Genomic evidence for adaptive evolution of the invasive Asian tiger mosquito towards temperate environment

Clément Goubert[A], Hélène Henri[A], Guillaume Minard[B,1], Claire Valiente Moro[B], Patrick Mavingui[B,C], Cristina Vieira[A], and Matthieu Boulesteix[A]

[A]Université de Lyon, F-69622, Lyon, France; Université Claude Bernard Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive, UMR5558, F-69100 Villeurbanne

[B]Université de Lyon, F-69622, Lyon, France; Université Lyon 1, Villeurbanne, France; CNRS, UMR 5557, Ecologie Microbienne, Villeurbanne, France; INRA, UMR 1418, Villeurbanne, France

[C]Université de La Réunion, UMR PIMIT, INSERM 1187, CNRS 9192, IRD 249, Plateforme Technologique CYROI, Sainte-Clotilde, La Réunion.

Corresponding author: Matthieu Boulesteix, Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, INRIA, VetAgro Sup, Université Claude Bernard Lyon 1, Villeurbanne, France, +334 72 43 29 16, matthieu.boulesteix@univ-lyon1.fr

1Present address: Metapopulation Research Center, Department of Biosciences, University of Helsinki, Helsinki, Finland

Abstract

Invasive species represent unique opportunities to evaluate the role of local adaptation during colonization of new environments. Among these, the Asian tiger mosquito, *Aedes albopictus*, is a threatening vector of several human viral diseases, including dengue, chikungunya and the emerging Zika fevers. Its broad presence in both temperate and tropical environments has sometimes been considered as the reflect of a great "ecological plasticity". However, no study has been conducted to assess the role of adaptive evolution in the ecological success of *Ae. albopictus* at the molecular level. In the present study we performed a genomic scan to search for potential signatures of selection leading to local adaptation in a hundred of field collected mosquitoes from native populations of Vietnam and temperate invasive populations of Europe. High throughput genotyping of transposable element insertions generated more than 120 000 polymorphic loci, which in their great majority revealed a virtual absence of structure between biogeographic areas. Nevertheless, 92 outlier loci show a high level of differentiation between temperate and tropical populations. The majority of these loci segregates at high insertion frequencies among European populations, indicating that this pattern could have been caused by recent events of adaptive evolution in temperate areas. Six outliers were located near putative diapause effector genes, suggesting fine tunning of this critical pathway during local adaptation.

Keywords: Invasive species, *Aedes albopictus*, local adaptation, genome scan, Transposable Elements, diapause

Significance statement

Few empirical data have been gathered to evaluate the importance of rapid adaptation in the ecological success of invasive species. We investigated whether adaptation has facilitated the invasion of *Aedes albopictus* in temperate environments. This species, that already transmits Yellow fever and Chikungunya viruses, is also competent for the Zika virus, posing dramatic sanitary consequences given the current species distribution. Using mobile genetic elements as dense genetic markers, we identified footprints of rapid adaptation in wild temperate populations. In depth analyses suggests that the diapause pathway could be a target of natural selection during invasion.

Biological invasions represent unique opportunities to study fast evolutionary changes such as adaptive evolution. Indeed, settlement in a novel area represents a biological challenge that invasive species have successfully overcome. The underlying processes could be studied at the molecular level, particularly to gather empirical knowledge about the genetics of invasions, a field of study that has produced extensive theoretical predictions, but for which there is still little evidence in nature (1). Some of the main concerns are to disentangle the effects of neutral processes during colonization, such as founder events or allele surfing at the migration front, from adaptive evolution (i.e. local adaptation, (1–3)).

Adaptation can arise either through the appearance and spread of a new beneficial mutation, the spread of a favorable allele from standing genetic variation, or from hybridization in the introduction area (1, 4, 5). Detection of the footprint of natural selection is however dependent on the availability of informative genetic markers, which should provide a substantial coverage of the genome to allow selection scans and be easily and confidently scored across many individuals. Unfortunately, invasive organisms are rarely model species, making the development of a reliable and efficient marker challenging.

The Asian tiger mosquito, *Aedes* (*Stegomya*) *albopictus* (Diptera:Culicidae) is currently one of the most threatening invasive species (Invasive Species Specialist Group); originating from South-Eastern Asia, this species is one of the primary vectors of Dengue and Chikungunya viruses, and is also involved in the transmission of other threatening arboviruses (6), in particular the newly emerging Zika virus (7–9). *Ae. albopictus* has now settled in every continent except Antarctica, and is found both under tropical and temperate climates (10). While this species is supposed to originate from rain forests of South-Eastern Asia (11), the native area of *Ae. albopictus* encompasses contrasted environments including temperate regions of Japan and China, offering a

large potential of fit towards newly colonized environments. For example, the induction of photoperiodic diapause in temperate areas, that has a genetic basis in *Ae. albopictus* (12, 13), is decisive to ensure invasive success in Europe or Northern America. Indeed it allows the sensible populations to survive through winter at the larval stage into the eggs. Such a trait appears governed by a “genetic toolkit” involving numerous genes and metabolic networks, for which however the genetic polymorphism between diapausing and non-diapausing strains remains to be elucidated (14). In addition, the colonization of new areas that look similar at first glance can still involve *de novo* adaptation: indeed, even environment sharing climatic variables are not necessarily similar regarding edaphic and biotic interactions (1). Hence, this suggests that whatever are the native and settled environment, it might be possible to find evidence of adaptive evolution in invasive populations of *Ae. albopictus*.

