

Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression

Emily A. Beck,¹ Aaron C. Thompson,² Joel Sharbrough,² Evgeny Brud,² and Ana Llopart^{1,2,3}

¹Interdisciplinary Graduate Program in Genetics, The University of Iowa, Iowa City, Iowa 52242

²The Department of Biology, The University of Iowa, Iowa City, IA 52242

³E-mail: ana-llopart@uiowa.edu

Received January 12, 2014

Accepted June 16, 2015

Introgression is the effective exchange of genetic information between species through natural hybridization. Previous genetic analyses of the *Drosophila yakuba*–*D. santomea* hybrid zone showed that the mitochondrial genome of *D. yakuba* had introgressed into *D. santomea* and completely replaced its native form. Since mitochondrial proteins work intimately with nuclear-encoded proteins in the oxidative phosphorylation (OXPHOS) pathway, we hypothesized that some nuclear genes in OXPHOS cointrogressed along with the mitochondrial genome. We analyzed nucleotide variation in the 12 nuclear genes that form cytochrome c oxidase (COX) in 33 *Drosophila* lines. COX is an OXPHOS enzyme composed of both nuclear- and mitochondrial-encoded proteins and shows evidence of cytonuclear coadaptation in some species. Using maximum-likelihood methods, we detected significant gene flow from *D. yakuba* to *D. santomea* for the entire COX complex. Interestingly, the signal of introgression is concentrated in the three nuclear genes composing subunit V, which shows population migration rates significantly greater than the background level of introgression in these species. The detection of introgression in three proteins that work together, interact directly with the mitochondrial-encoded core, and are critical for early COX assembly suggests this could be a case of cytonuclear cointrogression.

KEY WORDS: Coevolution, coadaptation, cytochrome c oxidase, gene flow, introgression.

Speciation is the process by which one interbreeding population is split into two descendant populations that, in time, acquire reproductive isolation (RI) (Mayr 1963). When RI is not complete, sympatric species can still effectively exchange genetic information through natural hybridization, also called introgression or introgressive hybridization (Anderson and Hubricht 1938; Riley 1939). Initially, it was believed that hybridization between different animal species was a rare accident. We now know, however, that hybridization is much more common in nature than previously suspected and introgression has been studied in many animal systems, including primates (Ferris et al. 1983; Hey 2003; Becquet and Przeworski 2007; Hey 2010a), mice (Gerald et al. 2008; Teeter et al. 2008; Staubach et al. 2012; Janousek et al. 2015;

Liu et al. 2015), birds (Ellegren et al. 2012; Alcaide et al. 2014; Poelstra et al. 2014), butterflies (Bull et al. 2006; Dasmahapatra et al. 2007; Mallet et al. 2007; Putnam et al. 2007; Giraldo et al. 2008; Pardo-Diaz et al. 2012; The *Heliconius* Genome 2012; Martin et al. 2013; Nadeau et al. 2014; Kronforst and Papa 2015), mosquitoes (Besansky et al. 1994; Marsden et al. 2011; Weetman et al. 2012; Clarkson et al. 2014; Neafsey et al. 2015; Norris et al. 2015), and *Drosophila* (Machado et al. 2002; Hey and Nielsen 2004; Llopart et al. 2005b; Bachtrog et al. 2006; Kurlath et al. 2009; Nunes et al. 2010; Garrigan et al. 2012; Herrig et al. 2014).

The study of introgression is important for at least three reasons. First, understanding introgression and the mechanisms



behind it is paramount in determining the evolutionary independence of taxa. Second, introgression can be a source of species' adaptations (Arnold 1997; Arnold 2007). Through adaptive introgression, species can potentially benefit from genetic variation that arose in another species (Castric et al. 2008; Kim et al. 2008; Rieseberg 2011; Song et al. 2011; Pardo-Díaz et al. 2012; Staubach et al. 2012). Additionally, it has been suggested that the acquisition of new traits may be much more common via introgression than by *de novo* mutation, particularly with respect to traits that depend on multiple substitutions (Rieseberg 2009). Finally, gene products that function well in their own species' genetic background can generate incompatibilities when interacting with products from another species and result in hybrid sterility and inviability (i.e., DM incompatibilities) (Bateson 1909; Dobzhansky 1937; Dobzhansky and Queal 1938; Muller 1942; Coyne and Orr 2004). The study of introgression can provide insights into the genomic location of DM incompatibilities, local adaptation, or alternatively, uncover candidates for cointrogression.

In regions of sympatry where species meet and hybridize, the potential for introgression exists. These hybrid zones are "natural laboratories" where new genic combinations are generated and tested by selection (Hewitt 1988). Unfortunately, occurrences of these hybrid zones in genetically amenable model systems, like *Drosophila*, are not very common. A hybrid zone of two *Drosophila* species (*D. yakuba* and *D. santomea*) was discovered by Daniel Lachaise on the volcanic island of São Tomé and later characterized (Lachaise et al. 2000; Llopart et al. 2005b, a). This hybrid zone is located on the slopes of Pico de São Tomé where *D. yakuba* is found at low elevations, below 1450 meters, while *D. santomea* stays in shady regions covered by the rainforest, between 1150 and 2024 m (Lachaise et al. 2000; Llopart et al. 2005a). At middle elevations, the two species meet and hybridize creating a hybrid zone. Hybrids have been reported at a frequency of approximately 1–3%, as estimated by the presence of pigmentation patterns intermediate between the two parent species (Llopart et al. 2002) and confirmed by SNP genotyping (Llopart et al. 2005a). This frequency is ~100 times higher than the reported frequency of hybridization between *D. pseudoobscura* and *D. persimilis*, another common pair used to study speciation (Dobzhansky 1973). *D. yakuba* and *D. santomea* began diverging genetically approximately ~400,000 years ago (Cariou et al. 2001; Llopart et al. 2002; Llopart et al. 2005b) but are still able to produce fertile hybrid females (Coyne et al. 2004), which makes them capable of backcrossing in nature (F_1 hybrid males are completely sterile), and are therefore an ideal system for studying introgression.

Previous work based on the analysis of 47 complete mitochondrial genomes indicates that mtDNA introgression is recurrent in this species system (Llopart et al. 2014). Studies have

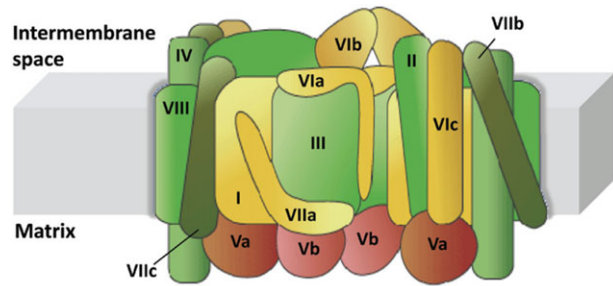


Figure 1. Cytochrome *c* oxidase (COX) subunits of oxidative phosphorylation (bovine). Mitochondrially encoded subunits (I–III) form the core of COX. Nuclear subunits Va and Vb showed evidence of cointrogression in *D. yakuba* and *D. santomea*. Adapted with permission from <http://www.genome.jp/kegg/pathway.html>, map00190 (Kanehisa and Goto 2000).

shown that the *D. santomea* mitochondrial genome has been completely replaced by that of *D. yakuba* (Llopart et al. 2005b; Bachtrog et al. 2006; Llopart et al. 2014). This replacement is of particular interest because it raises the possibility that nuclear-encoded proteins interacting with mitochondrially encoded proteins may have cointrogressed. The proteins encoded in the mitochondrial genome are heavily involved in oxidative phosphorylation (OXPHOS), a metabolic pathway that is responsible for the production of most of the energy in an animal's cells and is remarkably important to cell survival (Smeitink et al. 2001). Five protein complexes, which are all localized in the lipid bilayer of the inner mitochondrial membrane, play key roles during OXPHOS. While Complex II is composed solely of proteins encoded by nuclear DNA (nDNA), the four remaining complexes are composed of interacting proteins that are encoded in both nDNA and mtDNA (Tripoli et al. 2005). Complex IV of OXPHOS, often referred to as cytochrome *c* oxidase or COX, is of special interest because its core is composed entirely of mitochondrial-encoded subunits (I, II, and III) and is surrounded by multiple nuclear-encoded proteins with which it must interact (Saraste 1999) (Fig. 1). These functional interactions in COX have been proposed to result in cytonuclear coadaptation in some species (Blier et al. 2001; Rand et al. 2004; Osada and Akashi 2012).

