Biol Invasions
https://doi.org/10.1007/s10530-020-02217-5

ORIGINAL PAPER

# Population genomic and phenotype diversity of invasive *Drosophila suzukii* in Hawai'i

**Jonathan B. Koch · Julian R. Dupuis · Marcel-Kate Jardeleza · Nadiatou Ouedraogo · Scott M. Geib · Peter A. Follett · Donald K. Price**

Received: 15 May 2019 / Accepted: 6 February 2020


**Abstract** In the context of evolutionary theory, invasion biology provides a fantastic enigma: how does a species with limited standing genetic variation survive and adapt to a novel environment? Reduced genetic diversity is typically associated with low fitness and evolutionary potential, yet some introduced species have proven to be successful invaders despite undergoing a genetic bottleneck during the early stages of colonization. Our goal in this study was to characterize population genomic and phenotype diversity of invasive *Drosophila suzukii* (Diptera: Drosophilidae) since colonizing the Hawaiian archipelago as early as the 1980s. Wing phenotype analysis revealed that high altitude populations possessed significantly larger wings than low altitude populations, supporting the hypothesis that insects cope with high altitude environments by developing larger wings. While we discovered low genetic diversity and differentiation in all Hawai'i populations, three unique genetic clusters were detected with a model-free, multivariate statistical approach. We identified 23 candidate loci under selection using two complementary analyses to detect $F_{ST}$ outliers across the genome. For 12 of these loci, predicted proteins are

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s10530-020-02217-5) contains supplementary material, which is available to authorized users.

J. B. Koch (✉) · M.-K. Jardeleza · N. Ouedraogo · D. K. Price
Department of Biology and Tropical Conservation Biology and Environmental Science Graduate Program, University of Hawai'i, Hilo, 200 W. Kāwili St., Hilo, HI 96720, USA
e-mail: kochj@hawaii.edu

J. R. Dupuis
Department of Plant and Environmental Protection Services, University of Hawai'i, Mānoa, 2500 Campus Rd, Honolulu, HI 96822, USA

J. R. Dupuis · S. M. Geib · P. A. Follett
United States Pacific Basin Agricultural Research Center, USDA-ARS, 64 Nowelo St., Hilo, HI 96720, USA

M.-K. Jardeleza
Department of Bioagricultural Sciences and Pest Management, Colorado State University, 307 University Ave, Fort Collins, CO 50523, USA

D. K. Price
School of Life Sciences, University of Nevada-Las Vegas, 4505 S. Pkwy, Las Vegas, NV 89154, USA

Published online: 12 February 2020

associated with *Drosophila* spp. chemosensation, amino acid and sodium ion transport, a Ras effector pathway, and cytidine deamination. Despite a genetic bottleneck, adventive *D. suzukii* populations are beginning to differentiate across the Hawaiian archipelago and selection for key behavioral and cellular processes are likely ongoing.

**Keywords** Population genomics · Outlier analysis · Island · Phenotype · Invasive species

## Introduction

Invasive species are a global dilemma that impose a significant burden on the world economy (Pimentel et al. 2005). Within a few generations of establishment, some introduced species have exhibited increased growth rates, adjusted to novel climate niches, and acquired behavioral traits that promoted their invasion success (Pattison et al. 1998; Lee 2002; Wilson et al. 2009; Strange et al. 2011). The evolution of these invaders is posited to be driven in large part by both natural selection and genetic drift, which promote increased capacity of invasive species to persist and flourish in foreign environments (Lee 2002). Substantial theoretical and empirical advances have been made in the study of "Invasion Genetics" starting in the 1960s (Baker and Stebbins 1965). However, many questions remain unresolved in understanding the ecological and evolutionary processes responsible for the patterns and changes in genetic diversity that enable invasive species success (Barrett 2014). While many studies examine neutral genetic diversity of adventive populations during an invasion event, very few studies have examined non-neutral genetic diversity—and its role in driving speciation (Zayed and Whitfield 2008; Riquet et al. 2013; Batista et al. 2016; Dupuis et al. 2018).

Many invasive species exhibit novel phenotypes in their invaded range, primarily due to phenotype plasticity (Davidson et al. 2011; Fraimout et al. 2018). Phenotype plasticity is the ability for an organism to express different phenotypes (e.g, behaviors, body size) in response to changes in the biotic and abiotic environment (Agrawal 2001). Compared to non-invasive species, invasive species can show significantly higher phenotype plasticity (Davidson et al. 2011). In some circumstances, high phenotype plasticity may be associated with a fitness benefit, allowing the invasive species to survive and reproduce in its new environment (Davidson et al. 2011; Shearer et al. 2016). A meta-analysis across 75 invasive/non-invasive species pairs found evidence for invasive species to exhibit increased phenotype plasticity for 11 biological traits relative to non-invasive species (Davidson et al. 2011). The plastic nature of wings may enable an insect species to colonize novel environments (Shearer et al. 2016). The phenotype changes of an invasive species associated with new environmental conditions support the inference that these changes may be linked to increased fitness. For example, increased wing size of invasive insects along temperature, altitude, and latitude gradients are well documented (Huey et al. 2000; Fraimout et al. 2018). Wing clines (i.e, the phenomenon where an insect species exhibits small to large wings across a gradient) may be attributed to the effects of temperature changes and stress tolerance in changing environments (Hoffmann et al. 2003). In addition, genetic changes leading to the evolution of adaptive wing clines may also emerge, thus supporting the inference that there is a genetic basis for wing shape differentiation in some species (Huey et al. 2000; Hoffmann et al. 2003).

Invasive species present an opportunity to study the genetic processes associated with the evolution of a species in novel environments in real time (Sax et al. 2005; Rius et al. 2015). How a newly introduced species responds to a novel environment can vary (Tsutsui et al. 2000; Wilson et al. 2009), and depends on several population-level factors including the number of founders, the number of founding populations, and the rate of gene flow between adventive and non-adventive populations (Lee 2002; Barrett 2014; Rius et al. 2015). If multiple colonization events are associated with an invasive species, it is likely that genetic diversity will increase across time due to admixture (Kolbe et al. 2004), and potentially increase the fitness of the invader (Keller and Taylor 2010). However, if an invasive species undergoes a genetic bottleneck outside of its native range, it might also exhibit changes in traits over a relatively short time frame that enable it to survive and thrive in novel conditions (Tsutsui et al. 2000; Wilson et al. 2009).

The Spotted Wing Drosophila, *Drosophila suzukii* (Diptera: Drosophilidae), has become a world-wide invasive species and agricultural pest, expanding into

the Americas and Europe, and is causing significant economic damage (Bolda et al. 2010; Adrion et al. 2014; Fraimout et al. 2017). The traits that make *D. suzukii* a particularly nefarious pest is its preference to use ripening, as opposed to decaying fruit, to rear offspring (Mitsui et al. 2006), and the serrated ovipositor of the female which is used to puncture the peel of the host fruit for oviposition (Asplen et al. 2015). Furthermore, adventive *D. suzukii* populations exhibit a high degree of phenotype plasticity in response to changes in seasonality (Shearer et al. 2016; Stockton et al. 2018; Fraimout et al. 2018). There is compelling evidence that cool temperatures induce a ‘winter phenotype’ in *D. suzukii*, which increases their survival in the winter months, by increasing their body size, melanization of their integument, and differential regulation of key genes associated with cold-hardiness and reproductive activities (Shearer et al. 2016; Stockton et al. 2018). In addition to the significant economic loss associated with a *D. suzukii* invasion, recent studies have also shown that *D. suzukii* has the capacity to use a diversity of wild and ornamental host plants (Lee et al. 2015b).

Native to Asia, *D. suzukii* was first documented outside of its native range in the Wai‘anae mountains on O‘ahu, Hawai‘i, USA on 7 October 1980 (Kaneshiro 1983). Based on microsatellite data and approximate Bayesian computation analyses, the *D. suzukii* population in Hawai‘i are predicted to have originated from Japan, and show no evidence of multiple genetic introductions into the archipelago (Fraimout et al. 2017). In Hawai‘i, *D. suzukii* is not considered to be a pest of any food crops and there are no management programs being implemented or losses to the pest being reported. However, *D. suzukii* has been documented to use the fruits of plants native to Hawai‘i for reproductive activities, namely ‘ōhelo (*Vaccinium reticulatum*) and ‘ākala (*Rubus hawaiiensis*) (Magnacca et al. 2008). Unlike invasive populations documented in temperate environments, populations in Hawai‘i, South America, and La Réunion are situated in tropical environments where there is no climate-related reason to undergo diapause (Shearer et al. 2016). Thus, we predict that populations in Hawai‘i can reproduce year-round. For nearly three decades, *D. suzukii* continued its establishment in Hawai‘i, but was not reported to be expanding its geographic distribution to new locations outside of Hawai‘i and Asia. However, in September 2008, *D. suzukii* was detected in Santa Cruz, California in a field of raspberries (*Rubus idaeus*) (Hauser 2011), and then rapidly expanded throughout North America, South America, and Europe over the span of five years (Adrion et al. 2014; Fraimout et al. 2017). Rapid expansion of *D. suzukii* across the globe is estimated to be the product of five invasion events, with each colonizing population undergoing a genetic bottleneck (Fraimout et al. 2017).

