

# Historical environmental change in Africa drives divergence and admixture of *Aedes aegypti* mosquitoes: a precursor to successful worldwide colonization?

KELLY LOUISE BENNETT,\* FORTUNATE SHIJA,\*† YVONNE-MARIE LINTON,‡ §¶\*\*  
 GERALD MISINZO,† MARTHA KADDUMUKASA,†† ROUSSEAU DJOUAKA,‡‡  
 OKORIE ANYAELE,§§ ANGELA HARRIS,¶¶ SETH IRISH,\*\*\*<sup>1</sup> THAUNG HLAING,†††  
 ANIL PRAKASH,‡‡‡ JULIUS LUTWAMA†† and CATHERINE WALTON\*

\*Computational Evolutionary Biology Group, Faculty of Life Sciences, University of Manchester, Manchester, UK, †Department of Veterinary Microbiology and Parasitology, Sokoine University of Agriculture, Morogoro, Tanzania, ‡Walter Reed Biosystematics Unit, Smithsonian Institution Museum Support Center, Suitland, MD, USA, §Walter Reed Army Institute of Research, Silver Spring, MD, USA, ¶Uniformed Services University of Health Sciences, Bethesda, MD, USA, \*\*Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA, ††Department of Arbovirology, Emerging and Re-emerging Infections, Uganda Virus Research Institute, Entebbe, Uganda, ‡‡Agro-Eco-Health Platform for West and Central Africa, International Institute of Tropical Agriculture, Cotonou, Republic of Benin, §§Entomology Unit, Department of Zoology, University of Ibadan, Ibadan, Nigeria, ¶¶Mosquito Research & Control Unit, Cayman Islands Government, Grand Cayman, Cayman Islands, \*\*\*London School of Hygiene and Tropical Medicine, London, UK, †††Medical Entomology Research Division, Department of Medical Research (Lower Myanmar), Ministry of Health, Yangon, Myanmar, ‡‡‡National Institute for Research in Environmental Health, Ministry of H & FW Government of India, Bhopal, India

## Abstract

Increasing globalization has promoted the spread of exotic species, including disease vectors. Understanding the evolutionary processes involved in such colonizations is both of intrinsic biological interest and important to predict and mitigate future disease risks. The *Aedes aegypti* mosquito is a major vector of dengue, chikungunya and Zika, the worldwide spread of which has been facilitated by *Ae. aegypti*'s adaption to human-modified environments. Understanding the evolutionary processes involved in this invasion requires characterization of the genetic make-up of the source population (s). The application of approximate Bayesian computation (ABC) to sequence data from four nuclear and one mitochondrial marker revealed that African populations of *Ae. aegypti* best fit a demographic model of lineage diversification, historical admixture and recent population structuring. As ancestral *Ae. aegypti* were dependent on forests, this population history is consistent with the effects of forest fragmentation and expansion driven by Pleistocene climatic change. Alternatively, or additionally, historical human movement across the continent may have facilitated their recent spread and mixing. ABC analysis and haplotype networks support earlier inferences of a single out-of-Africa colonization event, while a cline of decreasing genetic diversity indicates that *Ae. aegypti* moved first from Africa to the Americas and then to Asia. ABC analysis was unable to verify this colonization route, possibly because the genetic signal of admixture obscures the true colonization pathway. By increasing genetic diversity and forming novel allelic combinations, divergence and historical admixture within Africa could have provided the adaptive potential needed for the successful worldwide spread of *Ae. aegypti*.

Correspondence: Catherine Walton, Fax: +44 161 275 5082; E-mail: Catherine.walton@manchester.ac.uk

<sup>1</sup>Present address: Entomology Branch, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

*Keywords:* approximate Bayesian computation, arbovirus, domestication, forest fragmentation, invasive species, Pleistocene climatic change

*Received 25 August 2015; revision received 10 June 2016; accepted 22 June 2016*

## Introduction

Invasions of non-native species have caused widespread declines in biodiversity, impacted on fisheries, agriculture and forestry and had repercussions for public health (Clavero & García-Berthou 2005; Pejchar & Mooney 2009). The latter includes the introduction of exotic disease vectors such as fleas, lice, kissing bugs and mosquitoes which have resulted in devastating epidemics in areas previously free of disease (Lounibos 2002; Tatem *et al.* 2006). Insight into how these species have become successful invaders is needed to assess current and future disease risks and to develop mitigation strategies for vector invasion and disease spread (Lee 2002; Allendorf & Lundquist 2003; Wilson *et al.* 2009; Estoup & Guillemaud 2010). This can be achieved by identifying the key components of evolutionary history contributing to invasion success. For example, high levels of genetic diversity can be beneficial to invasion as this has the potential to mitigate founder effects and facilitate adaptation to new environments (Rius & Darling 2014). The degree of genetic variability will depend on several factors, for example how rapidly populations have expanded, the size of the introduced population and the number of colonization events (Suarez & Tsutsui 2008). Genetic admixture is a particularly effective means to increase genetic diversity and has been incriminated in a number of biogeographic invasions previously (Facon *et al.* 2006; Kolbe *et al.* 2008; Keller & Taylor 2010; Lombaert *et al.* 2010; Lucek *et al.* 2010; Abbott *et al.* 2013; Rius & Darling 2014). Here, we investigated whether admixture characterizes the population history of the invasive mosquito vector, *Aedes aegypti*, and if so, whether this occurred prior to, or during, its worldwide spread.

The *Aedes aegypti* mosquito originates from Africa but colonized the rest of the tropics via intensive shipping trade during the 17th to 20th century (Hlaing *et al.* 2009; Brown *et al.* 2011, 2014; Powell & Tabachnick 2013). As a major vector of arthropod-borne viruses (arboviruses), the invasion of this mosquito coincided with the appearance of yellow fever in the Americas and dengue in the Americas and Asia (Smith 1956; Tabachnick 1991; Urdaneta-Marquez & Failloux 2011; Weaver & Lecuit 2015). Today, *Ae. aegypti* continues to contribute to the growing global threat from arboviral disease as evidenced by the recent pandemics of chikungunya and Zika spread by *Ae. aegypti* in Latin

America (Musso *et al.* 2015; Weaver & Forrester 2015; Weaver & Lecuit 2015). The evolution of human commensalism in *Ae. aegypti* is likely to have contributed to this mosquito's invasion success and its role as a major vector of arboviral disease.

The ancestral form of *Ae. aegypti* in Africa bred in tree holes and fed on nonhuman animals, as do sylvan populations of *Ae. aegypti* and many other members of the *Aedes* genus today (Mattingly 1957, 1967; Huang 2004; Powell & Tabachnick 2013). Presently, *Ae. aegypti* are also found breeding in domestic and peridomestic environments, where they feed primarily on humans (Gouck 1972; Harrington *et al.* 2001). Based on classical definitions by Mattingly (1957, 1967), a distinction is often made between the dark-coloured sylvan form, *Aedes aegypti formosus* and the light-coloured domestic form, *Aedes aegypti aegypti*, despite variation in these characteristics (McClelland 1974). While the sylvan form is commonly described within Africa, the domestic form predominates out of Africa (Mattingly 1957). This domestic form exhibits several adaptations to human habitats including: strong anthropophily that stems at least in part from changes to an odorant receptor; an ability to lay eggs and survive better in nutrient poor water; and a propensity to readily enter houses (McBride *et al.* 2014). At present, we do not know to what extent such adaptive traits were acquired in Africa prior to invasion, or during the transportation and colonization processes.

Microsatellite, SNP and nuclear sequence data have shown that populations within Africa form one genetic group that is genetically differentiated from a second group comprised of all pantropical populations outside Africa (Brown *et al.* 2011, 2014). The sole exception to this was in Rabai, coastal Kenya, which clustered with the pantropical populations and may represent a re-invasion of Africa (Brown *et al.* 2011, 2014). Based on their findings, Brown *et al.* (2014) have suggested that a single subspeciation event from the ancestral subspecies within Africa, referred to as *Ae. aegypti formosus*, gave rise to all pantropical domestic populations, which they refer to as *Ae. aegypti aegypti*.

That one colonization event is inferred indicates that domesticity in *Ae. aegypti aegypti* arose only once, and was integral to the formation and worldwide spread of this form (Powell & Tabachnick 2013); otherwise, we might expect to have seen evidence for multiple colonization events. As there are domestic populations of

*Ae. aegypti* in Africa that are genetically indistinguishable from the sylvatic populations (Paupy *et al.* 2008, 2010; Brown *et al.* 2011), it has been further inferred that there have been multiple independent adaptations to domestic habitats within Africa (Brown *et al.* 2011). Alternatively, these mosquitoes could be somewhat plastic in their use of larval habitats (Mattingly 1957; Chadee *et al.* 1998; Powell & Tabachnick 2013). Understanding when and where domestication has taken place in *Ae. aegypti*, cannot be fully understood without first comprehending the evolutionary history of ancestral *Ae. aegypti* within Africa.

Africa has a complex climatological past with repeated dry, cool glacial periods punctuated by warm and humid interglacials (deMenocal 1995, 2004). Climate-associated habitat change is reflected in pollen and sediment core analysis showing that forests were once more expansive, stretching as far as North Africa during wet and humid interglacial periods (Olago 2001; Hewitt 2004a; Castaneda *et al.* 2009; Watrin *et al.* 2009). During dry glacial periods, rainforest cover was reduced to small habitat refugia in which populations could potentially diverge in allopatry (Diamond & Hamilton 1980; Hamilton 1981; Maley 1991). Genetically distinct populations would subsequently have had opportunities for secondary contact and admixture during intermittent wet interglacials. Phylogeographic studies have revealed that many African species carry signals of past demographic change relating to Pleistocene climate change (Hewitt 2004b). These signals include population or species divergence dating to the Plio-Pleistocene, genetic structure associated with hypothesized refugia and signals of population expansion. Genetic structure promoted by historic forest refugia within Africa during glacial periods has been demonstrated in mammals (Anthony *et al.* 2007 and refs therein; Hassanin *et al.* 2015; Huhndorf *et al.* 2007; Johnson *et al.* 2007; Nicolas *et al.* 2008 and refs therein; Tosi 2008), tropical trees and plants (Kebede *et al.* 2007; Koffi *et al.* 2011 and refs therein; Lowe *et al.* 2010; Plana *et al.* 2004), arboreal reptiles (Measey & Tolley 2011 and refs therein) and rainforest birds (Fjelds  & Bowie 2008; Smith *et al.* 2011). Conversely, the inference that grassland taxa, including insects (Sezonlin *et al.* 2006) and savannah mammals (Barnett *et al.* 2014; Hewitt 2004a; Lorenzen *et al.* 2006 and refs therein), diverged in savannah refugia is consistent with substantial forest expansion during interglacial periods.

Demographic history associated with past climate change is generally complex, involving multiple events which can obscure earlier histories (Guillemaud *et al.* 2009; Estoup & Guillemaud 2010). In addition, stochastic genetic drift and mutation rate variation complicates demographic inference (Nielsen & Beaumont 2009).

