nature ecology & evolution

Supplementary information

https://doi.org/10.1038/s41559-025-02643-5

Adaptive genomic signatures of globally invasive populations of the yellow fever mosquito *Aedes aegypti*

In the format provided by the authors and unedited

CONTENT

Supplementary Methods

Confirmation of species identity for sequenced samples

We reconstructed the sequences of the ribosomal Internal Transcribed Spacer 1 and 2 (ITS1 and ITS2) for 27 samples with <50% of the reads aligned to the reference assembly AaegL5 [1], by using *LT_finder* from ViR pipeline [2] and the 5.8S rRNA sequence of *Aedes aegypti* as reference (accession M95126, region coordinates 574-755 in AagL5). The reconstructed ITS1-ITS2 sequences were searched with *blastn* and default parameters against the NCBI 'nt database' (https://blast.ncbi.nlm.nih.gov/) to confirm species identity. A total of 27 whole genome sequences (WGS) datasets were excluded, including all 18 samples from Ethiopia that showed the highest identity (>97%, E-value ≈ 0) and coverage (100%) to ITS sequences of *Ae. simpsoni*, and 9 WGS datasets whose ITS1-ITS2 sequences did not yield *blastn* results.

Custom "golden dataset" for SNPs identification

To improve alignments and variant discovery, we recalibrated the variant caller algorithm *GATK* v3.8.1.0 [3] with a set of known indels and Single Nucleotide Polymorphisms (SNPs) obtained from: (1) known SNPs from literature; and (2) *de novo* SNPs estimates from our sequenced mosquitoes. Both procedures are described next. See also Supplementary Data 11.

SNPs collected from the literature

A collection of 485,431 SNPs was obtained in variant calling format (vcf) files from nine published studies [4–12]. A total of 112,969 SNPs (23.27%) has experimental support, such as high-throughput genotyping chip 'Axiom_aegypti1' [5–9], while 372,462 SNPs (76.73%) were estimated through bioinformatic analyses only [4,10–12]. The updated genomic coordinates of all nine SNPs datasets for the assembly AaegL5 were obtained from VectorBase (https://vectorbase.org/). An SNPs dataset originated from exome analysis [13] was also added to our study, and required the re-mapping of SNPs coordinates from AaegL1 to AaegL5 with the *Flo* pipeline [14]. Also, *Crossmap* [15] was used to perform the coordinates conversion for the set of known SNPs over AaegL5. After merging all sites and removing overlapped redundant variants, a total of 304,428 SNPs from 10 datasets were mapped across 10,185 genes and 278 contigs annotated in AaegL5 (Supplementary Data 11).

De novo SNPs discovery

Two variant callers, *GATK* v3.8.1.0 [3] and *Freebayes* v1.3.1 [16], were used to improve SNPs prediction for each of the 634 WGS samples mapped to AagL5, given that both tools have high false discovery rate (FDR) [17–19] and they score the quality of parameter measurements differentially [17,20]. We further implemented a strong filtering of raw variants with both tools to

increase specificity and sensitivity. Filtering of SNPs in *Freebayes* was performed with *BCFtools* [21] and parameters: DP ≥ 5, MQ ≥ 20, QUAL ≥ 20, and MAF ≥ 0.01. Filtering of SNPs in *GATK* was done with parameters: QUAL < 30.0, QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0; and for INDELs: QUAL < 30.0, QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0. Any SNP at proximity of 10 base pairs (bp) to an indel and with a strong strand bias (p<0.001) were removed with *BCFtools*. We merged all sites and removed overlapped redundant variants from both 'known' and *'de novo'* approaches, generating a custom "golden dataset" with a high-quality collection of 82,686,298 indels and 207,724,254 SNPs that we used in the recalibration step, as described below.

A final refined variant caller prediction was performed only with the *GATK* protocol and our custom "golden dataset" for all recalibrated alignments and each of the 39 populations. Raw SNPs and indels were extracted and filtered with the same filtering parameters using *GATK*, as described previously. A total of 314,365,358 high-confidence SNPs was obtained as the core dataset of our analyses; indels were not further considered in our study. See also Supplementary Table 33 in this document.