To better understand the invasive success of this species, we genotyped 140 field individuals, collected from three Vietnamese (native tropical area) and five European (invasive temperate area) populations, aiming to identify genomic regions involved in local adaptation. To do so, we developed new genetic markers, based on high throughput genotyping of the insertion of Transposable Elements (TEs), which are highly prevalent and polymorphic in the *Ae. albopictus* genome.

To distinguish between neutral demographic effects and adaptive evolution, we first performed population genetic analyses to reveal the global genetic structure of the studied populations. We then performed a genomic scan for selection and identified 92 candidate loci under directional selection, among which several can be located in the neighborhood of diapause related genes.

Results

Rationale for using TE markers to identify regions under natural selection

Transposable Elements are mobile genetic elements, capable to insert at new loci from one generation to another throughout transposition in the germ line. TEs represent at least one third of the genome of *Ae. albopictus* and include recently active families that can reach thousands of copies (15). Identification of these markers represent a seducing alternative to other methods of diversity reduction, such as RAD-sequencing (16), that could be less efficient in species with high TE load (17) . Finally, in mosquitoes, TEs have been shown to be powerful markers for both population structure analysis (18–21) and genome scans (22).

Identification of TE insertions is particularly efficient to obtain a large number of genetic markers throughout a genome (22), especially if few genomic resources are available (23), which, until recently, was case for the Asian tiger mosquito. We hypothesized that some TE insertion sites could be located at the neighborhood of natural selection targets and thus could reach a high levels of differentiation between native and invasive populations if selective sweeps occurred during local adaptation. In addition, some TEs could also insert near or inside coding regions and thence be directly involved in environmental adaptation (24), eventually contributing to the success of invasive species (25).

High throughput TE insertion genotyping. A total of 140 individuals were collected in Europe (invasive temperate populations) and Vietnam (native tropical), and screened for their insertion polymorphism of five highly repeated families of TEs using paired-end Illumina sequencing (see Material and Methods). Briefly, individual insertions of five TE

families (IL1, L2B, RTE4, RTE5 and Lian1) were genotyped using Transposon Display (TD) (26), a TE insertion specific PCR method, combined with Illumina sequencing of all TD amplification products. TEs used for this study are non-LTR (LINE) retrotransposons that usually do not show a specific insertion site preference (31), making them likely to be well dispersed in the genome. Sequencing produced a total of 102,319,300 paired-end reads (2x101bp). After quality and specificity filtering 24,332,715 reads were suitable for analyses. Because of read coverage variation between individuals, we applied a read sampling procedure before the recovery of individual insertion loci by clustering (see SI Material and Methods); to ensure the consistency of this procedure, sampling and subsequent analysis were performed independently three times. On average, a total number of 128,491 polymorphic insertion loci were available for each of the three sampling replicates. The mean number of loci per individual and per TE family ranged from 1025 ± 290 s.d. (IL1 family, mean and s.d. averaged over the three replicates) to 3266 ± 766 s.d. (RTE5 family). Details are given in Table S1. While our read sampling procedure could have artificially lowered the mean insertion frequency of the loci, this effect should be small because in our final datasets, the TE insertion frequencies (i. e. the number of individuals that share an insertion) are not correlated with the mean number of read per individual at the considered locus (Figure S1).

Population structure. Principal Coordinate Analyses (PcoAs) were performed independently for each of the five TEs (see Figure 1 for an example with RTE4 and Figure S2 for all other TEs and replicates). Among the three main Principal Coordinates (PCs), individuals tend to be grouped according to their respective populations with little overlap between groups. However, the three main PCs represent only a small fraction of total genetic variation (< 10%), suggesting a weak genetic structuring between the populations. Overall, individuals from Vietnamese populations (HCM, TA, VT) tend to be

grouped together in a single cluster, at the exception of 13 to 14 individuals from HCM when using L2B and RTE5 TE families (Figure S2), along with six individuals of VT with the RTE4 TE family (Figure 1) that can not be clearly distinguished from European samples. BCN individuals (Spain) represent the most homogeneous group, well differentiated from Vietnamese and French individuals (SP, CGN, NCE and PLV). In agreement with PCoAs, Analyses of Molecular Variance (AMOVAs (27)) attributed very few genetic variances among groups (Vietnam-Europe) and between populations within groups (Table 1). In the studied populations, most of the genetic variance was distributed among individuals within groups.

Measures of genetic differentiation among pairs of populations were consistent with PCoAs and AMOVAs (SI File 1): the BCN population shows the highest F_{ST} values with the other populations for each of the five TEs ($0.051 < F_{ST} < 0.148$), while Vietnamese populations were the most closely related ($0.011 < F_{ST} < 0.032$). While VT is located 100 km away from TA and HCM (both sampled in the same city, Hô Chi Minh, Vietnam) the F_{ST} values are very similar between the three Vietnamese populations, suggesting no influence of geography at this scale. CGN and NCE, sampled in the same urban area (Nice agglomeration), are also little or not significantly differentiated depending on the TE family. The previously identified intermediate pattern of HCM with some European populations at L2B and RTE5 loci (PCoAs analyses) is also found at the F_{ST} level, especially regarding the low differentiation with the PLV population for these markers ($0.011 < F_{ST} < 0.020$).

