PLOS GENETICS

RESEARCH ARTICLE

Mosquito genomes are frequently invaded by transposable elements through horizontal transfer

Elverson Soares de Melo (ID), **Gabriel Luz Wallau** (ID)*

Department of Entomology, Aggeu Magalhães Institute–Oswaldo Cruz Foundation (Fiocruz), Recife, Pernambuco, Brazil

* gabriel.wallau@cpqam.fiocruz.br

Abstract

Transposable elements (TEs) are mobile genetic elements that parasitize basically all eukaryotic species genomes. Due to their complexity, an in-depth TE characterization is only available for a handful of model organisms. In the present study, we performed a *de novo* and homology-based characterization of TEs in the genomes of 24 mosquito species and investigated their mode of inheritance. More than 40% of the genome of *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* is composed of TEs, while it varied substantially among *Anopheles* species (0.13%–19.55%). Class I TEs are the most abundant among mosquitoes and at least 24 TE superfamilies were found. Interestingly, TEs have been extensively exchanged by horizontal transfer (172 TE families of 16 different superfamilies) among mosquitoes in the last 30 million years. Horizontally transferred TEs represents around 7% of the genome in *Aedes* species and a small fraction in *Anopheles* genomes. Most of these horizontally transferred TEs are from the three ubiquitous LTR superfamilies: Gypsy, Bel-Pao and Copia. Searching more than 32,000 genomes, we also uncovered transfers between mosquitoes and two different Phyla—Cnidaria and Nematoda—and two subphyla—Chelicerata and Crustacea, identifying a vector, the worm *Wuchereria bancrofti*, that enabled the horizontal spread of a Tc1-mariner element among various *Anopheles* species. These data also allowed us to reconstruct the horizontal transfer network of this TE involving more than 40 species. In summary, our results suggest that TEs are frequently exchanged by horizontal transfers among mosquitoes, influencing mosquito's genome size and variability.

Author summary

A large fraction of most eukaryotic genomes is derived from DNA fragments that can multiply by inserting themselves in other regions of the genome, generating genomic instability and variability. These fragments are called Transposable Elements (TEs). Since they are a constituent part of the eukaryote genomes, these DNA pieces are normally inherited vertically along with other genes to the host offspring. Host genomes have an

arsenal of molecular mechanisms that were naturally selected to control TEs replication
avoiding most deleterious damage to the genome and hence leading to long-term TEs
inactivation and degeneration. However, TEs sometimes get out of control and invade
other species through horizontal transfer, the transfer of genetic information through
other means than sexual intercourse. This phenomenon is poorly studied in mosquitoes, a
highly important insect taxon that transmit many human pathogens. We characterized
the whole TE content of 24 mosquito genomes and investigated the TE inheritance mode
uncovering hundreds of horizontal transfer events among these species and distantly
related ones. We also identify a mosquito parasitic filarial worm, that is involved as a
transfer vector between many species of mosquitoes and that horizontally transferred TEs
contributed significantly to mosquito's genome expansion and variation.

Introduction

Transposable elements (TEs) are DNA sequences that can move from one genomic locus to
another. They were discovered by Barbara McClintock in the late 1940s, in the maize genome,
but it was not until the 70s that they were rediscovered in other species and their genetics and
evolutionary importance revealed in all branches of life [1,2]. It is now known that TEs pro-
foundly reshape the genomes of all species in a myriad of ways. The repetitive nature of TEs
generates abundant sites for chromosomal rearrangements and their transposition may have
serious consequences for the host genomes by way of insertion into host gene coding regions
or introns. This may give rise to new poly(A) tails, causing exon skipping or creating alterna-
tive promoters [3,4]. Although TEs are part of the host genome, they can replicate indepen-
dently of the host genome, generating new copies of themselves using the host's molecular
machinery, thereby triggering a coevolutionary arms race. Such intertwined interactions
between TEs and hosts may have wide-ranging consequences for the organisms' evolution
[5–7].

There is a large diversity in sequence-structure among TEs. Classification schemes distin-
guish two large groups of TEs based on the intermediate transposition molecule: TEs that are
mobilized via an RNA intermediate or retrotransposons and TEs that transpose via a DNA
intermediate or DNA transposons [8,9]. Many retrotransposons are large elements ranging
from 10 to 20kb in size showing a number of domains and *long terminal repeats* (LTRs) which
are also found in viral genomes hence sharing a common origin with different viral taxa. Ret-
rotransposons are also composed of highly diversified elements that have poly(A) tail and an
endonuclease or apurinic endonuclease domains, usually known as non-LTR retrotransposons
[10]. DNA transposons are divided into two subclasses distinguished by its transposition
mechanism: subclass I, are small elements (2 to 7 kb) normally coding for a single protein
comprising elements that transpose by cleaving both strands of DNA, and subclass II is com-
posed of large (Mavericks reaching 40 Kb) and medium-sized TEs (Helitrons reaching 15 Kb)
that transpose by single DNA strand cleavage [8,11,12].

TEs are mostly inherited by vertical transfer, the transfer of genetic information from
ancestral to descendant species, to host offsprings [13]. However, there is growing evidence
that TEs can also move horizontally between independent species by a phenomenon known as
horizontal transfer (HT) [14]. Recent large-scale studies on Insects and Vertebrates have
revealed thousands of Horizontal Transposon Transfer (HTT) events and general patterns,
such as the occurrence of HTT more frequently between closely related host species which
share a spatiotemporal overlap; the substantial contribution of horizontally transferred TEs to

the genome size of the host species; and that DNA transposons are transferred horizontally much more frequently than Retrotransposons [15]. However, there are still many open questions that need further assessment in order to fully understand HTT in different host taxa, its functional impact on host genomes, and by which mechanisms HTT occurs. Taking this last topic as an example, HTT might occur by direct transfer of a TE between species or mediated by a vector/intermediate species that transport TEs between hosts [16]. Currently, there is evidence scattered on different host/TE systems that these elements can 'hitch a ride' on other parasite genomes/particles, such as those of parasitic nematodes, macroscopic blood-sucking insects, and viruses that interact intimately with different host species [17–19]. However, the exact mechanisms by which HTT occurs has remained a large mystery since it is difficult to reproduce this phenomenon in controlled conditions. Moreover, reconstructing TEs horizontal steps from sequence data is not a trivial task with many pitfalls mainly due to incomplete sampling of host-intermediate species [16,20].

A significant portion of the genome of some species is composed by TEs: 80% of the model plant organism *Zea mays* [21], and almost half of the genome of the mosquito *Aedes aegypti* [22]. The Culicidae family encompasses all mosquitoes and has more than 3500 known species dispersed trough all continents except Antarctica. It is divided into two large subfamilies, Culicinae and Anophelinae, the former is composed of 11 tribes with several genera, while the latter has only 3 genera, having the *Anopheles* genus as the most species-rich group [23]. Many mosquito species of the Culicidae family transmit important human pathogens such as dengue, chikungunya, Zika, and yellow fever viruses as well as other pathogens, such as unicellular eukaryotes of the *Plasmodium* genus and filarial nematodes such as *Wuchereria bancrofti* and *Brugia malayi* [24]. In order to devise new methodologies to control such pests, at least 24 mosquito genomes have been sequenced from three Culicidae genera: *Aedes*, *Anopheles*, and *Culex* [22,25–29]. TE content varies substantially in mosquito genomes, with a greater abundance of Class I over Class II TEs [22,25], but well-characterized mobilomes are restricted to the most extensively studied species, such as *An. gambiae* and *Ae. aegypti* and even for those species there is no TE data accessible and/or standardized to allow more in-depth comparative analysis [30]. Therefore, there is little research regarding the inheritance mechanism of TEs among mosquitoes and only four well-documented horizontal transfer cases have been reported so far [31].

In this study, we recharacterized the TE content of 24 mosquito species and evaluated the horizontal spread of these TEs inside and outside of the Culicidae family. As a result, we found hundreds of horizontal TE transfers that occurred in the past 30 million years, which contributed significantly to the genomic expansion of mosquito genomes. Some of these horizontally transferred TEs are involved in HTTs with distantly related species from other Phyla. We also uncovered a vector species, the nematode *Wuchereria bancrofti*, that facilitated the horizontal spread of a specific TE among mosquitoes. Lastly, we found an absence of a positive correlation between HTT events and spatiotemporal overlap. Overall, our data brought a substantial contribution to understanding the HTT phenomenon and its impact on host genomes.

Results

TE content is highly variable among different mosquito species

We performed a recharacterization of TEs in 24 mosquito genomes available (S1 Table) using a *de novo* and a homology-based approach. For the former, we used the TEdenovo pipeline (part of REPET package), while for the latter we applied pairwise-alignment Blast searches using two different TE databases—Repbase and TEfam. All the resulting sequences were classified by three different programs, PASTEC, RepeatClassifier and TEsorter. A substantial
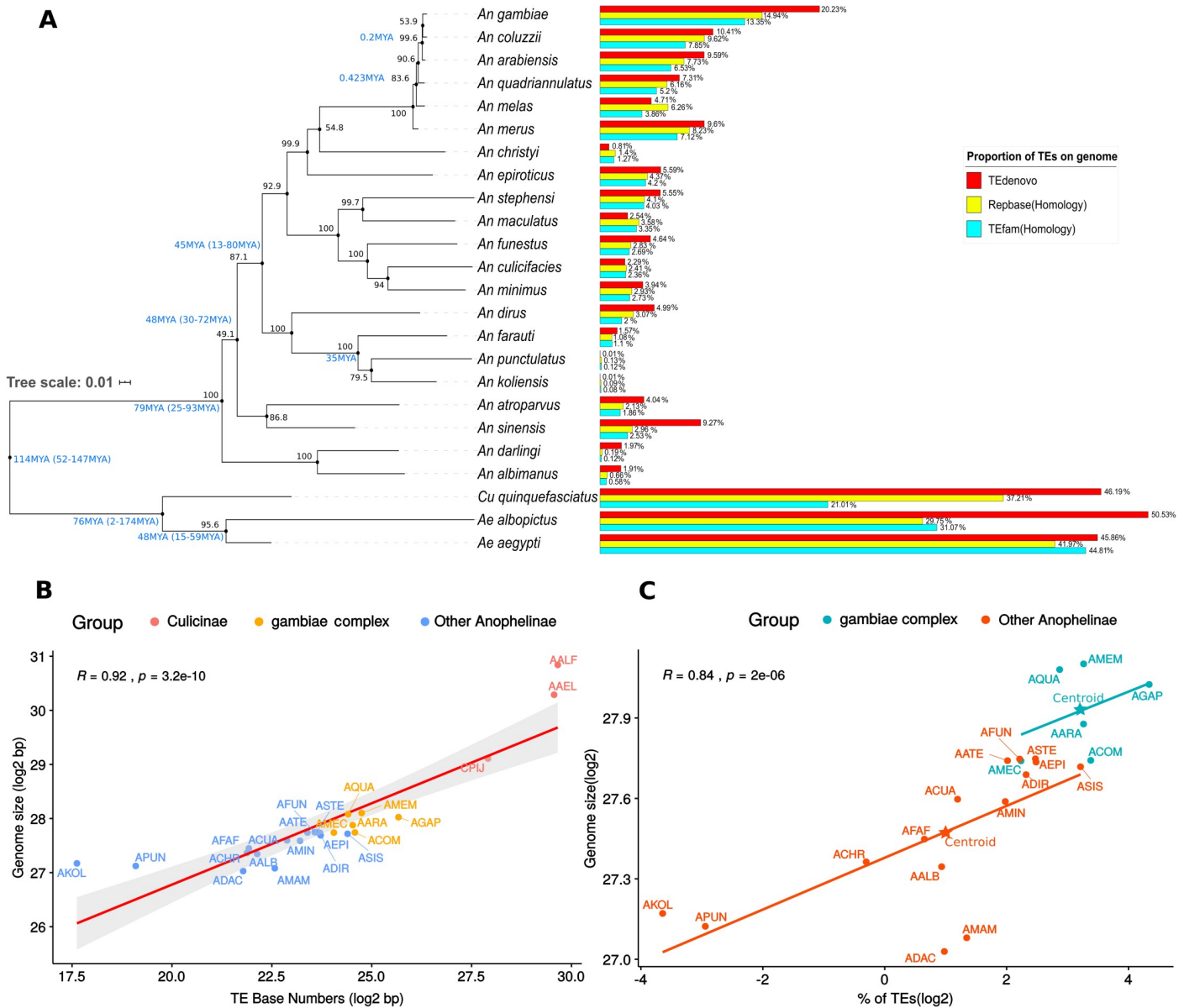
Fig 1. Transposable elements of mosquitoes. (A) A maximum likelihood-based phylogenetic inference, with bootstrap support over the nodes (see Materials and Methods - Genomes Analyzed), reporting the proportion of genome size occupied by transposable elements; blue letters values near bootstrap support represent medium divergence time between species obtained from TimeTree [104], colored horizontal bars represent the proportion of TEs characterized by different methods. (B) Correlation analysis of mosquito genome sizes and TE fraction. Four-letter species abbreviations follow Table 1.

https://doi.org/10.1371/journal.pgen.1008946.g001

variation was detected in the TE content of several species, for instance, *An. gambiae*—the most extensively studied mosquito species and the one that has the largest number of characterized TEs stored in databanks showed a TE content of 20.23% by TEdenovo, 14.94% by homology search against Repbase and 13.35% against TEfam (Fig 1A). The discrepancy between the three methods is more evident in one of the least studied species, *An. darlingi*. A 10-fold difference was observed between TEdenovo (1.97%) and homology-based search against Repbase (0.19%) (Fig 1A). To generate a final dataset of the TEs recovered using each methodology, we used all TEs recovered by the three different approaches and masked them using RepeatMasker. The final proportion of TEs in each genome is shown in S1 Fig.