Although the mitochondrial genome represents only a small fraction of the genetic material of an organism, the multiple interactions between mitochondrial and nuclear partners provide unique opportunities to study DM incompatibilities and their effects in hybrids (Rand et al. 2004; Burton and Barreto 2012). For example, the F_2 hybrid males of the parasitoid wasps *Nasonia giraulti* and *N. vitripennis* experience increased mortality and this hybrid breakdown is also associated with disruption of mitochondrial function and a reduction in ATP production (Breeuwer and Werren 1995; Ellison et al. 2008; Niehuis et al. 2008; Gibson et al. 2013). Interpopulation hybrids of the

copepod *Tigriopus californicus* show variable degrees of reduced fitness that map to mitochondrial-nuclear incompatibilities (Edmands and Burton 1999; Burton et al. 2006; Ellison and Burton 2006, 2008; Burton and Barreto 2012). A recent study of the genomic trajectory of hybrid swarms in this same species shows a decrease in cytonuclear mismatch over time, which is consistent with coordinated changes of mitochondrial and nuclear interactors (Pritchard and Edmands 2013). In *Drosophila*, the search for DM incompatibilities between species involving the mitochondrial genome has produced mixed results, possibly reflecting the examination of different species/strains (Sackton et al. 2003; Montooth et al. 2010; Hoekstra et al. 2013; Meiklejohn et al. 2013).

Here, we evaluate levels of introgression between *D. yakuba* and *D. santomea* in the COX nuclear genes using multilocus approaches based on maximum-likelihood population genetics methods. Our results revealed significant gene flow from *D. yakuba* to *D. santomea* for the entire COX complex and more specifically for COX subunit V. Most important, the population migration rate estimated for this subunit is significantly greater than the background migration rate obtained from 25 genes unrelated to OXPHOS and randomly distributed across the genomes of the two species (Llopart et al. 2005b). This supports the notion that introgression in subunit V cannot be fully explained by baseline levels of gene flow between *D. yakuba* and *D. santomea*. The signature of gene flow in subunit V is not limited to a single gene. We detected the same distinct pattern typical of introgression in the three nuclear genes encoding subunit V (*CoVa*, *CoVb*, and *CG11043*): no fixed difference between species and abundant shared variation, a pattern not seen in any of the other COX loci analyzed. Because these three genes are part of the same main subunit, which interacts with the mitochondrial-encoded core early on in COX assembly, we suggest that this pattern is the result of coordinated events of nuclear and mitochondrial cointrogression.

Materials and Methods

FLY LINES STUDIED

We established a total of 33 isofemale lines of *Drosophila* in the laboratory, including 16 *D. yakuba*, 16 *D. santomea*, and one *D. teissieri*. Allopatric samples of *D. yakuba* included two lines from Tanzania and one from Zimbabwe. The *D. yakuba* sample also included lines collected from regions of sympatry or parapatry with *D. santomea*, eight from the São Nicolau waterfall (São Tomé) and five from or near the hybrid zone of Pico de São Tomé. All *D. santomea* samples were collected from regions of sympatry with *D. yakuba* and included four from Rio Quija in southwest São Tomé, five from Bom Sucesso, the field station on Pico de São Tomé, and seven from or near the hybrid zone. The *D. teissieri* line was established from a female collected in

Brazzaville, Republic of Congo. All flies were maintained in the laboratory at 24° C on standard cornmeal-yeast agar medium and kept in 12 hour light/dark cycles.

LOCI ANALYZED

We analyzed patterns of polymorphism and divergence in the 12 genes encoding COX nuclear subunits (14 regions total because we included the 5' and 3' noncoding flanking regions of the *CoVa* gene; Table 1). We chose COX due to the intimate nature of protein-protein interactions between the nuclear-encoded subunits and the three mitochondrial-encoded subunits composing the complex. We also focused on this complex because it has the highest ratio of mitochondrial to nuclear subunits out of all the OXPHOS complexes, and it is the rate-limiting step of OXPHOS. While ten subunits are found in COX in mammals, so far only eight have been identified in *Drosophila*, which lacks VIIb and VIII nuclear-encoded subunits (Tripoli et al. 2005) (<http://www.mitocomp.uniba.it/>). We designed primers for each of these regions using the *D. yakuba* reference genome sequence obtained from the 12 *Drosophila* Genome Project (Clark et al. 2007) and Primer3Plus (Untergasser et al. 2007). Thirteen of the regions under analysis are autosomal and one is X-linked (Table 1).

According to the *D. yakuba* genome project sequence, *CG10396* is actually duplicated (denoted *GE15061* and *GE14691* in *D. yakuba*) and we were able to detect the duplication in every single *D. yakuba* line we studied. Whether these two copies remain functional is currently unknown. In contrast, we found no evidence of the *CG10396* duplication in any of the *D. santomea* lines investigated. Due to the lack of divergence information with *D. santomea*, the duplicated copy of *CG10396* (i.e., *GE14691*) was excluded from all our analyses. In addition, one line of *D. santomea* shows a premature stop codon in *CG30093*, resulting in a protein two amino acids short.

DNA EXTRACTION, SEQUENCING, AND DESCRIPTION OF INTER- AND INTRASPECIFIC VARIATION

We isolated DNA from single male flies from each line using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) and amplified the regions of interest using PCR. PCR products were then purified using the Wizard Magnesium PCR clean-up system (Promega, Madison, WI). Sequencing reactions were performed using the Big Dye 3.1 chemistry (Applied Biosystems, Foster City, CA), cleaned-up using a 6.25% Sephadex column, and run in an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). To maximize accuracy, both DNA strands were sequenced and assembled into single contigs using Sequencher 5.0 (Gene Codes, Ann Arbor, MI). A pool of haplotypes was then reconstructed for each locus containing heterozygous sites using the PHASE 2.0 program excluding indels (Stephens et al. 2001). For each

Table 1. Gene names, chromosomal locations, and subunit associations of nuclear genes of OXPHOS COX.

Gene name in <i>D. melanogaster</i>	Gene name in <i>D. yakuba</i>	Location in <i>D. melanogaster</i>	Cytogenetic map ¹	Location in <i>D. yakuba</i>	Subunit ²
<i>CG10396</i> ³	<i>GE15061</i>	2R	41E5	<i>Un</i>	IV
	<i>GE14691</i>	2R	—	<i>Un</i>	—
<i>CoIV</i>	<i>GE13288</i>	2L	38A8	2R	IV
<i>CoVa</i>	<i>GE24204</i>	3R	86F9	3R	Va
<i>CoVb</i>	<i>GE13966</i>	2L	26E3	2L	Vb
<i>CG11043</i>	<i>GE13977</i>	2L	26E3	2L	Vb
<i>CG30093</i>	<i>GE11738</i>	2R	52D3-D4	2R	VIa
<i>levy</i>	<i>GE14301</i>	2R	59E3	2R	VIa
<i>CoVIb</i>	<i>GE17355</i>	X	18E5	X	VIb
<i>Cype</i>	<i>GE25385</i>	2L	25D6	2L	Vlc
<i>CG9603</i>	<i>GE24854</i>	3R	84F13	3R	VIIa
<i>CG18193</i>	<i>GE25872</i>	3R	84F13	3R	VIIa
<i>CoVIIc</i>	<i>GE19330</i>	2R	46D8-D9	3L	VIIc

¹cytogenetic map in *D. melanogaster*.

²Subunit association of OXPHOS COX (<http://www.mitocomp.uniba.it/>).

³Gene *CG10396* is duplicated in the *D. yakuba* genome project sequence (Clark et al. 2007).

line a single haplotype was selected for further analysis from the pool based on highest confidence as calculated by PHASE 2.0 (Stephens et al. 2001).

To calculate the number of synonymous (d_S) and nonsynonymous (d_N) changes per site and d_N/d_S , we used the codeml program within the PAML v4.5 package (Yang 2007). We calculated the average nucleotide frequencies from third codon positions ($F_3 \times 4$), and we assumed a single d_N/d_S for all the lineages (Model 0) and among sites (NSites = 0). Gene trees were generated using MEGA 5 (Tamura et al. 2011), with bootstrap values (Felsenstein 1996) based on 10,000 replicates. Analyses of polymorphism were performed using the program DnaSP 5.1 (Librado and Rozas 2009).

DETECTION OF INTROGRESSION

To detect introgression between *D. yakuba* and *D. santomea*, we used the multilocus method based on the Isolation-with-Migration (IM) model (Nielsen and Wakeley 2001; Hey and Nielsen 2004) implemented in the program IMA2 (Hey and Nielsen 2004, 2007). This program generates simulations of gene genealogies using a Markov-Chain Monte-Carlo (MCMC) approach to obtain posterior distributions of population demographic parameters (Nielsen and Wakeley 2001; Hey and Nielsen 2004). The IM model assumes that one panmictic ancestral population divides into two descendant populations that may (or may not) have exchanged genes after their split. To determine whether a population model with gene flow fits better to the data than an alternative model with 0 migration, IMA2 performs likelihood ratio tests based on the maximal posterior probability for a migration parameter and the posterior probability when that parameter is set at zero (Nielsen and Wakeley 2001).