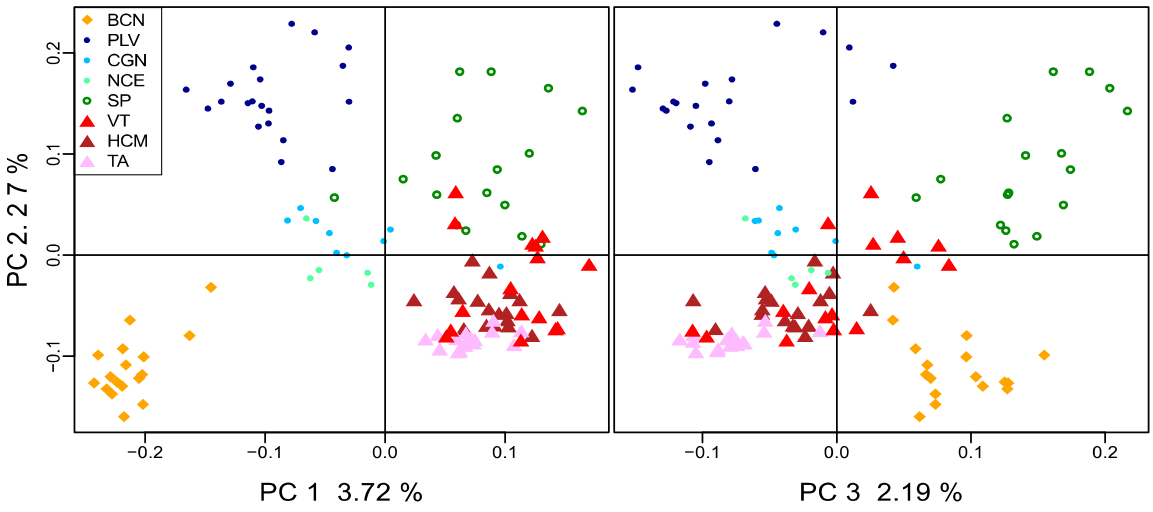

Figure 1. Projection of individuals over the three first principal coordinates (PC) of principal Coordinates Analyses (PCoAs) computed over all RTE4 loci for the replicates (M1). Proportion of inertia represented by each axes is noted in %. circles: European populations; triangles: Vietnamese populations. Results for other TE families and each sampling replicate can be found in Figure S2.

Table 1. Analyses of Molecular Variance (AMOVAs) for the three replicates (M1,M2,M3) of read sampling for the five TE families (IL1, L2B, RTE5, RTE4, Lian1). Values are given in percentage of the total genetic variance

	IL1	L2B	RTE5	RTE4	Lian1
Among groups	[0.59-0.70]	[1.22-1.29]	[1.08-1.10]	[1.97-2.04]	[0.67-0.74]
Among populations within groups	[5.15-5.37]	[3.58-3.63]	[3.36-3.40]	[6.67-6.78]	[4.47-4.56]
Within populations	[94.04-94.16]	[95.08-95.18]	[95.51-95.55]	[91.18-91.30]	[94.77-94.81]

intervals reports min and max values among the 3 sampling replicates

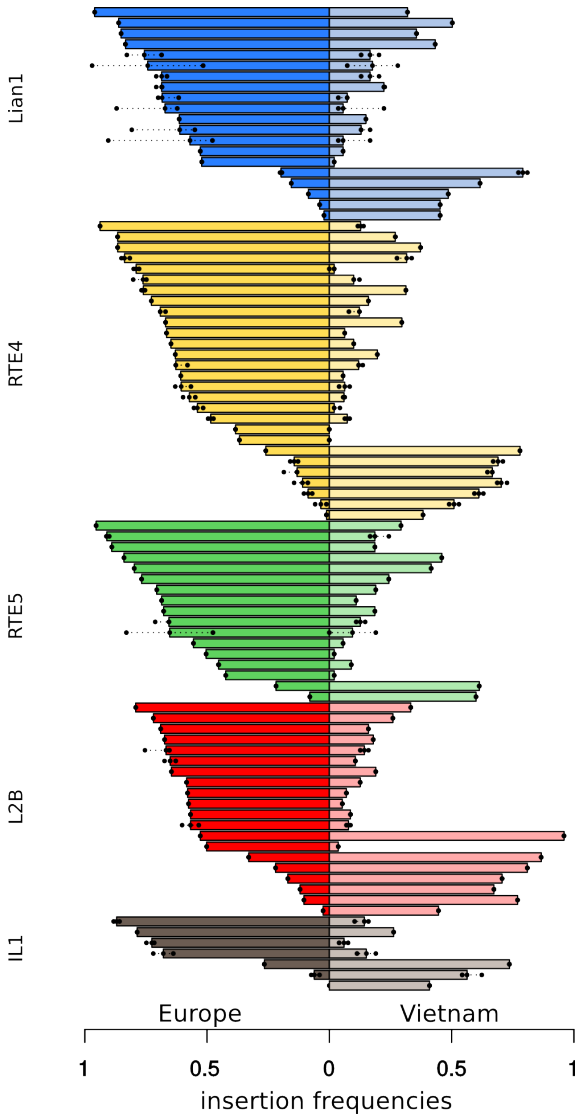

Figure 2. Insertion frequencies in Europe and Vietnam for the 92 outlier loci. Bars represent the median value from the three reads sampling replicates and dots the values from the other replicates (if outlier found in replicate). Colors correspond to each of the 5 TE families.