The *An. gambiae* complex, which contains eight African species [32], six of each were analyzed here—*An. gambiae*, *An. coluzzii*, *An. arabiensis*, *An. quadriannulatus*, *An. melas*, and *An. merus*—showed the highest proportion of genome covered by mobile elements among all *Anopheles* species investigated. At the other extreme, the New Guinean mosquitoes *An. farauti*, *An. punctulatus*, and *An. koliensis* showed a very low TE content, between 0.13% and 1.57%, considering the highest estimates for all methodologies. Since neither a homology-based search nor the *de novo* approach identified a TE content greater than 1% in the last two species, we tried other methods of TE characterization using raw sequencing reads to ascertain whether these estimates might have been biased by TE removal steps during the original genome assembly. Tedna software found many repetitive sequences, however few of them are transposable elements (S5 File). On the other hand, the three mosquito species of the Culicinae subfamily (*Ae. aegypti*, *Ae. albopictus*, and *Culex quinquefasciatus*) have a much higher TE content than Anophelinae species, ranging from around 44% to more than 50% in *Ae. albopictus* and *Ae. aegypti* respectively (S1 Fig).

The final TE content obtained from masking all TEs recovered was used to evaluate whether the mosquito's genome size correlates with the mobilome content. The analysis indicated a strong positive correlation (R = 0.92) between TE content and mosquito genome size (Fig 1B). Looking only to *Anopheles* species there are two patterns, species of the *An. gambiae* complex have a larger genome size for similar TE content compared to other Anophelinae mosquitoes (Fig 1C), suggesting that both TEs and non-TE sequences play an important role in driving genome size in the *An. gambiae* complex. Other phenomena at the genomic scale, such as segmental duplication and/or expansion of the repetitive content not related to TEs such as satellites and multi-copy gene family expansion might also have contributed to genome expansion while TE driven expansion has a more significant role in driving genome size in the remaining *Anopheles* species.

Diverse TE superfamilies are found in the mosquitoes

At least 24 TE superfamilies were found in the mosquito genomes (Fig 2). This diverse mobilome comprises most of the transposable element orders established by Wicker et al. (2007) and some superfamilies that had not been described at that time (S2 Fig). LTR superfamilies, such as Copia, Gypsy, and Bel-Pao, are ubiquitous in mosquito genomes (Fig 2), accounting for at least 6% of the *Ae. aegypti* genome and 4.3% of *An. gambiae*. LINE elements are abundant in most species, they represent 14.5% of the genome of *Ae. aegypti*, 15.7% of that of *Ae. albopictus*, and 6.2% of that of *Cu. quinquefasciatus*. RTE and Jockey are the most abundant superfamilies in mosquitos' genomes. Twelve different TIR superfamilies are well represented in mosquitoes. As expected, Tc1-Mariner—the most widespread superfamily of TEs among arthropods—is present in all species of mosquitoes, even those with a small TE content. Non-autonomous TIR elements, such as putative MITEs and SINEs, are also present in the majority of species. It was not possible to classify some elements at the superfamily level, most of these have been classified only as Class I and Class II (S2 Fig). These elements represent around 15–20% of all TEs in mosquitoes of *An. gambiae* species complex and 18% of the *Cu. quinquefasciatus* genome, but represent only a small proportion of other mosquito genomes ($2E^{-4}$ to 3%). Chimeric elements, formed by nested elements from two or more distinct superfamilies, are present in almost all species, but a higher proportion of chimeric TEs was found in the *Ae. aegypti*, *An. arabiensis*, *An. gambiae*, and *An. quadriannulatus* genomes (S2 Fig).

Both Class I and II superfamilies are heterogeneously distributed among mosquitoes. The Penelope superfamily was only found in the three Culicinae mosquitoes studied and covered only a small fraction of these genomes (Fig 2). The same pattern can be seen with Mutator,
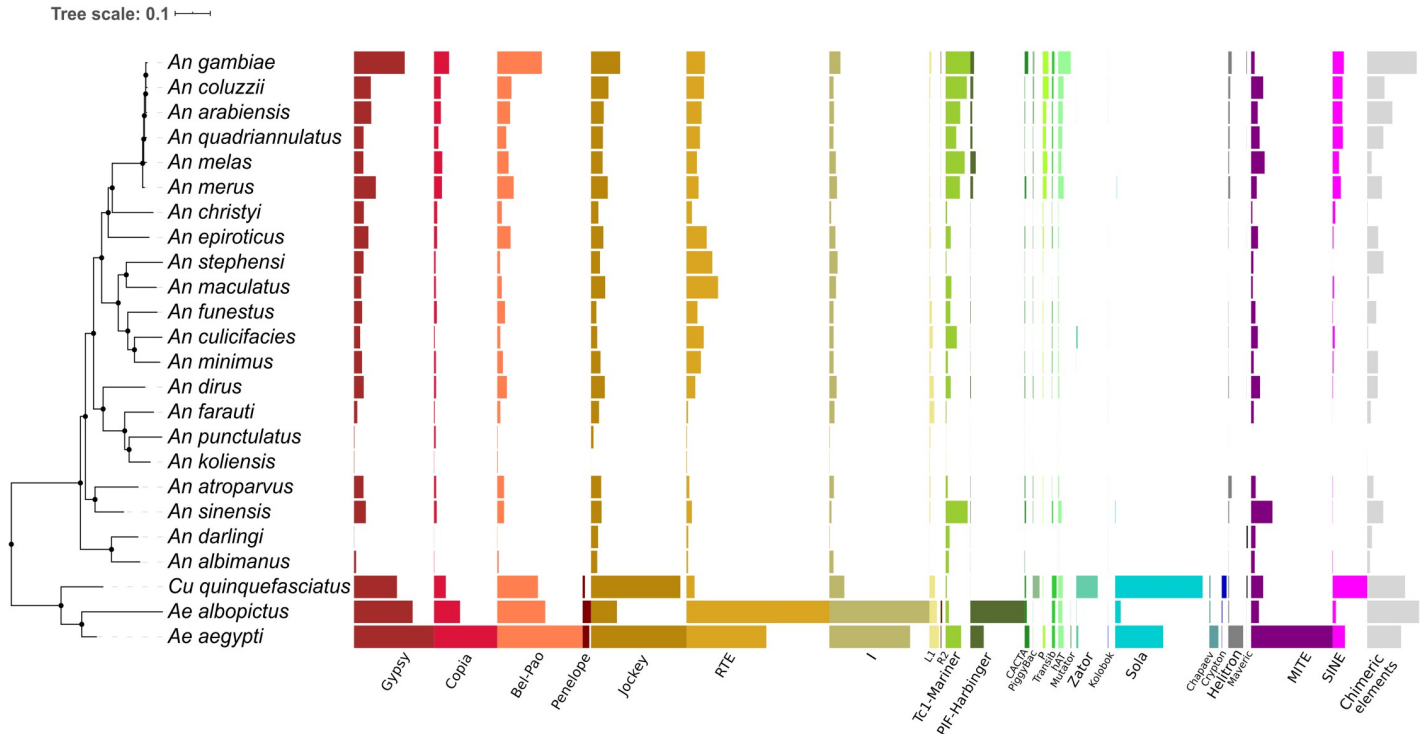
Fig 2. Proportion of each TE superfamily in the genomes of mosquitoes. A phylogenetic tree of 24 mosquito species showing the distribution of different TE superfamilies in mosquitoes. Each colored bar represents the proportion of each TE superfamily (names in X-axis) following Wicker et al. (2007) hierarchical scheme. Uncharacterized or order-only-characterized TEs are not shown (see S2 Fig).

https://doi.org/10.1371/journal.pgen.1008946.g002

Chapaev, and Crypton-like elements. The most remarkable difference in superfamily distribution was found in the case of Sola elements. This element is one of the most abundant TEs in the *Cu. quinquefasciatus* genome (2.86%). It is also present in *Aedes* mosquitoes and only one Anophelinae species—*An. sinensis* (representing nearly 0.01% of its genome). Another clear difference is the absence of elements from the CACTA, P, and Transib superfamilies among neotropical and New Guinean anophelines. R2 is another superfamily that has an intriguing patchy distribution among mosquitoes. It is present only in *Aedes* mosquitoes and in species from the *An. gambiae* complex (Fig 2).

From the host mosquitos' genomes perspective, the *An. gambiae* species complex has a similar TE superfamily landscape, while there is substantial variation in distribution and abundance between *Aedes* and *Anopheles* mosquitoes (Fig 2). However, it is clear that *Cu. quinquefasciatus* has the most distinct TE superfamily landscape of the Culicidae family, with the greatest abundance of Zator, Sola, and Crypton-like elements compared to all other mosquito species studied. Another distinctive feature of the *Cu. quinquefasciatus* mobilome is the low genome proportion of ubiquitous superfamilies among mosquitoes such as Tc1-Mariner, RTE, and I (Fig 2).

Species from the *An. gambiae* complex shows more recent TEs activity

To examine the landscape of TEs within species genomes over time, we estimated the Kimura-2-parameter (K2P) distance of all TEs. We clustered TEs in the four highest abundant orders: LTR, LINE, TIR, and SINE. The boxplots in S3 Fig reveals interesting patterns: mosquitoes from the *An. gambiae* complex tend to have more recent TE families from the TIR, LINE, and

LTR orders than other Anophelinae. Low K2P values in these six mosquitoes indicate that many TE families are currently active or had been recently active. In contrast, the vast majority of LINE families in *An. darlingi*, *An. albimanus*, *An. punctulatus* and *An. koliensis*—which are species with a very low TE content—are old and probably no longer active.

Most SINE elements in the genome of *Anopheles* mosquitoes seems to have undergone ancient transposition bursts and no longer appear to be active in these mosquitoes (S3D Fig). A distinct pattern can be seen in Culicinae species, where we detected recent expansion of SINEs. In general, the K2P pattern provides clues that TEs are younger and more recently active in Culicinae than Anophelinae families.

Hundreds of HTT events have occurred among the mosquito species

We found 172 TE families with significant HT signals among the species studied, using VHICA package [33]. However, it is important to note that this is certainly an underestimate of the true number of HTT events since a number of TE families have more than one significant pairwise HTT signal (Fig 3A). We, therefore, decided to describe it in more general terms, retaining at least one HT event for each TE family, in view of the complexity of the signals found and the lack of currently available algorithms for determining the most likely HTT scenario and the minimum number of events required to explain these signals. Fourteen additional families showed a positive HTT signal. However, these families are not present in all genome assembly versions of one species or we cannot differentiate HTT event from an introgression event (S6 File).

More than three-quarters of transposons involved in HT events belong to Class I TEs, covering eight different superfamilies (Fig 3B) while eight Class II TE superfamilies were detected undergoing HT. The superfamily with the largest number of families undergoing HT is Gypsy, in which HT accounts for more than 28% of all transfers, followed by the Class II Tc1-Mariner elements. Tc1-Mariner superfamily deserves special mention since it has 9 distinct families involved in HT events between four or more mosquito species.

