Heredity (2015) 115, 471–479

www.nature.com/hdy

ORIGINAL ARTICLE

Genome-wide QTL mapping of saltwater tolerance in sibling species of *Anopheles* (malaria vector) mosquitoes

HA Smith[1,2,6], BJ White[1,2,3,6], P Kundert[1,2,7], C Cheng[1,2,8], J Romero-Severson[1,2], P Andolfatto[4,5] and NJ Besansky[1,2]

Although freshwater (FW) is the ancestral habitat for larval mosquitoes, multiple species independently evolved the ability to survive in saltwater (SW). Here, we use quantitative trait locus (QTL) mapping to investigate the genetic architecture of osmoregulation in *Anopheles* mosquitoes, vectors of human malaria. We analyzed 1134 backcross progeny from a cross between the obligate FW species *An. coluzzii*, and its closely related euryhaline sibling species *An. merus*. Tests of 2387 markers with Bayesian interval mapping and machine learning (random forests) yielded six genomic regions associated with SW tolerance. Overlap in QTL regions from both approaches enhances confidence in QTL identification. Evidence exists for synergistic as well as disruptive epistasis among loci. Intriguingly, one QTL region containing ion transporters spans the 2Rop chromosomal inversion that distinguishes these species. Rather than a simple trait controlled by one or a few loci, our data are most consistent with a complex, polygenic mode of inheritance.
Heredity (2015) **115,** 471–479; doi:10.1038/hdy.2015.39; published online 29 April 2015

INTRODUCTION

Multiple independent acquisitions and losses led to interspersion of saltwater (SW) tolerance across mosquito genera, and even among closely related members of species complexes (Albers and Bradley, 2011; Surendran *et al.*, 2011). Freshwater (FW) is thought to be the ancestral habitat for the aquatic stage of mosquitoes, with larvae for 95% of all species found in FW (Albers and Bradley, 2011). Transitions between FW and SW occupancy may represent cases of rapid evolution. The *Anopheles gambiae* species complex presents one example, where speciation a mere ~2 million years ago led to three main clades or phylogenetic branches of SW-tolerant and intolerant species (Fontaine *et al.*, 2015). One branch consists of the obligate FW human malaria vectors *An. coluzzii* and *An. gambiae* s.s., another branch contains the euryhaline vector *An. merus*, and a third combines both FW (non-vector *An. quadriannulatus*, vector *An. arabiensis*) and euryhaline members (vector *An. melas*) (Fontaine *et al.*, 2015).

Salinity transitions are found in other insects as well. Multiple acquisitions of salinity tolerance occurred among species of water beetles, where tolerance may be co-opted from mechanisms originally evolved for drought resistance (Arribas *et al.*, 2014). Thus, SW tolerance may represent an exaptation, a trait originally evolved not in direct response to selection as is typical of an adaptation, but rather via co-expression with other traits that were under selection (Ketterson and Nolan, 1999). For aquatic systems in general, salinity is a key determinant of biodiversity and species' distributions (Arribas *et al.*, 2014). Research on SW tolerance has added importance with mosquitoes, because environmental features that influence vector species' biogeography may in turn impact disease transmission (Ramasamy *et al.*, 2014). Osmoregulatory systems even have been proposed as targets for insect pest control (Cohen, 2013).

Recent studies of anopheline larval osmoregulation have focused on the role of the rectum in modifying the primary urine produced in the Malpighian (renal) tubules. The hindgut and Malpighian tubules are key components of insects' excretory systems (Dow, 2009). The working physiological model is that in an ion-poor FW medium, coupling of activity of a Na^+/K^+-ATPase and V-ATPase in non-dorsal anterior rectum (DAR) cells leads to hyperpolarization of the apical membrane, and transport of sodium ions into the cells (Smith *et al.*, 2008, 2010). When placed in saline water, tolerant anophelines exhibit downregulation of Na^+/K^+-ATPase within the non-DAR and upregulation in the DAR cells, presumably disrupting the enzyme's coupling with V-ATPase (Smith *et al.*, 2008, 2010). The V-ATPase in the non-DAR cells then may energize transporters to move ions from cells into the lumen for later excretion (Smith *et al.*, 2008). Na^+/H^+ antiporters also are proposed to have a role in sodium transport based on the voltage gradient generated by H^+ V-ATPases (Xiang *et al.*, 2012).

Among mosquitoes, *An. gambiae* is the most advanced in terms of genome annotation and assembly, with initial publication of the genome over a decade ago (Holt *et al.*, 2002), but even for this species little is known about the molecular basis of larval osmoregulation.