Genomic scan. Research of outlier loci for both selection signature using Bayescan (non-hierarchical island model), and for significant F_{CT} (between Europe-Vietnam group differentiation) identified 92 candidate insertion loci (Figure 2). Most of these insertions

are found in both areas (no private allele), except for RTE4_6 and RTE4_7 that were not found in Vietnam. In addition, a majority of outliers corresponds to high frequency insertions in Europe, while the same trend is not observed at 92 randomly chosen loci among those having the same minimum insertion frequency (≥ 20 individuals/locus) between Europe and Vietnam (Figure 3). PCR amplification of the outlier loci were carried out on a representative panel of 47 individuals to validate the insertion pattern detected by TD (see SI Material and Method). For loci where the amplification was successful, the insertion pattern observed by PCR always confirmed the results from TD (Figure S3).

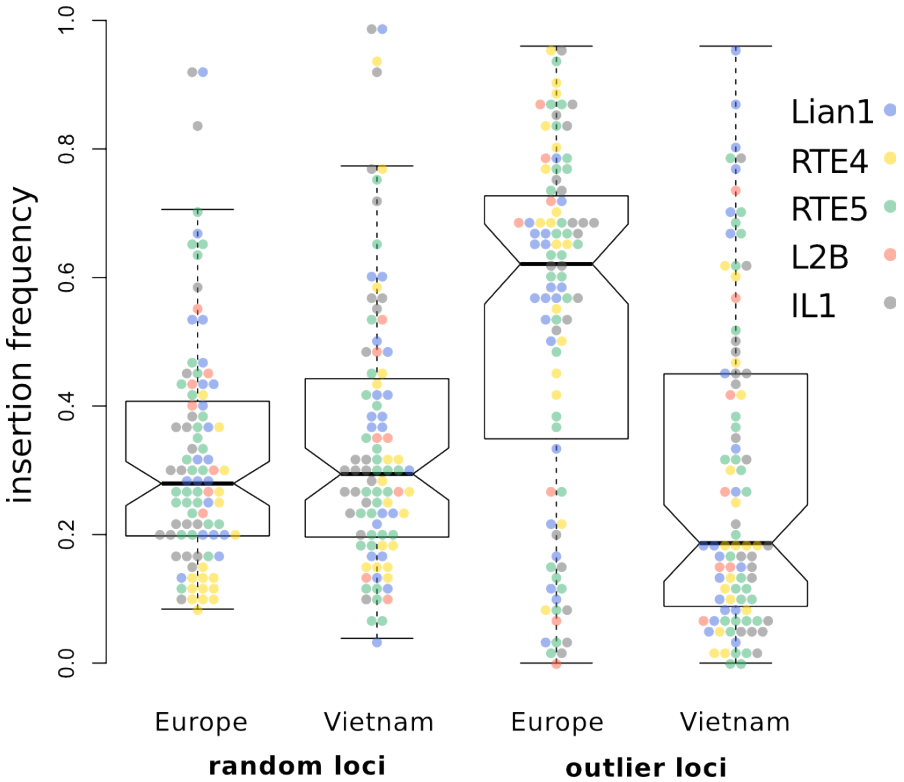

Figure 3. Insertion frequencies of 92 randomly chosen loci among those having the same minimum insertion frequency (≥20 individuals) as outliers compared to the 92 outlier loci. Random loci were taken from the first replicate (M1) and values for outliers are median values obtained among the three replicates. Non-overlapping notches indicate a significant difference between the true medians (thick dark horizontal bars).

From 92 outlier loci, 21 could be attributed to a unique position on the *Ae. albopictus* genome (28). Annotation and distance to surrounding genes are reported on SI File 2. We found that six outliers (SI File S2, highlighted) loci are located on contigs that harbor genes previously identified in *Ae. albopictus* as being differentially expressed between diapause-induced and non diapause-induced samples , and for which orthologs in *Drosophila melanogaster* are known to be part of well-identified functional networks (14). All these six loci are found to be outliers because of their high insertion frequencies in Europe compared with Vietnam.

Discussion

The goal of our study was to identify genomic regions involved in adaptive evolution of *Ae. albopictus* thanks to the development of new genetic markers. Through high-throughput genotyping of five TE families insertion polymorphisms, we identified up to 128,617 polymorphic loci among a hundred of individuals from eight sampling sites. The estimated genome size of *Ae. albopictus* exceeds one billion base-pairs (15, 29, 30). Accordingly, the amount of markers scored in this study offers a comfortable genomic density of one marker every 10 kb.

We provide here a new and cost efficient method to quickly generate a large amount of polymorphic markers without extensive knowledge about one species genome. Specifically, this strategy could be extremely valuable for species with a large genome size, where TE density could severely compromise the development of more classical approaches, such as the very popular RAD-sequencing (16).

The genetic structure of the studied populations showed strong consistency between sampling replicates of individuals' reads, demonstrating the robustness of the method in spite of an initial substantial coverage variation among individuals. Population genetics analyses revealed a very low level of genetic structuring between European and

Vietnamese populations. Among the studied populations, AMOVAs showed that most of the genetic variation is distributed between individuals within populations (> 90%), and as suggested by pairwise F_{ST} and PCoAs, only a small part (< 10%) of the genetic variance is due to differentiation between populations. The genetic differentiation we measured is indeed as high among European populations as it is between populations from Europe and Vietnam.

