

# Multilocus Analysis of Divergence and Introgression in Sympatric and Allopatric Sibling Species of the *Lutzomyia longipalpis* Complex in Brazil

Alejandra S. Araki<sup>1,9\*</sup>, Gabriel E. M. Ferreira<sup>1,2,9</sup>, Camila J. Mazzoni<sup>3,4</sup>, Nataly A. Souza<sup>5</sup>, Ricardo C. Machado<sup>1</sup>, Rafaela V. Bruno<sup>1,6</sup>, Alexandre A. Peixoto<sup>1,6†</sup>

**1** Laboratório de Biologia Molecular de Insetos, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, **2** Laboratório de Pesquisa em Leishmaniose, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, **3** Institut für Zoo- und Wildtierforschung, Berlin, Germany, **4** Berlin Center for Genomics in Biodiversity Research, Berlin, Germany, **5** Laboratório de Transmissores de Leishmanioses, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, **6** Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular (INCT-EM), Rio de Janeiro, Brazil

## Abstract

**Background:** *Lutzomyia longipalpis*, the main vector of visceral leishmaniasis in Latin America, is a complex of sibling species. In Brazil, a number of very closely related sibling species have been revealed by the analyses of copulation songs, sex pheromones and molecular markers. However, the level of divergence and gene flow between the sibling species remains unclear. Brazilian populations of this vector can be divided in two main groups: one producing Burst-type songs and the Cembrene-1 pheromone and a second more diverse group producing various Pulse song subtypes and different pheromones.

**Methodology/Principal Findings:** We analyzed 21 nuclear loci in two pairs of Brazilian populations: two sympatric populations from the Sobral locality (1S and 2S) in northeastern Brazil and two allopatric populations from the Lapinha and Pancas localities in southeastern Brazil. Pancas and Sobral 2S are populations of the Burst/Cembrene-1 species while Lapinha and Sobral 1S are two putative incipient species producing the same pheromone and similar Pulse song subtypes. The multilocus analysis strongly suggests the occurrence of gene flow during the divergence between the sibling species, with different levels of introgression between loci. Moreover, this differential introgression is asymmetrical, with estimated gene flow being higher in the direction of the Burst/Cembrene-1 species.

**Conclusions/Significance:** The results indicate that introgressive hybridization has been a crucial phenomenon in shaping the genome of the *L. longipalpis* complex. This has possible epidemiological implications and is particularly interesting considering the potential for increased introgression caused by man-made environmental changes and the current trend of leishmaniasis urbanization in Brazil.

**Citation:** Araki AS, Ferreira GEM, Mazzoni CJ, Souza NA, Machado RC, et al. (2013) Multilocus Analysis of Divergence and Introgression in Sympatric and Allopatric Sibling Species of the *Lutzomyia longipalpis* Complex in Brazil. *PLoS Negl Trop Dis* 7(10): e2495. doi:10.1371/journal.pntd.0002495

**Editor:** Alon Warburg, The Faculty of Medicine, The Hebrew University of Jerusalem, Israel

**Received:** May 1, 2013; **Accepted:** September 8, 2013; **Published:** October 17, 2013

**Copyright:** © 2013 Araki et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grants from the Howard Hughes Medical Institute, CNPq, FAPERJ, CAPES and FIOCRUZ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: saori@ioc.fiocruz.br

9 These authors contributed equally to this work.

† Deceased.

## Introduction

Speciation events are the result of a complex array of interesting and dynamic biological processes that remain only partially understood [1]. The intricate balance among mutation, recombination, gene flow, genetic drift and natural selection can either unify or differentiate genetic variation, and this consequently may or may not promote the appearance of new species [2–5]. The evolutionary history described by individual genes may be quite variable, and the pattern of relationships among closely related species can be discordant [6–8]. In principle, the use of multiple loci should give a more complete picture of the history of divergence of species complexes, and comparisons across genes

can reveal whether all loci fit a simple model of dichotomic phylogeny. Use of multiple loci can also reveal whether retention of ancestral polymorphisms, introgression or selective pressures can explain incongruities in the evolutionary histories of species complexes [5,8,9].

Gene flow during a speciation process can be evidenced when some loci show little or no differentiation, while others show a large level. In the last decade, many studies have been carried out in species that have recently diverged, and it appears that divergence and speciation may often occur in the presence of gene flow [10–21]. This is also true for insect disease vectors, where the process of divergence with gene flow is particularly interesting beyond the standard evolutionary viewpoint. Genes

## Author Summary

The sand fly *Lutzomyia longipalpis*, the most important vector of visceral leishmaniasis in the Americas, is a complex of cryptic species distributed from Argentina to Mexico. There is evidence for the existence of a number of closely related sibling species of this complex in Brazil that differ in their male mating songs, sex pheromones and molecular markers. We compared the levels of divergence and gene-flow (introgression) in sympatric and allopatric pairs of sibling species of this complex using several molecular markers. Our results suggest that the *L. longipalpis* species complex in Brazil has an intricate evolutionary history with a crucial role for introgression, which varies between loci and is asymmetrical between the sibling species. This introgression potentially has important epidemiological implications because it could affect genes that influence the relative role that the different sibling species play as vectors. In addition, our results are particularly relevant considering the potential for increased introgression caused by man-made changes to the environment and the current trend of urbanization of visceral leishmaniasis in Brazil.

involved in vectorial capacity, insecticide resistance and adaptation to different environmental conditions could be introgressing between sibling species with important epidemiological consequences [20,22–26].

The most important neotropical vector of *Leishmania infantum*, the causative agent of American visceral leishmaniasis, is the sand fly *Lutzomyia longipalpis* (Diptera: Psychodidae), a complex of sibling species with a large distribution area ranging from northern Argentina and Uruguay to Mexico [27–36]. In Brazil, despite some incongruence among genetic markers [32,34], an integrative analysis using a combination of biochemical, behavioral and molecular traits strongly supports a number of sibling species having different levels of differentiation [33].

The Brazilian populations of *L. longipalpis* s.l. can be divided in two main groups. The first group is genetically homogeneous and widely spread, and probably represents a single species. Males of this species produce Burst-type copulation songs and the Cembrene-1 pheromone. The second group is very heterogeneous and likely represents a number of putative incipient species with more restricted distributions. These sibling species produce different subtypes of Pulse-type copulation song in combination with different sex pheromones (Germacrene, Himachalene, Cembrene-1 and Cembrene-2) [33].

The coexistence of sibling species in an overlapping geographic area is one of the best pieces of evidence for the existence of a species complex [37,38]; in at least three localities in Brazil, siblings of *L. longipalpis* s.l. are present in sympatry [33]. This is true for the Brazilian municipality of Sobral (Ceará State, Northeast Brazil), where two *L. longipalpis* sibling species were observed. In this locality, males of these two species can be distinguished by the number of pale spots on the abdomen (one or two pairs of spots: “Sobral 1S” and “Sobral 2S,” respectively). Crossing experiments show that these two siblings have strong reproductive isolation, which is consistent with the fact that their males produce different pheromones and copulation songs [29,35,39,40]. In addition, molecular markers such as microsatellites and nuclear genes clearly indicate that Sobral 1S and Sobral 2S represent two sympatric species in Sobral [41–44]. However, some of these molecular markers also show signs of introgression [41,43], and this could explain the differences in

interpretation among early studies regarding the status of the Brazilian populations [32,34].

In the current study, we conducted a multilocus approach using 21 nuclear loci to estimate and compare levels of divergence and gene flow between the sympatric siblings from Sobral and two allopatric species from the localities of Lapinha (Minas Gerais State) and Pancas (Espírito Santo State) in Southeast Brazil. Pancas and Sobral 2S are probably populations of the same species, whose males produce Burst-type songs and the Cembrene-1 pheromone, while Lapinha and Sobral 1S are two putative incipient species that share the same pheromone (Germacrene) and produce, respectively, Pulse song subtypes P2 and P3 [33]. In the absence of interbreeding, sympatric populations of two species should not be more similar to each other genetically than to allopatric populations of these two species. This kind of comparison using a larger number of unlinked loci allows to distinguish among common ancestry and the effects of introgression. Moreover, it brings new insights of how this is shaping the genome of the *L. longipalpis* species complex.

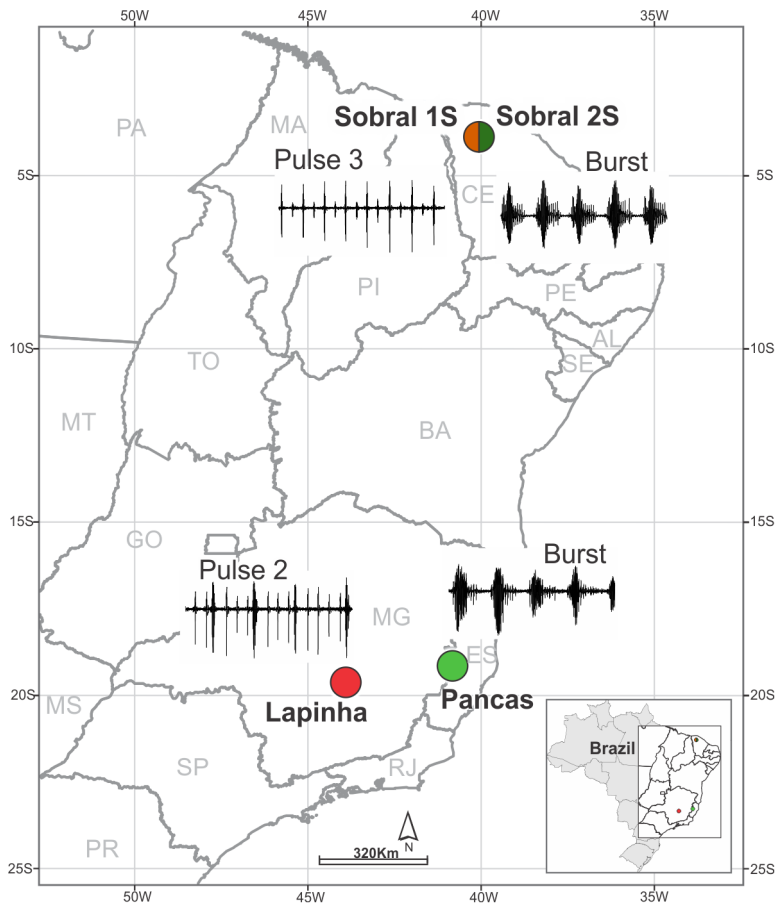
## Methods

Sand flies were collected from Sobral (Ceará state) (3°41'S, 40°20'W) in northeastern Brazil, Pancas (Espírito Santo state) (19°13'S, 40°51'W) and Lapinha (Lagoa Santa, Minas Gerais state) (19°38'S, 43°53'W) in southeastern Brazil (Figure 1). A permit for sand fly collection in Brazil was obtained from the Brazilian Ministry of Environment (SISBIO #26066-1). Sand flies were captured using CDC light-traps near human habitation with permission from local homeowners. In addition, the collections were usually supported by the local vector surveillance authorities from local State Health Departments. Samples were identified according to [45], and only males were used to avoid misidentification, as females are difficult to distinguish from other closely related species.

