Cryptic genetic differentiation of the sex-determining chromosome in the mosquito *Aedes aegypti*

Albin Fontaine[1,2,3], Igor Filipović[4], Thanyalak Fansiri[5], Ary A. Hoffmann[4], Gordana Rašić[4,§], Louis Lambrechts[1,3,§]

[1] *Insect-Virus Interactions Group, Department of Genomes and Genetics, Institut Pasteur, Paris, France*

[2] *Equipe Résidente de Recherche d'Infectiologie Tropicale, Division Expertise, Institut de Recherche Biomédicale des Armées, 91223 Brétigny-sur-Orge, France*

[3] *Centre National de la Recherche Scientifique, Unité de Recherche Associée 3012, Paris, France*

[4] *Pest and Environmental Adaptation Research Group, School of BioSciences, The University of Melbourne, Victoria 3010, Australia.*

[5] *Department of Entomology, Armed Forces Research Institute of Medical Sciences, Bangkok 10400, Thailand*

§ Authors for correspondence: Gordana Rašić (gordana.rasic@unimelb.edu.au), Louis Lambrechts (louis.lambrechts@pasteur.fr)

Abstract

The evolution of genetic sex determination in eukaryotes is often accompanied by the morphological and genetic differentiation (heteromorphy) of sex chromosomes. Sex determination systems are of particular interest in insect vectors of human pathogens like mosquitoes, for which novel control strategies aim to convert pathogen-transmitting females into non-biting males, or rely on accurate sexing for the release of sterile males. In the major arbovirus vector, the mosquito *Aedes aegypti*, sex determination is thought to be governed by a dominant male-determining locus (M-locus) spanning only a small portion of an otherwise homomorphic chromosome 1. Here, we provide evidence that the *Ae. aegypti* sex-determining chromosome is differentiated between males and females over a region considerably larger than the M-locus, showing the features of an XY chromosomal system despite the apparent homomorphy. In laboratory F_2 intercrosses, we could not detect recombination events in F_1 males along at least 28% of the physical length of chromosome 1, corresponding to 62% of its cytogenetic length. Sex-specific distortions from the expected genotype ratios in the F_2 progeny were consistent with the XY system and were not found on distal parts of chromosome 1 or on the other two chromosomes. The same chromosomal region showed substantial genetic differentiation between males and females in unrelated wild populations from Australia and Brazil, pointing to the commonality of these chromosomal features in *Ae. aegypti*. Our discovery of cryptic sex-chromosome differentiation in *Ae. aegypti* has important implications for linkage mapping studies, for analyses of population structure, and for the crossing practices to randomize the genetic background of populations in mosquito control strategies.

Key words: *Aedes aegypti*; differentiated sex chromosomes; sex-determining locus; recombination; linkage analysis; RAD markers.

Author summary

Sex is genetically determined in many species, but the genetic mechanisms underlying sex determination can evolve rapidly even among closely related species. Sex determination is of particular interest in mosquito vectors of human pathogens, for which novel control strategies aim to convert pathogen-transmitting females into non-biting males, or rely on accurate sexing for the release of sterile males. Whereas anopheline malaria vectors have fully differentiated XY sex chromosomes, culicine mosquitoes such as the major arbovirus vector *Aedes aegypti* have a homologous pair of sex chromosomes that are morphologically identical. Until now, *Ae. aegypti* sex chromosomes were also considered genetically undifferentiated with the exception of a narrow sex-determining locus. Our study demonstrates substantial genetic divergence of the sex chromosome between males and females over a much larger region than the sex-determining locus, showing the features of an XY chromosomal system despite the similar morphology. This finding has several practical implications for genetic studies of *Ae. aegypti* and calls for investigations of equivalent features of an XY chromosomal system in other culicine mosquito species.

Introduction

Sex is nearly universal to eukaryotic life, yet there is an enormous array of sex-determination mechanisms even between closely related species, suggesting their potential to evolve rapidly [1-3]. In many taxa where sex is genetically determined, sex-determining genes co-segregate with an entire chromosome that often evolves into morphologically and genetically distinct (heteromorphic) sex chromosomes [2,4,5]. Molecular and cytological analyses have revealed diverse evolutionary histories and stages of sex chromosomes across taxonomic groups, but our understanding of the evolutionary forces driving such a diversity of sex-determination systems remains limited [6]. Flies and mosquitoes (order Diptera) represent one of the most studied invertebrate groups, where frequent transitions in sex-chromosome structure and identity have been documented [7,8]. The underlying mechanisms and evolutionary dynamics of sex determination are of particular interest in mosquitoes given that only females take blood meals and transmit human pathogens such as malaria parasites and dengue viruses. For instance, some novel strategies for controlling mosquito-borne diseases aim to convert pathogen-transmitting females into non-biting males [9], or rely on accurate sexing for the release of sterile males [10,11].

Aedes and *Culex* mosquitoes in the Culicinae subfamily have homomorphic sex-determining chromosomes, which is considered the ancestral state of this character in the mosquito family [12]. *Anopheles* mosquitoes in the Anophelinae subfamily, however, have acquired fully differentiated X and Y sex chromosomes [8,12]. Why heteromorphy of sex chromosomes evolved in some mosquito lineages but not others remains unclear. Evolutionary models suggest progression of autosomes into heteromorphic sex chromosomes after the acquisition of a sex-determining locus [13,14]. The selective advantage of linkage between sex-determining genes and sexually antagonistic genes promotes initial suppression of recombination between homologous chromosomes, followed by expansion of the non-

89 recombining region [15]. An evolving pair of neo-sex chromosomes further differentiates
90 through changes in gene content, gene decay and epigenetic modifications [4]. Yet, recent
91 analyses of fly genomes revealed complex evolutionary trajectories where sex chromosomes
92 have been gained, lost, replaced and rearranged multiple times over the Dipteran evolutionary
93 history [7].

94 *Aedes aegypti* is the main vector of dengue, yellow fever and Zika viruses worldwide.
95 It has a large genome of 1.39 million base pairs (Mbp) and, like other mosquitoes of the
96 Culicinae subfamily, does not have morphologically distinct sex chromosomes [16]. Sex
97 determination in *Ae. aegypti* is controlled by a locus of the smallest chromosome
98 (chromosome 1) called the M-locus, with a dominant (M) allele conferring the male
99 phenotype [17]. A male-specific gene named *nix* and behaving as an M-factor was recently
100 identified within the M-locus [9], along with other male-biased (M-linked) sequences that are
101 present almost exclusively in the male genome [9,18]. The sex-determination locus resides in
102 a non-recombining region mapped to band 1q21 on the q arm of chromosome 1 [19], outside
103 which recombination is thought to occur in an autosome-like fashion, maintaining the overall
104 homomorphic structure of this chromosome [18] . Motara and Rai [20] proposed a
105 nomenclature to define chromosome 1 with the M-locus as the M-chromosome, and the copy
106 without the M-locus as the m-chromosome. Interestingly, they noticed some cytological
107 differences consistent with clear differentiation between chromosomal regions where the
108 M-locus and the m-locus are located [20].

109 Here, we provide genetic evidence that chromosome 1 in *Ae. aegypti* is differentiated
110 over a region much larger than the M-locus, showing features of an XY chromosomal system
111 despite the apparent homomorphy. Specifically, while carrying out linkage mapping
112 intercrosses to locate quantitative trait loci (QTL) of dengue vector competence, we could not
113 detect recombination events in F_1 males along at least 38% of the physical length of

chromosome 1. Sex-specific distortions from the expected genotype ratios in the F_2 progeny were consistent with the XY system and were not found on distal parts of chromosome 1 or on the other two chromosomes. Analyses of genomic variation in unrelated wild *Ae. aegypti* populations provided further evidence for substantial differentiation between the M- and m-chromosomes. These findings have important implications for the mapping and population genetic analyses of *Ae. aegypti*, as well as for some strategies of mosquito-borne disease prevention.