Despite the genetic bottleneck that occurs at the onset of a colonization event, many invasive species are able to thrive in novel environments (Tsutsui et al. 2000; Wilson et al. 2009). Low genetic diversity negatively impacts fitness, survival, and evolutionary potential, yet invasive species continue to be successful competitors during the early stages of a colonization event. The documented colonization history and rapid spread of *D. suzukii* in Hawai‘i make the species an ideal model to study how the genome evolves in a new environment. Here, we sampled invasive populations of wild *D. suzukii* across the archipelago to record phenotype data and to generate a genome-scale single nucleotide polymorphism (SNP) dataset using double digest restriction-site associated DNA sequencing (ddRAD). We measured phenotype variation of invasive populations as a proxy of *D. suzukii*’s response to climate variation and used the ddRAD dataset to perform two hierarchical analyses to characterize population structure. Finally, to identify candidate loci under selection, we used a frequency-based outlier approach to detect loci under selection and an environmental correlation-based approach to identify alleles that are correlated to altitude. These analyses let us ask the following questions: (1) Is there a relationship between island origin and altitude on phenotype? (2) Are there differences in genetic diversity and structure across island populations? and (3) Is there evidence for selection (non-neutral genetic variation) occurring across the genome of *D. suzukii* in response to altitude and geographic isolation? By studying fine-scale genetic diversity and structure of invasive *D. suzukii* populations in the Hawaiian archipelago, we will determine how the genome responds to a novel environment.

## Materials and methods

### Specimen collection

A total of 23 unique locations were surveyed for wild *D. suzukii* across the high islands in Hawaiian archipelago (Table 1; Fig. 1). All islands were surveyed for *D. suzukii* except for Ni'ihau, Moloka'i, Lāna'i, and Kaho'olawe. At each location, *D. suzukii* were collected by either sweeping insects flying over decaying fruit (e.g, *Psidium cattleyanum*, *Vaccinium reticulatum*) with a net, or by deploying three Haviland traps baited with a mixture of ~ 7 g of yeast, ~ 14 g of sugar, and 250 mL of water (Lee et al. 2012). The Haviland traps were left out for 3–5 days. Wild flies captured by sweep net or Haviland trap were placed into 95% EtoH. Flies collected from all 23 locations were used for wing phenotype analysis, whereas only 18 populations were selected for population genomic analysis. Of the 18 populations, two populations were found to have unsuitable DNA based on poor quality read mapping and alignment to the *D. suzukii* genome and were excluded from final population genomic analysis. Flies were identified by biological sex based on the presence/absence of a serrated ovipositor and spots on the distal region of the wings (Asplen et al. 2015). All *D. suzukii* specimens were kept at − 20 °C until needed for wing phenotype analysis and genomic DNA extraction. All analyses were done on wild-caught individuals, and no populations were reared in the laboratory.

**Table 1** Survey locations of *Drosophila suzukii* across the Hawaiian archipelago

| Location name | LN code | Island | Lat. | Lon. | Ele. (m) | Temp. (°C) | F | M | PGA |
|---|---|---|---|---|---|---|---|---|---|
| Haleakalā | HALE | Mau'i | 20.7659 | − 156.2410 | 2109 | 14.2 | 8 | 8 | * |
| Hanalei | HANA | Kaua'i | 22.1986 | − 159.4770 | 165 | 23.1 | 8 | 8 | |
| Hau'ula | HAUU | O'ahu | 21.6089 | − 157.9182 | 125 | 22.0 | 7 | 9 | * |
| Havo, Low | HALO | Hawai'i | 19.2983 | − 155.0989 | 42 | 22.2 | 20 | 0 | |
| Ka'ala | KAAL | O'ahu | 21.4958 | − 158.1587 | 521 | 22.6 | 7 | 6 | * |
| Kapolei | KAPO | O'ahu | 21.3354 | − 158.0760 | 30 | 22.6 | 8 | 8 | |
| Ka'u Desert Trail | KAUD | Hawai'i | 19.3662 | − 155.3650 | 948 | 13.6 | 25 | 27 | * |
| Kaumana | KAUM | Hawai'i | 19.6743 | − 155.3301 | 1533 | 16.5 | 18 | 18 | |
| Kukuiopane | KUKU | Hawai'i | 19.3033 | − 155.8189 | 1410 | 12.8 | 0 | 16 | * |
| Mākaha Ridge | MILO | Kaua'i | 22.1143 | − 159.6723 | 1179 | 20.8 | 8 | 8 | * |
| Makapipi | MAKP | Maui | 20.8072 | − 156.0960 | 332 | 18.5 | 8 | 8 | * |
| Makawao | MAKW | Maui | 20.8304 | − 156.2780 | 873 | 14.2 | 8 | 8 | * |
| Mānana | MANA | O'ahu | 21.4321 | − 157.9318 | 323 | 22.8 | 8 | 8 | * |
| Mānoa | MANO | O'ahu | 21.2988 | − 157.8138 | 24 | 23.3 | 1 | 15 | * |
| Mauna Loa | MAUN | Hawai'i | 19.6745 | − 155.4653 | 2034 | 13.6 | 27 | 23 | * |
| No Dump | NODU | Hawai'i | 19.6942 | − 155.2183 | 807 | 16.5 | 20 | 16 | |
| Nounou | SLEEP | Kaua'i | 22.0617 | − 159.3467 | 367 | 21.7 | 8 | 8 | * |
| Pu'u Ō'ō | PUOO | Hawai'i | 19.6713 | − 155.3845 | 1772 | 10.1 | 19 | 20 | |
| UHH | UHH | Hawai'i | 19.6971 | − 155.0892 | 282 | 22.0 | 19 | 18 | |
| Waihe'e | WAIH | Maui | 20.9498 | − 156.5363 | 291 | 20.6 | 8 | 8 | * |
| Waipi'o | WAIP | Hawai'i | 20.1111 | − 155.5495 | 262 | 19.4 | 0 | 16 | * |
| Waldron Ledge | WALD | Hawai'i | 19.4203 | − 155.2538 | 1203 | 16.7 | 19 | 32 | * |
| Wiliwili Camp | WILI | Kaua'i | 22.0630 | − 159.6424 | 216 | 18.2 | 5 | 11 | * |

LN code corresponds to the location name (LN). LN codes are found in Tables 2, 3, and Fig. 1a

*Lat.* latitude, *Lon.* longitude, *Ele.* elevation in meters, *Temp. (C)* annual mean (1970–2000) temperature °C (extrapolated from BIO1 in the WorldClim database, http://www.worldclim.org/), *F* female, *M* male

Locations used for population genetic analysis (PGA) are indicated with an *

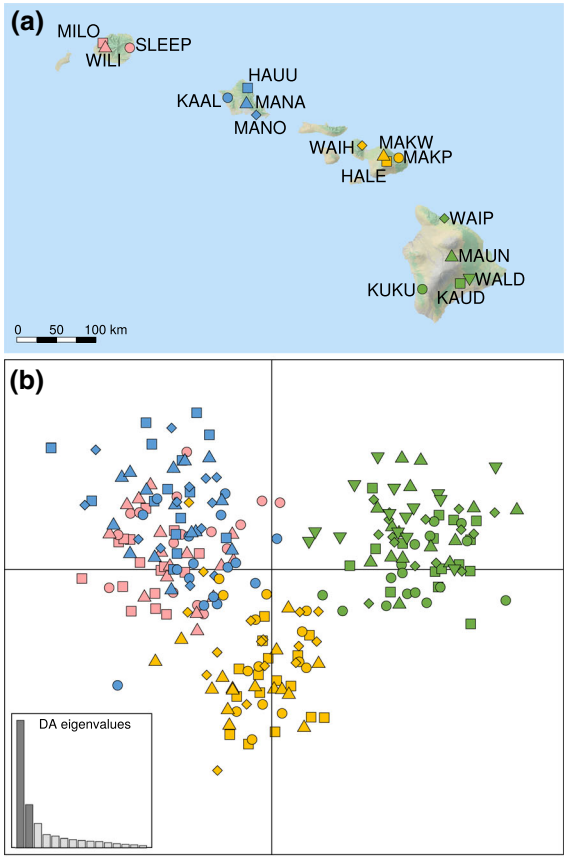


**Fig. 1** Distribution and genetic structure of adventive *Drosophila suzukii* populations in the Hawaiian archipelago. **a** Map of current *D. suzukii* sampling (n = 16) and **b** results of DAPC ($n$ = 246). Inset in DAPC ordination shows relative contribution of discriminant functions. Populations are defined in column LN code in Table 1. Map layers from http://naturalearthdata.com

