



Genetic evidence for the origin of *Aedes aegypti*, the yellow fever mosquito, in the southwestern Indian Ocean

John Soghigian^{1,2} | Andrea Gloria-Soria^{1,3} | Vincent Robert⁴ | Gilbert Le Goff⁴ | Anna-Bella Failloux⁵ | Jeffrey R. Powell¹

¹Yale University, New Haven, CT, USA

²Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, NC, USA

³Center for Vector Biology & Zoonotic Diseases, Department of Environmental Sciences, The Connecticut Agricultural Experiment Station, New Haven, CT, USA

⁴MIVEGEC Unit, IRD, Montpellier Univ., CNRS, Montpellier, France

⁵Institut Pasteur, Arboviruses and Insect Vectors, Paris, France

Correspondence

John Soghigian, 3114 Gardner Hall Box 7613, North Carolina State University, Raleigh NC 27695-7613, USA.
Email: john.soghigian@gmail.com

Funding information

Centers for Disease Control and Prevention, Grant/Award Number: U01CK000509; Division of Environmental Biology, Grant/Award Number: 1754376; National Institute of Allergy and Infectious Diseases, Grant/Award Number: AI101112

Abstract

Aedes aegypti is among the best-studied mosquitoes due to its critical role as a vector of human pathogens and ease of laboratory rearing. Until now, this species was thought to have originated in continental Africa, and subsequently colonized much of the world following the establishment of global trade routes. However, populations of this mosquito on the islands in the southwestern Indian Ocean (SWIO), where the species occurs with its nearest relatives referred to as the Aegypti Group, have received little study. We re-evaluated the evolutionary history of *Ae. aegypti* and these relatives, using three data sets: nucleotide sequence data, 18,489 SNPs and 12 microsatellites. We found that: (a) the Aegypti Group diverged 16 MYA (95% HPD: 7–28 MYA) from its nearest African/Asian ancestor; (b) SWIO populations of *Ae. aegypti* are basal to continental African populations; (c) after diverging 7 MYA (95% HPD: 4–15 MYA) from its nearest formally described relative (*Ae. mascarensis*), *Ae. aegypti* moved to continental Africa less than 85,000 years ago, where it recently (<1,000 years ago) split into two recognized subspecies *Ae. aegypti formosus* and a human commensal, *Ae. aegypti aegypti*; (d) the Madagascar samples form a clade more distant from all other *Ae. aegypti* than the named species *Ae. mascarensis*, implying that Madagascar may harbour a new cryptic species; and (e) there is evidence of introgression between *Ae. mascarensis* and *Ae. aegypti* on Réunion, and between the two subspecies elsewhere in the SWIO, a likely consequence of recent introductions of domestic *Ae. aegypti aegypti* from Asia.

KEYWORDS

insects, invasive species, population genetics—empirical, systematics

1 | INTRODUCTION

The *Aedes aegypti* mosquito is often considered to be among the most dangerous animal in the world due to its ability to transmit several arboviruses (yellow fever, dengue, chikungunya and Zika) that

historically have taken a heavy toll on human health and continue to do so today (Powell, 2016). The European colonization of the New World was strongly affected by *Ae. aegypti*, and these events define the Americas today (McNeill, 2010). Beyond its direct role in disease transmission, this mosquito is easy to rear in the laboratory

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Molecular Ecology published by John Wiley & Sons Ltd

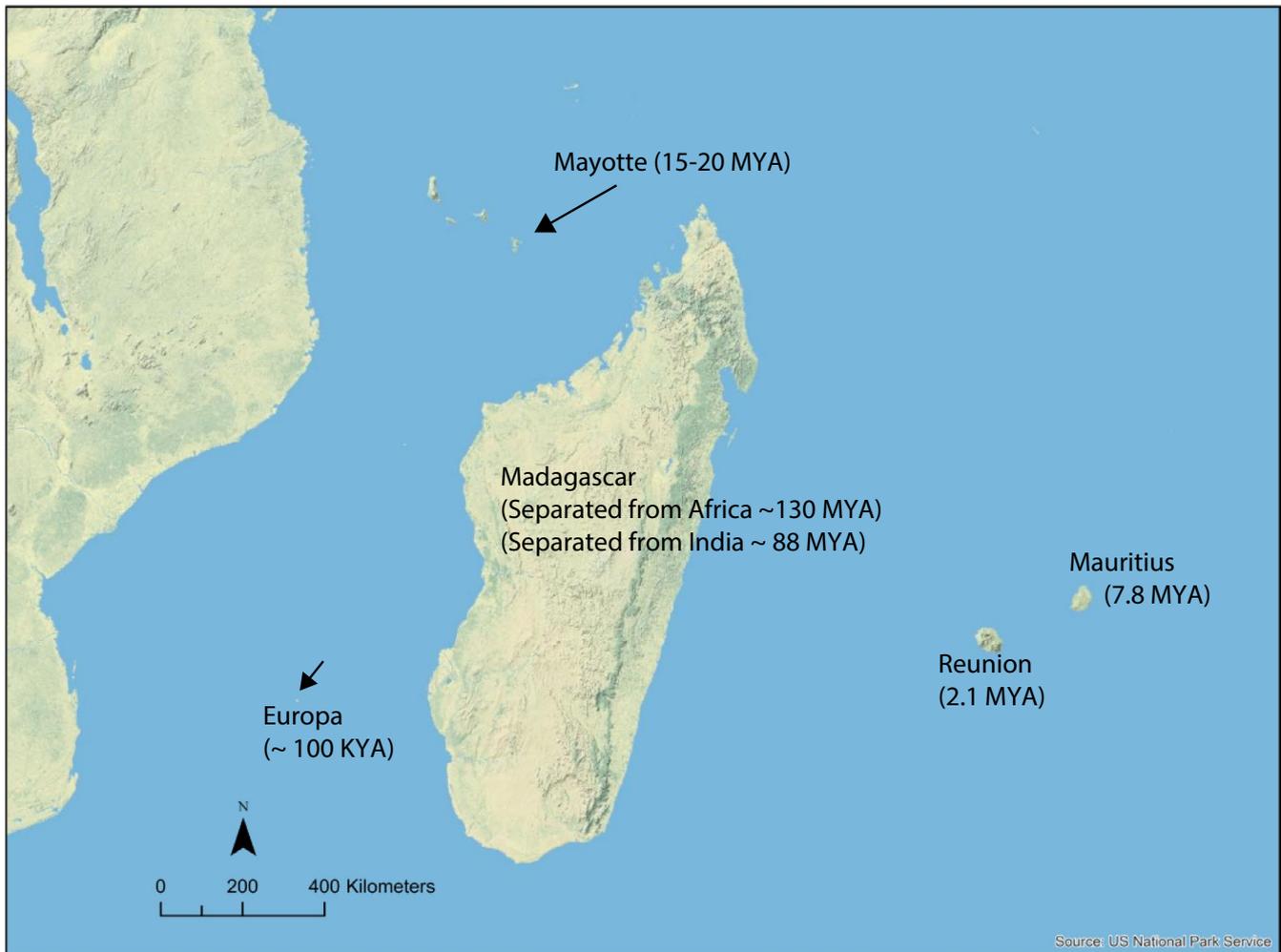


FIGURE 1 Map of the southwestern Indian Ocean, east of Africa. Estimates of island ages are from Ashwal et al. (2017), Warren et al. (2003) and Michon (2016). While *Ae. aegypti* is found on Madagascar and all islands highlighted, *Ae. mascarensis* is endemic to Mauritius, and *Ae. pia* is endemic to Mayotte [Colour figure can be viewed at wileyonlinelibrary.com]

and thus has been a model organism for laboratory-based disciplines such as physiology, neurobiology and development (Christophers, 1960; Clements, 1999a, 1999b, 2012). Its evolutionary genetics has been well studied for four decades (Crawford et al., 2017; Gloria-Soria et al., 2016; Paupy, Vazeille-Falcoz, Mousson, Rodhain, & Failloux, 2000; Rašić, Filipović, Weeks, & Hoffmann, 2014; Tabachnick & Powell, 1979), and it has been a model for invasive species biology (Braks, Honório, Lounibos, Lourenço-de-Oliveira, & Juliano, 2004; Brown et al., 2014).

Aedes aegypti is globally distributed in the tropics and subtropics. Ancestrally, there is little doubt that it occupied sub-Saharan Africa, where populations are still found in tropical rainforests, with larvae breeding in tree holes and female adults taking bloodmeals from nonhuman mammals (Lounibos, 1981; McBride et al., 2014). As human settlements grew in Africa, populations of this mosquito evolved to become associated with human habitats, where larvae can be found in human-generated containers and females prefer humans for bloodmeals. About 500 years ago, this human-associated form left Africa, likely via slave trade, and first invaded the New World, then subsequently Asia and the Pacific Islands, including

Australia (reviewed in Powell, Gloria-Soria, & Kotsakiozi, 2018). These two forms, one sylvatic in continental Africa and one domestic, primarily outside Africa, have been given subspecific names: *Ae. aegypti formosus* (abbreviated Aaf) and *Ae. aegypti aegypti* (Aaa), respectively. The vast majority of work has been on these two subspecies that we here refer to as *Aedes aegypti sensu stricto* (s.s.), to distinguish them from the newly studied populations from the islands in the southwestern Indian Ocean, which we will refer to here as *Aedes aegypti sensu lato* (s.l.). We note that while the clear-cut distinction of the subspecies designations is questionable in some instances (e.g. Powell & Tabachnick, 2013), here we use Aaf and Aaa as convenient shorthand for African native and outside Africa populations, respectively.