Approximate Bayesian computation (ABC) analysis has been developed which allows complex demographic scenarios to be specified which take into account the stochasticity of the coalescent and mutational processes (Nielsen & Beaumont 2009). The Bayesian nature of ABC allows the input of prior information for the comparison of model scenarios, for which data are simulated and represented using a selection of summary statistics. The selection of the best fitting model is based on comparison of posterior probabilities, and the data generated can be used to gain estimates of demographic parameter values. This approach has been successfully used in reconstructing complex colonization scenarios (Guillemaud *et al.* 2009; Fountain *et al.* 2014).

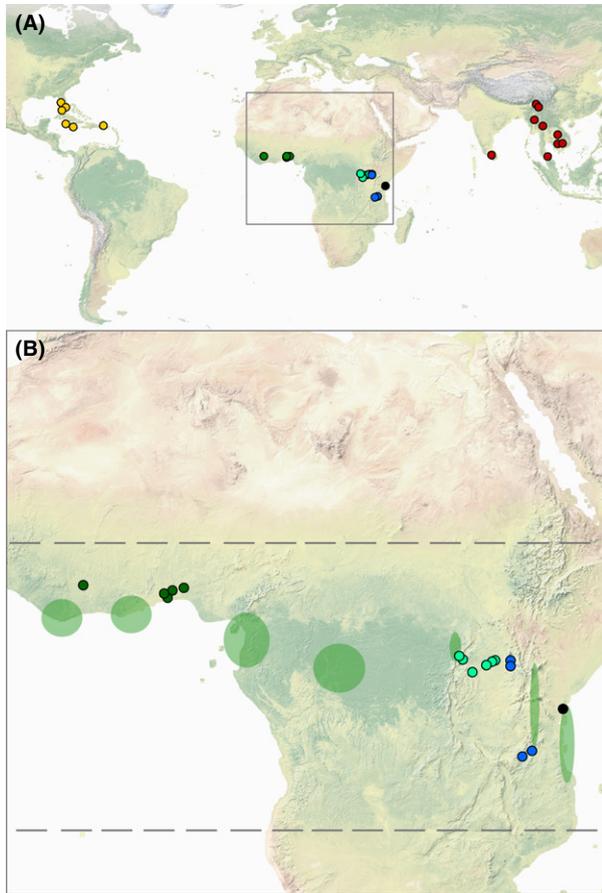
Using primarily an approximate Bayesian computation (ABC) approach, we use genetic information from four nuclear and one mitochondrial marker to analyse the evolutionary history of *Ae. aegypti*. We aim specifically to (i) determine the genetic structure of *Ae. aegypti* within Africa (Gorochotegui-Escalante *et al.* 2000) (ii) infer whether past climate-associated habitat change influenced the genetic structure of African populations (iii) reconstruct the colonization route of *Ae. aegypti* out of Africa and (iv) determine whether any admixture that occurred was older and associated with habitat change or more recent and associated with colonization.

## Methods

### Sample collection

A total of 167 *Aedes aegypti* individuals were collected from 19 localities over West, Central and East Africa (Fig. 1 and Table S1, Supporting information). In addition, 97 individuals were sampled from 17 pantropical populations across Asia and the Americas (Fig. 1 and Table S1, Supporting information). Mosquitoes were collected as eggs using oviposition traps or as larvae from breeding sites in forests and around human settlements. All but five individuals of *Ae. aegypti* originated from domestic habitats (Table S1, Supporting information). As we found the latter to be genetically indistinguishable from domestic individuals, mosquitoes were not separated by breeding site for any analyses.

Larvae were reared to adulthood for identification when possible using the approach detailed by the Walter Reed Biosystematics unit ([www.wrbu.org/techniques.html](http://www.wrbu.org/techniques.html)). Larvae within each individual sample site, that is domestic container, are likely to include siblings. Therefore, larvae from individual sample sites were reared as separate collections from which one individual was taken for genetic analysis. Adult mosquitoes were desiccated and stored with silica in BEEM capsules or pinned and larvae were stored in 100% ethanol



**Fig. 1** (A) Locations of collection sites for *Aedes aegypti* mosquitoes. Regional groups for DIYABC analysis included collections from West Africa (dark green), Central Africa (light green), East Africa (blue), the Americas (yellow) and Asia (pink). The coastal collection from Mombasa, Kenya, excluded from DIYABC analysis is also shown (black). (B) African collection sites in relation to the locations of hypothesized forest refugia present during glacial climate conditions (green ellipses) as described in the Introduction and the extent of forest cover during interglacial periods (represented by dotted lines).

to preserve sample DNA. Mosquitoes were identified using a morphological key (Huang 2004).

### Experimental procedures

DNA was extracted from larvae or legs of the adult mosquito using the modified phenol–chloroform method in Surendran *et al.* (2013). For mosquitoes from the Americas, DNA was extracted from whole adults using the method of Livak (1984). PCR was performed using three newly developed exon-primed intron-crossing (EPIC) primer sets for amplification of a 685-bp region of the enolase gene (*Enol-1* 5'-GTGCGTCACTCAACAGAAGG-3' and 5'-CTCCGAACCGATCTTCATGG-3'); a 776-bp region of the heat shock protein 70 gene (*HSP70* 5'-TGAATGGC

AAACGGATAGC-3' and 5'-CCAGCGAAATGGTTTTA TCC-3'); and 775 bp of the isocitrate dehydrogenase gene (*IDH2* 5'-CAGCAGTGCCTTCTTTTTCC-3' and 5'-CAGTATCGATGCCCTTGTGG-3'). A fourth EPIC marker, the ribosomal protein 30 gene (*RpL30b*), developed by White *et al.* (2015) was also used to amplify a region of approximately 300 bp. PCR amplification was performed in a 25  $\mu$ l volume with a final concentration of 1 $\times$  NH<sub>4</sub> reaction buffer, 2 mM MgCl<sub>2</sub> solution, 0.8 mM dNTP, 0.5  $\mu$ M of each forward and reverse primer with 1.25 units of BIOTAQ™ DNA Polymerase (Bioline, UK) and 1–10 ng of template DNA. Thermal cycling conditions for all EPIC primer sets were as follows: 95 °C for 2 min followed by 35 cycles of 95 °C for 15 s, 60 °C for 30 s, 72 °C for 30 s and a final elongation step at 72 °C for 10 min. Mitochondrial *COI* sequences were amplified with universal primers LCO1490 (5'-GGTCAACAAATC ATAAAGATATTGG-3') and HCO2198 (5'-TAAACTTC AGGGTGACCAAAAAATCA-3') (Folmer *et al.* 1994) and the protocol provided by the Consortium for the Barcode of Life (<http://barcoding.si.edu/dnabarcoding.htm>).

PCR products were purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich Co., LLC, UK) and sequenced on an Applied Biosystems 3730 automated sequencer in both directions using the BIGDYE TERMINATOR v3.1 cycle sequencing kit (Applied Biosystems, UK). Sequences were generated by direct sequencing of PCR products using the amplification primers for *COI* and for the nuclear loci when possible. PCR products from nuclear loci where the direct sequences contained indels or with low haplotype phase probability were cloned using the pGEM-T vector system (Promega, UK). For this, PCR products were generated with a final concentration of 1 $\times$  MyFi reaction buffer, 0.8  $\mu$ M of each primer with 1 unit high fidelity MyFi™ DNA Polymerase (Bioline, UK) and 1–10 ng of template DNA. A 30% ramp speed was used to implement the following thermal cycle: 95 °C for 3 min followed by 30 cycles of 95 °C for 30 s, 62 °C for 45 s and 72 °C for 45 s with no final extension. Product inserts were amplified with M13 primers (5'-TGTAACACGACGCCAGT-3' and 5'-CAGGAAACAGCTATGAC-3') (Messing 1983) using BIOTAQ™ DNA Polymerase and purified and sequenced as above. For the highly variable marker *HSP70*, PCR products were cloned directly from 124 individuals without prior direct sequencing. Two clones per individual were sequenced in the forward direction initially and the individual accepted as being a heterozygote if a minimum of four base changes were apparent on comparison of ABI trace files. At least two further clones were sequenced from remaining individuals before accepting an individual with four congruent sequences as a homozygote. We estimate to have misscored no more

than one individual as a homozygote using this procedure.

#### Data processing

Sequence ABI trace files were aligned and edited in GENIOUS V 5.4.7 (Kearse *et al.* 2012). To ensure homology, compound deletions and insertions affecting less than 5% of sequences were removed from the alignment. A total of 12 of 85 *COI* sequences were found to contain double peaks indicative of nuclear mitochondrial pseudogenes (NUMTs) and were therefore removed from analysis. Additional mitochondrial *COI* sequences were obtained from GenBank as detailed in Table S2 (Supporting information). Haplotype reconstruction for the nuclear loci was achieved with PHASE (Stephens *et al.* 2001). Haplotypes with lower than 0.8 probability of being correct were confirmed through cloning of PCR products as described above or removed. Less than 2% of sequences were removed from analysis due to low probability phase and should therefore have a negligible effect on analysis (Garrick *et al.* 2010). Sequences are available on GenBank (KX444686–KX446469).

#### Population genetic analysis

**Recombination.** Sequence alignments were tested for recombination using the pairwise homoplasy index in PhiPack (Bruen *et al.* 2006) and Max X2, Chimaera and GENECONV in RDP4 (Martin *et al.* 2010).

**Genetic diversity and population structure.** Nucleotide diversity ( $\pi$ ), Watterson's  $\theta$  ( $\theta_w$ ), number of segregating sites (S) and number of haplotypes were computed in DNASP v5 (Librado & Rozas 2009). Tests of a null model of neutrality and constant population size, Tajima's D and Fu's  $F_s$ , and estimates of genetic differentiation ( $F_{ST}$ ) were computed in ARLEQUIN v3.5 (Excoffier & Lischer 2010). Isolation by distance was tested using a Mantel test to assess the correlation between genetic and geographic distance in R v3.1.3 (RCoreTeam 2015) using the packages ADEGENET (Jombart 2008) and APE (Paradis *et al.* 2004). Isolation by distance was first tested within Africa and secondly to include all worldwide data. As the *COI* data set includes sequences from GenBank, samples were used if coordinates or specific locality were known.

To estimate the genealogical relationships among haplotypes, we constructed median-joining networks using NETWORK v4.6.1.3 (Fluxus Technology Ltd) and resolved any reticulation using the rules in Crandall & Templeton (1993). As missing data affects the accuracy of haplotype networks, the *COI* data set was trimmed to 493 bp to include more sequences as many were

taken from GenBank and were shorter than the consensus. Otherwise, indels are annotated on networks.

