Biology Faculty Publications   Biology

12-28-2020

Molecular Signatures of Sexual Communication in the Phlebotomine Sand Flies

Paul V. Hickner
University of Kentucky

Nataliya Timoshevskaya
University of Kentucky, nti225@uky.edu

Ronald J. Nowling
Milwaukee School of Engineering

Frédéric Labbé
University of Notre Dame

Andrew D. Nguyen
University of Notre Dame

See next page for additional authors

Follow this and additional works at: https://uknowledge.uky.edu/biology_facpub

Part of the Cell Biology Commons, and the Entomology Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Molecular Signatures of Sexual Communication in the Phlebotomine Sand Flies

Notes/Citation Information

Authors

Paul V. Hickner, Nataliya Timoshevskaya, Ronald J. Nowling, Frédéric Labbé, Andrew D. Nguyen, Mary Ann McDowell, Carolina N. Spiegel, and Zainulabeuddin Syed

Molecular signatures of sexual communication in the phlebotomine sand flies

Paul V. Hickner[1], **Nataliya Timoshevskaya**[2], **Ronald J. Nowling**[3], **Frédéric Labbé**[4], **Andrew D. Nguyen**[4], **Mary Ann McDowell**[4], **Carolina N. Spiegel**[5], **Zainulabeuddin Syed**[1] *

1 Department of Entomology, University of Kentucky, Lexington, Kentucky, United States of America, **2** Department of Biology, University of Kentucky, Lexington, Kentucky, United States of America, **3** Electrical Engineering and Computer Science, Milwaukee School of Engineering, Milwaukee, Wisconsin, United States of America, **4** Department of Biological Sciences, Eck Institute for Global Health, University of Notre Dame, Notre Dame, Indiana, United States of America, **5** Department of Molecular and Cell Biology, Institute of Biology, Fluminense Federal University, Niterói, Rio de Janeiro, Brazil

* zsy224@uky.edu

Abstract

Phlebotomine sand flies employ an elaborate system of pheromone communication wherein males produce pheromones that attract other males to leks (thus acting as an aggregation pheromone) and females to the lekking males (sex pheromone). In addition, the type of pheromone produced varies among populations. Despite the numerous studies on sand fly chemical communication, little is known of their chemosensory genome. Chemoreceptors interact with chemicals in an organism's environment to elicit essential behaviors such as the identification of suitable mates and food sources. Thus, they play important roles during adaptation and speciation. Major chemoreceptor gene families, odorant receptors (ORs), gustatory receptors (GRs) and ionotropic receptors (IRs) together detect and discriminate the chemical landscape. Here, we annotated the chemoreceptor repertoire in the genomes of *Lutzomyia longipalpis* and *Phlebotomus papatasi*, major phlebotomine vectors in the New World and Old World, respectively. Comparison with other sequenced Diptera revealed a large and unique expansion where over 80% of the ~140 ORs belong to a single, taxonomically restricted clade. We next conducted a comprehensive analysis of the chemoreceptors in 63 *L. longipalpis* individuals from four different locations in Brazil representing allopatric and sympatric populations and three sex-aggregation pheromone types (chemotypes). Population structure based on single nucleotide polymorphisms (SNPs) and gene copy number in the chemoreceptors corresponded with their putative chemotypes, and corroborate previous studies that identified multiple populations. Our work provides genomic insights into the underlying behavioral evolution of sexual communication in the *L. longipalpis* species complex in Brazil, and highlights the importance of accounting for the ongoing speciation in central and South American *Lutzomyia* that could have important implications for vectorial capacity.

Author summary

Phlebotomine sand flies are the primary vectors of *Leishmania* parasites, the causative agents of cutaneous and visceral leishmaniasis. Due to the lack of vaccines, control of leishmaniasis relies upon reducing human exposure to sand flies. Sand flies produce sex-aggregation pheromones that elicit robust olfactory behaviors, and the molecular targets for pheromone detection remain unknown. We identified chemoreceptors in the genomes of *L. longipalpis* and *P. papatasi*, and used these gene models to explore chemoreceptor evolution in 63 *L. longipalpis* individuals representing different pheromone types. These analyses identified genomic loci underlying chemosensory behavior in sand flies. This paves the way for understanding the sand fly species diversity at the molecular level, and functional characterization of these candidate genes will isolate and identify chemostimuli that can directly be tested as potential attractants for odor-baited traps.

Introduction

Globally, vector-borne diseases account for more than 17% of all infectious diseases every year. One such disease, leishmaniasis, is endemic in 98 countries with an estimated 700,000 to 1 million new cases, leading to 26–65,000 deaths each year[1]. Leishmaniasis is a group of vector-borne diseases caused by protozoan parasites in the genus *Leishmania* and is considered among the most important neglected tropical diseases [2]. Sand flies (Diptera:Psychodoidae) of the genus *Phlebotomus* in the Old World, and *Lutzomyia* in the New World are major vectors of these parasites. *L. longipalpis* is found in a wide but discontinuous geographical distribution from Mexico to Argentina, where they inhabit diverse ecological environments, while *Phlebotomus* has a wide geographical distribution (from southern Europe, northern Africa, the Middle East, and India) and inhabits a variety of ecological niches from tropical climates to arid desert [3]. It is widely accepted that *L. longipalpis* is a species complex, and genetic variability has potential implications for disease transmission [4].

Sand flies of both genera display robust olfactory behaviors to locate suitable hosts, oviposition sites and mates [5]. In contrast to most disease vectors, which do not employ long-range chemical communication to locate potential mates [6], sand flies of the *Lutzomyia* species complex employ an elaborate pheromone communication system [5,7], wherein males produce pheromone(s) that attract conspecific males to courtship aggregations (leks), and attract females to the lekking males. In *L. longipalpis*, these sex-aggregation pheromones are produced in tergal glands that appear as pale patches or "spots" on the abdomen [7]. Historically, the number of spots—one spot (1S) or two spots (2S)—served as a potential phenotypic marker for the cryptic species complex in *L. longipalpis* [7] even though they are increasingly being questioned as reliable markers. The first evidence of the existence of the *L. longipalpis* species complex was obtained in Brazil [4], and genetic variability in sand flies with potential implication in leishmaniasis has long been emphasized [8]. Different sex-aggregation pheromones have been described from male *L. longipalpis* from Brazil: *S*-9-methylgermacrene-B (9MGB), (1*S*,3*S*,7*R*) 3-methyl-α-himachalene (3MαH), two cembrene isomers (Cemb-1 and Cemb-2), and a novel chemotype based on variation in the quantity of 9MGB produced (9MGB+) [9,10]. Cemb-1 was recently reclassified as a novel diterpene and named sobralene [11]. In addition to pheromone communication, *L. longipalpis* males produce a song during copulation—an acoustic signal generated by vibrating their wings—that varies among populations and that can be broadly categorized as "pulse-type" or "burst-type". Variation in copulation song and sex-aggregation pheromone, together with subsequent crossing studies and genetic

differentiation, have provided compelling evidence for *L. longipalpis* comprising a species complex [7,9].

Signaling and reception evolve in synchrony and shape the resulting behavior. Broadly defined as 'sensory drive', especially in the context of environmental conditions [12], this process is prominent in chemosensation across phyla [13]. Genomic changes underlie behavioral evolution; thus, studies of sensory systems and their genetic correlates provide insights into the patterns of ecological evolution [14]. This is particularly evident in arthropod vectors that transmit various life-threatening diseases [15–17]. For example, multiple anopheline species altered their behavioral repertoire following sustained use of bed nets and indoor spraying [18]; ancestral African aedenine populations evolved to be human commensals facilitated by behavioral and genetic changes [19,20] such as house-entering behavior [21], and enhanced human preference [19] facilitated by chemoreceptor gene families [19,22,23]. Since the communication system–comprising songs and pheromones–is well defined in the *L. longipalpis* complex, we undertook a study to explore the genetic basis of the pheromone communication component. This work further offers a framework to study the molecular basis of the acoustic component of the communication. Though the communication system remains to be fully explored, recent work demonstrated the existence of at least two distinct acoustic signals during copulation in the Old World *P. argentipus* populations [24]. Here we report a molecular evolutionary analysis of the sand fly chemoreceptor genomes comprising the ORs, GRs, and IRs, which are among the largest gene families, and together define the reception and perception of odors associated with hosts, mates, and oviposition sites in insects [25,26]. We annotated the chemoreceptors in the whole genome assemblies of two phlebotomine sand flies, *L. longipalpis* and *P. papatasi*, and conducted comparative analyses with other Diptera. In addition, we investigated variation in the chemoreceptor genomes of 63 *L. longipalpis* individuals collected from four locations in Brazil, representing sympatric and allopatric populations and three different sex-aggregation pheromones. Investigations into the molecular signatures of genetic variation have traditionally focused on single nucleotide polymorphisms (SNPs) [27], but recently there has been an emphasis on gene copy number variation (CNV) as an additional source of genomic variation in the insect chemoreceptors [26, 28]. Our analysis of SNPs and CNV provides novel insights into the evolution of chemosensation, thus providing a framework for future studies on the molecular basis of chemical communication in the phlebotomine sand flies.