While the evolutionary history and genetic diversity of *Ae. aegypti* s.s. have received substantial study, populations of *Ae. aegypti* (and close relatives) on the islands east of Africa have received comparatively little attention (Failloux, Vazeille, & Rodhain, 2002; Kotsakiozi, Evans, et al., 2018; Vazeille et al., 2001). In this region of the southwestern Indian Ocean, *Ae. aegypti* occupies Madagascar and numerous smaller islands between Africa and Madagascar, such

as Mayotte and Europa, as well as islands east of Madagascar, such as Réunion and Mauritius (Figure 1). A previous study estimated that Malagasy *Ae. aegypti* diverged from their continental African counterparts approximately 10 MYA (Fort et al., 2012).

In addition to *Ae. aegypti*, two closely related species from these islands have been formally described. These are *Ae. mascarensis*, endemic on Mauritius (MacGregor, 1924); and *Ae. pia*, endemic on Mayotte (Le Goff, Brengues, & Robert, 2013). Together with *Ae. aegypti* s.s., these three species, in the subgenus *Stegomyia*, compose the Aegypti Group (Huang, 2004). *Aedes mascarensis* and *Ae. aegypti* s.s. can mate in the laboratory forming fertile F_1 offspring, but hybrid viability breaks down after the first generation (Hartberg & Craig, 1970). Genetic analysis of *Ae. aegypti* s.l. from Réunion has suggested the possibility of natural introgression from *Ae. mascarensis* (Kotsakiozi, Evans, et al., 2018). *Aedes pia* was described in 2013 (Le Goff et al., 2013), and is basal in phylogenetic analyses to all other members of the Aegypti Group (Soghigian, Andreadis, & Livdahl, 2017). The reproductive compatibility of *Ae. pia* with other members of the Aegypti Group is unknown.

In this paper, we push back considerably longer in time our understanding of the evolutionary history of *Ae. aegypti*, to the origin of *Ae. aegypti* s.s. on continental Africa. New data from populations on southwestern Indian Ocean islands provide evidence for novel insights.

2 | METHODS

2.1 | Mosquito collection and DNA extraction

Aedes aegypti were collected by our group, as well as collaborators from various geographic locations worldwide, particularly from the islands in the southwestern Indian Ocean (Table S1) and were processed at Yale University, New Haven, USA. DNA was extracted for all samples, other than *Ae. pia*, with the Qiagen DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's recommendations and including the optional RNase A step, and stored at -20°C until further use. *Aedes pia* is a morphologically distinct member of the subgenus *Stegomyia* and is basal to the other two members of the Aegypti species group: *Aedes aegypti* and *Aedes mascarensis* (Soghigian et al., 2017). DNA from two adult *Aedes pia* was extracted from pinned specimens, using phenol–chloroform and genotyped as described for the other specimens. We also collected adult specimens of *Culiseta melanura*, *Culiseta inornata* and *Culex erraticus* from Worcester, for use in divergence time analyses, below.

2.2 | Estimating the evolutionary history of the Aegypti Group with nucleotide data

To estimate the evolutionary relationships and date the divergence times in the Aegypti Group, we generated new sequence data for *Ae. aegypti* from Madagascar and other species (see below), and we used

previously published sequence data from Soghigian et al. (2017) for mosquitoes from the genus *Aedes*, subgenus *Stegomyia*. The subset of Soghigian et al. (2017) contained nucleotide data from *Stegomyia* for 25 species from 12 species groups (Table S2), including five of the ten species groups found in Africa (Huang, 2004). Soghigian et al. (2017) analysed an alignment of sequence data from 200 *Aedes* mosquitoes, generated from both sequencing of specimens and data mining of GenBank, to study the evolution of larval habitat specialization in *Aedes*; this work did not focus on *Stegomyia* specifically, but contained numerous species from this subgenus. From this original data set, we also used outgroups from *Psorophora* and *Culex* (supplemented by additional outgroups sequenced for this study; see below) for which fossil calibrations are known. We used nucleotide sequence data for estimating the evolutionary relationships of *Stegomyia*, rather than SNP data detailed elsewhere, as the SNP chip data captured from samples more distant from *Ae. aegypti* s.s. is limited (see supplemental information for missing data discussion) and because fossil calibrations are not available for the Aegypti Group but are for *Aedes* and other Culicidae (see below and supplemental information). The alignment derived from Soghigian et al. (2017) included partitions for seven markers: 18S, 28S, enolase, arginine kinase, cytochrome oxidase I and II, and ITS2. This alignment was merged with additional sequence data for *Ae. aegypti* from Madagascar and species for which additional fossil calibrations were available (see below and supplemental information), generated following sequencing protocols detailed more thoroughly in Soghigian et al. (2017). In brief, the aforementioned markers were amplified using previously published primers (Cook, Diallo, Sall, Cooper, & Holmes, 2005; Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994; Reidenbach et al., 2009; Soghigian et al., 2017; Wesson, Porter, & Collins, 1992), given in Table S3, then sequenced in both directions by Macrogen (Macrogen Boston USA). Using Geneious version 9 (Drummond et al., 2012) and the MAFFT version 7 (Katoh, Misawa, Kuma, & Miyata, 2002) alignment software, new sequences from *Ae. aegypti* from Madagascar, *Cs. melanura*, *Cs. inornata* and *Cx. erraticus*, were added to the existing alignment of each marker from Soghigian et al. (2017), such that the new alignment was 6,471 base pairs from 33 taxa. Hereafter, we will refer to this as our nucleotide data set.

We estimated the evolutionary history of the Aegypti Group, and the rest of the subgenus *Stegomyia*, with the nucleotide data set in IQ-Tree (Hoang, Chernomor, von Haeseler, Minh, & Vinh, 2018; Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermini, 2017; Nguyen, Schmidt, von Haeseler, & Minh, 2015). We partitioned our data set according to marker, and by codon position for protein-coding genes. We then allowed IQ-Tree to choose the best-fitting substitution model per partition, and we assessed clade support using ultrafast bootstrap values. In order to determine whether our mitochondrial markers (cytochrome oxidase I and II) had concordant topologies with our nuclear data (18S, 28S, ITS2, arginine kinase and enolase), we repeated this analysis with only the mitochondrial partitions and with only nuclear partitions, and compared subsequent results with our topology from all seven markers, as well as compared our results with the phylogeny of

the *Aegypti* Group obtained from SNPs (see below). We included several *Stegomyia* species missing loci (see Table S2) in order to increase the breadth of our taxonomic sampling of the subgenus. Although missing data are generally thought not to influence phylogenetic results in terms of topology or support values (Pyron, Burbrink, & Wiens, 2013; Roure, Baurain, & Philippe, 2013; Wiens & Morrill, 2011; Wiens & Tiu, 2012), we evaluated whether the degree of missing data at some markers might influence relative branch lengths, as these branch lengths would be essential for accurate divergence time estimates. Following Pyron et al. (2013), we used Pearson's correlation to test whether the terminal branch length from the ML analysis of the full nucleotide data set was correlated for (a) the number of loci present in the analysis, and (b) the proportion of nucleotides in the alignment.

Next, we used BEAST2 (Bouckaert et al., 2014) to estimate divergence times among members of the *Aegypti* Group. We used bModelTest (Bouckaert & Drummond, 2017) to estimate substitution models for each partition for ribosomal RNA genes, and by codon position for protein-coding genes. We used a birth–death tree process and a relaxed clock model following recommendations for BEAST2 divergence time analyses (Barido-Sottani et al., 2018). We placed four fossil calibrations, including one fossil from the genus *Aedes* and three outside of the genus, with uniform priors set to minimum ages based on minimum fossil age, and maximum ages set to previous estimates from Reidenbach et al. (2009; see Table S16 for fossils, constraints and explanation). We ran the BEAST2 analysis for 150,000,000 generations on CIPRES (Miller, Pfeiffer, & Schwartz, 2010), after which we evaluated logs to ensure convergence of parameter estimates and subsequently generated a maximum clade credibility tree with TreeAnnotator from BEAST2 with common ancestor heights and 95% highest posterior density (HPD) intervals. We ran two additional chains to confirm consistent estimates of divergence time across chains.

2.3 | SNP genotyping, calling and filtering

For fine-scale population structure and admixture analysis, we generated a SNP data set with populations of *Ae. aegypti* from throughout its global distribution, and from other members of the *Aegypti* Group, with a focus on the southwestern Indian Ocean. Approximately 200 ng of genomic DNA per individual mosquito was genotyped with the *Ae. aegypti* Axiom_*aegypti*1 SNP chip (Life Technologies Corporation CAT#550481; Evans et al., 2015), at the University of North Carolina Functional Genomics Core, Chapel Hill. Additionally, we used previously described samples from Gloria-Soria et al. (2018) and Kotsakiozi, Evans, et al. (2018) from several populations of *Aedes aegypti* collected worldwide, as well as for *Ae. mascarensis* from the island of Mauritius (Table S1). For population analyses, we classified populations into five groups corresponding to currently accepted species and subspecies designations (e.g. we divided *Aedes aegypti* s.s. into typically recognized groups of *Ae. aegypti aegypti* and *Ae. aegypti formosus*), and a group for "island"

populations of *Ae. aegypti* originating from Europa, Madagascar, Mayotte and Réunion.