Twenty-one loci were used in the analysis (Table S1). Three of these loci, *period* (*per*) (accession numbers AY082911–AY082957, AF446186–AF446164, EU713179–EU713153), *cacophony* (*cac*) (accession numbers AF493078–AF493142, AF528963–AF529009, AF493120–AF493101) and *paralytic* (*para*), were used in previous studies of the *L. longipalpis* complex (accession numbers EU746318–EU746365, JQ359112–JQ359114, JQ359116, JQ359118, JQ359120, JQ359122, JQ359124–JQ359134, JQ359136, JQ359138–JQ359139, JQ359264–JQ359266, JQ359268–JQ359272, JQ359274–JQ359276, JQ359278–JQ359295) [33,36,42–44,46].

The remaining 18 loci (Table S2) are new markers randomly selected from a screen of cDNA sequences performed in our lab. The selected loci are responsible for different functions (Table S1), show a high similarity at the protein level when compared to *Anopheles gambiae* and/or *Drosophila melanogaster* and contain one intron in the latter species under the assumption that they might also be present in the *L. longipalpis* studied fragment, although that was not always the case.

DNA samples used for PCR were prepared as previously described [24]. PCR was carried out using proofreading *Pfu* DNA Polymerase (Biotools) for 35 cycles: 95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. PCR products were purified using the Wizard PCR Preps kit (Promega) or Micro Spin S-400 HR (GE Healthcare) and were cloned into the pMOS Blunt-ended Vector (GE Healthcare). The plasmid DNA was prepared using the alkaline lysis method in 96 well micro-plates [47] and filtered in multiscreen filter plates. Sequencing was carried out using the ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction V3.1 kit (Applied Biosystems) in an ABI 3730 DNA Sequencer by



**Figure 1. Geographic locations of the Brazilian populations of *Lutzomyia longipalpis* species complex studied.** The circles show the approximate location of the two sympatric sibling species from Sobral (Ceará State) and the two allopatric sibling species from Pancas (Espírito Santo State) and Lapinha (Minas Gerais State). Divided circle represents the two sympatric species Sobral 2S (green) and Sobral 1S (orange), with Burst and Pulse (subtype 3) types of copulation songs, respectively. Circles in light green and red correspond to the allopatric species Pancas and Lapinha that produce Burst and Pulse (subtype 2) types of copulation songs, respectively.  
doi:10.1371/journal.pntd.0002495.g001

the PDTIS (FIOCRUZ, Rio de Janeiro, Brazil) sequencing service [48]. The sequences have been submitted to GenBank (accession numbers JX301771–JX303202).

The DNA sequences were edited and aligned using Bioedit v.7.0.9.0 [49] or MEGA5 [50]. The polymorphism summaries, the neutrality tests and the differentiation among populations were estimated using DnaSP5 [51] and ProSeq 2.91 software [52]. The HKA multilocus test [53] of neutral molecular evolution was carried out using HKA software (<http://genfaculty.rutgers.edu/hey/home>). Haplotype networks were constructed using statistical parsimony [54] implemented in the TCS software [55].

We analyzed our data set under the isolation-with-migration model available with the IM software [12] to discriminate between the relative effects of divergence and gene flow. This multilocus analysis was performed using non-recombining blocks (NRB) of the 21 different nuclear loci. Sites within indels or ambiguous alignment were removed (Table S3). To select the NRB, we used IMgc software [56] for the Sobral 1S and 2S sequences and manually selected the blocks and sequences of Pancas and Lapinha according to the NRB from Sobral to ensure consistency. Some putative recombinant sequences were excluded from the data set. The IM model considers parameters for splitting time ( $t$ ), bidirectional gene flow after splitting ( $m_1$  and  $m_2$ ) and current effective population sizes ( $\theta_1$ ,  $\theta_2$  and  $\theta_A$ ). To fit the model to our

data, the IM software used a Bayesian framework that gave estimates for the posterior probability densities of the model parameters using a Markov chain Monte Carlo simulation [12]. We carried out two sets of analyses, one comparison involving the sympatric siblings from Sobral (Sobral 1S *vs.* Sobral 2S) and one for the two allopatric siblings (Pancas *vs.* Lapinha). For each analysis, we ran simulations assuming the Infinite Site model (IS) [57] of sequence evolution recommended for nuclear loci. In addition, for each comparison (sympatric or allopatric) we ran simulations that estimated the population migration rates and simulations that estimated the per locus migration rates. We conducted preliminary runs to determine appropriate priors for subsequent runs. After that we performed three to five independent runs with different starting points to ensure that the values converged on similar estimates, with ten or twenty independent chains under Metropolis coupling. Each chain was initiated with a burn-in period of 200,000 updates, and the total run length of each analysis was between 30 million and 40 million updates, using the following input parameters for simulations:  $m = 25$  and 100,  $\theta = 5$  and 10,  $t = 5$ . Convergence was assessed by estimating the effective sample size, which was always over 50. The six demographic parameters were calculated from the values of the bin with the highest count; these values corresponded to the average calculated from three to five independent runs. For

credibility intervals for each parameter, we recorded the 90% posterior density interval, the shortest span that includes 90% of the probability density of a parameter. To convert the parameter estimates into demographic values, we used *Drosophila melanogaster* synonymous and non-synonymous substitution rates for nuclear genes,  $1.56 \times 10^{-8}$  and  $1.91 \times 10^{-9}$  per site per year, respectively [58]. From the 21 loci used in this study, we calculated a geometric mean of mutation rate per loci per year,  $\mu = 1.77 \times 10^{-6}$ , that was used to rescale the IM parameter estimates. This value was used to convert the parameter estimate  $t$  to the number of years since population splitting ( $t$ ), assuming that *L. longipalpis s.l.* has about ten generations per year, and to convert the estimates of the population mutation rates ( $\theta_1$ ,  $\theta_2$  and  $\theta_A$ ) into estimates of the effective population sizes ( $N_1$ ,  $N_2$  and  $N_A$ ).

## Results

### Intra-population nucleotide variation

Our data set included 21 loci. The markers are responsible for different functions and potentially spread throughout the genome of *L. longipalpis*, based on their positions in *A. gambiae* and *D. melanogaster* (Table S1, see also Methods for more details).

Table 1 shows a summary of the multilocus DNA polymorphisms for the four *L. longipalpis s.l.* populations sampled in eastern Brazil (Figure 1). In general, for each molecular marker, the levels of variation within populations were similar. However, we observed considerable variation in the levels among loci. The *CG9297*, *rpl36*, *tflIAL* and *obp19a* loci were the most polymorphic,

and *sesB*, *eno*, *tropC* and *CG9769* were the least polymorphic of the markers analyzed. Indels were observed in thirteen markers, and all were located in introns.

Table 2 shows the Tajima's  $D$  statistics [59]. Although not significantly different from zero after Bonferroni's correction, values were negative in most cases, possibly indicating population expansion. Additional tests, such as the Fu's  $F_s$  [60] and Ramos-Onsins and Rozas  $R^2$  [61] (considered more powerful to detect population growth [62]) also support the expansion hypothesis, mainly in Sobral (Table S4). Nevertheless, the HKA multilocus test showed no significant deviations from the neutral expectations in the sympatric pair ( $X^2_{\text{Sobral 1S-Sobral 2S}} = 9.26$ ,  $df = 40$ ,  $P > 0.99$ ) or in the allopatric pair of species comparisons ( $X^2_{\text{Lapinha-Pancas}} = 26.15$ ,  $df = 40$ ,  $P > 0.95$ ), indicating no obvious departures from neutrality in either case.