Almost 20% of the HT signals detected (34 families) occurred only between *Ae. aegypti* and *Ae. albopictus*. One hundred and thirty-three TE families involved in HT events were detected among species of the *Anopheles* genus, accounting for 18 of the 21 species. Only three mosquito species of this genus, *An. darlingi*, *An. punctulatus* and *An. koliensis*, were not involved in HTT events. Among *Anopheles* species, *An. merus*—an early divergent species of the *A. gambiae* complex—was involved in the largest number of HTTs (Fig 3A). The majority of species from the *Anopheles* genus are involved in HTTs from both Class I and Class II TEs (Table 1). In three species we detected only HTTs of Class I TEs: non-LTR elements in *An. albimanus* and *An. farauti*, and one non-LTR and one LTR in *An. christyi*. Concerning HTT between different mosquito genera, six events were found. Five out of those belong to the Tc1-Mariner superfamily, while only one R4—a non-LTR retrotransposon—was transferred from *Ae. albopictus* to the ancestor of *An. gambiae* complex mosquitoes, as already described in a previous study [34]. No horizontal transfer events involving the *Cu. quinquefasciatus* species were found. It is important to note that all HTT events showed very low p-values (S4 Fig) and were associated with several additional HTT evidence as patchy distribution and phylogenetic incongruences between TE and host phylogeny, as further discussed below (**2.8—Horizontal transfer of TEs involving distantly related eukaryotic species**).

Around 10% of the HTT networks involve at least four species. This proportion is small, particularly among LTR families, where four or more mosquito species were detected in less than 10% of the total HTTs involving LTR families. Of all the superfamilies studied, Tc1-Mariner is the one whose elements are most often horizontally transferred across many species, 9
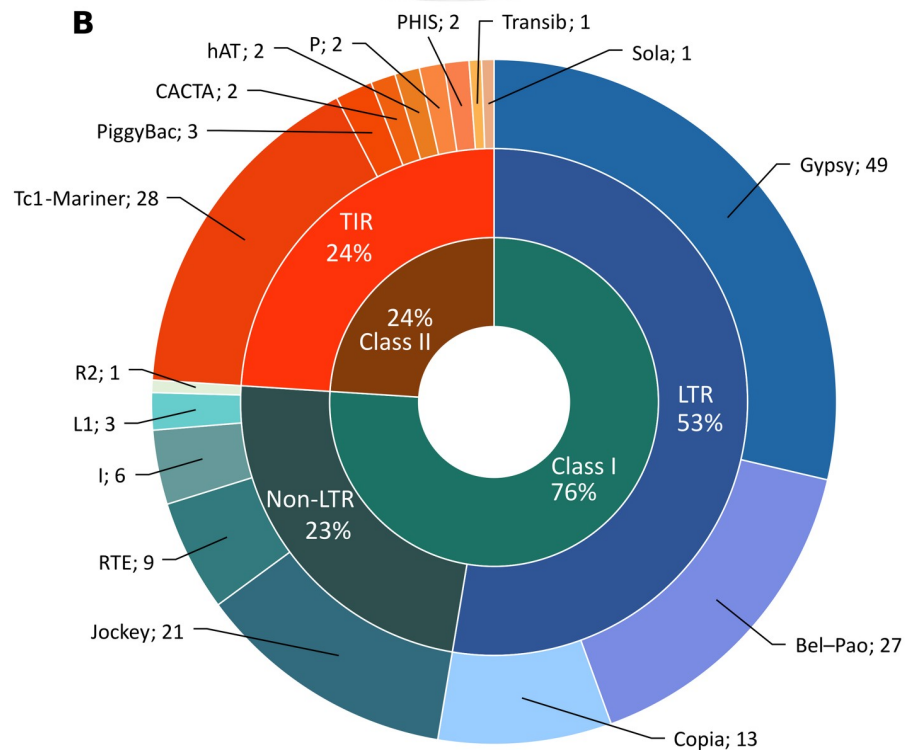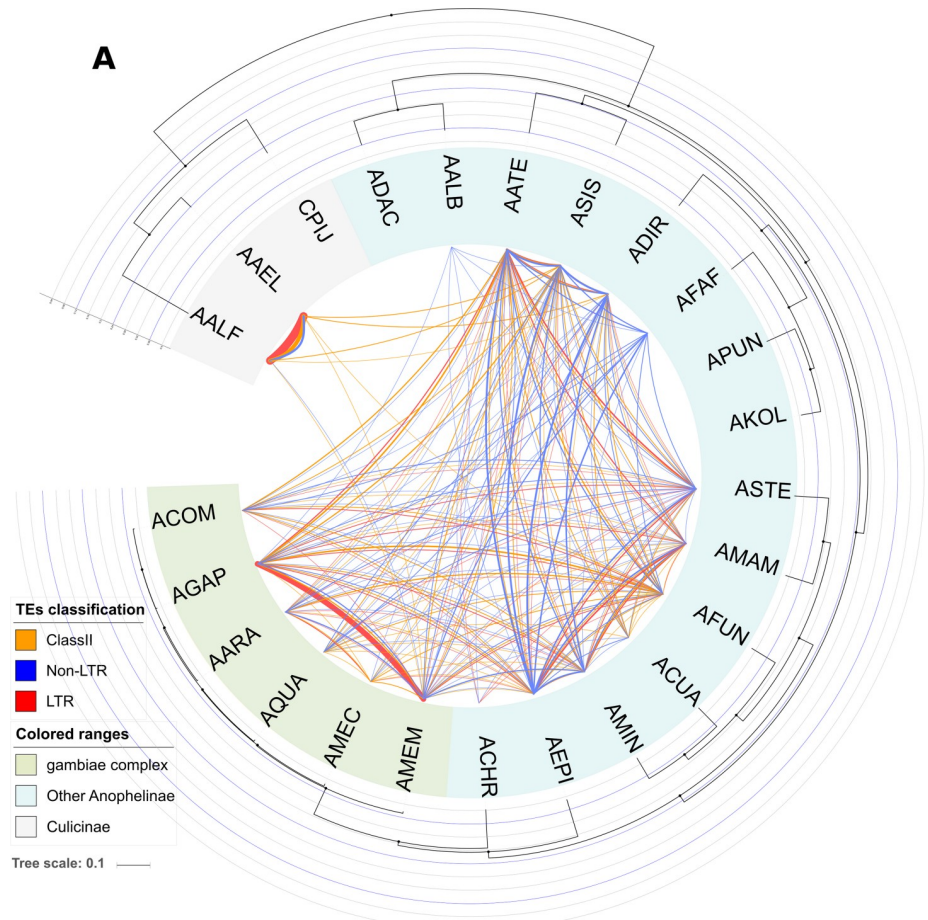
Fig 3. HTT among mosquitoes and representative TE superfamilies. (A) A phylogenetic tree (see Materials and Methods - Genomes Analyzed) whose edges represent pairwise HTT events and thickness is proportional to the number of events. (B) Classification of these 172 TE families involved in HTT among mosquitoes.

https://doi.org/10.1371/journal.pgen.1008946.g003

TE families of this superfamily involved in HTT occurred between more than 2 mosquito species. Details about the participation of species in each TE family horizontal transfer network can be seen in S2 File, which lists all VHICA output images of positive HTT cases. The relative intragenomic age of each TE family undergoing HT per mosquito genome is given in S5 Fig.

## Horizontal transposon transfers directly impact the genome size of mosquitoes

TEs horizontally transferred between species represent a significant fraction of some mosquito genomes (Fig 4A). We estimated that about 6.7% of *Ae. aegypti* and 7.8% *Ae. albopictus* genomes are covered by horizontally transferred TEs representing more than 10% of the total TE content of these species. In the *Anopheles* genus, the impact of horizontal transfer on genome size was considerably smaller. Only *An. gambiae*, *An. coluzzii*, and *An. stephensi* showed more than 1% of their genome covered with copies derived from horizontally transferred TEs.

The proportion of genome covered by horizontally transferred TEs showed a positive correlation with the genome size of the host species Fig 4B. Therefore, TE invasion by way of

Table 1. Number of TE families in each species that is involved in HTT events.

Species	Code	Class II	Class I		Total
			Non-LTR	LTR	
Anopheles merus	AMEM	9	13	21	43
Anopheles gambiae	AGAP	10	6	26	42
Aedes aegypti	AAEL	12	6	22	40
Aedes albopictus	AALF	11	7	22	40
Anopheles epiroticus	AEPI	5	15	15	35
Anopheles atroparvus	AATE	9	7	17	33
Anopheles maculatus	AMAM	5	9	12	26
Anopheles sinensis	ASIS	10	7	9	26
Anopheles dirus	ADIR	8	9	6	23
Anopheles minimus	AMIN	7	9	7	23
Anopheles stephensi	ASTE	4	7	12	23
Anopheles arabiensis	AARA	6	5	7	18
Anopheles funestus	AFUN	8	3	7	18
Anopheles coluzzii	ACOM	4	2	5	11
Anopheles melas	AMEC	6	1	3	10
Anopheles culicifacies	ACUA	5	3	1	9
Anopheles quadriannulatus	AQUA	4	3	1	8
Anopheles farauti	AFAF	0	6	0	6
Anopheles albimanus	AALB	0	2	0	2
Anopheles christyi	ACHR	0	1	1	2
Anopheles darlingi	ADAC	0	0	0	0
Anopheles koliensis	AKOL	0	0	0	0
Anopheles punctulatus	APUN	0	0	0	0
Culex quinquefasciatus	CPIJ	0	0	0	0

https://doi.org/10.1371/journal.pgen.1008946.t001
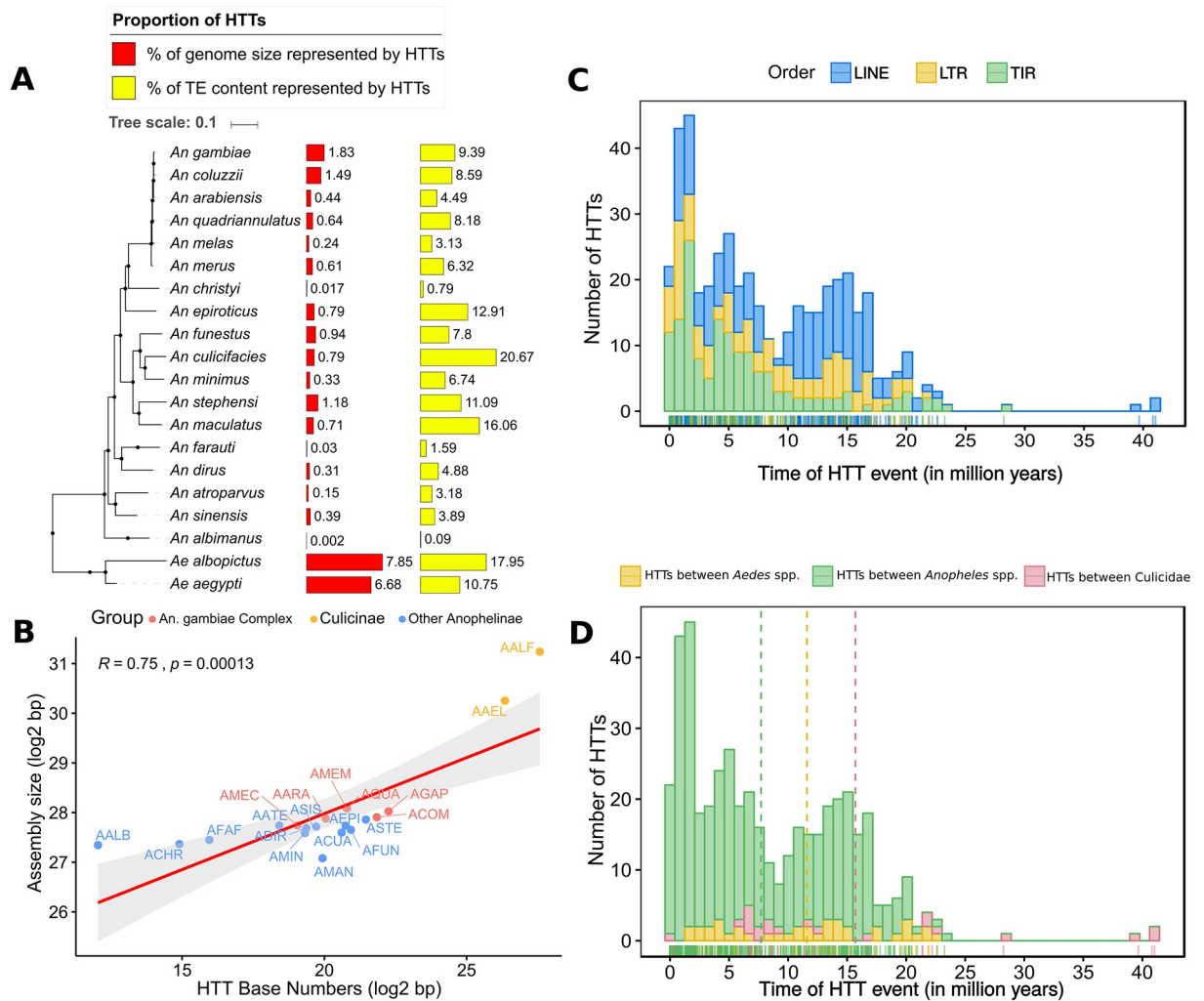
Fig 4. Influence of HTT on the host genome and dating. (A) The proportion of the host genomes composed of TEs involved in HT. (B) The correlation between the number of nucleotides covered by TEs involved in HTT events and genome size of mosquitoes. The age of each pairwise HTT event is estimated and represented by TE groups (C) and host taxa (D).

horizontal transfer may significantly contribute to genomic expansion and also impact the size of the mosquito genome. Thus, the more frequent HTT events, the larger the genome.