Genotype calls were made using Affymetrix Axiom Analysis Suite 3.1 (Life Technologies–Thermo Fisher Scientific), for both newly genotyped samples and previously studied populations. This integrated software suite calls genotypes from probe fluorescence values, excluding probes (and thus SNPs) whose fluorescence deviates from expectations, as might result from probe misbinding. For probe quality control and SNP calling, we used default software settings, save for two exceptions: first, in order to include SNPs called from *Ae. pia*, we overrode default software settings on allowable missing data for a given sample (this applied only to the two *Ae. pia* samples; see supplemental materials for a discussion on missing data); and second, we utilized the OTV Caller plugin for Axiom Analysis Suite to exclude genotype calls from SNPs that might contain off-target variants, such as those that could occur due to insertions, deletions or mutations at probe sites. Finally, we exported all polymorphic SNPs for downstream analyses.

Resulting polymorphic SNPs were then filtered using Plink version 1.9 (Chang et al., 2015). We removed SNPs that Evans et al. (2015) found not to follow Mendelian inheritance patterns. We also removed variants present missing in more than 25% of individuals (flag --geno 0.25), alleles with a minor allele frequency lower than 0.01 (--maf 0.01) and linked loci within a 75-kb window that exceeded a variance inflation factor of 2 (--indep 75k 1 2). This window size was chosen as it exceeded the linkage decay previously reported in *Ae. aegypti* (Matthews et al., 2018). We refer to this data set throughout the text as the SNP data set.

2.4 | Microsatellite genotyping

For complementary estimations of population history and demographic parameters, we generated a microsatellite data set comprised of samples from Europa Island (hereafter Europa) and Africa. Microsatellite genotyping was performed as described previously (Gloria-Soria et al., 2016). We used previously published microsatellite calls (Brown et al., 2011; Gloria-Soria et al., 2016), available on VectorBase (Giraldo-Calderón et al., 2015), for all but populations from Burkina Faso and Europa, for which data were generated in this study.

2.5 | Genetic diversity and population structure with the SNP data set

We used our population-level SNP data set to infer genetic diversity and population structure of *Ae. aegypti* in the southwestern Indian Ocean. Unless otherwise described, all analyses were performed in R version 3.5 (R Core Team, 2018), and maps were generated using Google Maps, or PaleoMap PaleoAtlas for GPlates (available from <http://www.gplates.org/>). We used a nested analysis of molecular variance (AMOVA) from the package poppr (Kamvar, Tabima, & Grünwald, 2014), defining nested strata according to (a) species/

subspecies and (b) population (see Table S1) to assess whether regions and populations were significantly different from one another. We also calculated pairwise F_{st} values with the R package StAMPP (Pembleton, Cogan, & Forster, 2013) for subspecies and population pairs (see Table S1 for populations and subspecies), regardless of the sample size of each population, as F_{st} can be reliably estimated from as few as two individuals, so long as there are many thousands of markers (Nazareno, Bemmels, Dick, & Lohmann, 2017; Patterson, Price, & Reich, 2006). Next, we used sparse non-negative matrix factorization (hereafter SNMF) as implemented in the package LEA (Frichot & François, 2015) to analyse population structure. We used the minimal cross-entropy criterion to determine the optimal number of ancestral populations (K values), and also visualized alternative K values based on biologically reasonable ancestral clusters due to similarity in K values past $K = 4$. We also assessed whether the removal of the genetic cluster associated with *Aedes pia* would provide qualitatively different results from the inclusion of this outgroup species. Additionally, we performed principal component analysis to evaluate genetic structure of populations and species as implemented in the R package adegenet (Jombart & Ahmed, 2011).

2.6 | Phylogenetic analysis with the SNP data set

We evaluated the evolutionary relationships among populations of *Ae. aegypti*, as well as other members of the Aegypti Group, using maximum-likelihood inference on the population-level SNP data set with ascertainment bias correction for invariant sites, as implemented in IQ-Tree (Hoang et al., 2018; Kalyaanamoorthy et al., 2017; Nguyen et al., 2015). We allowed IQ-Tree to select the best-fitting substitution model, and we assessed clade support using ultrafast bootstrap values. We rooted the resulting topology on the branch leading to *Aedes pia* based on our BEAST2 analysis (above), which was consistent with previous results in Soghigian et al. (2017) that suggested *Aedes pia* was sister to *Aedes mascarensis* + *Aedes aegypti*. This choice was further supported by all other analyses that suggested *Aedes pia* was more distant from *Aedes aegypti* than *Aedes mascarensis* (or any other population), for example relative missing data (see online supplemental information) and population clustering results.

2.7 | Admixture analysis with the SNP data set

Because our initial analysis of population structure supported a scenario of potential admixture, we used the F3 test (Reich, Thangaraj, Patterson, Price, & Singh, 2009) implemented in Treemix (Pickrell & Pritchard, 2012) to evaluate whether there was significant evidence of admixture among populations in our SNP data set. We compared population triplets, considering whether there was significant non-tree-like behaviour for all sets of populations, where negative Z scores were indicative of admixture between two parent populations to produce the third sibling population. p -values were calculated

from these Z scores with false discovery rate correction for multiple comparisons with the R function `p.adjust` (R Core Team, 2018).

To ensure that specific characteristics of our data set in terms of missing data and uneven sampling did not bias our admixture analysis, we estimated two additional F3 statistics per population triplet and compared them with our complete data set. First, to evaluate whether missing data influenced our Treemix results, we refiltered our data set as above, but with the flag `-geno 0` in Plink, thus requiring a SNP position be present in all samples to be included. We then used this data set in Treemix, as above, to estimate F3 statistics and p -values. Next, we evaluated whether the uneven sample size of some populations (e.g. *Aedes mascarensis* with four individuals relative to most populations with >11) could bias estimates of allelic variation between populations, as Treemix uses allele frequency estimates to infer admixture through the F3 test. Using the R function `dplyr`, we resampled, without replacement, each *Ae. aegypti* population to four individuals. For each resampling, we conducted F3 tests as implemented in Treemix as we describe above. We resampled our populations 100 times, ran Treemix on each resampled data set and then summarized the resampled mean estimates of the F3 statistic with parametric 95% confidence intervals. This summary method, rather than p -values, was chosen due to the large number of three-population comparisons (~204,000) during resampling.

2.8 | Estimates of demographic parameters and population history with microsatellites

We complemented our inferences on population history using microsatellite markers to further investigate the divergence events that involved African *Ae. aegypti* s.s. and their relatives on Europa. Particularly, we were interested in those events that occurred more recently in evolutionary history and for which nuclear and mitochondrial genes did not resolve the divergence times or directionality. Furthermore, *Aedes aegypti* SNP chip only assays two alleles at each locus and was designed to deliberately emphasize population-level differences among *Ae. aegypti aegypti*, possibly neglecting informative variation in African *Ae. aegypti*. Based on the admixture detected in other island populations (see below), we focused this analysis on comparisons between the Europa population, and populations from East and West Africa. Allele numbers, percentage of missing data, percentage of polymorphic loci, average observed (H_o) and expected (uHe) heterozygosities, and F_{is} values were estimated in GenAIEx (Peakall & Smouse, 2012). All microsatellite loci were analysed for within-population deviations from Hardy–Weinberg equilibrium (HWE) in Genodive 2.0b.27 (Meirmans & Tienderen, 2004) with the Heterozygosity-based (Nei) test and 10,000 permutations. A Bonferroni correction was applied to the resulting matrices of HWE to correct for multiple testing. Next, we inferred population history and demographic patterns using approximate Bayesian computation methods, ABC (Beaumont, Zhang, & Balding, 2002) as implemented in DIYABC v.2.0.4 (Cornuet et al., 2014), estimating 10 generations per year and a mutation rate ranging from 9×10^{-6} to 1×10^{-5} ,

consistent with previous publications (Gloria-Soria et al., 2016) and rates reported in the literature for other Diptera species (Pfeiler, Flores-López, Mada-Vélez, Escalante-Verdugo, & Markow, 2013; Schug et al., 1998). ABC allowed us to explicitly contrast different hypotheses on the population history of these species, namely whether Europa diverged prior to East and West Africa, or after East or West Africa.

The ABC analysis was performed as five independent runs on microsatellite data sets, with each run containing 24 individuals belonging to populations of *Ae. aegypti* in West Africa (WA), East Africa (EA) and Europa. The data sets were from: (1) La Lope Forest (Gabon—WA) and Lunyo (Uganda—EA); (2) Ouagadougou (Burkina Faso—WA) and Kisumu (Kenya—EA); and (3–5) independent combinations of randomly drawn individuals from WA (Ngari—SE, Yaoundé—CM, La Lope Forest—GB, Ouagadougou—BF and Bijagos—GB) and EA populations (Lunyo—UG, Bundibugyo—UG, Kichwamba—UG and Kisumu—KE). We tested three scenarios (Figure S3) to determine the history of *Ae. aegypti* in continental Africa relative to Europa. Scenario 1: Europa populations gave rise to African populations; Scenario 2: WA gave rise to both Europa and EA; and Scenario 3: EA gave rise to both Europa and WA populations. Priors used for on the analysis are shown in Tables S14 and S15.

3 | RESULTS

We estimated the evolutionary history and divergence times of the members of the Aegypti Group in the southwestern Indian Ocean

(Figure 2; Figures S1 and S2) with nucleotide sequence data from seven markers. Our maximum-likelihood phylogenetic analyses on the full nucleotide data set, the mitochondrial-only data set and the nuclear-only data set differed in clade support values but overall recovered concordant topologies for the major lineages of *Stegomyia*, including monophyly of the Aegypti Group (Figure S1). Moreover, there was no correlation between missing data in the alignment and terminal branch length (for the number of loci in the alignment per species, $t_{31} = 0.12$, p -value >0.9 ; and for the proportion of missing data in the alignment per species, $t_{31} = -0.6$, p -value >0.5). As such, we performed divergence time estimation on the combined nuclear and mitochondrial nucleotide data set. The topology recovered from our divergence time estimate was similar to the likelihood analyses; thus, we report only the results from the divergence time analysis here (but see Figure S1).