### Differentiation and introgression between sympatric and allopatric pairs of sibling species

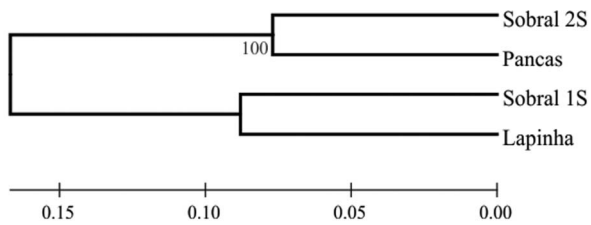
We carried out a comparison of the differentiation between the two pairs of sibling species: sympatric (Sobral 1S vs. Sobral 2S) and allopatric (Lapinha vs. Pancas). As mentioned before, Pancas and Sobral 2S are populations of the *Burst/Cembrene* species while Lapinha and Sobral 1S are two putative incipient species producing the same pheromone (Germacrene) and Pulse song subtypes P2 and P3 [33]. Table 3 shows the  $F_{ST}$  values for each locus for the Sobral 1S vs. Sobral 2S and Lapinha vs. Pancas comparisons. Table S5 shows the values for the other pairwise comparisons while Table S6 exhibits Nei's genetic distance ( $D_{xy}$ )

**Table 1.** Polymorphism summary for the 21 loci in four siblings of *L. longipalpis* complex from Brazil.

Locus	Length (bp) <sup>1</sup>	N				S				$\pi$				$\Theta$			
		S1S	S2S	Lap	Pan	S1S	S2S	Lap	Pan	S1S	S2S	Lap	Pan	S1S	S2S	Lap	Pan
<i>CG9297</i>	474	32	29	18	25	36	37	42	32	0.024	0.019	0.028	0.009	0.019	0.020	0.027	0.019
<i>CG9769</i>	378	24	24	20	14	12	8	1	3	0.005	0.035	0.000	0.002	0.009	0.006	0.001	0.003
<i>eno</i>	245	28	32	29	30	4	9	3	5	0.002	0.003	0.003	0.002	0.004	0.009	0.003	0.005
<i>kinC</i>	664	17	23	24	24	25	28	32	31	0.010	0.009	0.012	0.010	0.011	0.012	0.013	0.013
<i>mlcc</i>	217	30	28	17	24	12	10	7	7	0.005	0.006	0.008	0.007	0.012	0.014	0.010	0.009
<i>norpA</i>	104	32	30	38	25	8	5	9	9	0.020	0.014	0.017	0.024	0.019	0.012	0.021	0.023
<i>obp19a</i>	164	32	31	28	26	16	17	13	13	0.029	0.033	0.026	0.028	0.024	0.026	0.020	0.021
<i>rpl17A</i>	302	31	31	25	14	13	22	7	2	0.013	0.014	0.007	0.003	0.011	0.019	0.006	0.002
<i>rpl36</i>	450	21	16	32	23	44	48	25	26	0.025	0.031	0.020	0.015	0.029	0.034	0.015	0.017
<i>rpS19</i>	300	33	40	27	10	20	27	17	10	0.016	0.019	0.017	0.010	0.017	0.021	0.015	0.012
<i>sesB</i>	89	23	26	21	15	2	1	0	2	0.004	0.006	0.000	0.004	0.006	0.003	0.000	0.007
<i>slh</i>	258	27	30	19	16	22	24	11	12	0.028	0.018	0.011	0.015	0.023	0.024	0.013	0.015
<i>sec22</i>	429	17	15	29	16	26	19	14	24	0.013	0.015	0.007	0.020	0.018	0.014	0.008	0.017
<i>sod2</i>	310	26	27	27	16	10	15	5	3	0.008	0.008	0.006	0.002	0.009	0.013	0.005	0.003
<i>tflIAL</i>	377	14	18	20	38	23	33	29	28	0.020	0.030	0.020	0.020	0.022	0.029	0.025	0.019
<i>tropC</i>	483	27	27	29	31	12	12	3	10	0.003	0.004	0.000	0.003	0.007	0.007	0.002	0.005
<i>up</i>	404	21	22	14	16	19	11	14	10	0.010	0.006	0.011	0.009	0.013	0.008	0.011	0.008
<i>ζcop</i>	386	30	16	28	21	14	10	6	7	0.005	0.007	0.003	0.004	0.010	0.008	0.004	0.005
<i>cac</i>	109	23	22	20	26	12	5	14	6	0.025	0.025	0.066	0.026	0.043	0.018	0.053	0.021
<i>para</i>	377	27	22	21	29	12	8	2	10	0.004	0.004	0.002	0.005	0.009	0.006	0.002	0.007
<i>per</i>	266	24	23	23	19	27	22	20	18	0.024	0.017	0.018	0.014	0.027	0.022	0.020	0.019

<sup>1</sup>Aligned sequence length in base-pairs; N, number of sequences; S, segregating sites;  $\pi$ , nucleotide divergence average number of pair-wise differences;  $\Theta$ , nucleotide divergence based on the number of segregating sites. S1S, Sobral one spot; S2S, Sobral two spot; Lap, Lapinha; Pan, Pancas.

doi:10.1371/journal.pntd.0002495.t001



**Figure 2. Neighbor joining tree of *Lutzomyia longipalpis* populations based on the pairwise  $F_{ST}$  of 21 loci.** The bootstrap value is based on 100 replicates in which the 21 loci were bootstrapped. doi:10.1371/journal.pntd.0002495.g002

for all pairwise comparisons. A tree of these four populations based on the mean  $F_{ST}$  values clearly shows that the closer genetic similarity of the two Burst-type populations (Pancas and Sobral 2S) at one side and the two Pulse-type populations (Lapinha and Sobral 1S) in the other (Figure 2). In addition, it also shows that although Lapinha and Sobral 1S produce different subtypes of Pulse songs and probably represent incipient species, the overall level of genetic divergence between them is only slightly higher than between the two Burst-type populations.

The sympatric pair shared polymorphisms in almost all loci and only exhibited fixed differences in *para* (Table 3). On the other hand, the allopatric pair of species only shared polymorphisms in about half of the 21 loci and a total of 17 fixed differences were observed in six loci. The lower level of divergence between

sympatric species than allopatric species was also observed after inspection of the observed  $F_{ST}$  values (Table 3 and Figure 3). The mean pairwise  $F_{ST}$  was more than twice as high for the allopatric pair ( $0.453 \pm 0.245$ ) than for the sympatric pair ( $0.213 \pm 0.209$ ), and this difference was significant ( $t = -3.413$ ;  $p < 0.01$ ). While all 21  $F_{ST}$  values were significant in the case of the allopatric pair, five of the values for the sympatric siblings failed to reach significance. Out of 21 loci, only two (*sec22* and *up*) showed higher  $F_{ST}$  values between the sympatric pair than between the allopatric populations.

The data in Table 3 and Figure 3 strongly suggest the occurrence of introgression between the sympatric populations. In addition, the estimated  $F_{ST}$  statistics were highly variable between loci in both comparisons. Part of the variation was likely caused by differences in the rate of evolution of the different loci. However, as shown in Figure 3, differences in  $F_{ST}$  values between the allopatric and sympatric comparisons tend to be lower for loci that have high  $F_{ST}$  values in the sympatric pair. In fact, Figure 4A shows that there is a highly significant negative correlation ( $r = -0.896$ ;  $p < 0.001$ ) between the ranked (from the smallest to the highest) normalized differences in  $F_{ST}$  values between the sympatric and allopatric pairs at each loci and the respective rank of  $F_{ST}$  values in the sympatric pair at the same loci (see figure legend for more details). No correlation is observed in the case of the allopatric populations ( $r = 0.231$ ;  $p = 0.3137$ ) (Figure 4B). These results suggest that selection is acting as a filter on gene flow and, as a result, introgression is differential, affecting some loci to a much greater extent than others. This shows that, while one can predict that a molecular marker with a high  $F_{ST}$  value in a sympatric comparison will most likely show a similar value between two allopatric sibling species, the same is not true for many of the high  $F_{ST}$  values observed between allopatric populations. A gene that differentiates sympatric species is likely a good general marker for the complex. The same is not necessarily true for markers that differentiate allopatric species.

Figure 5 illustrates this further by comparing the haplotype networks of a few markers (*ζcop*, *rpL17A*, *per*, *up*, *sec22* and *para*) in sympatric and allopatric populations. The first two, *rpL17A* and *ζcop* (Figure 5A), represent examples of markers which showed high  $F_{ST}$  values in allopatry but low values in sympatry (Figure 3). The allopatric networks of these two loci show much better separation between the haplotypes, especially in the case of *rpL17A*, which shows no separation at all in sympatry. On the other hand, in the case of the four loci with high  $F_{ST}$  values in sympatry (*per*, *up*, *sec22* and *para*), the networks are somewhat similar in the allopatric and sympatric comparisons (Figure 5B). Therefore, these markers seem to be more reliable for the study of the divergence between populations of the *L. longipalpis* complex as they avoid the problems caused by strong introgression.

The results presented above show a complex pattern of divergence and gene flow between the Brazilian species of the *L. longipalpis* complex. In particular, we found evidence of differential introgression among loci in the sympatric pair of species. To explore in more detail the gene flow between sympatric and allopatric relatives, we performed a multilocus analysis using the IM software.