**Table 2** Descriptive population genetic statistics for populations and islands [as well as Kaua'i (KA) + O'ahu (OA)]

| LN code | Island | $H_{OBS}$ | $H_{EXP}$ | GenDiv | $G_{IS}$ |
|---|---|---|---|---|---|
| MILO | Kaua'i | 0.347 | 0.347 | 0.347 | − 0.002 |
| SLEEP | Kaua'i | 0.351 | 0.344 | 0.344 | − 0.019 |
| WILI | Kaua'i | 0.336 | 0.360 | 0.360 | 0.067 |
| HAUU | O'ahu | 0.330 | 0.342 | 0.342 | 0.035 |
| KAAL | O'ahu | 0.339 | 0.346 | 0.346 | 0.021 |
| MANA | O'ahu | 0.337 | 0.343 | 0.343 | 0.017 |
| MANO | O'ahu | 0.332 | 0.353 | 0.353 | 0.058 |
| HALE | Maui | 0.334 | 0.340 | 0.340 | 0.019 |
| MAKP | Maui | 0.340 | 0.337 | 0.337 | − 0.009 |
| MAKW | Maui | 0.354 | 0.355 | 0.355 | 0.002 |
| WAIH | Maui | 0.336 | 0.343 | 0.343 | 0.021 |
| KAUD | Hawai'i | 0.345 | 0.348 | 0.348 | 0.007 |
| KUKU | Hawai'i | 0.317 | 0.340 | 0.340 | 0.066 |
| MAUN | Hawai'i | 0.322 | 0.338 | 0.338 | 0.047 |
| WAIP | Hawai'i | 0.322 | 0.339 | 0.339 | 0.051 |
| WALD | Hawai'i | 0.330 | 0.346 | 0.346 | 0.047 |
| Kaua'i | NA | 0.345 | 0.352 | 0.352 | 0.020 |
| O'ahu | NA | 0.335 | 0.347 | 0.347 | 0.035 |
| Maui | NA | 0.341 | 0.345 | 0.345 | 0.012 |
| Hawai'i | NA | 0.327 | 0.342 | 0.342 | 0.043 |
| KA + OA | NA | 0.339 | 0.350 | 0.350 | 0.030 |

*LN code* location name code (see Table 1), $H_{OBS}$ observed heterozygosity, $H_{EXP}$ expected heterozygosity, *GenDiv* gene diversity, $G_{IS}$ inbreeding coefficient, *Hawai'i* Hawai'i Island

## Wing phenotype analysis

To examine differences in wing phenotype across invasive *D. suzukii* populations, the right wing of both female and male *D. suzukii* was excised from the thorax and mounted on a clean microscope slide in 70% EtOH. Wing centroid size, hereafter described as wing size, is a useful measure of wing phenotype, and is correlated with body size and thorax length (Azevedo et al. 1998). Wing centroid size is calculated as the square root of the sum of squared distances of a set of landmarks from their centroid (Dryden and Mardia 1998). Specimens were placed under a Leica microscope with a Canon camera attachment. Photographs of each wing were taken at 3.2× magnification on a stage micrometer. We digitized 15 coordinates (i.e., landmarks) on the wing of each specimen from the photographs and then scaled the image using the 1 mm ruler on the stage micrometer as a standard using ImageJ (Abràmoff et al. 2004). The 15 points measured on ImageJ were then analyzed using MorphoJ (Klingenberg 2011), which allowed us to determine the wing centroid size. With MorphoJ, we used a Procrustes fit, a statistical shape analysis that scales, rotates, and superimpose landmark data across specimens. The goal of this analysis is to minimize shape differences across the wings scored by obtaining similar placement and size of all wings scored.

Previous studies on *D. suzukii* and other *Drosophila* spp. have found there to be a significant difference in wing size between male and female flies (Huey et al. 2000; Shearer et al. 2016). Therefore, we elected to construct generalized linear models (GLMs) on each

**Table 3** Pairwise $F_{ST}$ values between populations

| | MILO | SLEEP | WILI | HAUU | KAAL | MANA | MANO | HALE | MAKP | MAKW | WAIH | KAUD | KUKU | MAUN | WAIP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SLEEP | 0.0024 | | | | | | | | | | | | | | |
| WILI | 0.0059 | 0.0071 | | | | | | | | | | | | | |
| HAUU | 0.0141 | 0.0180 | 0.0189 | | | | | | | | | | | | |
| KAAL | 0.0050 | 0.0017 | 0.0018 | 0.0173 | | | | | | | | | | | |
| MANA | 0.0032 | 0.0030 | 0.0100 | 0.0055 | − 0.0004 | | | | | | | | | | |
| MANO | 0.0023 | 0.0087 | 0.0055 | 0.0032 | − 0.0003 | 0.0026 | | | | | | | | | |
| HALE | 0.0115 | 0.0097 | 0.0159 | 0.0190 | 0.0049 | 0.0100 | 0.0135 | | | | | | | | |
| MAKP | 0.0075 | 0.0085 | 0.0072 | 0.0191 | 0.0005 | 0.0058 | 0.0077 | 0.0029 | | | | | | | |
| MAKW | 0.0037 | 0.0076 | 0.0054 | 0.0111 | 0.0054 | 0.0079 | 0.0024 | 0.0088 | 0.0021 | | | | | | |
| WAIH | 0.0046 | 0.0073 | 0.0168 | 0.0198 | 0.0061 | 0.0054 | 0.0107 | 0.0049 | 0.0038 | 0.0029 | | | | | |
| KAUD | 0.0198 | 0.0141 | 0.0206 | **0.0272** | 0.0182 | 0.0175 | 0.0187 | 0.0145 | 0.0133 | 0.0108 | 0.0118 | | | | |
| KUKU | 0.0143 | 0.0126 | 0.0150 | **0.0258** | 0.0083 | 0.0170 | 0.0138 | 0.0112 | 0.0049 | 0.0061 | 0.0098 | − 0.0024 | | | |
| MAUN | 0.0205 | 0.0151 | 0.0219 | 0.0186 | 0.0112 | 0.0126 | 0.0143 | 0.0097 | 0.0124 | 0.0102 | 0.0129 | − 0.0011 | − 0.0017 | | |
| WAIP | 0.0197 | 0.0119 | 0.0147 | **0.0351** | 0.0097 | 0.0134 | **0.0195** | 0.0106 | 0.0083 | 0.0158 | 0.0117 | − 0.0015 | − 0.0019 | 0.0013 | |
| WALD | 0.0135 | 0.0121 | 0.0193 | 0.0160 | 0.0148 | 0.0118 | 0.0082 | 0.0157 | 0.0087 | 0.0067 | 0.0059 | − 0.0042 | 0.0007 | − 0.0016 | 0.0055 |

Significant comparisons after Bonferroni correction are bolded ($p \leq 0.05$). Populations are defined in item LN code in Table 1

sex separately when testing for the effect of altitude (continuous independent variable) and island origin (categorical independent variable) with the "identity" link function. Altitude is used as a proxy for mean annual temperature as there is a significant and negative correlation between the two variables in our sampling of invasive *D. suzukii* populations ($r = -0.84$, $df = 22$, $p < 0.001$) (Table 1). Data curation, management, visualization, and analysis were conducted with the R statistical programming language (R Core Development Team 2018).

## DNA extraction, ddRAD library preparation, and sequencing

A total of 288 specimens ($\sim$ 16 males and $\sim$ 16 females per site) were selected for population genomic analysis. After the right wings of each fly was extracted for the wing phenotype analysis, entire flies were homogenized in tissue lysis buffer using a 3.175 mm 18/10 stainless steel bearing in a FastPrep 24 homogenizer (MP Biomedical, Santa Ana, CA) set to a speed of 4.0 m/s for 60 s. Proteinase K was added to the resulting homogenate and incubated overnight at 55 °C. DNA was extracted using the NucleoMag® Tissue Kit (Macherey–Nagel) and a KingFisher Flex-96 automated extraction instrument (Thermo Scientific, Waltham, MA) following manufacturer's protocol including the RNase A treatment. DNA was eluted into 100 uL of elution buffer and quantified using a picogreen assay on a SpectraMax M2 plate reader. Because of the low concentration of DNA found in our extractions, we elected to not dilute or normalize the extractions before downstream library preparation.

ddRAD libraries were prepared following the protocol outlined in Peterson et al. (2012) using the restriction enzymes *Nla*III and *Mlu*CI. In total, 48 unique barcode adapters were used in initial ligation, and sub-pools of these 48 barcodes were size selected using a 1.5% agarose gel cassette and target size selection of "narrow 450 bp" on a Blue Pippin electrophoresis unit (Sage Science, Beverly, MA). Illumina i7 barcodes were added to each sub-pool in a final PCR, followed by a clean-up step using a 1.5:1 ratio of polyethylene glycol containing solid-phase reversible immobilization beads to sample volume (DeAngelis et al. 1995). To quantify DNA, size-selected and cleaned products were placed on a 2100 Bioanalyzer with a high sensitivity DNA kit (Agilent, Santa Clara, CA). The DNA sub-pools were then combined at equal molar ratios to generate final libraries of 144 individuals (3 sub-pools per library); in total 288 samples were prepared this way. One hundred base pair single-end sequencing of two final, 144-plex libraries were performed on two lanes of Illumina HiSeq 4000.

## ddRAD data processing

We used the *process_radtags* utility in Stacks v2.1 (Catchen et al. 2011, 2013) to remove reads that had low-quality scores or uncalled bases, demultiplex reads by individual, and rescue any reads with errors in the barcode or cut site. We mapped the cleaned and demultiplexed reads to the *D. suzukii* genome assembly (NCBI: GCA_000472105.1) using the Burrow's Wheeler Aligner v0.7.17 (Li and Durbin 2009) MEM algorithm (Li 2013). The *ref_map.pl* utility in *Stacks* was used to assemble loci and call SNPs with default parameters. We used *populations* to filter the final *Stacks* output, requiring each locus to be present in 10 populations and in 75% of individuals per population, and generated a vcf output containing a random SNP per catalog locus. We then used VCFtools v0.1.15 (Danecek et al. 2011) to first identify individuals with more than 50% missing data (after removing loci with $> 50\%$ missing data), and then to filter the raw *populations* output by removing those individuals, as well as removing loci that had $> 10\%$ missing data, a minimum average read depth of $< 20$, a minor allele frequency $< 5\%$, and were within 10,000 bp of each other on a given scaffold of the reference genome. We converted the resulting vcf file into other formats using PGDSpider v2.1.0.3 (Lischer and Excoffier 2012).