Pairwise individual relatedness from *Ae. aegypti* populations

To avoid biases due to close relatedness among the 634 individuals analyzed, we first removed all SNPs detected over the annotated repetitive regions in AaegL5 [8] and then extracted all biallelic SNPs present in ≥90% of individuals within each population. Next, highly linked loci were eliminated using *snpgdsLDpruning* (ld.threshold=0.01) from *SNPRelated* [22], and the corresponding matrix of 'relatedness coefficients' (k) was generated with the Identity-by-Descent measurement based on maximum likelihood (ML) estimation using *snpgdsIBDMLE* from *SNPRelated*. Highly genetically related individuals were classified and filtered on each

population using two previously reported cutoffs for African and out-of-Africa populations [23]: *first cousin and closer relationships* ($k \leq 0.05$) for African populations and *siblings* ($k \leq 0.20$) for out-of-Africa populations. We repeated this analysis with a dataset of 1,000,000 randomly sampled SNPs across the whole genome *per* population, and a comparison of both approaches led us to remove 95 individuals (Supplementary Table 12). Because 15 of the removed individuals corresponded to the previously called 'domesticated' mosquitoes of the Rabai population [23], an analysis on the covariance of genotypes was performed with *pca* from *plink* v2.0 [24] to re-evaluate their relatedness. The results led us to reintroduce all 15 domestic individuals from Rabai (RABd) to our dataset (Extended Data Fig. 4d), resulting in a final set of 554 *Ae. aegypti* genomes and 40 African populations for further analyses (see also Supplementary Table 33 in this document).

Filtering of SNPs datasets for different analyses

Due to the large and highly repetitive nature of the *Ae. aegypti* genome (*i.e.*, repetitive sequences occupy >50% of 1.25 Gigabases [Gb] of the complete genome) [1], we created several filtered SNP datasets from the 314,365,358 high-confidence SNPs. First, we removed a total of 223,328,290 SNPs located across all repetitive sequences annotated in AaegL5 [1] (Supplementary Table 29), which generated a dataset of 91,037,068 biallelic and multiallelic SNPs located in non-repetitive regions only (NR-SNPs). Then, all "chromosomal biallelic SNPs" were extracted from the ~91 million 'NR-SNPs dataset', generating a second dataset of 89,613,991 biallelic NR-SNPs. This dataset was further filtered using plink v2.0: (a) to avoid sampling genotyping errors [25]; (b) to remove SNPs under high linkage equilibrium (LD), by applying the parameters for indep-pairwise: window size=50, step size=10, and R^2=0.1; (c) to remove SNPs with a minor allele frequency (MAF) <0.01; and (d) to retain only SNPs that are found in >80% of all individuals *per* population (option: -geno 0.2). This procedure led to a

dataset of 1,530,512 biallelic NR-SNPs. A "core-exome-SNPs" dataset of 2,929 NR-SNPs was further generated by mapping our 1.5 million biallelic NR-SNPs over protein-coding exons and considering only SNPs that are present in all individuals across all populations.

These filtering processes generated four main SNP datasets that we used to perform the different analyses of this study (see details in Methods): (1) the full dataset of ~314.4 million SNPs was used for global genome-wide statistics; (2) the dataset of ~89.6 million biallelic NR-SNPs including all individuals and populations was also used for genome-wide statistics, in addition to genetic diversity estimates, genetic relatedness, outlier detection and gene selection analyses; (3) the dataset of ~1.5 million biallelic NR-SNPs with different coverages (>80% and >90%) of individuals *per* population was used for population structure and genetic distance analyses; and (4) the "core-exome-SNPs" dataset of ~3,000 biallelic NR-SNPs was used for phylogenetic and population divergence analyses.

