


RESEARCH ARTICLE

Open Access



Genome-wide divergence among invasive populations of *Aedes aegypti* in California

Yoosook Lee^{1*} , Hanno Schmidt¹, Travis C. Collier², William R. Conner³, Mark J. Hanemaaijer¹, Montgomery Slatkin⁴, John M. Marshall⁵, Joanna C. Chiu³, Chelsea T. Smartt⁶, Gregory C. Lanzaro¹, F. Steve Mulligan⁷ and Anthony J. Cornel^{3,8}

Abstract

Background: In the summer of 2013, *Aedes aegypti* Linnaeus was first detected in three cities in central California (Clovis, Madera and Menlo Park). It has now been detected in multiple locations in central and southern CA as far south as San Diego and Imperial Counties. A number of published reports suggest that CA populations have been established from multiple independent introductions.

Results: Here we report the first population genomics analyses of *Ae. aegypti* based on individual, field collected whole genome sequences. We analyzed 46 *Ae. aegypti* genomes to establish genetic relationships among populations from sites in California, Florida and South Africa. Based on 4.65 million high quality biallelic SNPs, we identified 3 major genetic clusters within California; one that includes all sample sites in the southern part of the state (South of Tehachapi mountain range) plus the town of Exeter in central California and two additional clusters in central California.

Conclusions: A lack of concordance between mitochondrial and nuclear genealogies suggests that the three founding populations were polymorphic for two main mitochondrial haplotypes prior to being introduced to California. One of these has been lost in the Clovis populations, possibly by a founder effect. Genome-wide comparisons indicate extensive differentiation between genetic clusters. Our observations support recent introductions of *Ae. aegypti* into California from multiple, genetically diverged source populations. Our data reveal signs of hybridization among diverged populations within CA. Genetic markers identified in this study will be of great value in pursuing classical population genetic studies which require larger sample sizes.

Keywords: *Aedes aegypti*, Invasive species, Population genomics, California

Background

Aedes aegypti has a short flight range, usually not actively moving more than 200 m from their breeding source [1], but is exquisitely adapted to hitchhiking in transport vehicles [2]. One central question concerning the populations dynamics of *Ae. aegypti* in California (CA) therefore is whether the established populations at different locations are founded from one source population that spread across the state or if they are the result of other kinds of founding effects. A recent study revealed several, genetically distinct *Ae. aegypti* populations in CA presumably originating from multiple

introductions from other sites in the U.S. and/or northern Mexico [3]. Insight into the population structure of CA *Ae. aegypti* beyond this will be necessary to fully understand the dynamics that shape the current pattern of distribution and continuing spread of this invasive vector species.

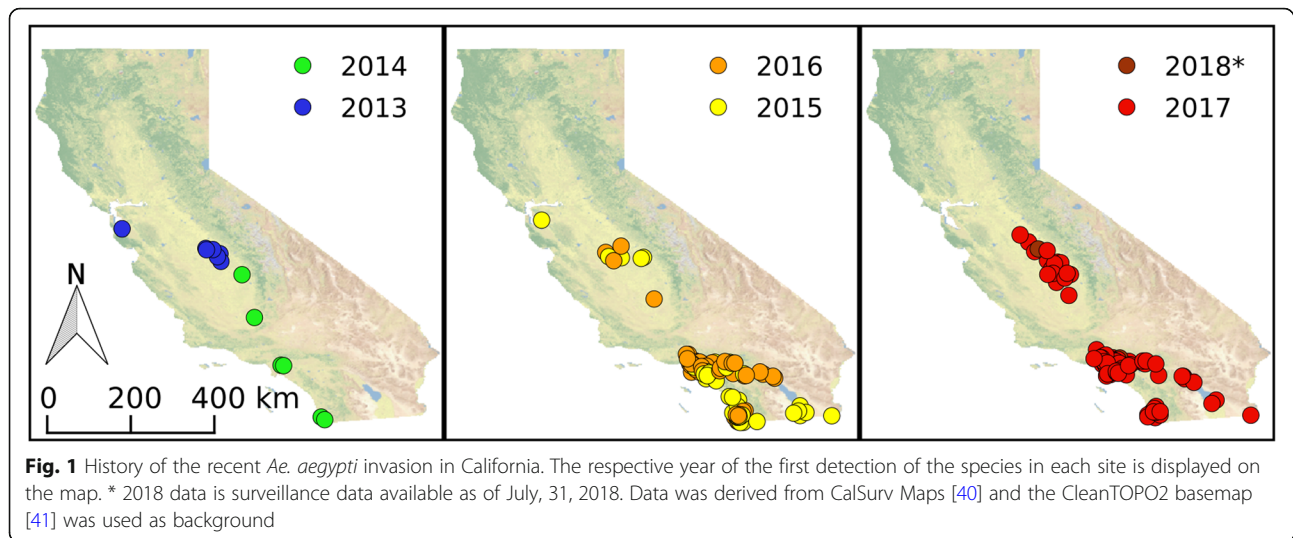
California had no known established local populations of *Ae. aegypti* prior to the summer of 2013 when it was detected in three cities in central California: Clovis, Madera and Menlo Park [4–6]. In the spring and summer of the following year, this mosquito was again found in the same three California locations and for the first time in additional communities in central California and further south in San Diego County (Fig. 1). *Ae. aegypti* specimens have been collected over multiple years from some sites and from additional locations in both central

* Correspondence: yoslee@ucdavis.edu

¹Vector Genetics Laboratory, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California Davis, Davis, CA 95616, USA

Full list of author information is available at the end of the article





and southern CA each year, indicating that *Ae. aegypti* has now become established and is spreading through large parts of the state (Fig. 1).