Results

Intercrosses reveal low recombination along a large region of chromosome 1 in males

We initially delineated a genomic region of reduced recombination between the M- and m-chromosomes in male meiosis using a smaller F_2 intercross, referred to as Cross #1 hereafter. Twenty-two F_2 males and 22 F_2 females from an isofemale line originating from Kamphaeng Phet, Thailand, were genotyped using double-digest RAD sequencing. Cross #1 generated a total of 347,080 single nucleotide polymorphism (SNP) markers, of which 1,010 were selected for linkage analysis based on the following criteria: (*i*) full informativeness (i.e., F_0 founders homozygous for alternative alleles), and (*ii*) sequencing depth $\geq$12×.

We ordered a total of 102 markers unambiguously mapped to a single position on a linkage map generated *de novo* from Cross #1 with both sexes combined, spanning 372.2 centiMorgans (cM). The Cross #1 linkage map consisted of three linkage groups of 25, 36 and 41 markers, covering 83.4, 209.7 and 79.0 cM for chromosomes 1, 2 and 3 respectively (Table 1). The average spacing between markers on chromosomes 1, 2 and 3 was 3.5, 6 and 2 cM, respectively. Linkage group assignments of supercontigs were generally in agreement with a previously published chromosome map [19,21] (Figure 1A). A total of ten

138 supercontigs (9.8%) were assigned to different chromosomes between the Cross #1 linkage
139 map and the published chromosome map, and a few discrepancies in the linear order of
140 supercontigs within the same chromosome were detected.

141 We analyzed deviations of genotype frequencies from the expected Mendelian ratios
142 (i.e., segregation distortion) in F_2 individuals using a χ^2 test. In Cross #1, we observed
143 significant deviations from the expected 1:2:1 segregation ratio for sections of chromosomes
144 1 and 2, but only chromosome 1 contained a set of markers with a different pattern of
145 segregation distortion in males *versus* females. Specifically, on chromosome 1 we detected an
146 absence of AA genotypes in F_2 females that was mirrored by an absence of BB genotypes in
147 F_2 males (Figure 1A). This sex-specific segregation distortion is expected in the vicinity of
148 the sex-determining locus because sex-linked markers co-segregate with the M-locus during
149 meiosis in F_1 males. For such sex-linked markers, F_0 paternal A alleles preferentially
150 segregate in F_2 males (Supplementary figure 1). We detected a complete absence of AA
151 genotypes (i.e., analogous to the lack of XY pairs) in all 22 F_2 females at 6 markers spanning
152 4.6 cM, and an absence of BB genotypes (i.e., analogous to the lack of XX pairs) in all 22 F_2
153 males at 13 markers spanning 25.8 cM on chromosome 1. Based on the Cross #1 linkage
154 map, markers in the region showing reduced recombination in males spanned over 12.0 Mbp,
155 representing 28% of the cumulative length of all supercontigs assigned to chromosome 1.

156 Because the probability to detect low-frequency recombinants increases with larger
157 sample size, we further analyzed 197 F_2 females from an independent F_2 intercross, referred
158 to as Cross #2 hereafter, which was initially carried out to locate QTL for dengue vector
159 competence. There was a high degree of synteny between the linkage maps generated from
160 Cross #1 and from Cross #2 (Supplementary figure 2). Misassemblies identified with both
161 linkage maps are reported in Supplementary files 1 and 2. In Cross #2 again, we recorded a
162 complete absence of AA genotypes in all 197 F_2 females at 12 markers spanning 36.8 cM on

chromosome 1 (Figure 1B). Cross #2 assigned 10 additional supercontigs representing 16.8 Mbp to the region of reduced recombination in males compared to Cross #1 (Supplementary file 3). Based on the synteny with the published chromosome map [19], the estimated size of the region showing reduced recombination in males in both Cross #1 and Cross #2 (highlighted in yellow in Figure 1) corresponded to 62% of the cytological length of chromosome 1.

We performed QTL mapping to confirm that the chromosome 1 region showing reduced recombination in males contained the M-locus. Using standard interval mapping based on the Cross #1 linkage map and a binary trait model, we found a major QTL associated with sex on chromosome 1 (Figure 1A). The highest logarithm of odds (LOD) score for this QTL was 7.6 at 56.4 cM with a 1.5 LOD support interval spanning from 54.1 to 69.2 cM. The genome-wide LOD threshold for statistical significance ($\alpha = 0.05$) calculated from 1,000 permutation tests was 3.10. The identified sex QTL accounted for 79.7% of phenotypic variation.

We also estimated local recombination rates by comparing the Cross #1 linkage map with the physical map of the genome, using the cumulative length of supercontigs that contained uniquely mapped markers as a proxy for physical distance. Even though this method underestimates physical distances and results in overestimation of recombination rates [22], it is suitable for comparisons of local recombination rates within and between chromosomes. The estimated recombination rate varied across the genome with marked recombination cold spots on all three chromosomes (Supplementary figure 3). Notably, local recombination rates were most variable on chromosome 1, ranging from 0.04 cM/Mbp to 8.43 cM/Mbp. On this chromosome, 14.6 Mbp of genomic sequence (38% of the chromosome physical length) were mapped to a 5.7 cM region (only 7% of the chromosome genetic length) that spanned between 54 and 60 cM. Local recombination rates for

188 chromosomes 2 and 3 varied from 1.56 to 7.71 cM/Mbp and 0.26 to 4.57 cM/Mbp,
189 respectively, showing more consistent relationships between physical and genetic distances.

190

191 **M- and m-chromosomes are genetically differentiated in wild *Ae. aegypti* populations**

192 Given that rare recombination events between the M- and m-chromosomes in male
193 meiosis cannot be detected unless mapping families are very large, sequences from natural
194 populations can be used to infer historical recombination events and their potential
195 consequences. We characterized a recombination-dependent metric (linkage disequilibrium
196 [LD]), genetic divergence, and heterozygosity in field-caught mosquitoes from a large,
197 panmictic population [23] in Rio de Janeiro, Brazil. Earlier studies have also shown that this
198 population is highly genetically divergent from *Ae. aegypti* found in Southeast Asia and
199 Australia [23,24]. We genotyped 69 males and 33 females at double-digest RAD tags,
200 providing genome-wide sequence data across 69 M-chromosomes, 135 m-chromosomes, and
201 204 copies of each autosome (chromosomes 2 and 3).

202 RAD tags unambiguously mapped to a single genomic position were selected if they
203 were (*i*) polymorphic in at least one sex (minor allele frequency [MAF] ≥5%), (*ii*) shared
204 among ≥70% of all individuals and (*iii*) sequenced at a depth ≥5×. To order the RAD
205 markers, we used a previously published genetic map [25] that allowed more supercontigs to
206 be positioned than with the maps generated from Cross #1 and Cross #2. Based on this
207 linkage map, we ordered 713 RAD markers on chromosome 1, 703 markers on chromosome
208 2, and 1,330 markers on chromosome 3 (Supplementary file 4). Median distance between
209 markers was 50 kbp on chromosomes 1 and 3, and 100 kbp on the largest chromosome 2.

210 We analyzed the pattern of LD between RAD markers for each sex and chromosome
211 pair separately. LD estimates (*ALD*, see Materials and Methods) indicated that strong allelic

association (r >0.5) does not generally extend beyond 1 Mbp. Males displayed slightly elevated LD along 5 Mbp in the approximate centromeric region of chromosome 1 (Figure 2), but it is difficult determine the exact extent of local LD due to the scarcity of RAD markers along a chromosome. Overall, the observed patterns suggest that some degree of recombination between the M- and m-chromosomes occurs in this *Ae. aegypti* population.