Methods

Annotation of the chemoreceptor repertoires

Genes were manually annotated as described previously [29]. Briefly, genomic loci encoding odorant receptors (ORs), gustatory receptors (GRs), and ionotropic receptors (IRs) were identified by tBLASTn analysis of the *L. longipalpis* (LlonJ1) and *P. papatasi* (PpapI1) genome assemblies in VectorBase [30] using *Anopheles gambiae* and *Drosophila melanogaster* peptides as queries. All BLAST analyses was conducted using the BLOSM62 scoring matrix and a maximum E value of 10000, with a tBLASTn word size of 3 and a BLASTn word size of 11. *An. gambiae* gene models were downloaded from VectorBase (AgamP4.9), and *D. melanogaster* gene models were downloaded from FlyBase (release FB2017_05). An exhaustive screen of the *L. longipalpis* and *P. papatasi* genome scaffolds was performed by reciprocal BLAST analyses using the sand fly chemoreceptor gene models. Genes were prefixed with either Llon (*L. longipalpis*) or Ppap (*P. papatasi*). The ORs and GRs were numbered arbitrarily with the following exceptions: the putative CO_2 receptors were numbered Gr1 and Gr2, followed by the sugar receptors; and several, but not all, 1:1 orthologs between *L. longipalpis* and *P. papatasi* were

numbered the same (e.g. Gr1, Gr2, Gr13, Gr26). Most IRs were named based on homology to *D. melanogaster*, while Ir101 and Ir102 were named based on their homology to Ir101 in *An. gambiae*.

Genes with gaps, premature stop codons, and indels that were "fixed" (when possible) by BLASTn analysis of the SRAs were suffixed with "_fix". Indels and premature stop codons that could not be fixed or were confirmed with SRAs were considered pseudogenes and suffixed with "P". Genes spanning two or more different scaffolds were suffixed with "_join", and genes that were on the same scaffold but on different strands were suffixed with "_strand". Genes annotated from the *de novo* assemblies are suffixed with "_denovo". Partial gene models encoding ≤330 amino acids were omitted from the dataset.

Phylogenetic analysis

A species phylogeny was estimated to illustrate the evolutionary relationships among the Diptera used for comparative analysis of chemoreceptor repertoire size. The OrthoFinder v2.3.3 program was used for orthologous group selection among peptides in *An. gambiae*, *Aedes aegypti*, *Culex quinquefasciatus*, *P. papatasi*, *L. longipalpis*, *Mayetiola destructor*, *Musca domestica*, *D. melanogaster*, *Glossina morsitans*, and the outgroup *Bombyx mori* [31]. The species tree was estimated with OrthoFinder using the FastME distance-based program [32].

OR, GR and IR gene trees were estimated by first conducting multiple sequence alignments of *An. gambiae*, *M. destructor*, *P. papatasi*, *L. longipalpis*, and *D. melanogaster* peptides [33–37] using Muscle v3.8.31 [38]. The multiple sequence alignments were trimmed using the *automated1* option in Trimal v1.4 [39]. Maximum likelihood trees were inferred using the JTT model of protein substitution in RAxML v.8.2.4, which was chosen based on *protgammaauto* model selection [40]. Branch support was estimated using 500 bootstrap replications. The OR, GR, and IR trees were rooted with *Orco*, the CO_2 receptor clade, and the *Ir8a*/*Ir25a* clade, respectively.

Analysis of *L. longipalpis* populations in Brazil

The SRA toolkit v2.9.2 [41] was used to download fastq files from NCBI containing paired-end reads (101 bp x 2) for 63 *L. longipalpis* individuals from Jacobina (n = 14), Lapinha Cave (n = 11), Marajó (n = 9), and Sobral (n = 29) Brazil. Accession numbers for SRA downloads are reported in S1 Table. The individuals from Sobral were further grouped based on the number of spots on the abdomen—Sobral with one spot (1S, n = 13) and Sobral with two spots (2S, n = 16). Due to the number of fixes made to the gene models, several redundant genes, extraneous fragments, and six missing genes, we used BEDtools v2.28.0 [42] to hard-mask all chemoreceptor loci in the LlonJ1 genome assembly, then appended the manually curated OR, GR and IR gene regions. This "revised" assembly (Supplemental data, rev_assembly.fa) was used for subsequent analyses. Chemoreceptor gene regions comprised introns, exons, and 300–500 bp flanking regions (when possible). We used BWA mem v0.7.17 [43] to map the reads to the revised genome assembly. PCR duplicates were removed using SAMtools v1.9 [44]. Reads with soft-clipping on both ends (marked with S in CIGAR string) were removed using custom awk scripts. Single nucleotide polymorphisms (SNPs) were called using BCFtools [44]. Sites with a PHRED score <30 and sequencing depth below 0.5 and above 1.5 modal coverage for that individual were omitted. For example, sites below 25X and above 75X were omitted in an individual with a 50X modal coverage.

Phylogenetic relationships among the field isolates from Brazil were estimated with 18,254 SNPs in the exons of all 100 single-copy orthologs using the neighbor joining method in TASSEL v5.2.57 [45]. The principal component analysis (PCA) approach in PCadapt [46] was used

to identify chemoreceptors associated with differences in chemotype and/or population structure. We limited the analysis to SNPs (n = 18,254) in the exons of genes that were single-copy in all 63 individuals (n = 100). Preliminary analysis with K = 20 and a minimum minor allele frequency of 0.1 was used to select the appropriate number of K (principal components) for subsequent analyses, which was determined to be K = 5 based on the scree plot (S4A Fig). Component-wise analysis with K = 5 and a minimum minor allele frequency of 0.1 resulted in 6,759 SNPs passing criteria and 502 outlier SNPs after Bonferroni correction (qvalue<0.05) for multiple tests (S4B Fig). Due to the large number of genes with only 1 or 2 outlier SNPs, we highlighted the top three genes with the highest number of SNPs based on SNPs per kb of CDS. The entire list of outlier SNPs and their associated PC can be found in S3 Table.

We used the program ADMIXTURE v1.3 [47] to estimate genetic introgression among populations and chemotypes. ADMIXTURE was run for K 1 through 7 (number of ancestral populations) with 5-fold cross-validation. Each ADMIXTURE analysis was repeated 30 times with different seeds, resulting in a total of 210 runs. To better understand the different solutions reported by ADMIXTURE, for each value of K we compared solutions and produced a major cluster of solutions that give similar results using the online version of CLUMPAK with default settings [48]. We used the ADMIXTURE cross-validation procedure to estimate the number of K [49].

The Tablet software v1.19.05.28 program [50] was used to visualize sequence read alignments, which revealed many apparent gene absences/losses. To estimate the extent of copy number variation in 245 chemoreceptor genes in all 63 individuals (~15,000 genes) we used background normalized sequencing depth to estimate gene copy number (CN). Read depth at each position in the chemoreceptor exons was extracted using SAMtools utility *depth* with option -*a* to take into account bases with zero depth of coverage [44]. The mean read depth for each chemoreceptor gene was normalized to modal depth across all exons as follows:

$$\text{normalized depth (ND)} = \frac{\bar{x} \text{ gene depth}}{\text{modal depth of all genes}} \times 2$$

Gene CN was calculated by rounding normalized depth (ND) to the nearest whole number (CN = $\|$ND$\|$). Modal depth ranged from 35 to 133 (mean 67.8) across individuals. A heatmap with hierarchical clustering (one minus the cosine similarity with complete linkage) using CN of chemoreceptor genes in all 63 individuals was calculated using the Morpheus software program (https://software.broadinstitute.org/morpheus).

PCA of normalized sequencing depth was conducted using the covariance method in SigmaPlot v14.0 (Systat Software, San Jose, CA). The one-way analysis of variance and Holm-Šídák test for post-hoc pairwise analysis was conducted using SigmaPlot v14.0 to test for differences in the mean number of absent (CN = 0), single-copy (CN = 2), and duplicated (CN >2) gene lineages among and between chemotypes. Pairwise V_{ST}, a measure of population differentiation analogous to F_{ST} but based on CNV, was calculated using methods described previously [51]. For pairwise analyses of V_{ST}, populations were combined according to their putative sex-aggregation pheromone (chemotype). Specifically, Marajó, Sobral 2S and Jacobina-B (sobralene); Sobral 1S-A and Lapinha (9MGB); and Sobral 1S-B and Jacobina-A (3MαH).

To evaluate ND as a proxy for CN, we used Megahit [52] to construct a *de novo* genome assembly for each individual with the parameters—min-count 2,—k-min 21,—k-max 141,—k-step 10. We annotated a subset of the chemoreceptors in the *de novo* assemblies using Geneious v6.1.8 (https://www.geneious.com) for BLAST analysis and methods previously described for the reference genomes. Seven genes (Or67, Or115, Or116, Or109, Or137, Or138, and Or139) were annotated or confirmed missing/nonfunctional in the *de novo* assemblies of

all 63 individuals. Genes and alleles that are full-length based on read coverage and/or manual annotation were presumed to be functional and are hereafter referred to as "intact".