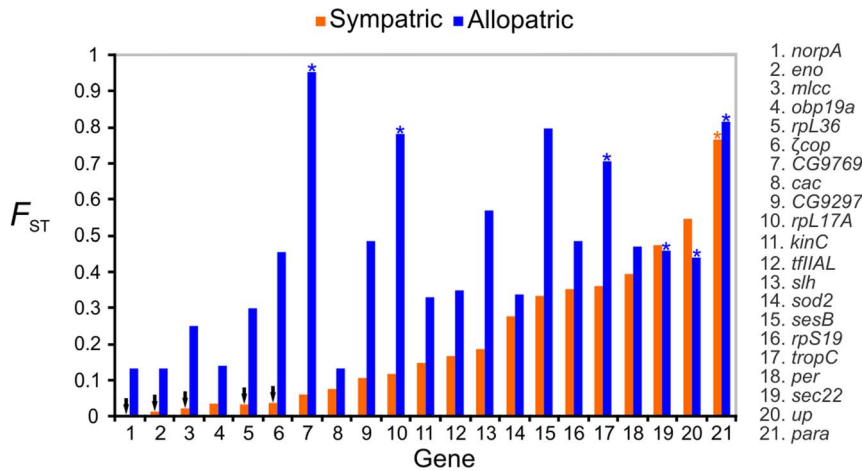
### Isolation with migration analysis

Because the IM software requires the absence of recombination within loci, a non-recombining block (NRB) was extracted from each gene alignment, and some putative recombinant sequences were excluded from the data set (Table S3, see Methods). To evaluate the effects of the decrease in fragment length and/or the number of sequences per loci, we conducted polymorphism and

**Table 2. Tajima's neutrality test of 21 loci for the four siblings of *L. longipalpis* from Brazil.**

Locus	Tajima's D			
	S1S	S2S	Lap	Pan
<i>CG9297</i>	0.484	-0.708	0.676	-1.970
<i>CG9769</i>	-1.620	-1.244	-1.164	-0.886
<i>eno</i>	-1.721	-1.977	0.075	-1.749
<i>kinC</i>	-0.511	-0.970	-0.613	-0.974
<i>mlcc</i>	-1.710	-1.998	-0.461	-0.449
<i>norpA</i>	-0.319	-0.112	-0.510	-0.602
<i>obp19a</i>	0.196	0.192	0.603	1.122
<i>rpL17A</i>	0.211	-1.072	0.155	0.985
<i>rpL36</i>	-0.753	-0.598	1.257	-0.247
<i>rpS19</i>	-0.447	-0.600	0.287	-0.782
<i>sesB</i>	-0.636	1.591	-	-1.002
<i>slh</i>	0.550	-1.230	-0.639	-0.035
<i>sec22</i>	-1.219	-0.884	0.118	0.118
<i>sod2</i>	-0.279	-1.270	0.645	-0.525
<i>tflIAL</i>	-0.742	0.012	-0.736	0.158
<i>tropC</i>	-1.823	-1.371	-1.733	-1.507
<i>up</i>	-1.082	-0.988	0.045	0.486
<i>ζcop</i>	-1.451	-0.929	-0.786	-0.608
<i>cac</i>	-1.484	0.726	0.176	-1.970
<i>para</i>	-1.878	-1.081	0.222	-0.832
<i>per</i>	-0.538	-0.852	-0.432	-1.007

S1S, Sobral 1S; S2S, Sobral 2S; Lap, Lapinha; Pan, Pancas. doi:10.1371/journal.pntd.0002495.t002



**Figure 3. The distribution of  $F_{ST}$  values comparing sympatric and allopatric *Lutzomyia longipalpis* species from Brazil.** Bars indicate the  $F_{ST}$  values for 21 nuclear markers from sympatric (orange) and allopatric (blue) pairs of species. The  $F_{ST}$  values were ordered according to ascending values from the sympatric comparisons. Black arrows indicate non-statistically significant values, and asterisks show markers with fixed polymorphic sites.

doi:10.1371/journal.pntd.0002495.g003

**Table 3. Differentiation between sympatric and allopatric siblings of *L. longipalpis* complex from Brazil.**

Locus	Sympatric species			Allopatric species		
	$F_{ST}$	Ss	Sf	$F_{ST}$	Ss	Sf
CG9297	0.106 <sup>***</sup>	26	0	0.486 <sup>***</sup>	9	0
CG9769	0.059 <sup>*</sup>	3	0	0.954 <sup>***</sup>	0	8
eno	0.007 <sup>ns</sup>	2	0	0.134 <sup>***</sup>	0	0
kinC	0.145 <sup>**</sup>	10	0	0.331 <sup>***</sup>	10	0
mlcc	0.021 <sup>ns</sup>	3	0	0.250 <sup>***</sup>	0	0
norpA	0.000 <sup>ns</sup>	6	0	0.130 <sup>***</sup>	3	0
obp19a	0.032 <sup>*</sup>	14	0	0.138 <sup>***</sup>	8	0
rpl17A	0.115 <sup>***</sup>	11	0	0.782 <sup>***</sup>	0	2
rpl36	0.034 <sup>ns</sup>	29	0	0.300 <sup>***</sup>	9	0
rpS19	0.351 <sup>***</sup>	10	0	0.486 <sup>***</sup>	3	0
sesB	0.334 <sup>***</sup>	0	0	0.796 <sup>***</sup>	0	0
slh	0.184 <sup>***</sup>	13	0	0.568 <sup>***</sup>	0	0
sec22	0.475 <sup>***</sup>	3	0	0.461 <sup>***</sup>	1	1
sod2	0.276 <sup>***</sup>	7	0	0.337 <sup>***</sup>	0	0
tflIAL	0.165 <sup>***</sup>	15	0	0.350 <sup>***</sup>	20	0
tropC	0.359 <sup>***</sup>	5	0	0.705 <sup>***</sup>	0	1
up	0.547 <sup>***</sup>	5	0	0.438 <sup>***</sup>	0	1
ζcop	0.036 <sup>ns</sup>	5	0	0.454 <sup>***</sup>	0	0
cac	0.076 <sup>*</sup>	4	0	0.130 <sup>***</sup>	4	0
para	0.766 <sup>***</sup>	1	3	0.815 <sup>***</sup>	0	4
per	0.395 <sup>***</sup>	6	0	0.471 <sup>***</sup>	3	0

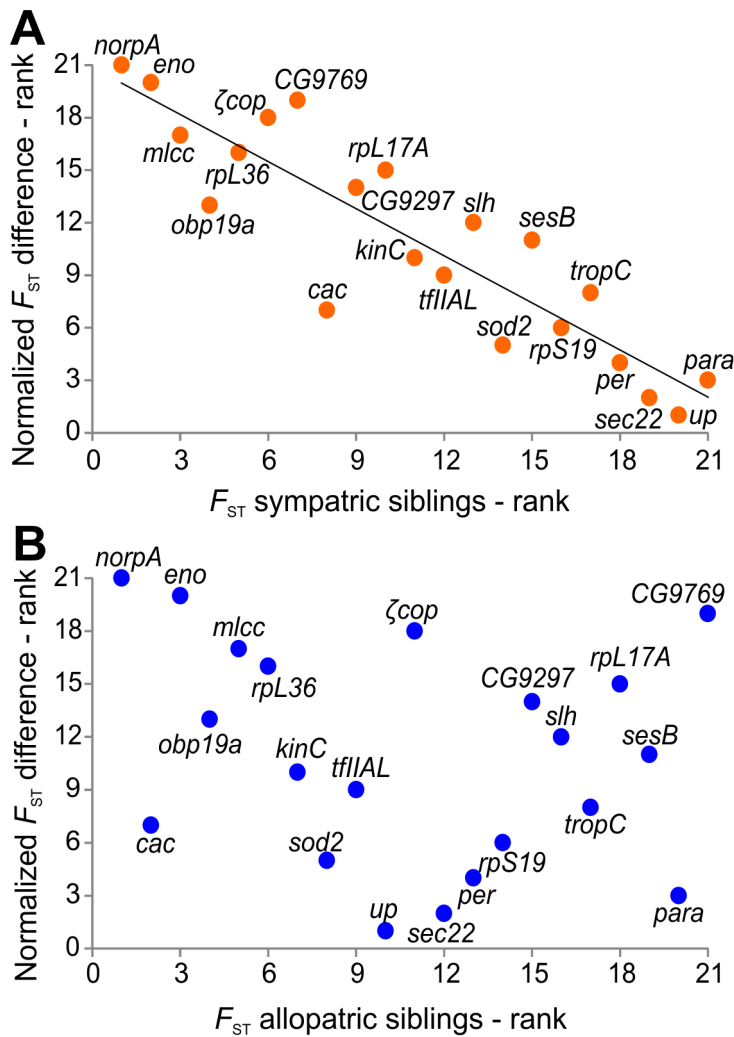
$F_{ST}$ , pairwise fixation index. Significance evaluated with 1000 permutations; <sup>\*\*\*</sup>, significant at  $P < 0.001$ ; <sup>\*\*</sup>, significant at  $P < 0.01$ ; <sup>\*</sup>, significant at  $P < 0.05$ ; <sup>ns</sup>, non-significant  $P > 0.05$ . Ss, shared sites; Sf, fixed sites.

doi:10.1371/journal.pntd.0002495.t003

divergence analyses (Table S7 and S8) similar to that carried out for the whole fragment (WF). The within-population variation in NRB was not significantly different from WF in any of the populations (Wilcoxon test:  $\pi_{Sobral\ 1S}$ :  $p \geq 0.782$ ;  $\theta_{Sobral\ 1S}$ :  $p \geq 0.890$ ;  $\pi_{Sobral\ 2S}$ :  $p \geq 0.818$ ;  $\theta_{Sobral\ 2S}$ :  $p \geq 0.495$ ;  $\pi_{Lapinha}$ :  $p \geq 0.891$ ;  $\theta_{Lapinha}$ :  $p \geq 0.546$ ;  $\pi_{Pancas}$ :  $p \geq 0.323$ ;  $\theta_{Pancas}$ :  $p \geq 0.312$ ). In addition, the comparisons between pairs of  $F_{ST}$  values also showed that the differences between NRB and WF remained small and unbiased in most cases (Wilcoxon test:  $F_{ST\ Sobral\ 1S\ vs.\ Sobral\ 2S}$ :  $p \geq 0.159$  and  $F_{ST\ Lapinha\ vs.\ Pancas}$ :  $p \geq 0.444$ ).