Frequency of heterozygotes (H) was higher in males than in females on chromosome 1 (H_{males}=0.291, $H_{females}$=0.270, one-tailed t_{1424}=1.908, p=0.028), but not on chromosome 2 (H_{males}=0.270, $H_{females}$=0.277, one-tailed t_{1408}=0.692, p=0.755), or on chromosome 3 (H_{males}=0.271, $H_{females}$=0.271, one-tailed t_{2655}=0.083, p=0.533). The fixation index F_{ST} is a common measure of divergence in allelic frequencies between populations [26]. F_{ST} between males and females is expected to have a maximum value of 0.5 for fully sex-linked markers where M (analogous to Y) and m (analogous to X) chromosome regions have been fixed for different alleles. Again, only chromosome 1 showed elevated F_{ST} values across a pericentric 50-Mbp region of chromosome 1 (Figure 2). Eleven supercontigs within this region have been previously physically mapped to chromosome 1 bands 1p25, 1q11, 1q13-14, 1q31-33, which are well outside the M-locus position (1q21).

To assess if genomic composition of M- and m-chromosomes is sufficiently differentiated to predict phenotypic sex in *Ae. aegypti*, we applied a multivariate clustering method (discriminant analysis of principal components [DAPC]) [27]. In addition to the sample from Brazil, this analysis included a smaller sample of 41 field-caught adult females and 15 adult males from Gordonvale, Australia. DAPC was performed separately for each geographic sample and chromosome. Individuals were assigned to two groups based on a discriminant analysis of five retained PCs in order to avoid model overfitting [28]. Genetic variation along chromosome 1 was sufficient to identifyassign most individuals to their correct sex: 89% (91/102) of individuals from Brazil and 95% (53/56) of individuals from

237 Australia were correctly identified as males or females. Conversely, genetic separation based
238 on variation on chromosomes 2 and 3 was no better than random (33% and 55% accuracy for
239 chromosome 2, 44% and 64% for chromosome 3 in the Brazilian and Australian samples,
240 respectively). The frequency distribution of individual mosquitoes according to their
241 individual DAPC scores demonstrates clear separation of females and males based only on
242 chromosome 1 variation (Figure 3).

243 Given that DAPC finds linear combinations of alleles (i.e., discriminant functions)
244 which best separate the clusters [27], sex-linked markers can be identified as those with the
245 highest contribution to discrimination between males and females. In agreement with an XY
246 sex-determination system where one sex is expected to be heterozygous, 30 RAD markers
247 with the highest contribution to the discriminant function were heterozygous (analogous to
248 XY) in (nearly) all males and homozygous (analogous to XX) in all females of both
249 populations. Importantly, these markers are located on supercontigs that have been mapped
250 outside of the M-locus chromosomal region (Supplementary file 4). Variants were annotated
251 as being in the intergenic, downstream and intron sequences (with modifier effects) rather
252 than in the coding sequences (Supplementary file 4).

253

254 ## Discussion

255 We provide evidence that the sex-determining chromosome in *Ae. aegypti* is
256 sufficiently genetically differentiated to show features of an XY chromosomal system despite
257 apparent homomorphy. Results from the laboratory crosses and unrelated wild populations
258 point to the commonality of these chromosomal features in *Ae. aegypti* from the New World
259 and Austral-Asia. Our findings challenge the traditional view that the sex-determining

chromosome in *Ae. aegypti* behaves like an autosome outside a small, non-recombining sex-determining region (i.e., the M-locus).

Synteny analyses between *Ae. aegypti* and the malaria mosquito *Anopheles gambiae* revealed a complex evolutionary history of their sex-determining chromosomes [21]. Even though both species have the same number of chromosomes (2n=6), *An. gambiae* has clearly distinguishable heteromorphic (X and Y) sex chromosomes that contain large amounts of heterochromatin [21]. In *Ae. aegypti*, chromosome rearrangements have substantially reshuffled the genetic material between 1p and 1q arms, but there are still notable homologies between the 1p arm of *Ae. aegypti* and the X chromosome of *An. gambiae* [21]. The two mosquito lineages diverged around 145-200 million years ago [29] and their common ancestor most likely had homomorphic sex chromosomes [12]. This suggests that the sex-chromosome homomorphy in *Ae. aegypti* is old and has been maintained through ongoing recombination [18].

However, in our linkage mapping intercrosses, recombination in males was undetectable across a large region of the sex-determining chromosome. Synteny analysis of supercontigs mapped to this region indicated that they were distributed between bands 1p22 and 1q41 of the chromosome map, representing 62% of the total ideogram length (Figure 1). This lack of recombination in males was a hurdle for generating a high-resolution genetic map for QTL mapping of dengue vector competence. In a recent study, Juneja and colleagues also reported low recombination across a large fraction of the chromosome 1 in their mapping female population, which prevented identification of a causal variant for resistance to a nematode parasite [25]. From a practical perspective, reduced recombination along chromosome 1 in males may limit the ability to identify genomic regions affecting phenotypic traits of interest through QTL mapping studies. Rather than increasing the number

of crossing generations in mapping families, genome-wide association studies could be more fruitful when narrowing down genomic regions of interest for a particular trait.

Mapping families provide a contemporary measure of recombination rate, but they do not necessarily reflect historical recombination events. For example, recombination between undifferentiated sex chromosomes in male tree frogs (*Hyla spp.*) was only evident from an analysis of molecular data from populations when coupled with simulations [30]. Our population analyses of chromosome-wide LD in *Ae. aegypti* suggested that haplotype blocks on chromosome 1 in males were not longer when compared to females, in agreement with data from the other two chromosomes. However, the low density of RAD markers along chromosomes meant that we could only examine statistical associations between alleles tens of kbp (or more) apart. The absence of haplotype blocks extending beyond 1-5 Mbp is indicative of some degree of ongoing recombination between M- and m-chromosomes. This is consistent with rare recombination events in the vicinity of the M-locus detected by Hall and colleagues who screened several thousands of individuals using a sensitive transgene-assisted approach [18]. A low recombination rate in males could be sufficient to prevent long-range LD on the *Ae. aegypti* M-chromosome. In the hylid frogs, X and Y chromosomes remain indistinguishable even though recombination is estimated to occur in only 1 in 10^5 males [30]. Simulation work by Grossen and colleagues showed that recombination rates of 10^{-4} could keep sex chromosomes homomorphic [31]. If we assume this rate of recombination, a large *Ae. aegypti* population like the one in Rio de Janeiro [23] would likely have tens of males with recombinant sperm in every generation.

While sex chromosomes in *Ae. aegypti* are morphologically undifferentiated, they were nevertheless genetically differentiated between sexes in our wild population samples. High F_{ST} values between wild males and females from the same population spanned a pericentromeric 50-Mbp region from band 1p25 to band 1q33 (Figure 2), which matches the

309 region of reduced recombination delineated in our laboratory crosses (Figure 1). Furthermore,
310 genotypes at the most differentiated (i.e., putative sex-linked) markers were in agreement
311 with the XY sex-determination system, with homozygous (i.e., analogous to XX) females and
312 heterozygous (i.e., analogous to XY) males.

313 Our analyses give conservative estimates of sequence differentiation between sex
314 chromosomes in *Ae. aegypti* because the datasets consisted of RAD tags found in both sexes;
315 any male-specific sequences without gametologs (i.e., homologous sequences on the non-
316 recombining opposite sex chromosome) were not considered. Male-specific sequences were
317 previously identified in the Liverpool strain of *Ae. aegypti* (used to generate the reference
318 genome sequence) as largely missing from the current genome assembly [9,18]. Male-
319 specific sequences were detected in our double-digest RAD sequencing datasets from wild
320 populations (Supplementary file 5). We created reference-free *de novo* RAD tags from
321 sequences that failed to align to the female-biased genome assembly, and retained tags that
322 were only present in males. We found 185 putative male-specific RAD tags that aligned to
323 the *Ae. aegypti* sequences in the NCBI database with a BLAST *E*-value $<10^{-5}$
324 (Supplementary file 5). Moreover, one RAD tag was identified as the *myo-sex* sequence, an
325 extremely male-biased gene linked to the M-locus in the *Ae. aegypti* Liverpool strain [18].
326 Our results support Y-like sequences in *Ae. aegypti* males not included in the current genome
327 assembly. A long-read sequencing technology was recently used to assemble the repeat-rich
328 Y chromosome of *Anopheles* mosquitoes [32]. The same approach could be used to identify
329 Y-like sequences and incorporate them into an improved assembly of *Ae. aegypti*.