**Approximate Bayesian computation.** The comparison of ABC scenarios was achieved in DIYABC v2.0.4 (Cornuet *et al.* 2008) with combined analysis of all four nuclear sequence data sets and independent analysis of the mitochondrial *COI* sequence data because previous findings suggest this may have a different population history to that of nuclear markers (Moore *et al.* 2013; Brown *et al.* 2014). To test for concordance of results, each nuclear sequence data set was also analysed separately in DIYABC v2.1.0. In each case we simulated a total of 2 000 000 data sets for analyses of nuclear loci. A larger number of data sets (3 000 000) were simulated on analysis of the *IDH2* marker alone as several scenarios displayed high posterior probabilities. On analysis of mitochondrial *COI*, 6 000 000 data sets were simulated under the hypothesized ancestral population history scenarios, while 3 000 000 data sets were simulated under colonization and admixture scenarios with equal probability given to each. Trials were run under a range of prior values to assess the ability of simulations to produce summary statistics close to the observed data. The final prior definitions and distributions are provided in Tables S3–S4 (Supporting information). For nuclear and mitochondrial genes, the number of segregating sites (S) and variance of pairwise differences within populations were chosen for one sample summary statistics. In addition, the mean of pairwise differences between populations (W) and  $F_{ST}$  were chosen as two sample summary statistics to assess scenarios of ancestral divergence and colonization. A logistic regression approach (Cornuet *et al.* 2008) implemented within DIYABC was used to estimate the posterior probabilities of scenarios using 1% of data sets closest to the observed data. To evaluate confidence in the posterior probabilities of scenarios, a model checking computation was performed through generation of 1000 pseudo-observed data sets with parameters drawn from the posterior distribution. To avoid overestimating the fit of the scenario, we used different statistics during model checking to those used for ABC analysis (Cornuet *et al.* 2010). To assess model fit, the one sample summary statistics used were mean of pairwise differences within a population, Tajima's D and the number of private segregating sites. The two sample summary statistics used were S and mean of pairwise differences between populations (B). Pseudo-observed data sets ( $n = 500$ ) were also used to calculate the posterior-based error rate; that is, the proportion of times the most probable scenario is chosen incorrectly, using the scenario choice and parameter values of the nearest 500 simulated data sets to the observed data.

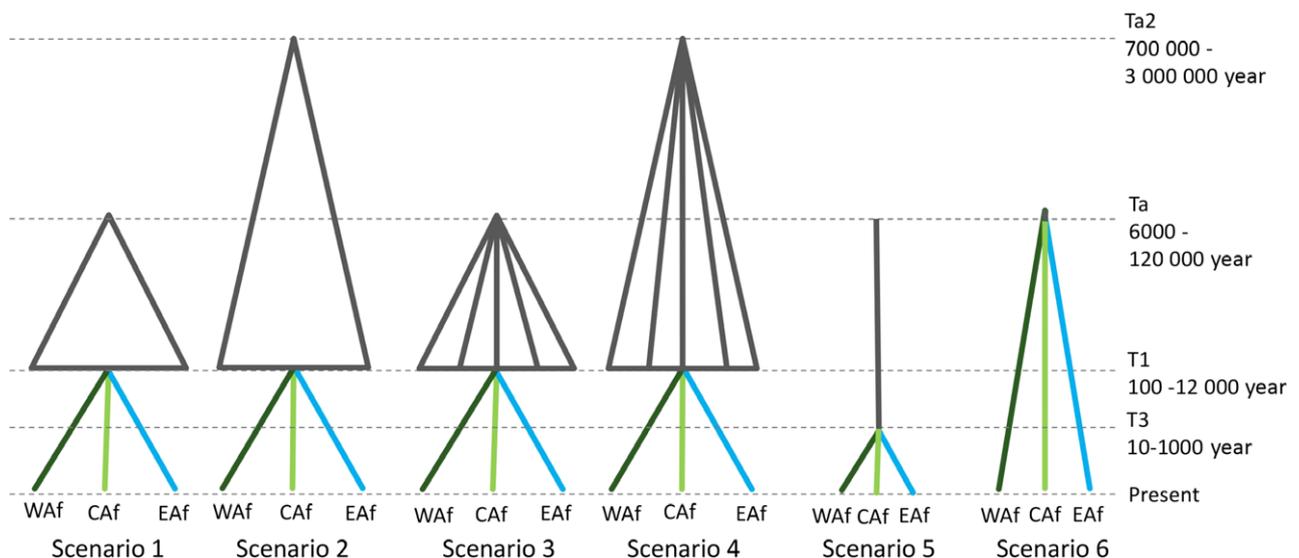
*Hypotheses of ancestral population history in Africa.* Based on previous findings that historical climate change was important in the evolution of African flora and fauna (see Introduction), we hypothesize that historical climatic oscillations have impacted the evolutionary history of *Ae. aegypti*. If so, we expect to detect signals of population structure consistent with allopatric divergence during glacial maxima and signals of secondary contact during hypothesized interglacial periods. We consider four alternative hypotheses involving divergence and admixture for the ABC analysis (scenarios 1–4, Fig. 2) that vary in the number of putative forest refugia and in the timing of divergence events, informed by available data as detailed below.

Two highly divergent mtDNA lineages (3%) are present in *Ae. aegypti* in Africa and around the world (Gorochotegui-Escalante *et al.* 2000; Bosio *et al.* 2005; Bracco *et al.* 2007; Scarpassa *et al.* 2008; Paupy *et al.* 2012), from which we infer that *Ae. aegypti* potentially diverged in two independent forest refugia. Scenarios 1 and 2 therefore first consider that populations of *Ae. aegypti* diverged in two independent forest refugia. Alternatively, *Ae. aegypti* may have diverged in multiple habitat fragments as several forest refugia have been postulated across Africa based on pollen and sediment core analysis (Diamond & Hamilton 1980; Maley 1991). Up to five putative refugia have been identified throughout West and Central Africa including the western Albertine Rift with other possible forest refugia in East Africa in the Eastern Arc mountains of Tanzania and Kenya and the East African Rift mountains of Kenya (Diamond & Hamilton 1980; Maley 1991; Roy 1997; Ray & Adams 2001; Fjelds a *et al.* 2007). In addition, the coastal forests

of East Africa could have provided refugia given that sylvan *Ae. aegypti* are still found there alongside the domestic form (Tabachnick *et al.* 1979; Powell & Tabachnick 2013). Based on these data, in scenarios 3 and 4 we model five allopatric divergence events to capture a reasonable level of population divergence associated with multiple refugia.

We also consider two time periods for ancestral population divergence. Scenarios 1 and 3 simulate a period of divergence within the late Pleistocene. For these scenarios, we place the upper bound of ancestral population divergence ( $T_a$ ) at 120 000 years ago to encompass the last two glacial periods which occurred in between estimated humid interglacials 10 000 and 45–50 000 years ago, and 45–50 000 and 110–120 000 years ago (Castaneda *et al.* 2009). Although there is the potential for populations of *Ae. aegypti* to be impacted by multiple divergence and admixture events throughout the Plio-Pleistocene, we only test for the most recent events because admixture can obscure previous demographic signal, making inference difficult (Qu *et al.* 2011, 2012; Sonsthagen *et al.* 2012).

In comparison, scenarios 2 and 4 simulate an older divergence event occurring in the mid-Pleistocene based on the observation of highly distinct mtDNA lineages in African *Ae. aegypti*. By applying divergence rates of  $2.3\% \text{ My}^{-1}$  (Papadopoulou *et al.* 2010) and  $2.54\% \text{ My}^{-1}$  (Brower 1994) we estimate these lineages diverged approximately 700 000 to 3 000 000 years ago. Based on this and preliminary runs to assess the ability of prior ranges to simulate data close to the observed data, we set our time prior ( $T_{a2}$ ) at 700 000 to 3 000 000 years ago. All models designed to test hypotheses of



**Fig. 2** Six demographic scenarios compared within DIYABC for African populations including West Africa (Waf), East Africa (Neafsey *et al.* 2015) and Central Africa (Caf). Scenarios 1 and 3 were the models with the highest posterior probabilities.

historical climate change (scenarios 1–4) include a period of admixture placed at a maximum of 12 000 years ago (T1) to encompass the last interglacial period of forest expansion in Africa, estimated at 6000 to 10 000 years ago (Castaneda *et al.* 2009).

We also consider a model where *Ae. aegypti* represents a single panmictic population unaffected by historical population divergence. As in all scenarios, recent divergence (T3) in the last 10 to 1000 years is included to allow comparison with other hypothesized models, which require the same number of populations (scenario 5, Figs 2 and S1, Supporting information). Finally we test a null model of independent divergence without mixing of populations with the prior range for ancestral population divergence (Ta) set to the same as in scenarios 1 and 3, at 6 to 120 000 years ago (scenario 6, Figs 2 and S1, Supporting information).

We designate regional groups for analysis based on geographic proximity. Within the West African group we include Ivory Coast, Benin and Nigeria, the Central Africa group comprises Uganda and the East African group includes Kenya and Tanzania. In this last group we have excluded sequences originating from coastal Kenya and Tanzania because these locations may be subject to re-introduction from outside of Africa (Brown *et al.* 2011).

Split times for our models were translated into years using a generation time of 0.1 years based on estimates in *Drosophila* (Cutter 2008) and *Ae. aegypti* (Irvin *et al.* 2004; Sowilem *et al.* 2013). Minimum and maximum effective population size priors were set to cover a wide range and to encompass values approximated from  $\Theta = 4N_e\mu$  using the diversity statistics reported for nuclear genes in Brown *et al.* (2014) and the substitution rate for *Drosophila* of  $1.6 \times 10^{-8}$  per site per year (Moriyama & Gojoberi 1992).

*Hypotheses of colonization routes out of Africa.* We test for the involvement of admixture in the invasion of worldwide *Ae. aegypti* in a hierarchical manner. We first test 12 scenarios to identify the most probable route of colonization without considering admixture (Fig. S2, Supporting information). Colonization from West Africa could have occurred with the intensive shipping trade in place between West Africa and ports throughout South and Central America during the late 15th century to early 19th century (Tabachnick 1991; Urdaneta-Marquez & Failloux 2011; Powell & Tabachnick 2013). Similarly migration could have occurred from East Africa via historic trading routes to India (Powell & Tabachnick 2013). In keeping with historical information, the colonization scenarios consider that migration out of Africa could have occurred from one or both these regions between 1 and 700 years ago (T1 and T2).

Scenarios 1–6 consider independent colonizations of Asia and the Americas from either West or East Africa within the last 700 years. Scenarios 7 through 10, consider stepping stone introduction as a possibility. Scenario 7 is designed to specifically test the findings of Brown *et al.* (2014) that *Ae. aegypti* travelled from Africa westwards to the Americas and from there to Asia and the Pacific region based on a cline of decreasing genetic diversity indicative of successive founding events. In addition to the above hypotheses, it is possible that both American and Asian populations are independently derived from a particularly successful source within Africa, for example, if the acquisition of domestic adaptation was involved in geographic spread. Scenarios 11 and 12 test this possibility.

We secondly test six models to determine the most probable admixture scenario in an independent analysis (Fig. S3, Supporting information). To do this, we consider that all populations outside Africa stem from an admixed population, which could include admixture between East and West Africa (S3, scenarios 1–3) or admixture between unsampled populations (S3, scenarios 4–6). As in the previous analysis, it is considered that colonization could have occurred in a stepping stone fashion (scenarios 1, 2, 4 and 5) or independently (scenarios 3 and 6) from the admixed population. We finally test the most probable scenarios from each of the two analyses together to determine whether the scenarios involving admixture are favoured as detailed in the Results (Fig. S4, Supporting information).