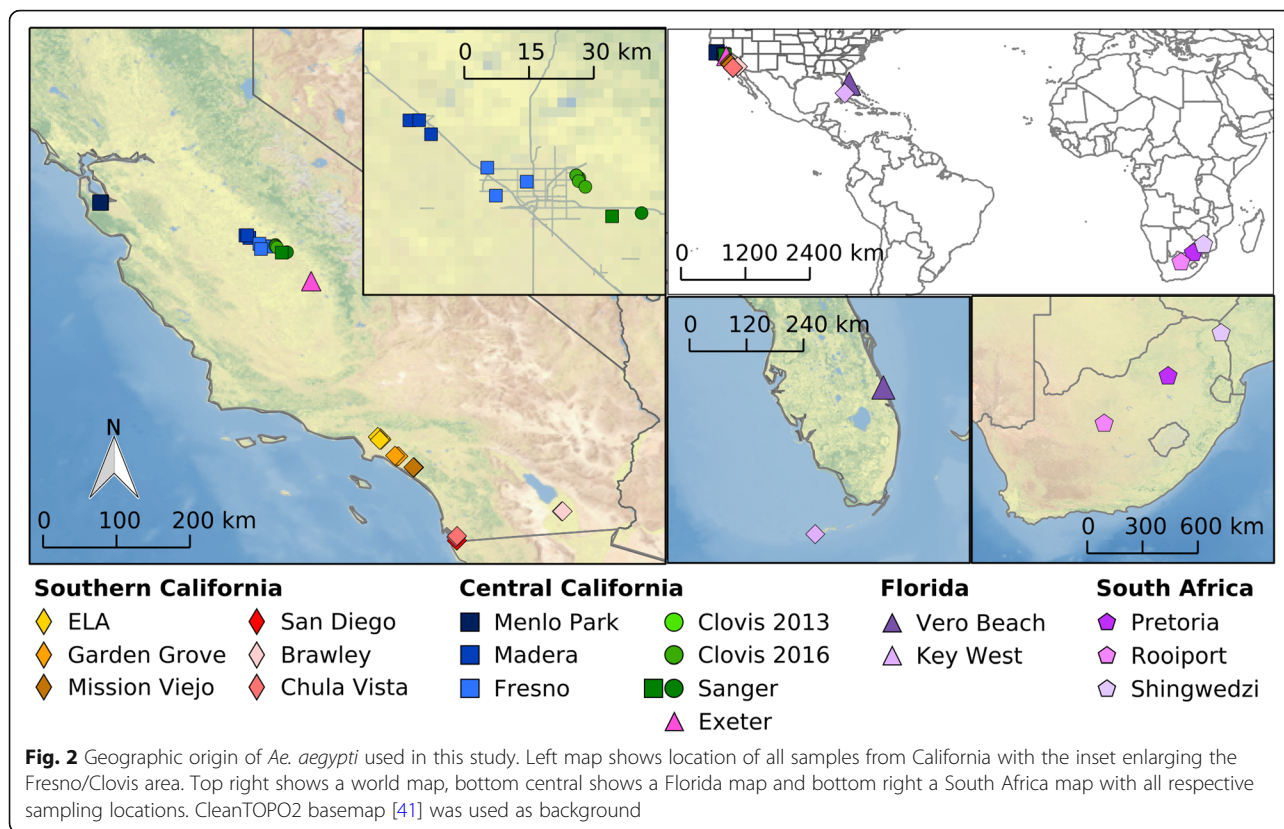
Successful vector control can benefit from population genetics and genomics analyses which can provide estimates of gene flow and identify the genetic basis of phenotypes such as insecticide resistance [7] and host preference [8]. Population genomics studies are especially critical to the development of control strategies based on genetic manipulation of vectors, which is a matter of growing interest. Modelling, planning and monitoring activities associated with control programs require affordable and rapid assays to distinguish vector sub-populations within a species and a deep understanding of the processes that shape their genetic structure. It is becoming increasingly apparent that hybridization between diverged vector populations may be an important source of new genetic material including alleles that mediate adaptations to facilitate range expansion [9] or that promote the evolution of resistance to insecticides [10]. Analysis of whole-genome sequencing data is the most powerful method to detect even minor admixture [11]. Therefore, we have applied a population genomics approach to study invasive *Ae. aegypti* populations. Here we report a preliminary analysis based on genome sequences of 39 individual *Ae. aegypti*, collected from twelve locations throughout CA, and, for comparison, four specimens from Florida and three from South Africa. The analyses presented here should serve as a step toward expanded population genomics studies aimed at understanding how invasive mosquito species become established in new locations and how distinct populations interact on the genetic level.

Results

We sequenced the genomes of 46 specimens of *Ae. aegypti* from California ($N = 39$), Florida ($N = 4$) and

