Sequential adaptive introgression of the mitochondrial genome in Drosophila yakuba and Drosophila santomea

Ana Llopart
University of Iowa

Danielle Herrig
University of Iowa

Evgeny Brud
University of Iowa

Please see article for additional authors.

DOI: https://doi.org/10.1111/mec.12678

Copyright: © 2014 The Authors

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.


Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.
Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*

ANA LLOPART,*† DANIELLE HERRIG,† EVGENY BRUD* and ZACHARY STECKLEIN*

*Department of Biology, University of Iowa, Iowa City, IA 52242, USA, †Interdisciplinary Graduate Program in Genetics, University of Iowa, Iowa City, IA 52242, USA

Abstract

Interspecific hybridization provides the unique opportunity for species to tap into genetic variation present in a closely related species and potentially take advantage of beneficial alleles. It has become increasingly clear that when hybridization occurs, mitochondrial DNA (mtDNA) often crosses species boundaries, raising the possibility that it could serve as a recurrent target of natural selection and source of species’ adaptations. Here we report the sequences of 46 complete mitochondrial genomes of *Drosophila yakuba* and *Drosophila santomea*, two sister species known to produce hybrids in nature (~3%). At least two independent events of mtDNA introgression are uncovered in this study, including an early invasion of the *D. yakuba* mitochondrial genome that fully replaced the *D. santomea* mtDNA native haplotypes and a more recent, ongoing event centered in the hybrid zone. Interestingly, this recent introgression event bears the signature of Darwinian natural selection, and the selective haplotype can be found at low frequency in Africa mainland populations of *D. yakuba*. We put forward the possibility that, because the effective population size of *D. santomea* is smaller than that of *D. yakuba*, the faster accumulation of mildly deleterious mutations associated with Muller’s ratchet in the former species may have facilitated the replacement of the mutationally loaded mitochondrial genome of *D. santomea* by that of *D. yakuba*.

Keywords: gene flow, positive selection, recurrent adaptation, speciation

Received 30 September 2013; revision received 9 January 2014; accepted 15 January 2014

Introduction

Shifting environments challenge species as they often throw them off their fitness optimum, and the role of adaptation is to move them to a new phenotypic optimum (Orr 2002, 2005). This can typically be achieved using standing genetic variation (i.e. variation present in the species at the time of the environmental change) or *de novo* mutations. Yet there is another way for species to acquire beneficial alleles: introgression via inter-specific hybridization (Rieseberg & Wendel 1993; Arnold 1997; Whitney et al. 2006). Adaptive introgression was first recognized by evolutionary botanists (Anderson 1949; Stebbins 1959) and later detected in animal species (Lewontin & Birch 1966). Although the initial hybridization is frequently deleterious, introgression provides the opportunity for species to tap into additional genetic variation of a closely related species and potentially take advantage of raw material for adaptation (Seehauseen 2004; Mallet 2005; Whitney et al. 2006; Arnold 2007; Baack & Rieseberg 2007; Castric et al. 2008; Rieseberg 2009; Consortium 2012). Universally beneficial alleles can spread easily across species boundaries when isolating barriers are incomplete (Coyne & Orr 2004) and produce the signature of a trans-species selective sweep (Barton & Gale 1993; Hilton et al. 1994; Stephan et al. 1998; Machado & Hey 2003; Llopart et al. 2005b; Teeter et al. 2008; Staubach et al. 2012; Brand et al. 2013).

For several reasons introgression may represent a particularly efficient way for species to adapt. First, introgression can enable populations to respond rapidly to sudden environmental changes where fast response is crucial. Cases of resistance to anthropogenic agents, such as the anticoagulant rodenticide warfarin,
constitute textbook examples (Rieseberg 2011; Song et al. 2011). Interestingly, adaptation through introgression often involves a change in the fitness value of already existing alleles that were originally either neutral or weakly deleterious before the environmental shift. This change in the fitness value is an attribute also observed in adaptation through standing genetic variation within species (Orr & Betancourt 2001; Hermisson & Pennings 2005; Barrett & Schluter 2008; Peter et al. 2012; Hedrick 2013). Second, alleles lost to random genetic drift or deleterious alleles fixed in a population may be easier to replace with alleles from a closely related species that have been pretested by natural selection (Kim et al. 2008; Rieseberg 2009). This could be particularly relevant to the introgression of nonrecombining genomes with high mutation rates (e.g. mitochondrial DNA in Drosophila) in species with small effective population sizes (N_e) (e.g. endemics) as Muller’s ratchet, the irreversible accumulation of mildly deleterious mutations, is expected to proceed faster in such species and thus result in mutationally loaded nonrecombining genomes (Lynch & Gabriel 1990; Charlesworth et al. 1993b; Gabriel et al. 1993; Loewe 2006). Lastly, the observation of novel traits in hybrids that are not present in either of the parental species (i.e. transgressive segregation) (Rieseberg et al. 1999, 2003) raises the interesting possibility that the placement of several genes of one species in the genetic background of another through introgression may result in new traits. The pervasive, nonadditive nature of gene regulation suggests that changes in expression could be underlying these novel phenotypes (Gibson & Weir 2005; Lai et al. 2006; Landry et al. 2007).