330 Overall, our findings show that a large section of chromosome 1 in *Ae. aegypti*
331 displays features of genetically but not morphologically differentiated X and Y
332 chromosomes. These features could mean that M- and m- chromosomes represent neo-Y and
333 neo-X chromosomes on their path to morphological differentiation, where cessation of

334 recombination helps to physically separate sexually antagonistic alleles [15]. Alternatively,
335 chromosomal features in *Ae. aegypti* could be old, and sexually antagonistic selection may
336 have been resolved by evolving sexually-biased gene expression instead of eliminating
337 recombination [33]. Perhaps mosquitoes, like birds [33], have found different evolutionary
338 solutions to deal with deleterious effects of sexually antagonistic mutations. Some lineages
339 may have maintained homomorphic sex chromosomes (e.g., *Ae. aegypti* and other Culicinae),
340 while others evolved heteromorphic sex chromosomes (e.g., *Anopheles* mosquitoes).

341 The results have implications not only for genetic mapping studies but also for *Ae.*
342 *aegypti* population genetics. Sex-specific differentiation over a large region of the genome
343 needs to be considered when using genetic markers for assessing population genetic structure.
344 To date, genetic analyses of *Aedes* populations have proven challenging due to problems
345 associated with null alleles and other factors that cause deviations from Hardy-Weinberg
346 equilibrium [e.g. 34], and some of these issues may stem from markers located within the
347 XY-like region of chromosome 1. Sexes should therefore always be clearly distinguished in
348 such studies and the chromosomal location of markers should be established. Where sex
349 separation based on morphological characters is difficult (e.g., in immature stages or
350 damaged material), molecular sex-specific markers can be used (e.g., Supplementary file 5).
351 Consideration of the XY chromosomal features in *Ae. aegypi* is also warranted for the
352 development of vector-control strategies such as the field deployment of *Wolbachia*-infected
353 mosquitoes [35] where the release stocks undergo several generations of backcrossing. The
354 backcrossing is done to create favorable combinations of alleles that facilitate artificial
355 rearing and increase fitness of the release stock under natural conditions. Because *Wolbachia*
356 is transmitted maternally and causes cytoplasmic incompatibility, *Wolbachia*-infected
357 females are crossed to males from a target field population [36]. Low recombination in male

meiosis means that males from the release colony are expected to maintain the background of the natural population along a significant portion of the M-chromosome.

In conclusion, our discovery of reduced recombination and genetic differentiation between otherwise homomorphic sex chromosomes of *Ae. aegypti* calls for investigations of similar features of an XY chromosomal system in other species of the Culicinae subfamily. For instance, the M-locus of *Culex pipiens* has a common origin with that of *Ae. aegypti* [37]. Elucidating the evolutionary history of sex-determining chromosomes in the Culicinae subfamily would help to determine whether the XY-like features we observed in *Ae. aegypti* have been maintained over long evolutionary times, or whether they indicate that M- and m-chromosomes are nascent X and Y heteromorphic chromosomes. Finally, a thorough understanding of *Ae. aegypti* sex determination will also require an improved assembly of the genome sequence that incorporates male-biased and male-specific sequences.

Materials and Methods

Mosquito collection and crosses

This study used two independent laboratory crosses of wild-type *Ae. aegypti* mosquitoes originally collected in February 2011 from Kamphaeng Phet, Thailand. Cross #2 was an F_2 intercross between a pair of field-collected mosquito founders collected as eggs using ovitraps as previously described [38]. Briefly, F_0 eggs were allowed to hatch in filtered tap water and pupae emerged in individual vials. *Aedes aegypti* adults were identified by visual inspection and maintained in an insectary under controlled conditions (28±1°C, 75±5% relative humidity and 12:12 hour light-dark cycle). The male and female of each mating pair were chosen from different collection sites to avoid that F_0 parents were siblings from the same wild mother [39,40]. Virgin F_0 adults were allowed to mate for 2-3 days following

382 emergence and the inseminated female was blood fed and allowed to lay eggs. Egg batches
383 from the same female were merged to obtain a pool of F_1 eggs. F_0 founders were saved for
384 later DNA extraction and genotyping. F_2 progeny was produced by mass sib-mating and
385 collective oviposition of the F_1 offspring. A total of 197 female individuals of the F_2 progeny
386 were used as a mapping population to generate a linkage map.

387 Cross #1 was an F_2 intercross between a pair of mosquitoes from two different
388 isofemale lines derived from the same wild *Ae. aegypti* population in Kamphaeng Phet,
389 Thailand. Both isofemale lines were established in February 2011 from F_0 founders as
390 described above and maintained in the laboratory by mass sib-mating and collective
391 oviposition until the 19th generation. At each generation, females were fed on commercial
392 sheep or rabbit blood through an artificial membrane feeding system. Institutional Animal
393 Care and Use Committee approval was not required because blood collection took place
394 postmortem as a by-product of a commercial enterprise. Mosquito eggs were collected and
395 stored on dry pieces of paper towel and maintained under high relative humidity no longer
396 than 6 months. Cross #1 resulted from mating between a single virgin male from one
397 isofemale line with a single virgin female from another isofemale line. Maintenance of an
398 isofemale line at a small population size in the laboratory for 19 generations is expected to
399 lead to inbreeding and maximize homozygosity, one important criterion to generate
400 informative markers for linkage mapping. A total of 22 males and 22 females from the F_2
401 progeny of Cross #1 were used to generate a linkage map and subsequently map the male-
402 determining locus (M-locus).

403 Independent samples were analyzed from two wild *Ae. aegypti* populations from Rio
404 de Janeiro, Brazil and Queensland, Australia. Samples of 41 adult females and 15 adult males
405 from Australia were caught using Biogents sentinel traps set up in Gordonvale, Queensland in
406 December 2010. Adult *Ae. aegypti* were identified as males or females based on the sexually

dimorphic antennae and external genitalia structure [41]. Mosquitoes from Rio de Janeiro, Brazil were collected from ovitraps within a single 3-week period in November-December 2011. Larvae were reared until the third instar in an insectary under controlled conditions (25±1°C, 80±10% relative humidity and 12:12 hour light-dark cycle). Only one individual per ovitrap was retained to avoid analyzing siblings that tend to co-occur within the same trap [39,40]. Sex of each individual was determined based on the presence or absence of the *myo-sex* sequence [18] and confirmed with two additional male-specific sequences identified in this study (Supplementary file 5; Supplementary figure 4). The final dataset from Brazil consisted of 69 males and 33 females.