Horizontal transfer of transposable elements occurred in the last 30 million years

We estimated the HTT dating based on all pairwise 463 significant comparisons found. The vast majority of HTT events detected occurred in the last 30 million years (Fig 4C and 4D), a time during which most *Anopheles* species underwent speciation [35]. Only one transfer is older than this: a transfer of a R4 clade involving species from the *An. gambiae* complex and both of the *Aedes* species studied that took place around 40 million years ago (S6 Fig). LINE and LTR elements have the most uniform number of HTTs over time, while HTT of TIR elements intensified more recently (Fig 4C). Horizontal transfers occurring between species of the *Anopheles* genus and between species from the *Aedes* genus showed no major difference in the estimated time (Fig 4D).

Geographical location does not affect the number of transfers between mosquito species

In order to investigate whether geographical overlap of mosquito distribution range favors the occurrence of HTTs, we compared the distribution of species across the globe with the number of HTTs between each pair of species. No correlation was found between these variables (R = 0.1, p-value = 0.25) (S7 Fig).

Tc1-Mariner elements were horizontally transferred to distantly related eukaryotic species

We also examined whether some of the transposable elements that are horizontally transferred among mosquitoes might also be involved in transfers to other eukaryotes. Five TE families from the *Tc1-Mariner* superfamily showed a broader horizontal transfer network with other species. Three of them have already been described in previous studies, showing transfers between at least two species (Fig 5). Our results further expand this network, adding twenty new species. The other HTT events in the remaining two TE families are described here for the first time. The similarity between mosquito elements and those of other eukaryotic species can be seen in S3 File.

The species involved in HTT events belong to three metazoan phyla. The first phylum was Arthropoda, which covers the majority of species. It is represented by the class Arachnida, by the common house spider *Parasteatoda tepidariorum*, by Hexanauplia, represented by the copepod *Lepeophtheirus salmonis*, and by the class Insecta, which includes most of the mosquito species involved in HTT. The second phylum was Cnidaria, with only one representative, *Hydra vulgaris*. The third was Nematoda, where we found a horizontal transfer case involving the parasitic worm *Wuchereria bancrofti* and other *Anopheles* species that transmit this nematode to humans. Besides the high identity between mosquito TEs and those of distantly related species, sometimes separated by more than 700 million years, we also observed a patchy distribution of these elements, providing further support for the horizontal transfer events.

Two of these TE lineages (HTa and HTe) belong to the ITmD37E group, which includes TEs that have the DD37E motif on their transposase sequences (Fig 6A). One lineage (HTd) clustered with previously described elements from the ITmD37D family, also known as the *mat* family. The other two lineaged (HTb and HTf) are from the DD34D or *Mariner* family. Elements from the HTf network belongs to the *mauritiana* subfamily, and the HTb network elements, which are involved in most HTT events with species from outside of the Culicidae family, belong to the *irritans* subfamily. To obtain a more in-depth understanding of the HTb network, we performed a second BLAST search using as queries all sequences recovered from the search against non-mosquito genomes. We found HTb related elements to be present in 42 species of divergent invertebrates, including *W. bancrofti* (Fig 6B). Intragenomic dating of the horizontally transferred TEs allows a more accurate interpretation of the HT events, including the source and receptor species. The relative intragenomic age of TEs showed that the most ancient elements belong to the Lepidoptera species *Adoxophyes honmai*. Most of the flies' elements are ancient and have similar ages. The elements of Asian mosquitoes (*An. maculatus*, *An. epiroticus*, *An. dirus*, *An. sinensis*) are more ancient than those of African ones (*An. gambiae*, *An. coluzzii*, *An. arabiensis*, *An. funestus*). Interestingly, the age of the *W. bancrofti* element is very similar to that of those of the Asian mosquitoes, suggesting that *W. bancrofti* acquired the element from the Asian *Anopheles* species and donated it to the African species of *Anopheles* more recently (Fig 6C). Further evidence that this element is prone to transpose horizontally is provided by the multiple pairwise HTT significant signal among mosquitoes (Fig 6D).
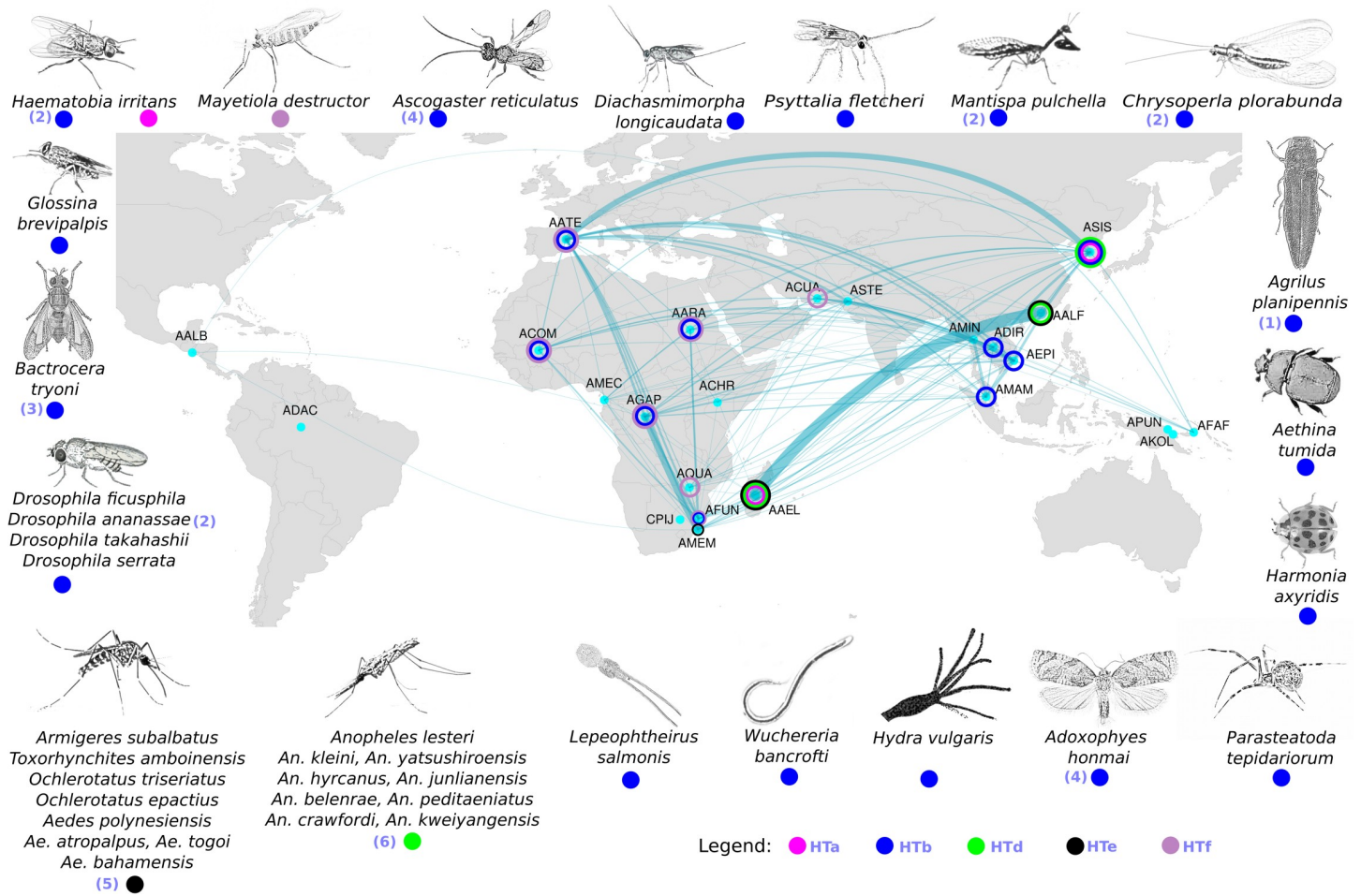
Fig 5. Mosquito HTT network from a geographical perspective. Light blue dots represent the species sampling site for the genomes used in this study or the most probable site of a species origin. The edges connecting dots represent significant pairwise HTT events, the thickness of those edges is proportional to the number of HTT events. Five out of 172 TEs are horizontally transmitted to other species outside the group of 24 studied species, these TE families are represented by colored circles: HTa (pink), HTb (blue), HTd (green), The (black), and HTf (light purple). The species involved in a horizontal transfer with at least one of the 24 mosquitoes are shown in the map border. The colored circle around light blue dots indicates that this mosquito species is involved in a long-range HTT network of at least one specific TE (HTa to HTf). The numbers inside the round brackets represent some preliminary evidence of HTTs involving these TEs and species in other published articles. 1 represents the APmar1 element already described by Rivera-Vega et al. [105], 2 represents HTT transfers among species described by Robertson and Lampe [106], 3 represents an element described by Green and Frommer [107], 4 represents HTT transfer between *Ascogaster reticulatus* and *Adoxophyes honmai* [108], 5 represents HTT events already described in mosquitoes [109], 6 represents HTT events already described in other *Anophelinae* species [110] not investigated in our study.