## Population genetic statistics

We used GenoDive v2.0b27 (Meirmans and Van Tienderen 2004) to calculate general population genetic statistics (e.g, heterozygosity, gene diversity, inbreeding) and to test for pairwise population differentiation using 10,000 permutations of the analysis of molecular variance (AMOVA) $F_{ST}$ method (Excoffier et al. 1992; Michalakis and Excoffier 1996). To account for multiple comparisons in this test, we applied a Bonferroni correction. We used the poppr library v2.8.1 (Kamvar et al. 2014, 2015) in R to identify private alleles within populations, and tested

for isolation by distance (IBD) using 10,000 permutations of *mantel.randtest* function in the adegenet R library v2.1.1 (Jombart 2008; Jombart and Ahmed 2011). For the genetic data in this test, we manually standardized the $F_{ST}$ values from GenoDive ($F_{ST}/1 - F_{ST}$), and for geographic data, we calculated a distance matrix using the *distm* function (Vincenty ellipsoid method) in the geosphere library v1.5 (Hijmans et al. 2017) in R and performed matrix formatting using the sna library v2.4 (Butts and Butts 2016). Finally, one-way analysis of variance (ANOVA) was used to test for the effect of island origin on estimates of genetic diversity.

### Population structure

We used two methods to identify population structure in this dataset: individual-based clustering in a Bayesian framework using STRUCTURE v2.3.4 (Pritchard et al. 2000) and discriminant analysis of principal components (DAPC) (Jombart et al. 2010). STRUCTURE assigns individuals to a set number of genetic clusters which maximizes Hardy–Weinberg and gametic equilibria. In our study, we ran 20 replicates of STRUCTURE for each of $K$ = 1–20 with 250,000 Markov Chain Monte Carlo (MCMC) generations (following 50,000 burn-in), the admixture model, and correlated allele frequencies (Falush et al. 2003). We also ran identical analyses using two different location priors, corresponding to collection localities and islands. We used CLUMPAK v1.1 (Kopelman et al. 2015) to average results across replicates and calculate Ln Pr($X|K$) (Pritchard et al. 2000) and $\Delta K$ (Evanno et al. 2005), and considered both statistics to choose the best value of $K$. In contrast to STRUCTURE, DAPC is a multivariate method which minimizes within-group variability and maximizes between-group variability of a predefined set of groups by transforming population allele frequencies using principal components analysis and conducting a subsequent discriminant analysis on those principal components (Jombart et al. 2010). DAPC is known to be effective at uncovering cases of complex and fine-scale population structure (Jombart et al. 2010; Kanno et al. 2011; Dupuis et al. 2018), especially when there are a priori hypotheses of population groupings. We conducted DAPC using the adegenet library in R, using the *find.clusters* function to estimate $K$ (considering all principal components) and *xvaldapc* to perform cross-validation in order to determine the optimal number of principal components to include in the discriminant analysis.

### Candidate loci

We used complementary methods to identify candidate loci under selection, a frequency-based outlier approach with BayeScan (Foll and Gaggiotti 2008) and an environmental correlation-based approach with BayEnv (Coop et al. 2010). Bayescan uses an outlier approach that compares locus-specific $F_{ST}$ to an observed or expected neutral $F_{ST}$ distribution to detect loci under selection. We ran BayeScan v2.1 using 10,000 output generations, 50,000 iterations of burn-in, and a thinning interval of 50, and following the author's recommendations, a false discovery rate of 0.05 and a prior odds value of 100 (Foll and Gaggiotti 2008). We ran five independent analyses, and only considered outliers that were predicted in all five analyses. In contrast to outlier analysis, BayEnv controls for population structure and identifies alleles that are correlated to environmental characteristics and assumed to be under environmental selection. We ran BayEnv v2.0 using the PyBayEnv wrapper (Ring 2015), with 10 replicates per analysis and 100,000 MCMC generations per replicate, and used altitude (determined via GPS points at collection sites, Geographic projection = WGS1984) as an environmental variable. Given high reported between-run variability of this method (Blair et al. 2014), we conducted 10 independent analyses as described, and considered candidate loci that were predicted in at least two analyses. Finally, we assessed autosomes and allosomes independently as the null distribution of $F_{ST}$ under neutral processes will systematically differ between the two chromosome types. As allosomes have a smaller effective population size than autosomes, we would expect there to be more statistical outliers on the allosomes (Charlesworth 2009).

We used the *intersect* function in BEDTools v2.25.0 (Quinlan and Hall 2010) to extract functional annotation information for candidate loci from the *D. suzukii* genome assembly. However, given that the *D. suzukii* assembly is far from chromosome-scale (8680 scaffolds) and that there is overall high synteny between *D. suzukii* and *Drosophila melanogaster* (Chiu et al. 2013), we used the *D. melanogaster* genome assembly to place the ddRAD loci into a

chromosomal context. The use of model organisms with chromosome-level resolution to study the genomics of non-model organisms is a legitimate approach to functional and population genomic assessments (O'Brien et al. 1999; Sadd et al. 2015; Toews et al. 2016). For each ddRAD SNP, we used samtools v1.9 (Li et al. 2009) to extract 4000 bp, centered on the SNP, from the *D. suzukii* reference genome, and used *blastn* in BLAST v2.7.1+ (Altschul et al. 1990) to match these sequences to a database of the *D. melanogaster* genome assembly (NCBI: GCA_000001215.4). We set both the "maximum number of aligned sequences to keep" (-max_target_seqs) and the "maximum number of high scoring pairs" (-max_hsps) to one for this search. We calculated locus-specific Weir and Cockerham's $F_{ST}$ using VCFtools (considering three populations matching the DAPC results) and plotted locus-specific $F_{ST}$ and -$\log_{10}$($q$-value), calculated from BayeScan, using the qqman R library (Turner et al. 2014).

## Results

### Wing phenotype analysis

A total of 533 flies ($n_{female}$ = 253, $n_{male}$ = 280) were available for wing phenotype analysis. As predicted, female fly wings ($\bar{x}$ = 3.51 mm ± 0.03 SE) were significantly larger than male fly wings ($\bar{x}$ = 3.10 mm ± 0.02 SE) (two sample $t$ test (equal variances), $t$ = 12.79, $df$ = 573, $p < 0.001$). The sex-specific GLM analyses found a significant and positive effect of altitude on mean wing size in both female ($\beta_{Altitude}$ = 0.00026 mm ± 0.00008 SE, $t$ = 3.09, $df$ = 17, $p$ = 0.007) and male flies ($\beta_{Altitude}$ = 0.00038 mm ± 0.00004 SE, $t$ = 8.64, $df$ = 18, $p < 0.001$) (Fig. 2). Specifically, for each 1 km increase in altitude, we predict a 0.26 mm and 0.38 mm increase in wing size for female and male *D. suzukii*, respectively. Because the adventive population on O'ahu is suspected to be the initial colonist population (Kaneshiro 1983), we elected to do pairwise comparisons of mean wing length of each island with O'ahu. We failed to detect a significant difference in mean female wing size between O'ahu and Kaua'i ($\beta_{O'ahuVsKaua'i}$ = 0.26 mm ± 0.15 SE, $t$ = 1.72, $df$ = 17, $p$ = 0.10), O'ahu and Hawai'i Island ($\beta_{O'ahuVsHawai'i}$ = 0.04 mm ± 0.15 SE, $t$ = 0.25, $df$ = 17, $p$ = 0.80), or O'ahu and Maui ($\beta_{O'ahuVsMaui}$ = 0.19 mm ± 0.16 SE, $t$ = 1.20, $df$ = 17, $p$ = 0.24) populations. We did find a significant difference in mean male wing size between O'ahu and Kaua'i populations ($\beta_{O'ahuVsKaua'i}$ = 0.35 mm ± 0.08 SE, $t$ = 4.39, $df$ = 18, $p$ = 0.0004). However, we failed to detect a significant difference in mean male wing size between O'ahu and Maui populations ($\beta_{O'ahuVsMaui}$ = 0.17 mm ± 0.08 SE, $t$ = 2.06, $df$ = 18, $p$ = 0.054), and between O'ahu and Hawai'i Island populations ($\beta_{O'ahuVsHawai'i}$ = − 0.07 mm ± 0.08 SE, $t$ = − 0.93, $df$ = 18, $p$ = 0.36).