Analysis of *Ae. aegypti* nrEVEs

The frequency of each of the 252 nrEVEs annotated in AaegL5 was established in each of the analyzed populations using the SVD pipeline [26–28]. The same procedure was used to verify the occurrence of *Ae. aegypti* nrEVEs in 4 WGS datasets from *Ae. mascarensis* [29]. The Vy-PER [30] and ViR [27] pipelines were used to search for novel viral integrations using a viral database assembled in October 2020, which encompasses a total of 1778 taxids and 3677 nucleotide sequences of both DNA and RNA viruses. To build the viral database, viral taxids were extracted from the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) according to three criteria: (*i*) main known arboviral genera; (*ii*) Insect-Specific Viruses (ISVs) identified from a search of NCBI PubMed publications between 2015 and 2020 using the keyword 'insect-specific viruses'; (*iii*) viruses having Diptera as host. Nucleotide sequences for the selected taxids were retrieved with NCBI E-utility tool [31]. After removing duplicates,

sequences were clustered at 97% identity using CD-HIT [32]. Sequences were BLAST searched against a database of conserved eukaryotic and bacterial ribosomal sequences developed for SortmeRNA [33] and of Diptera ribosomal sequences extracted from Quast *et al.* (2013) [34]. Entire or parts of sequences matching ribosomal DNA were removed or masked, respectively. Homopolymers and repeats were identified using a custom script and masked with 'Ns'. The presence of novel nrEVEs was further tested in all WGS data used for SNP discovery with ViR_LTFinder [27] to confirm their widespread distribution in African vs out-of-Africa samples (Supplementary Table 30).

A total of 64 new integrations were identified (Supplementary Table 7); all *new* nrEVEs were similar to ISVs, apart from three integrations from the *Liao Ning* virus of the *Seadornavirus* genus (Reoviridae family), which includes emerging pathogens [35] (Supplementary Table 31, Supplementary Data 2). A subset of 26 out of the 64 novel nrEVEs was molecularly validated by PCR and Sanger sequencing in the mosquito DNA samples that had been used for WGS, using primers designed from predictions by ViR (Supplementary Table 32). PCRs were performed in a volume of 50 mL containing 25 mL of 2X DreamTaq Green PCR Master Mix (ThermoFisher), 2.5 mL of 10 mM forward and reverse primers, 1 mL of DNA diluted 1:10, and sterile water to volume. Reactions without a template served as negative controls. PCR products were visualized through electrophoresis on 2% agarose gels stained with ethidium bromide. Since multiple bands were often observed, bands of the expected size were cut from the gels and purified using the GeneJET PCR Purification Kit (Thermo Scientific), following the manufacturer's protocol. Purified PCR products were sent for Sanger sequencing to Macrogen Europe (Netherlands) to confirm nrEVE sequence (Extended Data Fig. 5c). We also observed frequent rearrangement events following integration, especially for nrEVEs with similarity to flaviviruses that map in piRNA clusters

(Supplementary Tables 7 and 30), supporting the conclusion that nrEVEs contribute to the genetic flexibility of *Ae. aegypti* piRNA clusters.