Although sparsely sampled outside the Aegypti Group, these results (Figure 2; Figures S1 and S2) indicate a common ancestry for *Stegomyia* 52 MYA (95% HPD: 29–84 MYA) and African *Stegomyia* 39 MYA (95% HPD: 21–64 MYA). Our results support the monophyly of the Aegypti Group, and its affinity to other Afrotropic *Stegomyia* (Figure 2; Figure S1 and S2). Our estimates suggest a common ancestor of the Aegypti Group in the southwestern Indian Ocean 16 MYA (95% HPD: 7–28 MYA). Subsequent divergence times in the Aegypti Group suggest a divergence 7 MYA (95% HPD: 4–15 MYA) between *Ae. aegypti* on Madagascar, *Ae. mascarensis* and *Ae. aegypti* s.s., consistent with Fort et al. (2012). Of the three named species, *Ae. pia* is clearly the most distantly related and thus serves as an outgroup in later analyses.

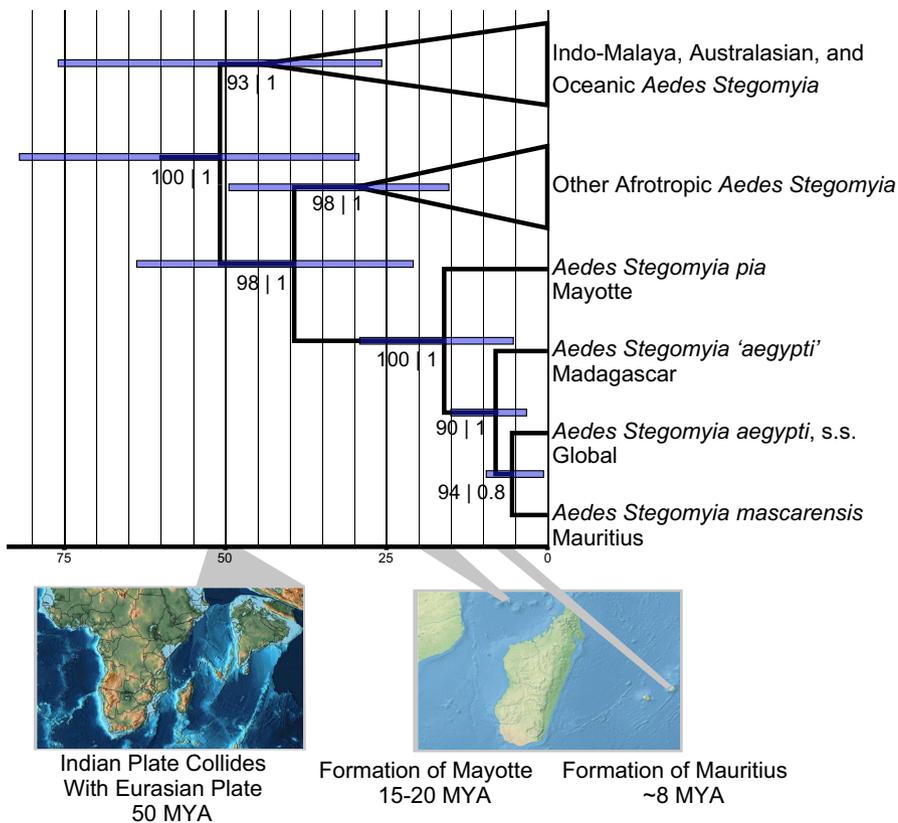


FIGURE 2 Divergence times in the Aegypti Group correlate with island formations for *Aedes pia* and *Aedes mascarensis*. Divergence times shown here based on the relaxed clock analysis in BEAST2 for members of the *Stegomyia* subgenus *Aedes*. Relationships among clades in the subgenus *Stegomyia* show here are concordant with the maximum-likelihood analyses performed, as well (see Figure S1). The first support value indicates is the ultrafast bootstrap support value from IQ-Tree, while the second is the posterior probability from BEAST2 [Colour figure can be viewed at wileyonlinelibrary.com]

In order to describe within- and between-species genetic variation in the Aegypti Group, we genotyped individuals from global populations of *Ae. aegypti* s.s. (Table S1), including samples from three islands and Madagascar in the southwestern Indian Ocean, and two species closely related to *Ae. aegypti* (*Ae. pia* and *Ae. mascarensis*), with the *Ae. aegypti* SNP chip (Evans et al., 2015). Following filtering (see Section 2), we retained 18,489 SNPs for subsequent analyses with an overall genotyping rate of 0.95. Among subspecies and populations, missing data were highest in *Ae. pia*, and lowest in populations of *Ae. aegypti* (see our discussion of missing data in the online supplemental materials). An analysis of molecular variation on our SNP data found significant genetic differentiation among subspecies Aaf and Aaa, island populations and species in the Aegypti Group (Table S4). Pairwise F_{st} values among these groups indicated that *Ae. pia* was the most distinct, with island populations having lower mean F_{st} values to *Ae. mascarensis* than to Aaa and Aaf (Tables S4–S6). These results are consistent with the PCA of the SNP data, which supports the uniqueness of these island populations as distinct from all other *Ae. aegypti* s.s. (Figure 3a). Population structure as assessed by the STRUCTURE-like SNMF was also consistent with the PCA and phylogenetic results (see Figure S4 for SNMF). At the optimal K determined by cross-entropy, $K = 3$, island populations had highest identity with the genetic clusters associated with *Ae. aegypti* from Madagascar, *Ae. mascarensis* and *Ae. pia*, than with the clusters associated with Aaa or Aaf. However, at this and higher K values (Figure S4), individuals from island populations frequently contained a minority of ancestry to ancestral populations associated with Aaa or Aaf, potentially indicating admixture.

A maximum-likelihood phylogenetic analysis of the SNP data indicated that all southwestern Indian Ocean populations are basal to *Ae. aegypti* s.s. (Figure 4). Consistent with the maximum-likelihood-based phylogenetic and Bayesian divergence time analyses above, *Ae. aegypti* from Madagascar resolved as basal to all other *Ae. aegypti* populations and *Ae. mascarensis*. The Réunion population of *Ae. aegypti* resolved as sister to *Ae. mascarensis*, while Mayotte and Europa occupied intermediate regions of the phylogenetic tree. *Ae. aegypti* s.s. formed a monophyletic group that includes all sampled Aaf populations from continental Africa and all Aaa populations from outside Africa.

Due to the aforementioned analyses indicating evidence of admixture in the southwestern Indian Ocean, we utilized F3 tests implemented in Treemix to formally test for potential admixture, with three separate analyses: the primary full SNP data set (18,489 SNPs), a data set with no missing data derived from this data set (5,866 SNPs) and a data set where populations of *Ae. aegypti* were subsampled to four individuals (and resampled 100 times). Since the results of our primary SNP data set achieved qualitatively the same conclusion to that of the other two data sets investigated, and because of the increased power of the larger sample size in terms of SNPs and number of individuals per population, we report the specific results from the primary data set here (but see supplemental tables). Tests of admixture (Figure 3b; Tables S7–S9) found strong evidence that three populations were the result of admixture: Mayotte, with evidence of admixture between *Aedes aegypti* from Madagascar and a population of Aaa, with Madagascar + Jeddah assigned the lowest Z score ($Z = -21.34$, $p = 5.58E-101$); Réunion, with evidence of admixture between *Aedes aegypti* from Réunion and a population of Aaa, with Réunion + Jeddah assigned the lowest Z score ($Z = -21.34$, $p = 5.58E-101$); and Mombasa, with evidence of admixture between *Aedes aegypti* from Mombasa and a population of Aaa, with Mombasa + Jeddah assigned the lowest Z score ($Z = -21.34$, $p = 5.58E-101$).