Traditionally viewed as fortuitous accidents of nature, hybrid zones provide the unique opportunity to study introgression (Barton & Hewitt 1985, 1989; Hewitt 1988; Harrison & Rand 1989; Harrison 1993). On the small African island of São Tomé, Drosophila yakuba and Drosophila santomea, a pair of very recently diverged sister species in the melanogaster subgroup, form a well-characterized hybrid zone where hybrids are abundant (~3–5%) (Llopart et al. 2005a). The two species began diverging allopatrically ~400 000 years ago (Llopart et al. 2005b), but today are both present on São Tomé due to secondary invasion by D. yakuba. (D. santomea is endemic to São Tomé.) Despite their recent divergence, D. yakuba and D. santomea show traditional forms of reproductive isolation, including sexual isolation (Lachaise et al. 2000; Coyne et al. 2002; Moehring et al. 2006b), conspecific sperm precedence (Chang 2004), prezygotic reinforcement (Matute 2010) and hybrid sterility that conforms to Haldane’s rule (Cariou et al. 2001; Coyne et al. 2004; Moehring et al. 2006a; Matute & Coyne 2009).

Two previous studies in these species reported introgression of mitochondrial DNA (mtDNA) based on lack of reciprocal monophony (Llopart et al. 2005b; Bachtrog et al. 2006). Neither study, however, had adequate statistical power to assess the selective nature of the process, rule out alternative neutral scenarios (i.e. population expansion or strong purifying selection) and obtain a detailed characterization (Llopart et al. 2005b; Bachtrog et al. 2006). To determine the evolutionary forces leading to mtDNA introgression in D. yakuba and D. santomea and understand their temporal dynamics, we sequenced 46 complete mitochondrial genomes.

These sequences represent East, Central and West African populations of D. yakuba as well as populations of both species from São Tomé. We found that introgression of the mitochondrial genome is recurrent in these species, with an ancient event in which the D. yakuba mtDNA completely replaced the D. santomea native mitochondrial genome and an ongoing introgression event centred in the D. yakuba–D. santomea hybrid zone. Interestingly, the ongoing introgression shows the signature of positive natural selection and is associated with a haplotype present at low frequency on the Africa mainland. We propose that in species with very different N_e, Muller’s ratchet (Lynch & Gabriel 1990; Charlesworth et al. 1993b; Gabriel et al. 1993; Loewe 2006) can have a significant influence on mitochondrial introgression resulting in the rescue of the more mutationally loaded mtDNA of the species with small N_e by the mtDNA of the species with large N_e.

**Methods**

*Drosophila* lines

A total of 46 isofemale lines, 29 Drosophila yakuba and 17 Drosophila santomea, were used in this study. To assess the direction of the mitochondrial introgression, we obtained a geographically diversified sample of D. yakuba, including 13 allopatric lines from mainland Africa (West Coast: Ivory Coast, Cameroon and Gabon; Central: Zimbabwe; East Coast: Kenya and Tanzania) and 16 lines collected in São Tomé. One of these island lines was collected in a garden outside São Tomé City, in an area where only D. yakuba is found, and the remaining 15 lines were established from females collected in the D. yakuba–D. santomea hybrid zones (Llopart et al. 2005a,b). To take into account possible population structure in D. santomea, our sampling strategy covered two different geographical areas where D. yakuba is also abundant, the hybrid zone in the Obo Natural Reserve (13 lines) and Rio Queijo in southwest São Tomé (4 lines). A detailed description of the lines is included in Table S1 (Supporting information).
DNA extraction and sequencing strategies