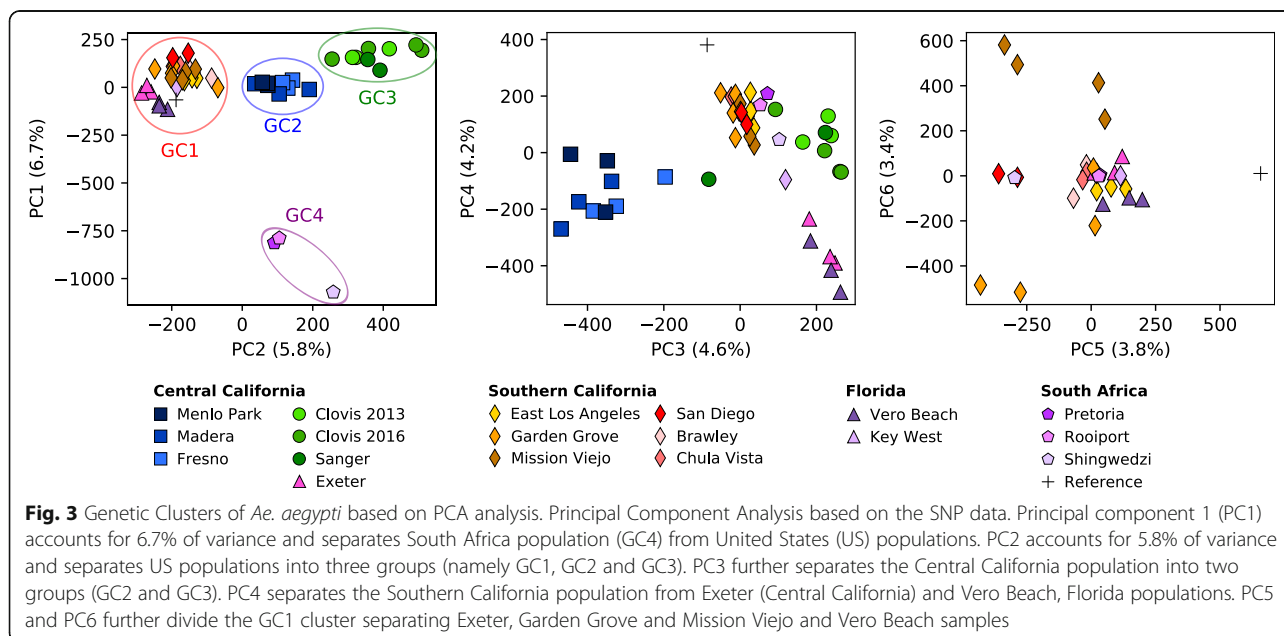
South Africa ($N = 3$) with median genome coverage of 9.6X per sample (Fig. 2, Additional file 1: Table S1). Filtering for biallelic SNPs and a minimum depth of 8 with at most 20% missing data yielded 4,653,297 high-quality SNPs.

Principal Components Analysis based on the genotypes of these SNPs revealed four distinct genetic clusters (Fig. 3). The genetic cluster designated GC1 includes all six cities in southern CA and includes one central CA site, Exeter. Samples from Florida also fall within this cluster. The GC2 cluster includes the northernmost sites near the coast in Menlo Park and includes some central CA populations at Madera and Fresno. Populations from the restricted area around Clovis and Sanger in central CA form the GC3 cluster. The GC4 cluster includes all South African *Ae. aegypti* samples. Overall, the distribution of the three *Ae. aegypti* genetic clusters containing CA *Ae. aegypti* have a nearly parapatric distribution with the three groups potentially converging in the Central Valley. The population at Sanger appears to have multiple genetic clusters occurring in sympatry. One specimen from Sanger (Ae17CON058) could be GC2 or a hybrid of GC2 and GC3 given its values for PC2 and PC3 relative to other samples (Fig. 3). Other Principal components (PC5 and PC6 in Fig. 3) indicate additional population subdivisions further dividing Mission Viejo, Garden Grove, Exeter and Vero Beach samples from the rest of GC2 (Fig. 3). This is consistent with a previous report suggesting highly structured populations within southern CA *Ae. aegypti* [3]. Our genome-wide SNP-based clustering results show clear subdivision between Clovis and Fresno/Madera/Menlo Park. This may seem slightly different from published SNPchip data [3] but is consistent with the microsatellite-based genetic clustering reported in the same study.



Windowed- F_{ST} analysis indicates genome-wide differentiation among the four genetic clusters (Fig. 4a -c). The GC1 and GC4 clusters are the most highly diverged (genome-wide average $F_{ST} = 0.159$). Genetic distance between GC1 and the other clusters is intermediate (Fig. 4 a and b). Nucleotide diversity (π) is highly variable but lowest in the

middle of chromosome 3 (Fig. 4). These regions correspond to the location of the centromeres (coordinates obtained from personal communications with M. Sharakhova at Virginia Polytechnic Institute) and include a relatively high density of repeated and difficult to sequence regions which are excluded from the SNP set analyzed. Overall nucleotide