IM parameters (Table 4) were estimated to infer and compare the population history of the two sympatric and the two allopatric sibling species of the *L. longipalpis* complex. Figure 6 shows the posterior probability distributions for six parameters with single narrow peaks and bounds that fell within the prior distributions. In all cases, each of the replicates yielded a posterior distribution with identical and well-defined modes. The marginal posterior distribution for  $\theta$  of the sympatric species showed slightly different distributions (Figure 6A). The current effective population sizes estimated from these values indicate a three-fold increase for Sobral 1S and a two-fold increase for Sobral 2S since their splitting from the ancestral population. On the other hand, the  $\theta$  parameters for Pancas and Lapinha indicate that these two allopatric relatives and the ancestral population have similar effective population sizes (Figure 6A). The marginal posterior distributions for  $t$  suggest that, as expected, the sympatric pair of species split from the ancestral population at approximately the same time as the allopatric pair (Figure 6B). Therefore, assuming mutation rates similar to *Drosophila* (see Material and Methods), the splitting event that separated the Burst-type (Sobral 2S and Pancas) and Pulse-type song (Sobral 1S and Lapinha) lineages occurred approximately 0.5 MYA (Table 4).

Shared variation may be the result of gene flow or the incomplete sorting of ancestral polymorphism. The distinction between both possibilities is the main goal of several statistical tests [63] like the one implemented by IM through coalescent simulations [12]. Based on multilocus estimates of population migration rates showing nonzero peaks, our results strongly suggested that after separation, the two lineages have undergone gene flow in both directions for the sympatric and for the allopatric



**Figure 4. Correlation between species'  $F_{ST}$  value ranks and the normalized  $F_{ST}$  differences ranks.** (A) Correlation between the ranks of the normalized differences in  $F_{ST}$  values  $[(F_{ST} \text{ Allopatric} - F_{ST} \text{ sympatric})/F_{ST} \text{ allopatric}]$  and the ranks of  $F_{ST}$  values for the sympatric species pair at 21 markers. High significant negative correlation was observed, and the trend line for all data calculated by ordinary least squares regression is shown by a solid line. (B) Correlation between the ranks of the normalized differences in  $F_{ST}$  and the ranks of  $F_{ST}$  values for the allopatric species pair at 21 markers.

doi:10.1371/journal.pntd.0002495.g004

species pairs (Table 4, Figure 6C). Although the effective calculated number of gene migrants per generation indicates a bidirectional migration in both cases, gene flow is five to ten times higher in the sympatric species. In addition, the observed level of introgression is highly asymmetrical, with Sobral 2S receiving about seven times more migrants than Sobral 1S. The estimates of the mean time of migration events [13] indicate that most of them occurred between 0.13 MYA and 0.20 MYA, suggesting lower current levels of gene flow.

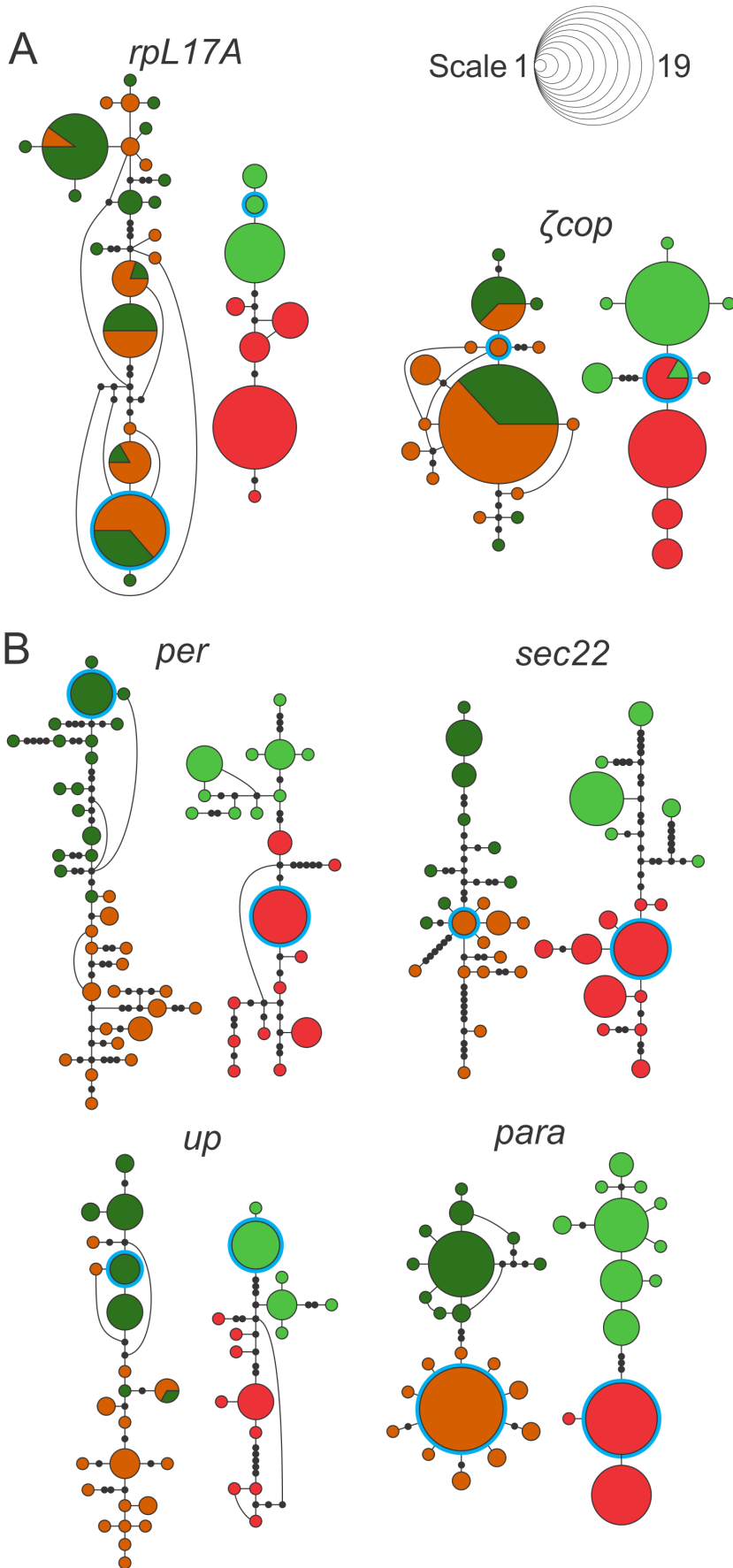
We also estimated the per locus migration rates. Figure 7 shows the marginal posterior distributions for each gene. Migration at several loci was observed in both pairs of species. However, although sympatric and allopatric comparisons showed a similar number of loci that exhibited signals of gene flow, in general the allopatric comparison showed much lower levels of gene flow. In addition, most loci showed unidirectional gene flow and introgression from Sobral 1S to Sobral 2S. Weak divergence and extensive gene flow may result in a flat and diffuse posterior

distribution [64], and this most likely explains the results observed for a number of loci in the case of the Sobral sympatric siblings.

## Discussion

The analysis of the divergence and gene flow at 21 nuclear loci in sympatric and allopatric Brazilian populations of *L. longipalpis* s.l. shows an intricate evolutionary history. The results suggest that, at least for the Brazilian species, introgressive hybridization has played a crucial role in the speciation process in this complex. In closely related species, this phenomenon seems to be common, and multilocus analyses have been useful in these cases [4,5]. Overall, our results strongly suggest that introgression has played an active role in shaping the genomes of species in the *L. longipalpis* complex, as previously shown, for example, in *Anopheles*, *Heliconius* and *Drosophila* closely related species (e.g. [8,22,65–67]).

The occurrence of sympatric species of the *L. longipalpis* complex in Sobral was strongly supported by our analysis, and this finding





**Figure 5. Haplotype networks from sympatric and allopatric comparisons of selected loci with extreme  $F_{ST}$  values.** (A) The network of the loci  $\zeta cop$  and  $rpL17A$  from the sympatric species of the *Lutzomyia longipalpis* complex shows mixed a haplotype distribution, unlike the two well-separated cluster for the allopatric species. This is in agreement with the low degree of divergence between sympatric and high divergence between allopatric species. (B) The loci *per*, *sec22*, *up* and *para* presented sympatric and allopatric networks with haplotypes separated by species group, which also corroborates the high values of pairwise  $F_{ST}$  (see Table 2). Each *L. longipalpis* population is represented by one color: Sobral 1S (orange), Sobral 2S (green), Lapinha (red) and Pancas (light green). Colored circles represent unique haplotypes, and their sizes are proportional to their frequencies. Black circles are hypothetical haplotypes. Curved lines represent alternative branching between haplotypes.  
doi:10.1371/journal.pntd.0002495.g005

agrees with previous work [32]. Moreover, the high number of shared polymorphisms, some full shared haplotypes in a number of loci and the lack of fixed polymorphisms (except those previously reported for *para* [44] were the outstanding genetic traits observed in those two sibling species. The extremely variable levels of divergence among loci were the first clue that an evolutionary model of isolation with migration might fit our data [5,14,68].

The comparison between sympatric *vs.* allopatric populations provide further evidence for the occurrence of introgression between the sympatric species. If the sympatric pair alone had been analyzed, it might have been more difficult to determine whether the low divergence observed in some loci was merely a consequence of retention of ancestral polymorphisms or was the result of gene flow. However, as in some other studies (e.g., [23,69,70], the sympatric/allopatric comparison provided a clearer picture of the divergence process and indicated the occurrence of gene flow, which was confirmed by the IM analysis [12].

Interestingly, the allopatric sibling data also fit the isolation with migration model, which might reflect gene flow coming from populations of the same species subject to introgression in sympatry. The estimates of migration rates, time of divergence and mean  $F_{ST}$  between the allopatric sibling species, all satisfy the tentative threshold criteria for species diagnosis proposed by Hey and Pinho [71]. However, in the case of the sympatric species, the situation is a bit less clear with introgression affecting the mean  $F_{ST}$  and migration rates.