[1]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA; [2]Eck Institute for Global Health, University of Notre Dame, Notre Dame, IN, USA; [3]Department of Entomology, University of California-Riverside, Riverside, CA, USA; [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA and [5]The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA
Correspondence: Dr HA Smith, Department of Biological Sciences, University of Notre Dame, 329 Galvin Life Sciences, Notre Dame, IN 46556, USA.
E-mail: hasmith.biology@gmail.com
[6]These authors contributed equally to this work.
[7]Current address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77054, USA.
[8]Current address: Center for Computational Biology and Bioinformatics, University of Texas at Austin, Austin, TX 78731, USA.
Received 30 November 2014; revised 11 March 2015; accepted 20 March 2015; published online 29 April 2015

Screening for osmoregulatory candidate genes and pathways based on studies of other taxa can provide insight, as some mechanisms are widespread. However, a candidate gene approach may fail to identify novel networks specific to mosquitoes. Further complicating such inferential studies is the fact that the model organism most closely related to mosquitoes, *Drosophila melanogaster*, lacks an aquatic stage. Studies attempting to understand osmotic stress responses in fruit flies often examine responses to dietary salt (Stergiopoulos *et al.*, 2009). This dietary exposure is different from the complete immersion of mosquito larvae in saline waters. As adult stages of female mosquitoes possess mechanisms for rapid excretion of water and ions following a blood meal (Coast *et al.*, 2005), there also exists the potential for larval osmoregulation to have evolved as an exaptation, potentially distinguishing mosquitoes' evolutionary trajectory from other SW-tolerant taxa.

Recently White *et al.* (2013) laid the groundwork for quantitative trait locus (QTL) mapping of larval anopheline osmoregulation by reporting that survivorship in 50% seawater is a discriminating trait, with salinities this high tolerated by *An. merus* but lethal to *An. coluzzii* and their F_1 hybrids. Albeit not without limitations, QTL mapping is a powerful approach for associating phenotypes with genomic regions, and ultimately genotypes. Moreover, such approaches do not rely on *a priori* information of candidate genes. Many recent single-nucleotide polymorphism (SNP)-based QTL analyses have leveraged next-generation sequencing technologies for genome-wide mapping. Here, we use an analogous approach to RAD-tag sequencing (Andolfatto *et al.*, 2011), creating a dense genetic map for associating genomic locations with SW tolerance. We performed QTL mapping based on a backcross of *An. merus* and *An. coluzzii*, and screened >1000 progeny for SW tolerance. This research serves as an important first step in revealing the complex genetic architecture of anopheline osmoregulation. Although we map tolerance as a simple binary trait by scoring survivorship or mortality in SW, we detect multiple QTL across several chromosomal arms, as well as evidence for epistasis.

MATERIALS AND METHODS

Experimental crosses and phenotype

The FW taxon *Anopheles coluzzii* strain Mali-NIH, formerly called *An. gambiae* 'M' form, and the SW-tolerant species *An. merus* strain MAF, were crossed to generate F_1 hybrids. Eggs of each species were obtained from the Malaria Research and Reference Reagent Center or MR4 (MRA-860 and MRA-1156, MR4, ATCC, Manassas, VA, USA). Protocols for mating and colony maintenance were performed as previously described (White *et al.*, 2013). Briefly, colonies were housed within insectaries at the University of Notre Dame with a 12-h light: 12-h dark cycle, 27 °C and 85% relative humidity. Larvae were kept in trays with a surface area $\sim 432\,cm^2$, at 200 larvae l^{-1}. The parental *An. coluzzii* colony was maintained in reverse-osmosis water, whereas in concordance with its maintenance since colonization, parental *An. merus* was reared at 50% salinity of seawater (addition of $15.85\,g\,l^{-1}$ commercial NaCl). Larval food consisted of 2:1 ground fish pellets to baker's yeast. To generate F_1 hybrids, a mass mating was performed of female *An. merus* with *An. coluzzii* males. As male hybrids are sterile (Davidson, 1964), F_1 hybrid females were backcrossed to males from the *An. merus* parental colony. The resulting backcross progeny were phenotyped for tolerance at 50% SW ($15.85\,g\,l^{-1}$ NaCl). Animals were scored as tolerant if they survived to adulthood, or susceptible if they died as first instars. Mosquitoes were frozen (−80 °C) until DNA extraction of individuals with the DNeasy 96 Blood and Tissue Kit (Qiagen, Valencia, CA, USA).

Adults were sexed by morphology, but this is not possible for first instar larvae. Thus, for larvae, PCR amplification of a Y-chromosome fragment was performed to determine which larvae were male, using primers Y-339-For (5′-CGATCAATAATGCGGCAGCTC-3′) and Y-339-Rev (5′-GTTGCGGTC TGCGAAGAGAA-3′). As a positive control, we developed a multiplex PCR design that also amplified X-linked ribosomal DNA (rDNA) in hemizygous males and females of each species, using the published primers 'UN', 'ME' and 'GA' (Scott *et al.*, 1993). The PCR reactions occurred in 10 µl volumes with 1 µl individual mosquito DNA, ~0.4 U Taq and final concentrations of 0.2 mM each dNTP, 28 mM TrisCl pH 8.3, 70 mM KCl, 1.5 mM $MgCl_2$, 0.24 µM each of primers UN, ME and GA, and 1.2 µM each of the forward and reverse Y-339 primers. Thermal cycling was performed with 2 min at 94 °C, followed by 35 cycles of 94 °C for 30 s, 54 °C for 30 s and 72 °C for 30 s, then ending with 72 °C for 2 min. Gel electrophoresis (2% agarose) revealed amplification of the 134-bp Y-339 fragment plus rDNA (~400 bp) in the presence of male DNA, but only amplification of the rDNA fragment for females. We tested the robustness of the multiplex PCR over a range of input male and female DNA, with as little as 0.1 ng DNA per reaction yielding Y-339 and rDNA amplicons. Based on tests using morphologically sexed specimens, the Y-339 fragment consistently and exclusively amplified in males, while the rDNA amplified in both sexes. For sexing backcross first instar larvae, reactions were performed with 1 µl undiluted DNA extract (~0.5–1 ng).