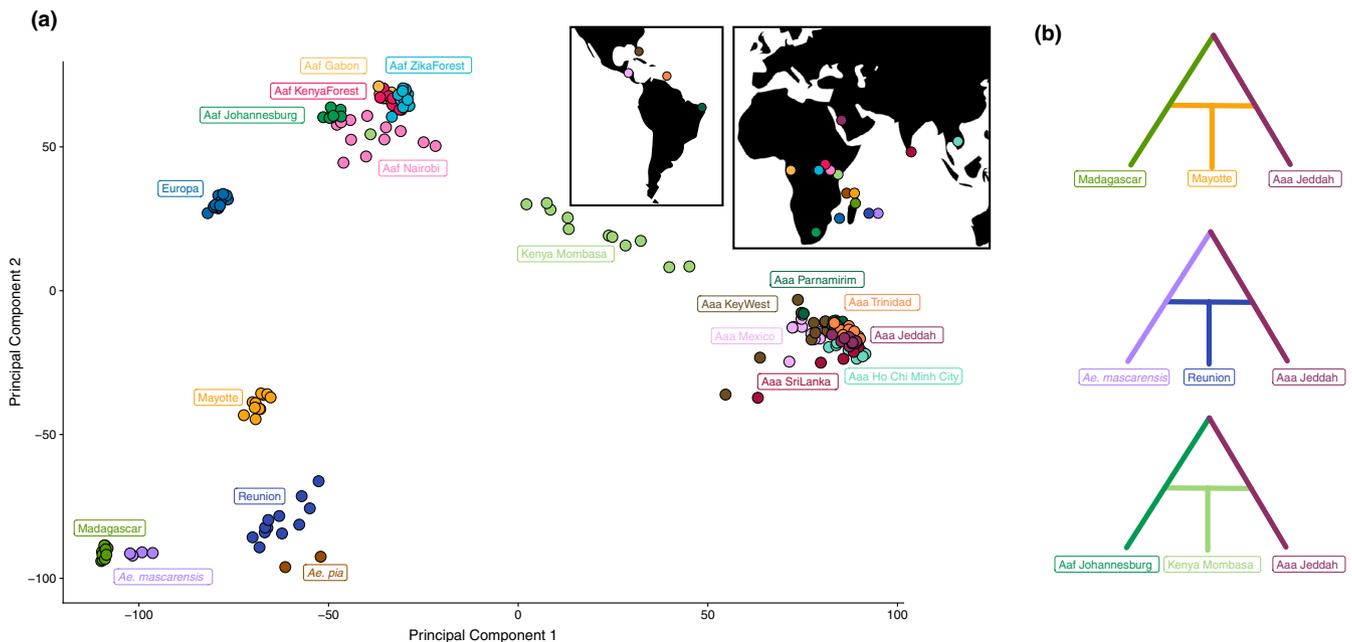


FIGURE 3 Genetic diversity in the Aegypti Group. (a) A PCA demonstrating that island populations cluster separately from the recognized subspecies *Aedes aegypti aegypti* (Aaa) and *Aedes aegypti formosus* (Aaf). (b) A visual depiction of TreeMix results demonstrating the most likely admixture scenarios involving the three admixed populations of Mayotte, Réunion and Mombasa in Kenya. In all three cases, the most likely scenario had one parental "donor" from Jeddah, although other Asian populations are possible, as well; see Sections 3 and 4 for additional details



FIGURE 4 The maximum-likelihood phylogeny estimated from Aegypti Group SNP data demonstrates the deep divergence of *Aedes aegypti* from Madagascar from populations of *Aedes aegypti* s.s. Narrow coloured vertical bars reflect species or population of origin. Black and grey bars reflect populations of *Aedes aegypti* s.s. assigned to one of the two subspecies *Aedes aegypti formosus* (Aaf) or *Aedes aegypti aegypti* (Aaa). The phylogeny was estimated in IQ-Tree and was rooted on the branch leading to *Aedes pia*. Support values are ultrafast bootstraps estimated in IQ-Tree, and branches with support values below 90 have been collapsed. Scale is in substitutions per site

with evidence of admixture between *Ae. mascarensis* and a population of Aaa, with *Ae. mascarensis* + Jeddah again assigned the lowest Z score ($Z = -10.09$, $p = 6.36E-24$); and Mombasa, the likely result of admixture between a population of Aaf and a population of Aaa, with Johannesburg + Jeddah assigned the lowest Z score ($Z = -41.77$, $p < 2E-200$). Both the filtered-for-completeness data set and the resampled subset of four *Ae. aegypti* individuals per population found the same likely admixed populations and same originating parent populations having the lowest F3 statistics as in the primary analysis. However, the F3 statistics estimated were

lower, concordant with the reduced power from fewer SNPs and/or fewer individuals per population from which to estimate allele frequencies.

Although Jeddah was a common contributor in the lowest F3 scenario across all analyses, all Asian populations tested were suggested as putative ancestors of these admixed populations (Tables S7, S8 and S19). As such, any Asian population in our analysis could also have been a source of admixture in this region. Unlike other island populations, we found no evidence of admixture within mosquito samples from Madagascar or Europa.

Next, we used microsatellite markers and approximate Bayesian computation (ABC) to confirm that the population of Europa, sister to *Ae. aegypti* s.s. in our maximum-likelihood analysis, diverged prior to continental *Aedes aegypti* diversifying within Africa, rather than representing a re-introduction from the continent. We used representative populations from East and West Africa to explicitly model whether populations in Europa diverged prior to, or after, the divergence of East and West African populations of *Aedes aegypti* on continental Africa (Kotsakiozi, Evans, et al., 2018). As our main focus was on populations in the southwestern Indian Ocean, and because the population structure of African *Ae. aegypti* has been recently reported from these microsatellites (Gloria-Soria et al., 2016), we report detailed population statistics from microsatellites in our supplement but summarize our descriptive statistics on the microsatellite data set briefly here. The main microsatellite data set used for this analysis included genotypes from 12 loci with 159 total alleles (Table S10). This data set contained individuals from six populations from West Africa and four from East Africa, plus the Europa population, and had 1.02% missing data (Table S11). Population genetic statistics for individual locations are reported in Table S12. The average observed heterozygosity (H_o) across the data set was 0.567, with $H_o = 0.458$ in Europa, $H_o = 0.579$ in West Africa and $H_o = 0.577$ in East Africa; see Table S12. A total of 14 out of 143 (9.7%) population-by-locus comparisons deviate significantly from Hardy-Weinberg expectations, after sequential Bonferroni correction (Table S13). The ABC analysis on the African and Europa Island samples of *Ae. aegypti* supports a scenario where the colonization of Africa occurred from the Indian Ocean (represented by Europa) less than 85,000 years ago (Mean: 36,500 YA, 95% CI across runs: 9,780–84,300 YA; see Tables S14 and S15), with a predicted split between West and East Africa no older than 50,000 years ago (mean: 21,000 YA, 95% CI across runs: 7,810–49,400 YA; see Tables S14 and S15), assuming an average of 10 generations per year (Scenario 1: $p = .8864$ and $p = .8481$ in Tables S14 and S15; Figure S3). Alternative scenarios were poorly supported by the analysis ($p < .11$; Tables S14 and S15; Figure S3). The estimated mutation rate under the best-fit scenario was $\sim 9.5 \times 10^{-6}$, and falls within the range of microsatellite mutation rates estimated for other Diptera (Pfeiler et al., 2013; Schug et al., 1998).

4 | DISCUSSION

4.1 | The Origin of the Aegypti Group and *Ae. aegypti* s.s

Research in the evolutionary history of this medically important group of mosquitoes has long focused on continental Africa and the invasive range throughout the subtropics and tropics, where most epidemics of the diseases transmitted by *Ae. aegypti* have occurred. But to fully understand the history of this mosquito, we need to consider its relationship to other members of the subgenus *Stegomyia* and, in particular, other members of the Aegypti Group.

Our analysis (Figure 2; Figures S1 and S2) found two well-supported clades of *Stegomyia*, corresponding to an Afrotropic lineage and an Indo-Malaya/Oceanic lineage (following the biogeographic regions proposed in Olson et al., 2001), with a last common ancestor approximately 50 MYA, although the large HPD intervals on our divergence estimates indicate a high degree of uncertainty on this estimate. The estimated age is consistent with the time period in which the Asian and Indian plates collided. Given the distribution of many *Stegomyia* species around the Indian Ocean, it is tempting to associate this geologic event with the subsequent distribution of *Stegomyia*, but our limited sampling and large intervals around divergence time estimates preclude strong conclusions on the biogeographic origin of the subgenus as a whole. Our reliance on uniform prior densities for fossil calibrations, due to the relatively poor record of fossil mosquitoes, uncertainty regarding the age of mosquito clades and the limited sequence data at hand to place additional fossils, likely resulted in large intervals around our divergence time estimates. Nevertheless, our analyses indicate that one lineage of Afrotropic *Stegomyia* diversified in Africa (leading to *Ae. africanus*, *Ae. simpsoni*, and others; see Figure 2). Another lineage, which entered the islands of the southwestern Indian Ocean more than 16 MYA, diversified there and gave rise to the Aegypti Group, before entering continental Africa as the *Ae. aegypti* s.s. lineage.

Our analyses on the divergence times of species in the Aegypti Group on Indian Ocean islands (Figure 2) correspond well with the ages of the islands on which these species are found, as the age of Mayotte (where *Ae. pia* is endemic) is less than 20 MY (Michon, 2016) and Mauritius (where *Ae. mascarensis* is endemic) is less than 9 MY (Ashwal, Wiedenbeck, & Torsvik, 2017; Warren, Bermingham, Bowie, Prys-Jones, & Thébaud, 2003). Our analyses support that ancestral populations in the Aegypti Group on Madagascar diverged prior to the divergence of *Ae. mascarensis* on Mauritius, east of Madagascar, a result consistent with previous single-marker analysis with limited taxon sampling that found populations in the Aegypti Group on Madagascar were basal to continental African *Aedes aegypti* (Delatte et al., 2011). This observation is also consistent with the maximum-likelihood analysis on the primary SNP data set reported here, although branch lengths from our SNP-based likelihood analysis may be underestimating variation in more distant species and populations, as the SNP chip was designed to capture variation within *Ae. aegypti*. Nonetheless, the concordance between the analyses performed on our nucleotide sequence data set, our primary SNP dataset and the microsatellite estimation of population history (see below) indicates that the common ancestor of the Aegypti Group was most likely found throughout the islands in the southwestern Indian Ocean.

Our results suggest that *Ae. aegypti* s.s. entered Africa less than 85,000 years ago, as the divergence of Europa and continental African populations was younger than 85,000 years across all five ABC analyses (Figure S3; Tables S14 and S15). Pollen analyses of cores from Lake Malawi in East Africa at about the same latitude as Madagascar suggest frequent fluctuations in dry- and wet-associated flora until about 80,000 years ago when present-day vegetation

stabilized (Ivory, Lézine, Vincens, & Cohen, 2018; Lyons et al., 2015). This is consistent with the unusually high and stable levels that Lake Malawi experienced over the last 70,000 years, indicative of a period of steady rain (Ivory et al., 2018; Lyons et al., 2015) that would have provided a suitable habitat typical of Aaf on continental Africa.