Supplementary References

1. Matthews, B. J. *et al.* Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* **563**, 501–507 (2018).
2. Pischedda, E. *et al.* ViR: a tool to solve intrasample variability in the prediction of viral integration sites using whole genome sequencing data. *BMC Bioinformatics* **22**, 1–15 (2021).
3. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
4. Faucon, F. *et al.* Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. *Genome Res* **25**, 1347–1359 (2015).
5. Evans, B. R. *et al.* A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3: Genes, Genomes, Genetics* **5**, 711–718 (2015).
6. Saarman, N. P. *et al.* Effective population sizes of a major vector of human diseases, *Aedes aegypti*. *Evol Appl* **10**, 1031–1039 (2017).
7. Gloria-Soria, A. *et al.* Origin of a high-latitude population of *Aedes aegypti* in Washington, DC. *American Journal of Tropical Medicine and Hygiene* **98**, 445–452 (2018).
8. Kotsakiozi, P., Gloria-Soria, A., Schaffner, F., Robert, V. & Powell, J. R. *Aedes aegypti* in the Black Sea: Recent introduction or ancient remnant? *Parasit Vectors* **11**, 1–13 (2018).
9. Kotsakiozi, P. *et al.* Population structure of a vector of human diseases: *Aedes aegypti* in its ancestral range, Africa. *Ecol Evol* **8**, 7835–7848 (2018).
10. Bonizzoni, M. *et al.* Probing functional polymorphisms in the dengue vector, *Aedes aegypti*. *BMC Genomics* **14**, 1–10 (2013).
11. David, J. P. *et al.* Comparative analysis of response to selection with three insecticides in the dengue mosquito *Aedes aegypti* using mRNA sequencing. *BMC Genomics* **15**, 1–15 (2014).
12. Pless, E. *et al.* Multiple introductions of the dengue vector, *Aedes aegypti*, into California. *PLoS Negl Trop Dis* **11**, e0005718 (2017).
13. Crawford, J. E. *et al.* Population genomics reveals that an anthropophilic population of *Aedes aegypti* mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biol* **15**, 1–16 (2017).
14. Pracana, R., Priyam, A., Levantis, I., Nichols, R. A. & Wurm, Y. The fire ant social chromosome supergene variant Sb shows low diversity but high divergence from SB. *Mol Ecol* **26**, 2864–2879 (2017).
15. Zhao, H. *et al.* CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
16. Garrison, E. FreeBayes source repository: https://github.com/freebayes/freebayes. Preprint at https://arxiv.org/abs/1207.3907 (2012). p. 1–9.
17. Bian, X. *et al.* Comparing the performance of selected variant callers using synthetic data and genome segmentation. *BMC Bioinformatics* **19**, 1–11 (2018).
18. Kumaran, M., Subramanian, U. & Devarajan, B. Performance assessment of variant calling pipelines using human whole exome sequencing and simulated data. *BMC Bioinformatics* **20**, 1–11 (2019).
19. Schilbert, H. M., Rempel, A. & Pucker, B. Comparison of read mapping and variant calling tools for the analysis of plant NGS data. *Plants* **9**, 1–14 (2020).
20. Liu, X., Han, S., Wang, Z., Gelernter, J. & Yang, B. Z. Variant Callers for Next-Generation Sequencing Data: A Comparison Study. *PLoS One* **8**, e75619 (2013).
21. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008, 1–4 (2021).
22. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
23. Rose, N. H. *et al.* Climate and Urbanization Drive Mosquito Preference for Humans. *Current Biology* **30**, 3570–3579 (2020).

24. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, 1–16 (2015).
25. Graffelman, J. & Weir, B. S. Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity (Edinb)* **116**, 558–568 (2016).
26. Dwivedi, S. B. *et al.* Brain proteomics of *Anopheles gambiae*. *OMICS* **18**, 421–437 (2014).
27. Suzuki, Y. *et al.* Non-retroviral Endogenous Viral Element Limits Cognate Virus Replication in *Aedes aegypti* Ovaries. *Current Biology* **30**, 3495–3506 (2020).
28. Palatini, U., Pischedda, E. & Bonizzoni, M. Computational methods for the discovery and annotation of viral integrations. in *piRNA: Methods and Protocols*, p. 293–313 (Springer, 2022).
29. Soghigian, J. *et al.* Genetic evidence for the origin of *Aedes aegypti*, the yellow fever mosquito, in the southwestern Indian Ocean. *Mol Ecol* **29**, 3593–3606 (2020).
30. Forster, M. *et al.* Vy-PER: Eliminating false positive detection of virus integration events in next generation sequencing data. *Sci Rep* **5**, 1–13 (2015).
31. Kans, J. Entrez direct: e-utilities on the UNIX command line. in *Entrez Programming Utilities Help [Internet]* (National Center for Biotechnology Information (US), 2022).
32. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
33. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
34. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2013).
35. Zhang, J. *et al.* Origin and evolution of emerging *Liao ning Virus* (genus Seadornavirus, family Reoviridae). *Virol J* **17**, 105 (2020).