The IM model assumes that all detected variation is neutral (i.e., not affected by directional or balancing selection), no intralocus recombination and free recombination between loci (Hey and Nielsen 2004). To test the assumption of neutrality we applied Tajima's D (Tajima 1989), McDonald-Kreitman (MK) (McDonald and Kreitman 1991), and Hudson-Kreitman-Aguade (HKA) (Hudson et al. 1987) tests. Tajima's D statistic rejected neutrality in only 2 out of 28 tests performed (14 regions analyzed in two species) (Table S1). None remained significant after multiple-test correction. There was an overall trend toward negative values of Tajima's D , which indicated a deficit of polymorphisms at intermediate frequencies (Tajima 1989). These results potentially indicated a recent population expansion (Tajima 1989) and/or purifying selection (Fu 1997; Fay and Wu 1999; Gordo et al. 2002), in agreement with previous findings in the same pair of species (Llopart et al. 2005b; Bachtrog et al. 2006). MK tests detected no significant departure from neutral expectations (McDonald and Kreitman 1991) (Table S2). The multilocus HKA test (<https://bio.cst.temple.edu/~hey/software/software.htm#HKA>) showed consistency in the polymorphism to divergence ratio across loci ($P = 0.16$; Table S3). From these findings combined, we conclude that overall there is no evidence of positive or balancing selection in our dataset. To address the issue of no intralocus recombination, IMA2 was run additionally using the largest sequence block of each region showing no evidence of intralocus recombination assessed using the four gamete type test in DNASP 5.1 (Librado and Rozas 2009).

IMA2 was run for all 14 COX regions, as well as groups of 3–5 loci corresponding to each main nuclear subunit of COX (IV, V, VI, and VII). We also used IMA2 to calculate migration

rates for a subset of 25 OXPPOS-unrelated genes randomly distributed throughout the *D. yakuba*/*D. santomea* genomes (Llopart et al. 2005b). These genes provide a background signal of introgression in these species. For all datasets the burn-in period was at least $1 \times 10^{+6}$ steps under the HKY model (Hasegawa et al. 1985) with 150 Markov chains ($a = .99$ $b = .75$). Prior values were set following the recommendations in the IMA2 documentation (Table S4). Mixing was assessed by monitoring the Effective Sample Sizes (ESS) and parameter trends over a run. Large ESS and lack of long-term parameter trends (zero autocorrelations) indicate good mixing. To ensure convergence, two independent sets of simulations were obtained, each run for at least $1 \times 10^{+5}$ genealogies. Posterior probabilities of the different parameters were calculated using both runs combined. To convert model parameters to more familiar demographic quantities (e.g., the population migration rate, $2N_e m$, where N_e is the effective population size and m is the migration rate per locus and per generation) we followed Won and Hey (2005). Model parameters and converted units are shown in Table S5.

To obtain posterior probability distributions of migration rates additional to those provided by IMA2, we applied the program MIMAR, which can incorporate intralocus recombination (Becquet and Przeworski 2007). The program GenCo was used to estimate intralocus recombination for each gene (Gay et al. 2007) (Table S6). For each dataset we performed at least three MIMAR runs with different seeds, each carried out independently utilizing a burn-in period of $1 \times 10^{+5}$ steps and at least $1 \times 10^{+6}$ steps after burn-in. The ranges of the prior distributions for MIMAR were U [0.002, 0.05] for the population mutation rates (in per base pair units), U [$2.5 \times 10^{+6}$, $4.5 \times 10^{+6}$] for splitting times (in number of generations), and logarithm from U [4.54×10^{-5} , 2.72] for population migration rates. Contrary to IMA2, MIMAR requires information from a third species used as outgroup. We used *D. erecta* as outgroup instead of *D. teissieri* for two reasons. First, it has been proposed that *D. teissieri* has exchanged genes with *D. yakuba*–*D. santomea*, at least for the mitochondrial genome (Llopart et al. 2005b; Bachtrog et al. 2006). Gene flow between *D. teissieri* and *D. yakuba* can severely affect the inference of the ancestral/derived allele upon which MIMAR is based. Second, the whole genome of *D. erecta* is available as part of the 12 *Drosophila* genome project and the species constitutes a valid outgroup for *D. yakuba* and *D. santomea* (Clark et al. 2007). For the 14 COX loci and the 25 random genes, MIMAR runs produced acceptance rates that were extremely low (7.78×10^{-6} – 3.27×10^{-5}) and thus considered unacceptable based on MIMAR documentation. MIMAR runs with individual main subunits showed good acceptance rates for subunits IV (0.19–0.21), VI (0.12–0.13), and VII (0.07–0.14), and acceptance rates that were marginally acceptable for subunit V (0.014–0.017).

Table 2. Average divergence between *D. yakuba* and *D. santomea* at nonsynonymous (d_N) and synonymous (d_S) sites of OXPPOS COX nuclear genes.

Locus ¹	d_N	d_S	d_N/d_S
<i>CG10396</i>	0.003	0.02	0.15
<i>CoIV</i>	0	0.03	0
<i>CoVa</i>	0	0.06	0
<i>CoVb</i>	0	0.03	0
<i>CG11043</i>	0.003	0.06	0.05
<i>CG30093</i>	0.01	0.20	0.05
<i>levy</i>	0	0.18	0
<i>CoVIb</i>	0	0.010	0
<i>Cypc</i>	0	0.03	0
<i>CG9603</i>	0.005	0.09	0.06
<i>CG18193</i>	0.004	0.06	0.07
<i>CoVIIc</i>	0	0.05	0
All	0.002	0.07	0.03

¹Gene symbols from *D. melanogaster*.

Results and Discussion

INTER- AND INTRASPECIFIC VARIATION

To obtain an overall view of nucleotide variation associated with the 12 nuclear genes of COX in *D. yakuba* and *D. santomea*, we investigated between-species divergence and within-species polymorphism levels. Table 2 shows the number of nonsynonymous (d_N) and synonymous (d_S) changes per site between *D. yakuba* and *D. santomea* for each gene. The average d_N of all 12 genes was consistent with previous estimates for this same pair of species (Llopart et al. 2005b) suggesting that the nuclear genes of COX are not particularly conserved when considered as a class. Levels of nucleotide polymorphism were also obtained. The average nucleotide heterozygosity at synonymous sites is 1.65% for *D. santomea* and 2.32% for *D. yakuba* (Table 3), indicating only a modest reduction in polymorphism in *D. santomea* relative to *D. yakuba* consistent with previous findings (Llopart et al. 2005b; Bachtrog et al. 2006). Because *D. santomea* is an island endemic and *D. yakuba* can be found throughout sub-Saharan Africa, the two species are predicted to have very different census population sizes based on their current geographical distribution, and population genetics theory would predict a more substantial difference in intraspecific variation. This same trend of similar levels of polymorphism, however, has been seen previously in other closely related species with apparently large differences in census population size, and has been attributed, at least partially, to ancestral polymorphism and very recent species split (Kliman et al. 2000; Ramos-Onsins et al. 2004; Llopart et al. 2005b; Leffler et al. 2012).

Table 3. Polymorphism data for *D. yakuba* and *D. santomea* in the nuclear genes of OXPPOS COX.

Locus ¹	Species	<i>n</i> ²	π ³	θ_w ⁴	θ_{syn} ⁵	θ_{sil} ⁶	θ_N ⁷	Syn ⁸	Rep ⁹
<i>CG10396</i>	<i>D. santomea</i>	16	0.08	0.19	0	0.40	0.074	0	1
	<i>D. yakuba</i>	14	0	0	0	0	0	0	0
<i>CoIV</i>	<i>D. santomea</i>	15	0.80	0.99	1.17	1.67	0	5	0
	<i>D. yakuba</i>	16	0.41	0.49	2.07	0.84	0	9	0
<i>CoVa</i>	<i>D. santomea</i>	14	0.54	0.82	2.55	1.60	0	9	0
	<i>D. yakuba</i>	15	0.53	0.87	2.50	1.43	0.18	9	2
<i>CoVa3'</i>	<i>D. santomea</i>	15	1.16	1.15	-	1.20	-	-	-
	<i>D. yakuba</i>	15	0.59	1.12	-	1.12	-	-	-
<i>CoVa5'</i>	<i>D. santomea</i>	15	1.03	1.34	-	1.34	-	-	-
	<i>D. yakuba</i>	15	1.25	1.84	-	1.84	-	-	-
<i>CoVb</i>	<i>D. santomea</i>	15	1.71	2.11	2.72	2.64	0.11	7	1
	<i>D. yakuba</i>	14	1.78	1.90	0.80	2.5	0	2	0
<i>CG11043</i> ¹⁰	<i>D. santomea</i>	14	1.43	1.55	4.14	1.90	0.55	16	7
	<i>D. yakuba</i>	14	1.22	1.53	4.14	1.81	0.27	17	5
<i>CG30093</i>	<i>D. santomea</i>	13	1.27	1.47	2.81	1.88	0.91	6	6
	<i>D. yakuba</i>	15	2.60	2.68	9.07	4.12	0.57	20	4
<i>levy</i>	<i>D. santomea</i>	16	0.42	0.69	1.90	1.06	0	5	0
	<i>D. yakuba</i>	16	0.92	1.08	4.19	1.70	0	11	0
<i>CoVIb</i>	<i>D. santomea</i>	16	0.11	0.20	1.21	0.26	0	2	0
	<i>D. yakuba</i>	16	0.42	0.64	0.61	0.87	0	1	0
<i>Cype</i>	<i>D. santomea</i>	16	0.75	0.86	0.99	1.11	0.18	2	1
	<i>D. yakuba</i>	14	1.23	1.34	2.06	1.78	0	4	0
<i>CG9603</i>	<i>D. santomea</i>	16	0.76	0.98	1.85	1.47	0.15	4	1
	<i>D. yakuba</i>	16	0.2	0.34	0	0.47	0	0	0
<i>CG18193</i>	<i>D. santomea</i>	14	1.35	1.31	0.43	1.77	0.67	1	5
	<i>D. yakuba</i>	16	0.13	0.23	0.41	0.29	0.26	1	1
<i>CoVIIc</i>	<i>D. santomea</i>	16	0.33	0.53	0	0.67	0	0	0
	<i>D. yakuba</i>	14	0.28	0.33	1.93	0.42	0	3	0

¹Gene symbols from *D. melanogaster*.²Number of sequences analyzed.³Nucleotide diversity (%).⁴Watterson's (1975) estimate of heterozygosity (%).⁵Heterozygosity (%) at synonymous sites.⁶Heterozygosity (%) at silent (synonymous and noncoding) sites.⁷Heterozygosity (%) at nonsynonymous sites.⁸Number of synonymous polymorphisms.⁹Number of amino acid replacement polymorphisms.¹⁰A single codon with multiple changes at a site was excluded.