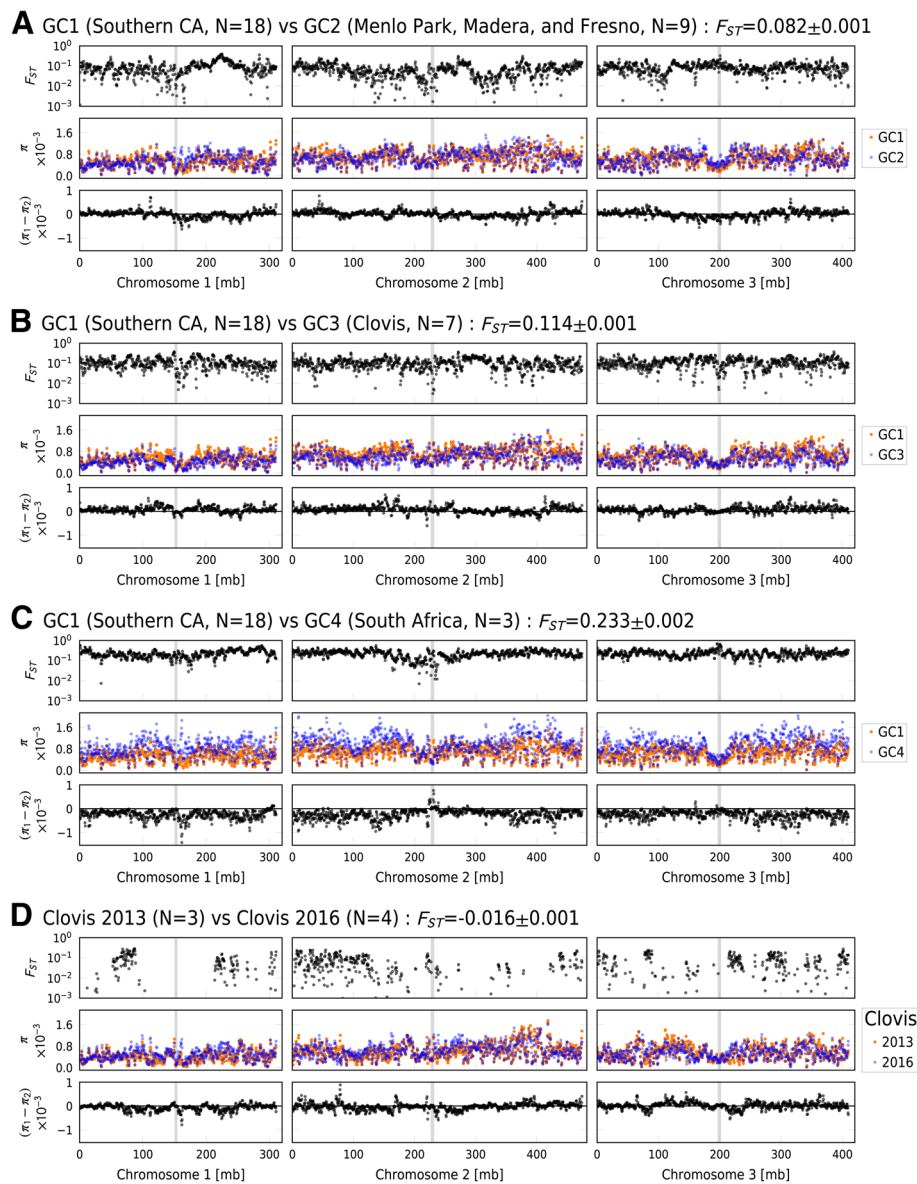


Fig. 4 Genome-wide comparison of *Ae. aegypti* populations revealing broad highly differentiated ($F_{ST} > 0.1$) genomic regions. In each panel: the top subpanel reports Hudson F_{ST} estimator between groups, the middle subpanel shows nucleotide diversity (π) within each group, and the bottom subpanel shows the difference in nucleotide diversity ($=\pi_1 - \pi_2$). Overall $F_{ST} \pm$ estimated standard error between groups is given in the title of each pane. **a:** GC1 (Southern California, $N = 18$) vs GC2 (Central California – Menlo Park, Madera and Fresno, $N = 9$). **b:** GC1 (Southern California, $N = 18$) vs GC3 (Central California – Clovis, $N = 7$). **c:** GC1 (Southern California, $N = 18$) vs GC4 (South Africa, $N = 3$). **d:** Clovis 2013 ($N = 3$) vs 2016 ($N = 4$). Values were calculated using 1Mbp windows with 500Kbp steps. Vertical gray bars indicate the location of centromeres. Other comparisons of GC2 vs GC3, GC2 vs GC4 and GC3 vs GC4 are provided in Additional file 3: Figure S1

diversity is lowest on Chromosome 1, which contains the sex determining locus in *Ae. aegypti*. Overall nucleotide diversity is similar in all population comparisons as indicated by difference in nucleotide diversity ($\Delta\pi = \pi_1 - \pi_2$). As a peculiarity, the South African populations have noticeably higher nucleotide diversity in a region around 160–170 Mbp of Chromosome 1 compared to samples from the southern CA GC1.

For our only temporal comparison, we compared the genomes of samples obtained in Clovis in 2013 with those from samples collected in 2016. Overall genome divergence is negligible ($F_{ST} = -0.025 \pm 0.002$). However, a whole genome scan using 1 Mbp windows for F_{ST} values indicates a number of genomic regions with markedly elevated F_{ST} values (> 0.1) (Fig. 4d). The difference in nucleotide diversity between 2013 and 2016 samples shows an increase over time in chromosome 1

and 2 (mean $\pi_{2016} - \pi_{2013}$ value of 7.22×10^{-5} and 2.32×10^{-5} , respectively) but a decrease on chromosome 3 ($\pi_{2016} - \pi_{2013}$ value of -1.70×10^{-5}). However, regions with a relatively large change in nucleotide diversity between 2016 and 2013 are visible on all three chromosomes, some of which also coincide with highly differentiated ($F_{ST} > 0.1$) regions. These highly differentiated regions with a relatively large nucleotide diversity change may indicate genomic regions under selection presumably as the founding population adapts to local environmental conditions [7].

Differentiation among populations within GC2 ($F_{ST} = -0.043$ to 0.007) and GC3 ($F_{ST} = -0.084$ to 0.000) is minimal (Additional file 2: Table S2). However, differentiation between some populations within GC1 is fairly high ($F_{ST} = 0.000$ to 0.210), especially between Exeter and the southern CA samples ($F_{ST} = 0.132$ to 0.228). This suggests potential substructure within GC1 separating Exeter and Florida samples from other southern CA samples (Fig. 3 and Fig. 5), so they are removed from GC1 in subsequent analysis (Fig. 4 and Additional file 3: Figure S1 panel D). The F_{ST} distance between one sample from the town of Sanger (sample Ae17CON058) to GC2 and GC3 clusters were equivalent (Additional file 2: Table S2) and its placement in the PCA (Fig. 3) and phylogenetic tree (Fig. 5) suggest that this individual could be a GC2/GC3 hybrid.