Supplementary Table 33. Descriptive statistics for the 314,365,358 high-confidence SNPs identified in this study across the complete and different genomic regions of the reference assembly AaegL5 and for each analyzed *Ae. aegypti* population.

No	Continent	Location	Country (Acronym)	Population (Acronym)	No. samples *	SNPs Genome (%) †	NR-SNPs (%) ‡	R-SNPs (%) §	Exons	CDS ‖	5'UTR	3'UTR	Singletons	No. Indels
1	**Africa**	**Western**	**Senegal (SEN)**	Bantata (BAN)	11 (11)	50,915,637 (3.98)	16,102,135 (31.63%)	34,813,502 (68.37%)	1,705,289 (3.43%)	1,055,724	208,393	233,077	16,194,719	7,296,641
2				Kedougou (KEN)	20 (19)	62,102,819 (4.86)	19,366,750 (31.18%)	42,736,069 (68.82%)	2,125,967 (3.54%)	1,327,093	258,005	287,478	17,812,101	10,299,313
3				Mindin (MIN)	20 (19)	61,835,584 (4.84)	19,294,616 (31.20%)	42,540,968 (68.80%)	2,102,856 (3.51%)	1,309,145	255,403	285,408	17,319,911	9,834,817
4				Ngoye (NGY)	20 (19)	53,512,353 (4.18)	16,538,705 (30.91%)	36,973,648 (69.09%)	1,776,672 (3.41%)	1,105,744	204,600	227,930	12,022,034	8,482,085
5				PK10 (PK10)	20 (17)	59,791,110 (4.68)	18,664,300 (31.22%)	41,126,810 (68.78%)	2,028,999 (3.50%)	1,263,257	246,173	275,049	17,413,512	9,551,089
6				Thiès (THI)	20 (11)	46,203,180 (3.61)	14,427,309 (31.23%)	31,775,871 (68.77%)	1,521,750 (3.36%)	943,659	183,961	204,556	11,642,221	6,779,311
7			**Burkina Faso (BKF)**	Dori (DOR)	18 (14)	52,829,396 (4.13)	16,687,945 (31.59%)	36,141,451 (68.41%)	1,854,562 (3.61%)	1,166,623	221,039	245,593	15,438,922	9,483,310
8				Ouagadougou (OGD)	20 (19)	56,323,619 (4.4)	17,539,688 (31.14%)	38,783,931 (68.86%)	1,891,072 (3.45%)	1,175,209	229,571	256,051	12,872,274	8,948,614
9				Ouahigouya (OHI)	16 (16)	57,293,151 (4.48)	17,970,578 (31.37%)	39,322,573 (68.63%)	1,939,706 (3.48%)	1,206,524	235,619	263,339	15,994,878	9,027,189
10			**Ghana (GHA)**	Larabanga (LRB)	16 (13)	52,202,429 (4.08)	16,322,075 (31.27%)	35,880,354 (68.73%)	1,792,497 (3.53%)	1,122,581	214,721	238,249	15,904,999	9,693,001
11				BoabengFiema (BBF)	2 (2)	16,599,457 (1.3)	5,963,110 (35.92%)	10,636,347 (64.08%)	678,753 (4.11%)	424,898	82,156	90,675	10,985,786	1,934,484
12				Kintampo (KIN)	20 (19)	60,694,927 (4.75)	18,910,506 (31.16%)	41,784,421 (68.84%)	2,052,330 (3.49%)	1,276,275	248,954	278,758	15,965,774	9,709,928
13				Kumasi (KUM)	20 (20)	59,896,221 (4.68)	18,713,467 (31.24%)	41,182,754 (68.76%)	2,018,830 (3.47%)	1,251,906	245,412	275,672	14,516,591	9,127,011
14			**Nigeria (NIG)**	Awka (AWK)	19 (19)	61,031,081 (4.77)	19,145,085 (31.37%)	41,885,996 (68.63%)	2,064,168 (3.48%)	1,279,189	252,562	282,331	15,951,307	9,188,886
15		**Central**	**Cameroon (CAM)**	Bénoué (BEN)	16 (10)	37,474,493 (2.93)	11,862,149 (31.65%)	25,612,344 (68.35%)	1,294,187 (3.53%)	812,368	152,671	170,333	8,832,296	7,059,479
16			**Gabon (GAB)**	Franceville (FRV)	29 (28)	64,094,393 (5.01)	20,069,234 (31.31%)	44,025,159 (68.69%)	2,276,121 (3.69%)	1,434,178	271,525	303,886	13,413,936	12,503,406