https://doi.org/10.1371/journal.pgen.1008946.g005

Discussion

Most evolutionary biologists are still puzzled by the abundance and variability of the mobilome, even in closely related species. These rather simple sections of DNA are able to override the rules of Mendelian inheritance, that govern the rest of the genome, by two main ways: hijacking the host molecular machinery to generate more copies of themselves increasing in copy number while the majority of host genes remains stable and through horizontal transfer, a phenomenon that allows TEs to invade the genomes of other species despite being inherited vertically to all original host descendants [3,36]. More than hijacking the host molecular machinery, TEs also can be detrimental to their host due to transposition and recombination [4,37]. Nonetheless, host genomes possess a full arsenal of molecular mechanisms that can silence these elements [38–40]. Therefore, the abundance and diversity of the mobilome in the
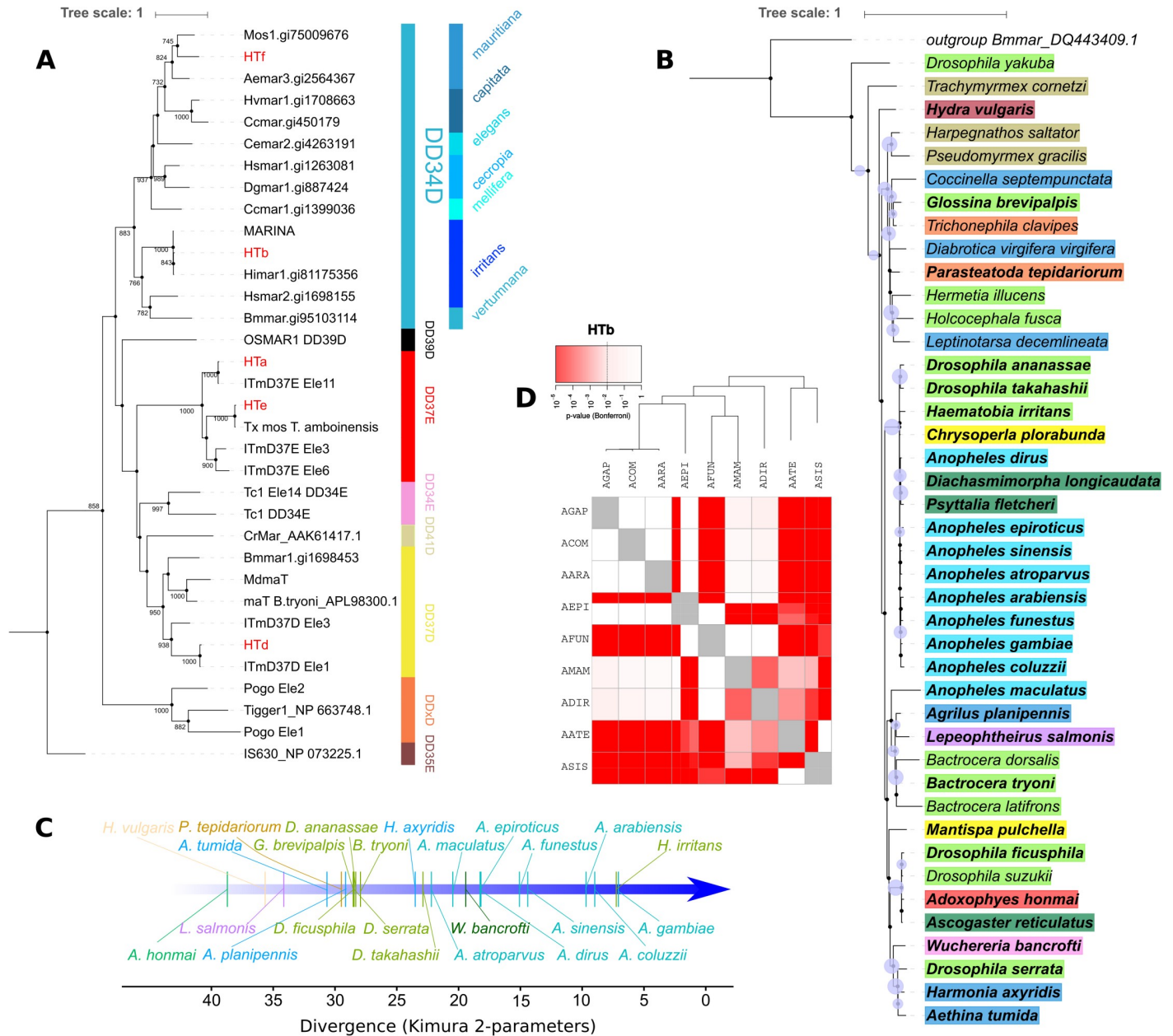
Fig 6. Transposons horizontally transmitted to other species. (A) A phylogenetic tree based on the transposase amino acid sequences of multiple ITm superfamily elements (bootstrap values are showed at nodes), including five Anophelinae TEs (labeled in red) involved in HTT events with other metazoan species and many elements that have already been described. The tree topology indicates that these five TEs belong to different families. HTb represents an element that is widespread in many metazoan genomes, as shown in the nucleotide-based phylogenetic tree of part B. A species name in bold represents elements that share more than 80% of similarity to *Anopheles* ones. Brachycera is highlighted in light green, cnidarians in dark brown, ants in light brown, spiders in orange, beetle in blue, elements from the *Anopheles* genus in light blue, parasitoid wasps in dark green, Neuroptera in yellow, Lepidoptera in red, crustaceans in purple, nematodes in light purple. (C) A timeline built up using K2P distances showing the average relative age of elements in the genome of each species. (D) VHICA graphical representation of the HTb element in mosquitoes. The red-colored boxes represent statistically significant HTT signals.

https://doi.org/10.1371/journal.pgen.1008946.g006

genome of each species is thus the result of an endless arms race between TEs and host [6,41,42]. Although much knowledge has been discovered about the TE "life cycle" since the genomic revolution, there are still several open questions about the specific role of horizontal transfer in reshuffling TEs between different eukaryotic species and their impact on receptor

species. In this study, we performed an in-depth mobilome characterization in all available mosquito genomes and detected a very large and variable TE content that is constantly exchanged by HTs between mosquitoes and distantly related species. Moreover, we found strong evidence of an intermediate worm species that mediated HT between mosquitoes and showed that horizontally transferred TEs significantly impacted the mosquito genome size.

The 24 mosquito genomes available have been scarcely studied from the mobilome perspective, except for the two model species *An. gambiae* and *Ae. aegypti* [22,43]. However, even though some TE sequences can be found in databases, there is no standard annotation and the use of different methodologies for TE detection hinders a broad-scale comparative genomic analysis of the mosquito mobilome. Recently, 16 new *Anopheles* genomes have been published [25], but very little is known about the mobilome of these mosquitoes and annotated TEs have not been made available to the scientific community. To improve detection and standardized classification of TEs, we used a range of softwares that perform de novo TE detection and homology-based search followed by three different classification programs, PASTEC, TEsorter and RepeatClassifier. In line with other studies, the different methodological approaches generate different and yet complementary results [44,45]. On the one hand, *de novo* approaches result in more extensive TE content characterization in the case of most mosquito genomes. On the other hand, this method does not recover the whole mobilome. Some low-copy and more divergent TE families were recovered only by homology search. The most striking impact of using complementary methodologies in terms of overall TE content occurred in the case of the *Cu. quinquefasciatus* genome. Arensburger et al. (2010) showed that 29% of the *Cu. quinquefasciatus* genome was composed of TEs. A further study using a *de novo* method based on structural search found other TEs not described at the time in this genome [46]. Here, we further expand the TE annotation for *Cu. quinquefasciatus* reaching a TE content of 43.55% and describe some TE superfamilies that had not yet been described in some mosquito species, as we report for the first time the occurrence of Penelope elements in the genome of *Cu. quinquefasciatus* and Crypton-like elements in the *Ae. albopictus* genome. Thus, a consistent characterization of the TE content of a given species requires that both *de novo* and homology-based approaches (with different reference TE database) be used for TE identification and classification.

Our in-depth mobilome characterization allowed us to investigate the influence of TEs on the size of the mosquito genomes. We observed that the mosquito genome fraction covered by TE correlates significantly with genome size although it is much more pronounced in species from the *Aedes* genus. Our results corroborate findings in arthropods [47] and vertebrates [48], However, it is interesting that the *An. gambiae* species complex showed a higher genome size to TEs ratio than the other *Anopheles* species. This suggests that other sequences are also responsible for genomic expansion in these species such as microsatellites and segmental duplications as shown for other species [49,50].

Given the high variability and uneven distribution of various TE superfamilies/families found in the genomes of mosquitoes, we next investigated the role of HTT in the evolutionary history of the mosquito mobilome. More than five thousand HTTs have been reported among eukaryotes so far, but only 4 HTT events involving mosquitoes have been well-documented, while 257 cases have been reported among drosophilids—a sister taxon [31]. Here, we report a forty-fold increase (172) in the number of HTT events described among mosquitoes and investigate several hypotheses regarding the influence of host and TEs biological features on HTTs. One of the basic requirements for a given TE transfer to a new host species is the spatiotemporal overlap of the hosts [16]. Recent findings on insects supported a higher HTT rate between taxa that shared the same realm [15]. However, in our fine-grained analysis of mosquito genomes, we found no association of ancestral species habitat range and HTT rate (S7 Fig)

showing that different patterns may emerge driving HTT rate at different host taxonomic levels. Association between TE types or classes and the horizontal transfer rate have been linked showing that LTR and DNA transposons transfer much more frequently than non-LTR retrotransposons [36,51]. In mosquitoes, we found that half of the superfamilies that have undergone HTT belong to the LTR order of TEs, with an almost equal number of families of TIR and non-LTR elements in the remaining fraction. This result differs somewhat from other studies of insects investigating a broader set of taxa and a narrower time frame (last 10 MYA), which report a majority DNA transposons TEs involved in HTT events followed by LTR and non-LTR retrotransposons [15].

Another host intrinsic feature that may impact HTT rate is the phylogenetic relatedness of the host species. A number of studies have reported that horizontal transfer occurs more frequently in more closely related species [15,52]. However, long-range HTTs, between distantly related taxa, have also been observed [53,54]. These occur frequently with specific TE groups, such as Class II elements from the Tc1-Mariner [55] and hAT superfamilies [56]. Our data corroborate these findings, showing that most HTTs of Retrotransposons occurred between closely related species of the same genus, and long-range transfers among species of *Aedes* and *Anopheles* genus occurred mostly with DNA transposons. Five TEs that underwent transfer to even more distant metazoan species were also Class II TEs from the Tc1-mariner superfamily. A similar pattern was observed in a study that investigated HTT in 195 insect species [15]. In general, HTTs occur more frequently between closely related species, but Class II TEs are more prone to long-range transfers probably due to host factor transposition independency and promoters that make them able to transpose in a wide set of divergent host species [57].

An interesting aspect of the large number of HTT events found among mosquitoes is that it is heterogeneously distributed on the mosquito evolutionary tree, showing a much higher prevalence of HT between species from the *Aedes* genus (Fig 2). Although there is a clear bias of more genomes available from *Anopheles* genus, we still found much more HTTs between *Aedes* species, the largest mosquito genomes sequenced so far, and not a single HT involving *Culex quinquefasciatus*, a species with intermediate genome size (~600Mb) but with a large TE content (around 44%). More unbiased studies including a diverse set of mosquito species should be performed to test these patterns, but our findings highlight that some taxa are more prone to exchange TEs horizontally than others, which is corroborated by other large-scale studies [58].

One of the most debated topics concerning HTT is how exactly do TEs move from one species to another and which vector species facilitate HTT bridging the gap between species. Several long-range HTT events across phyla were found to involve a Tc1-Mariner element of the *irritans* subfamily (Fig 6B). Interestingly, we found a number of copies of these elements in four different genome assemblies of *W. bancrofti*—a parasitic nematode transmitted by mosquitoes that causes lymphatic filariasis in humans [59]. At least one of these assemblies was reconstructed using only DNA sequences from adult worms isolated from the blood of an infected human patient (GCA_001555675.1), excluding any possibility of mosquito DNA contamination. To further evaluate the possibility of mosquito DNA contamination in the *W. bancrofti* genome assemblies, we Blasted the four genomes available against the *An. gambiae* genome and confirmed that highly similar elements with large alignment size were recovered only for the HTb Tc1-Mariner element (S7 File). These pieces of evidence together confirm that these elements are true components of the genome of this species. The *W. bancrofti* element has a similar age to that of those found in *Anopheles* species from Europe and Asia, although this element is younger in African *Anopheles* species. This suggests that *W. bancrofti* acquired this element from European and Asian *Anopheles* species and transferred them, more recently, to African *Anopheles*. Considering that *Anopheles* species do not migrate long distances, it is possible that these intercontinental transfers of TEs occurred by way of worms

hosted by migrating animals, such as birds [60]. In addition, there is evidence showing that *W. bancrofti* is transmitted by different species of the *Aedes*, *Culex*, *Mansonia*, and *Anopheles* genera [61], suggesting that *W. bancrofti* likely mediated the transfer of HTb elements among at least eight species of *Anopheles* genus. The discovery of this parasitic nematode strengthens the evidence of parasitic species as an important agent of TE transfer among multicellular eukaryotic species.

Measuring the direct influence of horizontally transferred TEs is a difficult task since precise gene annotation of all genomes is required to associate TE insertions with the impact on surrounding genes. However, it is possible to estimate the proportion of the host genome that is derived from horizontally transferred TEs. This can be used as a proxy for estimating the likely impact/burden on the host genome. We found that horizontally transferred TEs contribute to genome size in varying degrees, depending on the mosquito genome. For both species of the *Aedes* genus, TEs involved in HT represent a substantial portion of the genome (6.68–7.85%), that is, around 199Mb in the *Ae. albopictus* genome (2.5 GB). This is larger than the genome size of *Drosophila melanogaster* (143.7Mb) [62] and the two-spotted mite *Tetranychus urticae* (91Mb) [63]. These results show that horizontally transferred TEs reach a high copy number since they replicate unchecked by the host genome after their invasion and can contribute substantially to the expansion and structure of the host genome. In line with that, genome size is also correlated with the proportion of the genome covered by horizontally transferred TEs, confirming that HTT contributed substantially to the mosquito genome size (Fig 4B).