Estimates of F_{ST} derived from whole genome sequence data have been shown to be accurate even with very small sample sizes (i.e. $N = 2/\text{population}$ [12]). This is due to the very large number of SNPs (i.e. $n \gg 1000$

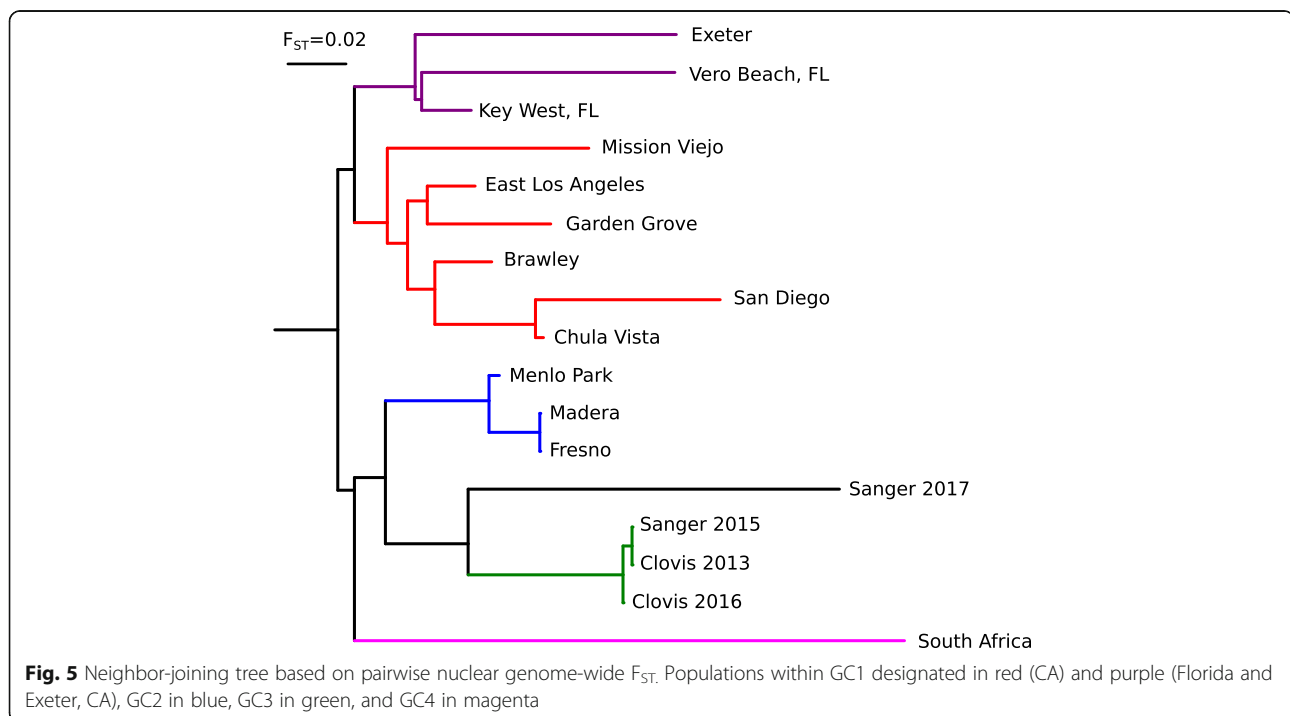
loci) used in these analyses. The 4.65 million loci used in our analysis is well in excess of the number of loci required for an accurate assessment of F_{ST} . In addition, we evaluated various minimum read depths and missing data ratios and observed results consistent with those we report here (data not shown).

Mitogenome sequence analysis revealed two major mitochondrial lineages in CA *Ae. aegypti* (Fig. 6). One lineage includes the *Ae. aegypti* reference sequence and is represented in all four genetic clusters, the other was present in samples from GC1, GC2 and GC4. Samples from Florida and South Africa are distributed among the two major lineages suggesting that these lineages might be present throughout the global range of *Ae. aegypti*.

Discussion

Using nuclear genome sequence data, we identified three major genetic clusters among CA *Ae. aegypti*. These correspond roughly to geographic regions in the state (Figs. 2 and 3). Our data support the hypothesis that *Ae. aegypti* in CA currently exists as multiple, mostly isolated populations. High genetic distance ($F_{ST} > 0.1$) as well as genome-wide differentiation (Fig. 4) support multiple introductions into CA from genetically distinct source populations as the most plausible history of this invasion.

Two major mitochondrial lineages are present within California populations, probably corresponding to previously described global clades [13, 14]. However, their genealogy differs from the nuclear genome genealogy



Contrary to the microsatellite data, the SNP chip data from the same study [3] groups the Exeter population apart from all other CA populations including those in central CA, consistent with our genome-wide SNP data. Unfortunately, their SNP chip data clustering results did not include samples from the southeast USA preventing direct comparison with their SNP clustering result. This, however, could support the view that the Exeter population, introduced in 2014 is distinct from all other CA populations and that it was introduced independently, rather than resulting from local spread of *Ae. aegypti* within CA.

PCA analyses of the SNP chip data separated Clovis (GC3) from the GC2 cluster with some overlap [3]. The larger number of SNPs used in our analysis (> 2.9 million biallelic SNPs compared to 15,698 SNPs) may have increased the resolution, allowing us to confidently separate the two. Our data together with previous reports strongly support multiple introductions of *Ae. aegypti* in California. The most likely scenario includes four independent introductions: (i) Clovis area; probably in 2013 (ii) Madera area; probably in 2013 (iii) southern CA, probably in 2014 (iv) Exeter, probably in 2014 introduced from someplace in the southeast USA like Florida. The years are based on reports from the California vector control districts. This scenario is also in line with most of the results published based on microsatellites and SNP chip data [3]. From our data the exact origin of the introductions remains uncertain with only the Exeter population showing signs of presumable derivation from the southeast USA.

The degree of genetic differentiation found in the Clovis population between the years 2013 and 2016 (Fig. 4d) indicates the population is undergoing rapid changes in its genome, potentially reflecting local adaptation, or, less likely, drift. The only other longitudinal investigation of a CA population of *Ae. aegypti* that we are aware of compares genotypes of samples from 2013 and 2015 from Madera, detecting almost no change within these two years [3]. Further investigation on genic features showing significant differences between the two time points may shed light on the genes involved in local adaptation at Clovis and the particular circumstances that drove it. Because *Ae. aegypti* chromosomes do not produce clearly visible polytene chromosomes like e.g. *Anopheles gambiae*, the detection of chromosome inversions has been challenging and the identification of precise location is at infant stage [22]. Approximate location of diverged regions and the potential chromosome inversions noted by Bernhardt et al. [22] did not provide clear indication that the diverged regions we observed are due to chromosome inversions. Future studies of linkage disequilibrium could illuminate the potential role of chromosome structures in adaptation as it has been demonstrated in *Anopheles* mosquitoes [23].