Sequencing was carried out using a combination of two different strategies: standard cycle sequencing preceded by PCR-based amplification (14 complete mitochondrial genomes) and next-generation sequencing (32 complete mitochondrial genomes). In the standard approach, mtDNA was isolated following Afonso et al. (1988) with minor modifications. This is a crude method for the selective isolation of covalently closed circular DNA molecules. The first step involves an alkaline denaturation followed by an acid/phenol extraction. To amplify the entire mtDNA prior to sequencing, we generated 12 overlapping PCR fragments using primers designed on a complete sequence of the D. yakuba mitochondrial genome [accession no. in GenBank X03240; (Clary & Wolstenholme 1985)]. (Primer sequences are available upon request from the authors.) PCR products were purified using the Wizard MagneSil PCR clean-up system (Promega, Madison, WI). Both strands of DNA were sequenced directly using the BigDye Terminator kit v3.1 chemistry and resolved in an ABI 3730 DNA Analyzer (Life Technologies, Carlsbad, CA). Sequences of each line were assembled using the software SEQUENCHER 5.10 (Gene Codes, Ann Arbor, MI), and multiple alignments were obtained using ClustalX (Thompson et al. 1997).

Next-generation technology was also used to obtain the mtDNA sequence of 32 lines. RNA-free DNA extracted from single flies using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA) was fragmented in a Bioruptor UCD-200 (Diagenode, Denville, NJ). We prepared genomic DNA libraries following the Illumina protocol (‘Preparing Samples for Sequencing Genomic DNA’, Illumina, San Diego, CA). We used custom-designed adapters that have unique 7-nucleotide tags, which were added to standard Illumina adapters, and allowed multiplexing of several samples prior to sequencing. The sequences and a detailed description of these custom-designed adapters can be found in Comeron et al. (2012). Prior to multiplexing for sequencing, library quantitation was performed using the Quant-iT PicoGreen kit (Invitrogen, Carlsbad, CA). Cluster generation and sequencing (75 bp) were carried out at the Iowa State University DNA Facility (Ames, IA) using a GAIL instrument. Filtering of reads, mapping and generation of consensus sequences were carried out using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit), BWA (Li & Durbin 2009), SAMtools v1.4 (Li et al. 2009) and custom scripts. After mapping, the average coverage and consensus quality per nucleotide of the mitochondrial genomes were 130.2 x (median 115.1 x) and 222.1 (median 242.8), respectively. SNPs were called with quality Q40 or greater.

Analysis of introgression

To determine whether there has been introgression of the mitochondrial genome between D. yakuba and D. santomea and explore posterior probabilities of demographic parameters under an Isolation-with-Migration model (IM) (Nielsen & Wakeley 2001), we performed simulations using the program IMa2 (Hey & Nielsen 2004, 2007). Our IMa2 analysis included GenBank sequence X03240 and excluded the control region of the mitochondrial genome (positions 14943–16019 in X03240, positions 14990–16127 in the multiple sequence alignment). We assumed a simple demographic model in which ancestral populations do not experience migration and there is one migration parameter between each pair of sampled populations. Log-likelihood-ratio (LLR) tests were used to evaluate how two different models, with and without gene flow between populations, fit the data (Nielsen & Wakeley 2001; Hey 2010). Several preliminary runs were conducted to assess the mixing of the Markov chain. We tried different numbers of Metropolis-coupled chains and heating terms, always using a geometric scheme. We determined that 100–180 chains with a = 0.99 and β = 0.75 produce high swapping rates between adjacent chains and lead to good mixing (i.e. low autocorrelations, high effective sample sizes and no trends in the plots over the course of a run). Upper bounds of prior distributions were originally set following recommendations in the IMa2 manual and adjusted afterwards, as needed, based on the results of a complete first run. The burn-in period was carried out until the stationary phase was reached, which typically took between 1–4 million steps depending on the specific population model (e.g. two populations vs. three populations). To ensure convergence, we performed two well-mixed independent runs using different random number seeds, each with its own burn-in period, and confirmed that they produced similar estimates for all parameters. Results from two well-mixed runs based on no less than 100 000 genealogies each, saved after the burn-in period, were combined to obtain parameter estimates. In all analyses, we used the HKY mutational model (Hasegawa et al. 1985), which is adequate for Drosophila mitochondrial sequences. To determine the role of selection on mitochondrial introgression, we performed coalescent simulations with population growth using the program ms and obtained expectations under strict neutrality (Hudson 2002).