Construction of pseudo-reference parental sequences

For *An. coluzzii,* we constructed a standard genomic DNA library with ~400–500-bp inserts, and collected 157 million paired 101-bp reads from sequencing one lane on the Illumina (San Diego, CA, USA) HiSeq. We trimmed the reads to base quality $\geqslant 20$ and a minimum length of 30 bp. For *An. merus*, we accessed 309 million previously published Illumina sequence reads (SRR403835, SRR403841–3). Sequences from both species were mapped to the *An. gambiae* PEST reference assembly AgamP3 (Holt *et al.*, 2002; Megy *et al.*, 2012) using Stampy (http://www.well.ox.ac.uk/project-stampy) with substitution rate $=0.02$. As both *An. merus* and *An. coluzzii* are fixed for the 2La chromosomal inversion (Coluzzi *et al.*, 2002) while the PEST sequence is in the standard arrangement, before mapping this region was identified using previously reported breakpoints (Sharakhov *et al.*, 2006) and replaced with the inverted *An. coluzzii* sequence. Otherwise genomic coordinates reported in this study correspond to those in PEST. We realigned insertions and deletions relative to the PEST reference genome using the GATK IndelRealigner (http://www.broadinstitute.org/gatk). Base calling was performed with SAMtools (http://samtools.sourceforge.net/), setting the bcftools 'scaled mutation rate for variant calling' to –t 0.01. To generate species-specific pseudo-reference sequences, we used a custom script to identify all single-nucleotide differences relative to the PEST reference genome that had genotype quality $\geqslant 20$ and depth coverage of 10–1000 (all insertion–deletion variation was ignored). We then used the mutfa function of the script seqtk (https://github.com/lh3/seqtk) to incorporate these into the PEST reference sequence to create a pseudo-reference sequence.

We prepared multiplex shotgun genotyping (MSG) libraries (Andolfatto *et al.*, 2011) from 192 individuals of each parental species. These libraries were sequenced on 1.5 HiSeq lanes with 101-bp single-end sequencing yielding ~222 million reads, which allowed us to tune the MSG error model. (The half lane was split with backcross progeny; see below). We retained information from the best-covered individuals, using 100 000 reads each from 180 *An. merus* and 155 *An. coluzzii* individuals to determine sites polymorphic within the two parental species. The 216 660 polymorphic sites in *An. coluzzii*, and 136 577 in *An. merus,* were masked in analysis of backcross progeny. This left >6 million ancestry-informative sites on the autosomes and >0.7 million on the X chromosome. We constructed MSG libraries from 1152 backcross progeny with 2.5 HiSeq lanes. The half lane consisted of 101-bp single-end sequencing run in standard mode; subsequent lanes were run in rapid mode and generated 141-bp single-end reads, which we trimmed to 101 bp before analysis for consistency. We obtained ~179 million reads from the backcross progeny. Library construction from approximately 10–15 ng DNA per individual, sequencing and assignment of ancestry was performed using the MSG library protocol (Andolfatto *et al.*, 2011) and associated software pipeline (https://github.com/JaneliaSciComp/msg).

Genotyping and mapping

We used the MSG software pipeline (Andolfatto *et al.*, 2011) to genotype 576 SW-susceptible larvae (288 male and 288 female) and 576 SW-tolerant adults

(288 male and 288 female) from the F_1 backcross of *An. merus* and *An. coluzzii*. The MSG parameters were set as follows: the priors on ancestry homozygosity for *An. coluzzii* = 0, heterozygosity = 0.5 and homozygosity for *An. merus* = 0.5; the ancestry error probabilities for *An. coluzzii* = 0.03 and for *An. merus* = 0.02; the genome-wide recombination rate = 3 crossovers per genome per generation; and the arbitrary recombination scaling factor, rfac = 10^{-7}. Sequence files have been deposited at NCBI SRA under accession SRP050537.

From the MSG backcross libraries, 766 723 autosomal and 51 910 X chromosome SNP markers (Table 1) were assigned probabilities of homozygosity for the *An. merus* sequence, heterozygosity or hemizygosity on the X chromosome, based on Hidden Markov Models as described in Andolfatto *et al.* (2011). Markers that contained the same sequence as their neighbors (differing by <10% in the probability of *An. merus* homozygosity) were removed to reduce redundancy and improve computation efficiency of mapping software, following Cande *et al.* (2012). Individuals with poor-quality sequences (for example, <3000 markers) were excluded from analysis ($N = 18$ individuals, ~1%). A sharp drop in linkage levels between markers occurred in the MSG output profiles toward the centromere of the X chromosome. This region primarily is heterochromatin, which tends to have a high level of repetitive DNA and transposable elements (Sharakhova *et al.*, 2010). Owing to the potential that the abnormal linkage levels represent assembly errors or other inaccuracies, we excluded this centromere-proximal ~5.1 Mb. Hence, our analysis of the X chromosome focuses on the distal ~19.3 Mb. In addition, we removed markers missing sequence information in ⩾50% of the mosquitoes. Sequence coverage was lowest at telomeres (Supplementary 1). Thus, our final QTL scans included 2387 markers and 1134 backcross progeny (Table 1), consisting of 574 SW-tolerant and 560 susceptible individuals.

