

# A Test for Gene Flow among Sympatric and Allopatric Hawaiian Picture-Winged *Drosophila*

Lin Kang<sup>1</sup> · Harold R. Garner<sup>2,3</sup> · Donald K. Price<sup>4</sup> · Pawel Michalak<sup>1,2,5</sup>

Received: 28 April 2017 / Accepted: 4 May 2017 / Published online: 10 May 2017  
© Springer Science+Business Media New York 2017

**Abstract** The Hawaiian *Drosophila* are one of the most species-rich endemic groups in Hawaii and a spectacular example of adaptive radiation. *Drosophila silvestris* and *D. heteroneura* are two closely related picture-winged *Drosophila* species that occur sympatrically on Hawaii Island and are known to hybridize in nature, yet exhibit highly divergent behavioral and morphological traits driven largely through sexual selection. Their closest-related allopatric species, *D. planitibia* from Maui, exhibits hybrid male sterility and reduced behavioral reproductive isolation when crossed experimentally with *D. silvestris* or *D. heteroneura*. A modified four-taxon test for gene flow was applied to recently obtained genomes of the three Hawaiian *Drosophila* species. The analysis indicates recent gene flow in sympatry, but also, although less extensive, between allopatric species. This study underscores the prevalence of gene flow, even in taxonomic groups considered classic examples of allopatric

speciation on islands. The potential confounding effects of gene flow in phylogenetic and population genetics inference are discussed, as well as the implications for conservation.

**Keywords** Reproductive isolation · Hybridization · *Drosophila silvestris* · *Drosophila heteroneura* · Hawaii

## Introduction

The likelihood of speciation-with-gene-flow is still one of the most debated topics in evolutionary biology with the difficulty of isolating the effect of gene flow and the effect of time since population divergence (Nosil 2008), while gene flow is frequently considered as a constraining force in evolution. Nevertheless, a growing number of examples of speciation-with-gene-flow are being reported, and are advancing our understanding of speciation in the absence of geographic obstacles (Jonsson et al. 2014; Keller et al. 2013; Martin et al. 2013; Stacy et al. 2017). Several conceptual frameworks have been developed for the comparison of speciation with and without gene flow (Feder et al. 2014; Smadja and Butlin 2011). In the speciation-with-gene-flow process, the effective migration experienced across the genome depends on many factors such as recombination rate, strength of selection, and number of selected loci. These dynamic changes occurring across the genome indicate that whole genomic analysis can provide better empirical understanding of the genetic differentiation and speciation (Gagnaire et al. 2013).

Due to the sequential geological formation of islands in the Hawaiian archipelago, founder events, and speciation within the Hawaiian *Drosophila*, the iconic example of adaptive radiation, appears to have occurred in sequential order from the oldest (northwest) to the youngest (southeast) islands (Price and Clague 2002). *Drosophila silvestris* (SIL) and *D.*

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-017-9795-7) contains supplementary material, which is available to authorized users.

✉ Pawel Michalak  
pmichalak@vcom.vt.edu

<sup>1</sup> Biocomplexity Institute, Virginia Tech, Washington Street, MC 0477, Blacksburg, VA 24061-0477, USA

<sup>2</sup> Edward Via College of Osteopathic Medicine, Blacksburg, VA 24060, USA

<sup>3</sup> Gibbs Cancer Center & Research Institute, Spartanburg, SC 29303, USA

<sup>4</sup> School of Life Sciences, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

<sup>5</sup> Center for Molecular Medicine and Infectious Diseases, Virginia-Maryland College of Veterinary Medicine, Blacksburg, VA 24060, USA

*heteroneura* (HET) from the Big Island (Hawai'i) are presumably the youngest species in the group, and *D. planitibia* (PLA) from Maui is their closest sister species living in allopatry (Magnacca and Price 2015). The two sympatric species, *D. silvestris* and *D. heteroneura*, have been known to hybridize in nature in spite of mate discrimination (Carson 1989). However, neither the extent of hybridization footprints at the whole genome level nor a possibility of gene flow between allopatric species has yet been tested in this system. Here, gene flow between allopatric species would entail surviving the crossing a 26-miles wide channel between Maui and Hawai'i and then finding a suitable mate.

One of the most productive attempts to test for interspecies gene flow explores a genome-wide excess of shared derived alleles between taxa using Patterson's D statistic (Durand et al. 2011). Known as the four-taxon test, or ABBA/BABA test, the method was originally developed to determine whether there was a genetic admixture between modern human and neanderthal populations (Green et al. 2010). We use a modified ABBA/BABA method which incorporates the fixation index ( $F_{ST}$ ) with genome sequences of *D. silvestris*, *D. heteroneura*, and *D. planitibia* that we recently obtained (Kang et al. 2016) and the genome sequences of *D. grimshawi*, a more distantly related Hawaiian *Drosophila*, that had its genome first reported by the *Drosophila* 12 Genomes Consortium (Clark et al. 2007). We test for gene flow in sympatry, as well as allopatry, between the three Hawaiian picture-winged *Drosophila* species.