Dating introgression events

To estimate the time to most recent common ancestor (i.e. TMRCA), we used the BEAST v.1.7.4 software package...
(Drummond & Rambaut 2007; Drummond et al. 2012). BEAST, with its interface program BEAUti, performs Bayesian MCMC analysis of molecular sequences related by a tree that allows estimates of divergence times (Drummond & Rambaut 2007; Drummond et al. 2012). To this end, the models implemented in BEAST fix the external nodes of the tree to a specified date or rate and then sample the time of the internal nodes from their posterior probability distribution using MCMC. We assumed a strict molecular clock with no rate variation among lineages, which is reasonable with very closely related species such as D. yakuba and D. santomea, a gamma heterogeneity site model with a fraction of invariant sites (initial pinv. = 0.64) and a coalescent-based framework. The initial value of the fraction of invariant sites was determined identifying the model that best fit the data using Akaike information criterion tests in the software jModelTest (Guindon & Gascuel 2003; Darriba et al. 2012).

To calibrate the molecular clock, we considered that D. yakuba–D. santomea split from Drosophila erecta 10.4 Mya (Tamura et al. 2004), which correspond to 1.4446 synonymous changes per synonymous site as estimated using the program PAML (Yang 1997, 2007). We conducted two independent runs each of 100 million steps with 10 million steps of burn-in. In each run, we sampled genealogies every 1000 steps to produce 90 000 trees that were analysed with Tracer v.1.5 (http://beast.bio.ed.ac.uk/Tracer). This resulted in effective sample sizes greater than 2000 for all the parameters and in all analyses.

Gene genealogy reconstruction

Phylogenetic analyses were conducted using the program MEGA5 (Tamura et al. 2011) and the available sequence of D. erecta as the outgroup (Clark et al. 2007). We obtained gene genealogies using (1) the neighbour-joining (NJ) algorithm (Saitou & Nei 1987) with evolutionary distances corrected for multiple hits at a site (Kimura 1980) and (2) maximum-likelihood and maximum-parsimony methods. Statistical confidence in the nodes of the reconstructed gene tree was assessed using bootstrap (Felsenstein 1985) with 1000 replicates.

Results

Drosophila yakuba and Drosophila santomea mitochondrial genomes are not genetically different

We sequenced the complete mitochondrial genomes of 29 Drosophila yakuba (13 from mainland Africa and 16 from São Tomé) and 17 Drosophila santomea lines. Levels of polymorphism across the different genes are similar in D. yakuba and D. santomea (Table 1), in agreement with previous observations in the same species pair (Llopart et al. 2005b; Bachtrog et al. 2006). We detect population structure within D. yakuba, for there is significant genetic differentiation between the Africa mainland and São Tomé populations of this species (nearest-neighbour statistic $S_{nn} = 0.79$, permutation test $P < 0.001$) (Hudson 2000). However, this differentiation is based on differences in SNP frequencies rather than on the presence of fixed differences between populations. In striking contrast, mitochondrial haplotypes of D. santomea are not genetically distinct from those of the São Tomé population of D. yakuba ($S_{nn} = 0.54$, $P = 0.24$), the first hint of mtDNA introgression. This becomes particularly informative when we consider that genes on the Y chromosome, the other nonrecombining, uni-parentally inherited genomic element, do show significant genetic differentiation between D. yakuba and D. santomea ($S_{nn} = 1.0$, $P < 0.0001$), with numerous fixed differences between species (Llopart et al. 2005b). Our analysis of population differentiation strongly suggests that mitochondrial gene flow has indeed occurred from D. yakuba to D. santomea.

Significant post-split mitochondrial gene flow

To formally investigate whether gene flow between D. yakuba and D. santomea occurred after the colonization of São Tomé by the ancestral species (i.e. post-split gene flow), we fit an IM model to our data of complete mtDNA sequences using the software IMa2 (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). As polymorphism levels are low in the Drosophila mitochondrial genome due to strong functional constraints (Nabholz et al. 2013), maternal inheritance, and the effects of hitchhiking (Maynard Smith & Haigh 1974; Kaplan et al. 1989; Sella et al. 2009) and/or background selection (Charlesworth et al. 1993a; Charlesworth 2012) associated with the lack of recombination, there is not enough polymorphism information in the data to obtain precise maximum-likelihood estimates of population mutation rate parameters (i.e. $h$). To assess whether our results are robust to different $h$ values, we tested the null hypothesis of no gene flow under three different models using $h$ set to 1, 2.5 and 5× the value observed in D. yakuba mtDNA ($h$ per sequence = 29.79). In all conditions investigated, LLR tests indicated that a model in which post-split gene flow between D. yakuba and D. santomea had occurred was significantly more likely than a scenario with 0 migration (LLR test statistic $> 14.08$ in all cases, $P < 0.001$). In addition, marginal posterior probability distributions of the migration rate parameter ($m$) excluded the smallest value significantly, producing the unequivocal signal of gene flow (posterior $P < 1.2 \times 10^{-8}$, $P < 3.8 \times 10^{-6}$ and $P < 9.9 \times 10^{-5}$.