The geographic origin of CA *Ae. aegypti* populations and the means by which they were introduced remains unclear. Perhaps the most interesting open question is what conditions facilitated multiple introductions? Answering these questions is beyond the scope of this study and requires additional data. Investigating samples from different origins using the same NGS platform may provide a clearer description of *Ae. aegypti* invasion history in CA. In addition, investigation describing genomic changes over time may provide information on local adaptation and potentially will be useful for the control of the species in California.

Conclusion

The mosquito species *Aedes aegypti*, introduced in 2013, has now been detected in multiple locations throughout California. Our genome analyses identified 3 distinct population groups loosely corresponding to different regions within California. Genome-wide comparisons indicate extensive differentiation between genetic clusters. Samples collected from Clovis in two different years (2013 and 2016) reveal genomic signatures of potential selection. Our mitogenome analysis suggests that founding populations were polymorphic for two mitochondrial lineages with one or the other lost in the various extant populations. These observations support recent multiple introductions of *Ae. aegypti* into California. This is the first paper that utilizes the whole genome sequences of *Aedes aegypti* field isolates. Our dataset serves an important step toward future studies aimed at understanding population divergence, gene-environment interactions, and dispersal of this invasive species.

Methods

Mosquito collection

Adult female *Ae. aegypti* were collected from 13 cities by personnel from Mosquito Abatement Districts in Fresno, San Diego, and Orange Counties (Fig. 2 and Additional file 1: Table S1). These mosquitoes are collected using BG Sentinel traps baited with CO₂. All collections on private properties were conducted after obtaining permission from residents and/or owners. Mosquito samples were individually stored in 80% ethanol prior to DNA extraction.

Whole genome sequencing

Genomic DNA was extracted using established protocols [24, 25]. DNA concentrations for each sample were measured using the Qubit dsDNA HS Assay Kit (Life Technologies) on a Qubit instrument (Life Technologies). A genomic DNA library was constructed for each individual mosquito using 20 ng DNA, Qiaseq FX 96 (Qiagen, Valencia, CA), and Ampure SPRI beads (Beckman) following an established protocol [25]. Library concentrations were measured using Qubit (Life Technologies) as

described above. Libraries were sequenced as 150 bp paired-end reads using a HiSeq 4000 instrument (Illumina) at the UC Davis DNA Technologies Core.

Sequence analysis

Raw reads were trimmed using Trimmomatic [26] version 0.36 and mapped to the *AeegL5* reference genome [27] using BWA-MEM [28] version 0.7.15. Mapping statistics were calculated using Qualimap version 2.2 [29] (Additional file 1: Table S1). Joint variant calling using all samples was done using Freebayes [30] version 1.0.1 with standard filters and population priors disabled. We required a minimum depth of 8 to call variants for each individual following the recommendation of Crawford and Lazzaro to minimize bias in population inference [31]. To improve the reliability of calls, we required variants to be supported by both forward and reverse reads overlapping the loci (Erik Garrison, Wellcome Trust Sanger Institute and Cambridge University, personal communication, Dec. 2014). The repeat regions are “soft-masked” in the *AeegL5* reference genome and SNPs in these regions were excluded from analysis. Only biallelic SNPs were used for further analysis. A missing data threshold of 20% was used to filter SNPs. A phylogenetic tree based on the polymorphism data was constructed using the neighbor-joining algorithm as implemented in PHYLIP [32] version 3.696. Hudson F_{ST} [33], nucleotide diversity (π) and Principal Component Analysis (PCA) analyses were done in Python version 3.6.6 using the scikit-allel module version 1.2.0 [34].

The presence of mitochondrial pseudogenes in the nuclear genomes of *Ae. aegypti* could potentially confound SNP calling [35]. Thus we followed the mapping recommendations suggested by Schmidt et al. [14] and mapped raw reads to the mitochondrial reference genome prior to mapping unmapped reads to the nuclear genome.

We used Ae13CLOV028MT (Genbank ID: MH348176) as a reference for mapping the mitochondrial genome because all our specimens contained a deletion between position 14,522 and 14,659 compared to the *AeegL5* reference genome [14]. Variants in the mitochondrial genome were called with Freebayes as described for the nuclear genome, but set to single ploidy. Mitochondrial coverage was on average 160 times greater than the nuclear genome coverage with a minimum of 25-fold difference (Additional file 1: Table S1). Use of properly paired reads for variant calling reduced errors generated by failing to recognize mitochondrial pseudogenes present in the nuclear genome. The Vcf2fasta program [36] was used to extract mitogenome sequences from the VCF file to FASTA format. MEGA version 7.0.26 [37] was used for mitogenome alignment. Mitogenome reference sequences of *Culex quinquefasciatus* (Genbank accession number = HQ724617), *Aedes notoscriptus* (KM676219), and *Aedes albopictus* (NC_006817)

were obtained from GenBank and added to the alignment. Sequences for the thirteen mitochondrial protein-coding genes in *Ae. aegypti* were obtained from GenBank [38], extracted from our dataset, and concatenated for tree construction with the maximum likelihood algorithm implemented in MEGA.

Data visualization

QGIS version 2.18 was used to create maps. Python matplotlib version 3.0.2 (<https://matplotlib.org/>) was used for generating plots. Inkscape (<https://inkscape.org/>) version 0.92 was used to edit images.