Interestingly, our F3 tests did not indicate admixture has occurred in *Ae. aegypti* from Europa, suggesting that the apparent admixture signatures detected by the STRUCTURE-like SNMF in this population (Figure S4) may represent instead ancestral polymorphism for this lineage. If true, *Ae. aegypti* from Europa could be a remnant of the island-dwelling lineage that originally colonized continental Africa, although testing this hypothesis requires further sampling and analysis. Moreover, we estimated a strikingly early last common ancestor for continental *Ae. aegypti* s.s. in Africa, at approximately 17,000 to 25,000 years ago. This time period coincides with the end of the last glacial maximum (Clark et al., 2009), where stable Pleistocene forest refugia along the east African coast (Finch, Leng, & Marchant, 2009; Fjeldsaa & Lovett, 1997; Mumbi, Marchant, Hooghiemstra, & Wooller, 2008) may have provided suitable habitat for *Ae. aegypti* to more widely disperse across Africa. Following the end of the last glacial maximum, forests spread across Africa and reached their extent approximately 6,000 years ago, covering most of central Africa (Hoelzmann et al., 1998; Watrin, Lézine, & Hély, 2009) and enabling dispersal westward. Our estimates for the divergence of East and West African populations of *Ae. aegypti* are consistent with Bennett et al. (2016) who presented molecular data suggesting the deepest divergences of populations of *Aedes aegypti* on continental Africa occurred between 6,000 and 120,000 years ago. Much more recently, the continental *Ae. aegypti* s.s. lineage separated into the two recognized subspecies *Ae. aegypti aegypti* (Aaa) and *Ae. aegypti formosus* (Aaf). This has been speculated to have begun when human settlements developed in sub-Saharan Africa and water was stored during the dry season. The full separation of subspecies occurred ~500 years when European slave trade brought *Ae. aegypti* s.s. to the New World beginning in the 16th century (Powell et al., 2018).

4.2 | A divergent lineage of *Aedes aegypti* on Madagascar

We find strong evidence that *Ae. aegypti* on Madagascar are genetically distinct, and deeply diverged from *Ae. aegypti* s.s., at a level equal to or greater than what is seen in the named species *Ae. mascarensis* (Figures 2 and 3). A previous study, including only putative *Ae. aegypti* samples, suggested Madagascar *Ae. aegypti* separated from Europa and continental populations 7–15 MYA (Fort et al., 2012), similar to our estimate (Figure 2). Phylogenetic analyses using mtDNA from mosquitoes from northern Madagascar that were morphologically identified as Aaf indicated that these mosquitoes were genetically distinct from other populations of *Ae. aegypti* (Mousson et al., 2005), and basal to *Ae. mascarensis*, consistent with our results. To date, no phenotypic differences are known between *Ae.*

aegypti from Madagascar and *Ae. aegypti* s.s., but ecological observations of *Ae. aegypti* on Madagascar suggest a sylvan mosquito similar in appearance and behaviour to Aaf (Fontenille & Rodhain, 1989; Raharimalala et al., 2012), with differences in vector competence relative to *Ae. aegypti* s.s. (Failloux et al., 2002; Vazeille et al., 2001). A cryptic lineage—particularly one that has shown some evidence of differences in vector competence relative to *Ae. aegypti* s.s.—may present a unique challenge to control efforts. Surveillance efforts may need to account for divergent lineage, whose risk to public health may be significantly different from *Ae. aegypti* s.s., made doubly hard by the cryptic nature of this taxa and its largely unknown ecology. However, in the absence of tests for reproductive isolation and detailed morphological analysis, it seems premature at this time to formally describe a new species. Furthermore, it is important to note that the sample from Madagascar we analysed came from a single locality. Madagascar is well known as a biodiversity hot spot, and thus, it is conceivable, maybe even likely, that other cryptic taxa in the Aegypti Group occur on Madagascar. This is especially true as two sources of molecular data (Fort et al. and nucleotide data from this study) indicate these populations have been on Madagascar for >7 million years.

4.3 | A recent history of introgression in *Ae. aegypti*

On the islands in the southwestern Indian Ocean, we find evidence for recent introgression among members of the Aegypti Group. These results indicate a component of the admixture is Aaa from Asia, rather than from New World Aaa or from Aaf. In our analysis, Aaa from Saudi Arabia was most often identified as the Aaa parental population in admixture analyses, but we note that the inference of ancestral population origins with Treemix is limited to the populations used in the analysis, and thus, future sampling may yield different results; additionally, all Asian populations we tested had similar strength of support as Saudi Arabia to be the origin of the parental Aaa involved in introgression in these island populations (Tables S13–S15). This inference is consistent with a hypothesis of Aaa expanding eastward after the Suez Canal opened in 1869 to colonize Asia (Kotsakiozi, Gloria-Soria, Schaffner, Robert, & Powell, 2018; Powell et al., 2018). Here, we infer they also entered islands east of Africa where they encountered other species in the Aegypti Group and, in some cases, interbred. It is also possible due to the small sample sizes in some populations (e.g. *Ae. mascarensis* or *Ae. aegypti* from Sri Lanka) we have underestimated or missed potential admixture events, as the relative F3 statistics inferred from our analysis of resampled populations were slightly lower than those from the full SNP data set. However, our resampling provided qualitatively the same results as our full data set. This indicates that small sample sizes can detect admixture that is consistent with observations from larger samples of the same population, at least given the relatively large number of SNPs in our study.

The consequences of this introgression remain to be tested. However, some *Ae. aegypti* on Réunion exhibit a particular scutal

morph that appears intermediate between *Ae. aegypti* and *Ae. mascarensis* (Le Goff et al., in preparation). Other *Ae. aegypti* from Réunion—differentiated from Aaf in scaling patterns—have been reported to be primarily a sylvatic mosquito (Bagny, Delatte, Elissa, Quilici, & Fontenille, 2009; Salvan & Mouchet, 1994); that is, most are morphologically Aaa but ecologically Aaf. Previous studies speculated that the presence of Aaa in rural locations on Réunion, rather than in urban locations typical of Aaa, was due to the widespread use of pesticides, and larval competition with *Ae. albopictus* (Bagny et al., 2009). We propose here that admixture with *Ae. mascarensis*, or a mascarensis-like population, could have resulted in behavioural shifts due to the sylvatic nature of *Ae. mascarensis*, while the morphology typical of Aaa was maintained. Admixture may complicate efforts to control *Ae. aegypti* by shifting typical behaviours associated with this mosquito in a way that may allow them to evade control measures that are typically effective against them. Vector control efforts in this region thus may need to account for not only a cryptic species, but also admixed populations, and will likely require more detailed information on ecological differences between species and populations to be successful. Moreover, the primary vector of the chikungunya virus during the outbreak of 2005/6 on Réunion was attributed to *Ae. albopictus* (Delatte et al., 2008), rather than *Ae. aegypti*, the more competent vector present throughout the rest of the world. Does this mean *Ae. aegypti* on Réunion is less competent than most *Ae. aegypti* populations for chikungunya? While some information is available for Madagascar (Failloux et al., 2002; Vazeille et al., 2001), the vector competence of other members of the Aegypti Group on southwestern Indian Ocean islands is largely unknown.

5 | CONCLUSION

Ae. aegypti is among the best-studied mosquitoes, yet from an evolutionary standpoint, the results presented here highlight how much more there is still to be learned. We present the first solid evidence that the common ancestor *Ae. aegypti* s.s. and its nearest relatives were occupying Madagascar and/or islands in the southwestern Indian Ocean, indicating that these islands gave rise only relatively recently (<85,000 YA) to *Ae. aegypti* s.s. on continental Africa. Moreover, we find strong evidence of a cryptic genetic lineage on Madagascar. It is also clear, however, that while our limited sampling in this region of the world has revealed important insights, more intense sampling and analyses of collections are likely to reveal a more complete picture of the evolutionary history of the Aegypti Group.

ACKNOWLEDGEMENTS

We thank T. Chiodo for laboratory assistance and processing of some samples used in this study. We also thank J. Baltzegar and F. Gould for comments on an earlier version of this manuscript. Jessica Thompson and Roderick McIntosh provided insights into the Pleistocene climate of Africa. The funding for the SNP chip

used in this study came from NIH award RO1 AI101112. JS was partially funded on NSF DEB #1754376 and AG-S by the Grant or Cooperative Agreement Number U01CK000509, funded by the Centers for Disease Control and Prevention. The publication contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

AUTHOR CONTRIBUTIONS

JS, AGS and JRP designed research. JS, AGS, VR and GLG performed research. AGS, VR, GLG, ABF and JRP contributed samples. JS and AGS analysed data. JS, AGS, VR, GLG, ABF and JRP wrote the manuscript.

DATA AVAILABILITY STATEMENT

The nucleotide alignment in Nexus format is available at FigShare.com, <https://doi.org/10.6084/m9.figshare.12476189>. VCF file containing SNP calls is available at VectorBase.org, Population Biology Project ID: VBP0000672. Microsatellite raw allele frequencies are available at VectorBase.org, Population Biology Project ID: VBP0000672.

[dataset] Soghigian, Gloria-Soria, Robert, Le Goff, Failloux, and Powell; 2020; VCF SNP calls for *Ae. aegypti* and related species from the southwestern Indian Ocean; Vector Base; Population Biology Project identifier VBP0000672.

[dataset] Soghigian, Gloria-Soria, Robert, Le Goff, Failloux, and Powell; 2020; Raw allele frequencies for *Ae. aegypti* from the southwestern Indian Ocean; Vector Base; Population Biology Project VBP0000672.