PATTERNS OF VARIATION IN GENE TREE TOPOLOGY

Introgression violates the simple bifurcating model of species divergence upon which phylogenetic analyses are based, and as such, is expected to leave a phylogenetic signal. We obtained gene trees for each locus individually using Neighbor-Joining and maximum-likelihood methods (Figs. S1 and S2). As expected for closely related species, we detect a variety of gene tree topologies, ranging from two reciprocally monophyletic clades with good bootstrap support (e.g., gene *levy*) to single groups with sequences from both species clustering together (e.g., *CoVa*) (Fig. 2). The observed heterogeneity in gene trees across loci is consistent with either introgression between species, incomplete

lineage sorting, or high sequence conservation across species due to strong functional constraints. Additional analyses are thus required to determine whether there is evidence of introgression between *D. yakuba* and *D. santomea* in the COX loci.

EVIDENCE OF GENE FLOW FROM *D. YAKUBA* TO *D. SANTOMEA* IN COX NUCLEAR GENES

To investigate whether there has been genetic exchange between *D. yakuba* and *D. santomea* in COX, we first determined the number of shared polymorphisms and fixed differences in each nuclear gene under study (Table 4). There was substantial heterogeneity in the distribution of shared polymorphisms and fixed

Table 4. Shared and fixed variation in nuclear genes of OXPPOS COX.

Locus ¹	L(bp) ²	S _{x1} ³	S _{x2} ⁴	F ⁵	S _s ⁶	Main subunit
<i>CG10396</i>	942	4	0	7	0(0)	IV
<i>CoIV</i>	1138	31	16	6	0(0.44)	IV
<i>CoVa</i>	924	12	12	0	8(0.16)	V
<i>CoVa3'</i>	858	18	18	0	8(0.38)	V
<i>CoVa5'</i>	850	30	42	10	1(1.48)	V
<i>CoVb</i>	1487	64	61	0	22(2.63)	V
<i>CG11043</i>	1805	53	53	0	25(1.56)	V
<i>CG30093</i>	883	17	48	6	7(0.92)	VI
<i>levy</i>	816	11	22	5	5(0.30)	VI
<i>CoVlb</i>	789	4	14	2	1(0.07)	VI
<i>Cype</i>	932	18	28	2	4(0.54)	VI
<i>CG9603</i>	785	23	6	9	1(0.18)	VII
<i>CG18193</i>	850	26	4	8	1(0.12)	VII
<i>CoVIIc</i>	892	12	7	9	0(0.09)	VII

¹Locus symbols from *D. melanogaster*.

²Length in base pairs.

³Polymorphisms exclusive to *D. santomea*.

⁴Polymorphisms exclusive to *D. yakuba*.

⁵Differences fixed between species.

⁶Number of shared polymorphisms (Expected number of shared polymorphisms from parallel mutation).

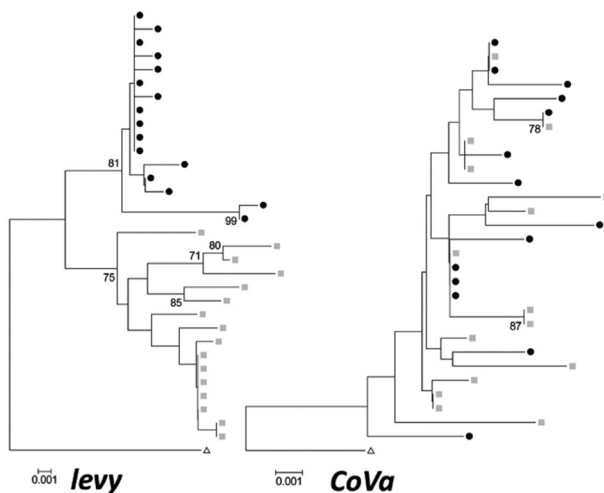


Figure 2. Gene trees reconstructed with the Neighbor-Joining algorithm and bootstrap values based on 10,000 replicates for *levy* and *CoVa*. Bootstrap values higher than 70% are shown. *D. santomea*, *D. yakuba*, and *D. teisseri* are shown with filled circles, shaded squares, and open triangles, respectively.

differences across loci, which suggests the possibility of introgression at some but not all of the genes investigated. Shared variation can occur from (1) introgression (i.e., gene flow), (2) parallel mutations, or (3) retention of ancestral polymorphism due to incomplete lineage sorting in closely related species (Clark 1997). To rule out parallel mutation, we calculated the expected number of shared polymorphisms produced by independent convergent mutations in each of the two species (Kliman et al. 2000).

In all cases, the number of observed shared polymorphisms was substantially greater than that expected from parallel mutation (Table 4). The other evolutionary source of shared variation in very closely related species is ancestral polymorphism, which can persist due to incomplete lineage sorting. However, the genealogies under introgression and incomplete lineage sorting are expected to have different branch lengths and topologies. Population genetic models, like the IM model (Nielsen and Wakeley 2001; Hey and Nielsen 2004) implemented in the software IMA2, can differentiate between introgression and incomplete lineage sorting using likelihood methods (Hey and Nielsen 2004, 2007).

To determine whether the patterns of intra- and interspecific variation in the nuclear genes of COX could be fully explained in absence of gene flow between *D. yakuba* and *D. santomea*, we applied the likelihood ratio (LLR) tests (Nielsen and Wakeley 2001; Hey 2010b) implemented in IMA2 (Hey and Nielsen 2004, 2007). Simulations yielded contained posterior probabilities for all demographic parameters estimated, suggesting that our dataset contained sufficient information (Fig. 3). The LLR tests indicated that a population model in which there was gene flow from *D. yakuba* to *D. santomea* fits significantly better to our data (LLR = 7.07, $P < 0.01$ and LLR = 7.22, $P < 0.01$ for seeds 1 and 2, respectively) than a model with 0 migration. In contrast, there was no evidence of gene flow in the opposite direction (LLR = 1.03, $P > 0.05$ and LLR = 1.11, $P > 0.05$ for seeds 1 and 2, respectively). Qualitatively similar results were obtained but with significance in both directions when we used the largest block of each region with no evidence of intralocus recombination

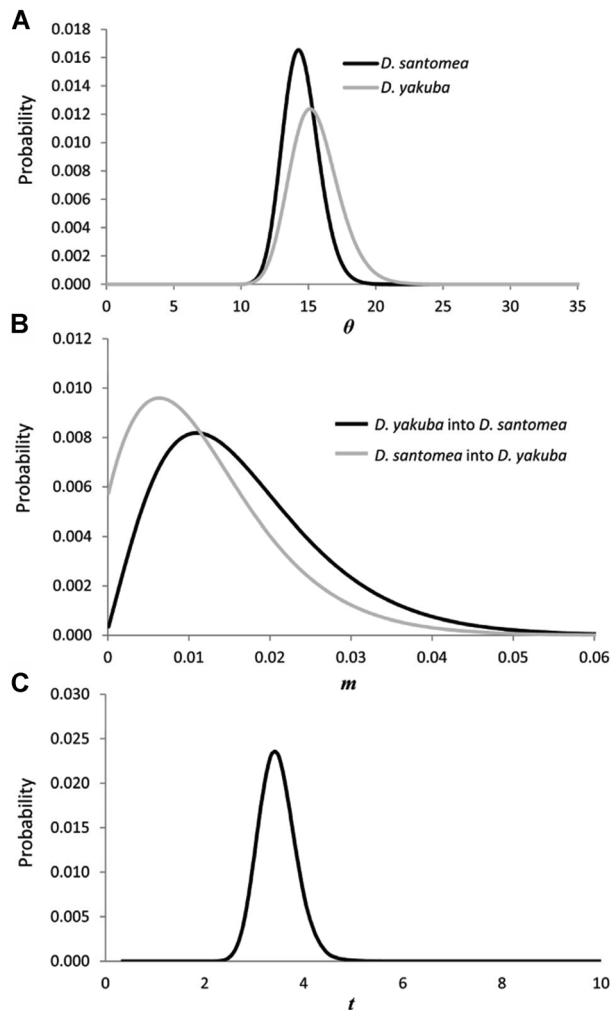


Figure 3. The marginal posterior probability distributions for speciation parameters scaled by the neutral mutation rate and obtained by fitting the IM model to a 14-region dataset using IMA2. (A) Population mutation rates, $\theta = 4N_e u$, where N_e is the effective population size and u is the neutral mutation rate per locus and per generation. (B) Rates of migration for each gene copy per mutation event, $m = m/u$, where m is the migration rate per locus and per generation. (C) Time since population split in terms of mutations, $t = t u_y$, where t is time since population splitting in years and u_y is the neutral mutation rate per locus and per year.