It is important to highlight that the genome assembly quality might have a large impact on TE characterization, annotation and further analysis. Mosquito genomes are highly variable in terms of assembly quality and many *Anopheles* genomes are fragmented (S1 Table). However, we used very stringent filters and validation with new genome assemblies' versions produced using long-reads to avoid inconsistent results for the species with available complementary data *(Ae. aegypti*, *Ae. albopictus*, *An. funestus*, *An. stephensi* and *An. coluzzii)* confirming the great majority of the findings based on more fragmented genome versions.

Our study shows that complementary methodologies should be used for the precise characterization of the mobilome of a host species. Moreover, mosquito TEs are being constantly reshuffled among mosquito species by way of multiple horizontal transfer events facilitated by mosquito parasites. The emergence of horizontally transferred TEs into naïve genomes have substantially impacted the mosquito genome size.

Materials and methods

Genomes analyzed

All the mosquito genomes used in this study were obtained from VectorBase [64] and NCBI Assembly [65] as a result of many studies [22,25–29,43,66–72]. The GenBank accession codes and other features of the assemblies are listed in S1 Table. The Mitochondrial DNA (mtDNA) was assembled using the raw reads for some species on Mitobin version 1.9.1 [73]. These sequences, along with mtDNA genomes already available in the NCBI, were used to build up a phylogenetic tree of studied species. The tree was constructed by way of maximum likelihood using PhyML [74] with 1000 bootstrap replicates, according to the GTR +G+I model based on the Akaike Information Criterion on SMS software [75].

Transposable element identification

Two main types of approaches are used to identify TEs in genomic assembled sequences: homology-based and *de novo* approaches. The most commonly used method involves the

detection of homology between previously characterized TEs, normally obtained from databases such as Repbase, and the newly assembled sequences [30]. RepeatMasker is the software of the first choice for providing an overview of the TEs in any given genome [76]. Alternatively, *de novo* approaches identify TEs by their structural features or by repetitive/multiple-copy characteristics in the host genome [77–79]. There are various programs developed over the years, implementing these kinds of approaches, such as RECON [80], RepeatScout [81], PILER [82], LTR Finder [79], RepeatExplorer [83] and so forth. As a way of avoiding underestimation, the REPET pipeline [84] comes close to proposing a full pipeline for TE identification and annotation. Although each approach has its strengths, the use of either one of these methodologies in isolation almost always leads to an underestimation of the true TE content and adoption of a complementary strategy is, therefore, advised [44].

Identification of transposable elements using a *de novo* approach

A *de novo* approach was employed to identify mosquito TEs, using the TEdenovo pipeline from the REPET package [84]. See S1 File for details of pipeline execution. To complement our initial identification, we also ran RepeatScout, another software that uses a *de novo* approach [81], on the genome of each species using the default parameters. The repeated elements identified were then passed to the TEdenovo pipeline.

Given the virtual absence of elements characterized by TEdenovo or RepeatScout in some genomes, we examined the raw reads of these species further using Tedna [77], to ascertain whether TEs could have been removed before or during the genome assembly step.

Identification of transposable elements using a homology-based approach

At this stage, the TEs were identified based on homology with previously described TEs in the RepBase and TEfam databases. Initially, we executed a blastn and a tblastn (BLAST 2.2.31+ version) search of the two databases, individually, against the genome of each species. Only HSPs (high-scoring segment pairs) with a bit-score above 200 were retained, to exclude random hits. Thereafter, blastn and tblastn-derived hits were merged into a single file to reconstruct the copies derived from HSPs that matched elements from the same family spaced no more than 1000 nucleotides apart.

All TE copies in a given genome were clustered by the CD-HIT-Est [85] algorithm with 80% identity and coverage using a global alignment strategy. Copies of each cluster were then extracted and aligned using the MAP algorithm, which reconstructed the representative consensus of each structural variant. Finally, a file containing all the TE consensuses for each species was generated for each of the databases used.

Consensus mapping and calculation of TE family divergence

The consensus sequences derived either from TEdenovo and homology-based search were classified using three different programs: PASTEC [86], a component of the REPET package; TEsorter, a recently described program that classifies TEs according to their conserved protein domains [87] and RepeatClassifier, a component of the RepeatModeler [88]. The consensus was then processed using RepeatMasker, which obtained the number of copies and base pairs covered by each TE of each genome. In the final TE dataset of each mosquito, we kept only TE consensus that masks regions of mosquito genome. This is needed to remove false-positive TE consensus. The dataset consisted of TEdenovo consensus and representative copies found by homology search (Available in Figshare repository - https://figshare.com/s/1ea991f1a2004c3f8fa0). We also extracted the Kimura 2-parameter distance (K2P) [89] from

each TE family using RepeatMasker's auxiliary scripts. This distance can be used to estimate the intragenomic age of these elements.

We performed an analysis of the correlation between the fraction of transposons and the genome size of each species studied. This was achieved using the cor() function (method = "spearman") in the R software. The correlation coefficient was used to test the strength of the correlation. Resulting graphics were created using the ggpubr package [90].

Horizontal transfer analysis

To determine the inheritance mode of the TEs characterized, either vertical or horizontal, we used the R software's VHICA package [33]. This package compares the relation between dS ratio and codon usage bias (CUB) linear regression of vertically-inherited single-copy orthologous host genes with transposable elements. A vertical transfer is the most likely scenario if the dS-CUB of a TE is not significantly different from that of the host genes. By contrast, a significant deviation in the host genes' dS-CUB values indicates horizontal transfer. We certified that the statistical assumptions of VHICA package are met since the residues of orthologous genes regression show a Gaussian distribution (Figure A and Figure B in S8 Fig), the pairwise species comparisons also show a linear correlation between dS and CUB (Figure C, Figure D and Figure F in S8 Fig) and linearity of the relationship between dS and CUB considering all pairwise comparisons (S9 Fig).

Fifty randomly selected single-copy orthologous genes from mosquito genomes were obtained from OrthoDB [91]. The ID of each of these genes was used to retrieve the nucleotide coding sequence for each gene from VectorBase (Single_copy_orthologous_genes.tar.gz File, Available in Figshare repository - https://figshare.com/s/1ea991f1a2004c3f8fa0). The sequences for each ortholog gene set were then codon-aligned using the MACSE software [92].

The TE sequence consensuses of all species recovered using TEdenovo, excluding chimeric elements, were submitted to clustering by the CD-HIT-est algorithm, with 80% identity and 80% coverage using the refinement parameter (-g 1) and global alignment. As a validation method, we extracted the copy that was most similar to each consensus (based on the best bit-score hit) by conducting a BLAST search for each consensus against the genome of its respective species. This created a second set of clusters with the same structure as the consensus clusters, using the TE copies instead of the TE consensus. To evaluate HTT, we also use some copies derived from the homology-based approach (see S1 File). Clusters of representative TE copies remaining in the analysis were as follows: i) those having sequences in at least two species; ii) those in which all sequences had at least 600 nucleotides; iii) those having at least one sequence with ORF codifying a polypeptide greater than 300 and 200 amino acids in size, for Class I and Class II TEs respectively. We used MACSE software to perform codon alignment for each cluster/family (as defined by Wicker et al. 2007), taking the nucleotide sequence of the largest ORF found among the clustered sequences as a reference (-seq parameter) and the remaining sequences as a FASTA file in the -seq_lr parameter. The flanking regions of sequences, based on the beginning and end of the reference sequence after alignment, were trimmed, as only the coding region was of interest for the dS-CUB analysis.

The alignment of the orthologous genes and TEs were then passed as input to the VHICA package to ascertain whether vertical/horizontal transfers occurred. Those TE clusters whose p-value was less than 0.01 in a one-tailed statistical test were considered horizontal transfer events. Additionally, the percentage of the genome involved in horizontal transfer events, as well as the K2P parameter for each TE family in each species, was calculated using the RepeatMasker software and its auxiliary scripts.

Although the species that make up the *An. gambiae* complex are morphologically indistinguishable, we decided to investigate each one separately, in view of the many natural pre-mating barriers that restrict species hybridization. When mating does occur, there is also evidence that male progeny is non-viable or sterile [93,94]. However, some possible introgression between some species was reported in previous studies, in these cases we compared the flanking region of TEs. In an HTT event, only the transposable element sequence is transferred to other species, in cases of introgression both TEs and flanking regions were exchanged [95]. HTT was considered if we found a positive signal of horizontal transfer calculated by VHICA and different TE flanking regions (3-5kb) between species.

To investigate whether any of the TE families that were involved in horizontal transfer events among the 24 studied mosquito species could also be involved in horizontal transfers to other species, we performed a blastn (dc-megablast) of the sequences of these TEs against: the NCBI nt database; all genomes of protostomes, plants, fungi, protists, flatworms, viruses, Echinodermata, Hemichordata, and Chordata organisms present in NCBI as of January 2019 (S4 File). Those matches that had a high degree of identity at the nucleotide level, with more than 80% of the mosquito's TE sequence coverage, and those that had a copy number greater than five were considered to be probable HTTs. When a species had more than one assembly from different samples, we dispensed with the need for a copy number if the matches were present in the majority of the different genome assemblies of this species.

Dating horizontal transfers

The dating of HTT events was performed by applying the formula $T = k/2r$ [96], where **T** is time, **k** is the synonym substitution rate (dS) between TE copies from two species, and **r** is the evolutionary rate of the species groups. We obtained dS estimates per mosquito taxon from the 50 single-copy ortholog genes as follows: 17.567×10^{-3} mutations per million years for transfers among mosquitoes from the *Anophelinae* subfamily, 9.205×10^{-3} for transfers within Culicinae subfamily of mosquitoes, and 10.006×10^{-3} for transfers between the Anophelinae and Culicinae subfamilies (S1 File). It should be noted that we found very few orthologous TE copies and these were restricted to species of the *An. gambiae* complex. As the speciation time is not well defined among these species, we performed HTT dating using host gene estimates.

Analysis of the geographical distribution of vector mosquitoes

The distribution region of each mosquito species of the genus *Anopheles* was taken from the distributions presented by Sinka et al. [97–100] and the distribution predicted using the *Malaria Atlas Project* [101]. These regions were considered the ancestral habitats of these mosquitoes since most *Anopheles* mosquitoes are not invasive species and thus disperse very little. We consider the islands in the southwestern Indian Ocean to be the ancestral habitat of *Ae. aegypti* [102] and the east and southeast Asian region extending to India to be the native habitat of *Ae. albopictus* [103]. To evaluate whether the overlap and/or proximity of the mosquitoes species distribution has any impact on the likelihood of the horizontal transfer, a point-biserial correlation analysis was performed using the cor.test () function of the R software, considering the number of transfers that occurred between two species of mosquitoes and their geographical overlap or non-overlapping distribution.

Supporting information

S1 File. Supplementary methods.
(DOCX)

S2 File. VHICA output images. Figures of VHICA graphs for each case of horizontal transposon transfer detected.
(PDF)

S3 File. Similarity matrix of Tc1-mariner. Pairwise similarity matrix of five TE families that are horizontally transferred among several of 24 studied species to different species.
(XLSX)

S4 File. Accession numbers and names of species that are searched for HHT by blast. The genome IDs, taxa, and species/virus names of more than 32,000 genomes that were searched for horizontally transferred mosquito TEs.
(PDF)

S5 File. Results of analysis performed on raw reads of *An. punctulatus* e *An. koliensis.*
(PDF)

S6 File. Fourteen additional putative horizontal transfer. The file shows HTT cases when we cannot differentiate HTT from an introgression, or HTT that represent removed in different genome assemblies of the same species.
(PDF)

S7 File. Results of blastn analysis between *Anopheles gambiae* and *Wuchereria bancrofti* genomes.
(XLSX)

S1 Table. List of mosquito genome assemblies used in this study.
(DOCX)

S1 Fig. Fraction of genome occupied by TEs characterized by both methods, *de novo* and homology-based.
(PNG)

S2 Fig. Mobilome fraction of each Wicher's system Orders.
(PNG)

S3 Fig. Intragenomic dating of the four most abundant TE Orders of mosquito genomes.
(PNG)

S4 Fig. P-values distribution of all positive pairwise comparisons. The majority of detected HTT signals are very significant.
(PNG)

S5 Fig. Landscape of TE families that undergone horizontal transfers. Each dot represents the relative age of each family in a mosquito genome.
(PNG)

S6 Fig. Horizontal transfer of a TE of R4 clade of R2 superfamily. Each red square represents a significant HTT pairwise comparison.
(PNG)

S7 Fig. Lack of correlation between geographical overlap and the number of horizontal transfer cases in mosquitoes. Each dot represents the number of horizontal transfers between to species.
(PNG)

S8 Fig. Mosquito dataset follow VHICA assumptions. Histogram (A) and cumulative probability curve (B) showing that the distribution of pooled residuals of linear regressions are Gaussian. Histograms (C, D, E, F) of some pairwise comparison were also showed, p-values of Kolmogorov–Smirnov test of normality confirm that we cannot reject normality. Relation between codon bias and dS can also be seen.
(PNG)

S9 Fig. Linearity relationship between dS and CUB. The plot shows the correlation of CUB of each pairwise comparison (Y axis) and synonymous substitution rate (X axis).
(JPEG)

Acknowledgments

Author Contributions

Conceptualization: Gabriel Luz Wallau.