[dataset] Soghigian, Gloria-Soria, Robert, Le Goff, Failloux, and Powell; 2020; Nucleotide alignment for *Ae. aegypti* and other Culicidae; Fig Share; <https://doi.org/10.6084/m9.figshare.12476189>.

ORCID

John Soghigian  <https://orcid.org/0000-0003-4469-2072>

Andrea Gloria-Soria  <https://orcid.org/0000-0002-5401-3988>

REFERENCES

- Ashwal, L. D., Wiedenbeck, M., & Torsvik, T. H. (2017). Archaean zircons in Miocene oceanic hotspot rocks establish ancient continental crust beneath Mauritius. *Nature Communications*, 8, 14086. <https://doi.org/10.1038/ncomms14086>
- Bagny, L., Delatte, H., Elissa, N., Quilici, S., & Fontenille, D. (2009). *Aedes* (Diptera: Culicidae) vectors of arboviruses in Mayotte (Indian Ocean): Distribution area and larval habitats. *Journal of Medical Entomology*, 46(2), 198–207.
- Barido-Sottani, J., Bošková, V., Plessis, L. D., Kühnert, D., Magnus, C., Mitov, V., ... Stadler, T. (2018). Taming the BEAST—A community teaching material resource for BEAST 2. *Systematic Biology*, 67(1), 170–174. <https://doi.org/10.1093/sysbio/syx060>
- Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, 162(4), 2025–2035.
- Bennett, K. L., Shija, F., Linton, Y.-M., Misinzo, G., Kaddumukasa, M., Djouaka, R., ... Walton, C. (2016). Historical environmental change in Africa drives divergence and admixture of *Aedes aegypti* mosquitoes:

- A precursor to successful worldwide colonization? *Molecular Ecology*, 25(17), 4337–4354. <https://doi.org/10.1111/mec.13762>
- Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42. <https://doi.org/10.1186/s12862-017-0890-6>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Braks, M. A. H., Honório, N. A., Lounibos, L. P., Lourenço-de-Oliveira, R., & Juliano, S. A. (2004). Interspecific competition between two invasive species of container mosquitoes, *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae), in Brazil. *Annals of the Entomological Society of America*, 97(1), 130–139.
- Brown, J. E., Evans, B. R., Zheng, W., Obas, V., Barrera-Martinez, L., Egizi, A., ... Powell, J. R. (2014). Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution; International Journal of Organic Evolution*, 68(2), 514–525. <https://doi.org/10.1111/evo.12281>
- Brown, J. E., McBride, C. S., Petrina, J., Scott, R., Christophe, P., Hervé, B., ... Powell, J. R. (2011). Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. *Proceedings of the Royal Society B: Biological Sciences*, 278(1717), 2446–2454. <https://doi.org/10.1098/rspb.2010.2469>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 7. <https://doi.org/10.1186/s13742-015-0047-8>
- Christophers, R. (1960). *Aedes aegypti: The yellow fever mosquito*. Cambridge, UK: Cambridge University Press.
- Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., ... McCabe, A. M. (2009). The last glacial maximum. *Science*, 325(5941), 710–714. <https://doi.org/10.1126/science.1172873>
- Clements, A. N. (1999a). *The biology of mosquitoes, volume 1: Development, nutrition and reproduction*. New York, NY: CABI.
- Clements, A. N. (1999b). *The biology of mosquitoes, volume 2: Sensory reception and behaviour*. Chapman & Hall.
- Clements, A. N. (2012). *The biology of mosquitoes, volume 3: Transmission of viruses and interactions with bacteria*. Cambridge, UK: CABI.
- Cook, S., Diallo, M., Sall, A. A., Cooper, A., & Holmes, E. C. (2005). Mitochondrial markers for molecular identification of *Aedes* mosquitoes (Diptera: Culicidae) involved in transmission of arboviral disease in West Africa. *Journal of Medical Entomology*, 42(1), 19–28. <https://doi.org/10.1093/jmedent/42.1.19>
- Cornuet, J.-M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... Estoup, A. (2014). DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, 30(8), 1187–1189. <https://doi.org/10.1093/bioinformatics/btt763>
- Crawford, J. E., Alves, J. M., Palmer, W. J., Day, J. P., Sylla, M., Ramasamy, R., ... Jiggins, F. M. (2017). Population genomics reveals that an anthropophilic population of *Aedes aegypti* mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biology*, 15(1), 16. <https://doi.org/10.1186/s12915-017-0351-0>
- Delatte, H., Bagny, L., Brengue, C., Bouetard, A., Paupy, C., & Fontenille, D. (2011). The invaders: Phylogeography of dengue and chikungunya viruses *Aedes* vectors, on the South West islands of the Indian Ocean. *Infection, Genetics and Evolution*, 11(7), 1769–1781. <https://doi.org/10.1016/j.meegid.2011.07.016>
- Delatte, H., Paupy, C., Dehecq, J. S., Thiria, J., Failloux, A. B., & Fontenille, D. (2008). *Aedes albopictus*, vecteur des virus du chikungunya et de la dengue à la Réunion: Biologie et contrôle. *Parasite*, 15(1), 3–13. <https://doi.org/10.1051/parasite/2008151003>
- Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., & Wilson, A. (2012). Geneious version 5.6. Retrieved from <http://www.geneious.com>
- Evans, B. R., Gloria-Soria, A., Hou, L., McBride, C., Bonizzoni, M., Zhao, H., & Powell, J. R. (2015). A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3 (Bethesda)*, 5(5), 711–718. <https://doi.org/10.1534/g3.114.016196>
- Failloux, A.-B., Vazeille, M., & Rodhain, F. (2002). Geographic genetic variation in populations of the dengue virus vector *Aedes aegypti*. *Journal of Molecular Evolution*, 55(6), 653–663. <https://doi.org/10.1007/s00239-002-2360-y>
- Finch, J., Leng, M. J., & Marchant, R. (2009). Late Quaternary vegetation dynamics in a biodiversity hotspot, the Uluguru Mountains of Tanzania. *Quaternary Research*, 72(1), 111–122. <https://doi.org/10.1016/j.yqres.2009.02.005>
- Fjeldsaa, J., & Lovett, J. C. (1997). Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodiversity & Conservation*, 6(3), 325–346. <https://doi.org/10.1023/A:1018356506390>
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.
- Fontenille, D., & Rodhain, F. (1989). Biology and distribution of *Aedes albopictus* and *Aedes aegypti* in Madagascar. *Journal of the American Mosquito Control Association*, 5(2), 219–225.
- Fort, P., Albertini, A., Van-Hua, A., Berthomieu, A., Roche, S., Delsuc, F., ... Weill, M. (2012). Fossil rhabdoviral sequences integrated into arthropod genomes: Ontogeny, evolution, and potential functionality. *Molecular Biology and Evolution*, 29(1), 381–390. <https://doi.org/10.1093/molbev/msr226>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Giraldo-Calderón, G. I., Emrich, S. J., MacCallum, R. M., Maslen, G., Dialynas, E., Topalis, P., ... Lawson, D. (2015). VectorBase: An updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Research*, 43(Database issue), D707–D713. <https://doi.org/10.1093/nar/gku1117>
- Gloria-Soria, A., Ayala, D., Bheecarry, A., Calderon-Arguedas, O., Chadee, D. D., Chiappero, M., ... Powell, J. R. (2016). Global genetic diversity of *Aedes aegypti*. *Molecular Ecology*, 25(21), 5377–5395. <https://doi.org/10.1111/mec.13866>
- Gloria-Soria, A., Lima, A., Lovin, D. D., Cunningham, J. M., Severson, D. W., & Powell, J. R. (2018). Origin of a high-latitude population of *Aedes aegypti* in Washington, DC. *The American Journal of Tropical Medicine and Hygiene*, 98(2), 445–452. <https://doi.org/10.4269/ajtmh.17-0676>
- Hartberg, W. K., & Craig, G. B. (1970). Reproductive isolation in stegomyia mosquitoes. II. Hybrid breakdown between *Aedes aegypti* and *A. mascarensis*. *Evolution*, 24(4), 692–703. <https://doi.org/10.2307/2406550>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBboot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Hoelzmann, P., Jolly, D., Harrison, S. P., Laarif, F., Bonnefille, R., & Pachur, H.-J. (1998). Mid-Holocene land-surface conditions in northern Africa and the Arabian Peninsula: A data set for the analysis of biogeophysical feedbacks in the climate system. *Global Biogeochemical Cycles*, 12(1), 35–51. <https://doi.org/10.1029/97GB02733>
- Huang, Y.-M. (2004). The subgenus *Stegomyia* of *Aedes* in the afro-tropical region with keys to the species (Diptera: Culicidae). *Zootaxa*, 700(1), 1–120. <https://doi.org/10.11646/zootaxa.700.1.1>