© 2014 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd.
for \( m = 0.004 \) in the 1, 2.5 and 5× models, respectively; Fig. 1). Note that these posterior probabilities suggest that ancestral populations with large \( N_e \), which are expected to have more ancestral polymorphism, make the test to zero migration conservative. The lack of fixed differences between the island populations (\( D. \) yakuba and \( D. \) santomea combined) and mainland \( D. \) yakuba strongly indicates that some, if not all, \( D. \) santomea mitochondrial genomes have been replaced by haplotypes from its sister species \( D. \) yakuba.

Although we were able to rule out isolation between \( D. \) yakuba and \( D. \) santomea, a model in which each species is represented by a single population is not very realistic considering that there is population structure within \( D. \) yakuba (see above). A more realistic scenario can be modelled using three populations, two \( D. \) yakuba populations (one mainland and one island) and one population of \( D. \) santomea. This scenario is also expected to be more sensitive to recent events of introgression. We fit a three-population IM model to our data set and investigated the possibility of recent introgression on São Tomé. LLR tests produced the same results across replicates and population sizes. There is significant post-split gene flow between \( D. \) santomea and the island population of \( D. \) yakuba (LLR test statistics = 3.38–4.84, \( P < 0.05 \); Fig. 2a) but not between the two \( D. \) yakuba populations, or the mainland \( D. \) yakuba population and \( D. \) santomea (LLR test statistics = 0–0.5, \( P > 0.05 \); Fig. 2b, c). In summary, IMa2 results are consistent with previous findings that suggested gene flow

### Table 1 Polymorphism data summary

| Locus | Species | \( n \) | \( S \) | \( \Pi_S \times 10^{-3} \) | \( \Pi_N \times 10^{-3} \) | Fixed | Shared | \( L \) (bp)
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ATPase 6</td>
<td>yak</td>
<td>30</td>
<td>8</td>
<td>7.97</td>
<td>0.51</td>
<td>0</td>
<td>2</td>
<td>672</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>6</td>
<td>9.11</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPase 8</td>
<td>yak</td>
<td>30</td>
<td>1</td>
<td>2.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>159</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COI</td>
<td>yak</td>
<td>30</td>
<td>14</td>
<td>3.82</td>
<td>0.11</td>
<td>0</td>
<td>3</td>
<td>1533</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>9</td>
<td>5.57</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COII</td>
<td>yak</td>
<td>30</td>
<td>7</td>
<td>4.03</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>684</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>3</td>
<td>5.11</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COIII</td>
<td>yak</td>
<td>30</td>
<td>6</td>
<td>0.74</td>
<td>0.44</td>
<td>0</td>
<td>0</td>
<td>786</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>3</td>
<td>1.31</td>
<td>0.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cyt_b</td>
<td>yak</td>
<td>30</td>
<td>10</td>
<td>4.77</td>
<td>0.22</td>
<td>0</td>
<td>3</td>
<td>1134</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>6</td>
<td>3.03</td>
<td>0.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND1</td>
<td>yak</td>
<td>30</td>
<td>5</td>
<td>2.40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>972</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>2</td>
<td>0.55</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND2</td>
<td>yak</td>
<td>30</td>
<td>6</td>
<td>3.82</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
<td>1023</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>3</td>
<td>3.10</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND3</td>
<td>yak</td>
<td>30</td>
<td>3</td>
<td>1.88</td>
<td>0.47</td>
<td>0</td>
<td>0</td>
<td>351</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>1</td>
<td>1.66</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND4</td>
<td>yak</td>
<td>30</td>
<td>14</td>
<td>3.19</td>
<td>0.38</td>
<td>0</td>
<td>3</td>
<td>1338</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>6</td>
<td>3.35</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND4L</td>
<td>yak</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>0.29</td>
<td>0</td>
<td>0</td>
<td>288</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND5</td>
<td>yak</td>
<td>30</td>
<td>27</td>
<td>6.62</td>
<td>0.58</td>
<td>0</td>
<td>10</td>
<td>1719</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>23</td>
<td>10.59</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND6</td>
<td>yak</td>
<td>30</td>
<td>5</td>
<td>5.96</td>
<td>0.32</td>
<td>0</td>
<td>1</td>
<td>522</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>2</td>
<td>6.0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total coding</td>
<td>yak</td>
<td>30</td>
<td>107</td>
<td>4.13</td>
<td>0.32</td>
<td>0</td>
<td>24</td>
<td>11181</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>65</td>
<td>4.76</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total noncoding</td>
<td>yak</td>
<td>30</td>
<td>31</td>
<td>0.81</td>
<td>–</td>
<td>0</td>
<td>11</td>
<td>4949</td>
</tr>
<tr>
<td></td>
<td>san</td>
<td>17</td>
<td>25</td>
<td>1.13</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \text{yak}, \text{Drosophila yakuba and san, Drosophila santomea.} \)

\*Sample size.