Double digest Restriction-site Associated DNA (RAD) sequencing

Mosquito genomic DNA was extracted using the NucleoSpin 96 Tissue Core Kit (Macherey-Nagel, Düren, Germany) and whole genome amplified by Multiple Displacement Amplification using the Repli-g Mini kit (Qiagen, Hilden, Germany) to obtain a sufficient amount of DNA. All DNA concentrations were measured with Qubit fluorometer and Quant-iT dsDNA Assay kit (Life technologies, Paisley, UK). An adaptation of the original double-digest Restriction-site Associated DNA (ddRAD) sequencing protocol [42] was used as previously described [24] with minor additional modifications. Briefly, a standardized quantity of 500 ng of genomic DNA from each mosquito was digested in a 50 µl reaction containing 50 units each of *NlaIII* and *MluCI* restriction enzymes (New England Biolabs, Herts, UK), 1× CutSmart® Buffer and water for 3 hours at 37°C, without a heat-kill step. The digestion products were cleaned with 1.5× volume of Ampure XP paramagnetic beads (Beckman Coulter, Brea, CA, USA) and ligated to the modified Illumina P1 and P2 adapters with overhangs complementary to *NlaIII* and *MluCI* cutting sites, respectively. Each mosquito was uniquely labeled with a combination of P1 and P2 barcodes of variable lengths

to increase library diversity at 5' and 3' ends (Supplementary file 6). This method allows the multiplexing of up to 60 mosquitoes using 12 P1 and 5 P2 adapters. Ligation reactions were set up in a 45 µl volume with 2 µl of 4 µM P1 and 12 µM P2 adapters, 1,000 units of T4 ligase and 1× T4 buffer (New England Biolabs) and were incubated at 16°C overnight. Ligations were heat-inactivated at 65°C for 10 minutes and cooled down to room temperature in a thermocycler at a rate of 1.5°C per 2 minutes. Adapter ligated DNA fragments from all individuals were then pooled and cleaned with 1.5× bead solution. Size selection of fragments between 350–440 base pairs (bp) for the crosses or 300-450 bp for the field population was performed using a Pippin-Prep 2% gel cassette (Sage Sciences, Beverly, MA, USA). Finally, 1 µl of the size selected DNA was used as a template in a 10 µl PCR reaction with 5 µl of Phusion High Fidelity 2× Master mix (New England Biolabs) and 1 µl of 50 µM P1 and P2 primers. To reduce PCR duplicates bias, 8 PCR reactions were run in parallel, pooled, and cleaned with a 0.8× bead solution to make the final library. If sequence reads for particular loci are enriched due to PCR duplication in a single PCR reaction, their final over-representation in the library would be limited with the pool of 8 different PCR reactions. At this step, final libraries were quantified by quantitative PCR using the QPCR NGS Library Quantification Kit (Agilent technologies, Palo Alto, CA, USA). For the mapping crosses, libraries containing multiplexed DNA fragments from 48 to 50 mosquitoes were sequenced on an Illumina NextSeq platform using a NextSeq 500 High Output 300 cycles v1 kit (Illumina, San Diego, CA, USA) to obtain 150-bp paired-end reads. An optimized final library concentration of 1.1 pM, spiked with 15% PhiX, was loaded onto the flow cell. For the field populations, three ddRAD libraries each containing 52-56 mosquitoes were sequenced in three lanes of the Illumina HiSeq platform with a 100-bp paired-end chemistry (SRA# pending).

Sequence processing and SNP calling

A previously developed bash script pipeline [24] was used to process raw sequence reads with minor modifications. Briefly, the DDemux program was used for demultiplexing fastq files according to the P1 and P2 barcodes combinations. Sequence quality scores were automatically converted into Sanger format. Sequences were filtered with FASTX-Toolkit. Reads were trimmed to 90 bp (HiSeq) and 140 bp (NextSeq) on both P1 and P2 ends. In addition, we trimmed the first 4 bp of P2 NextSeq reads to decrease the probability of sequencing errors in the low diversity *MluCI* cutting site. A higher sequencing error rate was initially detected in this AATT sequence compared to other parts of the sequence reads. This could be due to a lower accuracy of the two-channel SBS technology to distinguish A from T nucleotides in a region of very low diversity (a known weakness of the NextSeq 500 v1 kits). All reads with Phred scores <25 were discarded. P1 and P2 reads were then matched and unpaired reads were sorted as orphans.

Paired reads were aligned to the *Ae. aegypti* genome (AaegL1, February 2013) [16] using Bowtie version 0.12.7 [43]. Parameters for the ungapped alignment included a maximum of three mismatches allowed in the seed, suppression of alignments if more than one reportable alignment exists, and a "try-hard" option to find valid alignments. Orphans were joined with all unaligned paired-reads and single-end alignment was attempted. All aligned Bowtie output files were merged per individual and were imported into the Stacks pipeline. A catalogue of RAD loci used for single nucleotide polymorphism (SNP) discovery was created using the ref_map.pl pipeline in Stacks version 1.19 [44,45]. First, sequences aligned to the same genomic location were stacked together and merged to form loci using Pstacks. Only loci with a sequencing depth ≥5 reads per individual were retained. Cstacks was then used to create a catalog of consensus loci, merging alleles together and Sstacks was used to match all identified loci. For the mapping crosses, we used the "genotypes" module to

export F_2 mosquito genotypes for all markers homozygous for alternative alleles in the F_0 parents (i.e., AA in the F_0 male and BB in the F_0 female) with a sequencing depth ≥12× in ≥60% of the mapping population. The automated correction option was enabled to correct false-negative heterozygote alleles.

In samples from the wild populations, we selected polymorphic RAD tags (MAF ≥1%) shared among ≥70% of individuals. Where present, multiple SNPs per RAD tag were collated into mini-haplotypes with ≥2 "haplo-alleles" (ver. Stacks 1.35). This was done to retain information about the variation within a 90-bp sequence, but to avoid treating such nearby SNPs as independent variants in downstream analyses. Therefore, our data set consisted of bi-allelic SNPs (i.e., extracted from tags that contained only 1 SNP), and multiallelic markers (i.e., "haplo-alleles" extracted from tags that contained more than 1 SNP). The final number of markers was 2,748 in the sample from Brazil (Chr1 = 713, Chr2 = 705, Chr3 = 1,330), and 1,183 in the sample from Australia (Chr1 = 295, Chr2 = 326, Chr3 = 562).

Linkage map construction

OneMap v2.0-3 [46], implemented as a package in the R environment [47], was used to construct linkage maps based on recombination fractions among RAD markers in the mapping populations. Based on the selection criteria described above, every marker is expected to segregate at a frequency of 25% for homozygotes (i.e., AA and BB genotypes) and of 50% for heterozygotes (i.e., AB genotypes) in an F_2 mapping population that includes both males and females. When only females are genotyped in the F_2 progeny (i.e., Cross #2), fully sex-linked markers are expected to segregate with equal frequency (50%) of AB and BB genotypes, because the F_0 paternal AA genotype only occurs in F_2 males (Supplementary

figure 1). Reciprocally, fully sex-linked markers in F_2 males are expected to lack F_0 maternal BB genotypes. The chromosome 1 section encompassing markers that appear fully sex-linked based on the sample size of the study is referred to here as the region of reduced recombination (RRR) in males. As the genetic distance from the M-locus increases, the probability to detect recombinant AA genotypes in F_2 females and BB genotypes in F_2 males increases on both sides of the RRR.

In Cross #2, markers exhibiting extreme Mendelian segregation distortion in F_2 females were excluded from further analysis. Markers were included if they had heterozygous (AB) genotype frequencies inside the]20% - 65%[range, F_0 maternal (BB) genotype frequencies inside]5% - 65%[, and F_0 paternal (AA) genotype frequencies <70%. These arbitrary limits for initial marker selection were largely permissive for markers segregating according to theoretical proportions (25% AA : 50% AB : 25% BB at autosomal loci or a maximum of 50% AB : 50% BB at sex-linked loci) and facilitated the initial assignment of markers to linkage groups.

The occurrence of RAD markers in multiple copies (i.e., different loci sharing the same nucleotide sequence) was examined among all RAD markers identified. Only 0.5% of all reads aligned in multiple locations after suppression of multiple copy alignments in Bowtie, and the corresponding markers were discarded. RAD markers with a sequencing depth $\leq 12\times$ in F_0 parents were discarded to minimize the risk of false homozygous calls.