The observed level of introgression is highly asymmetrical with Sobral 2S receiving about seven times more migrants than Sobral 1S. Asymmetry in gene flow might be caused by differences in population density, and the gene flow direction is often predominantly from the most abundant species to the least [72–74]. In fact, our estimates of the  $\theta$  population size parameters suggest that Sobral 1S is larger than Sobral 2S. The evidence for differential introgression among loci fits the mosaic model of speciation (e.g. [8,22,75,76] in sympatric and allopatric species and can be used to identify genomic regions containing genes involved in speciation [77]. Some genes may have freely crossed the species boundaries; these genes include housekeeping and other genes that may not be associated with reproductive isolation or species-specific adaptation (e.g., *sod2* and *tjIAL*) and were therefore not subjected to selective pressure against introgression [78]. Among the loci used in this study, three (*per*, *cac*, and *para*) are known to control courtship songs in *Drosophila* [79] and, interestingly, two (*per* and *para*) are among those with highest  $F_{ST}$  values. Future analysis of other song genes may reveal whether these genes tend to show higher levels of divergence and lower levels of gene flow between the *L. longipalpis* sibling species.

Assuming that the speciation process between the Burst/Cembrene and Pulse/Germacrene lineages occurred in allopatry (~0.5 MYA), followed by secondary contact in localities such as Sobral, the evidence for introgression provided by the data suggests that, at first, the period of separation was not long enough to ensure the appearance of full reproductive isolation mechanisms. Following the secondary contact and a period of stronger

hybridization and introgression, reinforcement of reproductive isolation might have promoted the evolution of more efficient mate discrimination, and other mechanisms of isolation could have taken place [67,80]. Differences in male sex pheromones [35], copulation songs [39], locomotor activity rhythms [81] and life cycle traits [82] between Sobral 1S and 2S indicate that selection might be playing an active role in the divergence process of the two sibling species. In fact, there is evidence that the reproductive isolation between the sympatric Sobral siblings is stronger than between allopatric siblings of the *L. longipalpis* complex [29,40], and gene flow might therefore be currently diminished.

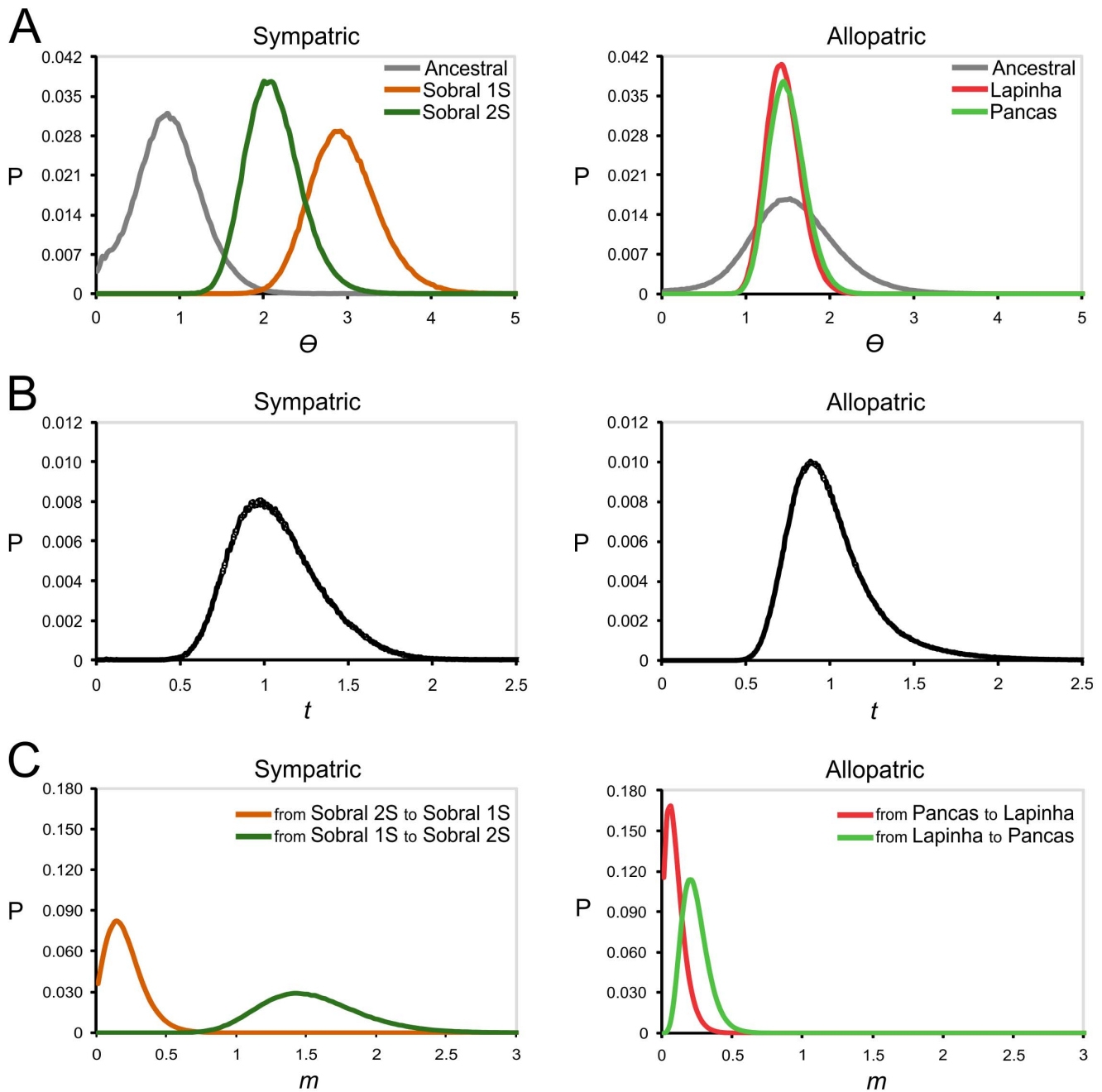
In addition to Sobral, two other pairs of sympatric species in Jaiba (Bahia State) and Estrela de Alagoas (Alagoas State) localities were previously described in Brazil. In both cases, a Burst-type song and Cembrene-1 population, coexist in sympatry with a different Pulse-type song sibling [33]. Unlike Sobral, in Jaiba the

**Table 4. Estimates of parameters from IM analysis of sympatric and allopatric comparisons.**

Comparison	Estimates of parameters
Sympatric	$\theta_{S1S}$ 2.9127 (2.2693–3.6479) $N_{S1S}$ 4,094,250 (3,189,942–5,127,705)
	$\theta_{S2S}$ 2.056 (1.5213–2.6660) $N_{S2S}$ 2,890,075 (2,138,459–3,747,539)
	$\theta_A$ 0.8661 (0.2109–1.5138) $N_A$ 1,217,493 (296,387–2,127,881)
	$t$ 0.9463 (0.6388–1.4613) $t$ 532,049 (359,151–821,619)
	$m_{S1S}$ 0.1375 (0.9125–2.0938) $2Nm$ 0.2002 (0.0142–0.7524)
	$m_{S2S}$ 1.425 (0.0125–0.4125) $2Nm$ 1.4649 (0.6941–2.7910)
Allopatric	$\theta_{Pan}$ 1.4355 (1.0803–1.8506) $N_{Pan}$ 2,017,852 (1,518,555–2,601,349)
	$\theta_{Lap}$ 1.4134 (1.0549–1.7923) $N_{Lap}$ 1,986,716 (1,482,851–2,519,328)
	$\theta_A$ 1.4901 (0.6760–2.4426) $N_A$ 2,094,637 (950,169–3,433,476)
	$t$ 0.8863 (0.6138–1.3863) $t$ 498,313 (345,094–779,449)
	$m_{Pan}$ 0.2125 (0.0625–0.4375) $2Nm$ 0.1525 (0.0338–0.4048)
	$m_{Lap}$ 0.0625 (0.0125–0.2625) $2Nm$ 0.0442 (0.0066–0.2352)

$\theta$ , the estimated population mutation rates;  $t$ , the estimated number of mutations since the time of splitting;  $m$ , the estimated migration rates;  $N$ , the estimated effective population sizes;  $t$ , the estimated time in years at which the ancestral population split into the sampled populations;  $2Nm$ , the estimated effective number of migrant gene copies per generation in each direction. The 90% lowest and highest posterior density intervals for each parameter are shown in brackets.  $S1S$ , Sobral 1S;  $S2S$ , Sobral 2S;  $Pan$ , Pancas;  $Lap$ , Lapinha;  $A$ , ancestral species.