- Ivory, S. J., Lézine, A.-M., Vincens, A., & Cohen, A. S. (2018). Waxing and waning of forests: Late quaternary biogeography of southeast Africa. *Global Change Biology*, 24(7), 2939–2951. <https://doi.org/10.1111/gcb.14150>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281. <https://doi.org/10.7717/peerj.281>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kotsakiozi, P., Evans, B. R., Gloria-Soria, A., Kamgang, B., Mayanja, M., Lutwama, J., ... Powell, J. R. (2018). Population structure of a vector of human diseases: *Aedes aegypti* in its ancestral range, Africa. *Ecology and Evolution*, 8(16), 7835–7848. <https://doi.org/10.1002/ece3.4278>
- Kotsakiozi, P., Gloria-Soria, A., Schaffner, F., Robert, V., & Powell, J. R. (2018). *Aedes aegypti* in the Black Sea: Recent introduction or ancient remnant? *Parasites & Vectors*, 11. <https://doi.org/10.1186/s13071-018-2933-2>
- Le Goff, G., Brengues, C., & Robert, V. (2013). *Stegomyia* mosquitoes in Mayotte, taxonomic study and description of *Stegomyia pia* n. sp. *Parasite*, 20, 31. <https://doi.org/10.1051/parasite/2013030>
- Lounibos, L. P. (1981). Habitat segregation among African treehole mosquitoes. *Ecological Entomology*, 6(2), 129–154. <https://doi.org/10.1111/j.1365-2311.1981.tb00601.x>
- Lyons, R. P., Scholz, C. A., Cohen, A. S., King, J. W., Brown, E. T., Ivory, S. J., ... Blome, M. W. (2015). Continuous 1.3-million-year record of East African hydroclimate, and implications for patterns of evolution and biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51), 15568–15573. <https://doi.org/10.1073/pnas.1512864112>
- MacGregor, M. E. (1924). *Aedes* (*Stegomyia*) *mascaensis*, MacGregor: A new Mosquito from Mauritius. *Bulletin of Entomological Research*, 14(4), 409–412. <https://doi.org/10.1017/S0007485300045648>
- Matthews, B. J., Dudchenko, O., Kingan, S. B., Koren, S., Antoshechkin, I., Crawford, J. E., ... Voshall, L. B. (2018). Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*, 563(7732), 501. <https://doi.org/10.1038/s41586-018-0692-z>
- McBride, C. S., Baier, F., Omondi, A. B., Spitzer, S. A., Lutomiya, J., Sang, R., ... Voshall, L. B. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*, 515(7526), 222–227. <https://doi.org/10.1038/nature13964>
- McNeill, J. R. (2010). *Mosquito Empires: Ecology and War in the Greater Caribbean, 1620–1914*. New York, NY: Cambridge University Press.
- Meirmans, P. G., & Tienderen, P. H. V. (2004). genotype and genotype: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>
- Michon, L. (2016). The volcanism of the Comoros archipelago integrated at a regional scale. In P. Bachelery, J.-F. Lenat, A. Di Muro, & L. Michon (Eds.), *Active Volcanoes of the Southwest Indian Ocean: Piton de la Fournaise and Karthala* (pp. 333–344). https://doi.org/10.1007/978-3-642-31395-0_21
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES science gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE)*, 2010, 1–8. <https://doi.org/10.1109/GCE.2010.5676129>
- Mousson, L., Dauga, C., Garrigues, T., Schaffner, F., Vazeille, M., & Failloux, A. B. (2005). Phylogeography of *Aedes* (*Stegomyia*) *aegypti* (L.) and *Aedes* (*Stegomyia*) *albopictus* (Skuse) (Diptera: Culicidae) based on mitochondrial DNA variations. *Genetical Research*, 86(1), 1–11.
- Mumbi, C. T., Marchant, R., Hooghiemstra, H., & Wooller, M. J. (2008). Late Quaternary vegetation reconstruction from the Eastern Arc Mountains. *Tanzania. Quaternary Research*, 69(2), 326–341. <https://doi.org/10.1016/j.yqres.2007.10.012>
- Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, 17(6), 1136–1147. <https://doi.org/10.1111/1755-0998.12654>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial ecoregions of the world: A New Map of Life on Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11), 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051\[0933:TEOTWA\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2)
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/journal.pgen.0020190>
- Paupy, C., Vazeille-Falcoz, M., Mousson, L., Rodhain, F., & Failloux, A. B. (2000). *Aedes aegypti* in Tahiti and Moorea (French Polynesia): Isoenzyme differentiation in the mosquito population according to human population density. *The American Journal of Tropical Medicine and Hygiene*, 62(2), 217–224. <https://doi.org/10.4269/ajtmh.2000.62.217>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, 13(5), 946–952. <https://doi.org/10.1111/1755-0998.12129>
- Pfeiler, E., Flores-López, C. A., Mada-Vélez, J. G., Escalante-Verdugo, J., & Markow, T. A. (2013). Genetic diversity and population genetics of mosquitoes (Diptera: Culicidae: *Culex* spp.) from the Sonoran Desert of North America. *The Scientific World Journal*, 2013. <https://doi.org/10.1155/2013/724609>
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, 8(11), e1002967. <https://doi.org/10.1371/journal.pgen.1002967>
- Powell, J. R. (2016). Mosquitoes: New contender for most lethal animal. *Nature*, 540(7634), 525. <https://doi.org/10.1038/540525c>
- Powell, J. R., Gloria-Soria, A., & Kotsakiozi, P. (2018). Recent history of *Aedes aegypti*: Vector genomics and epidemiology records. *BioScience*, 68(11), 854–860. <https://doi.org/10.1093/biosci/biy119>
- Powell, J. R., & Tabachnick, W. J. (2013). History of domestication and spread of *Aedes aegypti* – a review. *Memórias do Instituto Oswaldo Cruz*, 108(Suppl 1), 11–17. <https://doi.org/10.1590/0074-0276130395>
- Pyron, R. A., Burbrink, F. T., & Wiens, J. J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology*, 13, 93. <https://doi.org/10.1186/1471-2148-13-93>
- R Core Team. (2018). *R: A language and environment for statistical computing*. Retrieved from <http://www.R-project.org/>
- Raharimalala, F. N., Ravaomanarivo, L. H., Ravelonandro, P., Rafarasoa, L. S., Zouache, K., Tran-Van, V., ... Mavingui, P. (2012). Biogeography of the two major arbovirus mosquito vectors, *Aedes aegypti* and *Aedes*

- albopictus* (Diptera, Culicidae). *Madagascar. Parasites & Vectors*, 5(1), 56. <https://doi.org/10.1186/1756-3305-5-56>
- Rašić, G., Filipović, I., Weeks, A. R., & Hoffmann, A. A. (2014). Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*, 15(1), 275. <https://doi.org/10.1186/1471-2164-15-275>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489–494. <https://doi.org/10.1038/nature08365>
- Reidenbach, K. R., Cook, S., Bertone, M. A., Harbach, R. E., Wiegmann, B. M., & Besansky, N. J. (2009). Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. *BMC Evolutionary Biology*, 9(1), 298. <https://doi.org/10.1186/1471-2148-9-298>
- Roure, B., Baurain, D., & Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, 30(1), 197–214. <https://doi.org/10.1093/molbev/mss208>
- Salvan, M., & Mouchet, J. (1994). *Aedes albopictus* et *Aedes aegypti* à l'île de La Réunion. *Annales De La Societe Belge De Medecine Tropicale*, 74(4), 323–326.
- Schug, M. D., Hutter, C. M., Wetterstrand, K. A., Gaudette, M. S., Mackay, T. F., & Aquadro, C. F. (1998). The mutation rates of di-, tri- and tetra-nucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution*, 15(12), 1751–1760. <https://doi.org/10.1093/oxfordjournals.molbev.a025901>
- Soghigian, J., Andreadis, T. G., & Livdahl, T. P. (2017). From ground pools to treeholes: Convergent evolution of habitat and phenotype in *Aedes* mosquitoes. *BMC Evolutionary Biology*, 17(1), 262. <https://doi.org/10.1186/s12862-017-1092-y>
- Tabachnick, W. J., & Powell, J. R. (1979). A world-wide survey of genetic variation in the yellow fever mosquito, *Aedes aegypti*. *Genetical Research*, 34(3), 215–229.
- Vazeille, M., Mousson, L., Rakatoarivony, I., Villeret, R., Rodhain, F., Duchemin, J. B., & Failloux, A. B. (2001). Population genetic structure and competence as a vector for dengue type 2 virus of *Aedes aegypti* and *Aedes albopictus* from Madagascar. *The American Journal of Tropical Medicine and Hygiene*, 65(5), 491–497. <https://doi.org/10.4269/ajtmh.2001.65.491>
- Warren, B. H., Bermingham, E., Bowie, R. C. K., Prys-Jones, R. P., & Thébaud, C. (2003). Molecular phylogeography reveals island colonization history and diversification of western Indian Ocean sunbirds (Nectarinia: Nectariniidae). *Molecular Phylogenetics and Evolution*, 29(1), 67–85. [https://doi.org/10.1016/S1055-7903\(03\)00063-0](https://doi.org/10.1016/S1055-7903(03)00063-0)
- Watrin, J., Lézine, A.-M., & Hély, C. (2009). Plant migration and plant communities at the time of the “green Sahara”. *Comptes Rendus Geoscience*, 341(8), 656–670. <https://doi.org/10.1016/j.crte.2009.06.007>
- Wesson, D. M., Porter, C. H., & Collins, F. H. (1992). Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). *Molecular Phylogenetics and Evolution*, 1(4), 253–269. [https://doi.org/10.1016/1055-7903\(92\)90001-W](https://doi.org/10.1016/1055-7903(92)90001-W)
- Wiens, J. J., & Morrill, M. C. (2011). Missing data in phylogenetic analysis: Reconciling results from simulations and empirical data. *Systematic Biology*, 60(5), 719–731. <https://doi.org/10.1093/sysbio/syr025>
- Wiens, J. J., & Tiu, J. (2012). Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS ONE*, 7(8), e42925. <https://doi.org/10.1371/journal.pone.0042925>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Soghigian J, Gloria-Soria A, Robert V, Le Goff G, Failloux A-B, Powell JR. Genetic evidence for the origin of *Aedes aegypti*, the yellow fever mosquito, in the southwestern Indian Ocean. *Mol Ecol*. 2020;29:3593–3606. <https://doi.org/10.1111/mec.15590>