Recombination fractions between all pairs of selected markers were estimated using the rf.2pts function with default parameters. In Crosses #1 and #2, markers linked with a minimum LOD score of 8 and 14, respectively, were assigned to a same linkage group. Unlinked markers, if any, were removed from further analysis. For each cross, the minimum LOD score thresholds to declare linkage were chosen to maximize the number of markers in three distinct linkage groups. At this stage, linkage groups could be assigned to the three

distinct *Ae. aegypti* chromosomes based on markers located on supercontigs shared between the present study and a published linkage map [25]. For each marker, a χ^2 test was used to detect deviations of the observed genotype frequencies in the F_2 progeny from the theoretical 25% (homozygotes): 50% (heterozygote) expectations. In autosomal linkage groups, only markers whose genotype frequencies had Mendelian segregation ratios with *p*-values >0.1 and >0.05 were retained in the analysis of Cross #1 and Cross #2, respectively. Marker selection based on Mendelian segregation ratios was not applied to Cross #2 markers in the linkage group assigned to chromosome 1, and to Cross #1 markers in the linkage group assigned to chromosome 2 that displayed an unexplained deficit of AA genotypes in both sexes. Instead, iterative marker ordering was performed by sequentially removing markers with strong linkage discrepancies within the linkage group.

Cross #1 could be analyzed as a classical F_2 intercross design because both males and females were genotyped. Markers from each linkage group were provisionally ordered using the two-point based recombination counting and ordering (RECORD) algorithm [48]. Recombination fractions were converted into genetic distances in centiMorgans (cM) using the Kosambi mapping function [49]. Genetic distances among six equally spaced markers from this initial map were refined using multipoint likelihood of all possible orders. All remaining markers were then re-mapped to this initial frame in a stepwise fashion using the automated procedure of the order.seq function in the OneMap package. A “safe” linkage map was generated that only included markers uniquely mapped to a single position with a LOD score >3. The touchdown option was used to perform a second round of stepwise mapping for LOD scores >2. When possible, markers located on supercontigs that were previously mapped physically to the chromosomes [19] were substituted for markers located in supercontigs with unknown chromosome position. This step maximized connections between linkage and chromosome maps for synteny analysis.

Cross #2 could only be analyzed as a classical F_2 intercross design for autosomal linkage groups because only females were genotyped. A "safe" linkage map was generated as described above but chromosome 1 was further processed to account for the specific segregation patterns in the sex-linked region of F_2 females. Markers fully sex-linked lack F_0 paternal AA genotypes in F_2 females and segregate as in a backcross design in which F_1 AB heterozygotes are backcrossed to F_0 BB homozygotes. No linkage analysis method is readily available to deal with a chromosome that behaves partially as in a backcross (i.e., fully sex-linked loci) and partially as in an intercross (i.e., the remainder of chromosome 1 loci). Linkage analysis of chromosome 1 in Cross #2 was therefore restricted to the fully sex-linked region, referred to as the RRR. A new OneMap input file that only contained markers lacking AA genotypes was made by setting the population type as "backcross" instead of "F2 intercross". Markers were ordered in the RRR using the order.seq function in OneMap as described above. The final linkage map of Cross #2 combined the relative order of markers in the RRR obtained from a backcross analysis and the relative order of markers on chromosomes 2 and 3 obtained from an F_2 intercross analysis.

The linkage maps generated from Cross #1 and Cross #2 allowed a number of supercontigs to be identified that were mapped to distinct genome locations inside the same map. Intra- and inter- chromosome inconsistencies in the mapping of supercontigs can arise from supercontig misassemblies. The sequence of each RAD tag was mapped to supercontigs of the AaegL1 genome assembly using BWA [50] to provide anchors for integrating supercontigs into the genetic maps. The Chromonomer software [51] was used to detect and correct supercontigs misassemblies and orientation when possible (Supplementary file 1 and 2). Two out of the 4 supercontigs detected as misassembled in our maps had already been reported as misassembled in previous work [25,52]. Shotgun sequence assemblies of large genomes that are rich in sequence repeats are prone to supercontigs misassemblies [53]. As

reported by Juneja and colleagues [25], misassembled supercontigs are larger in size than correctly assembled ones. The size of the 4 misassembled supercontigs identified in this study was above the 95th percentile of the size distribution of all supercontigs of the current genome assembly. Chromosome rearrangements across different *Ae. aegypti* populations or inbred lines is an alternative hypothesis to explain discrepancies between supercontigs assignment to chromosomes.

Linkage disequilibrium and genetic differentiation analyses

Linkage disequilibrium (LD) between marker pairs on each chromosome was calculated as the asymmetric LD coefficient (*ALD*) in the R package "asymLD" [54]. *ALD* is a multiallelic extension of the r^2 metric [54] and is appropriate for our haplo-allelic dataset (see above). Heatmaps of the mean LD metric within a 1-Mbp window were plotted using the R package "lattice". The pairwise measure of genetic differentiation F_{ST} by Weir and Cockerham [55], as well as [55] and the observed heterozygosity *H* were calculated for the haplo-allele dataset using the program Genepop [56]. Discriminant analysis of principal components (DAPC) [27] was performed with the R package "adegenet" [57].

Acknowledgements

We thank Alongkot Ponlawat, Jason Richardson, Richard Jarman, Butsaya Thaisomboonsuk, Robert Gibbons, Stefan Fernandez, Timothy Endy, Anthony Schuster, and members of the Lambrechts lab for their insights. We are grateful to Eric Deveaud, Nicolas Joly, Olivia Doppelt-Azeroual and Véronique Legrand for assistance with computational analysis, and to the Nectar Research Cloud for computational resources. The opinions or assertions contained herein are the private views of the authors and are not to be construed as reflecting the

605 official views of the United States Army, Royal Thai Army, or the United States Department
606 of Defense.

Figure Legends

Figure 1: Synteny between linkage maps and chromosome idiograms reveals a large region of reduced recombination on the chromosome 1 of *Ae. aegypti* males. Circos plots [58] show syntenic links between linkage (left) and chromosome (right) maps. In panel (A), the linkage map was constructed using both F_2 males (n=22) and females (n=22) from Cross #1. In panel (B), the linkage map was generated from Cross #2 that only involved F_2 females (n=197). Linkage groups 1, 2 and 3 are represented in blue, orange and green, respectively. Markers are displayed with white internal ticks and their position in cM is indicated with a scale. Line graphs represent the observed frequency of AA genotypes (red), AB genotypes (blue) and BB genotypes (green) at each marker along the 3 linkage groups for males and females separately (A) or only females (B). The inset on the left side of panel (A) is an enlargement of the region displaying reduced recombination in males. In panel (A) the LOD curve for the sex QTL is displayed in purple in the outer track, and the black line represents the genome-wide statistical significance threshold. The 1.5 (dark purple) and 2 LOD (light purple) support intervals are represented on the top. Idiograms on the right side of the graphs represent *Ae. aegypti* mitotic chromosomes based on their banding patterns at early metaphase [19]. Centromeric and telomeric positions are indicated with long ticks. Bands are annotated according to their chromosome number followed by p (short arm) or q (long arm) and the number of the band. The band 1q21 harboring the M-locus is represented in red. Lines connecting the two sides of the graph show synteny between linkage and chromosome maps based on shared supercontigs located independently by fluorescent *in situ* hybridization on the chromosome map [19] and linkage mapping in the present study. The extent of the region of reduced recombination in males is highlighted in yellow on both linkage and chromosome maps. Supercontigs with conflicting location between the genetic and the chromosome maps are shown in grey.