Results

The sand fly OR repertoires are relatively large and uniquely expanded among the dipterans analyzed here, with 140 in *L. longipalpis* and 142 in *P. papatasi* (Fig 1A). Over 80% of the sand fly ORs belong to a single, taxonomically restricted lineage with no close relationship to other ORs included in our phylogenetic analysis (Figs 1B and S1). Only 5 lineages were conserved
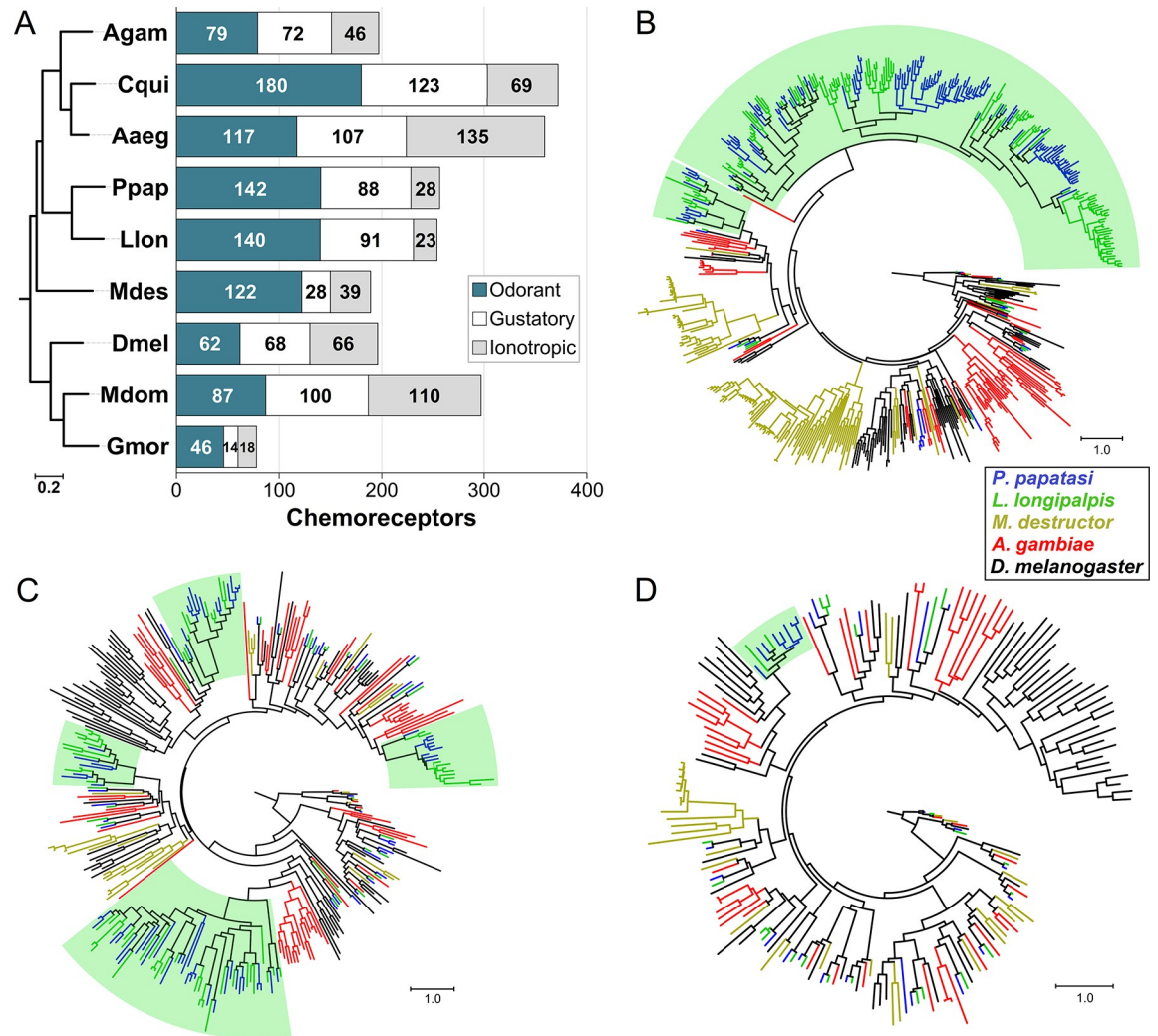

Fig 1. Chemoreceptors in the phlebotomine sand flies. (A) Chemoreceptor repertoire size in the sand flies *L. longipalpis* (Llon) and *P. papatasi* (Ppap) compared with *An. gambiae* (Agam), *C. quinquefasciatus* (Cqui), *Ae. aegypti* (Aaeg), *Ma. destructor* (Mdes), *Mu. domestica* (Mdom), *G. morsitans* (Gmor) and *D. melanogaster* (Dmel). Species tree estimated using OrthoFinder with the multiple sequence alignment option and 252 single-copy orthologs. **(B)** Phylogenetic analysis of chemoreceptors in five Diptera revealed a large taxonomically-restricted clade comprising over 80% of the *L. longipalpis* and *P. papatasi* ORs (highlighted in green). **(C)** Several smaller lineage expansions are evident in the GRs, while **(D)** only one IR lineage (Ir7c) was expanded in the sand flies. Phylogenies were estimated using *L. longipalpis*, *P. papatasi*, *A. gambiae*, *M. destructor* and *D. melanogaster* protein sequences aligned with ClustalX. The JTT model of protein substitution and Maximum Likelihood method in RAxML v.8.2.4 were used for tree estimation. The trees are rooted at the branch leading to Orco, the CO_2 receptors, and the Ir25a and Ir8a clades for ORs, GRs and IRs, respectively. Branch support based on 500 bootstrap replications. Scale bars indicate the number of amino acid substitutions per site.

https://doi.org/10.1371/journal.pntd.0008967.g001

throughout the Diptera suggesting that gene death through pseudogenization and/or deletion has helped shape the sand fly repertoire of ORs (S1 Fig). One notable difference between the two sand fly species is the *AgamOr1-DmelOr46a* lineage, which is lost in *L. longipalpis* but expanded to five copies in *P. papatasi* (S1 Fig). The GR repertoires include 82 genes in *L. longipalpis*, and 77 in *P. papatasi* (Figs 1C and S2). Interestingly, in contrast to most Diptera which have three CO_2 receptors conferring the detection of this important host odor cue (25), only two intact genes (*Gr1* and *Gr2*) were found in the sand fly genomes. Screening of sequence read archives (SRA) for the third CO_2 receptor revealed a highly degraded pseudogene in each sand fly species, thus suggesting the loss of *DmelGr63a/MdesGr3/AgamGr24* ortholog in a common ancestor. The 3rd major chemosensory gene family, the IRs, was smaller (23 and 28 IRs in *L. longipalpis* and *P. papatasi*, respectively) compared to mosquitoes and *D. melanogaster* (Figs 1D and S3). Of the three chemoreceptor families analyzed, the IRs were the least dynamic based on the number and extent of expanded and lost lineages. Coding sequences for ORs, GRs, and IRs in *L. longipalpis* and *P. papatasi* are provided in S2 Table.

New World sand flies in the *L. longipalpis* complex employ an exquisite communication system, whereby males produce pheromones and copulatory songs to identify their mates. Phylogenetic analysis based on SNPs (n = 18,254) in the single-copy chemoreceptor loci (n = 100) in 63 *L. longipalpis* individuals from four sites in Brazil (Fig 2A) revealed distinct clades that grouped individuals and populations broadly based on chemotypes and copulation songs. The most prominent separation was between the 'burst type' clade comprising Marajó, Sobral 2S and 6 Jacobina individuals with undetermined song type, and the remaining into 'pulse type' composed of Lapinha, Sobral 1S, and 8 Jacobina (Fig 2B).

We used PCA to further analyze the population structure and to identify chemoreceptors significantly associated with population differentiation. The first two principal components explained 73.6% of the total variation and grouped individuals into three discrete clusters (Fig 2C) that represented three putative pheromone types (chemotypes): a sobralene cluster composed of all individuals from Marajó and Sobral 2S, and 6 Jacobina (hereafter, JAC-B); a 3MαH cluster comprised of 7 Sobral 1S (hereafter, S1S-B) and 8 Jacobina individuals (hereafter, JAC-A). Interestingly, a third cluster of 9MGB was apparent that included 6 Sobral 1S (hereafter, S1S-A) and all Lapinha individuals. This serendipitous grouping is consistent with the reported findings by Hamilton et al. who classified Lapinha and S1S-A as different chemotypes based on quantitative differences in 9MGB producing individuals [10]. These populations were named 9MGB (Lapinha) or 9MGB+ (S1S-A). With over 47% of the total variation captured, PC1 broadly separates individuals based on the major song types (Fig 2C). It will be exciting to correlate chemotype separation from our analyses with the estimated divergence of burst and song which occurred ca. 0.5–0.7 mya [9,53].