This singular population structure is in agreement with previous results gathered in *Ae. albopictus* using different collections of allozymes, mtDNA or microsatellites markers (31–35). Moreover, a recent analysis performed with a set of 11 microsatellites on individuals from the same populations (at the exception of BCN) showed a similar distribution of genetic variation among hierarchical levels (36). These results demonstrate the reliability of our markers and confirm that a non-hierarchical island model can likely fit the global genetic structure. The observed genetic diversity is compatible with a scenario of multiple and independent introductions, as already suggested for *Ae. albopictus* (37–40). However, as previously suggested, this pattern could also be the result of founder events that may occur during colonization and/or a restriction of gene flow between populations consecutive to their introduction. Answering such a question would require an extended sampling all over the native area.

Outlier analysis revealed 92 loci with high posterior probabilities of being under positive selection between European and Vietnamese populations. When possible, the PCR amplification of the outlier loci using a set of representative individuals always confirmed a shift of insertion frequencies toward either the European or the Vietnamese sampling sites. This suggests that in spite of a reduced coverage, introduced by sampling in the dataset, the scored insertion polymorphisms are reliable. In addition, our method of analysis is likely to be conservative: the Bayescan outliers were selected for their consistency with a significant F_{CT} between European temperate and Vietnamese tropical

populations, which avoid retaining outliers that we were not looking for, for example those due to a population specific event.

We were able to assign a unique position for 21 of the outlier loci on the *Ae. albopictus* genome. As expected by the *Ae. albopictus* genomic composition (15, 28, 41), an important part of the other outlier loci were located in repeated regions (44,6% of total), despite our efforts to remove *a priori* loci occurring in known transposable elements. Since the *Ae albopictus* genome publication is very recent, no gene set or other genome annotation are currently available. We thus took advantage of the *Ae. albopictus* transcriptome data to annotate regions surrounding the detected ouliers.(14). We found six outliers located on contigs which also harbors genes that are differentially expressed between individuals induced for diapause and controls (14), two of them, being located either in an intron (RTE4_7442) or within 3kb (RTE4_17015) of these candidate genes. It is worth mentioning that these two genes belong to the the same functional group (GO:0005576 extracellular region). Diapause is a critical developmental stage found only in temperate populations of the Asian tiger mosquito. Interestingly, this functional pathway has been shown to benefit from fast adjustments thanks to local adaptation. For instance, Urbanski *et al.* (42) showed that invasive American populations originating from Japan have rapidly evolved a new adaptive clinal response to diapause induction, independent from that observed in the native area. Thus, adaption in the temperate regions could have led to several selective sweeps on gene or regulatory sequences involved in this critical pathway, allowing the settlement of the mosquito in new temperate areas.

Interestingly, and as it is the case for these six outliers, we found significantly more outlier loci with a high frequency in Europe and low frequency in Vietnam than the opposite pattern. This was unexpected regarding our initial assumptions: a favored allele selected in one or another environment has *a priori* no reason to be more often

associated with the presence or the absence of a TE insertion at linked sites. However, we found that the majority of the sequenced TE insertions segregates at low frequencies (around 10% of all individuals). When considering the linked region of one polymorphic TE insertion, if a favorable mutation appears in an individual where the insertion is absent, the increase of frequency of this “absence” haplotype will thus, most of the time, have a modest effect on the genetic differentiation at this marker, since it is already segregating at high frequency. By contrast, if a favorable mutation appears in a TE “presence” haplotype, the increase in frequency of the linked TE insertion would lead to high F_{ST} (F_{CT}) values. In absence of an alternative explanation, our outlier loci could thus indicate in which subset of populations the adaptive mutation occurred, and in the present case, this would have happened more frequently in the temperate populations.

Two scenarios, not mutually exclusive, could be invoked in the light of our data. A simple case would be a direct adaptive evolution in European invasive population that originated from tropical regions of the native area. A second hypothesis, could be that invasive temperate populations came from Northernmost territories of the native area such as northern China or Japan where *Ae. albopictus* populations are already cold-adapted. It would be thus interesting to know whether the observed signature of selection results from more “ancient” adaptations in the native area, or if it originates from more recent fine tuning of cold-related traits in the invasive areas. A recent study (33) suggested, using new variable *COI* mtDNA sequences and historical species range modeling, that Northern areas of the native range of *Ae. albopictus* would be the latest to have been colonized after a range expansion from Southern refugia following the last glacial around 21,000 years ago (34). The authors suggested that *Ae. albopictus* may have followed the human populations during their expansion from South to North in this area, that began approximatively 15,000 years ago. Thus wherever the origin of the invasive individuals sampled in Europe, it is likely that they are representatives of populations that had recently undergone a shift of selective pressure from tropical to

temperate climatic conditions. This could explain why so many outliers are associated with high insertion frequency in Europe, and that candidate genes in the diapause pathway are found in the neighborhood of some of these outliers. An easy way to distinguish between these possibilities would be to search if the same outlier insertions are present in several temperate populations from the native area.