\( \text{Number of polymorphic sites.} \)

\( \text{Nucleotide diversity at synonymous, or silent in noncoding regions (}\Pi_S, \text{and nonsynonymous (}\Pi_N\text{) sites calculated with DnaSP v.5 (Librado & Rozas 2009).} \)

\( \text{Differences fixed between } D. \text{ yakuba and } D. \text{ santomea and shared polymorphisms.} \)

\( \text{Size of the sequenced region (including alignment gaps).} \)
of mtDNA between D. yakuba and D. santomea (Llopart et al. 2005b; Bachtrog et al. 2006).

Ongoing adaptive introgression in the mitochondrial genome

To investigate the temporal dynamics of the mitochondrial introgression on São Tomé, we applied a phylogenetic approach. Gene trees obtained using different methods (i.e. distance, parsimony and maximum likelihood) revealed the same pattern: a distinct, starlike clade strongly supported by high bootstrap values (>90%) clusters together 13 D. yakuba and 8 D. santomea sequences (21 of 47 total sequences; 46 newly reported here and X03240 from GenBank) (Fig. 3). This clade is significantly enriched in sequences derived from the classic hybrid zone (HZ) of these two species (Llopart et al. 2005a) (16/21 in the HZ clade vs. 10/26 outside the HZ clade; FET, P = 0.017). The starlike topology is consistent with a rapid increase in frequency of a single haplotype (i.e. HZ haplotype) in the population due to either population growth or positive selection. The presence of both D. yakuba and D. santomea sequences in this clade indicates that the HZ haplotype spread through the population after it transferred from one species to the other. This suggests that the HZ clade potentially represents an ongoing adaptive introgression event of the mitochondrial genome.

To formally assess the adaptive nature of the ongoing introgression, we performed neutral coalescent simulations using the program ms (Hudson 2002). We simulated population growth and evaluated whether the severe reduction in nucleotide variation in the HZ clade could be explained by demographic factors. Our simulations are thus conditional on the observed number of total polymorphisms and the frequency spectrum, which is skewed towards rare variants, as captured by the summary statistic Tajima’s D (Tajima 1989) and in agreement with population growth. We analysed a subset of 10,000 genealogies showing an average $D = -2.20 \pm 1$ SD. (Note that this value in our simulations closely resembles our observation in the entire sample of D. yakuba and D. santomea mitochondrial genomes, $D = -2.24$.) For each simulation, we tracked...
the number of polymorphisms present in a cluster formed by the 21 most closely related sequences. Our results indicate that there is a significant deficit of polymorphisms in the HZ clade given the overall distribution of variation in the entire gene tree (23 or fewer polymorphisms in a cluster containing 21 of 47 sequences, \( P < 0.0001 \)). Similar results were obtained when we analysed \( D. yakuba \) (13 sequences with 15 or fewer polymorphisms in a clade of 21 sequences, \( P < 0.0001 \)) and \( D. santomea \) (eight sequences with eight or fewer polymorphisms in a clade of 21 sequences, \( P = 0.008 \)) separately. We conclude that the presence of a haplotype at a frequency of >40% cannot be explained by population expansion and is thus consistent with positive selection. The HZ clade constitutes a case of ‘caught-in-the-act’ mitochondrial adaptive introgression.