Having found discrete chemotype clusters based on SNPs, we aimed to identify genes contributing to the observed patterns using PCadapt [46]. We identified 502 SNPs, of which 164, 147, 170, 13 and 8 were associated with PC1 through PC5, respectively (S3 Table). Three genes with a significant contribution, defined as 'genes with greatest number outlier SNPs' in *pcadapt*, included *Ir60a*, *Or10* and *Or127* that were involved in the separation of Sobralene from the 9MGB and 3MαH chemotypes (Fig 2D). Or123, Ir68a, and Ir101 were involved in the separation between the 9MGB and 3MαH chemotypes in PC2 (Fig 2D). PC3 through PC5 separated populations, where JAC-A and S1S-B (3MαH) were separated by PC3, Marajó was separated from all others by PC4, and S1S-A and Lapinha were separated by PC5. Intriguingly, four of the 11 genes with the highest number of SNPs underscoring PC1-5 were IRs, which are increasingly being implicated in multimodal signaling. Admixture analysis revealed seven groups that were distinguishable at K = 7, consistent with the PCA and ML tree (Fig 2E). However, a clear modeling choice for the number of K was not indicated by cross-validation error
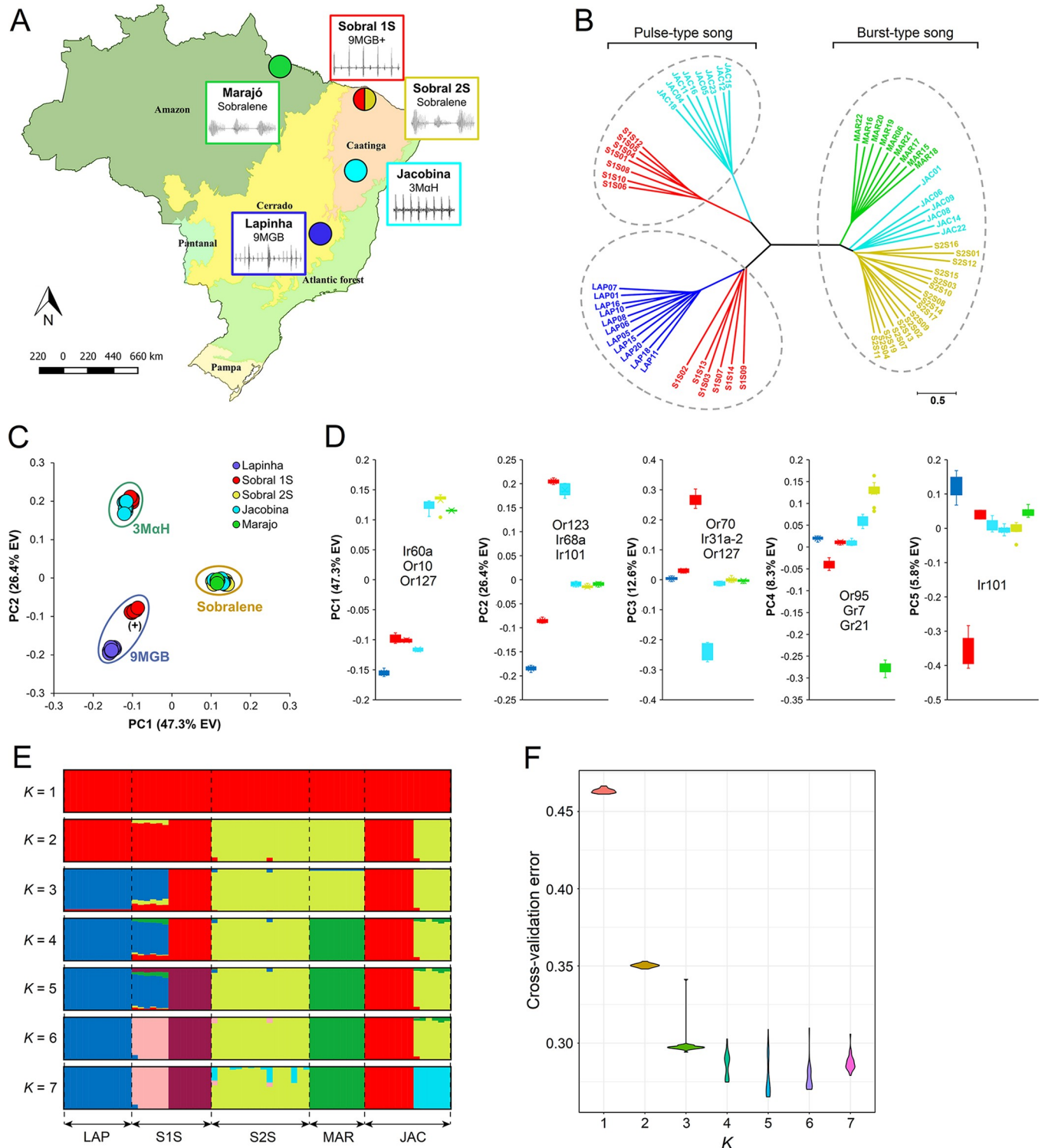
Fig 2. Population structure of 63 *L. longipalpis* from four sites in Brazil based on SNPs in 245 chemoreceptor genes. (A) Geographic distribution, sex-aggregation pheromones (9MGB, 3MαH, sobralene) and copulatory songs based on previous studies of *L. longipalpis* in Brazil (16). (B) Individuals from Sobral (with 2 spots: S2S), Marajo and Lapinha formed discrete clades, while individuals from Sobral (with one spot, S1S) and Jacobina split into two clades each. Unrooted tree estimated using SNPs in the exons of all 100 single-copy orthologs and neighbor-joining method in Tassel v5.2.57. (C) Principal component analysis (PCA) was used to identify loci associated with population structure and conducted using *pcadapt* (explained variance EV). The first two principal components accounted for 73.6% of the total variation and grouped individuals into four clusters. Putative chemotypes were assigned based on previous studies and PCA clustering patterns. (D) Principal components 1–5 and genes with the highest number of SNPs based on component-wise outlier analysis in *pcadapt*.

(E) Ancestry proportions within individual sand flies for ADMIXTURE models from K = 1 to K = 7 ancestral populations. Each vertical bar represents the proportion of ancestry within a single individual, with colors corresponding to ancestral populations. Data are the average of the major q-matrix clusters derived by CLUMPAK analysis. (F) Violin plot of ADMIXTURE cross-validation error for each of 30 replicates for each K value from 1 to 7.

analysis, which suggested K is between 3 and 7 (Fig 2F). Despite this limitation, little introgression was indicated, with most being among the sympatric populations in Sobral and Jacobina (Fig 2E).

While SNP analysis revealed population structuring that is largely consistent with previously described chemotypes, we noticed a surprisingly large number of missing genotypes, which prompted us to investigate potential CNV. Variation in gene copy number has been hypothesized to play an important role in the emergence of adaptive traits within and among populations [54]. Analysis of the read coverage revealed extensive variation in depth of coverage across loci (Fig 3A). For estimation of gene copy number (CN), we calculated a normalized sequencing depth based on the modal coverage of the gene coding regions of all 15,435 loci (63 individuals x 245 genes) (Fig 3B and S4 Table). Normalized depth (ND) ranged from 0 to 6.56 with a distribution having a central tendency ~2 in all three chemotypes (Fig 3C).

To validate our method for estimating gene copy number, we annotated a gene model or confirmed absence for 522 chemoreceptor genes in the *de novo* assemblies (S5 Table). We were unable to annotate any intact genes when CN = 0. Annotation of the duplications was not always possible, most likely due to high sequence similarity causing them to be collapsed during *de novo* assembly. Within-population CNV was apparent at several loci, evidenced by CN of approximately 2, 1, and 0, which we inferred as two intact alleles ($A^{+/+}$), one intact and one degraded allele ($A^{+/-}$), and two degraded alleles ($A^{-/-}$), respectively, which is further supported by sites with heterozygosity in $A^{+/+}$, but not in $A^{+/-}$ (S5A Fig). However, not all genes with CN = 1 represented a combination of an intact and a degraded allele. Some genes were degraded in regions of a gene, with approximately half of the gene intact (S5B Fig).

Of the 245 genes in the reference chemoreceptor genome, only 100 were single-copy (CN = 2) in all 63 individuals, while those that were at least single-copy (CN$\geq$2) in each of the three chemotypes ranged from 141 in Sobralene to 170 in 3MαH, with only 128 shared among all individuals (Fig 3D). Further, the mean number of absent (CN = 0) genes differed among the three chemotypes (p<0.001) with Sobralene individuals having 28.9 ± 4.5 (mean ± SD), followed by 9MGB with 18.9 ± 2.2, and 3MαH with 15.7± 2.3 (Fig 3E). Accordingly, the mean number of single-copy genes (CN = 2) differed among chemotypes (p<0.001), with 3MαH (193.6 ±4.3) followed by 9MGB (183.5 ±3.5) and Sobralene (179 ± 4.6). The mean number of duplicated genes (CN>2) was greatest in individuals from 3MαH (13.5 ± 3.2) compared to 9MGB (11.8 ± 2.7) and Sobralene (10.5 ± 3.0) which did not significantly differ (Fig 3E–3G).

Hierarchical analysis and PCA of CN grouped individuals according to their putative chemotype (Fig 4A and 4B). PC1 and PC2 explained 27.4% and 17.3% of the variation, respectively (Fig 4B). Of the three chemoreceptor gene families, the ORs were the most dynamic with 72.1% displaying CNV, followed by 46.3% of GRs and 21.7% of IRs.