($\text{LLR}_{\text{seed1}} = 11.87$, $P < 0.001$ and $\text{LLR}_{\text{seed2}} = 12.71$, $P < 0.001$ from *D. yakuba* to *D. santomea*, and $\text{LLR}_{\text{seed1}} = 7.28$, $P < 0.01$ and $\text{LLR}_{\text{seed2}} = 7.25$, $P < 0.01$ from *D. santomea* to *D. yakuba*). We conclude that the nuclear genes encoding COX show the molecular signature of introgression from *D. yakuba* into *D. santomea*.

GENE FLOW IN MAIN SUBUNIT V

To assess whether introgression occurs in clusters or is scattered throughout the different genes of COX, we performed independent IMA2 runs for each main nuclear-encoded subunit (IV, V, VI,

and VII; Table 4). The number of regions associated with each main subunit ranges from two (subunit IV) to five (subunit V, 3 genes and two noncoding regions). A recent study evaluating the accuracy of IMA2 has shown that, even when analyzing only two loci, the false-positive rate is smaller than 0.01 for species with divergence time greater than $2N_e$ generations, where N_e is the effective population size (Cruickshank and Hahn 2014). Based on the analysis of 25 loci unrelated to OXPHOS and randomly distributed across the *D. yakuba*/*D. santomea* genomes (Llopart et al. 2005b), the two species split from their common ancestor $\sim 4N_e$ generations ago (Table S5). This corresponds to ~ 0.43 Mya, in agreement with estimates from other studies (Cariou et al. 2001; Llopart et al. 2002; Llopart et al. 2005b; Bachtrog et al. 2006). Subunit V, consisting of Va and Vb, showed a significant rejection of an isolation model (i.e., gene flow = 0) in favor of an alternative model with gene flow from *D. yakuba* to *D. santomea* ($\text{LLR}_{\text{seed1}} = 5.11$, $0.01 < P < 0.05$ and $\text{LLR}_{\text{seed2}} = 4.89$, $0.01 < P < 0.05$) but not in the opposite direction ($\text{LLR}_{\text{seed1}} = 0$, $P > 0.05$ and $\text{LLR}_{\text{seed2}} = 0.0$, $P > 0.05$). In contrast, nucleotide variation in main subunits IV, VI, and VII is compatible with a model with no gene flow between species (all $\text{LLR} = 0$, $P > 0.05$). Equivalent results were obtained when the largest block with no evidence of intralocus recombination was used ($\text{LLR} = 3.33 - 3.42$, $P < 0.05$ for migration from *D. yakuba* to *D. santomea* in subunit V and $\text{LLR} = 0 - 0.23$, $P > 0.05$ in all the other subunits). Altogether these results suggest that the strongest signature of introgression in COX resides in subunit V, the innermost polypeptide in the mitochondrial membrane that interacts directly with the three mitochondrially encoded subunits (I, II, and III) (Fig. 1).

All three genes composing main subunit V (i.e., *CoVa*, *CoVb*, and *CG11043*) show the same pattern of variation: no fixed differences between species and many shared polymorphisms (Table 4). One could reason that shared polymorphisms in a group of functionally connected genes may indicate relaxed selection, which would result in abundant ancestral polymorphism. However, several observations argue against this possibility in subunit V. First, long-term relaxed functional constraints would produce fixed differences between species, but there are none. Second, d_N estimates for the three genes of subunit V analyzed in the phylogeny of *D. yakuba*, *D. teisseri*, *D. erecta*, *D. simulans*, *D. sechellia*, and *D. melanogaster* are not greater than those for other COX genes, indicating no evidence of relaxed functional constraints (Mann-Whitney U test $z = 0.83$, $P = 0.48$; Table S7). Note that in this analysis we did not use *D. santomea*, as introgression with *D. yakuba* would tend to reduce d_N . Altogether these observations suggest that the fraction of shared polymorphisms present in the three genes of COX subunit V is not due to relaxed selection.

The shared variation in all genes of subunit V supports the notion that introgression is not concentrated on a single gene but

spread throughout its three components. The probability of detecting introgression in a nuclear gene in these species is ~ 0.1 (5/49 genes) based on this study and two additional large-scale polymorphism surveys (Llopart et al. 2005b; Bachtrog et al. 2006). If we assume that this probability is similar across the genome, our observation of all three components in one main subunit showing abundant shared variation and no difference fixed between species is unlikely to be the result of unrelated introgression events ($P = 0.001$, binomial test). Altogether our results suggest the possibility that the three genes of subunit V may have cointrogressed with the *D. yakuba* mitochondrial genome.

SUBUNIT V IS A CANDIDATE FOR CYTONUCLEAR COINTROGRESSION

Mitochondrial introgression appears to be recurrent in the *D. yakuba* and *D. santomea* system in which an early invasion of the *D. yakuba* mitochondrial genome fully replaced the *D. santomea* native haplotypes (Llopart et al. 2014). In addition, a previous survey of nuclear loci randomly distributed throughout the genomes reported limited but detectable introgression between the two species (Llopart et al. 2005b). We hypothesize that genes involved in cytonuclear cointrogression will show migration rates significantly greater than the background signal from random genes in *D. yakuba* and *D. santomea*. To test this hypothesis, we inspected posterior probability distributions of migration rates for each of the main COX subunits. Only the curve corresponding to migration rates from *D. yakuba* to *D. santomea* for subunit V has a nonzero peak (Fig. 4). Similar posterior probability distributions for population migration rates from *D. yakuba* to *D. santomea* were obtained using MIMAR, which incorporates intralocus recombination (Fig. S3).

Subunit V shows an estimate of the population migration rate from *D. yakuba* to *D. santomea* that is incompatible with 0 migration ($2N_e m = 0.40$, 95%HPD: 0.079 – 3.98) and is more than one order of magnitude greater than the migration rate estimated from random genes ($2N_e m = 0.014$, 95%HPD: 0.002–0.11). We further applied Mann–Whitney tests to random subsamples from the two posterior probability distributions of $2N_e m$ using the number of genes within each category. This study indicates significantly greater migration rates for subunit V than for the random genes, with 99.88% and 98.7% of the subsamples showing significant Mann–Whitney tests at critical values of $P = 0.05$ and $P = 0.01$, respectively. In addition, we also applied a nonparametric test. We find that the ratio of shared to exclusive polymorphisms, which is an index of introgression, is significantly greater in the loci associated with subunit V than in all the other 34 genes combined (25 random genes and nine genes unrelated to COX subunit V) (Mann–Whitney test for small sample sizes, exact $P = 0.007$; Table S8). Equivalent results were obtained when the coding sequence of *CoVa* was

combined with the 5' and 3' flanking sequences into a single region (Mann–Whitney test for small sample sizes, exact $P = 0.011$; Table S8), and when only autosomal genes were analyzed (Mann–Whitney test for small sample sizes, exact $P = 0.020$). All these results combined indicate that (i) the significant introgression in subunit V is not due to a single region, and (ii) the signal of introgression in subunit V stands out from the background signal of introgression in *D. yakuba* and *D. santomea* nuclear genes.

To gain insight into the potential molecular causes of the introgression of main subunit V, we investigated its role in OXPHOS. The basic functions of subunits Va and Vb (Cox6p and Cox4p respectively in yeast) have been previously characterized in mouse and yeast, and more recently in *Drosophila* (Khalimonchuk and Rodel 2005; Fontanesi et al. 2006; Galati et al. 2009; Ren et al. 2010). Additionally, the specific protein–protein interactions of these subunits have also been characterized. These subunits are unique for several reasons. First, they are the innermost subunits of COX, and are extrinsic to the lipid bilayer (Khalimonchuk and Rodel 2005). These subunits are also unique because of their role in the formation of the COX holoenzyme. Subunit Vb (genes *CoVb* and *CG11043*) is the primary nuclear subunit involved in assembly of COX (Khalimonchuk and Rodel 2005; Fontanesi et al. 2006; Galati et al. 2009) and interacts with all three mitochondrially encoded subunits. More specifically, subunit Vb has been shown to interact intimately with mitochondrial protein Cox1 to initiate COX assembly (Tsukihara et al. 1996; Khalimonchuk and Rodel 2005). This is the first interaction that takes place in COX assembly, and once the Cox1/CoVb subcomplex is formed, the remaining subcomplexes of the holoenzyme come together to form COX (Galati et al. 2009). Subunit Vb also physically interacts tightly with mitochondrially encoded Cox3 (Tsukihara et al. 1996), is specifically regulated by the mitochondrial content of cardiolipin, and in turn enhances binding affinity of Cox2 for cytochrome c, which initiates COX activity (Galati et al. 2009). Cardiolipin is a mitochondrion-specific phospholipid required for COX activity (Khalimonchuk and Rodel 2005). Previous studies have also shown that a loss of subunit Vb prevents assembly of COX (Galati et al. 2009). Likewise, subunit Va (gene *CoVa*) has been shown to be a major regulator of COX by blocking inhibition of COX activity, and also plays a very important role in COX assembly and regulation. After the primary Cox1/CoVb interaction takes place, the CoIV/CoVa heterodimer is necessary for association with Cox2. This heterodimer then interacts with Cox1 (Fornuskova et al. 2010).