Computation of recombination rates and QTL mapping was conducted in R v. 3.0.2. We used the R/qtl package v. 1.31–9 to estimate the genetic map in centimorgans (cM) with the Kosambi map function (Broman and Sen, 2009). Sliding window (10 Mb) analyses of recombination rates were performed with Broman's R/xoi package v. 0.61–1 (http://www.biostat.wisc.edu/~kbroman/software.html). From chi-square tests run in R/qtl, segregation distortion was significant across 1202 of the 2387 markers, with a higher likelihood for recovering heterozygous genotypes on the autosomes. However, segregation distortion was relatively uniform across each chromosome arm, and thus we included all 2387 markers in our analysis. We detected QTL using Bayesian interval mapping and machine learning (random forests), two conceptually different approaches.

Scans for QTL

For QTL analysis, SW tolerance was treated as a binary trait. We first scanned for QTL associated with SW tolerance by performing Bayesian interval mapping with the R/qtlbim package v. 2.0.7 (Yandell *et al.*, 2007). Probabilities of homozygosity for *An. merus* markers from the MSG output were imported into a cross object for use with R/qtlbim, similar to Cande *et al.* (2012). Single-dimension scans were performed allowing for the presence of epistasis, and for each marker, we evaluated the contribution of the marker to the trait averaging across all models including that marker. This method adjusts for potential impacts of other loci, and thereby can reduce variation that is explained by other loci, as well as bias from linkage among markers. Although at the time of writing, this software does not allow analysis of sex chromosomes, we included gender as a fixed covariate to permit tests of gene–sex interactions, similar to the approach of Yi *et al.* (2007). Results were assessed with Bayesian log posterior density (LPD) profiles, which are analogous to LOD profiles and represent the logarithm of the posterior distribution for QTL locations (Sen and Churchill, 2001; Yandell *et al.*, 2007). Analyses were made of the mean from 10 separate Bayesian interval scans, each started from a different seed, to account for stochastic variability among runs. For calling QTL peaks, we chose as a threshold a mean sum LPD >10. Sum LPD scores were computed as the sum of the main, epistasis, and gene–sex LPD values. To test this threshold, we used 1000 permutations of the SW-tolerant phenotype. We did not alter the genotype or order of markers in these tests, and thus the permuted data retained the original structure of the observed data. The proportion of the 1000 runs of permuted data with sum LPD >10 remained <0.05 at all loci (<0.02 except at telomeres, where the proportion of missing data was greatest), indicating an LPD >10 corresponds to a *P*-value <0.05 (Supplementary 2).

A second, machine learning approach was performed by construction of random forests using the R package randomForest v. 4.6–7 (Liaw and Wiener, 2002). Random forests analysis is a non-parametric technique that can be used in the classification of individuals or phenotypes based on bootstrap sampling of large numbers of predictors, such as SNPs. Although not specifically testing for epistasis, it allows for and can detect genes that affect phenotypes via interactions (Chen *et al.*, 2011). We used random forests to classify the mosquito individuals as SW-tolerant or susceptible, applying the genotype information across markers as predictors. Markers were encoded as homozygous for *An. merus* or heterozygous (or hemizgyous on the X). That is, we used the discrete hard calls generated from the MSG output rather than genotype probabilities. We also included gender as male or female as an additional predictor. Missing values for genotypes were imputed using the function rfImpute.

Parameters for random forests were optimized by three separate runs with distinct seeds, with 500, 1000, 5000 or 10 000 trees, and varying the number of predictor markers sampled (mtry of 12, 24, 49, 98, 239, 796 and 1194). We selected values that minimized the out-of-bag error rate, which represents the error in classification accuracy (Liaw and Wiener, 2002; Chen *et al.*, 2011). This resulted in use of 5000 trees and 796 predictors (~1/3rd of the 2387 markers). The out-of-bag error showed little change for ⩾500 trees; we used 5000 trees to be conservative. As with qtlbim, we analyzed the mean of 10 separate, uniquely seeded runs of random forests. Importance of predictors for classification was assessed with unscaled permutation variable importance (VI), which is computed by the program from randomly permuting genotype values at markers and testing for a change in classification accuracy. Higher VI values suggest a greater role of the variable (SNP) in classifying individuals as SW-tolerant or susceptible. The threshold for calling QTL peaks was assigned as mean unscaled permutation VI >0.002. For this more computationally intensive approach, we used 100 permutations of the SW tolerance phenotype to test the chosen VI threshold. No genomic locus from the 100 permuted data sets had a VI >0.002, suggesting a *P*-value <0.01. Fifteen of the 100 runs yielded a VI >0.002 for the trait of gender, and thus our chosen threshold corresponded to a *P*-value cutoff of 0.15 for this variable. However, the maximum VI for gender was 0.005 for the permuted data set, well below the VI of 0.008 for the observed (non-permuted) data set.