Additional files

Additional file 1: Table S1. Metadata for each sample including geographic coordinates, collection date, genome sequencing related metrics, and genotypes of genetic-cluster-specific SNPs. (XLSX 20 kb)

Additional file 2: Table S2. Pairwise F_{ST} values based on nuclear genome SNP data. (XLSX 15 kb)

Additional file 3: Figure S1. Genome-wide comparison of *Ae. aegypti* populations comparing GC2 vs GC3, GC2 vs GC4, GC3 vs GC4, and GC1 vs GC1-B (Vero Beach and Exeter). (PDF 352 kb)

Additional file 4: Appendix S1. Comparison of Evans et al. 2015 SNPchip probe sequences and genotypes with our data. (DOCX 554 kb)

Abbreviations

CA: California; GC: genetic cluster; PCA: Principal Component Analysis; SNP: single nucleotide polymorphism

Acknowledgments

We thank Youki Yamasaki, Allison Chang, Parker Houston, Allison Weakley, Kendra Person, Hans Gripkey for assisting DNA extraction and library preparations for this study. Special thanks to Dr. Bradley Main for his comments to the manuscript. We thank personnel from Consolidated Mosquito Abatement District (Ms. Jodi Holeman and Ms. Katherine Ramirez), Delta, Greater LA County (Dr. Susan Klueh) and San Mateo County Vector Control Districts and Fresno, Madera County and Orange County Mosquito and Vector Control Districts (Mr. Michael Hearst, District Manager), Community Health Division of the Department of Environmental Health (Ms. Rebecca Lafreniere, Chief and Ms. Elizabeth Pozzebon, Director), San Diego County Dept. of Environmental Health, Vector Control, and Dr. Christopher Barker (UC Davis) for providing specimens used in this study. We also thank Dr. Leo Braack (University of Pretoria, South Africa) for assisting AJC in collection of South African samples and Dr. Danny Governer and SANParks for permitting collection from Shingwedzi in the Kruger National Park. We also thank Dr. Lutz Froenicke and his team at the UC Davis DNA technologies Core for genome sequencing.

Funding

We acknowledge funding support from the UC Davis Bridge Funding Program, UC Davis School of Veterinary Medicine Vector-borne Disease Pilot Grant Program, DARPA Safe Gene Program (HR0011-17-2-0047), and the Pacific Southwest Regional Center of Excellence for Vector-Borne Diseases funded by the U.S. Centers for Disease Control and Prevention (Cooperative Agreement 1U01CK000516). YL and MJH received salary support from the UC Davis Bridge Funding Program. YL received research supply support from the UC Davis School of Veterinary Medicine Vector-Borne Disease Pilot Program. YL, MS, JMM, GCL, and AJC received salary and material support from DARPA Safe Gene Program. HS received salary support from the Pacific Southwest Regional Center of Excellence for Vector-Borne Diseases. These funding bodies had no role in sample collection, data analysis, or writing the manuscript.

Availability of data and materials

Sequence data is available through the NCBI Sequence Read Archive (Study accession number: SRP106694). Data is also available through the UC Davis Popl OpenProjects 'AedesGenomes' page [39].

Authors' contributions

YL, GCL, FSM and AJC conceptualized experimental design. CTS, FSM and AJC collected specimens from Clovis and also obtained samples from other locations. YL and GCL assisted in obtaining specimens. YL and MJH carried out DNA extraction, library preparations and submission for genome sequencing. YL, HS, TCC and WRC conducted data analysis. MS, JMM, JCC assisted in data interpretation. All authors contributed to the writing and editing of this manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Vector Genetics Laboratory, Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, University of California Davis, Davis, CA 95616, USA. ²Daniel K. Inouye US Pacific Basin Agricultural Research Center (PBARC), United States Department of Agriculture, Agricultural Research Service, Hilo, Hawaii, USA. ³Department of Entomology and Nematology, College of Agricultural and Environmental Sciences, University of California - Davis, Davis, CA 95616, USA. ⁴Department of Integrative Biology, University of California - Berkeley, Berkeley, CA 94720, USA. ⁵School of Public Health, University of California - Berkeley, Berkeley, CA 94720, USA. ⁶Florida Medical Entomology Laboratory, Institute of Food and Agricultural Sciences, University of Florida, Vero Beach, FL 32962, USA. ⁷Consolidated Mosquito Abatement District, Clovis, CA 93648, USA. ⁸Mosquito Control Research Laboratory, Kearney Agricultural Center, Department of Entomology and Nematology, University of California -, Davis, CA 95616, USA.