Data curation: Elverson Soares de Melo, Gabriel Luz Wallau.

Formal analysis: Elverson Soares de Melo, Gabriel Luz Wallau.

Funding acquisition: Gabriel Luz Wallau.

Investigation: Elverson Soares de Melo.

Methodology: Elverson Soares de Melo.

Project administration: Gabriel Luz Wallau.

Resources: Gabriel Luz Wallau.

Software: Elverson Soares de Melo.

Supervision: Gabriel Luz Wallau.

Validation: Elverson Soares de Melo.

Visualization: Elverson Soares de Melo.

Writing – original draft: Elverson Soares de Melo, Gabriel Luz Wallau.

Writing – review & editing: Elverson Soares de Melo, Gabriel Luz Wallau.

References

1. McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci. 1950; 36: 344–355. https://doi.org/10.1073/pnas.36.6.344 PMID: 15430309

2. Ravindran S. Barbara McClintock and the discovery of jumping genes. Proc Natl Acad Sci. 2012; 109: 20198–20199. https://doi.org/10.1073/pnas.1219372109 PMID: 23236127

3. Biémont C, Vieira C. Genetics: Junk DNA as an evolutionary force. Nature. 2006; 443: 521–524. https://doi.org/10.1038/443521a PMID: 17024082

4. Cowley M, Oakey RJ. Transposable Elements Re-Wire and Fine-Tune the Transcriptome. PLOS Genet. 2013; 9: 1–7. https://doi.org/10.1371/journal.pgen.1003234 PMID: 23358118

5. Schrader L, Kim JW, Ence D, Zimin A, Klein A, Wyschetzki K, et al. Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat Commun. 2014; 5: 1–10. https://doi.org/10.1038/ncomms6495 PMID: 25510865

6. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017; 18: 71–86. https://doi.org/10.1038/nrg.2016.139 PMID: 27867194

7. Jangam D, Feschotte C, Betrán E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. 2017; 33: 817–831. https://doi.org/10.1016/j.tig.2017.07.011 PMID: 28844698

8. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007; 8: 973–982. https://doi.org/10.1038/nrg2165 PMID: 17984973

9. Kapitonov V V, Jurka J. A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet. 2008; 9: 411–412. https://doi.org/10.1038/nrg2165-c1 PMID: 18421312

10. Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst. 2019; 94: 233–252. https://doi.org/10.1266/ggs.18-00024 PMID: 30416149

11. Haapa-Paananen S, Wahlberg N, Savilahti H. Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. Mol Phylogenet Evol. 2014; 78: 271–274. https://doi.org/10.1016/j.ympev.2014.05.024 PMID: 24882428

12. Kapitonov V V., Jurka J. Rolling-circle transposons in eukaryotes. Proc Natl Acad Sci U S A. 2001; 98: 8714–8719. https://doi.org/10.1073/pnas.151269298 PMID: 11447285

13. McDonald JF. Evolution and consequences of transposable elements. Curr Opin Genet Dev. 1993; 3: 855–864. https://doi.org/10.1016/0959-437x(93)90005-a PMID: 8118210

14. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics. 1990; 124: 339–55. PMID: 2155157

15. Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci. 2017; 114: 4721–4726. https://doi.org/10.1073/pnas.1621178114 PMID: 28416702

16. Wallau GL, Vieira C, Loreto ÉLS. Genetic exchange in eukaryotes through horizontal transfer: Connected by the mobilome. Mob DNA. 2018; 9: 1–16. https://doi.org/10.1186/s13100-017-0106-z PMID: 29308092

17. Cummings MP. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol Evol. 1994; 9: 141–145. https://doi.org/10.1016/0169-5347(94)90179-1 PMID: 21236798

18. Wallau GL, Ortiz MF, Loreto ELS. Horizontal transposon transfer in eukarya: Detection, bias, and perspectives. Genome Biol Evol. 2012; 4: 689–699. https://doi.org/10.1093/gbe/evs055 PMID: 22798449

19. Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol. 2010; 25: 537–546. https://doi.org/10.1016/j.tree.2010.06.001 PMID: 20591532

20. Peccoud J, Cordaux R, Gilbert C. Analyzing Horizontal Transfer of Transposable Elements on a Large Scale: Challenges and Prospects. BioEssays. 2018; 40: 1–8. https://doi.org/10.1002/bies.201700177 PMID: 29283188

21. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics. Science (80-). 2009; 326: 1112–1115. https://doi.org/10.1126/science.1178534 PMID: 19965430

22. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu Z, et al. Genome Sequence of Aedes aegypti, a Major Arbovirus Vector. Science (80-). 2007; 316: 1718–1723. https://doi.org/10.1126/science.1138878 PMID: 17510324

23. Foster WA, Walker ED. Mosquitoes (Culicidae). Medical and Veterinary Entomology. Elsevier; 2019. pp. 261–325. https://doi.org/10.1016/B978-0-12-814043-7.00015–7

24. Benelli G, Mehlhorn H. Declining malaria, rising of dengue and Zika virus: insights for mosquito vector control. Parasitol Res. 2016; 115: 1747–1754. https://doi.org/10.1007/s00436-016-4971-z PMID: 26932263

25. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. Science (80-). 2015; 347: 1258522. https://doi.org/10.1126/science.1258522 PMID: 25554792

26. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, et al. Sequencing of Culex quinquefasciatus Establishes a Platform for Mosquito Comparative Genomics. Science (80-). 2010; 330: 86–88. https://doi.org/10.1126/science.1191864 PMID: 20929810

27. Chen X-G, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. Proc Natl Acad Sci. 2015; 112: E5907–E5915. https://doi.org/10.1073/pnas.1516410112 PMID: 26483478

28. Logue K, Small ST, Chan ER, Reimer L, Siba PM, Zimmerman PA, et al. Whole-genome sequencing reveals absence of recent gene flow and separate demographic histories for Anopheles punctulatus

mosquitoes in Papua New Guinea. Mol Ecol. 2015; 24: 1263–1274. https://doi.org/10.1111/mec.13107 PMID: 25677924

29. Marinotti O, Cerqueira GC, de Almeida LGP, Ferro MIT, Loreto EL da S, Zaha A, et al. The Genome of Anopheles darlingi, the main neotropical malaria vector. Nucleic Acids Res. 2013; 41: 7387–7400. https://doi.org/10.1093/nar/gkt484 PMID: 23761445

30. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015; 6: 11. https://doi.org/10.1186/s13100-015-0041-9 PMID: 26045719

31. Dotto BR, Carvalho EL, da Silva AF, Dezordi FZ, Pinto PM, Campos TDL, et al. HTT-DB: new features and updates. Database. 2018; 2018: 1–5. https://doi.org/10.1093/database/bax102 PMID: 29315358

32. Barrón MG, Paupy C, Rahola N, Akone-Ella O, Ngangue MF, Wilson-Bahun TA, et al. A new species in the major malaria vector complex sheds light on reticulated species evolution. Sci Rep. 2019; 9: 1–13. https://doi.org/10.1038/s41598-018-37186-2 PMID: 30626917

33. Wallau GL, Capy P, Loreto E, Le Rouzic A, Hua-Van A. VHICA, a New Method to Discriminate between Vertical and Horizontal Transposon Transfer: Application to the Mariner Family within Drosophila. Mol Biol Evol. 2016; 33: 1094–1109. https://doi.org/10.1093/molbev/msv341 PMID: 26685176

34. Biedler JK, Chen X, Tu Z. Horizontal transmission of an R4 clade non-long terminal repeat retrotransposon between the divergent Aedes and Anopheles mosquito genera. Insect Mol Biol. 2015; 24: 331–337. https://doi.org/10.1111/imb.12160 PMID: 25615532

35. Tang C, Davis KE, Delmer C, Yang D, Wills MA. Elevated atmospheric CO2 promoted speciation in mosquitoes (Diptera, Culicidae). Commun Biol. 2018; 1: 1–8. https://doi.org/10.1038/s42003-017-0002-6 PMID: 29809203

36. Gilbert C, Feschotte C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. Curr Opin Genet Dev. 2018; 49: 15–24. https://doi.org/10.1016/j.gde.2018.02.007 PMID: 29505963

37. Bennetzen JL, Wang H. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. Annu Rev Plant Biol. 2014; 65: 505–530. https://doi.org/10.1146/annurev-arplant-050213-035811 PMID: 24579996

38. Ito H, Kakutani T. Control of transposable elements in Arabidopsis thaliana. Chromosom Res. 2014; 22: 217–223. https://doi.org/10.1007/s10577-014-9417-9 PMID: 24801341

39. Hirano T, Siomi H. Small RNAs: Artificial piRNAs for Transcriptional Silencing. Curr Biol. 2015; 25: R280–R283. https://doi.org/10.1016/j.cub.2015.02.009 PMID: 25829012

40. Saito K, Siomi MC. Small RNA-Mediated Quiescence of Transposable Elements in Animals. Dev Cell. 2010; 19: 687–697. https://doi.org/10.1016/j.devcel.2010.10.011 PMID: 21074719

41. Platt RN, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. Chromosom Res. 2018; 26: 25–43. https://doi.org/10.1007/s10577-017-9570-z PMID: 29392473

42. Friedli M, Trono D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. Annu Rev Cell Dev Biol. 2015; 31: 429–451. https://doi.org/10.1146/annurev-cellbio-100814-125514 PMID: 26393776

43. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The Genome Sequence of the Malaria Mosquito Anopheles gambiae. Science (80-). 2002; 298: 129–149. https://doi.org/10.1126/science.1076181 PMID: 12364791

44. Goerner-Potvin P, Bourque G. Computational tools to unmask transposable elements. Nat Rev Genet. 2018; 19: 688–704. https://doi.org/10.1038/s41576-018-0050-x PMID: 30232369

45. Berthelier J, Casse N, Daccord N, Jamilloux V, Saint-Jean B, Carrier G. A transposable element annotation pipeline and expression analysis reveal potentially active elements in the microalga Tisochrysis lutea. BMC Genomics. 2018; 19: 1–14. https://doi.org/10.1186/s12864-017-4368-0 PMID: 29291715

46. Marsano RM, Leronni D, D'Addabbo P, Viggiano L, Tarasco E, Caizzi R. Mosquitoes LTR Retrotransposons: A Deeper View into the Genomic Sequence of Culex quinquefasciatus. PLoS One. 2012; 7: e30770. https://doi.org/10.1371/journal.pone.0030770 PMID: 22383973

47. Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 2019; 19: 11. https://doi.org/10.1186/s12862-018-1324-9 PMID: 30626321

48. Chalopin D, Naville M, Plard F, Galiana D, Volff JN. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol Evol. 2015; 7: 567–580. https://doi.org/10.1093/gbe/evv005 PMID: 25577199

49. Lu J, Peatman E, Tang H, Lewis J, Liu Z. Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. BMC Genomics. 2012; 13: 246. https://doi.org/10.1186/1471-2164-13-246 PMID: 22702965

50. Wang N, Xiang Y, Fang L, Wang Y, Xin H, Li S. Patterns of Gene Duplication and Their Contribution to Expansion of Gene Families in Grapevine. Plant Mol Biol Report. 2013; 31: 852–861. https://doi.org/10.1007/s11105-013-0556-5