## Materials and Methods

### Genome Sequencing and Analysis

Genomic sequences from *D. silvestris*, *D. heteroneura*, and *D. planitibia* were obtained, processed, pre-assembled, and annotated as described by us elsewhere (Kang et al. 2016). Briefly, pooled genomes from ten non-inbred males per species were used for construction of paired-end and mate pair libraries that were Illumina-sequenced at a total  $>80 \times$  coverage. De-duplicated reads were then assembled with Spades (Bankevich et al. 2012) and protein-coding genes were predicted using MAKER2 (Holt and Yandell 2011). The completeness of assembly estimated with CEGMA (Parra et al. 2007) was in the range of 93–98% ('complete') and 98–99% ('partial').

### $F_{ST}$ , Tajima's D, $\pi$ , Watterson's $\theta$ , and $N_e$ Calculations

Sequences were mapped using BWA (Li and Durbin 2010) with default parameters and *D. grimshawi* assembly as reference. Samtools (Li et al. 2009) was used to generate

the pileup result. SNPs within 10 bp of indel were discarded, and PoPoolation2 (Kofler et al. 2011b) was used to estimate the  $F_{ST}$  value for each SNP. All pairwise analyses used the maximum number of sites, i.e.,  $F_{ST}$  estimates were based on sites that were polymorphic in at least one of the three species or divergent (if monomorphic) between at least two species. The progressiveMauve (Darling et al. 2010) was used for multiple sequence alignments of *D. silvestris*, *D. heteroneura*, *D. planitibia*, and *D. grimshawi*. PoPoolation (Kofler et al. 2011a) was used to estimate Tajima's  $\pi$  and Watterson's  $\theta$ , with the window size set to 10 Kb.  $N_e$  was calculated as  $4 \mu/\theta$ , where mutation rate  $\mu = 3.5 \times 10^{-9}$  was taken from *Drosophila melanogaster* (Keightley et al. 2009).

### ABBA/BABA Approach

We used a modified ABBA/BABA method (Green et al. 2010) using *D. grimshawi* (GRI) as an outgroup. Reference *D. grimshawi* genome sequences were downloaded from FlyBase (r1.3, <http://flybase.org>), and PLA, HET, and SIL reads were mapped using BWA (Li and Durbin 2010) with default parameters. We generated the genotypes for each species using GATK (DePristo et al. 2011) with default parameters except for setting heterozygosity to 0.01. Only sites with genotyping quality greater than 30 and minimal depth 10 were kept. After genotyping, we determined the  $G_1G_2G_3G_4$  pattern for each SNP position, where  $G_1$ ,  $G_2$ ,  $G_3$ , and  $G_4$  represent genotypes in HET, SIL, PLA, and GRI, respectively. GRI alleles were considered ancestral (A), and sites with more than two alleles were filtered out. For each site, species were either assigned a "A" or "B" in the four taxa pattern according to the agreement of its genotype with GRI. The "AABA" pattern is the position in which PLA carries derived allele "B," but HET and SIL carry ancestral allele "A" as GRI. Similarly, the "BBAA" pattern indicates that HET and SIL carry derived allele, while PLA carries the ancestral allele, and so forth. We used similar D statistic (Green et al. 2010) for AABA and BBAA and other combinations to test if the two patterns were of equivalent abundance/counts using a jackknife block size of 2 Mb. To evaluate the linkage between the four taxa pattern and sequence diversity, we estimated  $F_{ST}$  values in the neighborhood of each pattern  $p$  ( $p \in \{AABA, ABAA, ABBA, BAAA, BABA, BBAA, BBBA\}$ ). First, for each SNP from the pattern  $p$ , we calculated  $R_s$  (where  $s \in \{HS, HP, SP\}$  and H, S, P stands for HET, SIL and PLA, respectively), which is an average normalized  $F_{ST}$  of  $N$  downstream and  $N$  upstream SNPs (excluding the SNP itself). Normalized  $F_{ST}$  is obtained from the raw  $F_{ST}$  value divided by the mean  $F_{ST}$  for each species pairwise comparison  $s$ . In each block of the size of  $n$  bases, the

statistic  $D$  for  $s_1$  and  $s_2$  ( $s_1, s_2 \in \{HS, HP, SP\}$ ) was calculated:

$$D_{s_1, s_2; p} = \frac{\sum_{i=1}^n B_{s_1, s_2}(i)}{\sum_{i=1}^n C_p(i)},$$

where  $C_p(i)$  can be 1 or 0 according to the pattern  $p$  is seen in position  $i$  or not,  $B_{s_1, s_2}(i)$  is given by:

$$B_{s_1, s_2}(i) = \begin{cases} 1, & \text{if } R_{s_1}(i) > R_{s_2}(i) \\ 0, & \text{if } R_{s_1}(i) = R_{s_2}(i) \\ -1, & \text{if } R_{s_1}(i) < R_{s_2}(i) \end{cases}$$

Other than Fig. 3 with various  $N_s$ , results for  $N = 100$  only are reported throughout the text. Mean  $D$  and its variance were used to obtain Z-scores for jackknife tests, with the block size of 2 Mb, and a total of 100 blocks were generated. To reduce the effect of recent HET–SIL ancestry as a confounding factor in  $F_{ST}$  calculations for the HET–SIL pair, we used only fixed SNPs because fixation permits more time ( $\sim 4N_c$ ) for eroding the association by HET–SIL ancestral LD under the neutral model. For comparison, we used the same method and tested for all sites including non-fixed genotypes (total 7,359,805 sites, including 5,216,097 sites found in SIL and 5,215,281 sites found in HET, thus indicating similar mutation rates in SIL and HET). We obtained very similar results, including  $F_{ST}$  dynamics and ancestral branch estimates (see below). We also simulated alternative demographic models on ABBA/BABA expectations using MSMS software (Ewing and Hermisson 2010) and confirmed that our results are best explained by gene flow between species (Supplementary Table 1 and Supplementary Fig. 1).