It is important to note that the results presented here only are restricted to a subset of the Asian tiger mosquito populations located in temperate and tropical environments. It is thus probable that some of the outliers detected could be specific to this particular comparison and do not reflect the global pattern of differentiation between tropical and temperate populations. Research of the same outliers between other tropical and temperate populations from the native and non-native areas would be extremely valuable to extrapolate our results at a larger scale. Should the same outlier insertions be found at high frequencies in temperate locations – such as in USA, Japan or China –, extended investigations about the origin of invasive populations would help clarify if those similarities are due to an ancestral sweep or parallel sweeps that occurred independently in several populations. This study already provides for some candidate loci a set of functional primers that could be directly used to answer this question in any DNA sample of *Ae. albopictus*.

We report here the first leads supporting adaptive evolution at the molecular level in the Asian tiger mosquito. Progress in the annotation of published genomes, and the looming availability of supplementary genomic resources will allow to gain the most from these results. We hope that this work will contribute to unravel the implication of adaptive processes during the invasion of disease vectors.

Material and Methods

Biological samples. A total number of 140 flying adult females *Ae. albopictus* were collected in the field at eight sampling sites in Europe and Vietnam during the summers of 2012 and 2013 (Table S2). Individuals were either sampled using a single trap or using aspirators through the sampling site within a 50 meters radius. When traps were used, live mosquitoes were collected after a maximum of two days.

High throughput Transposon Display (TD) genotyping. Insertion polymorphism of five transposable elements families: I Loner Ele1 (IL1), Loa Ele2B (L2B), RTE4, RTE5 and Lian 1 identified by Goubert *et al.* (15) in *Ae. albopictus* were characterized. These TE families were chosen according to their high estimate of copy number (from 513 to 4203 cp), high identity between copies, and a “copy and paste” mode of transposition (all these TEs are non-LTR Class I retrotransposons). The protocol was developed combining methods from previous studies (26, 43–45) with high throughput Illumina sequencing of TD products (Figure S4). For each TE family and each individual, three independent PCR run were performed after total DNA digestion, in order to reduce the risk of uneven amplification of TE insertions. TD products were then pooled by individuals and paired-end sequenced on an Illumina Hiseq 2000 (1 lane) at the GeT-PlaGe core facility (Genome and Transcriptome, Toulouse) using TruSeq PE Cluster Kit v3 (2x100 bp) and TruSeq SBS Kit v3.

Bioinformatic treatment of TD sequencing. The different steps of the informatics treatment from the raw sequencing dataset to population binary matrices for presence/absence of TE insertions per individual are described in Figure S5. A total number of 102,319,300 paired-end 101bp Illumina reads were produced by sequencing PCR products. First, the paired-end reads of each individual were quality checked and

trimmed using UrQt v. 1.0.17 (46) with standard parameters and a *t* quality threshold of 10. Reads pairs were then checked and trimmed for Illumina adapter contamination using cutadapt (47). Specific amplification of TE insertions was controlled by checking for the expected 3' TE sequence on the R1 read using Blat (48) with an identity threshold of 0.90. Only reads with an alignment-length/read-length ratio ≥ 0.90 were then retained. R2 reads for which the R1 mate passed this filter were then selected for the insertion loci construction, after the removal of the TD adapter on the 5' start using cutadapt and the removal of reads under 30 bp. Selected reads were separated in each individual according to the TE families for loci construction.

In order to correct the inter-individual coverage variations, we performed a sampling of the cleaned reads (See SI Material and Methods). Insertion polymorphism of each TE family was assessed in two steps: first, insertion loci were recovered for each TE family at the individual level byclustering R2 reads using CD-HIT-EST (49). Then, representative sequences of each insertion loci were clustered between individuals to score the insertion polymorphism. Finally, for each TE family, a scored locus had to be shared by at least 2.5% of individuals and its coverage as well as the variance of this coverage had to be below the 99th centile of their distribution among loci (excluding the "0"s). After those filters, remaining loci were turned to binary scoring (1 = presence, 0 = absence) for genetic analyses.

Genetic analyses and Genomic scan. Population structure analyses were performed independently for each TE family. Principal Coordinate Analysis (PCoAs) were performed to identify genetic clusters using the ade4 package (50) of R vers. 3.2.1 (R development core team 2015). S7 coefficient of Gower and Legendre was used as a genetic distance since it gives more weight to shared insertions. Shared absences were not used because they do not give information about the genetic distance between

individuals. Pairwise populations F_{ST} were computed using Arlequin 3.5 (27); significance of the index was assessed over 1000 permutations using a significance threshold of 0.05.

The genomic scan was performed in two steps for each of the sampling replicates of each TE. First, Bayescan 2.1 (51) was used to test for each locus deviation from neutrality. Bayescan consider a fission/island model where all subpopulations derive from a unique ancestral population. In this model, variance in allele frequencies between subpopulations is expected to be due either to the genetic drift that occurred independently in each subpopulation or to selection that is a locus-specific parameter. The differentiation at each locus in each subpopulation from the ancestral population is thus decomposed into a β component (shared by all loci in a subpopulation) and is related to genetic drift, and a α component (shared for a locus by all subpopulations) due to selection. Using a Bayesian framework, Bayescan tests for each locus the significance of the α component. Rejection of the neutral model at one locus is done using posterior Bayesian probabilities and controlled for multiple testing using false discovery rate. In addition, Bayescan manage uncertainty about allele frequency from dominant data such as the TD polymorphism, leaving the F_{IS} to freely vary during the estimation of parameters. Bayescan was used with default values except for the prior odds that were set to 100 (more compatible with datasets with a large number of loci, see Bayescan manual), and a significance q-value threshold of 0.05 was used to retain outliers loci. In a second step, only outliers loci suggesting divergent directional selection between, Europe and Vietnam were considered. To identify them, locus by locus Analyses of Molecular Variance (AMOVAs) were performed using Arlequin 3.5 for each TE family. Significance of the F_{CT} (inter group differentiation) between Vietnamese and European populations was assessed performing 10,000 permutations with a significance threshold of 0.05. For each dataset, Bayescan outliers were crossed with significant F_{CT} loci to retain candidate loci.