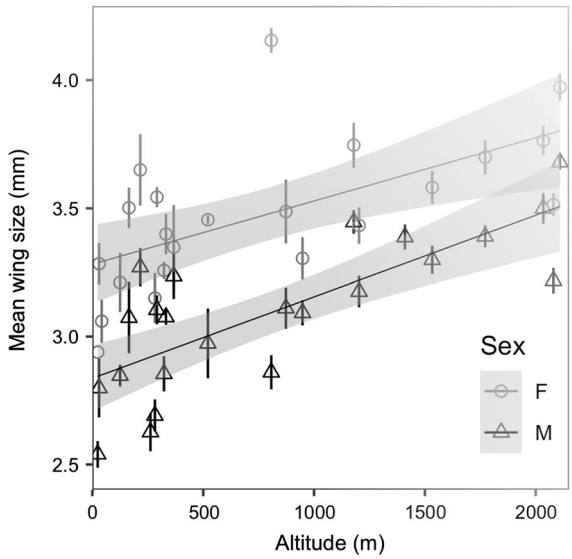


**Fig. 2** Mean wing size distribution of female and male *Drosophila suzukii* across an altitude gradient in the Hawaiian archipelago. Each point represents the mean wing size (± SE) of the fly population at a field site ($n$ = 23). Bold line represents generalized linear model fit of mean wing size as a function of altitude and island for the respective sex. Shaded regions represent 95% confidence interval of the model fit

### ddRAD data characteristics

We sampled 288 flies from 18 populations across four islands in the Hawaiian archipelago (Fig. 1). Two lanes of 100 bp single-end sequencing resulted in 659.9 million total reads for 288 individuals. After removing 12.4 million reads with low quality bases or missing RAD tags, we mapped 647.5 million reads to the 8680 scaffolds of the reference genome (minimum, mean, maximum, 54 thousand, 2.25 million, and 7.7 million, respectively), 581.8 million of which were

successfully mapped (minimum, mean, maximum: 49 thousand, 2.0 million, and 4.6 million, respectively). The raw output of *populations* consisted of 65,441 SNPs, and initial filtering with VCFtools identified 42 individuals with > 50% missing data (after temporarily removing 7879 loci with > 50% missing data), and most of these individuals (31 of 42) were from two localities. After removing these aberrant individuals and applying more strict filters in VCFtools, the resulting dataset consisted of 3484 SNPs across 246 individuals (16 localities, Table 1), with an average of 1.7% missing data per individual.

## Population genetic statistics

We calculated population genetic statistics at two spatial scales, one treating collection localities as populations and the other treating islands as populations, and we use "populations" and "islands" to disambiguate these units. At both population and island levels, heterozygosity and gene diversity were quite moderate and consistent between population units; observed heterozygosity ranged from 0.31 to 0.35 and 0.32 to 0.34 in populations and islands, respectively, and genetic diversity ranged from 0.33 to 0.36 and 0.34 to 0.35, respectively (Table 2). The inbreeding coefficient, $G_{IS}$, was positive in all but three populations, but generally quite close to zero (range: − 0.019 to 0.067). Across islands, $G_{IS}$ ranged from 0.012 to 0.043, and was highest on Hawai'i Island and lowest on Maui. Furthermore, one-way ANOVA failed to detect a significant difference in observed heterozygosity ($F_{3,12}$= 3.20, $p$ = 0.06), genetic diversity ($F_{3,12}$= 1.11, $p$ = 0.38), or $G_{IS}$ ($F_{3,12}$= 1.72, $p$ = 0.22) between islands. Pairwise $F_{ST}$ values between populations were low, averaging $F_{ST}$= 0.01, and significant differentiation, after Bonferroni correction, was only observed in four of the 120 comparisons (Table 3). Pairwise differentiation between islands was significant in all comparisons except for one, Kaua'i to O'ahu, although $F_{ST}$ values were still very low ($\bar{x}$ = 0.009, Table 4). We detected significant IBD across populations distributed throughout the entire archipelago ($r$ = 0.58, $p < 0.0001$) (Fig. 3a). However, we failed to detect significant IBD across populations within an island (Fig. 3b).

**Table 4** Pairwise $F_{ST}$ values between islands in the Hawaiian archipelago

| | KA + OA | KA | OA | MA |
|---|---|---|---|---|
| OA | | 0.0041 | | |
| MA | **0.0045** | **0.0055** | **0.0058** | |
| HA | 0.0139 | **0.0150** | **0.0151** | **0.0092** |

*KA* Kaua'i, *OA* O'ahu, *MA* Maui, *HA* Hawai'i Island

Significant comparisons after Bonferroni correction are bolded ($p \leq 0.05$)

## Population structure

STRUCTURE was unable to identify any meaningful population structure in this dataset. In all analyses, including those using a location prior of collection localities or islands, values of Ln Pr($X|K$) were approximately equal from $K$ = 1–10, and $\Delta K$ supported $K$ = 2 (this metric is unable to assess the probability of $K$ = 1) (Figure S1). Visualizing ancestry across all values of $K$ resulted in a pattern that strongly supported an absence of structure in the dataset (Figure S1). DAPC, on the other hand, was able to identify structure across the dataset, with *find.clusters* predicting $K$ = 3 and cross-validation with *xvaldapc* identifying 60 principal components as the optimal number to retain in the discriminant analysis. Clustering at $K$ = 3 generally corresponded to Kaua'i + O'ahu, Maui, and Hawai'i Island, although differentiation of these clusters was not complete; some intergradation was apparent between clusters, particularly with some Maui individuals falling in with the Kaua'i + O'ahu cluster (Fig. 1b). Independent analyses using either populations or islands as a priori groups resulted in the same general pattern, and there was no apparent pattern of structure between populations on the same island.

## Candidate loci

BayeScan predicted 19 candidate loci outliers ($WCF_{ST}(q)$), whereas BayEnv identified 8 loci ($BF(q)$) correlated to altitude across 10 independent runs (Fig. 4; Table 5). In the BayEnv analysis only five loci were supported by two or more independent analyses. Bayes factors ($BF(q)$) of these five loci were highly variable, ranging from 12.5 to 546.7

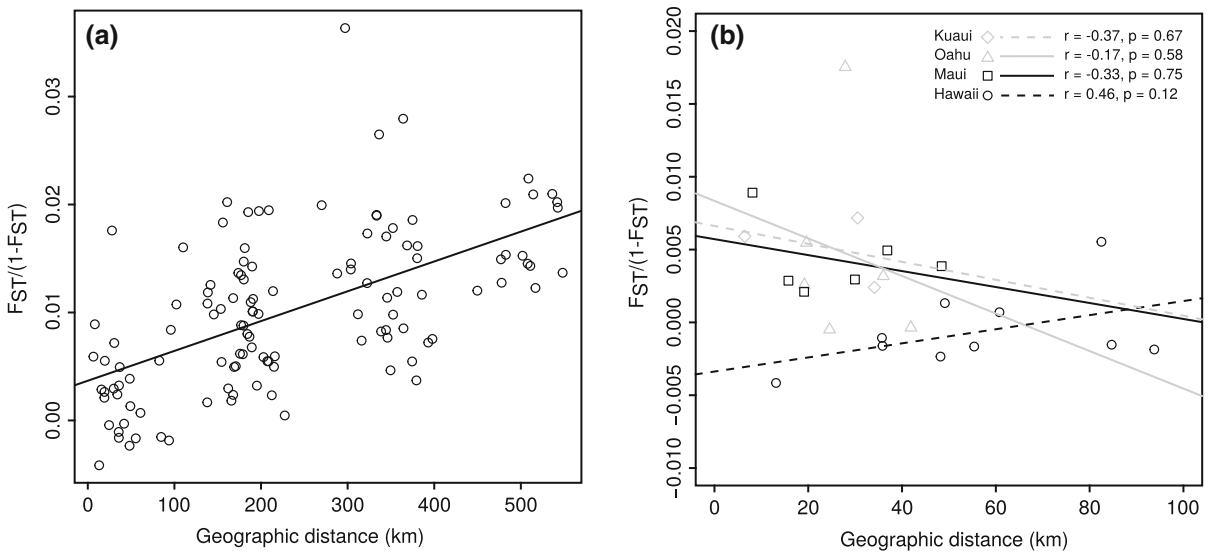


**Fig. 3** Isolation by distance (km) **a** across all islands ($n$ = 120), and **b** across locations within islands ($n$ = 26)

($\bar{x}$ = 121.6) (Table 5). When comparing the two well-supported sets of candidate loci (19 for BayeScan and five for BayEnv), only one candidate locus was in common between the analyses.

When placed in a chromosomal context (relative to *D. melanogaster* chromosomes), we failed to detect a significant difference in the number of candidate loci across the five autosomes and allosome ($X^2$ = 7.33, $df$ = 5, $p$ = 0.20) (Fig. 4). Of the 23 candidate loci exhibiting selection, only 12 loci were associated within predicted proteins of the *D. suzukii* genome. Of the 12 annotations, four predicted proteins are associated with *D. melanogaster* or *D. simulans* (Table 5). The four loci are found within the putative gustatory receptor 47b transcript variant (BayeScan), sodium-dependent nutrient amino acid transporter 1 transcript variant (BayeScan), protein sprint transcript variant (BayeScan), and cytidine deaminase (BayEnv) (Table 5). The remaining predicted proteins were found in other organisms including *Homo sapiens*, *Mycobacterium tuberculosis*, and *Nephila clavipes*. The candidate locus shared between BayeScan and BayEnv is found on the 2R chromosome and is an uncharacterized protein. Overall, the $F_{ST}$ values of candidate loci was high relative to the genome-wide values, but BayEnv candidate loci provided exceptions to this trend with many candidate loci showing lower differentiation.

## Discussion

Here we present the first population genomic study of *D. suzukii*, specifically characterizing phenotype variability, genetic diversity, and putative selection signatures across the genome. Focusing on invasive populations in the Hawaiian archipelago allows for a close examination of how genetic diversity and structure respond to ∼ 38 years of establishment (Kaneshiro 1983). We discovered that Hawaiian *D. suzukii* exhibit phenotype variability, with individuals at high altitude possessed significantly larger wings than individuals at low altitude. Furthermore, while there was little difference in genetic diversity across islands, population genetic structure is beginning to emerge relative to island origin. Specifically, distinct genetic clusters of *D. suzukii* were found on (1) Hawai'i Island, (2) Maui, and (3) O'ahu + Kaua'i. Based on the results of low genetic diversity and little genetic structure across islands, our study supports the hypothesis that there likely was a single genetic source associated with the colonization event of *D. suzukii* into Hawai'i from Japan (Adrion et al. 2014; Fraimout et al. 2017). Finally, we present a line of inference to support the hypothesis for ongoing selection for certain behavioral and cellular processes.