We observe similar patterns of shared variation and no fixed difference between species in the three proteins that form COX subunit V (CoVa and CoVb components). This observation, together with the complete replacement of the *D. santomea* mitochondrial genome by that of *D. yakuba* (Llopart et al. 2005b;

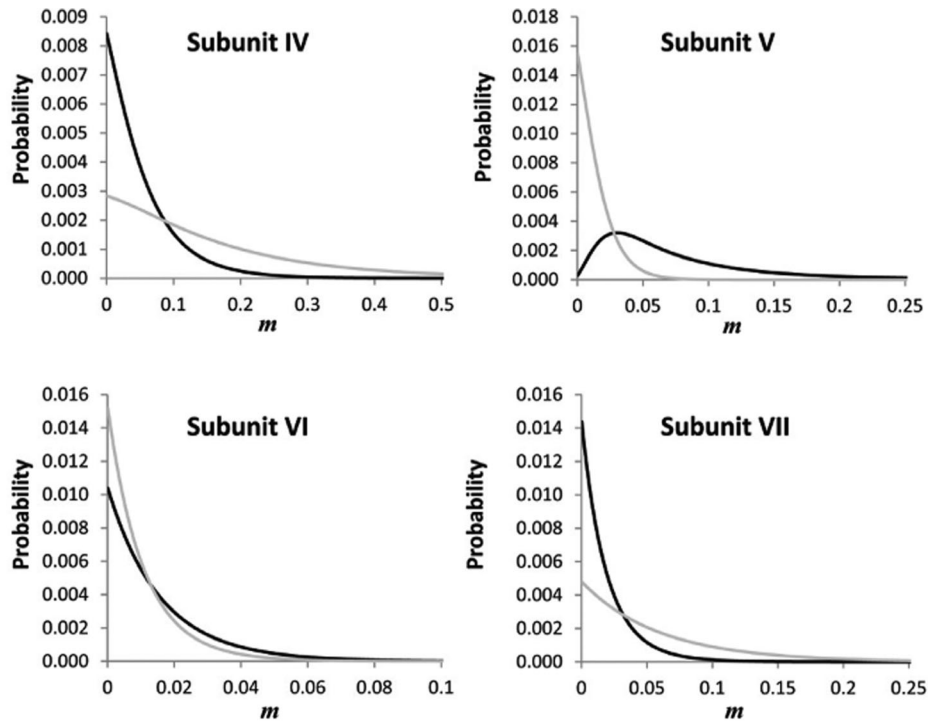


Figure 4. The marginal posterior probability distributions for migration rates ($m = m/u$) from *D. yakuba* to *D. santomea* (black line) and in the opposite direction (gray line) obtained by fitting the IM model to datasets of 2–5 regions corresponding to each main subunit using IMA2 (see legend of Fig. 3 for a detailed explanation of the units of m).

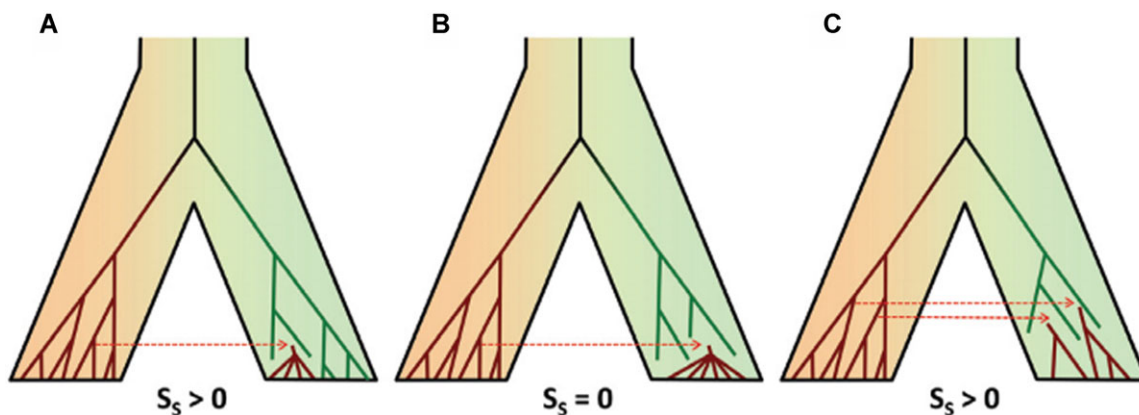


Figure 5. Schematic representation of different cases of introgression between species. (A) Adaptive introgression with partial replacement of variation in the recipient species. (B) Adaptive introgression with complete replacement of variation in the recipient species. (C) Introgression of multiple haplotypes from the donor species capturing a significant fraction of the standing variation. Note that in both (A) and (C) shared variation will be detected in the recipient species.

Llopart et al. 2014) and the unique roles of CoVa and CoVb in COX assembly, opens the possibility that the three genes of subunit V may have cointrogressed together with the mitochondrial genome. This is also supported by population migration rates for the entire subunit V that are significantly greater than the background signal of introgression in this system. However, none of the genes of subunit V shows the molecular fingerprints of a recent “hard” selective sweep (Maynard Smith and Haigh

1974; Kaplan et al. 1989; Sella et al. 2009), which has been invoked in some cases of adaptive introgression (Song et al. 2011; Pardo-Diaz et al. 2012; Staubach et al. 2012; Brand et al. 2013; Hedrick 2013; Clarkson et al. 2014; Huerta-Sanchez et al. 2014; Llopart et al. 2014). The examination of the distribution of polymorphisms across lines suggests that introgression in the genes of subunit V is not incomplete or partial, for all variation is shared and there is no evidence of diverged haplotypes (Fig. S4). Note

that, because our sample of *D. yakuba* includes sequences from allopatric populations of Africa mainland, our finding suggests that the variation present today in the three genes of subunit V consists of actually *D. yakuba* haplotypes, as it is the case for the mtDNA. We put forward the possibility that multiple *D. yakuba* nuclear haplotypes of subunit V cointrogressed into *D. santomea* capturing a fraction of the *D. yakuba* standing variation, which would easily explain the presence of shared polymorphisms and no fixed differences (Fig. 5C). At least qualitatively, this scenario would produce patterns of variation within the recipient species reminiscent of “soft” sweeps (Hermisson and Pennings 2005; Pritchard et al. 2010; Orr and Unckless 2014), as seen in *Mytilus* (Fraisse et al. 2014). Although quantitative studies of how much gene flow between species is necessary for this scenario to be realistic, the high frequency of hybridization of *D. yakuba* and *D. santomea* in nature today (~3%) (Llopart et al. 2005a) and the fertility of F₁ hybrid females (Coyne et al. 2004) suggests that abundant introgression may be allowed in some areas of the genome.

Conclusions

Our multilocus analyses indicate that a population model with gene flow from *D. yakuba* to *D. santomea* fits significantly better to the patterns of nucleotide variation in COX nuclear loci than an alternative model of isolation. More specifically, there is strong evidence that the genes composing major subunit V (genes *CoVa*, *CoVb*, and *CG11043*) are the main source of gene flow, with migration rates from *D. yakuba* to *D. santomea* significantly greater than the background signal of introgression detected in these two species. Because all three proteins show the same pattern of shared variation and lack of fixed differences between species, are part of the same main subunit, and interact with the mitochondrial-encoded core early on in COX assembly, we interpret this finding as a potential case of cointrogression. Based on what we know about the molecular function of these subunits, we suggest that cointrogression of these genes contributed to maintain proper activity of cytochrome *c* oxidase and ultimately OXPHOS.

ACKNOWLEDGMENTS

We extend special thanks to Matthew Hahn, Peter Andolfatto, and three anonymous reviewers for comments and suggestions that greatly improved this study. We thank Andrew Adrian and Josep Comeron for helpful comments on the manuscript. We are also grateful to the Carver Center for Genomics (CCG, University of Iowa) for technical assistance and Allison Valenzuela for design work. This work was partially funded by University of Iowa funds and NSF Grant 1354921 to A.LI.

DATA ARCHIVING

DNA Sequences in GenBank under accessions KF863239–KF863689.