Similar to the approach of Holliday *et al.* (2012), we tested for epistasis by omitting a subset of data, and then re-running the random forests analysis to determine whether the ability to detect QTL changed. Specifically, we computed the difference in the mean of 10 runs of random forests with all QTL data and the trait of gender, vs the mean of 10 runs after omitting all markers on a chromosome or gender identity. Higher VI scores in a predicted QTL after removing another chromosome suggest negative (disruptive) epistasis between the QTL and the omitted chromosome, whereas lower VI scores indicate a positive (synergistic) interaction. Genes within QTL peaks were identified by use of BioMart in VectorBase (Megy *et al.*, 2012), based on genomic coordinates of the peaks and the AgamP3.8 gene set.

Table 1 Number of informative markers for backcross progeny, and thinned filtered markers used in final QTL analysis

Chromosome	*Informative markers*	*Thinned markers*	*Physical length (Mb)*	*Genetic length (cM)*
2	412 329	952	110.9	78.7
3	354 394	1000	95.2	80.0
X	51 910	435	19.3	10.7

Abbreviation: QTL, quantitative trait locus.
Physical and genetic lengths for the X chromosome reflect data used in the analysis, after masking the final 5.1 Mb of the 24.4 Mb chromosome.

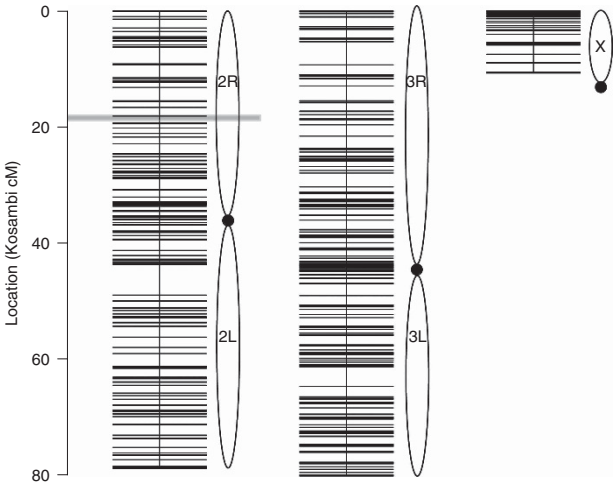

Figure 1 Genetic map showing marker density along chromosomes, with gray shading at the inversion 2Rop.

RESULTS

SNP coverage

We mapped a total of 818 613 informative markers in 1134 backcross individuals, creating a dense genetic map of 2387 markers after removing redundant data (Table 1, Figure 1). Distances of the chromosomes are approximately 78.7 and 80.0 Kosambi cM (111 and 95 Mb) for chromosomes 2 and 3, respectively, and 10.7 cM for the first 19.3 Mb of the X chromosome (Table 1, Figure 1). These values are remarkably similar to autosomal lengths proposed nearly two decades ago, of 72.1 Kosambi cM for chromosome 2 and 93.7 cM for chromosome 3 (Zheng *et al.*, 1996). A single, fixed inversion named 2 Rop that differentiates *An. coluzzii* and *An. merus* is found in *An. merus* on arm 2R at ~9.5–36 Mb (Fontaine *et al.*, 2015). As expected, recombination rates are greatly reduced in the inversion and at autosomal centromeres (Supplementary 3).

Location of QTL and gene interactions

We report strong congruence from both mapping approaches, with thresholds for calling QTL supported by permutation tests. All autosomal peaks found by random forests also were found by Bayesian interval mapping (Figures 2 and 3); therefore, we used the narrower peak intervals from random forests to delimit QTL (Table 2, Supplementary 4). The only notable autosomal peaks of elevated LPD found by Bayesian interval mapping that were not identified by random forests were located at the telomeres. Considering the high proportion of missing data at telomeres (Supplementary 1) and their higher LPD scores in the randomly permuted data (Supplementary 2), we conclude that these telomere peaks probably do not reflect true QTL and do not consider them further.

Two QTL regions were identified per chromosome for a total of six peaks, labeled QTL A through F (Figures 2 and 3). Together these peaks contain 2728 genes or roughly 20% of all annotated genes in the *An. gambiae* PEST assembly (Supplementary 5). Near and within the 2Rop inversion, we found markers with elevated VI in random forests tests at 7.19–7.24, 8.77, 9.41–9.42 and 36.88 Mb, with the entire region showing a high Bayesian LPD (Supplementary 4). Combining these intervals, we defined 7.19–36.88 Mb as the range of QTL A. Other autosomal QTL are at the centromere of chromosome 2 (QTL B) and within arm 3R (QTL C, D), whereas QTL E and F are on the X chromosome (Table 2).