Received: 24 October 2018 Accepted: 5 March 2019

Published online: 12 March 2019

References

- Honorio NA, Silva Wda C, Leite PJ, Goncalves JM, Lounibos LP, Lourenco-de-Oliveira R. Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the state of Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz*. 2003;98(2):191–8.
- Harrington LC, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, Jones JJ, Kittthawee S, Kittayapong P, Sithiprasasna R, et al. Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *Am J Trop Med Hyg*. 2005;72(2):209–20.
- Pless E, Gloria-Soria A, Evans BR, Kramer V, Bolling BG, Tabachnick WJ, Powell JR. Multiple introductions of the dengue vector, *Aedes aegypti*, into California. *PLoS Negl Trop Dis*. 2017;11(8):e0005718.
- Gloria-Soria A, Brown JE, Kramer V, Hardstone Yoshimizu M, Powell JR. Origin of the dengue fever mosquito, *Aedes aegypti*, in California. *PLoS Negl Trop Dis*. 2014;8(7):e3029.
- Cornel AJ, Holeman J, Nieman CC, Lee Y, Smith C, Amorino M, Brisco KK, Barrera R, Lanzaro GC, Mulligan III FS: surveillance, insecticide resistance and control of an invasive *Aedes aegypti* (Diptera: Culicidae) population in California. *F1000Res*. 2016;5:194.
- Jewell D, Grodhaus G. In: *Commerce and the Spread of Pests and Disease Vectors* edited by Laird M. New York: Praeger Publishers; 1984. p. 103–7.
- Main BJ, Lee Y, Collier TC, Norris LC, Brisco K, Fofana A, Cornel AJ, Lanzaro GC. Complex genome evolution in *Anopheles coluzzii* associated with increased insecticide usage in Mali. *Mol Ecol*. 2015;24(20):5145–57.
- Main BJ, Lee Y, Ferguson HM, Kreppel KS, Kihonda A, Govella NJ, Collier TC, Cornel AJ, Eskin E, Kang EY, et al. The genetic basis of host preference and resting behavior in the major African malaria vector, *Anopheles arabiensis*. *PLoS Genet*. 2016;12(9):e1006303.
- Vicente JL, Clarkson CS, Caputo B, Gomes B, Pombi M, Sousa CA, Antao T, Dinis J, Botta G, Mancini E, et al. Massive introgression drives species radiation at the range limit of *Anopheles gambiae*. *Sci Rep*. 2017;7:46451.
- Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc Natl Acad Sci U S A*. 2015;112(3):815–20.
- Hanemaaijer MJ, Collier TC, Chang A, Shott CC, Houston PD, Schmidt H, Main BJ, Cornel AJ, Lee Y, Lanzaro GC. The fate of genes that cross species boundaries after a major hybridization event in a natural mosquito population. *Mol Ecol*. 2018; In press.
- Nazareno AG, Bemmels JB, Dick CW, Lohmann LG. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Mol Ecol Resour*. 2017;17(6):1136–47.
- Moore M, Sylla M, Goss L, Burugu MW, Sang R, Kamau LW, Kenya EU, Bosio C, Munoz Mde L, Sharakova M, et al. Dual African origins of global *Aedes aegypti* s.l. populations revealed by mitochondrial DNA. *PLoS Negl Trop Dis*. 2013;7(4):e2175.
- Schmidt H, Hanemaaijer MJ, Cornel AJ, Lanzaro GC, Braack L, Lee Y. Complete mitogenome sequence of *Aedes (Stegomyia) aegypti* derived from field isolates from California and South Africa. *Mitochondrial DNA Part B*. 2018. <https://doi.org/10.1080/23802359.2018.1495117>.
- Damal K, Murrell EG, Juliano SA, Conn JE, Loew SS. Phylogeography of *Aedes aegypti* (yellow fever mosquito) in South Florida: mtDNA evidence for human-aided dispersal. *Am J Trop Med Hyg*. 2013;89(3):482–8.
- Lynch M. The origins of genome architecture. Sunderland, mass: Sinauer Associates; 2007.
- Havird JC, Sloan DB. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Mol Biol Evol*. 2016;33(12):3042–53.
- Molnar RI, Bartelmes G, Dinkelacker I, Witte H, Sommer RJ. Mutation rates and intraspecific divergence of the mitochondrial genome of *Pristionchus pacificus*. *Mol Biol Evol*. 2011;28(8):2317–26.
- Marsden CD, Lee Y, Kreppel K, Weakley A, Cornel A, Ferguson HM, Eskin E, Lanzaro GC. Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector *Anopheles arabiensis*. *G3 (Bethesda)*. 2014;4(1):121–31.
- Evans BR, Gloria-Soria A, Hou L, McBride C, Bonizzoni M, Zhao H, Powell JR. A multipurpose high throughput SNP Chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3 (Bethesda)*. 2015.
- Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*. 2007;316(5832):1718–23.
- Bernhardt SA, Blair C, Sylla M, Bosio C, Black WC: evidence of multiple chromosomal inversions in *Aedes aegypti formosus* from Senegal. *Insect Mol Biol*. 2009;18(5):557–69.
- Ayala D, Ullastres A, Gonzalez J. Adaptation through chromosomal inversions in *Anopheles*. *Front Genet*. 2014;5:129.
- Nieman CC, Yamasaki Y, Collier TC, Lee Y. A DNA extraction protocol for improved DNA yield from individual mosquitoes. *F1000Res*. 2015;4:1314.
- Yamasaki YK, Nieman CC, Chang AN, Collier TC, Main BJ, Lee Y. Improved tools for genomic DNA library construction of small insects. *F1000Res*. 2016;5:211.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014.
- Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH, et al. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*. 2018; 563(7732):501–7.
- Li H: Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In: Cornell University Library; 2013: arXiv:1303.3997v1302.
- Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32(2):292–4.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing; 2012. p. arXiv preprint.
- Crawford JE, Lazzaro BP. Assessing the accuracy and power of population genetic inference from low-pass next-generation sequencing data. *Front Genet*. 2012;3:66.

32. Felsenstein J. PHYLIP - phylogeny inference package (version 3.2). *Cladistics*. 1989;5:164–6.
33. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. *Genetics*. 1992;132(2):583–9.
34. Miles A, Harding N: scikit-allel - Explore and analyse genetic variation. In., 1. 2.0 edn. <https://github.com/cggh/scikit-allel>: GitHub; 2018.
35. Hlaing T, Tun-Lin W, Somboon P, Socheat D, Setha T, Min S, Chang MS, Walton C. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genet*. 2009;10:11.
36. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
37. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
38. Behura SK, Lobo NF, Haas B, deBruyn B, Lovin DD, Shumway MF, Puiu D, Romero-Severson J, Nene V, Severson DW. Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochem Mol Biol*. 2011;41(10):770–7.
39. UC Davis Popl OpenProjects- *AedesGenomes* [<https://popi.ucdavis.edu/PopulationData/OpenProjects/AedesGenomes/>].
40. CalSurv: California Surveillance Gateway Maps. In: California Vectorborne Disease Surveillance System; 2007.
41. Patterson T: CleanTOPO2: Edited SRTM30 Plus World Elevation Data. In. Online (author communicated that his data is published in public domain and free to use).; 2008.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