17				Libreville (LBV)	15 (12)	48,299,015 (3.78)	15,195,868 (31.46%)	33,103,147 (68.54%)	1,622,009 (3.44%)	1,009,147	195,102	219,066	11,557,858	7,491,976
18				LopeVillage (LPV)	15 (14)	54,778,534 (4.28)	17,276,876 (31.54%)	37,501,658 (68.46%)	1,866,198 (3.50%)	1,162,513	224,592	253,016	14,789,053	8,713,121
19		**Eastern**	**Uganda (UGA)**	Bundibugyo (BDB)	14 (3)	24,569,478 (1.92)	8,349,082 (33.98%)	16,220,396 (66.02%)	912,503 (3.76%)	571,550	110,033	121,527	11,952,216	3,620,480
20				Entebbe (ENT)	12 (8)	44,961,217 (3.52)	14,332,483 (31.88%)	30,628,734 (68.12%)	1,533,438 (3.48%)	956,443	184,044	206,552	15,676,414	6,643,509
21				Karenga (KRG)	3 (3)	23,972,914 (1.87)	8,160,190 (34.04%)	15,812,724 (65.96%)	892,938 (3.77%)	559,272	107,429	119,092	10,894,177	3,524,838
22				Kichwamba (KIG)	11 (3)	25,528,733 (2)	8,929,347 (34.98%)	16,599,386 (65.02%)	984,900 (3.90%)	616,651	118,594	133,255	13,356,934	3,170,583
23			**Kenya (KEN)**	Virhembe (VIR)	20 (10)	51,812,827 (4.05)	16,678,566 (32.19%)	35,134,261 (67.81%)	1,789,674 (3.54%)	1,114,262	216,678	243,497	18,162,017	7,630,673
24				Kakamega (KKM)	19 (12)	56,512,255 (4.42)	17,907,324 (31.69%)	38,604,931 (68.31%)	1,926,149 (3.50%)	1,198,705	233,646	260,887	18,743,667	8,526,763
25				Mbarakani village (MBK)	16 (15)	62,753,159 (4.91)	19,847,642 (31.63%)	42,905,517 (68.37%)	2,239,863 (3.70%)	1,413,233	266,842	296,611	21,733,493	11,528,803
26				Ganda (GND)	9 (8)	49,216,282 (3.85)	16,232,877 (32.98%)	32,983,405 (67.02%)	1,751,755 (3.64%)	1,092,941	211,839	239,300	19,458,170	6,967,291
27				Arabuko (ARA)	19 (19)	69,214,993 (5.41)	22,125,039 (31.97%)	47,089,954 (68.03%)	2,495,983 (3.74%)	1,573,474	300,036	334,893	22,091,234	11,914,929
28				KayaBomu (KYB)	19 (14)	61,448,172 (4.81)	19,707,563 (32.07%)	41,740,609 (67.93%)	2,156,246 (3.62%)	1,348,081	260,552	292,172	20,522,089	9,457,843
29				Kwale (KWA)	19 (19)	68,038,422 (5.32)	21,621,320 (31.78%)	46,417,102 (68.22%)	2,429,027 (3.70%)	1,530,106	291,485	325,449	21,472,110	11,695,552
30				ShimbaHills (SHH)	8 (7)	45,111,490 (3.53)	14,888,939 (33.00%)	30,222,551 (67.00%)	1,599,713 (3.62%)	996,180	193,378	218,130	18,815,962	6,491,976
31				Rabai domesticated (RABd)	15	65369724 (5.11)	20784994 (31.80%)	44584730 (68.20%)	1048465 (1.60%)	1,509,393	280,877	313,451	26,775,337	12,343,895
32				Rabai selvatic (RABg)	19	31131199 (2.43)	9263625 (29.76%)	21867574 (70.24%)	2373467 (7.62%)	695,331	111,764	123,060	10,612,053	7,791,455
33	**Americas**	**North America**	**Mexico (MEX)**	Tapachula (TAP)	16 (16)	22,994,712 (1.8)	7,244,143 (31.50%)	15,750,569 (68.50%)	785,605 (3.46%)	488,564	93,551	103,786	2,545,413	4,441,046
34		**South America**	**Brazil (BRZ)**	Bebedouro (BEB)	16 (12)	27,735,391 (2.17)	8,602,653 (31.02%)	19,132,738 (68.98%)	919,159 (3.35%)	568,211	108,830	122,236	3,483,460	4,578,433
35				Santarem (SAN)	18 (18)	27,688,035 (2.17)	8,239,613 (29.76%)	19,448,422 (70.24%)	868,365 (3.18%)	536,554	102,599	115,287	1,830,877	4,550,368
36	**Asia**	**Western Asia**	**Saudi Arabia (SAA)**	Jeddah (JED)	9 (9)	24,610,505 (1.92)	7,791,931 (31.66%)	16,818,574 (68.34%)	830,633 (3.41%)	514,465	99,035	110,225	4,382,323	4,043,669