### The Inference of Ancestral Relationships

Using a neutral coalescent model without gene flow, incomplete lineage sorting, or recurrent mutations, let us assume that AABA (a total of 118,181 mutations) accumulates only along the PLA branch (a–c, Supplementary Fig. 2). We also assume that XBAA (X can be either A or B) accumulates along the branch a–b–d leading to SIL (Let us ignore HET branch for now, as HET may have hybridized with an unknown species—see “Results and Discussion” section), and we note that lengths of branches a–c and a–b–d are the same (assuming equal mutation rates in all three species). Therefore, the number of XBAA should be equivalent to AABA. XBAA can be divided into three sets, with the first set accumulating divergence along the branch a–b belonging to the common SIL–HET ancestor, quantified as  $S_1$ , or the count of ancestral BBAA genotypes. The second set includes variants accumulated along the

branch b–d in SIL and exchanged with HET through hybridization ( $S_2$ , the count of recombinant BBAA genotypes). The third set includes variants accumulated along the branch b–d and unique to SIL, i.e., not exchanged with HET ( $S_3$ , the count of ABAA genotype = 42,187 mutations). Then  $S_1 + S_2 + S_3 = AABA$ , and thus  $S_2 = AABA - S_3 - S_1 = 75994 - S_1$ . If we assume that SIL and HET had equal contributions to the number of recombinant BBAA, then the BBAA count cannot exceed  $2S_2$ , that is,  $<151988 - 2S_1$ . Since the total number of BBAA is 132558 and should not exceed  $151988 - 2S_1 + S_1$ , we obtain  $132558 < 151988 - S_1 \Rightarrow S_1 < 19430$ . The estimate indicates that the ancestral branch a–b does not exceed 16.5% ( $19430/118181$ ) of the a–c PLA branch (see “Materials and Methods” section for details). We obtained a similar estimate (17.3%) including both fixed and non-fixed alleles.

### ABBA/BABA Simulations

MSMS software (Ewing and Hermisson 2010) was used to simulate different demographic models and their consequences on SNP patterns related to ABBA/BABA tests. Gene flow was simulated by adding a migration event (-m). A total of seven demographic models were simulated, each demographic model representing gene flow between.

- a) HET and SIL (m(HS));
- b) SIL and PLA (m(SP));
- c) HET and PLA (m(HP));
- d) SIL and PLA, and between HET and PLA (m(SP + HS));
- e) HET and PLA, and between HET and SIL (m(HP + HS));
- f) SIL and PLA, between HET and PLA, and between HET and SIL (m(HP + SP + HS)); and
- g) No gene flow (m(NULL)).

We assumed  $N_e$  to be 100,000, adjusted across species according to their estimated effective population sizes (-en 0.28 1 2 -en 0.28 2 1.40), and we simulated the recent decrease of population size (-en 0.002 1 0.20 -en 0.002 2 0.20 -en 0.002 3 0.04, (Foote 1995)). A total of 377,627 segregated sites were simulated, corresponding to the number of fixed genotype differences relative to *D. grimshawi*. Estimated divergence times from mcmctree were used to adjust the time of joint events (-ej). Average numbers of genotypes were counted based on 100 simulation cycles. The command line for m(HS) simulation was given as follows:

```
java -jar msms.jar -N 100000 -ms 3 100 -s 377627 -I 3 1 1 1 -t 1400 -r 1400 -en 0.28 1 2 -en 0.28 2 1.40 -ej 0.145 3
```

2 -ej 0.274 2 1 -en 0.002 1 0.20 -en 0.002 2 0.20 -en 0.002 3 0.04 -em 0.05 3 2 5 -threads 12.

We found the model m(HS) to be most consistent with our empirical results, thus corroborating the close sympatric relationship between these two species (Supplementary Table 1 and Supplementary Fig. 1).