We used pairwise V_{ST} to identify genes exhibiting the greatest CNV between chemotypes. Of the 60 genes with V_{ST} >0.5 in all three pairwise analyses, 43 are ORs, 16 are GRs, and only one is an IR (Fig 5A–5C). In most cases, genes with V_{ST} >0.5 had CN<2 in some individuals from both chemotypes. That is, few were intact (CN$\geq$2) in all individuals of a chemotype, suggesting that they were not under strong purifying selection in either chemotype. However, several genes were intact in one chemotype and largely absent in the other. In the pairwise analysis of 9MGB and Sobralene, Or37, Or43, Or115, and Or139 have CN$\geq$2 in all 9MGB, but CN<2 in some of the Sobralene individuals. In the pairwise analysis between 3MαH and Sobralene, Or115 and Gr37 have a CN$\geq$2 in all 3MαH, but are absent in some Sobralene (Fig
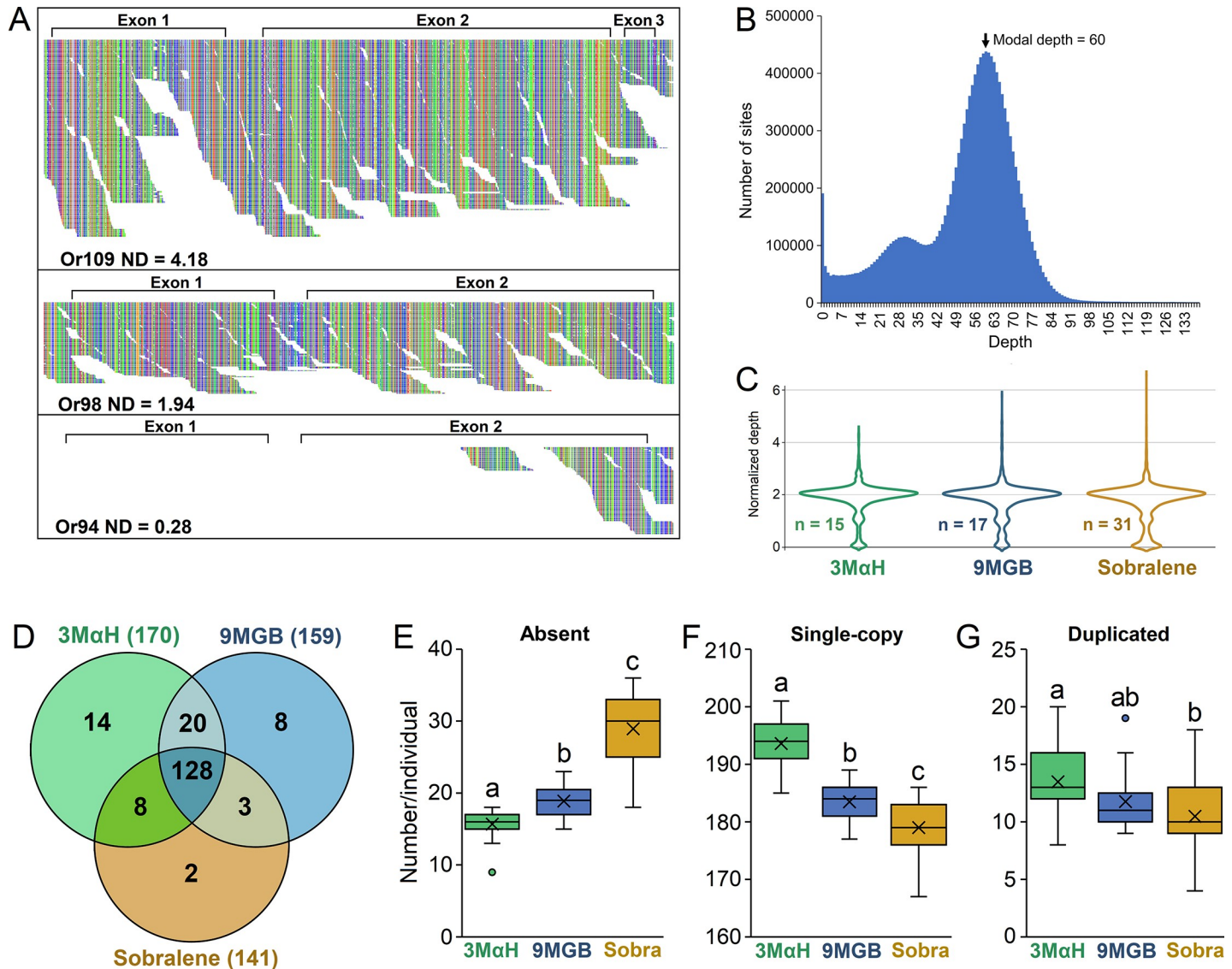
Fig 3. Visualization of reads aligned to chemoreceptor loci revealed variation in coverage that indicated potential copy number variation. (A) For example, JAC01 Or109 had much deeper coverage than Or98, while Or94 had reads mapped only at the end of the second exon. The Tablet software program was used for visualization of the mapped reads. **(B)** To quantify these differences for comprehensive analysis of all 245 chemoreceptor loci, we calculated background-normalized sequencing depth of each gene using the modal depth across the exons in all protein coding genes. **(C)** A central tendency of ~2 is expected for single copy genes with two intact alleles. Normalized depth was rounded to the nearest whole number as a proxy for copy number (CN). **(D)** The number of intact chemoreceptors (CN≥2) in all individuals of a chemotype ranged from 141 (Sobralene) to 170 (3MαH). **(E)** The mean number of absent (CN = 0) genes differed among all chemotypes ($P < 0.001$), with Sobralene individuals having the most and 3MαH individuals having the fewest. **(F)** Accordingly, the number of single-copy (CN = 2) genes differed among all three chemotypes ($P < 0.001$), with 3MαH individuals having the largest number and Sobralene individuals having the fewest. **(G)** The number of duplicated genes (CN>2) differed only between 3MαH and sobralene individuals.

5B). In the analysis between 9MGB and 3MαH, Or23, Or116, and Or118 are intact only in 9MGB, while Or14, Or43 and Gr71 are intact only in 3MαH (Fig 5C). Or115 and Or116 are noteworthy because they are paralogs and were among the highest V_{ST} in the pairwise analyses (Fig 5A–5C). Furthermore, they are in a neighboring clade to the *Mayetiola destructor* and *D. melanogaster* pheromone receptors (S1 Fig). While annotating Or115 and Or116 in the *de novo* assemblies, a third paralog (Or137) was found in all of the individuals but was missing in the reference assembly (S5 Table).
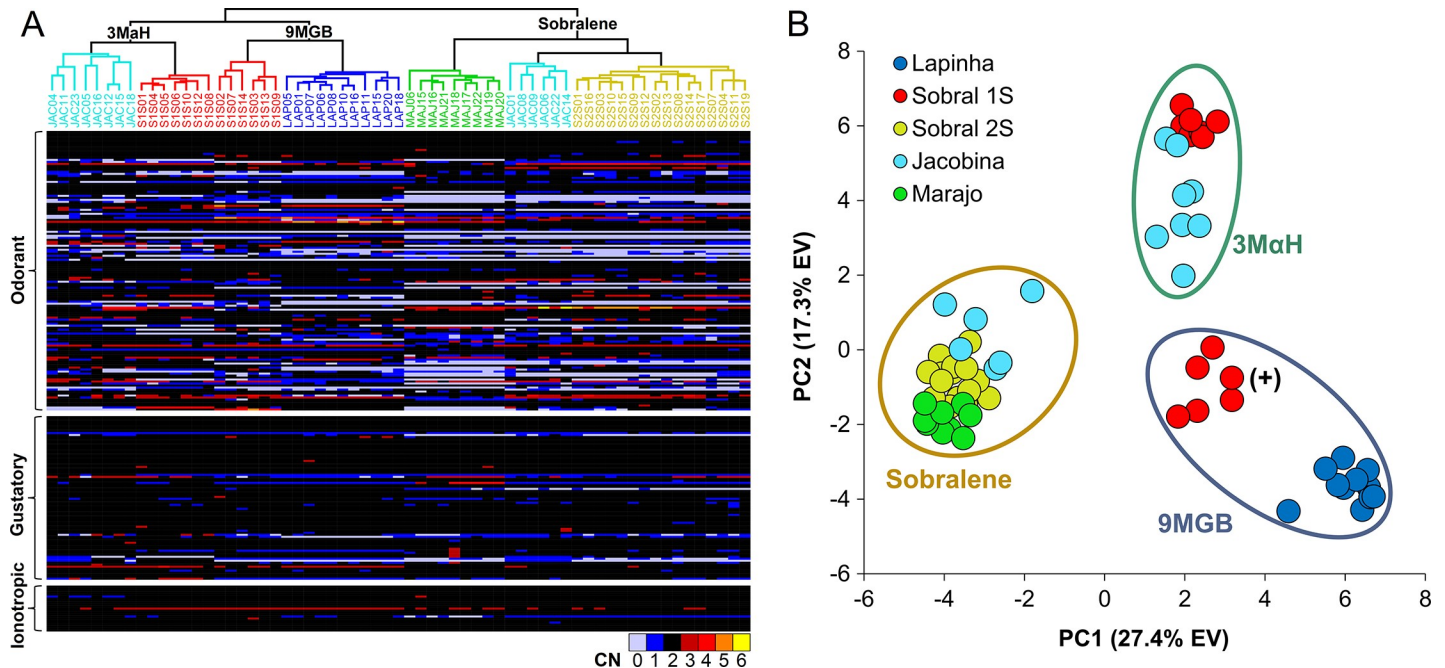
Fig 4. Relationships among 63 *L. longipalpis* from four sites in Brazil (Lapinha Cave, Sobral, Jacobina and Marajó) based on copy number (CN) of 245 chemoreceptor genes. (A) Hierarchical analysis of gene CN clustered individuals according to their putative chemotype as determined previously using SNPs. A heatmap of CN illustrates the large number of genes with CNV among the odorant receptors. **(B)** PCA of gene CN showed a similar pattern, wherein Sobral 2S (2 spots), Marajó and six Jacobina clustered together (Sobralene); seven Sobral 1S (1 spot) and eight Jacobina clustered together (3MαH); and Lapinha Cave and six Sobral 1S (one spot) were in separate clusters (9MGB), which is consistent with the genetic differentiation observed by Hamilton et al. (2005) thus leading them to classify these as different chemotypes (9MGB and 9MGB+) due to the larger quantity of 9MGB produced by males from Lapinha [10].

https://doi.org/10.1371/journal.pntd.0008967.g004

Discussion

Evidence of the genetic variability and its potential implication in vector management strategies was established early in sand flies: biting of *L. longipalpis* females from Costa Rica did not leave long-lasting erythemas that are characteristic in Brazil and Colombia even though parasites were indistinguishable from those locations [8,55]. This clinical pleomorphism in leishmaniasis—manifested as cutaneous or visceral—is still being debated as the outcome of genetic variability of either sand fly, parasite, or the combination thereof. Therefore, understanding the sand fly species diversity at the molecular level contributes to resolving these complex interactions that impact vectorial capacity. The population structure and genetic variability within and among different sand fly populations were found to influence vectorial capacity [56]. Consequently, an array of molecular and biochemical markers have been explored to identify the genotypes underlying phenotypes of interest, such as vector competence, that should be considered during the planning and implementation of integrated control strategies for leishmaniasis [57,58].