doi:10.1371/journal.pntd.0002495.t004

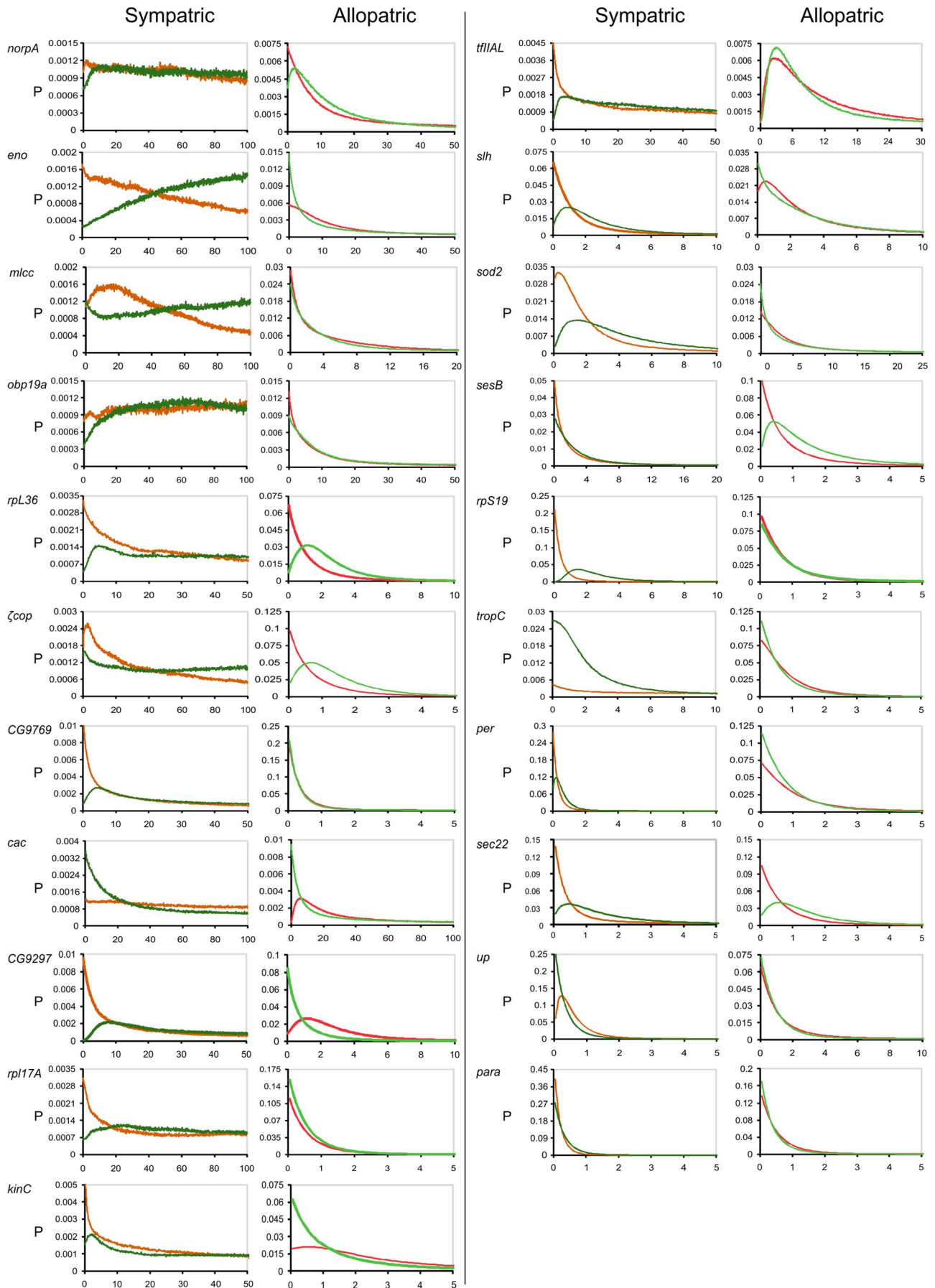


**Figure 6. Distributions of marginal posterior probabilities of model parameters of sympatric and allopatric species.** Probability densities (P) are shown in curves, which present single narrow peaks for six parameters of the isolation with migration model: (A) theta ( $\theta = 4N\mu$ ), (B) divergence time between species ( $t = t\mu$ ) and (C) migration rate ( $m = m/\mu$ ). doi:10.1371/journal.pntd.0002495.g006

two siblings differ in the type of diterpene isomers that they carry and in Estrela de Alagoas the siblings share the same type of pheromone [83]. Future work might reveal semipermeable species boundaries which would not necessarily involve the same genes, intensity and/or direction of gene flow.

Nevertheless, our results indicate that past gene flow has affected several areas of the *L. longipalpis s.l.* genome. In some cases, genomic regions with suppressed or reduced recombination, as a result of chromosomal rearrangements such as inversions, might have less introgression than colinear regions [84,85]. Genes located in such regions might be involved in adaptive differences

or reproductive divergence between siblings, and therefore be filtered by selection, while genes that are not linked to such regions might introgress more readily [8,66]. In *L. longipalpis s.l.*, putative chromosomal inversions between some siblings have been identified [86], but future mapping experiments are needed to reveal which genome regions of this species complex have undergone more introgression than others. Furthermore, the knowledge of introgression patterns throughout the genome is important to understand whether loci related to vectorial capacity can influence the transmission dynamics of *Leishmania* parasites by the different *L. longipalpis* sibling species. This will be particularly



**Figure 7. Distribution of marginal posterior probabilities of the migration rates at each locus.** Probability densities (P) of migration estimates for 21 nuclear loci for two comparisons, sympatric - Sobral 15 (orange) vs. Sobral 25 (green) - and allopatric - Lapinha (red) vs. Pancas (light green).

doi:10.1371/journal.pntd.0002495.g007

interesting under an epidemiological point of view considering the potential for increased introgression caused by human-made changes to the environment [87] and the current trend of visceral leishmaniasis urbanization in Brazil [88].

## Supporting Information

**Table S1** Chromosome positions of the 21 loci in *D. melanogaster* and *A. gambiae* and their biological functions and processes. (DOC)

**Table S2** List of primers used for 18 new markers of the 21 loci used in the multilocus analysis. (DOC)

**Table S3** Domains of the 21 loci and description of the non-recombining blocks constructed for the IM analyzes. (DOC)

**Table S4** Additional neutrality tests for 21 locus in four *L. longipalpis s.l.* populations of Brazil. (DOC)

**Table S5** Differentiation among siblings of *L. longipalpis* complex from Brazil. (DOC)

## References

1. Marie Curie SPECIATION Network, Butlin R, Debelle A, Kerth C, Snook RR, et al. (2012) What do we need to know about speciation? Trends Ecol Evol 27: 27–39.
2. Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. Science 296: 707–711.
3. Nei M, Nozawa M (2011) Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. Genome Biol Evol 3: 812–829.
4. Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. Annu Rev Ecol Syst 41: 215–230.
5. Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. Curr Opin Genet Dev 16: 592–596.
6. Wang RL, Wakeley J, Hey J (1997) Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147: 1091–1106.
7. Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, et al. (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics 156: 1913–1931.
8. Machado CA, Hey J (2003) The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc Biol Sci 270: 1193–1202.
9. Baack EJ, Rieseberg LH (2007) A genomic view of introgression and hybrid speciation. Curr Opin Genet Dev 17: 513–518.
10. Hey J (2001) The mind of the species problem. Trends Ecol Evol 16: 326–329.
11. Hey J, Machado CA (2003) The study of structured populations—new hope for a difficult and divided science. Nat Rev Genet 4: 535–543.
12. Hey J, Nielsen R (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.
13. Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. Mol Biol Evol 22: 297–307.
14. Bull V, Beltran M, Jiggins CD, McMillan WO, Bermingham E, et al. (2006) Polyphyly and gene flow between non-sibling *Heliconius* species. BMC Biol 4: 11.
15. Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). Genetics 173: 919–933.
16. Niemiller ML, Fitzpatrick BM, Miller BT (2008) Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Cyprinophilus*) inferred from gene genealogies. Mol Ecol 17: 2258–2275.
17. Nosil P (2008) Speciation with gene flow could be common. Mol Ecol 17: 2103–2106.
18. Salazar C, Jiggins CD, Taylor JE, Kronforst MR, Linares M (2008) Gene flow and the genealogical history of *Heliconius heurippa*. BMC Evol Biol 8: 132.
19. Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (2009) Speciation in the deep sea: multi-locus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. PLoS One 4: e6485.
20. della Torre A, Fanello C, Akogbeto M, Dossou-yovo J, Favia G, et al. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. Insect Mol Biol 10: 9–18.
21. Pardo-Diaz C, Salazar C, Baxter SW, Merot C, Figueiredo-Ready W, et al. (2012) Adaptive Introgression across Species Boundaries in *Heliconius* Butterflies. PLoS Genet 8: e1002752.
22. Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, et al. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proc Natl Acad Sci U S A 100: 10818–10823.
23. Donnelly MJ, Pinto J, Girod R, Besansky NJ, Lehmann T (2004) Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. Heredity 92: 61–68.
24. Mazzoni CJ, Araki AS, Ferreira GE, Azevedo RV, Barbuji G, et al. (2008) Multilocus analysis of introgression between two sand fly vectors of leishmaniasis. BMC Evol Biol 8: 141.
25. Djogbenou L, Chandre F, Berthomieu A, Dabire R, Koffi A, et al. (2008) Evidence of introgression of the *ace-1(R)* mutation and of the *ace-1* duplication in West African *Anopheles gambiae s.s.* PLoS One 3: e2172.
26. White BJ, Lawniczak MK, Cheng C, Coulibaly MB, Wilson MD, et al. (2011) Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. Proc Natl Acad Sci U S A 108: 244–249.
27. Salomon OD, Basmajdian Y, Fernandez MS, Santini MS (2011) *Lutzomyia longipalpis* in Uruguay: the first report and the potential of visceral leishmaniasis transmission. Mem Inst Oswaldo Cruz 106: 381–382.
28. Watts PC, Hamilton JG, Ward RD, Noyes HA, Souza NA, et al. (2005) Male sex pheromones and the phylogeographic structure of the *Lutzomyia longipalpis* species complex (Diptera: Psychodidae) from Brazil and Venezuela. Am J Trop Med Hyg 73: 734–743.
29. Ward RD, Ribeiro AL, Ready PD, Murtagh A (1983) Reproductive isolation between different forms of *Lutzomyia longipalpis* (Lutz & Neiva) (Diptera: Psychodidae), the vector of *Leishmania donovani chagasi* Cunha & Chagas and its significance to kala-azar distribution in South America. Mem Inst Oswaldo Cruz 78: 269–280.
30. Lanzaro GC, Ostrowska K, Herrero MV, Lawyer PG, Warburg A (1993) *Lutzomyia longipalpis* is a species complex: genetic divergence and interspecific hybrid sterility among three populations. Am J Trop Med Hyg 48: 839–847.