37		**Southeastern Asia**	**Thailand (THA)**	Samut Sakhon (SAS)	16 (16)	26,403,048 (2.06)	8,256,492 (31.27%)	18,146,556 (68.73%)	901,570 (3.46%)	560,837	107,307	118,800	3,524,705	4,889,932
38				Bangkok (BNK)	20 (19)	27,461,673 (2.15)	8,301,275 (30.23%)	19,160,398 (69.77%)	889,459 (3.28%)	550,636	106,309	117,606	2,517,081	4,918,596
39	**Oceania**	**Oceania**	**American Samoa (AMS)**	Tafuna Village (TFV)	16 (15)	23,464,111 (1.83)	7,372,772 (31.42%)	16,091,339 (68.58%)	791,587 (3.41%)	491,206	94,663	104,855	2,429,934	4,534,497
40	**Continent**	**Location**	**New Caledonia (NCA)**	Zac Panda (ZAP) ¶	12 (12)	26,365,057 (2.06)	8,345,473 (31.65%)	18,019,584 (68.35%)	891,787 (3.42%)	554,143	106,073	118,286	3,033,466	4,562,815

***** The number of samples before and after the relatedness analysis is shown outside and inside the parentheses, respectively. Note that some (but not all) samples from the Rabai population were discarded from our final working dataset (see section "Pairwise individual relatedness from *Ae. aegypti* populations" in this document).

† The total number of SNPs detected across the *Ae. aegypti* genome is shown, along with its percentage from the total number of nucleotides present in the complete AaegL5 genome (1.25 Gigabases, Gb) [1].

‡ The total number of SNPs detected in non-repetitive regions (NR-SNPs) across the genome is shown, along with its percentage (%) from the total SNPs count identified for the complete AaegL5 genome (1.25 Gigabases, Gb) [1].

§ The total number of SNPs detected in repetitive regions (R-SNPs) across the genome is shown, along with its percentage (%) from the total SNPs count identified for the complete AaegL5 genome (1.25 Gigabases, Gb) [1].

‖ The total number of SNPs detected in Protein CoDing Sequences (CDS) of the *Ae. aegypti* genome.

¶ Zac Panda: Zone d'Aménagement Concerté Panda (Development area 'Panda').