For all models, we provide an upper bound of 12 000 years for population divergence (Ta) to overlap with the period of ancestral population history that is pertinent to distinguishing the alternative colonization scenarios. The generation time used to convert divergence estimates into years and prior ranges for effective population sizes is as described previously. The lower bound of the prior for  $N_e$  for Asian and American populations was lowered to allow for a population bottleneck on colonization (Table S4, Supporting information). For DIYABC analysis East and West African populations were once again included as regional groups. Central Africa is not considered in colonization scenarios as this population is inland and unlikely to represent a source for global populations. The regional groups Asia and Americas include all populations sampled from these continents.

## Results

### *Genetic marker diversity*

Highest overall nucleotide diversity was provided by markers *HSP70* ( $\theta_w = 0.078$ ,  $\pi = 0.013$ ), *Enol-I*

( $\theta_W = 0.053$ ,  $\pi = 0.016$ ), and *RpL30b* ( $\theta_W = 0.0455$ ,  $\pi = 0.008$ ) while *IDH2* had the lowest genetic diversity ( $\theta_W = 0.020$ ,  $\pi = 0.008$ ). In comparison to the EPIC markers, overall genetic diversity values for mitochondrial *COI* were low ( $\theta_W = 0.017$ ,  $\pi = 0.007$ ).

### Population genetic analysis

**Recombination.** Tests for recombination including Phi, Max X2, Chimaera and GENECONV were not significant for all four nuclear genes indicating that the data sets are not affected by recombination.

**Genetic diversity and population structure.** Genetic diversity was higher in Africa than America or Asia for all genetic data sets. Measures of  $\theta_W$  were higher for all Africa-to-out-of-Africa regional pairwise comparisons, while  $\pi$  was higher in 18 of 20 comparisons (Table 1). Exceptions are attributed to reduced  $\pi$  at nuclear *IDH2* and mitochondrial *COI*. In addition, there is a cline of

decreasing genetic diversity moving from the Americas to Asia for 8 of 10 comparisons for  $\theta_W$  and  $\pi$  for each locus.

Tajima's D and Fu's  $F_s$  values for the combined African data set are significant and negative for all four nuclear data sets and are often significant for regional data sets including West, Central and East Africa (Table 1). No significant neutrality statistics were obtained from analysis of the mtDNA data for all regions including the Americas and Asia.

For all sequenced nuclear markers,  $F_{ST}$  values indicate that Africa is significantly differentiated from the Americas and Asia (Table 2). Conversely, three of four nuclear genes detect no differentiation between the Americas and Asia. In comparison, analysis of the *COI* data set produces significant  $F_{ST}$  differences between all geographic regions excepting the Americas and Central Africa.  $F_{ST}$  estimates for each of the four nuclear data sets and the *COI* gene reveal significant genetic differentiation between East and West Africa.

**Table 1** Genetic diversity values including number of sequences (No. seqs), number of haplotypes (No. haps), number of segregating sites (No. S), Watterson's theta ( $\theta_W$ ) and nucleotide diversity ( $\pi$ ) by region and continent. Significant values that deviate from the null hypothesis of neutral mutation for Tajima's D and Fu's  $F_s$  ( $P < 0.05$ ) are annotated (\*)

| Gene          | Region         | No. seqs | No. haps | No. S | $\theta_W$ | $\pi$ | Tajima's D | Fu's $F_s$ |
|---------------|----------------|----------|----------|-------|------------|-------|------------|------------|
| <i>Enol_1</i> | Africa         | 232      | 121      | 90    | 0.053      | 0.020 | -1.870*    | -24.507*   |
|               | West Africa    | 110      | 76       | 73    | 0.046      | 0.001 | -1.775*    | -25.104*   |
|               | Central Africa | 26       | 17       | 32    | 0.027      | 0.020 | -0.890     | -3.887     |
|               | East Africa    | 96       | 49       | 38    | 0.025      | 0.018 | -1.015     | -23.141*   |
|               | Asia           | 80       | 16       | 20    | 0.013      | 0.011 | -0.130     | -1.239     |
|               | Americas       | 114      | 16       | 21    | 0.014      | 0.009 | -0.874     | -2.625     |
| <i>HSP70</i>  | Africa         | 208      | 151      | 159   | 0.081      | 0.018 | -2.313*    | -24.501*   |
|               | West Africa    | 104      | 76       | 85    | 0.045      | 0.016 | -1.993*    | -24.940*   |
|               | Central Africa | 32       | 28       | 42    | 0.028      | 0.019 | -0.880     | -14.506*   |
|               | East Africa    | 72       | 61       | 107   | 0.057      | 0.020 | -2.181*    | -24.777*   |
|               | Asia           | 66       | 15       | 19    | 0.010      | 0.008 | -0.445     | -0.5654    |
|               | Americas       | 96       | 26       | 29    | 0.015      | 0.010 | -0.741     | -4.961     |
| <i>IDH2</i>   | Africa         | 270      | 51       | 49    | 0.019      | 0.007 | -1.709*    | -25.797*   |
|               | West Africa    | 102      | 23       | 25    | 0.011      | 0.007 | -0.997     | -7.608*    |
|               | Central Africa | 56       | 25       | 32    | 0.016      | 0.008 | -1.766*    | -12.936*   |
|               | East Africa    | 112      | 26       | 31    | 0.014      | 0.008 | -1.293     | -9.167*    |
|               | Asia           | 48       | 10       | 11    | 0.006      | 0.008 | 1.032      | 0.033      |
|               | Americas       | 70       | 19       | 21    | 0.010      | 0.010 | 0.125      | -3.242     |
| <i>RpL30b</i> | Africa         | 330      | 74       | 67    | 0.045      | 0.010 | -2.130*    | -25.997*   |
|               | West Africa    | 122      | 29       | 32    | 0.023      | 0.009 | -1.654*    | -10.659*   |
|               | Central Africa | 82       | 45       | 37    | 0.030      | 0.012 | -1.729*    | -26.464*   |
|               | East Africa    | 130      | 34       | 43    | 0.031      | 0.010 | -1.894*    | -18.437*   |
|               | Asia           | 80       | 8        | 6     | 0.004      | 0.004 | -0.847     | -2.130     |
|               | Americas       | 112      | 11       | 11    | 0.008      | 0.004 | -0.320     | -0.435     |
| <i>COI</i>    | Africa         | 169      | 29       | 22    | 0.014      | 0.007 | -0.678     | -6.233     |
|               | West Africa    | 34       | 9        | 10    | 0.006      | 0.004 | -0.867     | 1.096      |
|               | Central Africa | 39       | 14       | 13    | 0.010      | 0.007 | -0.519     | -1.529     |
|               | East Africa    | 96       | 20       | 24    | 0.011      | 0.010 | -0.195     | -1.493     |
|               | Asia           | 31       | 10       | 14    | 0.009      | 0.007 | -1.097     | 0.793      |
|               | Americas       | 30       | 11       | 17    | 0.009      | 0.013 | 1.991      | 2.436      |

**Table 2** Pairwise  $F_{ST}$  values by geographic region for each genetic marker. Significant values that deviate from the null hypothesis of no genetic differentiation are shown in bold

|                | Gene          | Africa       | West Africa  | Central Africa | East Africa  | Asia         |
|----------------|---------------|--------------|--------------|----------------|--------------|--------------|
| West Africa    | <i>Enol-I</i> | <b>0.013</b> |              |                |              |              |
|                | <i>HSP70</i>  | <b>0.069</b> |              |                |              |              |
|                | <i>IDH2</i>   | −0.001       |              |                |              |              |
|                | <i>RpL30b</i> | <b>0.006</b> |              |                |              |              |
|                | <i>COI</i>    | <b>0.124</b> |              |                |              |              |
| Central Africa | <i>Enol-I</i> | <b>0.025</b> | <b>0.080</b> |                |              |              |
|                | <i>HSP70</i>  | <b>0.029</b> | <b>0.170</b> |                |              |              |
|                | <i>IDH2</i>   | 0.003        | 0.006        |                |              |              |
|                | <i>RpL30b</i> | 0.000        | <b>0.013</b> |                |              |              |
|                | <i>COI</i>    | 0.009        | <b>0.075</b> |                |              |              |
| East Africa    | <i>Enol-I</i> | <b>0.008</b> | <b>0.057</b> | 0.015          |              |              |
|                | <i>HSP70</i>  | 0.003        | <b>0.079</b> | 0.005          |              |              |
|                | <i>IDH2</i>   | 0.002        | <b>0.015</b> | <b>0.021</b>   |              |              |
|                | <i>RpL30b</i> | 0.006        | <b>0.036</b> | <b>0.014</b>   |              |              |
|                | <i>COI</i>    | <b>0.026</b> | <b>0.252</b> | <b>0.083</b>   |              |              |
| Asia           | <i>Enol-I</i> | <b>0.086</b> | <b>0.095</b> | <b>0.189</b>   | <b>0.115</b> |              |
|                | <i>HSP70</i>  | <b>0.030</b> | <b>0.116</b> | −0.001         | 0.002        |              |
|                | <i>IDH2</i>   | <b>0.051</b> | <b>0.061</b> | <b>0.077</b>   | <b>0.040</b> |              |
|                | <i>RpL30b</i> | <b>0.289</b> | <b>0.358</b> | <b>0.308</b>   | <b>0.332</b> |              |
|                | <i>COI</i>    | <b>0.313</b> | <b>0.233</b> | <b>0.266</b>   | <b>0.416</b> |              |
| Americas       | <i>Enol-I</i> | <b>0.167</b> | <b>0.161</b> | <b>0.356</b>   | <b>0.230</b> | <b>0.056</b> |
|                | <i>HSP70</i>  | <b>0.038</b> | <b>0.117</b> | 0.017          | 0.007        | −0.008       |
|                | <i>IDH2</i>   | <b>0.070</b> | <b>0.080</b> | <b>0.099</b>   | <b>0.040</b> | 0.037        |
|                | <i>RpL30b</i> | <b>0.303</b> | <b>0.370</b> | <b>0.317</b>   | <b>0.341</b> | 0.009        |
|                | <i>COI</i>    | <b>0.066</b> | <b>0.225</b> | 0.042          | <b>0.095</b> | <b>0.285</b> |

Low, but significant isolation by distance was detected within Africa from the nuclear genes *Enol-I* ( $r = 0.058$ ,  $P = 0.01$ ) and *HSP70* ( $r = 0.050$ ,  $P = 0.01$ ). There is some evidence for isolation by distance among worldwide populations with significant signals detected with nuclear *IDH2* ( $r = 0.112$ ,  $P = 0.01$ ) and mitochondrial *COI* ( $r = 0.153$ ,  $P = 0.01$ ) genes.