**Table S6** The average number of nucleotide substitutions per site among siblings of *L. longipalpis* species complex of Brazil, Dxy (Nei 1987). (DOC)

**Table S7** Polymorphism summary of the non-recombining blocks for the 21 loci in four sibling species of the *L. longipalpis* complex from Brazil. (DOC)

**Table S8** Differentiation in the non-recombining blocks between sympatric and allopatric species of the *L. longipalpis* complex from Brazil. (DOC)

## Acknowledgments

We would like to thank Robson Costa da Silva for technical assistance, Elisa Cupolillo for comments on an earlier version of the manuscript and Renata Azevedo for helpful advice and suggestions. We dedicate this manuscript to Alexandre Afranio Peixoto *in memoriam*.

## Author Contributions

Conceived and designed the experiments: ASA GEMF AAP. Performed the experiments: ASA GEMF. Analyzed the data: ASA GEMF CJM AAP. Contributed reagents/materials/analysis tools: GEMF NAS RCM RVB. Wrote the paper: ASA GEMF AAP.

31. Arrivillaga J, Mutebi JP, Pinango H, Norris D, Alexander B, et al. (2003) The taxonomic status of genetically divergent populations of *Lutzomyia longipalpis* (Diptera: Psychodidae) based on the distribution of mitochondrial and isozyme variation. *J Med Entomol* 40: 615–627.
32. Bauzer LG, Souza NA, Maingon RD, Peixoto AA (2007) *Lutzomyia longipalpis* in Brazil: a complex or a single species? A mini-review. *Mem Inst Oswaldo Cruz* 102: 1–12.
33. Araki AS, Vigoder FM, Bauzer LG, Ferreira GE, Souza NA, et al. (2009) Molecular and behavioral differentiation among Brazilian populations of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *PLoS Negl Trop Dis* 3: e365.
34. Maingon RD, Ward RD, Hamilton JG, Bauzer LG, Peixoto AA (2008) The *Lutzomyia longipalpis* species complex: does population sub-structure matter to *Leishmania* transmission? *Trends Parasitol* 24: 12–17.
35. Ward RD, Phillips A, Burnet B, Marcondes CB (1988) The *Lutzomyia longipalpis* complex: reproduction and distribution. *Biosystematics of Haematophagus Insects*. Oxford: Clarendon Press. pp. 257–269.
36. Lins RMMA, Souza NA, Brazil RP, Maingon RDC, Peixoto AA (2012) Fixed differences in the *paralytic* gene define two lineages within the *Lutzomyia longipalpis* complex producing different types of courtship songs. *PLoS One* 7(9): e44323.
37. Coyne JA, Price TD (2000) Little evidence for sympatric speciation in island birds. *Evolution* 54: 2166–2171.
38. Llopart A, Lachaise D, Coyne JA (2005) Multilocus analysis of introgression between two sympatric sister species of *Drosophila*: *Drosophila yakuba* and *D. santomea*. *Genetics* 171: 197–210.
39. Souza NA, Vigoder FM, Araki AS, Ward RD, Kyriacou CP, et al. (2004) Analysis of the copulatory courtship songs of *Lutzomyia longipalpis* in six populations from Brazil. *J Med Entomol* 41: 906–913.
40. Souza NA, Andrade-Coelho CA, Vigoder FM, Ward RD, Peixoto AA (2008) Reproductive isolation between sympatric and allopatric Brazilian populations of *Lutzomyia longipalpis* s.l. (Diptera: Psychodidae). *Mem Inst Oswaldo Cruz* 103: 216–219.
41. Maingon RD, Ward RD, Hamilton JG, Noyes HA, Souza N, et al. (2003) Genetic identification of two sibling species of *Lutzomyia longipalpis* (Diptera: Psychodidae) that produce distinct male sex pheromones in Sobral, Ceara State, Brazil. *Mol Ecol* 12: 1879–1894.
42. Bauzer LG, Gesto JS, Souza NA, Ward RD, Hamilton JG, et al. (2002) Molecular divergence in the *period* gene between two putative sympatric species of the *Lutzomyia longipalpis* complex. *Mol Biol Evol* 19: 1624–1627.
43. Bottechia M, Oliveira SG, Bauzer LG, Souza NA, Ward RD, et al. (2004) Genetic divergence in the *cacophony IV/56* intron among five Brazilian populations of *Lutzomyia longipalpis*. *J Mol Evol* 58: 754–761.
44. Lins RM, Souza NA, Peixoto AA (2008) Genetic divergence between two sympatric species of the *Lutzomyia longipalpis* complex in the *paralytic* gene, a locus associated with insecticide resistance and lovesong production. *Mem Inst Oswaldo Cruz* 103: 736–740.
45. Young DG, Duncan MA (1994) Guide to the identification and geographic distribution of *Lutzomyia* sandflies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Mem Am Entomol Inst* 54: 881.
46. Bauzer LG, Souza NA, Ward RD, Kyriacou CP, Peixoto AA (2002) The *period* gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. *Insect Mol Biol* 11: 315–323.
47. Sambrook J, Russell D (2001) *Molecular Cloning: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press. 2100 p.
48. Otto TD, Vasconcelos EA, Gomes LH, Moreira AS, Degraive WM, et al. (2008) ChromaPipe: a pipeline for analysis, quality control and management for a DNA sequencing facility. *Genet Mol Res* 7: 861–871.
49. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
50. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739.
51. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
52. Filatov DA (2002) ProSeq: A software for preparation and evolutionary analysis of DNA sequence data sets. *Mol Ecol Notes* 2: 621–624.
53. Hudson RR, Kreitman M, Aguadé M (1987) A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* 116: 153–159.
54. Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132: 619–633.
55. Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9: 1657–1659.
56. Woerner AE, Cox MP, Hammer MF (2007) Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23: 1851–1853.
57. Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* 61: 893–903.
58. Li W-H (1997) *Molecular Evolution*; Associates S, editor. Sunderland, Massachusetts: Sinauer Associates.
59. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
60. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
61. Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19: 2092–2100.
62. Michel AP, Ingrassi MJ, Schemerhorn BJ, Kern M, Le Goff G, et al. (2005) Rangediver population genetic structure of the African malaria vector *Anopheles funestus*. *Mol Ecol* 14: 4235–4248.
63. Joly S, McLenachan PA, Lockhart PJ (2009) A Statistical approach for distinguishing hybridization and incomplete lineage sorting. *Am Nat* 174: E54–E70.
64. Garrigan D, Kingan SB, Pilkington MM, Wilder JA, Cox MP, et al. (2007) Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 177: 2195–2207.
65. Ayala FJ, Coluzzi M (2005) Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci U S A* 102 Suppl 1: 6535–6542.
66. Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 3: e285.
67. Machado CA, Haselkorn TS, Noor MA (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 175: 1289–1306.
68. Dopman EB, Perez L, Bogdanowicz SM, Harrison RG (2005) Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proc Natl Acad Sci U S A* 102: 14706–14711.
69. de Leon LF, Bermingham E, Podos J, Hendry AP (2010) Divergence with gene flow as facilitated by ecological differences: within-island variation in Darwin's finches. *Philos Trans R Soc Lond B Biol Sci* 365: 1041–1052.
70. Grant PR, Grant BR, Petren K (2005) Hybridization in the recent past. *Am Nat* 166: 56–67.
71. Hey J, Pinho C (2012) Population genetics and objectivity in species diagnosis. *Evolution* 66: 1413–1429.
72. Grant BR, Grant PR (2008) Fission and fusion of Darwin's finches populations. *Philos Trans R Soc Lond B Biol Sci* 363: 2821–2829.
73. Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. *Evolution* 62: 1908–1920.
74. Beysard M, Perrin N, Jaarola M, Heckel G, Vogel P (2012) Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (*Microtus agrestis*). *J Evol Biol* 25: 400–408.
75. Machado CA, Kliman RM, Markert JA, Hey J (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.
76. Wang-Sattler R, Blandin S, Ning Y, Blass C, Dolo G, et al. (2007) Mosaic genome architecture of the *Anopheles gambiae* species complex. *PLoS One* 2: e1249.
77. Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol Ecol Resour* 10: 806–820.
78. Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. *Evolution* 56: 1557–1565.
79. Gleason JM, Nuzhdin SV, Ritchie MG (2002) Quantitative trait loci affecting a courtship signal in *Drosophila melanogaster*. *Heredity* 89: 1–6.
80. Servedio MR (2004) The what and why of research on reinforcement. *PLoS Biol* 2: e420.
81. Rivas GB, Souza NA, Peixoto AA (2008) Analysis of the activity patterns of two sympatric sandfly siblings of the *Lutzomyia longipalpis* species complex from Brazil. *Med Vet Entomol* 22: 288–290.
82. Souza NA, Andrade-Coelho CA, Silva VC, Ward RD, Peixoto AA (2009) Life cycle differences among Brazilian sandflies of the *Lutzomyia longipalpis* sibling species complex. *Med Vet Entomol* 23: 287–292.
83. Hamilton JGC, Maingon R, Alexander B, Ward R, Brazil RP (2005) Analysis of the sex pheromone extract of individual male *Lutzomyia longipalpis* sandflies from six regions in Brazil. *Med Vet Entomol* 19: 480–488.
84. Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713–727.
85. Slotman MA, Reimer IJ, Thiemann T, Dolo G, Fondjo E, et al. (2006) Reduced recombination rate and genetic differentiation between the M and S forms of *Anopheles gambiae* s.s. *Genetics* 174: 2081–2093.
86. Yin H, Mutebi JP, Marriotti S, Lanzaro GC (1999) Metaphase karyotypes and G-banding in sandflies of the *Lutzomyia longipalpis* complex. *Med Vet Entomol* 13: 72–77.
87. Crispo E, Moore JS, Lee-Yaw JA, Gray SM, Haller BC (2011) Broken barriers: human-induced changes to gene flow and introgression in animals: an examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *Bioessays* 33: 508–518.
88. Harshay MO, Olliaro PL, Costa DL, Costa CH (2011) Urban parasitology: visceral leishmaniasis in Brazil. *Trends Parasitol* 27: 403–409.