LITERATURE CITED

- Alcaide, M., E. S. Scordato, T. D. Price and D. E. Irwin. 2014. Genomic divergence in a ring species complex. *Nature* 511:83–85.
- Anderson, E., and L. Hubricht. 1938. Hybridization in *Tradescantia* III. The evidence for introgressive hybridization. *Am. J. Bot.* 25:396–402.
- Arnold, M. L. 1997. Natural hybridization and evolution. Oxford Univ. Press, New York.
- . 2007. Evolution through genetic exchange. Oxford Univ. Press, New York.
- Bachtrog, D., K. Thornton, A. Clark and P. Andolfatto. 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. *Evolution* 60:292–302.
- Bateson, W. 1909. Heredity and variation in modern lights. Pp. 85–101 in A. C. Steward, ed. Darwin and modern science. Cambridge Univ. Press, Cambridge.
- Becquet, C. and M. Przeworski. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res.* 17:1505–1519.
- Besansky, N. J., J. R. Powell, A. Caccone, D. M. Hamm, J. A. Scott and F. H. Collins. 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. *Proc. Natl. Acad. Sci. USA* 91:6885–6888.
- Blier, P. U., F. Dufresne and R. S. Burton. 2001. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic coadaptation. *Trends Genet.* 17:400–406.
- Brand, C. L., S. B. Kingan, L. Wu and D. Garrigan. 2013. A selective sweep across species boundaries in *Drosophila*. *Mol. Biol. Evol.* 30:2177–2186.
- Breeuwer, J. A. J. and J. H. Werren. 1995. Hybrid breakdown between 2 haplodiploid species—the role of nuclear and cytoplasmic genes. *Evolution* 49:705–717.
- Bull, V., M. Beltran, C. D. Jiggins, W. O. McMillan, E. Bermingham and J. Mallet. 2006. Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol.* 4:11.
- Burton, R. S. and F. S. Barreto. 2012. A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? *Mol. Ecol.* 21:4942–4957.
- Burton, R. S., C. K. Ellison and J. S. Harrison. 2006. The sorry state of F2 hybrids: consequences of rapid mitochondrial DNA evolution in allopatric populations. *Am. Nat.* 168 (Suppl 6):S14–S24.
- Cariou, M. L., J. F. Silvain, V. Daubin, J. L. Da Lage and D. Lachaise. 2001. Divergence between *Drosophila santomea* and allopatric or sympatric populations of *D. yakuba* using paralogous amylase genes and migration scenarios along the Cameroon volcanic line. *Mol. Ecol.* 10:649–660.
- Castric, V., J. Bechsgaard, M. H. Schierup and X. Vekemans. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* 4:e1000168.
- Clark, A. G. 1997. Neutral behavior of shared polymorphism. *Proc. Natl. Acad. Sci. USA* 94:7730–7734.
- Clark, A. G., M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, W. Gelbart, V. N. Iyer, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Clarkson, C. S., D. Weetman, J. Essandoh, A. E. Yawson, G. Maslen, M. Manske, S. G. Field, M. Webster, T. Antão, B. MacInnis, et al. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat. Commun.* 5.
- Coyne, J. A., S. Elwyn, S. Y. Kim and A. Llopart. 2004. Genetic studies of two sister species in the *Drosophila melanogaster* subgroup, *D. yakuba* and *D. santomea*. *Genet. Res.* 84:11–26.
- Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

- Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* 23:3133–3157.
- Dasmahapatra, K. K., A. Silva-Vasquez, J. W. Chung and J. Mallet. 2007. Genetic analysis of a wild-caught hybrid between non-sister *Heliconius* butterfly species. *Biol. Lett.* 3:660–663.
- Dobzhansky, T. 1937. *Genetics and the origin of species*. Columbia Univ. Press, New York.
- Dobzhansky, T. 1973. Is there gene exchange between *Drosophila pseudoobscura* and *Drosophila persimilis* in their Natural Habitats? *Am. Nat.* 107:312–314.
- Dobzhansky, T., and M. L. Queal. 1938. Genetics of natural populations. II. Genic variation in populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. *Genetics* 23:463–484.
- Edmunds, S., and R. S. Burton. 1999. Cytochrome C oxidase activity in interpopulation hybrids of a marine copepod: a test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. *Evolution* 53:1972–1978.
- Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backstrom, T. Kawakami, A. Kunstner, H. Makinen, K. Nadachowska-Brzyska, A. Qvarnstrom, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* 491:756–760.
- Ellison, C. K., and R. S. Burton. 2006. Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* 60:1382–1391.
- Ellison, C. K., and R. S. Burton. 2008. Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution* 62:631–638.
- Ellison, C. K., O. Niehuis and J. Gadau. 2008. Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia* parasitoid wasps. *J. Evol. Biol.* 21:1844–1851.
- Fay, J. C., and C. I. Wu. 1999. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol. Biol. Evol.* 16:1003–1005.
- Felsenstein, J. 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* 266:418–427.
- Ferris, S. D., R. D. Sage, C. M. Huang, J. T. Nielsen, U. Ritte and A. C. Wilson. 1983. Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA* 80:2290–2294.
- Fontanesi, F., I. C. Soto, D. Horn and A. Barrientos. 2006. Assembly of mitochondrial cytochrome c-oxidase, a complicated and highly regulated cellular process. *Am. J. Physiol. Cell. Physiol.* 291:C1129–C1147.
- Fornuskova, D., L. Stiburek, L. Wenchich, K. Vinsova, H. Hansikova and J. Zeman. 2010. Novel insights into the assembly and function of human nuclear-encoded cytochrome c oxidase subunits 4, 5a, 6a, 7a and 7b. *Biochem. J.* 428:363–374.
- Fraisse, C., C. Roux, J. J. Welch and N. Bierne. 2014. Gene-flow in a mosaic hybrid zone: is local introgression adaptive? *Genetics* 197:939–951.
- Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.
- Galati, D., S. Srinivasan, H. Raza, S. K. Prabu, M. Hardy, K. Chandran, M. Lopez, B. Kalyanaraman and N. G. Avadhani. 2009. Role of nuclear-encoded subunit Vb in the assembly and stability of cytochrome c oxidase complex: implications in mitochondrial dysfunction and ROS production. *Biochem. J.* 420:439–449.
- Garrigan, D., S. B. Kingan, A. J. Geneva, P. Andolfatto, A. G. Clark, K. R. Thornton and D. C. Presgraves. 2012. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res.* 22:1499–1511.
- Gay, J., S. Myers and G. McVean. 2007. Estimating meiotic gene conversion rates from population genetic data. *Genetics* 177:881–894.
- Geraldes, A., P. Basset, B. Gibson, K. L. Smith, B. Harr, H. T. Yu, N. Bulatova, Y. Ziv and M. W. Nachman. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol. Ecol.* 17:5349–5363.
- Gibson, J. D., O. Niehuis, B. R. Peirson, E. I. Cash and J. Gadau. 2013. Genetic and developmental basis of F2 hybrid breakdown in *Nasonia* parasitoid wasps. *Evolution* 67:2124–2132.
- Giraldo, N., C. Salazar, C. D. Jiggins, E. Bermingham and M. Linares. 2008. Two sisters in the same dress: *Heliconius* cryptic species. *BMC Evol. Biol.* 8:324.
- Gordo, I., A. Navarro and B. Charlesworth. 2002. Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* 161:835–848.
- Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hedrick, P. W. 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* 22:4606–4618.
- Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.
- Herrig, D. K., A. J. Modrick, E. Brud and A. Llopart. 2014. Introgression in the *Drosophila subobscura*-*D. madeirensis* sister species: evidence of gene flow in nuclear genes despite mitochondrial differentiation. *Evolution* 68:705–719.
- Hewitt, G. M. 1988. Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol. Evol.* 3:158–167.
- Hey, J. 2003. Speciation and inversions: chimps and humans. *Bioessays* 25:825–828.
- . 2010a. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* 27:921–933.
- . 2010b. Isolation with migration models for more than two populations. *Mol. Biol. Evol.* 27:905–920.
- Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.
- Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci. USA* 104:2785–2790.
- Hoekstra, L. A., M. A. Siddiq and K. L. Montooth. 2013. Pleiotropic effects of a mitochondrial-nuclear incompatibility depend upon the accelerating effect of temperature in *Drosophila*. *Genetics* 195:1129–1139.
- Hudson, R. R., M. Kreitman and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159.
- Huerta-Sanchez, E., X. Jin, Asan, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel et al. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512:194–197.
- Janousek, V., P. Munclinger, L. Wang, K. C. Teeter and P. K. Tucker. 2015. Functional organization of the genome may shape the species boundary in the house mouse. *Mol. Biol. Evol.* 32:1208–1220.
- Kanehisa, M., and S. Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28:27–30.
- Kaplan, N. L., R. R. Hudson and C. H. Langley. 1989. The "hitchhiking effect" revisited. *Genetics* 123:887–899.
- Khalimonchuk, O., and G. Rodel. 2005. Biogenesis of cytochrome c oxidase. *Mitochondrion* 5:363–388.