The median-joining haplotype networks for nuclear and mitochondrial genes (Figs S5–S9, Supporting information) reveal that the genetic diversity of pantropical regions is composed of a few, common haplotypes sampled from the greater diversity within Africa. These haplotypes are generally shared between the Americas and Asia. There is some evidence for relatively recent, local geographic structuring; the strongest signal is seen in *Enol-I*, where at the end branches there are small clusters of closely related haplotypes specific to West Africa, East Africa and the Americas (Fig. S5, Supporting information). Consistent with the low genetic diversity of *IDH2* no local structuring is apparent at this locus and a large proportion of African haplotypes are shared by pantropical populations compared to the other haplotype networks (Fig. S8, Supporting information). The network for the mitochondrial gene *COI*

reveals a different genealogical history to the nuclear loci with the mtDNA network comprising two major haplogroups separated by 6–8 mutational steps (Fig. S9, Supporting information). Haplotypes from East Africa, Central Africa, the Americas and Asia are found in both lineages with local substructure apparent by the clustering of haplotypes specific to these geographic regions at the end branches for all but Central Africa. Haplotypes from West Africa are only found in haplogroup 2.

#### Approximate Bayesian computation analysis

*Ancestral divergence in Africa.* Six scenarios were considered to determine whether ancestral populations of *Ae. aegypti* were affected by past climate-associated habitat change within Africa. Posterior probabilities of scenarios and confidence intervals for model choice are given in Table 3 and Fig. S1, Supporting information. Comparison of posterior probabilities using local linear regression provides the highest support for scenario 3 on combined analysis of nuclear loci and scenario 1 for mitochondrial *COI*, which are both models for Holocene divergence and admixture (Fig. 2); all other models have low probability support. Individual analysis of nuclear loci showed support for either scenario 1 or 3, in three of four cases. The exception is for the gene *IDH2* where comparison of posterior probabilities suggests that scenario 5, inferring no ancestral divergence within populations, is the most probable scenario. It is possible that this results from positive selection acting on *IDH2* as this gene has the lowest genetic diversity of the four data sets and an HKA test involving comparisons to sequences of *Ae. hansfordi* and the nuclear *RpL30b* gene just failed to be significant at the 5% level ( $\chi^2 = 3.27$ ,  $P = 0.07$ ). Posterior error rates (Table 3) and model checking revealed that the chosen scenarios were a good fit to data simulated from the posterior distribution with only 1 of the summary statistics used for model checking falling outside the 95% confidence interval on analysis of the nuclear gene data set. Significant statistics indicate that observed data are not perfectly predicted by the estimated parameters; this can be attributed to low information content in the data rather than bad choice of model or priors distributions. Our data have a limited power to obtain parameter values as these have large confidence intervals (Table S3, Supporting information) and so should be interpreted with caution. However, model checking and posterior error rates suggest that the model selection is robust.

*Colonization out of Africa.* To determine the colonization route of *Ae. aegypti* out of Africa, 12 divergence scenarios were considered to determine the most probable scenario in the absence of admixture. For the combined

**Table 3** Highest probability scenarios for DIYABC analysis with their posterior probabilities, 95% confidence intervals and posterior error rates

| Question  | Data set | Description of selection scenario  | Post. | 95% CI      | Posterior |
|---|----------|--|-------|-------------|-----------|
|   |          |  | prob. |             | error     |
| 1. Were African populations impacted by past climate change?    | nDNA     | Scenario 3: Late Pleistocene divergence into two populations, Holocene admixture and recent divergence of East, West and Central Africa                              | 0.645 | 0.607–0.683 | 0.188     |
|   | MtDNA    | Scenario 1: Late Pleistocene divergence into two populations, Holocene admixture and recent divergence of East, West and Central Africa                              | 0.396 | 0.375–0.418 | 0.224     |
| 2. What is the most probable colonization route out of Africa?  | nDNA     | Colonization 8: Colonization out of West Africa to Asia and from Asia to the Americas  | 0.443 | 0.413–0.473 | 0.534     |
|   | MtDNA    | Colonization 4: Independent out of East Africa to Asia and to the Americas<br>Colonization 10: Colonization out of East Africa to Asia and from Asia to the Americas | 0.265 | 0.249–0.280 | 0.706     |
| 3. What is the most probable colonization route with admixture? | nDNA     | Admixture 2: Admixture between East and West Africa, colonization of Asia and migration from Asia to the Americas  | 0.322 | 0.306–0.338 | 0.494     |
|   | MtDNA    | Admixture 2: Admixture between East and West Africa, colonization of Asia and migration from Asia to the Americas  | 0.393 | 0.370–0.415 | 0.584     |
| 4. Did colonization involve recent admixture?                   | nDNA     | Colonization 8: Colonization out of West Africa to Asia and from Asia to the Americas  | 0.267 | 0.244–0.289 | 0.634     |
|   | MtDNA    | Colonization 4: Independent colonization out of East Africa to Asia and to the Americas  | 0.251 | 0.238–0.263 | 0.672     |

nuclear data set and individual analysis of three nuclear genes, *Enol-I*, *HSP70* and *RpL30b*, posterior probabilities identified scenarios 7 or 8, which model colonization from West Africa to either Asia or the Americas before stepping stone introduction, as the most probable colonization routes (Table 3 and Fig. S2, Supporting information). Conversely, individual analysis of the nuclear gene *IDH2* and mitochondrial *COI* showed support for an East African origin by providing the highest support for scenarios 9 or 10, and 4 respectively (Table 3 and Fig. S2, Supporting information). In contrast to the other nuclear markers, ABC analysis has difficulty in choosing a high probability model for *IDH2* and *COI* with the confidence intervals of posterior probabilities being either close or overlapping, indicating a lack of demographic signal in these markers.

For the ABC analysis of six evolutionary scenarios involving admixture, multiple scenarios had high posterior probabilities for nuclear and mitochondrial loci (Fig. S3, Supporting information). The most probable colonization scenarios with admixture (admixture scenarios 1, 2, 4 and 5) were tested against possible colonization scenarios without admixture (colonization scenarios 3, 4, 7, 8, 9 and 10) to determine whether recent population admixture was involved in the spread of *Ae. aegypti*. In all cases, excepting individual analysis of nuclear *IDH2*, the models previously chosen without admixture were preferred (Fig. S4, Supporting

information). That many of the models tested share characteristics in common is reflected in a relatively high posterior error rate (Table 3) and high posterior probabilities for more than one colonization scenario; for example, scenarios 7 and 8 both include stepping stone introduction from West Africa, while scenarios 9 to 10 both model stepping stone introduction from East Africa (Fig. 3). The model checking algorithm generally indicated a good fit of models to the observed data. However, for the analysis of nuclear genes the statistic, private segregating sites, was significantly outside the 95% confidence interval, suggesting that the observed data is not perfectly predicted in this case. These significant values may result from simplification of the model or because of limited power in the data to resolve evolutionary history. Parameter estimates from analysis of the combined nuclear data set are given in Table S4 (Supporting information).

## Discussion

As in previous studies based on microsatellites, SNPs and nuclear markers (Hlaing *et al.* 2010; Brown *et al.* 2011, 2014), we find weak geographic structure within Africa revealed by low  $F_{ST}$  between populations at four nuclear genes with genetic diversity shared among geographic regions, as can be seen in the haplotype networks for both mitochondrial and nuclear genes.

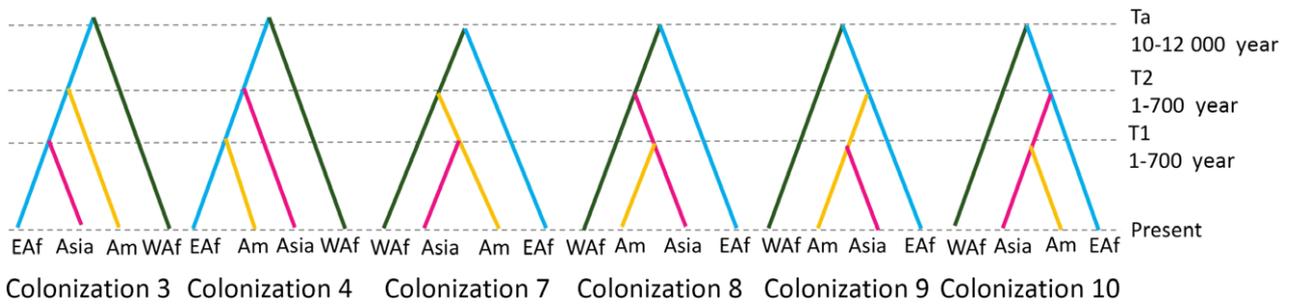


Fig. 3 Highest probability DIYABC scenarios in comparison of 12 models testing the colonization route of *Aedes aegypti* out of Africa, including West Africa (Waf), East Africa (Neafsey *et al.* 2015), Asia (Asia) and the Americas (Am).

Previously, because of the lack of genetic structure seen within African populations, it has been suggested that *Aedes aegypti* have remained panmictic over the course of history (Brown *et al.* 2014). However, ABC analysis for the combined nuclear markers and the mtDNA marker rejected a scenario of constant panmixia in favour of historical divergence, admixture and recent population differentiation. Admixture within ancestral populations obscures the colonization pathway leading to inference of six possible colonization scenarios inferred by different genetic markers. Nonetheless, it is apparent that all populations of *Ae. aegypti* stem from admixed populations of ancestral lineages, which may have played an important role in this mosquito's adaption to human-modified environments and success out of Africa.

#### Evolutionary history of *Aedes aegypti* within Africa

The phylogenetic signal detected by our ABC analysis is consistent with our hypothesis that the population structure of African *Ae. aegypti* has been influenced by historical climate change. The best fit model of evolution for both nuclear and mtDNA genes suggests that populations diverged during recent glacial maxima occurring between 6000 and 120 000 years ago when persistent forest fragments were separated by large expanses of dry savannah (Maley 1991; Walsh *et al.* 1993; Sang & Dunster 2001; Wilcox & Ellis 2006; Castaneda *et al.* 2009; Vasilakis *et al.* 2011). At this time period, gene flow was presumably restricted between populations of *Ae. aegypti* which were subject to genetic drift and ancestral lineage sorting. Unfortunately ABC analysis was unable to distinguish whether *Ae. aegypti* diverged in two or more refugia. This may be because the reality is more complex than our models, for example repeated episodes of divergence and admixture associated with periodic glacial/interglacial cycles and/or persistence in multiple microrefugia, as inferred for African *Begonia* species (Plana *et al.* 2004), within two major refugial areas in East and West/Central Africa. However, we can infer that allopatric divergence of

*Ae. aegypti* has occurred in forest refugia across Africa, although the number of these is as yet unknown.

The ABC analysis indicates that admixture occurred between divergent populations relatively recently during the Holocene 1000 to 12 000 years ago when forests expanded during a wet climate phase (Sang & Dunster 2001; Castaneda *et al.* 2009). This signal of admixture within Africa is also seen in the nuclear loci by low  $F_{ST}$  values between geographic regions and by haplotype networks which reveal that populations of *Ae. aegypti* are composed of mixtures of the genetic diversity present across Africa. The mitochondrial haplotype network also supports admixture. Moore *et al.* (2013) reported two mitochondrial clades that broadly represent East and West Africa. Our mtDNA haplotype network broadly concurs with this but additionally shows that historical east–west divergence has been followed by the migration of individuals from West and Central Africa to East Africa because a cluster of East African haplotypes are derived from the latter. This apparently unidirectional movement of mtDNA from west to east presumably occurred due to migration from large populations of *Ae. aegypti* in the expansive Guineo-Congolian forest block to the smaller areas of forest in East Africa. The problem of NUMTs in *Ae. aegypti* will always cast doubt over the ability of mitochondrial markers to resolve genuine population history (Bensasson *et al.* 2001; Hlaing *et al.* 2009). However, our findings with mtDNA appear to reflect a sensible population history which is compatible with the demographic history revealed by nuclear markers that there has been historical admixture within Africa.