Since the initial interactions between an organism and its chemical landscape are mediated by chemoreceptors, the role of these proteins in ecological adaptation is paramount. Genomic divergence in the form of SNPs and CNV is pervasive in chemosensory gene families in both mammals and insects [28,59,60]. Birth-and-death evolution of these gene families often manifests as divergent, taxonomically restricted gene lineages, which have been implicated in the evolution of sociality in the honey bee [61]. The sand flies present an extreme example among the Diptera, with over 80% of their ORs belonging to a single, highly expanded lineage. Aside from a few exceptions, the *P. papatasi* and *L. longipalpis* reference chemoreceptor genomes are
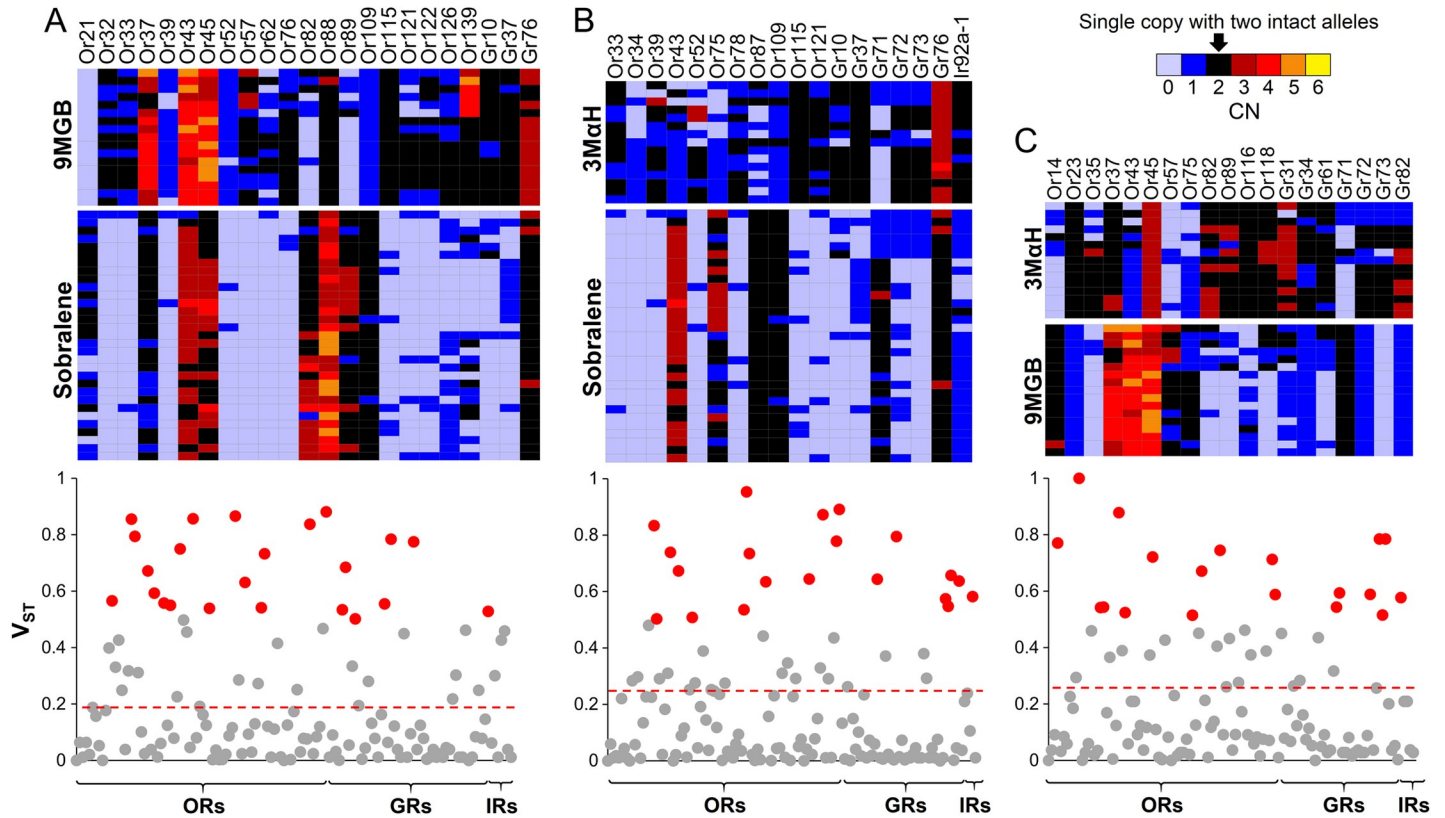
Fig 5. V_ST was used to identify the most differentiated genes based on CNV. Heatmaps of copy number (CN) and Manhattan plots of V_{ST} between **(A)** sobralene and 9MGB, **(B)** sobralene and 3MαH, and **(C)** 9MGB and 3MαH ($V_{ST} > 0.5$ highlighted in red). The dashed lines indicate the threshold for significance (0.99) based on 1,000 permutations. Heatmaps illustrate CN of genes with $V_{ST} > 0.5$. Of the 60 genes with $V_{ST} > 0.5$ in all three pairwise analyses, 43 were ORs, 16 were GRs and only one was an IR.

https://doi.org/10.1371/journal.pntd.0008967.g005

similar in size and content (Fig 1A). Therefore, the amount of CNV among and within chemo-types in Brazil was unexpected. Our results indicate differential CNV among the chemorecep-tors wherein ORs were the most dynamic, followed by the GRs and IRs (Fig 3). Similar observations are noted in pea aphids that were attributed to strong drift and selection [62]. A major role for CNV in adaptative innovation was proposed nearly 50 years ago [54]. Unequal crossover during meiosis is thought to be the primary mechanism leading to CNV in the insect chemoreceptors [26,28]. One of the outcomes of these duplications is the process of neofunc-tionalization, the development of novel function due to relaxed constraints on a paralog [63]. The adaptive radiation model of neofunctionalization predicts that the emergence of a dupli-cated gene produces additional variation through relaxed selection, which is followed by "com-petitive evolution" where the most favorable variant is preserved. Sexual communication in sand flies herein offers an exciting model to test these hypotheses.

Sex pheromones mediate intraspecific communication in many systems [64], and mediate behavioral isolation between insect populations [65]. In sand flies, two different chemotypes in São Paulo (9MGB and sobralene) showed that the expansion of the visceral leishmaniasis dis-ease (canine and human form) was only correlated with the dispersion route of the 9MGB che-motype [66]. Further analyses of the variation in chemical profiles of *Lutzomyia* populations [7] are warranted, and correlating this with both geographic location and genomic composi-tion will enhance our understanding of the evolution of chemical communication. Pheromone detecting ORs among the Diptera are not well known. In the fruit fly, DmelOr67d detects the

male pheromone cVA [67], which is phylogenetically similar to the hessian fly pheromone receptor *Mdes115* [68] (S1 Fig). Our analysis identifies a few promising candidate ORs from the sand flies that cluster with these pheromone receptors (S1 Fig). Identification and functional characterization of the sex-aggregation pheromone receptors in *L. longipalpis* will provide insights into the evolutionary mechanisms associated with assortative mating between chemotypes.

In addition to sex-aggregation pheromones, the molecular basis of other sand fly behaviors, such as host-seeking, oviposition, and sugar feeding, could provide insights into these life history traits. Several studies have investigated host odors in combination with sex-aggregation pheromones and have suggested additive and, or synergistic effects [7,69]. Members of the *L. longipalpis* species complex have an irregular distribution and have adapted to a variety of tropical habitats, ranging from rocky and arid to humid and forested areas. Consequently, insights from their chemoreceptor genome, especially IRs that are being increasingly implicated in diverse roles in environmental sensing [70], will be essential to understanding the role of chemosensory gene family evolution in sexual communication and ecological adaptation.