## Results and Discussion

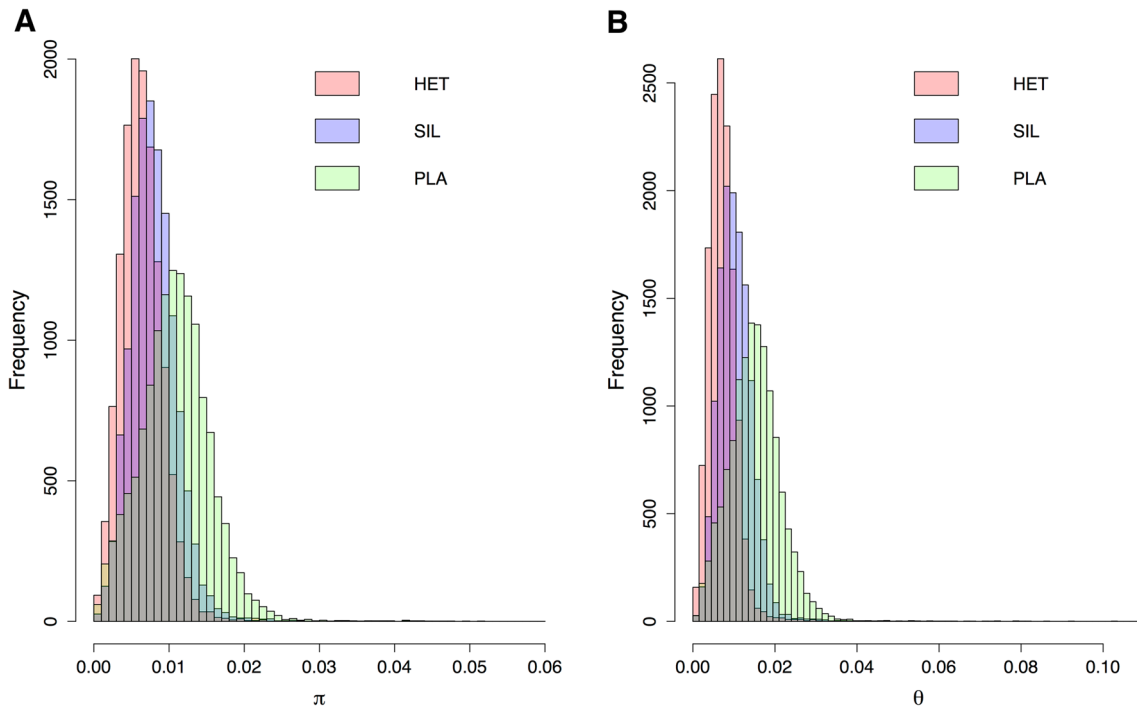
We used the ABBA/BABA method (Green et al. 2010) to contrast the two phylogenetic hypotheses, based on a total of 3,736,586 sites with fixed differences. To determine if paleo-PLA could be an ancestor of SIL and HET, we used a modified version of the ABBA/BABA method (Green et al. 2010) with *D. grimshawi* (GRI), a more distantly related Hawaiian picture-winged *Drosophila* (Clark et al. 2007) as an outgroup. We estimated the frequency of the BBAA (in the order: HET, SIL, PLA, GRI throughout the text) SNP pattern as a measure of the inter-node divergence time (a–b) between the most recent common SIL–HET ancestor and PLA ancestor (Supplementary Fig. 2). The BBAA pattern was abundant. It was enriched beyond a frequency easily attributable to inter-node ancestral divergence, as it exceeded the divergence of the PLA (AABA) branch by 11% (jackknifed Z-score = 3.78,  $P < 0.0002$ ). This also suggests that at least some SNPs shared by SIL and HET (BBAA) may have originated through introgression after sympatric interbreeding between SIL (ABAA) and HET (BAAA). This potential genetic admixture between SIL and HET is consistent with previous observations that up to 2% of  $F_1$  and backcross SIL–HET hybrids were found in some localities of Hawaii (Carson 1989). After subtracting the estimates of recombinant BBAA from the total BBAA count (see “Materials and Methods” section for details), we found that the length a–b (to HET–SIL putative ancestor) branch was 16.5% of the a–c (to PLA) branch or less. This coalesces PLA very close to the most recent ancestral node for HET and SIL, indicating that the PLA lineage was likely ancestral to both HET and SIL.

The relatively high frequency of the PLA-specific AABA SNP pattern indicates divergence time almost twice as long as that in the PLA branch relative to the HET (BAAA) and SIL (ABAA) branches. This is roughly consistent with the difference in the geological ages of Maui and Hawaii Islands and the hybrid sterility observed between PLA and both HET and SIL (Brill et al. 2016). Additionally, the HET-specific BAAA SNP pattern is enriched relative to the SIL-specific ABAA pattern (jackknifed Z-score = 10.54). This possibly reflects elevated genetic drift in HET due to the dramatically reduced

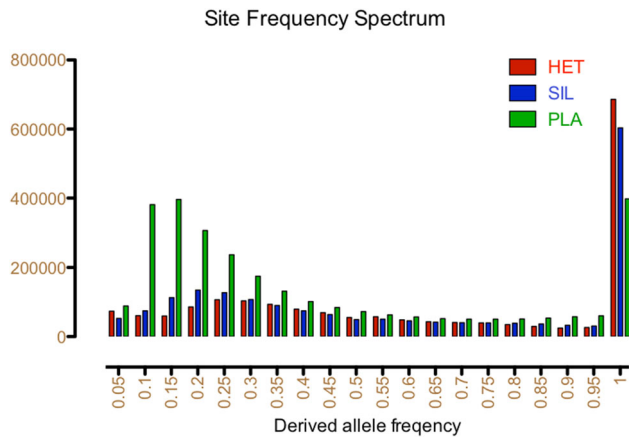
population size that has led to its endangered species status (Muir and Price 2008), or reflects the admixture between HET and an unknown species or population having occurred after the HET–SIL split. Using Tajima’s estimates of nucleotide diversity  $\pi$  and Watterson’s  $\theta$  (Fu 1994) in HET ( $\pi = 0.00647$ ,  $\theta = 0.0073$ ), SIL ( $\pi = 0.00798$ ,  $\theta = 0.0101$ ), and PLA ( $\pi = 0.0109$ ,  $\theta = 0.0125$ ) (Fig. 1), as well as mutation rate estimates  $\mu$  from *D. melanogaster* ( $\mu = 3.5 \times 10^{-9}$  (Keightley et al. 2009)), we estimated effective population sizes  $N_e$  for HET, SIL, and PLA to be 504,286, 728,571, and 1,057,143, respectively. This smaller  $N_e$  of *D. heteroneura* suggests that drift effects in this species may indeed be stronger compared with the other species. From the shapes of the site frequency spectrum (or allele frequency spectrum) of the three species, SIL and HET have more similar demographic configurations compared to that of PLA (Fig. 2).