Figure 2: F_{ST} (but not LD) patterns support the existence of a differentiated sex-determining chromosome in an unrelated, wild *Ae. aegypti* population. Upper: 1-Mbp sliding window median F_{ST} value for RAD markers on each of the three chromosomes in *Ae. aegypti* samples from Brazil. The approximate region of reduced recombination between M- and m-chromosomes (delineated in laboratory crosses) is highlighted in yellow. Lower: LD heatmaps for each chromosome in males (above diagonal) and females (below diagonal) from Brazil. LD was calculated as the asymmetric LD (*ALD*) coefficient for multiallelic loci. Median *ALD* values were calculated by 1-Mpb non-overlapping windows.

Figure 3: Discriminant analysis of principal components (DAPC) separates *Ae. aegypti* males and females based only on chromosome 1 genetic variation. DAPC for RAD markers on each of the three chromosomes in *Ae. aegypti* from Brazil (upper) and Australia (lower). Each graph represents the frequency distribution of individual mosquitoes according to their DAPC coordinates, stratified by sex.

Table 1: Summary statistics of linkage maps generated in this study.

	Chr1	Chr2	Chr3	Overall
Cross #1 map				
Number of markers mapped	25	36	41	102
Maximum marker spacing (cM)	18.1	29.6	11.0	29.6
Average marker spacing (cM)	3.5	6.0	2.0	3.8
Length of linkage group (cM)	83.4	209.7	79.0	372.2
Cross #2 map (restricted to the region of reduced recombination in males for Chr1)				
Number of markers mapped	12	32	33	77
Maximum marker spacing (cM)	10.0	17.5	8.6	17.5
Average marker spacing (cM)	3.3	2.2	2.1	2.3
Length of linkage group (cM)	36.8	67.8	66.5	171.2

651 **References**

652 1. Bell G (1982) The Masterpiece of Nature: the Evolution and Genetics of Sexuality.
653 London: University of California Press, Berkeley.
654 2. Bull JJ (1983) Evolution of Sex Determining Mechanisms. Menlo Park, CA:
655 Benjamin/Cummings.
656 3. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, et al. (2014) Sex
657 Determination: Why So Many Ways of Doing It? PLoS Biol 12: e1001899.
658 4. Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-
659 chromosome degeneration. Nat Rev Genet 14: 113-124.
660 5. Charlesworth B (1991) The evolution of sex chromosomes. Science 251: 1030-1033.
661 6. Ellegren H (2011) Sex-chromosome evolution: recent progress and the influence of male
662 and female heterogamety. Nat Rev Genet 12: 157-166.
663 7. Kaiser VB, Bachtrog D (2010) Evolution of sex chromosomes in insects. Annu Rev Genet
664 44: 91-112.
665 8. Vicoso B, Bachtrog D (2015) Numerous transitions of sex chromosomes in Diptera. PLoS
666 Biol 13: e1002078.
667 9. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, et al. (2015) A male-determining
668 factor in the mosquito Aedes aegypti. Science 348: 1268-1270.
669 10. Eckermann KN, Dippel S, KaramiNejadRanjbar M, Ahmed HM, Curril IM, et al. (2014)
670 Perspective on the combined use of an independent transgenic sexing and a
671 multifactorial reproductive sterility system to avoid resistance development against
672 transgenic Sterile Insect Technique approaches. BMC Genet 15 Suppl 2: S17.
673 11. Gilles JR, Schetelig MF, Scolari F, Marec F, Capurro ML, et al. (2014) Towards
674 mosquito sterile insect technique programmes: exploring genetic, molecular,
675 mechanical and behavioural methods of sex separation in mosquitoes. Acta Trop 132
676 Suppl: S178-187.
677 12. Toups MA, Hahn MW (2010) Retrogenes reveal the direction of sex-chromosome
678 evolution in mosquitoes. Genetics 186: 763-766.
679 13. Charlesworth B (1996) The evolution of chromosomal sex determination and dosage
680 compensation. Curr Biol 6: 149-162.
681 14. Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philos
682 Trans R Soc Lond B Biol Sci 355: 1563-1572.

683 15. Rice WR (1987) The accumulation of sexually antagonistic genes as a selective agent
684 promoting the evolution of reduced recombination between primitive sex
685 chromosomes. Evolution 41: 911-914.
686 16. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, et al. (2007) Genome sequence of
687 Aedes aegypti, a major arbovirus vector. Science 316: 1718-1723.
688 17. McClelland G (1962) Sex-linkage in Aedes aegypti. Trans roy Soc trop Med Hyg 56.
689 18. Hall AB, Timoshevskiy VA, Sharakhova MV, Jiang X, Basu S, et al. (2014) Insights into
690 the preservation of the homomorphic sex-determining chromosome of Aedes aegypti
691 from the discovery of a male-biased gene tightly linked to the M-locus. Genome Biol
692 Evol 6: 179-191.
693 19. Timoshevskiy VA, Severson DW, Debruyn BS, Black WC, Sharakhov IV, et al. (2013)
694 An integrated linkage, chromosome, and genome map for the yellow fever mosquito
695 Aedes aegypti. PLoS Negl Trop Dis 7: e2052.
696 20. Motara MA, Rai KS (1978) Giemsa C-banding patterns in Aedes (Stegomyia)
697 mosquitoes. Chromosoma 70: 51-58.
698 21. Timoshevskiy VA, Kinney NA, deBruyn BS, Mao C, Tu Z, et al. (2014) Genomic
699 composition and evolution of Aedes aegypti chromosomes revealed by the analysis of
700 physically mapped supercontigs. BMC Biol 12: 27.
701 22. Li N, Stephens M (2003) Modeling Linkage Disequilibrium and Identifying
702 Recombination Hotspots Using Single-Nucleotide Polymorphism Data. Genetics 165:
703 2213-2233.
704 23. Rasic G, Schama R, Powell R, Maciel-de Freitas R, Endersby-Harshman NM, et al.
705 (2015) Contrasting genetic structure between mitochondrial and nuclear markers in
706 the dengue fever mosquito from Rio de Janeiro: implications for vector control. Evol
707 Appl 8: 901-915.
708 24. Rasic G, Filipovic I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong
709 signals of geographic structure and relatedness patterns in the major arbovirus vector,
710 Aedes aegypti. BMC Genomics 15: 275.
711 25. Juneja P, Osei-Poku J, Ho YS, Ariani CV, Palmer WJ, et al. (2014) Assembly of the
712 genome of the disease vector Aedes aegypti onto a genetic linkage map allows
713 mapping of genes affecting disease transmission. PLoS Negl Trop Dis 8: e2652.
714 26. Holsinger KE, Weir BS (2009) Genetics in geographically structured populations:
715 defining, estimating and interpreting FST. Nat Rev Genet 10: 639-650.

716 27. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components:
717 a new method for the analysis of genetically structured populations. BMC Genet 11:
718 94.
719 28. Jombart T, Collins C (2015) A tutorial for Discriminant Analysis of Principal
720 Components (DAPC) using adegenet 2.0.0. http://adegenet.r-forge.r-
721 project.org/files/tutorial-dapc.pdf.
722 29. Krzywinski J, Grushko OG, Besansky NJ (2006) Analysis of the complete mitochondrial
723 DNA from Anopheles funestus: an improved dipteran mitochondrial genome
724 annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol 39:
725 417-423.
726 30. Guerrero RF, Kirkpatrick M, Perrin N (2012) Cryptic recombination in the ever-young
727 sex chromosomes of Hylid frogs. J Evol Biol 25: 1947-1954.
728 31. Grossen C, Neuenschwander S, Perrin N (2012) The evolution of XY recombination:
729 sexually antagonistic selection versus deleterious mutation load. Evolution 66: 3155-
730 3166.
731 32. Hall AB, Papathanos PA, Sharma A, Cheng C, Akbari OS, et al. (2016) Radical
732 remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. Proc
733 Natl Acad Sci U S A. In press.
734 33. Vicoso B, Kaiser VB, Bachtrog D (2013) Sex-biased gene expression at homomorphic
735 sex chromosomes in emus and its implication for sex chromosome evolution. Proc
736 Natl Acad Sci U S A 110: 6453-6458.
737 34. Olanratmanee P, Kittayapong P, Chansang C, Hoffmann AA, Weeks AR, et al. (2013)
738 Population genetic structure of Aedes (Stegomyia) aegypti (L.) at a micro-spatial
739 scale in Thailand: implications for a dengue suppression strategy. PLoS Negl Trop
740 Dis 7: e1913.
741 35. Lambrechts L, Ferguson NM, Harris E, Holmes EC, McGraw EA, et al. (2015) Assessing
742 the epidemiological effect of wolbachia for dengue control. Lancet Infect Dis 15: 862-
743 866.
744 36. Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, et al.
745 (2011) Successful establishment of Wolbachia in Aedes populations to suppress
746 dengue transmission. Nature 476: 454-457.
747 37. Mori A, Severson DW, Christensen BM (1999) Comparative linkage maps for the
748 mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci. J Hered
749 90: 160-164.