To identify the genomic environment of the candidate loci, the outlier sequences (reference R2 read) were mapped onto the assembled genome of *Ae. albopictus* (28) using Blastn. Blastn alignments were performed with default parameters and sorted according to alignment score and after visual inspection of each alignment. Outlier loci with multiple identical hits were discarded. To identify genes surrounding the mapped outliers, the complete transcriptome of *Ae. albopictus* (including eggs, larvae and adult females[1]) was mapped over the reference genome using blat with default parameters; after alignment, one best hit was retained per transcript according to the best alignment score. When a transcript had multiple best hits, all positions for the transcript were considered.

Acknowledgements

We thank Van Tran-Van, Christophe Bellet, Grégory Lambert, Huynh Kim Ly Khanh and Trang Huynh who made possible and contributed to the samplings in France and Vietnam. We grateful to Valèria Romero Soriano and her family for their help during sampling in Sant Cugat dèl Vallès. We thank Manon Vigneron for PCR validation experiments. We also thank Rita Rebollo who provided insightful comments and helps for the English review of the manuscript. This work was performed using the computing facilities of the CC LBBE/PRABI. C.G. received a grant from the French Ministry of Superior Education. This work was supported by the Centre National de la Recherche Scientifique, the Institut Universitaire de France, and preliminary experiments benefited from of grant of the Federation de Recherche 41 “Bio-Environnement et Santé”. ”. Funding for mosquito sampling in Vietnam was provided by grants from EC2CO CNRS and occurred within the framework of GDRI “Biodiversity and Infectious Diseases in Southeast Asia”.

1 downloaded at www.albopictusexpression.org (Armbruster *et al.*)

References

1. Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Mol Ecol* 24(9):1999–2017.
2. Lande R (2015) Evolution of phenotypic plasticity in colonizing species. *Mol Ecol*. doi:10.1111/mec.13037.
3. Peischl S, Excoffier L (2015) Expansion load: recessive mutations and the role of standing genetic variation. *Mol Ecol*. doi:10.1111/mec.13154.
4. Handley L-J, et al. (2011) Ecological genetics of invasive alien species. *BioControl* 56(4):409–428.
5. Bock DG, et al. (2015) What we still don't know about invasion genetics. *Mol Ecol*:n/a–n/a.
6. Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D (2009) Aedes albopictus, an arbovirus vector: from the darkness to the light. *Microbes Infect* 11(14-15):1177–85.
7. Grard G, et al. (2014) Zika Virus in Gabon (Central Africa) – 2007: A New Threat from Aedes albopictus? *PLoS Negl Trop Dis* 8(2):e2681.
8. Marcondes CB, Ximenes M de FF de M (2015) Zika virus in Brazil and the danger of infestation by Aedes (Stegomyia) mosquitoes. *Rev Soc Bras Med Trop* (AHEAD). doi:10.1590/0037-8682-0220-2015.
9. Chouin-Carneiro T, et al. (2016) Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. *PLoS Negl Trop Dis* 10(3):e0004543.
10. Bonizzoni M, Gasperi G, Chen X, James AA (2013) The invasive mosquito species Aedes albopictus: current knowledge and future perspectives. *Trends Parasitol* 29(9):460–468.
11. Hawley WA (1988) The biology of Aedes albopictus. *J Am Mosq Control Assoc Suppl* 1:1–39.
12. Hawley W, Reiter P, Copeland R, Pumpuni C, Craig G (1987) Aedes albopictus in North America: probable introduction in used tires from northern Asia. *Science (80-)* 236(4805):1114–1116.
13. Hanson SM, Craig GB (1994) Cold Acclimation, Diapause, and Geographic Origin Affect Cold Hardiness in Eggs of Aedes albopictus (Diptera: Culicidae). *J Med Entomol* 31(2):192–201.
14. Poelchau MF, Reynolds J a, Elsik CG, Denlinger DL, Armbruster P a (2013) Deep sequencing

reveals complex mechanisms of diapause preparation in the invasive mosquito, Aedes albopictus. *Proc Biol Sci* 280(1759):20130143.

15. Goubert C, et al. (2015) 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*:evv050–.

16. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* 17(2):240–8.

17. Davey JW, et al. (2012) Special features of RAD Sequencing data: implications for genotyping. *Mol Ecol*. doi:10.1111/mec.12084.

18. Biedler J, Tu Z (2003) Non-LTR retrotransposons in the African malaria mosquito, Anopheles gambiae: unprecedented diversity and evidence of recent activity. *Mol Biol Evol* 20(11):1811–1825.

19. Boulesteix M, et al. (2007) Insertion polymorphism of transposable elements and population structure of Anopheles gambiae M and S molecular forms in Cameroon. *Mol Ecol* 16(2):441–452.