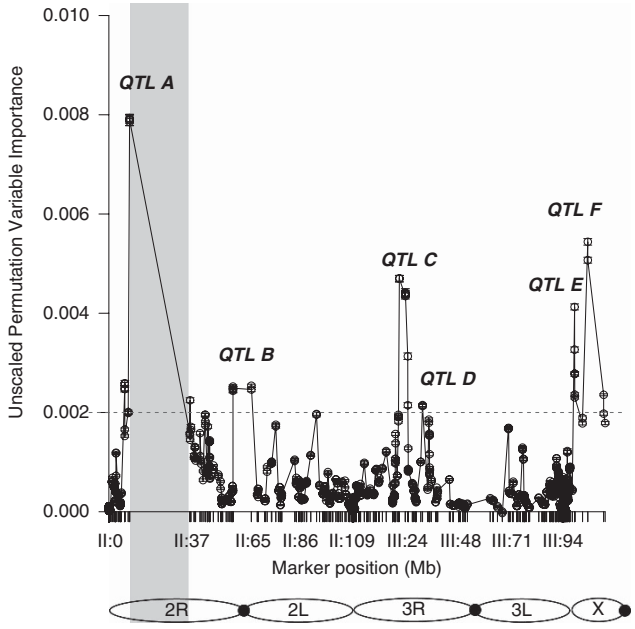

Figure 2 Mean VI (±1 s.e.) for predicting SW tolerance from 10 runs of random forests. Schematic below graph depicts positions of chromosome arms, with gray shading at the inversion 2Rop. Dashed horizontal line is the threshold chosen for identifying QTL peaks.

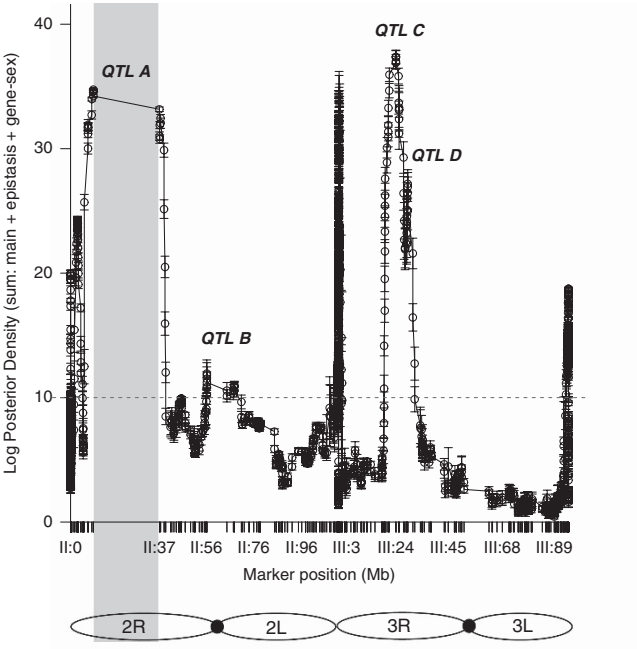

Figure 3 LPD for SW tolerance, shown as the mean (±1 s.e.) from 10 runs of Bayesian interval mapping. Schematic below graph depicts positions of chromosome arms, with gray shading at the inversion 2Rop. Dashed horizontal line is the threshold chosen for identifying QTL peaks.

As F_1 hybrids do not survive 50% salinity (White *et al.*, 2013), we expected the loci of greatest import to salinity tolerance would be homozygous for the *An. merus* genotype in the SW-tolerant backcross progeny. Conversely, we anticipated susceptible progeny to be heterozygous (or hemizygous for *An. coluzzii* on the X). Therefore,

we separately considered genotypes of the tolerant (Figure 4) and susceptible (Figure 5) backcross progeny at markers within the six QTL peaks, as these regions represent our strongest candidates for containing loci that confer SW tolerance. Among the 560 SW-susceptible progeny, only one individual is homozygous for the *An. merus* genotype at all 40 loci in the QTL peaks. Most tolerant individuals are homozygous (hemizygous) for the *An. merus* genotype at the markers within at least two of the six QTL peaks, although which two peaks are homozygous for the *An. merus* genotype varies. However, there are some exceptions. Among the 574 tolerant individuals, there are 47 (~8%) with heterozygous genotypes (or hemizygosity for *An. coluzzii*) at all 40 loci in the QTL peaks. These individuals were heterozygous across the majority of their genomes, but only 13 tolerant individuals are heterozygous (or hemizygous for *An. coluzzii*) across the entire set of 2387 analyzed loci. Most had at least a portion of their genome showing homozygosity for *An. merus*. Notably, the 47 individuals were scattered across three plates during library preparation, making a systematic error in genotyping unlikely. Also, this pattern is less pronounced when animals are separated by gender. The most striking result is seen for males. Of the 286 tolerant males, most ($N = 247$, or ~86%) are hemizygous for *An. merus* across all markers within the X chromosome peaks, QTL E and F (Supplementary 6). In contrast, only ~50% of tolerant females were homozygous for *An. merus* at all loci across QTL E and F.