Wing phenotype analysis of *D. suzukii* support the established consensus that *D. suzukii* and other *Drosophila* spp. are sensitive to climate variation,

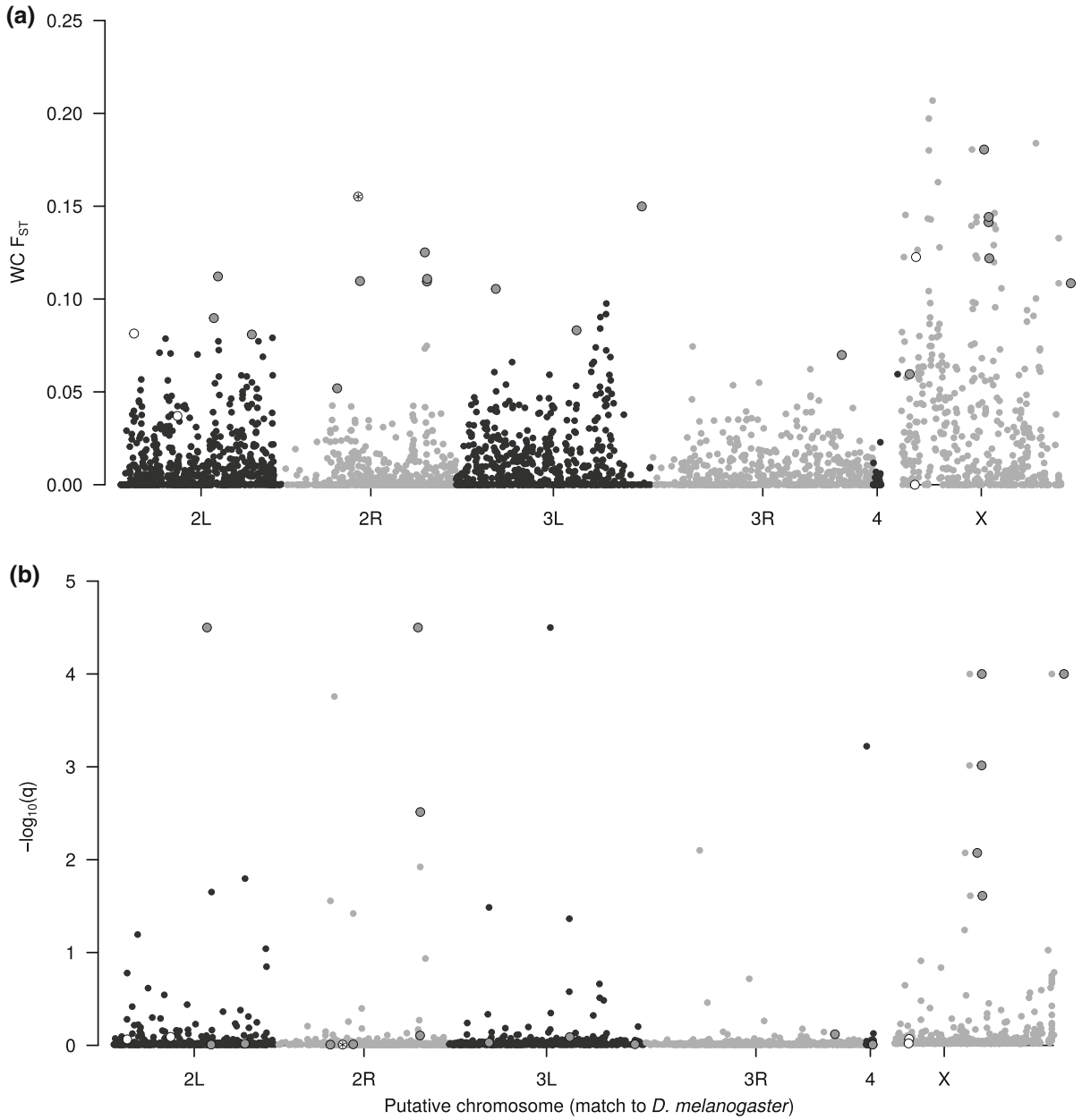


**Fig. 4** Genome-wide Weir and Cockerham's $F_{ST}$ (**a**) and $-\log_{10}(q)$ (**b**) of *Drosophila suzukii*, relative to chromosomes of the *D. melanogaster* genome ($n$ = 3484). Larger circles with black outlines are candidate loci; dark grey: well-supported BayeScan outliers, white: well-supported Bayenv candidate loci (predicted in more than two analyses), light-grey: Bayenv candidates predicted in only one analysis. Candidate locus in common between BayeScan and Bayenv noted with white circles and asterisks

and have the capacity to respond to changing climates (Huey et al. 2000; Gilchrist et al. 2001; Balanyá et al. 2006). Adventive populations of *D. subobscura* found outside of its native range are found to exhibit increased wing size in cooler latitudes, relative to populations in warmer latitudes (Gilchrist et al. 2004). Other studies of *Drosophila* have found a correlation between wing size and stress resistance, suggesting that body size is a significant physiological response to climate variation, and exhibits a degree of plasticity in

**Table 5** Outliers identified across *Drosophila suzukii* populations in the Hawaiian archipelago

| Scaffold | Position | Clocus | $WCF_{ST}$ ($q$) | BF ($q$) | Chr (pos) | Predicted protein |
|---|---|---|---|---|---|---|
| NW_016024447.1 | 6,038,738 | 313,736:28:+ | **0.155 (< 0.0001)** | **94.2 (< 0.0001)**, **56.1 (< 0.0001)**, **163.3 (0.02)** | 2R (11,093,743) | Uncharacterized protein DDB_G0271670 |
| NW_016019885.1 | 9,218,624 | 47,699:57:+ | **0.112 (0.022)** | | 2L (14,070,713) | Neuronal acetylcholine receptor subunit alpha-7 |
| NW_016019885.1 | 14,242,821 | 74,206:62:+ | **0.081 (0.016)** | 39.7 (0.031) | 2L (18,952,612) | Homocysteine S-methyltransferase |
| NW_016019885.1 | 19,175,422 | 100,367:67:+ | **0.089 (< 0.0001)** | | 2L (13,475,697) | Mucin-5AC |
| NW_016019886.1 | 1,198,382 | 124,673:82:+ | **0.141 (0.001)** | | X (11,142,396) | Uncharacterized LOC108004723 transcript variant |
| NW_016019886.1 | 1,275,423 | 124,936:47:− | **0.144 (0.0001)** | | X (11,171,137) | Spidroin-1-like |
| NW_016019886.1 | 1,359,059 | 125,335:84:+ | **0.122 (0.025)** | | X (11,252,135) | Uncharacterized LOC108004719 transcript variant |
| NW_016022043.1 | 28,988 | 202,537:17:+ | **0.060 (0.0006)** | | 4 (1,007,809) | No match |
| NW_016024447.1 | 6,330,989 | 315,514:69:− | **0.110 (0.038)** | 520.1 (0.012) | 2R (11,362,388) | Putative gustatory receptor 47b transcript variant |
| NW_016024447.1 | 6,460,778 | 316,155:85:+ | **0.125 (0.003)** | | 2R (20,725,927) | Uncharacterized LOC108010078 |
| NW_016024447.1 | 6,808,390 | 318,180:29:+ | **0.109 (0.012)** | | 2R (21,030,150) | No match |
| NW_016024447.1 | 6,844,902 | 318,372:76:+ | **0.111 (0.0002)** | | 2R (21,062,567) | Acidic mammalian chitinase |
| NW_016024475.1 | 10,437 | 371,456:15:+ | **0.108 (0.0001)** | | X (23,029,460) | No match |
| NW_016024625.1 | 369,579 | 408,285:96:− | **0.105 (0.033)** | | 3L (5,764,660) | Proliferation-associated protein 2G4 |
| NW_016024842.1 | 3,824,075 | 469,590:66:− | **0.083 (0.043)** | | 3L (17,416,662) | Sodium-dependent nutrient amino acid transporter 1 transcript variant |
| NW_016025501.1 | 75,918 | 615,319:62:− | **0.150 (< 0.0001)** | 281.1 (< 0.0001) | 3L (26,844,457) | No match |
| NW_016026917.1 | 1,320,394 | 764,406:37:− | **0.181 (0.008)** | | X (10,504,025) | Protein sprint transcript variant |
| NW_016027348.1 | 257,386 | 793,686:76:+ | **0.070 (0.008)** | | 3R (27,631,466) | Tubby-related protein 4 |
| NW_016027458.1 | 1,987,768 | 807,019:95:+ | **0.052 (0.028)** | | 2R (8,065,462) | Uncharacterized LOC108019654 transcript variant |
| NW_016019885.1 | 4,429,641 | 22,973:20:+ | 0.037 (0.811) | **31.3 (< 0.0001)**, **47.3 (< 0.0001)**, **59.4 (< 0.0001)** | 2L (8,214,508) | Cytidine deaminase |
| NW_016019885.1 | 8,433,069 | 43,676:88:+ | 0.081 (0.166) | **103.3 (0.048)**, **546.7 (0.010)** | 2L (1,944,037) | Uncharacterized LOC108008776 |
| NW_016025862.1 | 2,575,586 | 661,871:20:− | 0 (0.950) | **70.9 (< 0.0001)**, **112.7 (< 0.0001)**, **59.4 (< 0.0001)**, **59.3 (< 0.0001)**, **114.8 (< 0.0001)**, **57.8 (< 0.0001)**, **193.1 (< 0.0001)**, **139.2 (< 0.0001)**, **193.9 (< 0.0001)**, **78.9 (< 0.0001)** | X (512,062) | Arginase-1 transcript variant |

response to changing environmental conditions (Griffiths et al. 2005; Gilchrist et al. 2008; Shearer et al. 2016; Stockton et al. 2018; Fraimout et al. 2018). In *D. suzukii*, the onset of cooler temperatures triggers plastic changes in phenotype, where flies exposed to cooler temperatures exhibit increased body size, wing size, and melanization of the integument (Shearer et al. 2016; Clemente et al. 2018; Fraimout et al. 2018). While we observed *D. suzukii* collected at high altitude field sites in the Hawaiian archipelago to be more melanic relative to low altitude populations, this phenomenon remains to be formally tested.