We also found a high frequency of both ABBA and BABA, two SNP patterns that are incongruent with phylogeny. Their presence can be explained by (1) hybridization between PLA and SIL (ABBA) and PLA and HET (BABA), (2) recurrent mutations, or (3) incomplete lineage sorting (Green et al. 2010). Remarkably, the ABBA/BABA test (Green et al. 2010) indicates significant hybridization between HET and PLA (mean =  $-0.087$ , SE = 0.022, jackknifed Z-score =  $-4.03$ ). However, caution is needed in interpreting these results, since BAAA is overrepresented relative to ABAA, which could be a violation of substitution or fixation rate constancy. We notice that the ABBA/BABA test is in fact no longer significant (mean =  $-0.006$ , SE = 0.007, jackknifed Z =  $-0.92$ ) when both polymorphic and fixed sites are counted, as opposed to fixed sites only; and BAAA and ABAA numbers are then less dissimilar (301,326 and 226,184, respectively).

To explore hybridization as a source of ABBA/BABA enrichment, we reasoned that recently introgressed haplotypes have had insufficient time to be broken down by recombination, and that closely linked introgressed alleles should therefore occur in linkage-disequilibrium ‘islands’ (Martin et al. 2013; Sankararaman et al. 2012), which, in turn, should exhibit reduced sequence divergence in recipient-donor species comparisons relative to non-introgressed regions. To narrow down potential introgression ‘islands,’ we focused on the neighborhoods of all informative SNP patterns from the ABBA/BABA approach and their species pairwise  $F_{ST}$  values (Table 1). A number of very specific predictions can be tested using this new method. For example, if ABBA and BABA reflect interbreeding between PLA and, respectively, SIL and HET,  $F_{ST}$  values from comparisons involving PLA should be significantly decreased in the neighborhoods of these SNP patterns. Such a pattern was indeed observed. In fact,



**Fig. 1** **a** Distribution of the mean number of pairwise differences (Tajima’s  $\pi$ ) and **b** Watterson’s  $\theta$  for *D. heteroneura* (red), *D. silvestris* (blue), and *D. planitibia* (green). Both  $\pi$  and  $\theta$  values were calculated based on a window size of 10 Kb (Color figure online)



**Fig. 2** Site frequency spectrum of three species: *D. heteroneura*, *D. silvestris*, and *D. planitibia*

**Table 1** Frequencies of trans-species single nucleotide polymorphisms (SNPs) with their corresponding  $F_{ST}$  values (normalized by mean  $F_{ST}$  value for all three pairwise comparisons)

Genotype	Normalized $F_{ST}$ (divided by mean $F_{ST}$ )			
	SNPs	HET–SIL	HET–PLA	SIL–PLA
BAAA	71620	1.31123	1.10063	1.03658
ABAA	42187	1.39469	1.05738	1.10928
AABA	118181	1.05111	1.19512	1.15716
BBAA	132558	0.94353	1.08849	1.0768
ABBA	5969	1.33074	1.08408	1.03209
BABA	7112	1.48349	1.02171	1.09017
BBBA	3,358,959	0.98716	0.98653	0.989

Species order in each genotype: *Drosophila heteroneura* (HET), *D. silvestris* (SIL), *D. planitibia* (PLA), and *D. grimshawi*

