Maui, Hawaii, USA, in 2017 and Potential Impacts to the Hawaiian Entomofauna. *Proceedings of the Hawaiian Entomological Society* 50:55–65.

Yassin A, Capy P, Madi-Ravazzi L, Ogereau D, David JR. 2008. DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid Zaprionus indianus. *Molecular Ecology Resources* [Internet] 8:491–501. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-8286.2007.02020.x

Yassin A, David J. 2010. Revision of the Afrotropical species of Zaprionus (Diptera, Drosophilidae), with descriptions of two new species and notes on the internal reproductive structures and immature stages. *ZooKeys* 51:33–72.

Zanuncio-Junior JS, Fornazier MJ, Andreazza F, Culik MP, Mendonça L de P, Oliveira EE, Martins D dos S, Fornazier ML, Costa H, Ventura JA. 2018. Spread of Two Invasive Flies (Diptera: Drosophilidae) Infesting Commercial Fruits in Southeastern Brazil. *flen* 101:522–525.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28:3326–3328.