Historically, some of the most successful campaigns against vector-borne diseases have been those targeted against the vectors [71]. Control of leishmaniasis has largely depended on reducing human exposure to sand flies using residual insecticides, repellents, bed nets and population control strategies [72]. However, there is a growing interest in developing novel vector management strategies by exploiting the chemosensory behaviors of the vector insects [73], as recently demonstrated against sand flies [74]. Our data provides novel insights into complex population structure in Brazilian sand flies, and highlights the role of chemoreceptors in the evolution of novel pheromone types, thus adding to the theoretical framework of speciation by sexual selection. While our results offer insights into the enormous genetic diversity in the chemoreceptor genome repertoire in the *L. longipalpis* populations, similar analyses at the whole genome level will identify and illustrate loci related to critical traits such as vectorial capacity, host preference, and insecticide resistance.

Supporting information

S1 Table. Accession numbers for SRAs and assigned chemotypes for the 63 *L. longipalpis* individual genome sequences used in the study.
(DOCX)

S2 Table. Chemoreceptor gene models for ORs, GRs and IRs for *L. longipalpis* and *P. papatasi*.
(XLSX)

S3 Table. SNPs in the 100 single-copy orthologs among the *L. longipalpis* field collections.
(XLSX)

S4 Table. Predicted gene copy number for 245 chemoreceptor genes in 63 *L. longipalpis* individuals.
(XLSX)

S5 Table. Gene models and confirmed absences for 522 genes based on manual annotations of the 63 *de novo* assemblies.
(XLSX)

S1 Fig. Phylogenetic relationships among ORs in *L. longipalpis*, *P. papatasi*, *A. gambiae*, *M. destructor* and *D. melanogaster*. Expanded, conserved and lost OR lineages are shaded. The phylogeny was estimated using the JTT model of protein substitution and Maximum

Likelihood method in RAxML v.8.2.4 [40]. The tree is rooted at the branch leading to Orco. Branch support based on 500 bootstrap replications.
(TIF)

S2 Fig. Phylogenetic relationships among GRs in _L. longipalpis_, _P. papatasi_, _A. gambiae_, _M. destructor_ and _D. melanogaster_. Expanded and conserved GR lineages are shaded. The phylogeny was estimated using the JTT model of protein substitution and Maximum Likelihood method in RAxML v.8.2.4 [40]. The tree is rooted at the branch leading to the CO_2 receptors. Branch support based on 500 bootstrap replications.
(TIF)

S3 Fig. Phylogenetic relationships among IRs in in _L. longipalpis_, _P. papatasi_, _A. gambiae_, _M. destructor_ and _D. melanogaster_. Conserved IR lineages are shaded. The phylogeny was estimated using the JTT model of protein substitution and Maximum Likelihood method in RAxML v.8.2.4 [40]. The tree is rooted at the branch leading to Ir25a and Ir8a. Branch support based on 500 bootstrap replications.
(TIF)

S4 Fig. _pcadapt_ was used to identify loci associated with population structure based on SNPs in the exons of 245 chemoreceptor genes in 63 individuals. (A) An optimal number of PCs, $K = 5$, was selected following preliminary analysis with $K = 20$ and subsequent visualization of the scree plot. (B) Scatterplot showing the outliers after Bonferroni correction (0.05) and the PCs they are associated with based on component-wise analysis in PCadapt [46].
(TIF)

S5 Fig. Two distinct conditions were identified producing a copy number of 1 (CN = 1), which is half the expected CN for a single-copy gene with two intact alleles. (A) Or120 has CN = 2 in LAP05, CN = 1 in LAP06 and CN = 0 in LAP20, suggesting the presence of intra-population variation in the form of two intact alleles ($A^{+/+}$), an intact and a degraded allele ($A^{+/-}$), and two degraded alleles ($A^{-/-}$), respectively. Arrows indicate sites of heterozygosity in LAP05, while all sites are homozygous in LAP06. **(B)** Ir31a-1 in JAC05 illustrates the second situation where CN = 1 represents a moderately degraded pseudogene. Arrows indicate sites of heterozygosity indicating the parents likely had at least one allele with a similar degree of degradation. The Tablet software program [50] was used for visualization of read alignments.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Paul V. Hickner, Mary Ann McDowell, Zainulabeuddin Syed.

References

1. WHO. Leishmaniasis: https://www.who.int/news-room/fact-sheets/detail/leishmanreviewiasis; 2020 [cited 2020].

2. Bates PA, Depaquit J, Galati EA, Kamhawi S, Maroli M, McDowell MA, et al. Recent advances in phlebotomine sand fly research related to leishmaniasis control. Parasites & vectors. 2015; 8(1):131. https://doi.org/10.1186/s13071-015-0712-x PMID: 25885217

3. WHO. Leishmaniasis, The Vector: https://www.who.int/leishmaniasis/disease/vector/en/; 2020.

4. Souza NA, Brazil RP, Araki AS. The current status of the *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae) species complex. Memorias do Instituto Oswaldo Cruz. 2017; 112(3):161–74. https://doi.org/10.1590/0074-02760160463 PMID: 28225906

5. Bray DP, Ward RD, Hamilton JG. The chemical ecology of sandflies. In: Takken W, Knols B, editors. Ecology and control of vector-borne diseases2010. p. 203–16.

6. Stökl J, Steiger S. Evolutionary origin of insect pheromones. Current opinion in insect science. 2017; 24:36–42. https://doi.org/10.1016/j.cois.2017.09.004 PMID: 29208221

7. Spiegel CN, dos Santos Dias DB, Araki AS, Hamilton JG, Brazil RP, Jones TM. The *Lutzomyia longipalpis* complex: a brief natural history of aggregation-sex pheromone communication. Parasites & vectors. 2016; 9(1):580. https://doi.org/10.1186/s13071-016-1866-x PMID: 27842601

8. Lanzaro GC, Warburg A. Genetic variability in phlebotomine sandflies: possible implications for leishmaniasis epidemiology. Parasitology today. 1995; 11(4):151–4.

9. Araki AS, Ferreira GE, Mazzoni CJ, Souza NA, Machado RC, Bruno RV, et al. Multilocus analysis of divergence and introgression in sympatric and allopatric sibling species of the Lutzomyia longipalpis complex in Brazil. PLoS neglected tropical diseases. 2013; 7(10):e2495. https://doi.org/10.1371/journal.pntd.0002495 PMID: 24147172

10. Hamilton JG, Maingon RD, Alexander B, Ward RD, Brazil RP. Analysis of the sex pheromone extract of individual male *Lutzomyia longipalpis* sandflies from six regions in Brazil. Med Vet Entomol. 2005; 19 (4):480–8. Epub 2005/12/13. https://doi.org/10.1111/j.1365-2915.2005.00594.x PMID: 16336313.

11. Palframan MJ, Bandi KK, Hamilton JG, Pattenden GJTl. Sobralene, a new sex-aggregation pheromone and likely shunt metabolite of the taxadiene synthase cascade, produced by a member of the sand fly *Lutzomyia longipalpis* species complex. 2018; 59(20):1921–3. https://doi.org/10.1016/j.tetlet.2018.03.088 PMID: 29780183

12. Fuller RC, Endler JA. A perspective on sensory drive. Current zoology. 2018; 64(4):465–70. https://doi.org/10.1093/cz/zoy052 PMID: 30108627

13. Yohe LR, Brand P. Evolutionary ecology of chemosensation and its role in sensory drive. Current zoology. 2018; 64(4):525–33. https://doi.org/10.1093/cz/zoy048 PMID: 30108633

14. Stevens M. Sensory ecology, behaviour, and evolution: Oxford University Press; 2013.

15. Raji JI, DeGennaro M. Genetic analysis of mosquito detection of humans. Current opinion in insect science. 2017; 20:34–8. https://doi.org/10.1016/j.cois.2017.03.003 PMID: 28428935

16. Ruzzante L, Reijnders MJ, Waterhouse RM. Of genes and genomes: mosquito evolution and diversity. Trends in parasitology. 2019; 35(1):32–51. https://doi.org/10.1016/j.pt.2018.10.003 PMID: 30391118

17. Syed Z. Chemical ecology and olfaction in arthropod vectors of diseases. Current Opinion in Insect Science. 2015; 10:83–9. https://doi.org/10.1016/j.cois.2015.04.011 PMID: 29588018

18. Carrasco D, Lefèvre T, Moiroux N, Pennetier C, Chandre F, Cohuet A. Behavioural adaptations of mosquito vectors to insecticide control. Current opinion in insect science. 2019; 34:48–54. https://doi.org/10.1016/j.cois.2019.03.005 PMID: 31247417

19. Brown JE, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, et al. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. Evolution. 2014; 68(2):514–25. https://doi.org/10.1111/evo.12281 PMID: 24111703

20. Powell JR, Gloria-Soria A, Kotsakiozi P. Recent history of *Aedes aegypti*: Vector genomics and epidemiology records. Bioscience. 2018; 68(11):854–60. https://doi.org/10.1093/biosci/biy119 PMID: 30464351

21. Trpis M, Hausermann W. Genetics of house-entering behaviour in East African populations of *Aedes aegypti* (L.)(Diptera: Culicidae) and its relevance to speciation. Bulletin of Entomological Research. 1978; 68(3):521–32.

22. McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, et al. Evolution of mosquito preference for humans linked to an odorant receptor. Nature. 2014; 515(7526):222–7. https://doi.org/10.1038/nature13964 PMID: 25391959

23. Rose NH, Sylla M, Badolo A, Lutomiah J, Ayala D, Aribodor OB, et al. Climate and Urbanization Drive Mosquito Preference for Humans. Curr Biol. 2020. Epub 2020/07/25. https://doi.org/10.1016/j.cub.2020.06.092 PMID: 32707056.