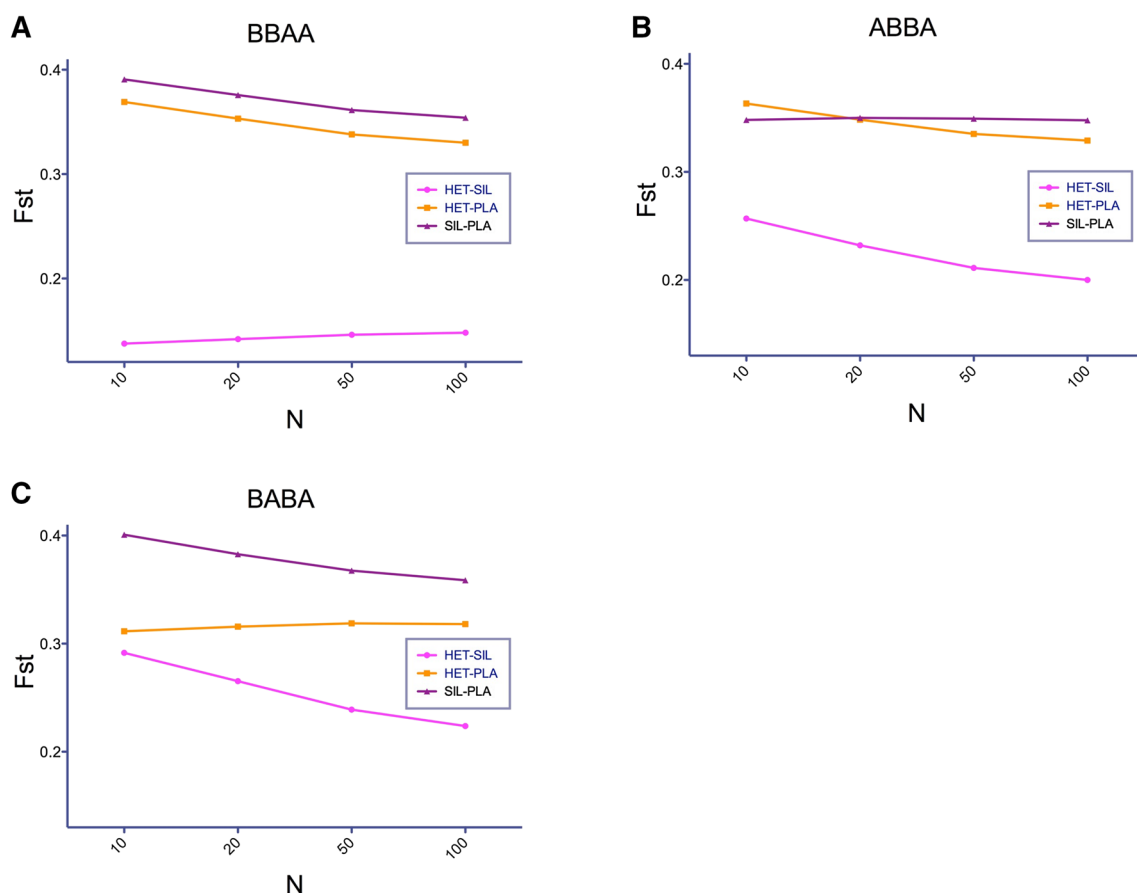
ABBA and BABA produced the largest and most statistically significant differences between normalized  $F_{ST}$  values (jackknifed  $7.05 < |Z\text{-scores}| < 17.58$ , Supplementary Table 2), exceeding the estimates for BBAA. This is expected only if BBAA was a mixture of ancestral and HET–SIL post-hybridization introgressive genotypes, whereas ABBA and BABA were mostly post-hybridization-introgressed genotypes. For comparison, relatively low differences between normalized  $F_{ST}$  values were found within AABA, a genotype specific of PLA divergence with  $F_{ST}$  estimates least likely to have been affected by

introgression/recombination on Hawai’i Island. We also predicted that if BBAA derives from recent interbreeding and introgression between HET and SIL, there should be a significant decrease in  $F_{ST}$  between HET and SIL, but not in PLA–SIL or PLA–HET, across the BBAA neighborhoods. Indeed, not only did we observe a dramatic ( $2.1\times$ ) decrease in mean HET–SIL  $F_{ST}$  values, but we also found that the lowest  $F_{ST}$  values within the HET–SIL comparison were produced by BBAA, as expected under introgressive hybridization (Table 1; jackknifed  $|Z\text{-scores}| > 11.03$ ).

Alternatively,  $F_{ST}$  values around BBAA will necessarily be biased downwards if the speciation time is much more recent than the ultimate time to coalesce or, in other words, BBAA are surrounded by neighborhoods of LD spikes originating from the HET–SIL common ancestor rather than from recent introgressions. To minimize this bias, we focused on fixed SNPs, which under a neutral model require on average  $\sim 4N_c$  generations for fixation, a long and likely sufficient time to erase most of the internal branches corresponding to a shallower coalescent history. Although this time-dependent homogenizing effect of recombination applies to both ancestral and newly introgressed linkage groups, our approach introduces an ascertainment bias in favor of recent LD spikes. Additional support for enrichment of the configurations BBAA, BABA, and ABBA due to introgressive hybridization comes from the analysis of species pairwise  $F_{ST}$  with respect to genomic window sizes used (Fig. 3). Unlike HET–PLA and SIL–PLA, HET–SIL  $F_{ST}$  values from BBAA neighborhoods tend to grow with the window size, consistent with introgression erosion due to recombination (Fig. 3a). Similarly, SIL–PLA  $F_{ST}$  values tend to increase

only in ABBA neighborhoods (Fig. 3b), while HET–SIL  $F_{ST}$  values rise only in BABA neighborhoods (Fig. 3c). This suggests possible introgressive hybridization between PLA and SIL, as well as between PLA and HET, respectively. All other SNP patterns produce  $F_{ST}$  values slightly declining with the window size, which is a typical genome-wide tendency due to the greater sampling error in smaller windows (Beissinger et al. 2015). Thus in addition to previously reported SIL–HET hybridization (Carson 1989), there may have been occasional interbreeding between PLA and the other two species, after migrants had successfully crossed the 26-mile wide Alenuihaha Channel separating Maui from Hawai'i. Although the direction of the inter-island migration and hybridization remains unknown, a Maui–Hawaii flyover by PLA is a more parsimonious scenario as compared to HET and SIL both migrating back to Maui.