**Sequential introgression events**

To determine whether we could formally rule out the possibility of a single introgression event, we obtained estimates of the time to the most recent common ancestor (TMRCA) of all \( D. yakuba \) and \( D. santomea \) sequences, as well as those in the HZ clade. We assumed a neutral mutation rate of 0.06945 mutations per site and million year based on synonymous divergence between \( D. erecta \) and \( D. yakuba/D. santomea \) in the 13 mitochondrial genes \( (d_S = 1.4446; \text{see materials and methods}) \) and a time since split of 10.4 My (Tamura et al. 2004). The MRCA of the entire data set of \( D. yakuba \) and \( D. santomea \) mitochondrial sequences was estimated to have lived 14 153 years ago [upper and lower 95% highest posterior density (HPD): 17 888–10 792]. The estimate appears unusually recent and therefore contradicts divergence times between \( D. yakuba \) and \( D. santomea \) based on nuclear genes (Llopart et al. 2005b). One could argue, however, that the comparison of times based on nuclear and mitochondrial genomes is unfair as the latter do not recombine, are maternally inherited and are thus susceptible to shorter coalescent times due to smaller \( N_c \) (Hudson & Coyne 2002; Hudson & Turelli 2003; Ballard & Whitlock 2004). The \( Y \) chromosome, instead, provides the optimal comparison and, in addition, is more resilient to introgression in the \( D. yakuba \) and \( D. santomea \) system, for it has a large effect on hybrid male sterility (Coyne et al. 2004; Llopart et al. 2005b). Divergence times estimated for the \( Y \) chromosome using BEAST are significantly older than those obtained using mtDNA \[116 100 years with 95% HPD: 187 900–51 730; nuclear mutation rate assumed to be 5× lower than the mitochondrial mutation rate (Moriyama & Powell 1997; Haag-Liautard et al. 2007, 2008)] uncovering an old event of mtDNA introgression that drove the \( D. santomea \) mtDNA to extinction. The inclusion of the \( Y \) chromosome data set in our mitochondrial IMa2 simulations results in a more than 8-fold increase in the maximum-likelihood estimate of \( m \) between \( D. yakuba \) and \( D. santomea \) \((m = 0.19 \text{ for mtDNA alone and } m = 1.6 \text{ for mtDNA and } Y \text{ chromosome data sets combined, respectively, both for the } 0 \text{ variation.}

© 2014 The Authors. Molecular Ecology Published by John Wiley & Sons Ltd.
This is expected as the power to detect older introgression events is hampered by the analysis of a single, albeit lengthy, locus that has experienced introgression. The incorporation of Y chromosome data provides information on the speciation time of *D. yakuba* and *D. santomea*, allowing IMa2 to estimate \( m \) more accurately.

The TMRCA for the HZ clade was estimated to be 2539 years ago (95% HPD: 3624–1557). Similar results were obtained when population expansion was allowed [13 360 years for all sequences (HPD: 16 936–10 008) and 2905 years for the HZ clade (95% HPD: 4010–1893)]. Although precise dating of the different clades depends on the particular assumptions of the underlying molecular model, nonoverlapping HPDs clearly indicate that there have been at least two distinct waves of mitochondrial gene flow between *D. yakuba* and *D. santomea* (Fig. 5).

**Discussion**

Although *Drosophila yakuba* and *Drosophila santomea* began diverging ~400 Kya (Cariou *et al.* 2001; Llopart *et al.* 2002, 2005b), today they share the same mitochondrial genome. This observation stands out when we consider several facts. First, multiple nuclear loci investigated in these same species, including some on the Y chromosome, do show extensive genetic differentiation between *D. yakuba* and *D. santomea* (Llopart *et al.* 2005b; Bachtrog *et al.* 2006). Second, mutations rates of mtDNA in *Drosophila* are ~5–10 times higher than those of nuclear genes (Moriyama & Powell 1997; Haag-Liautard *et al.* 2007, 2008). Thus, the observed genetic similarity can only be explained by mitochondrial exchange through interspecific hybridization.

Our gene tree reconstruction and molecular dating indicate two independent mitochondrial introgression events (Fig. 5). Estimates of the TMRCA for all *D. yakuba* and *D. santomea* mitochondrial genomes are more than 8-fold smaller than estimates of speciation times based on the Y chromosome and indicate a common ancestor as recent as ~14 000 years ago. This is not only evidence of an early mitochondrial introgression event in *D. yakuba* and *D. santomea*, but it also suggests that only the mitochondrial genome of one of the two species survived this early introgression. Because the sequences of the mainland and island populations of *D. yakuba* do not form distinct clades, we infer that the direction of the gene flow was from *D. yakuba* to *D. santomea*. (The alternative possibility requires introgression from *D. santomea* into island populations of *D. yakuba* and then a species-wide selective sweep of the *D. santomea* haplotype among mainland populations of *D. yakuba*.) The basal sequences of *D. santomea* shown in the gene tree of Fig. 3 may represent *D. yakuba*-like remnants of this early introgression event. The replacement of the mtDNA of one species by that of another has been also reported in *D. pseudoobscura* and *D. persimilis* (Powell 1983; Machado *et al.* 2002) but contrasts with the *D. simulans–D. mauritiana* case, wherein the mtDNA of both species is still present today in *D. mauritiana* (Ballard 2000; Nunes *et al.* 2010; Garrigan *et al.* 2012). It is tempting to speculate that the early replacement of the *D. santomea* mitochondrial genome was
driven by positive selection. However, as the signature of positive selection is typically short-lived (Przeworski 2002), it is not likely to detect its fingerprints. Indeed, we find no evidence in our data to support the possibility that the original *D. santomea* mtDNA replacement was adaptive. We detected a second, ongoing, mitochondrial exchange between *D. yakuba* and *D. santomea* centred in the classic hybrid zone (Llopart et al. 2005a). Contrary to the early introgression event, in this case, we were able to detect the fingerprints of natural selection using coalescent simulations. This event represents a prime example of adaptive introgression or the acquisition of beneficial alleles from closely related species (Rieseberg 2011; Song et al. 2011).