An alternative or additional explanation for the historical phylogeographic pattern seen is that populations were once isolated by a long-term geographic barrier to dispersal, for example the East African rift valley which restricts the dispersal of *Anopheles* mosquitoes (Lehmann *et al.* 1999, 2000, 2003; Braginetts *et al.* 2003). Admixture in the Holocene may then have been promoted by human migration across the continent. Human migrations are known to have occurred during this time with

favourable conditions facilitating migration across the continent, through the Sahara and out of Africa (Hewitt 1996; Castaneda *et al.* 2009). There is scarce archaeological data to identify the early migration routes of humans within Africa. However, one possibility for human-augmented mosquito dispersal is provided by the movement of Bantu farmers across the centre of Africa from Cameroon to Kenya beginning approximately five thousand years ago (Norris *et al.* 2010; Green *et al.* 2013). The Holocene era and migration of Bantu peoples is linked to the rise of agriculture which allowed permanent settlement and presumably water storage (Norris *et al.* 2010; Green *et al.* 2013; Kadu *et al.* 2013). Mosquitoes could have moved in parallel with human populations, transported as larvae in water containers. Alternatively, water storage and domestic habitats could have provided reliable water sources for breeding so that *Ae. aegypti* could expand into previously unsuitable habitat, connect isolated forest habitats and promote admixture. This dispersal could have occurred with or without adaptation to human habitats, although an interesting hypothesis is that the rise of permanent human settlements at this time triggered domestication in *Ae. aegypti*.

After admixture of populations in the early Holocene, recent population divergence between regional groups is indicated by the significant albeit low  $F_{ST}$  values for all loci. This can be seen in the haplotype networks of nuclear genes which show localized geographic structuring at the end branches. More recently, drier climatic conditions have reduced the distribution of forests which have also been subject to human induced deforestation (Maley 1991; Couvreur *et al.* 2008; Wagner *et al.* 2008; Weaver 2014). While the forests of West and Central Africa remain to some extent connected, contact between West and East African forest habitats is limited by unsuitable habitat and the Eastern Rift and Eastern Arc mountain ranges. We also found some evidence for isolation by distance within African *Ae. aegypti*, detected by the nuclear loci with the highest genetic diversity, *Enol-I* and *HSP70*. Therefore recent divergence is at least in part likely to be explained by restricted active dispersal, which is estimated to be <1 km in domestic *Ae. aegypti* (McDonald 1977; Tabachnick & Powell 1978; Tabachnick *et al.* 1979; Trpis & Hausermann 1986; Reiter *et al.* 1995).

#### Colonization of worldwide populations of *Aedes aegypti*

It is reasonable to surmise that introduction out of Africa occurred via the shipping trade of the 17th to early 20th century as dengue epidemics were rife, throughout the Americas from the 1600s and Asia after WW2, and *Ae. aegypti* was found outside Africa for the first time (Gubler 1998a, 2004; Urdaneta-Marquez & Failloux 2011; Brathwaite Dick *et al.* 2012). Significant

$F_{ST}$  values confirm previous reports that populations within Africa are genetically differentiated from those outside Africa (Brown *et al.* 2011, 2014) while haplotype networks for all loci show this is the result of a founding effect. Pantropical populations share the same subset of haplotypes sampled from high African diversity which, in agreement with Brown *et al.* (2014), provides convincing evidence for a single out-of-Africa colonization event. Colonizing populations of pantropical *Ae. aegypti* may therefore stem from a particularly successful source population, known as the invasive bridgehead effect, which often characterizes invasion and represents the most evolutionary parsimonious explanation for geographic spread (Downie 2002; Kolbe *et al.* 2004; Miller *et al.* 2005; Floerl *et al.* 2009; Lombaert *et al.* 2010). If stepping stone colonization did occur between Asia and the Americas, this could have occurred through historical shipping lanes connecting the continents via Europe (Eltis & Richardson 1997; Eltis 2007). Alternatively, or in addition, a direct route for trade in spices and silver linked Manila in the Philippines with Acapulco in Mexico, Panama and Lima in Peru between 1565 and 1815 (Schurz 1917, 1918; Giraldez 2015).

As in Brown *et al.* (2014), our results find evidence for a decline in genetic diversity moving westwards from Africa to the Americas and then to Asia, which could reflect the colonization route of *Ae. aegypti* out of Africa involving successive founding events. However, caution should be taken in inferring this route of colonization because standard diversity measures can be biased by multiple introduction events (Marchette *et al.* 1969; Kolbe *et al.* 2004; Lavergne & Molofsky 2007), selective sweeps (Pastorino *et al.* 2004; Demanou *et al.* 2010) and founding events associated with vector control of pantropical *Ae. aegypti* populations (Huelsenbeck *et al.* 1996; Gubler 1998b; Diallo *et al.* 2005; Paduan & Ribolla 2008; Urdaneta-Marquez & Failloux 2011; Ngoagouni *et al.* 2012). ABC analysis was unable to determine a definitive colonization route possibly because the historical admixture signal acts to obscure subsequent demographic signal; some nuclear loci could reveal a West African origin while others will favour an East African origin because of the stochastic nature of ancestral lineage sorting. A West African origin appears more likely given the high intensity of historical shipping traffic moving from West Africa to the Americas, compared to trade between East Africa and Asia (Cathey & Marr 2014). Further analyses are required to confirm this colonization route using more nuclear markers to increase demographic resolution. A challenge for demographic analysis is that of sampling; inclusion of data from North Africa, a potential source of worldwide colonization (Petersen 1977; Powell & Tabachnick 2013), and

from Brazil, a major port involved in the historical slave trade (Eltis & Richardson 1997; Eltis 2007), could improve selection of the correct model.

We have observed an exceptionally high genetic diversity in African *Ae. aegypti*. As admixture acts to raise genetic diversity (Laverigne & Molofsky 2007; Kolbe *et al.* 2008), this demographic process may explain the high levels of genetic diversity we observe. The fact that we find a lack of evidence for recent admixture involved in the colonization out of Africa supports our finding that admixture was historical and occurred within Africa. Although it is also possible that no particular adaptation to domesticity was required, this admixture could have increased adaptive potential (Facon *et al.* 2006; Laverigne & Molofsky 2007; Kolbe *et al.* 2008; Rius & Darling 2014) and facilitated the ability of *Ae. aegypti* to exploit human environments; admixture may therefore have predisposed *Ae. aegypti* to successful worldwide invasion. Alternatively, or additionally, adaption could have occurred during colonization as rare beneficial mutations can surf to high frequencies in expanding populations and so cause a shift in the adaptive landscape (Demanou *et al.* 2010).

## Conclusion

The population history of *Aedes aegypti* within Africa represents a complex situation of historical divergence, admixture and contemporary restricted gene flow, most likely in response to climate induced forest contraction and expansion. The relative importance of historical climate change and human movement in shaping the genetic diversity of *Ae. aegypti* could be resolved with a comparative phylogeographic approach using other forest-associated *Aedes* species. Historical admixture within Africa has generated high levels of genetic diversity in populations of *Ae. aegypti* throughout Africa. This high level of genetic diversity and the potential for novel allele combinations could have generated a high adaptive potential important for the success of expanding populations in novel environments. The colonization scenarios we explored in this study were simplified while the real demographic history is likely to have involved an array of events. Further analyses taking into account multiple introductions and founder events, while circumventing the difficulty of the pervading genetic signature of historical admixture, will be key to unravelling *Ae. aegypti*'s colonization past.

## Acknowledgements

The authors are grateful to field teams at the Uganda Virus Research Institute, International Institute for Tropical Agriculture, Vector Control Program for the Ministry of Health in

Jamaica, Manatee County Mosquito District, Public Works Division in Miami-Dade County, The Florida Keys Mosquito District, Key West, and Dengue Branch of CDC in Puerto Rico and to Miriam Richards, Luke B. Mitchell and John Morgan for assistance with collection of mosquito samples. This work was supported by the Natural Environment Research Council (NE/H525170/1, NE/1528134/1, NE/J500057/1) and a Royal Society and Leverhulme Trust Africa Award (AA110092). This manuscript was prepared while YML held a National Research Council (NRC) Research Associateship Award at the Walter Reed Army Institute of Research. The material to be published reflects the views of the authors and should not be construed to represent those of the Department of the Army or the Department of Defense. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## References

- Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.
- Allendorf FW, Lundquist LL (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology*, **17**, 24–30.
- Anthony NM, Johnson-Bawe M, Jeffery K *et al.* (2007) The role of Pleistocene refugia and rivers in shaping gorilla genetic diversity in central Africa. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 20432–20436.
- Barnett R, Yamaguchi N, Shapiro B *et al.* (2014) Revealing the maternal demographic history of *Panthera leo* using ancient DNA and a spatially explicit genealogical analysis. *BMC Evolutionary Biology*, **14**, 70.
- Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, **16**, 314–321.
- Bosio CF, Harrington LC, Jones JW *et al.* (2005) Genetic structure of *Aedes aegypti* populations in Thailand using mitochondrial DNA. *American Journal of Tropical Medicine and Hygiene*, **72**, 434–442.
- Bracco JE, Capurro ML, Lourenco-de-Oliveira R, Sallum MA (2007) Genetic variability of *Aedes aegypti* in the Americas using a mitochondrial gene: evidence of multiple introductions. *Memorias do Instituto Oswaldo Cruz*, **102**, 573–580.
- Braginets OP, Minakawa N, Mbogo CM, Yan G (2003) Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. *The American Journal of Tropical Medicine and Hygiene*, **69**, 303–308.
- Brathwaite Dick O, San Martín JL, Montoya RH *et al.* (2012) The history of dengue outbreaks in the Americas. *The American Journal of Tropical Medicine and Hygiene*, **87**, 584–593.
- Brower AV (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius Erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences*, **91**, 6491–6495.
- Brown JE, McBride CS, Johnson P *et al.* (2011) Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. *Proceedings of the Royal Society of London B: Biological Sciences*, **278**, 2446–2454.
- Brown JE, Evans BR, Zheng W *et al.* (2014) Human impacts have shaped historical and recent evolution in *Aedes aegypti*,