Table 2 Genomic regions of QTL for anopheline SW tolerance, based on random forests variable importance >0.002

QTL name	*Chromosome arm*	*Start (bp)*	*Stop (bp)*	*Length (Mb)*
A	2R	7 193 539	36 879 143	29.686
B	2R/2L	56 364 210	64 692 100	8.328
C	3R	21 002 600	24 851 978	3.849
D	3R	31 505 802	31 509 339	0.004
E	X	5 411 204	5 491 649	0.080
F	X	11 386 614	18 597 670	7.211

Abbreviations: LPD, log posterior density; QTL, quantitative trait locus; SW, saltwater.
Autosomal QTL also showed Bayesian LPD >10.

Besides identifying genomic regions associated with SW tolerance, we also aimed to enhance understanding of the genetic architecture of this trait. Bayesian interval mapping detected the QTL peak A at the 2Rop inversion as a main effect (Supplementary 7). The other peak on chromosome 2, QTL B, showed elevation of LPD scores in scans for epistatic (Supplementary 8) and for main effect loci. The peaks on arm 3R (QTL C, D) were most prominent in QTL mapping of the gene–sex interaction effect (Supplementary 9).

Random forests were used for a complementary investigation of genetic architecture. If removing a chromosome improves detection of a QTL peak on another chromosome (increase in the peak's VI score), this is evidence for disruptive epistasis between the QTL peak and

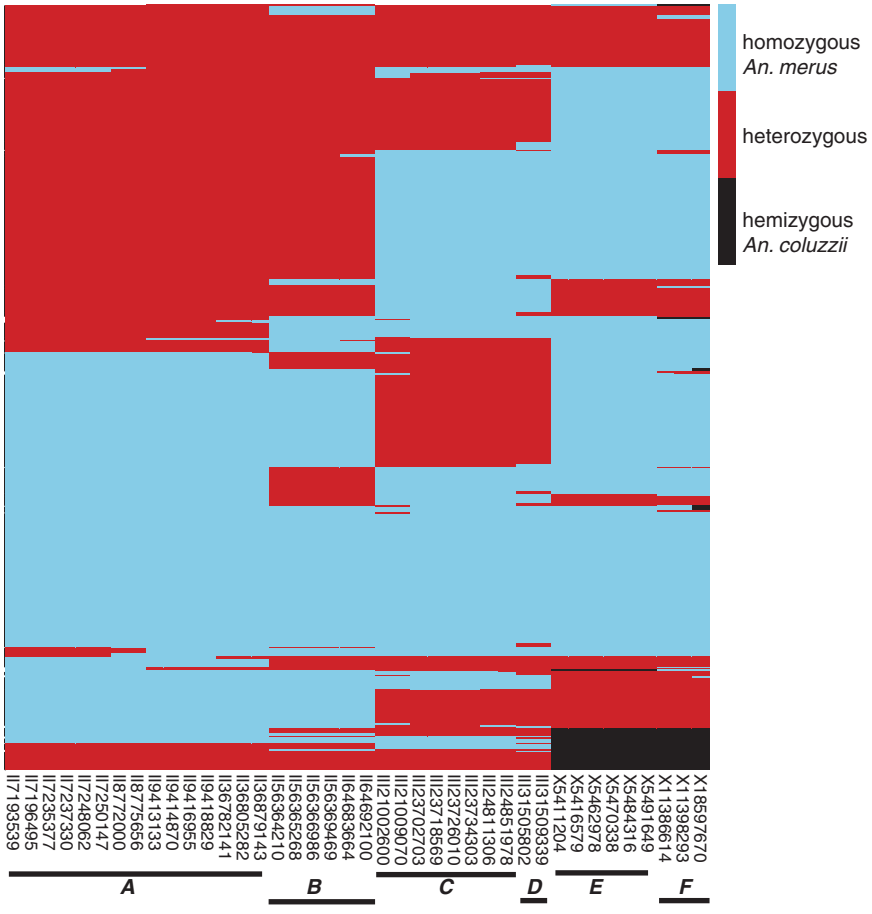

Figure 4 Heatmap of genotypes for SW-tolerant backcross progeny, showing SNP markers in the six QTL regions. Each row represents an individual mosquito, and each column a marker. Blue color on the X may represent homozygosity for *An. merus* in females, or hemizygosity for *An. merus* in males.