24. Araki AS, Brazil RP, Hamilton JG, Vigoder FM. characterization of copulatory courtship song in the old World sand fly species *Phlebotomus argentipes*. Scientific reports. 2020; 10(1):1–5. https://doi.org/10.1038/s41598-019-56847-4 PMID: 31913322

25. Benton R. Multigene family evolution: perspectives from insect chemoreceptors. Trends in ecology & evolution. 2015; 30(10):590–600. https://doi.org/10.1016/j.tree.2015.07.009 PMID: 26411616

26. Robertson HM. Molecular evolution of the major arthropod chemoreceptor gene families. Annual review of entomology. 2019; 64:227–42. https://doi.org/10.1146/annurev-ento-020117-043322 PMID: 30312552

27. Wellenreuther M, Mérot C, Berdan E, Bernatchez L. Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. Molecular ecology. 2019; 28(6):1203–9. https://doi.org/10.1111/mec.15066 PMID: 30834648

28. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nature Reviews Genetics. 2008; 9(12):951. https://doi.org/10.1038/nrg2480 PMID: 19002141

29. Hickner PV, Rivaldi CL, Johnson CM, Siddappaji M, Raster GJ, Syed Z. The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. 2016; 17(1):648.

30. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic acids research. 2015; 43(Database issue):D707–13. Epub 2014/12/17. https://doi.org/10.1093/nar/gku1117 PMID: 25510499; PubMed Central PMCID: PMC4383932.

31. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome biology. 2015; 16(1):157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

32. Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Molecular biology and evolution. 2015; 32(10):2798–800. https://doi.org/10.1093/molbev/msv150 PMID: 26130081

33. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS genetics. 2010; 6(8):e1001064. https://doi.org/10.1371/journal.pgen.1001064 PMID: 20808886

34. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, et al. G protein-coupled receptors in *Anopheles gambiae*. Science. 2002; 298(5591):176–8. https://doi.org/10.1126/science.1076196 PMID: 12364795

35. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. Nature. 2018; 563(7732):501. https://doi.org/10.1038/s41586-018-0692-z PMID: 30429615

36. Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences. 2003; 100(suppl 2):14537–42. https://doi.org/10.1073/pnas.2335847100 PMID: 14608037

37. Zhao C, Escalante LN, Chen H, Benatti TR, Qu J, Chellapilla S, et al. A massive expansion of effector genes underlies gall-formation in the wheat pest Mayetiola destructor. Current Biology. 2015; 25 (5):613–20. https://doi.org/10.1016/j.cub.2014.12.057 PMID: 25660540

38. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004; 32(5):1792–7. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

39. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25(15):1972–3. https://doi.org/10.1093/bioinformatics/btp348 PMID: 19505945

40. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

41. Leinonen R, Sugawara H, Shumway M, Collaboration INSD. The sequence read archive. Nucleic acids research. 2010; 39(suppl_1):D19–D21.

42. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

43. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.

44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25(16):2078–9. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

45. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007; 23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308 PMID: 17586829

46. Luu K, Bazin E, Blum MG. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular ecology resources. 2017; 17(1):67–77. https://doi.org/10.1111/1755-0998.12592 PMID: 27601374

47. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome research. 2009; 19(9):1655–64. https://doi.org/10.1101/gr.094052.109 PMID: 19648217

48. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular ecology resources. 2015; 15(5):1179–91. https://doi.org/10.1111/1755-0998.12387 PMID: 25684545

49. Jombart T, Pontier D, Dufour A-B. Genetic markers in the playground of multivariate analysis. Heredity. 2009; 102(4):330. https://doi.org/10.1038/hdy.2008.130 PMID: 19156164

50. Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, et al. Using Tablet for visual exploration of second-generation sequencing data. Briefings in bioinformatics. 2012; 14(2):193–202. https://doi.org/10.1093/bib/bbs012 PMID: 22445902

51. Rinker DC, Specian NK, Zhao S, Gibbons JG. Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. Proceedings of the National Academy of Sciences. 2019:201901093. https://doi.org/10.1073/pnas.1901093116 PMID: 31209046

52. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015; 31(10):1674–6. https://doi.org/10.1093/bioinformatics/btv033 PMID: 25609793

53. Pech-May A, Ramsey JM, Ittig REG, Giuliani M, Berrozpe P, Quintana MG, et al. Genetic diversity, phylogeography and molecular clock of the *Lutzomyia longipalpis* complex (Diptera: Psychodidae). PLoS

neglected tropical diseases. 2018; 12(7):e0006614. https://doi.org/10.1371/journal.pntd.0006614 PMID: 29975695

54. Ohno S. Evolution by gene duplication: Springer Science & Business Media; 2013.

55. Warburg A, Saraiva E, Lanzaro GC, Titus RG, Neva F. Saliva of *Lutzomyia longipalpis* sibling species differs in its composition and capacity to enhance leishmaniasis. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1994; 345(1312):223–30. https://doi.org/10.1098/rstb.1994.0097 PMID: 7972360

56. McCoy K. The population genetic structure of vectors and our understanding of disease epidemiology. EDP Sciences; 2008. https://doi.org/10.1051/parasite/2008153444 PMID: 18814720

57. Bates PA, Depaquit J, Galati EA, Kamhawi S, Maroli M, McDowell MA, et al. Recent advances in phlebotomine sand fly research related to leishmaniasis control. 2015; 8(1):131.

58. Ready PD, Vigoder FM, Rangel EF. Molecular and Biochemical Markers for Investigating the Vectorial Roles of Brazilian Sand Flies. Brazilian Sand Flies: Springer; 2018. p. 213–50.

59. Niimura Y, Nei M. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. Journal of human genetics. 2006; 51(6):505–17. https://doi.org/10.1007/s10038-006-0391-8 PMID: 16607462

60. Sánchez-Gracia A, Vieira F, Rozas J. Molecular evolution of the major chemosensory gene families in insects. Heredity. 2009; 103(3):208–16. https://doi.org/10.1038/hdy.2009.55 PMID: 19436326

61. Johnson BR, Tsutsui ND. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC genomics. 2011; 12(1):164. https://doi.org/10.1186/1471-2164-12-164 PMID: 21447185

62. Duvaux L, Geissmann Q, Gharbi K, Zhou J-J, Ferrari J, Smadja CM, et al. Dynamics of copy number variation in host races of the pea aphid. Molecular biology and evolution. 2014; 32(1):63–80. https://doi.org/10.1093/molbev/msu266 PMID: 25234705

63. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics. 2010; 11(2):97. https://doi.org/10.1038/nrg2689 PMID: 20051986

64. Wyatt TD. Pheromones and animal behavior: chemical signals and signatures: Cambridge University Press; 2014.

65. Smadja C, Butlin R. On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity. 2009; 102(1):77–97. https://doi.org/10.1038/hdy.2008.55 PMID: 18685572

66. Casanova C, Colla-Jacques FE, Hamilton JG, Brazil RP, Shaw JJ. Distribution of *Lutzomyia longipalpis* chemotype populations in São Paulo state, Brazil. PLoS Negl Trop Dis. 2015; 9(3):e0003620. https://doi.org/10.1371/journal.pntd.0003620 PMID: 25781320

67. van Naters WvdG Carlson JR. Receptors and neurons for fly odors in Drosophila. Current biology. 2007; 17(7):606–12. https://doi.org/10.1016/j.cub.2007.02.043 PMID: 17363256

68. Andersson MN, Corcoran JA, Zhang D-D, Hillbur Y, Newcomb RD, Löfstedt C. A sex pheromone receptor in the Hessian fly *Mayetiola destructor* (Diptera, Cecidomyiidae). Frontiers in cellular neuroscience. 2016; 10:212.

69. Hamilton J. Sandfly pheromones: their biology and potential for use in control programs. EDP Sciences; 2008.

70. Rimal S, Lee Y. The multidimensional ionotropic receptors of *Drosophila melanogaster*. Insect molecular biology. 2018; 27(1):1–7. https://doi.org/10.1111/imb.12347 PMID: 28857341

71. Shaw WR, Catteruccia F. Vector biology meets disease control: using basic research to fight vector-borne diseases. Nature microbiology. 2019; 4(1):20–34. https://doi.org/10.1038/s41564-018-0214-7 PMID: 30150735

72. Alexander B, Maroli M. Control of phlebotomine sandflies. Medical and veterinary entomology. 2003; 17 (1):1–18. https://doi.org/10.1046/j.1365-2915.2003.00420.x PMID: 12680919

73. Carey AF, Carlson JR. Insect olfaction from model systems to disease control. Proceedings of the National Academy of Sciences. 2011; 108(32):12987–95. https://doi.org/10.1073/pnas.1103472108 PMID: 21746926

74. Courtenay O, Dilger E, Calvo-Bado LA, Kravar-Garde L, Carter V, Bell MJ, et al. Sand fly synthetic sex-aggregation pheromone co-located with insecticide reduces the incidence of infection in the canine reservoir of visceral leishmaniasis: A stratified cluster randomised trial. PLoS neglected tropical diseases. 2019; 13(10):e0007767. https://doi.org/10.1371/journal.pntd.0007767 PMID: 31652261