Cooler temperatures and reduced air density are hypothesized to shape wing size as a mechanism to sustain flight (Dillon et al. 2006). Adventive Hawai'i populations of *D. suzukii* show clear differences in wing size across an altitude gradient suggesting that colder temperatures and reduced air density may induce increased wing size in wild populations (Table 1). To test this hypothesis there is a critical need to account for genotype by phenotype interactions by exposing lab-reared *D. suzukii* populations to temperature and air density treatments (Dillon et al. 2006, Huey et al. 2000). From these tests, one can begin to measure the interplay between developmental time, wing phenotypes, and gene expression patterns (Shearer et al. 2016; Fraimout et al. 2018).

Although we found no evidence of genetic structure across island populations using an individual-based Bayesian clustering approach (STRUCTURE), we identified three genetic clusters (Kaua'i + O'ahu, Maui, and Hawai'i Island) using the model-free approach of DAPC. Despite this population structure, and an overall signature of IBD across the archipelago, differentiation between these clusters was very low (maximum pairwise $F_{ST}$ between island pairs = 0.0151, between O'ahu and Hawai'i Island) (Table 4). Populations on offshore and oceanic islands are typically genetically differentiated due to reproductive isolation caused by limited dispersal opportunity (Jha and Kremen 2013; Lozier et al. 2013; Funk et al. 2016; Szalanski et al. 2016; Foster et al. 2018). Prolonged reproductive isolation due to the low frequency or inability to disperse results in genetic drift and the potential for local adaptation that can affect the evolutionary trajectory of island populations. While our study is unable to directly test whether there have been multiple colonization events in Hawai'i, the low genetic diversity observed here

**Table 5** continued

| Scaffold | Position | Clocus | $WCF_{ST}$ ($q$) | BF ($q$) | Chr (pos) | Predicted protein |
|---|---|---|---|---|---|---|
| NW_016025862.1 | 3,091,015 | 664,379:13:– | 0.122 (0.854) | **12.5 (0.044)**, **14.8 (0.044)** | X (670,244) | No match |

Scaffold, position, predicted protein, and mRNA_accession derived from the *D. suzukii* draft genome, and Chr (pos) refers to putative position relative to the *D. melanogaster* genome (position is the midpoint of the blast hit on the *D. melanogaster* genome, see Methods)

CLocus refers to the Catalog locus identified by Stacks, and significance in BayeScan results represents the final set of outliers from that analysis (see Methods)

Bayenv results consist of significant Bayes Factors (BF) and their corresponding $q$-values across five replicates of PyBayEnv, and WC $F_{ST}$ represents Weir and Cockerham's $F_{ST}$, calculated with VCFTools

Significant $q$-values indicated in bold

supports the hypothesis for a single colonization event. Other research on *D. suzukii* in Hawai'i has come to the same conclusion (Adrion et al. 2014; Fraimout et al. 2017). Operating under the single colonization hypothesis, we suggest that O'ahu and Kaua'i have been either colonized at similar times, or that there is ongoing dispersal between the islands. However, Maui and Hawai'i Island are forming unique genetic clusters suggesting that there is limited dispersal occurring between islands, unlike with Kaua'i and O'ahu.

Using both $F_{ST}$ outlier and environmental correlation approaches uncovered a line of inference to support the hypothesis of ongoing selection across the genome. Of the 12 predicted proteins we identified in our study, four predicted proteins are associated with either cellular or behavioral processes, including putative gustatory receptor 47b. Gustatory receptors provide a key function in chemosensory pathways in *Drosophila*, and enable the insect to make choices about nutrition, toxicity in the environment (carbon dioxide), mate selection, and oviposition (Clyne et al. 2000). In Hawai'i and across the globe, *D. suzukii* populations are exposed to novel host fruits that are not found in their native environment (Bellamy et al. 2013; Burrack et al. 2013; Keesey et al. 2015; Diepenbrock et al. 2016). Preference for certain host fruits are linked to olfactory cues, natal host fidelity, as well as the probability for successful oviposition (Burrack et al. 2013; Lee et al. 2015a; Diepenbrock et al. 2016). Of course, these results must be interpreted with caution due to well-known occurrences of false positives in outlier analyses (Narum and Hess 2011). However our results do provide guidance for future, targeted research on selection in *D. suzukii*, such as directed sequencing experiments (Pespeni and Palumbi 2013).

The invasion of *D. suzukii* in the Hawaiian archipelago is in the early stages of genetic diversification across a novel tropical environment. Although *D. suzukii* is estimated to have been present in the archipelago for $>$ 38 years (Kaneshiro 1983), we detected evidence for fine scale genetic structure across the islands and signatures of selection across the genome. Emerging research has revealed that *D. suzukii* exhibits a degree of phenotype variability that is directly linked to their ability to survive and reproduce in novel and changing environments. Our phenotype analysis discovered a cline in *D. suzukii* wing size relative to altitude. This result provides support for the hypothesis that high altitude populations are responding to cooler conditions and reduced air density characteristic of high altitudes. By having a larger wing area relative to body size, these high altitude populations have the wing phenotype required to efficiently sustain flight (Dudley 2002). Interestingly, the environmental correlation approach for detecting selection identified several genomic regions correlated with altitude, which may have a complex relationship with the phenotype variability observed in the system. To account for phenotype by genotype interactions, more research will be required to validate these candidate regions and follow up on this relationship through directed sequencing.

**Acknowledgements** We thank Ms. Angela Kauwe for her assistance in ddRAD library preparation, Mr. Daniel Martinez and Ms. Quinn Hamamoto for their assistance in wing phenotype analysis, Mr. Armen Armaghanyan for collection assistance, and three anonymous reviewers for comments on an earlier draft of this manuscript. JBK was supported by a National Science Foundation (NSF) Postdoctoral Research Fellowship in Biology (#1523661) and a David H. Smith Postdoctoral Research Fellowship. MKJ and NO were supported by an NSF-Research experience for undergraduates site grant awarded to Becky Ostertag (PI) and Noelani Puniwai (Co-PI) (#146130). Permission to survey *D. suzukii* was made possible by an endorsement from the Department of Land and Natural Resources (FHM16-410), Haleakalā National Park (HALE-2016-SCI-003), and Hawai'i Volcanoes National Park (HAVO-2016-SCI-015). The US Department of Agriculture, Agricultural Research Service is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

**Author contributions** JBK and DKP designed the research. JBK, MKJ, and NO made field surveys and collected samples. JBK and JRD conducted analyses. JBK and JRD led the writing of the manuscript. SG, PF, MKJ, and DKP provided useful comments and contributed to revisions.

**Data accessibility** Individual SNP genotype data and population locations assigned to each specimen analyzed in our study will be available from Dryad (https://doi.org/10.5061/dryad.60b67d5).

## References

Abràmoff MD, Magalhães PJ, Ram SJ (2004) Image processing with ImageJ. Biophotonics Int 11:36–42

Adrion JR, Kousathanas A, Pascual M et al (2014) *Drosophila suzukii*: the genetic footprint of a recent, worldwide invasion. Mol Biol Evol 31:3148–3163

Agrawal AA (2001) Phenotypic plasticity in the interactions and evolution of species. Science 294:321–326

Altschul SF, Gish W, Miller W et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Asplen MK, Anfora G, Biondi A et al (2015) Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci 88:469–494

Azevedo RBR, James AC, McCabe J, Partridge L (1998) Latitudinal variation of wing: thorax size ratio and wing-aspect ratio in *Drosophila melanogaster*. Evolution 52:1353–1362

Baker HG, Stebbins GL (eds) (1965) The genetics of colonizing species. Academic Press, New York

Balanyá J, Oller JM, Huey RB et al (2006) Global genetic change tracks global climate warming in *Drosophila subobscura*. Science 313:1773–1775

Barrett SCH (2014) Foundations of invasion genetics: the Baker and Stebbins legacy. Mol Ecol 24:1927–1941

Batista PD, Janes JK, Boone CK et al (2016) Adaptive and neutral markers both show continent-wide population structure of mountain pine beetle (*Dendroctonus ponderosae*). Ecol Evol 6:6292–6300

Bellamy DE, Sisterson MS, Walse SS (2013) Quantifying host potentials: indexing postharvest fresh fruits for spotted wing Drosophila, *Drosophila suzukii*. PLoS ONE 8:e61227

Blair LM, Granka JM, Feldman MW (2014) On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics 8:1