- the dengue and yellow fever mosquito. *Evolution*, **68**, 514–525.
- Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics*, **172**, 2665–2681.
- Castaneda IS, Mulitza S, Schefuss E *et al.* (2009) Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proceedings of the National Academy of Sciences*, **106**, 20159–20163.
- Cathey JT, Marr JS (2014) Yellow fever, Asia and the East African slave trade. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **108**, 252–257.
- Chadee DD, Ward RA, Novak RJ (1998) Natural habitats of *Aedes Aegypti* in the Caribbean—a review. *Journal of the American Mosquito Control Association*, **14**, 5–11.
- Clavero M, García-Berthou E (2005) Invasive species are a leading cause of animal extinctions. *Trends in Ecology & Evolution*, **20**, 110.
- Cornuet J-M, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Couvreur T, Chatrou L, Sosef M, Richardson J (2008) Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees. *BMC Biology*, **6**, 54.
- Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.
- Cutter AD (2008) Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Molecular Biology and Evolution*, **25**, 778–786.
- Demanou M, Antonio-Nkondjio C, Ngapana E *et al.* (2010) Chikungunya outbreak in a rural area of Western Cameroon in 2006: a retrospective serological and entomological survey. *BMC Research Notes*, **3**, 128.
- Diallo M, Sall AA, Moncayo AC *et al.* (2005) Potential role of sylvatic and domestic African mosquito species in dengue emergence. *The American Journal of Tropical Medicine and Hygiene*, **73**, 445–449.
- Diamond AW, Hamilton AC (1980) The distribution of forest passerine birds and Quaternary climatic change in tropical Africa. *Journal of Zoology*, **191**, 379–402.
- Downie DA (2002) Locating the sources of an invasive pest, grape phylloxera, using a mitochondrial DNA gene genealogy. *Molecular Ecology*, **11**, 2013–2026.
- Eltis D (2007) *A Brief Overview of the Trans-Atlantic Slave Trade*. Voyages: The Trans-Atlantic Slave Trade Database. <http://www.slavevoyages.org/assessment/essays>.
- Eltis D, Richardson D (1997) West Africa and the transatlantic slave trade: new evidence of long-run trends. *Slavery & Abolition*, **18**, 16–35.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Facon B, Genton BJ, Shykoff J *et al.* (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution*, **21**, 130–135.
- Fjeldså J, Bowie RCK (2008) New perspectives on the origin and diversification of Africa's forest avifauna. *African Journal of Ecology*, **46**, 235–247.
- Fjeldså J, Johansson U, Lokugalappatti LGS, Bowie RK (2007) Diversification of African greenbul in space and time: linking ecological and historical processes. *Journal of Ornithology*, **148**, 359–367.
- Floerl O, Inglis GJ, Dey K, Smith A (2009) The importance of transport hubs in stepping-stone invasions. *Journal of Applied Ecology*, **46**, 37–45.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.
- Fountain T, Duvaux L, Horsburgh G, Reinhardt K, Butlin RK (2014) Human-facilitated metapopulation dynamics in an emerging pest species, *Cimex lectularius*. *Molecular Ecology*, **23**, 1071–1084.
- Garrick RC, Sunnucks P, Dyer RJ (2010) Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, **10**, 118.
- Giraldez A (2015) *The Age of Trade: The Manila Galleons and the Dawn of the Global Economy*. Rowman & Littlefield, Lanham, Maryland.
- Gorrochotegui-Escalante N, Munoz ML, Fernandez-Salas I, Beaty BJ, Black WCT (2000) Genetic isolation by distance among *Aedes aegypti* populations along the northeastern coast of Mexico. *American Journal of Tropical Medicine and Hygiene*, **62**, 200–209.
- Gouck HK (1972) Host preferences of various strains of *Aedes aegypti* and *A. simpsoni* as determined by an olfactometer. *Bulletin of the World Health Organization*, **47**, 680–683.
- Green JMH, Larrosa C, Burgess ND *et al.* (2013) Deforestation in an African biodiversity hotspot: extent, variation and the effectiveness of protected areas. *Biological Conservation*, **164**, 62–72.
- Gubler DJ (1998a) Dengue and dengue hemorrhagic fever. *Clinical Microbiology Reviews*, **11**, 480–496.
- Gubler DJ (1998b) Resurgent vector-borne diseases as a global health problem. *Emerging Infectious Diseases*, **4**, 442–450.
- Gubler DJ (2004) The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? *Comparative Immunology, Microbiology and Infectious Diseases*, **27**, 319–330.
- Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2009) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.
- Hamilton AC (1981) The quaternary history of African forests: its relevance to conservation. *African Journal of Ecology*, **19**, 1–6.
- Harrington LC, Edman JD, Scott TW (2001) Why do female *Aedes aegypti* (Diptera: Culicidae) feed preferentially and frequently on human blood? *Journal of Medical Entomology*, **38**, 411–422.
- Hassanin A, Khouider S, Gembu G-C *et al.* (2015) The comparative phylogeography of fruit bats of the tribe Scotonycterini

- (Chiroptera, Pteropodidae) reveals cryptic species diversity related to African Pleistocene forest refugia. *Comptes Rendus Biologies*, **338**, 197–211.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
- Hewitt G (2004a) The structure of biodiversity – insights from molecular phylogeography. *Frontiers in Zoology*, **1**, 4.
- Hewitt GM (2004b) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**, 183–195.
- Hlaing T, Tun-Lin W, Somboon P *et al.* (2009) Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genetics*, **10**, 11.
- Hlaing T, Tun-Lin W, Somboon P *et al.* (2010) Spatial genetic structure of *Aedes aegypti* mosquitoes in mainland Southeast Asia. *Evolutionary Applications*, **3**, 319–339.
- Huang Y-M (2004) The subgenus *Stegomyia* of *Aedes* in the Afrotropical Region with keys to the species (Diptera: Culicidae). *Zootaxa*, **700**, 1–119.
- Huelsenbeck JP, Bull JJ, Cunningham CW (1996) Combining data in phylogenetic analysis. *Trends in Ecology & Evolution*, **11**, 152–158.
- Huhndorf MH, Kerbis Peterhans JC, Loew SS (2007) Comparative phylogeography of three endemic rodents from the Albertine Rift, east central Africa. *Molecular Ecology*, **16**, 663–674.
- Irvin N, Hoddle MS, O'Brochta DA, Carey B, Atkinson PW (2004) Assessing fitness costs for transgenic *Aedes aegypti* expressing the GFP marker and transposase genes. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 891–896.
- Johnson MB, Clifford SL, Goossens B *et al.* (2007) Complex phylogeographic history of central African forest elephants and its implications for taxonomy. *BMC Evolutionary Biology*, **7**, 244.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Kadu CAC, Konrad H, Schueler S *et al.* (2013) Divergent pattern of nuclear genetic diversity across the range of the afro-montane *Prunus africana* mirrors variable climate of African highlands. *Annals of Botany*, **111**, 47–60.
- Kearse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
- Kebede M, Ehrlich D, Taberlet P, Nemomissa S, Brochmann C (2007) Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopian and Tropical East African mountains. *Molecular Ecology*, **16**, 1233–1243.
- Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology*, **23**, 1720–1731.
- Koffi KG, Hardy OJ, Doumenge C, Cruaud C, Heuertz M (2011) Diversity gradients and phylogeographic patterns in *Santiria trimera* (Burseraceae), a widespread African tree typical of mature rainforests. *American Journal of Botany*, **98**, 254–264.
- Kolbe JJ, Glor RE, Rodriguez Schettino L *et al.* (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.
- Kolbe JJ, Larson A, Losos JB, de Queiroz K (2008) Admixture determines genetic diversity and population differentiation in the biological invasion of a lizard species. *Biology Letters*, **4**, 434–437.
- Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences*, **104**, 3883–3888.
- Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution*, **17**, 386–391.
- Lehmann T, Hawley WA, Grebert H *et al.* (1999) The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. *Journal of Heredity*, **90**, 613–621.
- Lehmann T, Blackston CR, Besansky NJ *et al.* (2000) The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: the mtDNA perspective. *Journal of Heredity*, **91**, 165–168.
- Lehmann T, Licht M, Elissa N *et al.* (2003) Population structure of *Anopheles gambiae* in Africa. *Journal of Heredity*, **94**, 133–147.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Livak KJ (1984) Organization and mapping of a sequence on the *Drosophila melanogaster* X and Y chromosomes that is transcribed during spermatogenesis. *Genetics*, **107**, 611–634.
- Lombaert E, Guillemaud T, Cornuet JM *et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE*, **5**, e9743.
- Lorenzen ED, Arctander P, Siegmund HR (2006) Regional genetic structuring and evolutionary history of the impala *Aepyceros melampus*. *Journal of Heredity*, **97**, 119–132.
- Lounibos LP (2002) Invasions by insect vectors of human disease. *Annual Review of Entomology*, **47**, 233–266.
- Lowe A, Harris D, Dormontt E, Dawson I (2010) Testing putative African tropical forest refugia using chloroplast and nuclear DNA phylogeography. *Tropical Plant Biology*, **3**, 50–58.
- Lucek KAY, Roy D, Bezaul E, Sivasundar A, Seehausen OLE (2010) Hybridization between distant lineages increases adaptive variation during a biological invasion: stickleback in Switzerland. *Molecular Ecology*, **19**, 3995–4011.
- Maley J (1991) The African rain forest vegetation and palaeoenvironments during late quaternary. *Climatic Change*, **19**, 79–98.
- Marchette NJ, Garcia R, Rudnick A (1969) Isolation of Zika virus from *Aedes aegypti* mosquitoes in Malaysia. *The American Journal of Tropical Medicine and Hygiene*, **18**, 411–415.
- Martin DP, Lemey P, Lott M *et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**, 2462–2463.
- Mattingly PF (1957) Genetical aspects of the *Aedes aegypti* problem. I. Taxonomy and bionomics. *Annals of Tropical Medicine and Parasitology*, **51**, 392–408.
- Mattingly PF (1967) Taxonomy of *Aedes aegypti* and related species. *Bulletin of the World Health Organization*, **36**, 552–554.
- McBride CS, Baier F, Omondi AB *et al.* (2014) Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*, **515**, 222–227.
- McClelland GAH (1974) A worldwide survey of variation in scale pattern of the abdominal tergum of *Aedes aegypti* (L.) (Diptera: Culicidae). *Transactions of the Royal Entomological Society of London*, **126**, 239–259.
- McDonald PT (1977) Population characteristics of domestic *Aedes aegypti* (Diptera: culicidae) in villages on the Kenya