51. Silva JC, Loreto EL, Clark JB. Factors that affect the horizontal transfer of transposable elements. Curr Issues Mol Biol. 2004; 6: 57–71. PMID: 14632259

52. Baidouri M El, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, et al. Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res. 2014; 24: 831–838. https://doi.org/10.1101/gr.164400.113 PMID: 24518071

53. Metzger MJ, Paynter AN, Siddall ME, Goff SP. Horizontal transfer of retrotransposons between bivalves and other aquatic species of multiple phyla. Proc Natl Acad Sci. 2018; 115: E4227–E4235. https://doi.org/10.1073/pnas.1717227115 PMID: 29669918

54. Gao D, Chu Y, Xia H, Xu C, Heyduk K, Abernathy B, et al. Horizontal transfer of Non-LTR retrotransposons from arthropods to flowering plants. Mol Biol Evol. 2018; 35: 354–364. https://doi.org/10.1093/molbev/msx275 PMID: 29069493

55. Oliveira SG, Bao W, Martins C, Jurka J. Horizontal transfers of Mariner transposons between mammals and insects. Mob DNA. 2012; 3: 1–6. https://doi.org/10.1186/1759-8753-3-1 PMID: 22277150

56. Gilbert C, Schaack S, Pace II JK, Brindley PJ, Feschotte C. A role for host–parasite interactions in the horizontal transfer of transposons across phyla. Nature. 2010; 464: 1347–1350. https://doi.org/10.1038/nature08939 PMID: 20428170

57. Palazzo A, Lorusso P, Miskey C, Walisko O, Gerbino A, Marobbio CMT, et al. Transcriptionally promiscuous "blurry" promoters in Tc1/mariner transposons allow transcription in distantly related genomes. Mob DNA. 2019; 10: 1–11. https://doi.org/10.1186/s13100-018-0144-1 PMID: 30622655

58. Reiss D, Mialdea G, Miele V, de Vienne DM, Peccoud J, Gilbert C, et al. Global survey of mobile DNA horizontal transfer in arthropods reveals Lepidoptera as a prime hotspot. Buerkle A, editor. PLOS Genet. 2019; 15: e1007965. https://doi.org/10.1371/journal.pgen.1007965 PMID: 30707693

59. Manguin S, Bangs MJ, Pothikasikorn J, Chareonviriyaphap T. Review on global co-transmission of human Plasmodium species and Wuchereria bancrofti by Anopheles mosquitoes. Infect Genet Evol. 2010; 10: 159–177. https://doi.org/10.1016/j.meegid.2009.11.014 PMID: 19941975

60. Suh A, Witt CC, Menger J, Sadanandan KR, Podsiadlowski L, Gerth M, et al. Ancient horizontal transfers of retrotransposons between birds and ancestors of human pathogenic nematodes. Nat Commun. 2016; 7: 11396. https://doi.org/10.1038/ncomms11396 PMID: 27097561

61. Van den Berg H, Kelly-Hope LA, Lindsay SW. Malaria and lymphatic filariasis: The case for integrated vector management. Lancet Infect Dis. 2013; 13: 89–94. https://doi.org/10.1016/S1473-3099(12)70148-2 PMID: 23084831

62. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al. FlyBase 2.0: the next generation. Nucleic Acids Res. 2019; 47: D759–D765. https://doi.org/10.1093/nar/gky1003 PMID: 30364959

63. Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature. 2011; 479: 487–492. https://doi.org/10.1038/nature10640 PMID: 22113690

64. Giraldo-Calderón 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 Res. 2015; 43: D707–D713. https://doi.org/10.1093/nar/gku1117 PMID: 25510499

65. Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, et al. Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res. 2016; 44: D73–D80. https://doi.org/10.1093/nar/gkv1226 PMID: 26578580

66. Sharakhova M V., Hammond MP, Lobo NF, Krzywinski J, Unger MF, Hillenmeyer ME, et al. Update of the Anopheles gambiae PEST genome assembly. Genome Biol. 2007; 8: R5. https://doi.org/10.1186/gb-2007-8-1-r5 PMID: 17210077

67. Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, et al. Widespread Divergence Between Incipient Anopheles gambiae Species Revealed by Whole Genome Sequences. Science (80-). 2010; 330: 512–514. https://doi.org/10.1126/science.1195755 PMID: 20966253

68. Ghurye J, Koren S, Small ST, Redmond S, Howell P, Phillippy AM, et al. A chromosome-scale assembly of the major African malaria vector Anopheles funestus. Gigascience. 2019; 8: 1–8. https://doi.org/10.1093/gigascience/giz063 PMID: 31157884

69. Chida AR, Ravi S, Jayaprasad S, Paul K, Saha J, Suresh C, et al. A near-chromosome level genome assembly of Anopheles stephensi. bioRxiv:063040 [preprint]. 2020. https://doi.org/10.1101/2020.04.27.063040

70. 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: 501–507. https://doi.org/10.1038/s41586-018-0692-z PMID: 30429615

71. Kingan SB, Heaton H, Cudini J, Lambert CC, Baybayan P, Galvin BD, et al. A high-quality de novo genome assembly from a single mosquito using pacbio sequencing. Genes (Basel). 2019; 10. https://doi.org/10.3390/genes10010062 PMID: 30669388

72. Palatini U, Masri RA, Cosme L V., Koren S, Thibaud-Nissen F, Biedler JK, et al. Improved reference genome of the arboviral vector Aedes albopictus. Genome Biol. 2020; 21: 215. https://doi.org/10.1186/s13059-020-02141-w PMID: 32847630

73. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 2013; 41: e129–e129. https://doi.org/10.1093/nar/gkt371 PMID: 23661685

74. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010; 59: 307–321. https://doi.org/10.1093/sysbio/syq010 PMID: 20525638

75. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017; 34: 2422–2424. https://doi.org/10.1093/molbev/msx149 PMID: 28472384

76. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2015. Available: http://www.repeatmasker.org

77. Zytnicki M, Akhunov E, Quesneville H. Tedna: A transposable element de novo assembler. Bioinformatics. 2014; 30: 2656–2658. https://doi.org/10.1093/bioinformatics/btu365 PMID: 24894500

78. Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). Genome Biol Evol. 2015; 7: 1192–1205. https://doi.org/10.1093/gbe/evv050 PMID: 25767248

79. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007; 35: W265–W268. https://doi.org/10.1093/nar/gkm286 PMID: 17485477

80. Bao Z, Eddy SR. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. 2002; 12: 1269–1276. https://doi.org/10.1101/gr.88502 PMID: 12176934

81. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005; 21: 351–358. https://doi.org/10.1093/bioinformatics/bti1018 PMID: 15961478

82. Edgar RC, Myers EW. PILER: Identification and classification of genomic repeats. Bioinformatics. 2005; 21: 152–158. https://doi.org/10.1093/bioinformatics/bth487 PMID: 15377504

83. Novák P, Neumann P, Pech J, Steinhaisl J, MacAs J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013; 29: 792–793. https://doi.org/10.1093/bioinformatics/btt054 PMID: 23376349

84. Flutre T, Duprat E, Feuillet C, Quesneville H. Considering transposable element diversification in de novo annotation approaches. PLoS One. 2011; 6: 1–15. https://doi.org/10.1371/journal.pone.0016526 PMID: 21304975

85. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22: 1658–1659. https://doi.org/10.1093/bioinformatics/btl158 PMID: 16731699

86. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, et al. PASTEC: An automatic transposable element classification tool. PLoS One. 2014; 9. https://doi.org/10.1371/journal.pone.0091929 PMID: 24786468

87. Zhang R-G, Wang Z-X, Ou S, Li G-Y. TEsorter: lineage-level classification of transposable elements using conserved protein domains. bioRxiv. 2019; 800177. https://doi.org/10.1101/800177

88. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020; 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117 PMID: 32300014

89. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980; 16: 111–120. https://doi.org/10.1007/BF01731581 PMID: 7463489

90. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. 2018. Available: https://rpkgs.datanovia.com/ggpubr/

91. Kriventseva E V., Rahman N, Espinosa O, Zdobnov EM. OrthoDB: the hierarchical catalog of eukaryotic orthologs. Nucleic Acids Res. 2007; 36: D271–5. https://doi.org/10.1093/nar/gkm845 PMID: 17947323

92. Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One. 2011;6. https://doi.org/10.1371/journal.pone.0022594 PMID: 21949676

93. Slotman M, della Torre A, Powell JR. The Genetics of Inviability and Male Sterility in Hybrids Between Anopheles gambiae and An. arabiensis. Genetics. 2004; 167: 275–287. https://doi.org/10.1534/genetics.167.1.275 PMID: 15166154

94. Davidson G, Paterson HE, Coluzzi M, Mason GF, Micks DW. The Anopheles gambiae complex. In: Wright JW, Pal R, editors. Genetics of Insect Vector of Disease. Amsterdam: Elsevier Publishers; 1967. pp. 211–250.

95. Fortune PM, Roulin A, Panaud O. Horizontal transfer of transposable elements in plants. Commun Integr Biol. 2008; 1: 74–77. https://doi.org/10.4161/cib.1.1.6328 PMID: 19513203

96. Graur D, Li W-H. Fundamentals of Molecular Evolution. 2nd ed. Sunderland: Sinauer Associates; 2000.

97. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al. The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010; 3: 117. https://doi.org/10.1186/1756-3305-3-117 PMID: 21129198

98. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, et al. The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2011; 4: 89. https://doi.org/10.1186/1756-3305-4-89 PMID: 21612587

99. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, et al. The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010; 3: 72. https://doi.org/10.1186/1756-3305-3-72 PMID: 20712879

100. Sinka ME. Global Distribution of the Dominant Vector Species of Malaria. Anopheles mosquitoes—New insights into malaria vectors. InTech; 2013. pp. 109–143. https://doi.org/10.5772/54163

101. Hay SI, Snow RW. The Malaria Atlas Project: Developing global maps of malaria risk. PLoS Med. 2006; 3: 2204–2208. https://doi.org/10.1371/journal.pmed.0030473 PMID: 17147467

102. Soghigian J, Gloria-Soria A, Robert V, Le Goff G, Failloux A, Powell JR. Genetic evidence for the origin of Aedes aegypti, the yellow fever mosquito, in the southwestern Indian Ocean. Mol Ecol. 2020; mec.15590. https://doi.org/10.1111/mec.15590

103. Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species Aedes albopictus: Current knowledge and future perspectives. Trends Parasitol. 2013; 29: 460–468. https://doi.org/10.1016/j.pt.2013.07.003 PMID: 23916878

104. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017; 34: 1812–1819. https://doi.org/10.1093/molbev/msx116 PMID: 28387841

105. Rivera-Vega L, Mittapalli O. Molecular characterization of mariner-like elements in emerald ash borer, agrilus planipennis (Coleoptera, Polyphaga). Arch Insect Biochem Physiol. 2010; 74: 205–216. https://doi.org/10.1002/arch.20357 PMID: 20602451

106. Robertson HM, Lampe DJ. Distribution of transposable elements in arthropods. Annu Rev Entomol. 1995; 40: 333–357. https://doi.org/10.1146/annurev.en.40.010195.002001 PMID: 7529010

107. Green CL, Frommer M. The genome of the Queensland fruit fly Bactrocera tryoni contains multiple representatives of the mariner family of transposable elements. Insect Mol Biol. 2001; 10: 371–386. https://doi.org/10.1046/j.0962-1075.2001.00275.x PMID: 11520360

108. Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T, Kimura K. Possible horizontal transfer of a transposable element from host to parasitoid. Mol Biol Evol. 2001; 18: 1952–1958. https://doi.org/10.1093/oxfordjournals.molbev.a003735 PMID: 11557800

109. Biedler JK, Shao H, Tu Z. Evolution and horizontal transfer of a DD37E DNA transposon in mosquitoes. Genetics. 2007; 177: 2553–2558. https://doi.org/10.1534/genetics.107.081109 PMID: 17947403

110. Diao Y, Qi Y, Ma Y, Xia A, Sharakhov I, Chen X, et al. Next-generation sequencing reveals recent horizontal transfer of a DNA transposon between divergent mosquitoes. PLoS One. 2011;6. https://doi.org/10.1371/journal.pone.0016743 PMID: 21379317