20. Santolamazza F, et al. (2008) Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms. *Malar J* 7(1):163.

21. Esnault C, et al. (2008) High genetic differentiation between the M and S molecular forms of Anopheles gambiae in Africa. *PLoS One* 3(4):e1968.

22. Bonin A, et al. (2008) A MITE-based genotyping method to reveal hundreds of DNA polymorphisms in an animal genome after a few generations of artificial selection. *BMC Genomics* 9:459.

23. Monden Y, Yamaguchi K, Tahara M (2014) Application of iPBS in high-throughput sequencing for the development of retrotransposon-based molecular markers. *Curr Plant Biol*. doi:10.1016/j.cpb.2014.09.001.

24. Casacuberta E, González J (2013) The impact of transposable elements in environmental adaptation. *Mol Ecol*:1503–1517.

25. Stapley J, Santure AW, Dennis SR (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol Ecol* 24(9):2241–52.

26. Roy AM, et al. (1999) Recently integrated human Alu repeats: finding needles in the haystack.

Genetica 107(1-3):149–161.

27. Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–7.

28. Chen X-G, et al. (2015) Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. *Proc Natl Acad Sci U S A* 112(44):E5907–5915.

29. Rao PN, Rai KS (1987) Inter and intraspecific variation in nuclear DNA content in Aedes mosquitoes. *Heredity (Edinb)* 59(2):253–258.

30. Kumar A, Rai KS (1990) Intraspecific variation in nuclear DNA content among world populations of a mosquito, Aedes albopictus (Skuse). *Theor Appl Genet* 79(6):748–52.

31. Black WICI V, Hawley WA, Rai KS, Craig GB (1988) Breeding structure of a colonizing species: Aedes albopictus (Skuse) in peninsular Malaysia and Borneo. *Heredity (Edinb)* 61(March):439–446.

32. Kambhampati S, Black WC, Rai KS (1991) Geographic origin of the US and Brazilian Aedes albopictus inferred from allozyme analysis. *Heredity (Edinb)* 67 (Pt 1)(September 1990):85–93.

33. Zhong D, et al. (2013) Genetic analysis of invasive Aedes albopictus populations in Los Angeles County, California and its potential public health impact. *PLoS One* 8(7):e68586.

34. Gupta S, Preet S (2014) Genetic differentiation of invasive Aedes albopictus by RAPD-PCR: Implications for effective vector control. *Parasitol Res* 113(6):2137–2142.

35. Manni M, et al. (2015) Molecular markers for analyses of intraspecific genetic diversity in the Asian Tiger mosquito, Aedes albopictus. *Parasit Vectors* 8(1):188.

36. Minard G, et al. (2015) French invasive Asian tiger mosquito populations harbor reduced bacterial microbiota and genetic diversity compared to Vietnamese autochthonous relatives. *Front Microbiol* 6. doi:10.3389/fmicb.2015.00970.

37. Urbanelli S, Bellini R, Carrieri M, Sallicandro P, Celli G (2000) Population structure of Aedes albopictus (Skuse): the mosquito which is colonizing Mediterranean countries. *Heredity (Edinb)* 84 (Pt 3)(November 1999):331–337.

38. Birungi J, Munstermann LE (2002) Genetic Structure of Aedes albopictus (Diptera: Culicidae) Populations Based on Mitochondrial ND5 Sequences: Evidence for an Independent Invasion into

Brazil and United States. *Ann Entomol Soc Am* 95(1):125–132.

39. Takumi K, et al. (2009) Introduction, scenarios for establishment and seasonal activity of Aedes albopictus in The Netherlands. *Vector Borne Zoonotic Dis* 9(2):191–6.

40. Becker N, et al. (2013) Repeated introduction of Aedes albopictus into Germany, July to October 2012. *Parasitol Res* 112(4):1787–90.

41. Dritsou V, et al. (2015) A draft genome sequence of an invasive mosquito: an Italian Aedes albopictus. *Pathog Glob Health*:2047773215Y0000000031.

42. Urbanski J, et al. (2012) Rapid adaptive evolution of photoperiodic response during invasion and range expansion across a climatic gradient. *Am Nat* 179(4):490–500.

43. Munroe DJ, et al. (1994) IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. *Genomics* 19(3):506–14.

44. Akkouche A, et al. (2012) tirant, a newly discovered active endogenous retrovirus in Drosophila simulans. *J Virol* 86(7):3675–81.

45. Carnelossi EAG, et al. (2014) Specific activation of an I-like element in Drosophila interspecific hybrids. *Genome Biol Evol* 6(7):1806–17.

46. Modolo L, Lerat E (2015) UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. *BMC Bioinformatics* 16(1):137.

47. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10.

48. Kent WJ (2002) BLAT---The BLAST-Like Alignment Tool. *Genome Res* 12(4):656–64.

49. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–9.

50. Dray S, Dufour A-B (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *J Stat Softw* 22(4):1–20.

51. Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180(2):977–93.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J*

Mol Biol 215(3):403–10.

Data Accessibility

Paired-end raw sequences are available through SRA at NCBI under SRP070185 (Bioproject PRJNA312147)

Final presence/absence matrices (including replicates) will be available at Dryad

Author contributions

CG, CV and MB conceived the experiments and conducted the analyses. CG and HH developed and performed the molecular experiments. GM, CVM and PM conducted the sampling in France and Vietnam. All authors contributed to the final version of the manuscript.