Bolda MP, Goodhue RE, Zalom FG (2010) Spotted wing Drosophila: potential economic impact of a newly established pest. Agric Resour Econ Update 13:5–8

Burrack HJ, Fernandez GE, Spivey T, Kraus DA (2013) Variation in selection and utilization of host crops in the field and laboratory by *Drosophila suzukii* Matsumara (Diptera: Drosophilidae), an invasive frugivore. Pest Manag Sci 69:1173–1180

Butts CT, Butts MCT (2016) SNA: social network analysis. R package version 2.5

Catchen JM, Amores A, Hohenlohe P et al (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genomes Genet 1:171–182

Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. Nat Rev Genet 10:195

Chiu JC, Jiang X, Zhao L et al (2013) Genome of *Drosophila suzukii*, the spotted wing drosophila. G3 Genes Genomes Genet 3:2257–2271

Clemente M, Fusco G, Tonina L, Giomi F (2018) Temperature-induced phenotypic plasticity in the ovipositor of the invasive species *Drosophila suzukii*. J Therm Biol 75:62–68

Clyne PJ, Warr CG, Carlson JR (2000) Candidate taste receptors in *Drosophila*. Science 287:1830–1834

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423

Core Development Team R (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

Davidson AM, Jennions M, Nicotra AB (2011) Do invasive species show higher phenotypic plasticity than native species and if so, is it adaptive? A meta-analysis. Ecol Lett 14:419–431

DeAngelis MM, Wang DG, Hawkins TL (1995) Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res 23:4742–4743

Diepenbrock LM, Swoboda-Bhattarai KA, Burrack HJ (2016) Ovipositional preference, fidelity, and fitness of *Drosophila suzukii* in a co-occurring crop and non-crop host system. J Pest Sci 89:761–769

Dillon ME, Frazier MR, Dudley R (2006) Into thin air: physiology and evolution of alpine insects. Integr Comp Biol 46:49–61

Dryden IL, Mardia KV (1998) Statistical shape analysis. Wiley, Chinchester

Dudley R (2002) The biomechanics of insect flight: form, function, evolution. Princeton University Press, Princeton

Dupuis JR, Sim SB, San Jose M et al (2018) Population genomics and comparisons of selective signatures in two invasions of melon fly, *Bactrocera cucurbitae* (Diptera: Tephritidae). Biol Invasions 20:1211–1228

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Foster JT, Walker FM, Rannals BD, Sanchez DE (2018) Population genetics of an island invasion by Japanese Bush-Warblers in Hawaii, USA. The Auk Ornithol Adv 135:171–180

Fraimout A, Debat V, Fellous S et al (2017) Deciphering the routes of invasion of *Drosophila suzukii* by means of ABC random forest. Mol Biol Evol 34:980–996

Fraimout A, Jacquemart P, Villarroel B et al (2018) Phenotypic plasticity of *Drosophila suzukii* wing to developmental temperature: implications for flight. J Exp Biol. https://doi.org/10.1242/jeb.166868

Funk WC, Lovich RE, Hohenlohe PA et al (2016) Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol Ecol 25:2176–2194

Gilchrist GW, Huey RB, Serra L (2001) Rapid evolution of wing size clines in *Drosophila subobscura*. Genetica 112–113:273–286

Gilchrist GW, Huey RB, Balanyà J et al (2004) A time series of evolution in action: a latitudinal cline in wing size in South American *Drosophila subobscura*. Evolution 58:768–780

Gilchrist GW, Jeffers LM, West B et al (2008) Clinal patterns of desiccation and starvation resistance in ancestral and invading populations of *Drosophila subobscura*. Evol Appl 1:513–523

Griffiths JA, Schiffer M, Hoffmann AA (2005) Clinal variation and laboratory adaptation in the rainforest species *Drosophila birchii* for stress resistance, wing size, wing shape and development time. J Evol Biol 18:213–222

Hauser M (2011) A historic account of the invasion of *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae) in the continental United States, with remarks on their identification. Pest Manag Sci 67:1352–1357

Hijmans RJ, Williams E, Vennes C (2017) Geosphere. R package version 1.5-10

Hoffmann AA, Sørensen JG, Loeschcke V (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. J Therm Biol 28:175–216

Huey RB, Gilchrist GW, Carlson ML et al (2000) Rapid evolution of a geographic cline in size in an introduced fly. Science 287:308–309

Jha S, Kremen C (2013) Urban land use limits regional bumble bee gene flow. Mol Ecol 22:2483–2495

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinform 24:1403–1405

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinform 27:3070–3071

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208

Kaneshiro K (1983) *Drosophila* (*Sophophora*) *suzukii* (Matsumura). Notes and exhibitions. Proc Hawaiian Entomol Soc 24:179

Kanno Y, Vokoun JC, Letcher BH (2011) Fine-scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. Mol Ecol 20:3711–3729

Keesey IW, Knaden M, Hansson BS (2015) Olfactory Specialization in *Drosophila suzukii* supports an ecological shift in host preference from rotten to fresh fruit. J Chem Ecol 41:121–128

Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. J Evol Biol 23:1720–1731

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. Mol Ecol Resour 11:353–357

Kolbe JJ, Glor RE, Rodríguez Schettino L et al (2004) Genetic variation increases during biological invasion by a Cuban lizard. Nature 431:177–181

Kopelman NM, Mayzel J, Jakobsson M et al (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191

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

Lee JC, Burrack HJ, Barrantes LD et al (2012) Evaluation of monitoring traps for *Drosophila suzukii* (Diptera: Drosophilidae) in North America. J Econ Entomol 105:1350–1357

Lee JC, Dalton DT, Swoboda-Bhattarai KA et al. (2015a) Characterization and manipulation of fruit susceptibility to *Drosophila suzukii*. J Pest Sci 89:771–780

Lee JC, Dreves AJ, Cave AM et al (2015b) Infestation of wild and ornamental noncrop fruits by *Drosophila suzukii* (Diptera: Drosophilidae). Ann Entomol Soc Am 108:117–129

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN]

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinform 25:1754–1760

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinform 25:2078–2079

Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299

Lozier JD, Strange JP, Koch JB (2013) Landscape heterogeneity predicts gene flow in a widespread polymorphic bumble bee, *Bombus bifarius* (Hymenoptera: Apidae). Conserv Genet 14:1099–1110

Magnacca KN, Foote D, O'Grady PM (2008) A review of the endemic Hawaiian Drosophilidae and their host plants. Zootaxa 1728:1–58

Meirmans PG, Van Tienderen PH (2004) genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792–794

Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061–1064

Mitsui H, Takahashi KH, Kimura MT (2006) Spatial distributions and clutch sizes of *Drosophila* species ovipositing on cherry fruits of different stages. Popul Ecol 48:233–237

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Mol Ecol Resour 11:184–194

O'Brien SJ, Menotti-Raymond M, Murphy WJ et al (1999) The promise of comparative genomics in mammals. Science 286(458–62):479–481

Pattison RR, Goldstein G, Ares A (1998) Growth, biomass allocation and photosynthesis of invasive and native Hawaiian rainforest species. Oecologia 117:449–459

Pespeni MH, Palumbi SR (2013) Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. Mol Ecol 22:3580–3597

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS one 7:e37135

Pimentel D, Zuniga R, Morrison D (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. Ecol Econ 52:273–288

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinform 26:841–842

Ring KH (2015) PyBayenv: A framework for interpreting, testing and optimizing Bayenv analyses. https://www.duo.uio.no/

Riquet F, Daguin-Thiébaut C, Ballenghien M et al (2013) Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusc *Crepidula fornicata*. Mol Ecol 22:1003–1018

Rius M, Bourne S, Hornsby HG (2015) Applications of next-generation sequencing to the study of biological invasions. Curr Zool 61:488–504

Sadd BM, Barribeau SM, Bloch G et al (2015) The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol 16:76

Sax DF, Stachowicz JJ, Gaines SD (2005) Species invasions: insights into ecology, evolution and biogeography. Sinauer Associates Incorporated, Sunderland

Shearer PW, West JD, Walton VM et al (2016) Seasonal cues induce phenotypic plasticity of *Drosophila suzukii* to enhance winter survival. BMC Ecol 16:11

Stockton DG, Wallingford AK, Loeb GM (2018) Phenotypic plasticity promotes overwintering survival in a globally invasive crop pest, *Drosophila suzukii*. Insects 9:105

Strange JP, Koch JB, Gonzalez VH, Nemelka L (2011) Global invasion by *Anthidium manicatum* (Linnaeus) (Hymenoptera: Megachilidae): assessing potential distribution in North America and beyond. Biol Invasions 13:2115–2133

Szalanski AL, Tripodi AD, Trammel CE, Downey D (2016) Mitochondrial DNA genetic diversity of honey bees, *Apis mellifera*, in Hawaii. Apidologie 47:679–687

Toews DPL, Campagna L, Taylor SA et al (2016) Genomic approaches to understanding population divergence and speciation in birds. Auk 133:13–30

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. Proc Natl Acad Sci USA 97:5948–5953

Turner KG, Hufbauer RA, Rieseberg LH (2014) Rapid evolution of an invasive weed. New Phytol 202:309–321

Wilson EE, Mullen LM, Holway DA (2009) Life history plasticity magnifies the ecological effects of a social wasp invasion. Proc Natl Acad Sci USA 106:12809–12813

Zayed A, Whitfield CW (2008) A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee *Apis mellifera*. Proc Natl Acad Sci USA 105:3421–3426