38. Fansiri T, Fontaine A, Diancourt L, Caro V, Thaisomboonsuk B, et al. (2013) Genetic Mapping of Specific Interactions between Aedes aegypti Mosquitoes and Dengue Viruses. PLoS Genet 9: e1003621.

39. Apostol BL, Black WC, Reiter P, Miller BR (1994) Use of randomly amplified polymorphic DNA amplified by polymerase chain reaction markers to estimate the number of Aedes aegypti families at oviposition sites in San Juan, Puerto Rico. Am J Trop Med Hyg 51: 89-97.

40. Hoffmann AA, Goundar AA, Long SA, Johnson PH, Ritchie SA (2014) Invasion of Wolbachia at the residential block level is associated with local abundance of Stegomyia aegypti, yellow fever mosquito, populations and property attributes. Med Vet Entomol 28 Suppl 1: 90-97.

41. Becker N (2003) Mosquitoes and their control. Second Edition. Heidelberg: Springer Verlag.

42. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7: e37135.

43. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25.

44. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22: 3124-3140.

45. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping Loci de novo from short-read sequences. G3 1: 171-182.

46. Margarido GR, Souza AP, Garcia AA (2007) OneMap: software for genetic mapping in outcrossing species. Hereditas 144: 78-79.

47. R Core Team. R Foundation for Statistical Computing V, Austria. (2015) R: A language and environment for statistical computing.

48. Van Os H, Stam P, Visser RG, Van Eck HJ (2005) RECORD: a novel method for ordering loci on a genetic linkage map. Theor Appl Genet 112: 30-40.

49. Kosambi DD (1943) The estimation of map distances from recombination values. Annals of Eugenics 12: 172-175.

50. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.

782 51. Amores A, Catchen J, Nanda I, Warren W, Walter R, et al. (2014) A RAD-tag genetic
783 map for the platyfish (Xiphophorus maculatus) reveals mechanisms of karyotype
784 evolution among teleost fish. Genetics 197: 625-641.
785 52. Bonin A, Paris M, Frerot H, Bianco E, Tetreau G, et al. (2015) The genetic architecture of
786 a complex trait: Resistance to multiple toxins produced by Bacillus thuringiensis
787 israelensis in the dengue and yellow fever vector, the mosquito Aedes aegypti. Infect
788 Genet Evol 35: 204-213.
789 53. Salzberg SL, Yorke JA (2005) Beware of mis-assembled genomes. Bioinformatics 21:
790 4320-4321.
791 54. Thomson G, Single RM (2014) Conditional asymmetric linkage disequilibrium (ALD):
792 extending the biallelic r2 measure. Genetics 198: 321-331.
793 55. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
794 structure. Evolution 38: 1358–1370.
795 56. Rousset F (2008) genepop'007: a complete re-implementation of the genepop software
796 for Windows and Linux. Mol Ecol Res 8: 103-106.
797 57. Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide
798 SNP data. Bioinformatics 27: 3070-3071.
799 58. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, et al. (2009) Circos: an
800 information aesthetic for comparative genomics. Genome Res 19: 1639-1645.
801 59. Rezvoy C, Charif D, Gueguen L, Marais GA (2007) MareyMap: an R-based tool with
802 graphical interface for estimating recombination rates. Bioinformatics 23: 2188-2189.

803

804

Supporting Information

Supplementary figure 1: Schematic of the expected sex-specific segregation of marker genotypes in an F_2 intercross. The reduced recombination region in males (represented as a black hatched area harboring the M-locus) is defined as the sex-linked region with a complete lack of observed AA genotypes in F_2 females and a complete lack of observed BB genotypes in F_2 males. The A allele from the M-chromosome is shown in red font.

Supplementary figure 2: Relationship between genetic positions of shared supercontigs between the genetic maps of Cross #1 and Cross #2. No major discrepancy in synteny or genetic distance was observed between the two maps. Coefficients of determination (R^2) and *p*-values were obtained from linear regressions.

Supplementary figure 3: Relationship between genetic (cM) and physical (Mbp) distances and estimated local recombination rates across the three *Ae. aegypti* chromosomes. The cumulative length of supercontigs captured with markers of the Cross #1 genetic map was used as a proxy for physical distances. The local recombination rate per nucleotide was calculated as the first derivative of a cubic smoothing spline function with the MareyMap R package [59]. A major recombination cold spot was observed on chromosome 1 around 55-60 cM.

Supplementary figure 4: Flowchart of bioinformatics analytical steps undertaken to generate a male-specific genome assembly and male-diagnostic RAD markers. Details of the analyses and results are presented in the Supplementary file 5.

Supplementary file 1: Chromonomer integration of the AaegL1 genome assembly with the Cross #1 genetic map. Markers are displayed on the charts as lines connecting the genetic map to the ordered supercontigs (colored boxes). The length of supercontigs in bp is displayed on the charts in squared brackets together with their orientation when known. Genetic distances in cM and the cumulative length of supercontigs in Mbp are indicated on the right side of supercontigs. Supercontigs that were incorrectly assembled according to the genetic map were split after Chromonomer integration as indicated in the "After Chromonomer Integration" section of each graph. Corrected assemblies are summarized in the tables provided for each linkage group.

838

839 **Supplementary file 2: Chromonomer integration of the AaegL1 genome assembly with**
840 **the Cross #2 genetic map.** Description is provided in the Supplementary file 1 caption.
841

842 **Supplementary file 3: Linkage maps generated from Cross #1 and Cross #2.** Marker
843 identification number, chromosome position, genetic position in cM, *p*-values of χ^2 tests of
844 expected Mendelian segregation ratios, percentage of missing genotypes, percentage of
845 marker genotypes, supercontig identification, marker position on supercontigs and
846 chromosomes and marker assignments to the region of reduced recombination in males
847 (RRR) are provided. Sex-specific information is given for Cross #1. Genotype proportions
848 were compared between males and females at each marker using χ^2 tests for Cross #1. Raw
849 and Bonferroni-corrected *p*-values are provided.
850

851 **Supplementary file 4: Marker and genotype information for the wild *Ae. aegypti***
852 **populations.** Sample name, origin and sex, marker position (on supercontigs and
853 chromosomes, based on the genetic map from [25]), average marker depth and observed
854 heterozygosity in each sex, and F_{ST} are provided. The variant calling format file (.vcf) for
855 each population sample is provided in a separate .xls spreadsheet.
856

857 **Supplementary file 5: Identification of male-diagnostic RAD markers.** Bioinformatics
858 analyses undertaken to assemble male-specific scaffolds and identify male-diagnostic RAD
859 markers are described. The resulting male-diagnostic scaffold sequences are listed in fasta
860 format. Analytical steps are depicted in the Supplementary figure 4.
861

862 **Supplementary file 6: Adapters and primer sequences used in this study, and schematic**
863 **of the final double-digest RAD sequencing assembly** (adapted from [42]).