In addition to *D. silvestris* and *D. heteroneura*, natural hybridization has been documented only for few other Hawaiian picture-winged *Drosophila* species: *D. setosimentum* and *D. ochrobasis*, and *D. hawaiiensis* and *D. engyochracia* (Carson et al. 1975; Kaneshiro 1990; Price



**Fig. 3** Average  $F_{ST}$  values in the neighborhoods of SNP patterns with respect to species pairwise comparisons and window sizes ( $N$ ). **a** BBAA. **b** ABBA. **c** BABA

and Muir 2008). However, the frequency of hybridization in Hawaiian picture-winged *Drosophila* in nature assessed with traditional methods has likely been underestimated because of the extensive analyses required to detect hybrids in the wild (Carson et al. 1975; Kaneshiro 1990; Price and Muir 2008). New genome-wide methods hold promise for investigating the extent and time course of genomic admixture, both before speciation and during different time periods following speciation (Martin et al. 2013; Sousa and Hey 2013). The rapid speciation and radiation of Hawaiian picture-winged *Drosophila* (Craddock 2000), and plausible changes in mating behavior in small populations (Kaneshiro 1983), suggest that the Hawaiian picture-winged *Drosophila* may be particularly prone to hybridization, which in turn may compromise the genetic integrity of some of the species, potentially posing a serious conservation challenge.

## References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455
- Beissinger TM, Rosa GJM, Kaepler SM, Gianola D, de Leon N (2015) Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet Sel Evol*. doi:10.1186/s12711-015-0105-9
- Brill E, Kang L, Michalak K, Michalak P, Price DK (2016) Hybrid sterility and evolution in Hawaiian *Drosophila*: differential gene and allele-specific expression analysis of backcross males. *Heredity (Edinb)* 117:100
- Carson HL, Nair PS, Sene FM (1975) *Drosophila* hybrids in nature—proof of gene exchange between sympatric species. *Science* 189:806
- Carson HL, Kaneshiro KY, Val FC (1989) Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura*. *Evolution* 43:190
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comeron JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipinski A, Findeiss S, Freyhult E, Fulton L, Fulton R, Garcia AC, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J, Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jaffe DB, Jagadeeshan S, Jeck WR, Johnson J, Jones CD, Jordan WC, Karpen GH, Kataoka E, Keightley PD, Kheradpour P, Kirkness EF, Koerich LB, Kristiansen K, Kudrna D, Kulathinal RJ, Kumar S, Kwok R, Lander E, Langley CH, Lapoint R, Lazzaro BP, Lee SJ, Levesque L, Li R, Lin CF, Lin MF, Lindblad-Toh K, Llopart A, Long M, Low L, Lozovsky E, Lu J, Luo M, Machado CA, Makalowski W, Marzo M, Matsuda M, Matzkin L, McAllister B, McBride CS, McKernan B, McKernan K, Mendez-Lago M, Minx P, Mollenhauer MU, Montooth K, Mount SM, Mu X, Myers E, Negre B, Newfield S, Nielsen R, Noor MA, O'Grady P, Pachter L, Papaceit M, Parisi MJ, Parisi M, Parts L, Pedersen JS, Pesole G, Phillippy AM, Ponting CP, Pop M, Porcelli D, Powell JR, Prohaska S, Pruitt K, Puig M, Quesneville H, Ram KR, Rand D, Rasmussen MD, Reed LK, Reenan R, Reily A, Remington KA, Rieger TT, Ritchie MG, Robin C, Rogers YH, Rohde C, Rozas J, Rubenfield MJ, Ruiz A, Russo S, Salzberg SL, Sanchez-Gracia A, Saranga DJ, Sato H, Schaeffer SW, Schatz MC, Schlenke T, Schwartz R, Segarra C, Singh RS, Sirot L, Sirota M, Sisneros NB, Smith CD, Smith TF, Spieth J, Stage DE, Stark A, Stephan W, Strausberg RL, Stempel S, Sturgill D, Sutton G, Sutton GG, Tao W, Teichmann S, Tobar YN, Tomimura Y, Tsolas JM, Valente VL, Venter E, Venter JC, Vicario S, Vieira FG, Vilella AJ, Villasante A, Walenz B, Wang J, Wasserman M, Watts T, Wilson D, Wilson RK, Wing RA, Wolfner MF, Wong A, Wong GK, Wu CI, Wu G, Yamamoto D, Yang HP, Yang SP, Yorke JA, Yoshida K, Zdobnov E, Zhang P, Zhang Y, Zimin AV, Baldwin J, Abdouelleil A, Abdulkadir J, Abebe A, Abere B, Abreu J, Acer SC, Aftuck L, Alexander A, An P, Anderson E, Anderson S, Arachi H, Azer M, Bachantsang P, Barry A, Bayul T, Berlin A, Bessette D, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Bourzgui I, Brown A, Cahill P, Channer S, Cheshatsang Y, Chuda L, Citroen M, Collymore A, Cooke P, Costello M, D'Aco K, Daza R, De Haan G, DeGray S, DeMaso C, Dhargay N, Dooley K, Dooley E, Doricent M, Dorje P, Dorjee K, Dupes A, Elong R, Falk J, Farina A, Faro S, Ferguson D, Fisher S, Foley CD, Franke A, Friedrich D, Gadbois L, Gearin G, Gearin CR, Giannoukos G, Goode T, Graham J, Grandbois E, Grewal S, Gyaltsen K, Hafez N, Hagos B, Hall J, Henson C, Hollinger A, Honan T, Huard MD, Hughes L, Hurhula B, Husby ME, Kamat A, Kanga B, Kashin S, Khazanovich D, Kisner P, Lance K, Lara M, Lee W, Lennon N, Letendre F, LeVine R, Lipovsky A, Liu X, Liu J, Liu S, Lokysang T, Lokysang Y, Lubonja R, Lui A, MacDonald P, Magnisalis V, Maru K, Matthews C, McCusker W, McDonough S, Mehta T, Meldrim J, Meneus L, Mihai O, Mihalev A, Mihova T, Mittelman R, Mlenga V, Montmayeur A, Mulrain L, Navidi A, Naylor J, Negash T, Nguyen T, Nguyen N, Nicol R, Norbu C, Norbu N, Novod N, O'Neill B, Osman S, Markiewicz E, Oyono OL, Patti C, Phunkhang P, Pierre F, Priest M, Raghuraman S, Rege F, Reyes R, Rise C, Rogov P, Ross K, Ryan E, Settupalli S, Shea T, Sherpa N, Shi L, Shih D, Sparrow T, Spaulding J, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Strader C, Tesfaye S, Thomson T, Thoulutsang Y, Thoulutsang D, Topham K, Topping I, Tsamla T, Vassiliev H, Vo A, Wangchuk T, Wangdi T, Weiland M, Wilkinson J, Wilson A, Yadav S, Young G, Yu Q, Zembek L, Zhong D, Zimmer A, Zwirko Z, Alvarez P, Brockman W, Butler J, Chin C, Grabherr M, Kleber M, Mauceli E, MacCallum I (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203
- Craddock EM (2000) Speciation processes in the adaptive radiation of Hawaiian plants and animals. *Evol Biol* 31(31):1
- Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna

- A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491
- Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28:2239
- Ewing G, Hermisson J (2010) MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26:2064
- Feder JL, Nosil P, Wacholder AC, Egan SP, Berlocher SH, Flaxman SM (2014) Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *J Hered* 105:810
- Foote D, Carson HL (1995) *Drosophila* as monitors of change in Hawaiian ecosystems. In: LaRoe ET, Farris GS, Puckett CE, Doran PD, Mac MJ (eds) *Our living resources: a report to the nation on the distribution, abundance, and health of US plants, animals, and ecosystems*. U.S. Department of the Interior, National Biological Service, Washington, DC, pp 368–372
- Fu YX (1994) Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. *Genetics* 138:1375
- Gagnaire PA, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution* 67:2483
- Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prufer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Hober B, Hoffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Z, Gusic I, Doronichev VB, Golovanova LV, Lalueva-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PL, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Paabo S (2010) A draft sequence of the Neandertal genome. *Science* 328:710
- Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491
- Jonsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliusen TS, Vilstrup JT, Lear T, Myka JL, Lundquist J, Miller DC, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Stagegaard J, Strauss G, Bertelsen MF, Sicheritz-Ponten T, Antczak DF, Bailey E, Nielsen R, Willerslev E, Orlando L (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A* 111:18655
- Kaneshiro KY (1983) Sexual selection and direction of evolution in the biosystematics of hawaiian *Drosophilidae*. *Annu Rev Entomol* 28:161
- Kaneshiro KY (1990) Natural hybridization in *Drosophila*, with special reference to species from Hawaii. *Can J Zool* 68:1800
- Kang L, Settlage R, McMahon W, Michalak K, Tae H, Garner HR, Stacy EA, Price DK, Michalak P (2016) Genomic signatures of speciation in sympatric and allopatric Hawaiian picture-winged *Drosophila*. *Genome Biol Evol* 8:1482
- Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res* 19:1195
- Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, Wittwer S, Seehausen O (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Mol Ecol* 22:2848
- Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlotterer C (2011a) PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* 6:e15925
- Kofler R, Pandey RV, Schlotterer C (2011b) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078
- Magnacca KN, Price DK (2015) Rapid adaptive radiation and host plant conservation in the Hawaiian picture wing *Drosophila* (Diptera: *Drosophilidae*). *Mol Phylogenet Evol* 92:226
- Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res* 23:1817
- Muir C, Price DK (2008) Population structure and genetic diversity in two species of Hawaiian picture-winged *Drosophila*. *Mol Phylogenet Evol* 47:1173
- Nosil P (2008) Speciation with gene flow could be common. *Mol Ecol* 17:2103
- Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061
- Price JP, Clague DA (2002) How old is the Hawaiian biota? Geology and phylogeny suggest recent divergence. *Proc Biol Sci* 269:2429
- Price DK, Muir C (2008) Conservation implications of hybridization in Hawaiian picture-winged *Drosophila*. *Mol Phylogenet Evol* 47:1217
- Sankararaman S, Patterson N, Li H, Paabo S, Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genet* 8:e1002947
- Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Mol Ecol* 20:5123
- Sousa V, Hey J (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nat Rev Genet* 14:404
- Stacy EA, Paritosh B, Johnson MA, Price DK (2017) Incipient ecological speciation between successional varieties of a dominant tree involves intrinsic postzygotic isolating barriers. *Ecol Evol* 7:2501