Several explanations could potentially account for the exchange of mitochondrial genomes among species. Theoretical population genetic models show that very small amounts of gene flow between two species are sufficient to homogenize neutral mtDNA, even when hybrids experience substantial fitness loss caused by nuclear incompatibilities (Takahata & Slatkin 1984). Recurrent hybridization between *D. yakuba* females and *D. santomea* males and selection against hybrid nuclear genotypes, but not against mtDNA variants, could lead to some of the observed patterns. Selectively advantageous factors also need to be considered. mtDNA could hitchhike with the spread of the endosymbiont *Wolbachia*, as reported in *D. simulans* (Turelli & Hoffmann 1991; Turelli et al. 1992), because they are both maternally transmitted, and infected females experience a selective advantage relative to uninfected ones due to cytoplasmic incompatibility (Hoffmann et al. 1986; Nigro & Prout 1990). Although to date *D. yakuba* and close relatives lack cytoplasmic incompatibility associated with *Wolbachia* (Charlat et al. 2004; Zabalou et al. 2004), the endosymbiont may induce other potentially beneficial alleles from closely related species (Rieseberg 2011; Song et al. 2011).

Lastly, we propose a third selective explanation for mitochondrial introgression and replacement that is based on Muller’s ratchet (Muller 1964; Felsenstein 1974; Maynard Smith 1978; Lynch & Gabriel 1990; Charlesworth et al. 1993b). While Muller’s ratchet has often been proposed to explain the irreversible eventual extinction of asexual species (or populations) (Gabriel et al. 1993; Lynch et al. 1993; Gordo & Charlesworth 2000), it can also operate in nonrecombining genetic elements, such as mtDNA and the Y chromosome (Gabriel et al. 1993; Charlesworth & Charlesworth 2000; Filatov et al. 2000; Loewe 2006; Kaiser & Charlesworth 2010) as well as in nuclear regions with severely reduced recombination rates (Dolgin & Charlesworth 2008). The larger *N_e* of *D. yakuba* relative to that of *D. santomea* likely resulted in a decreased rate of accumulation of mildly deleterious mutations (i.e. slower Muller’s ratchet) in the mtDNA. Interspecific hybridization after the secondary colonization of São Tomé by *D. yakuba* may have allowed *D. santomea* to tap into additional genetic variation. The lower mutational load of the *D. yakuba* mitochondrial genome may have driven the early replacement of the *D. santomea* mtDNA. Rieseberg (2009) proposed that introgression could serve as an efficient way to replace damaged or lost alleles using those of a closely related species. The early mtDNA replacement in *D. yakuba* and *D. santomea*, and perhaps even the ongoing introgression event, could well constitute examples of this scenario.

Acknowledgements

We thank all the members of the ‘Isolation with Migration Discussion Group’ (https://groups.google.com/forum/?from-groups#forum/Isolation-with-Migration) for tireless and helpful advice. We are also grateful to Josep Comeron and four anonymous reviewers for comments on the manuscript. This work was funded by University of Iowa funds to A.L. D.H. was partially supported by NIH Predoctoral Training Grant in Genetics T32 GM008629014.

References


not induce cytoplasmic incompatibility but fully rescue the wRi modification. Genetics, 167, 827–834.

A.L. conceived and supervised the project and also performed the analyses. A.L., D.H., E.B. and Z.S. carried out laboratory experiments and obtained DNA sequences. A.L. wrote the manuscript with comments from all the other authors.

Data accessibility

Supporting information
Additional supporting information may be found in the online version of this article.

Table S1 Detailed description of the lines used in this study.