- Kim, M., M. L. Cui, P. Cubas, A. Gillies, K. Lee, M. A. Chapman, R. J. Abbott and E. Coen. 2008. Regulatory genes control a key morphological and ecological trait transferred between species. *Science* 322:1116–1119.
- Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley and J. Hey. 2000. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156:1913–1931.
- Kronforst, M. R., and R. Papa. 2015. The functional basis of wing patterning in *Heliconius* butterflies: the molecules behind mimicry. *Genetics* 200:1–19.
- Kulathinal, R. J., L. S. Stevison and M. A. Noor. 2009. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet.* 5:e1000550.
- Lachaise, D., M. Harry, M. Solignac, F. Lemeunier, V. Benassi and M. L. Cariou. 2000. Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from Sao Tome. *Proc. Biol. Sci.* 267:1487–1495.
- Leffler, E. M., K. Bullaughey, D. R. Matute, W. K. Meyer, L. Segurel, A. Venkat, P. Andolfatto and M. Przeworski. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* 10:e1001388.
- Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Liu, K. J., E. Steinberg, A. Yozzo, Y. Song, M. H. Kohn and L. Nakhleh. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. *Proc. Natl. Acad. Sci. USA* 112:196–201.
- Llopart, A., S. Elwyn, D. Lachaise and J. A. Coyne. 2002. Genetics of a difference in pigmentation between *Drosophila yakuba* and *Drosophila santomea*. *Evolution* 56:2262–2277.
- Llopart, A., D. Herrig, E. Brud and Z. Stecklein. 2014. Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Mol. Ecol.* 23:1124–1136.
- Llopart, A., D. Lachaise and J. A. Coyne. 2005a. An anomalous hybrid zone in *Drosophila*. *Evolution* 59:2602–2607.
- . 2005b. Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* 171:197–210.
- Machado, C. A., R. M. Kliman, J. A. Markert and J. Hey. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Mol. Biol. Evol.* 19:472–488.
- Mallet, J., M. Beltran, W. Neukirchen and M. Linares. 2007. Natural hybridization in heliconiine butterflies: the species boundary as a continuum. *BMC Evol. Biol.* 7:28.
- Marsden, C. D., Y. Lee, C. C. Nieman, M. R. Sanford, J. Dinis, C. Martins, A. Rodrigues, A. J. Cornel and G. C. Lanzaro. 2011. Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization. *Mol. Ecol.* 20:4983–4994.
- Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet and C. D. Jiggins. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* 23:1817–1828.
- Maynard Smith, J., and J. Haigh. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* 23:23–35.
- Mayr, E. 1963. *Animal species and evolution*. Belknap Press of Harvard Univ. Press, Cambridge.
- McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- Meiklejohn, C. D., M. A. Holmbeck, M. A. Siddiq, D. N. Abt, D. M. Rand and K. L. Montooth. 2013. An Incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLoS Genet.* 9:e1003238.
- Montooth, K. L., C. D. Meiklejohn, D. N. Abt and D. M. Rand. 2010. Mitochondrial-nuclear epistasis affects fitness within species but does not contribute to fixed incompatibilities between species of *Drosophila*. *Evolution* 64:3364–3379.
- Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. *Biol. Symp.* 6:71–125.
- Nadeau, N. J., M. Ruiz, P. Salazar, B. Counterman, J. A. Medina, H. Ortiz-Zuazaga, A. Morrison, W. O. McMillan, C. D. Jiggins and R. Papa. 2014. Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res.* 24:1316–1333.
- Neafsey, D. E., R. M. Waterhouse, M. R. Abai, S. S. Aganezov, M. A. Alekseyev, J. E. Allen, J. Amon, B. Arca, P. Arensburger, G. Artemov et al. 2015. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* 347:1258522.
- Niehuis, O., A. K. Judson and J. Gadau. 2008. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics* 178:413–426.
- Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics* 158:885–896.
- Norris, L. C., B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel and G. C. Lanzaro. 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl. Acad. Sci. USA* 112:815–820.
- Nunes, M. D., P. O. Wengel, M. Kreissl and C. Schlotterer. 2010. Multiple hybridization events between *Drosophila simulans* and *Drosophila mauritiana* are supported by mtDNA introgression. *Mol. Ecol.* 19:4695–4707.
- Orr, H. A., and R. L. Unckless. 2014. The population genetics of evolutionary rescue. *PLoS Genet.* 10:e1004551.
- Osada, N., and H. Akashi. 2012. Mitochondrial-nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol. Biol. Evol.* 29:337–346.
- Pardo-Diaz, C., C. Salazar, S. W. Baxter, C. Merot, W. Figueiredo-Ready, M. Joron, W. O. McMillan and C. D. Jiggins. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* 8:e1002752.
- Poelstra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Muller, V. Baglione, P. Unneberg, M. Wikelski, M. G. Grabherr, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.
- Pritchard, J. K., J. K. Pickrell and G. Coop. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* 20:R208–R215.
- Pritchard, V. L., and S. Edmands. 2013. The genomic trajectory of hybrid swarms: outcomes of repeated crosses between populations of *Tigriopus californicus*. *Evolution* 67:774–791.
- Putnam, A. S., J. M. Scriber and P. Andolfatto. 2007. Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution* 61:912–927.
- Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds and M. Aguade. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* 166:373–388.
- Rand, D. M., R. A. Haney and A. J. Fry. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* 19:645–653.
- Ren, J. C., I. Rebrin, V. Klichko, W. C. Orr and R. S. Sohal. 2010. Cytochrome c oxidase loses catalytic activity and structural integrity during the aging

- process in *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* 401:64–68.
- Rieseberg, L. 2011. Adaptive introgression: the seeds of resistance. *Curr. Biol.* 21:R581–R583.
- Rieseberg, L. H. 2009. Evolution: replacing genes and traits through hybridization. *Curr. Biol.* 19:R119–R122.
- Riley, H. P. 1939. Introgressive hybridization in a natural population of *Tridescantia*. *Genetics* 24:753–769.
- Sackton, T. B., R. A. Haney and D. M. Rand. 2003. Cytonuclear coadaptation in *Drosophila*: disruption of cytochrome c oxidase activity in backcross genotypes. *Evolution* 57:2315–2325.
- Saraste, M. 1999. Oxidative phosphorylation at the fin de siècle. *Science* 283:1488–1493.
- Sella, G., D. A. Petrov, M. Przeworski and P. Andolfatto. 2009. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet.* 5: e1000495.
- Smeitink, J., L. vanden Heuvel and S. DiMauro. 2001. The genetics and pathology of oxidative phosphorylation. *Nat. Rev. Genet.* 2:342–352.
- Song, Y., S. Endepols, N. Klemann, D. Richter, F. R. Matuschka, C. H. Shih, M. W. Nachman and M. H. Kohn. 2011. Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr. Biol.* 21:1296–1301.
- Staubach, F., A. Lorenc, P. W. Messer, K. Tang, D. A. Petrov and D. Tautz. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet.* 8:e1002891.
- Stephens, M., N. J. Smith and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–2739.
- Teeter, K. C., B. A. Payseur, L. W. Harris, M. A. Bakewell, L. M. Thibodeau, J. E. O'Brien, J. G. Krenz, M. A. Sans-Fuentes, M. W. Nachman and P. K. Tucker. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* 18:67–76.
- The *Heliconius* Genome. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487: 94–98.
- Tripoli, G., D. D'Elia, P. Barsanti and C. Caggese. 2005. Comparison of the oxidative phosphorylation (OXPHOS) nuclear genes in the genomes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. *Genome Biol.* 6:R11.
- Tsukihara, T., H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono and S. Yoshikawa. 1996. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 272:1136–1144.
- Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts and J. A. Leunissen. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35:W71–W74.
- Waterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7:256–276.
- Weetman, D., C. S. Wilding, K. Steen, J. Pinto and M. J. Donnelly. 2012. Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Mol. Biol. Evol.* 29:279–291.
- Won, Y. J., and J. Hey. 2005. Divergence population genetics of chimpanzees. *Mol. Biol. Evol.* 22:297–307.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Associate Editor: M. Hahn
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Tajima's D test for selection in OXPHOS COX loci.

Table S2. McDonald–Kreitman test for selection in OXPHOS COX genes.

Table S3. HKA test for selection in OXPHOS COX genes.

Table S4. Values of uniform priors used in IMA2 runs.

Table S5. Parameters and converted demographic quantities obtained with IMA2.

Table S6. Estimates of recombination by GenCo.

Table S7. Number of nonsynonymous (d_N) and synonymous (d_S) changes per site obtained with PAML in COX genes analyzed in the phylogeny of *D. yakuba*, *D. teisseri*, *D. erecta*, *D. simulans*, *D. sechellia*, and *D. melanogaster*.

Table S8. Exclusive, fixed and shared variation in *D. yakuba* and *D. santomea* for the COX loci and random genes.

Figure S1. Gene trees reconstructed with the Neighbor-Joining algorithm with bootstrap values based on 10,000 replicates.

Figure S2. Gene trees reconstructed with the Maximum-Likelihood algorithm with bootstrap values based on 10,000 replicates.

Figure S3. Posterior probability distributions of the population migration rates from *D. yakuba* to *D. santomea* in the different COX subunits obtained with MIMAR.

Figure S4. Nucleotide variation (singletons not shown) in the three genes (*CoVa*, *CoVb*, and *CG11043*) of COX subunit V.