- Coast I. Adult survivorship and population size. *Journal of Medical Entomology*, **14**, 42–48.
- Measey GJ, Tolley KA (2011) Sequential fragmentation of pleistocene forests in an east Africa biodiversity hotspot: chameleons as a model to track forest history. *PLoS ONE*, **6**, e26606.
- deMenocal PB (1995) Plio-Pleistocene African climate. *Science*, **270**, 53–59.
- deMenocal PB (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, **220**, 3–24.
- Messing J (1983) New M13 vectors for cloning. *Methods in Enzymology*, **101**, 20–78.
- Miller N, Estoup A, Toepfer S *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science*, **310**, 992–992.
- Moore M, Sylla M, Goss L *et al.* (2013) Dual African origins of global *Aedes aegypti* s.l. populations revealed by mitochondrial DNA. *PLoS Neglected Tropical Diseases*, **7**, e2175.
- Moriyama EN, Gojobori T (1992) Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. *Genetics*, **130**, 855–864.
- Musso D, Cao-Lormeau VM, Gubler DJ (2015) Zika virus: following the path of dengue and chikungunya? *The Lancet*, **386**, 243–244.
- Neafsey DE, Waterhouse RM, Abai MR *et al.* (2015) Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science*, **347**, 1258522.
- Ngoagouni C, Kamgang B, Manirakiza A *et al.* (2012) Entomological profile of yellow fever epidemics in the Central African Republic, 2006–2010. *Parasites & Vectors*, **5**, 175.
- Nicolas V, Bryja J, Akpatou B *et al.* (2008) Comparative phylogeography of two sibling species of forest-dwelling rodent (*Praomys rostratus* and *P. tullbergi*) in West Africa: different reactions to past forest fragmentation. *Molecular Ecology*, **17**, 5118–5134.
- Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.
- Norris K, Asase A, Collen B *et al.* (2010) Biodiversity in a forest-agriculture mosaic – the changing face of West African rainforests. *Biological Conservation*, **143**, 2341–2350.
- Olago DO (2001) Vegetation changes over palaeo-time scales in Africa. *Climate Research*, **17**, 105–121.
- Paduan KDS, Ribolla PEM (2008) Mitochondrial DNA polymorphism and heteroplasmy in populations of *Aedes aegypti* in Brazil. *Journal of Medical Entomology*, **45**, 59–67.
- Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.
- Pastorino B, Muyembe-Tamfum JJ, Bessaud M *et al.* (2004) Epidemic resurgence of chikungunya virus in Democratic Republic of the Congo: identification of a new Central African strain. *Journal of Medical Virology*, **74**, 277–282.
- Paupy C, Brengues C, Kamgang B *et al.* (2008) Gene flow between domestic and sylvan populations of *Aedes aegypti* (Diptera: Culicidae) in North Cameroon. *Journal of Medical Entomology*, **45**, 391–400.
- Paupy C, Brengues C, Ndiath O *et al.* (2010) Morphological and genetic variability within *Aedes aegypti* in Niakhar, Senegal. *Infection, Genetics and Evolution*, **10**, 473–480.
- Paupy C, Le Goff G, Brengues C *et al.* (2012) Genetic structure and phylogeography of *Aedes aegypti*, the dengue and yellow-fever mosquito vector in Bolivia. *Infection Genetics and Evolution*, **12**, 1260–1269.
- Pejchar L, Mooney HA (2009) Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution*, **24**, 497–504.
- Petersen JL (1977) Behavioral differences in two subspecies of *Aedes aegypti* (L.) (Diptera: Culicidae) in East Africa.
- Plana V, Gascoigne A, Forrester LL, Harris D, Pennington RT (2004) Pleistocene and pre-Pleistocene *Begonia* speciation in Africa. *Molecular Phylogenetics and Evolution*, **31**, 449–461.
- Powell JR, Tabachnick WJ (2013) History of domestication and spread of *Aedes aegypti* – a review. *Memorias do Instituto Oswaldo Cruz*, **108**(Suppl 1), 11–17.
- Qu Y, Luo X, Zhang R *et al.* (2011) Lineage diversification and historical demography of a montane bird *Garrulax elliotii* – implications for the Pleistocene evolutionary history of the eastern Himalayas. *BMC Evolutionary Biology*, **11**, 174.
- Qu Y, Zhang R, Quan Q *et al.* (2012) Incomplete lineage sorting or secondary admixture: disentangling historical divergence from recent gene flow in the Vinous-throated parrotbill (*Paradoxornis webbianus*). *Molecular Ecology*, **21**, 6117–6133.
- Ray N, Adams J (2001) A GIS-based vegetation map of the world at the last glacial maximum (25,000–15,000 BP). *Internet Archaeology*, **11**.
- RCoreTeam (2015) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Reiter P, Amador MA, Anderson RA, Clark GG (1995) Short report: dispersal of *Aedes aegypti* in an urban area after blood feeding as demonstrated by rubidium-marked eggs. *American Journal of Tropical Medicine and Hygiene*, **52**, 177–179.
- Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology & Evolution*, **29**, 233–242.
- Roy MS (1997) Recent diversification in African greenbuls (Pycnonotidae: *Andropadus*) supports a montane speciation model. *Proceedings of the Royal Society B: Biological Sciences*, **264**, 1337–1344.
- Sang RC, Dunster LM (2001) The growing threat of arbovirus transmission and outbreaks in Kenya: a review. *East African Medical Journal*, **78**, 655–661.
- Scarpassa VM, Cardoza TB, Cardoso Junior RP (2008) Population genetics and phylogeography of *Aedes aegypti* (Diptera: Culicidae) from Brazil. *American Journal of Tropical Medicine and Hygiene*, **78**, 895–903.
- Schurz WL (1917) The Manila Galleon and California. *The Southwestern Historical Quarterly*, **21**, 107–126.
- Schurz WL (1918) Mexico, Peru, and the Manila Galleon. *The Hispanic American Historical Review*, **1**, 389–402.
- Sezonlin M, Dupas S, Le RÜB *et al.* (2006) Phylogeography and population genetics of the maize stalk borer *Busseola fusca* (Lepidoptera, Noctuidae) in sub-Saharan Africa. *Molecular Ecology*, **15**, 407–420.
- Smith CE (1956) The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. *Journal of Tropical Medicine and Hygiene*, **59**, 243–251.
- Smith TB, Thomassen HA, Freedman AH *et al.* (2011) Patterns of divergence in the olive sunbird *Cyanomitra*

- olivacea (Aves: Nectariniidae) across the African rainforest–savanna ecotone. *Biological Journal of the Linnean Society*, **103**, 821–835.
- Sonsthagen SA, Chesser RT, Bell DA, Dove CJ (2012) Hybridization among Arctic white-headed gulls (*Larus* spp.) obscures the genetic legacy of the Pleistocene. *Ecology and Evolution*, **2**, 1278–1295.
- Sowilem MM, Kamal HA, Khater EI (2013) Life table characteristics of *Aedes aegypti* (Diptera: Culicidae) from Saudi Arabia. *Tropical Biomedicine*, **30**, 301–314.
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.
- Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology*, **17**, 351–360.
- Surendran S, Sarma D, Jude P *et al.* (2013) Molecular characterization and identification of members of the *Anopheles subpictus* complex in Sri Lanka. *Malaria Journal*, **12**, 304.
- Tabachnick WJ (1991) Evolutionary genetics and arthropod-borne disease: the yellow fever mosquito. *American Entomologist*, **37**, 14–26.
- Tabachnick WJ, Powell JR (1978) Genetic structure of the East African domestic populations of *Aedes aegypti*. *Nature*, **272**, 535–537.
- Tabachnick WJ, Munstermann LE, Powell JR (1979) Genetic distinctness of sympatric forms of *Aedes aegypti* in east Africa. *Evolution*, **33**, 287–295.
- Tatem AJ, Hay SI, Rogers DJ (2006) Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences*, **103**, 6242–6247.
- Tosi AJ (2008) Forest monkeys and Pleistocene refugia: a phylogeographic window onto the disjunct distribution of the *Chlorocebus lhoesti* species group. *Zoological Journal of the Linnean Society*, **154**, 408–418.
- Trips M, Hausermann W (1986) Dispersal and other population parameters of *Aedes aegypti* in an African village and their possible significance in epidemiology of vector-borne diseases. *American Journal of Tropical Medicine and Hygiene*, **35**, 1263–1279.
- Urdaneta-Marquez L, Failloux AB (2011) Population genetic structure of *Aedes aegypti*, the principal vector of dengue viruses. *Infection Genetics and Evolution*, **11**, 253–261.
- Vasilakis N, Cardoso J, Hanley KA, Holmes EC, Weaver SC (2011) Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nature Reviews Microbiology*, **9**, 532–541.
- Wagner P, Köhler J, Schmitz A, Böhme W (2008) The biogeographical assignment of a West Kenyan rain forest remnant: further evidence from analysis of its reptile fauna. *Journal of Biogeography*, **35**, 1349–1361.
- Walsh JF, Molyneux DH, Birley MH (1993) Deforestation: effects on vector-borne disease. *Parasitology*, **106**, S55–S75.
- Watrjn J, Lézine A-M, Hély C (2009) Plant migration and plant communities at the time of the “green Sahara”. *Comptes Rendus Geoscience*, **341**, 656–670.
- Weaver SC (2014) Arrival of chikungunya virus in the new world: prospects for spread and impact on public health. *PLoS Neglected Tropical Diseases*, **8**, e2921.
- Weaver SC, Forrester NL (2015) Chikungunya: evolutionary history and recent epidemic spread. *Antiviral Research*, **120**, 32–39.
- Weaver SC, Lecuit M (2015) Chikungunya virus and the global spread of a mosquito-borne disease. *New England Journal of Medicine*, **372**, 1231–1239.
- White VL, Endersby NM, Chan J, Hoffmann AA, Weeks AR (2015) Developing Exon-Primed Intron-Crossing (EPIC) markers for population genetic studies in three *Aedes* disease vectors. *Insect Science*, **22**, 409–423.
- Wilcox BA, Ellis B (2006) Forests and emerging infectious diseases of humans. *UNASYLVA-FAO*, **57**, 11. Available at <http://www.fao.org/docrep/009/a0789e/a0789e03.html>
- Wilson JR, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution*, **24**, 136–144.

---

C.W. and K.L.B. designed the study; K.L.B., F.S., Y-M.L., G.M., M.K., R.D., O.A., A.H., S.I., T.H., A.P., J.L. and C.W. carried out the fieldwork and contributed samples; K.L.B. performed the research and analysed the data under the supervision of C.W.; Y-M.L. contributed to sequencing of the *COI* gene; and K.L.B. wrote the manuscript with help from C.W. All authors contributed with comments and approved the final version.

---

### Data accessibility

DNA sequences: GenBank Accession nos KX444686–KX444699. Sampling locations and/or online-only appendices uploaded as online supplemental material.

### Supporting information

Additional supporting information may be found in the online version of this article.

**Fig. S1** Six evolutionary scenarios for ancestral African *Ae. aegypti*.

**Fig. S2** Twelve evolutionary scenarios to test the colonisation route of *Ae. aegypti* out of Africa without admixture.

**Fig. S3** Six evolutionary scenarios to test the colonisation route of *Ae. aegypti* out of Africa with admixture.

**Fig. S4** Scenarios included in ABC analysis testing whether recent admixture was involved in colonisation out of Africa.

**Fig. S5** Haplotype network for the gene *Enol-I*.

**Fig. S6** Haplotype network for the gene *HSP70*.

**Fig. S7** Haplotype network for the gene *Rp30Lb*.

**Fig. S8** Haplotype network for the gene *IDH2*.

**Fig. S9** Haplotype network for the gene *COI*.

**Table S1** *Ae. aegypti* sample locations and information including the number of individuals sequenced for mitochondrial *COI* and nuclear genes *Rp30Lb*, *IDH2*, *Enol-I* and *HSP70*.

**Table S2** Information on mitochondrial *COI* sequence data in analysis, mined from GenBank.

**Table S3** Prior distributions of demographic and mutation rate parameters for DIYabc analysis of competing evolutionary scenarios for ancestral African populations.

**Table S4** Prior distributions of demographic and mutation rate parameters for DIYabc analysis of competing evolutionary scenarios for colonisation out of Africa.

**Appendix S1** Hypotheses of ancestral population history in Africa.