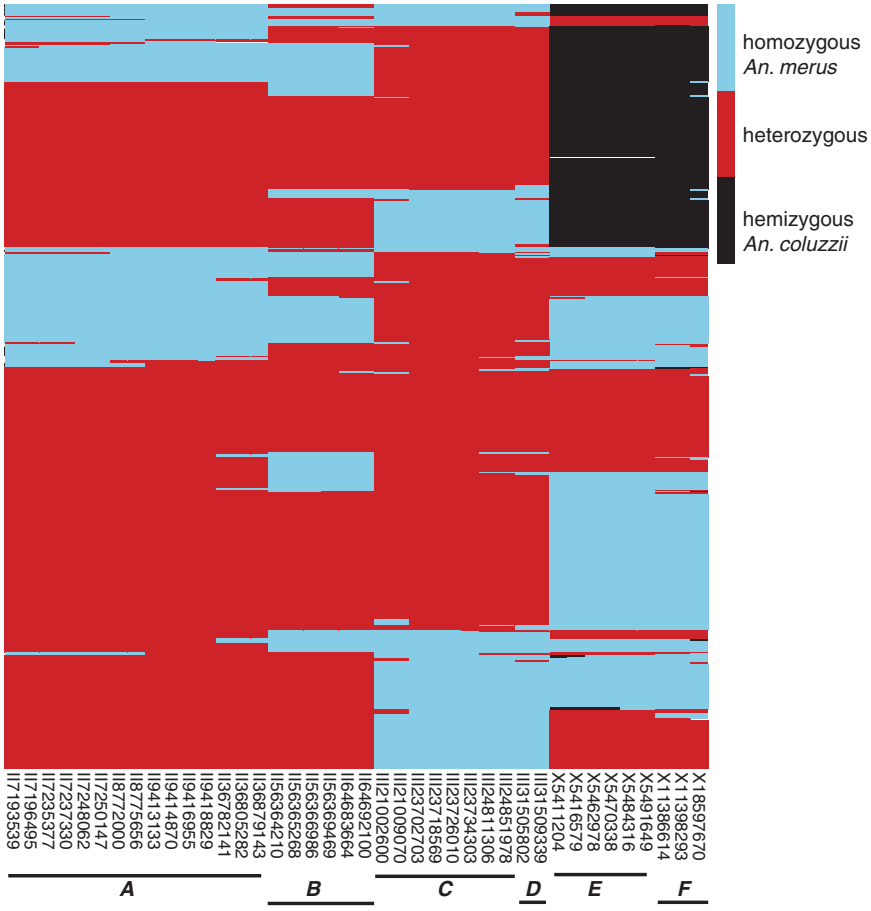

Figure 5 Heatmap of genotypes for SW-susceptible backcross progeny, showing SNP markers in the six QTL regions. Each row represents an individual mosquito, and each column a marker. Blue color on the X may represent homozygosity for *An. merus* in females, or hemizygosity for *An. merus* in males.

omitted chromosome. Conversely, a decline in detection ability (decrease in VI) indicates synergistic epistasis. Chromosome 2 shows evidence for synergistic epistasis with both QTL on chromosomes 3 and both QTL on the X, whereas chromosome 3 appears to be in synergistic epistasis with QTL A on arm 2R. Together, this suggests the potential for synergistic epistasis between loci within or near the 2Rop inversion in QTL A, and the third chromosome QTL peak C or D. The fact that QTL A only was detected as a main effect in the Bayesian scan may reflect that QTL A loci possess a large marginal effect, along with a weaker epistatic interaction. Alternatively, the Bayesian analysis simply may not be as powerful as random forests in testing for epistasis. Among loci, only QTL B showed evidence of disruptive epistasis. Omitting the trait of gender had negligible impact on VI scores for QTL C or D on arm 3R, but omission of the X chromosome decreased the VI score for QTL D (Figure 6). Although the QTL on arm 3R were detected as gene–sex interactive effect loci by Bayesian interval mapping, the Bayesian analysis may have confounded gender with the difference in number of X chromosomes for males ($N=1$) and females ($N=2$). Loci on the X chromosome, rather than simply the gender identity, may interact with loci in QTL C and/or D to influence SW tolerance.

Candidate osmoregulatory genes

Searches of VectorBase gene descriptions reveal several transporters, channels, and kinases in the QTL peaks (Supplementary 10). Two genes with the VectorBase gene description of *Solute carrier family 12 (sodium/potassium/chloride transporter), member 2* occur within QTL A, represented by AGAP003274 and AGAP003275, as well as *Diuretic hormone 44* (AGAP003269). Also, QTL A contains *Atrial natriuretic peptide receptor A* (AGAP003283), *Na^+/K^+-ATPase* (AGAP002858), and Na^+/H^+antiporters *AgNHA1* (AGAP002093) and *AgNHA2* (AGAP002324). The insecticide resistance gene *Para*, described as a *Voltage-gated sodium channel* (AGAP004707), is one of several genes in QTL B. Ion transporters are found in QTL C as well, including the *Solute carrier family 9 (sodium/hydrogen exchanger), member 3* (AGAP009036). The one gene (AGAP009332) in QTL D does not have a gene description in VectorBase, but is listed in VectorBase as an ortholog of the *Slowpoke binding protein, putative* in *Ixodes scapularis*; the gene may have a role in insulin signaling. An NCBI protein blast (default settings) reveals a hit of AGAP009332 to the *Drosophila Slowpoke binding protein* (e-value of 1e-19), but with a percent identity of merely ~35%. The QTL F peak contains two *Amiloride-sensitive sodium channels, other* (AGAP000657, AGAP000840), as well as a *Chloride intracellular channel* (AGAP000943). We searched VectorBase descriptions of genes within the six QTL for aquaporins, but did not find any. A handful of mitogen-activated protein kinases occur within the QTL peaks and could be important in signaling; these include three genes in QTL A (AGAP001867, AGAP013516 and AGAP003365), and one each in QTL E (AGAP000310) and QTL F (AGAP000747). The QTL A peak also contains the transcription factor *hepatocyte nuclear factor 4* (*HNF-4*; AGAP002155).

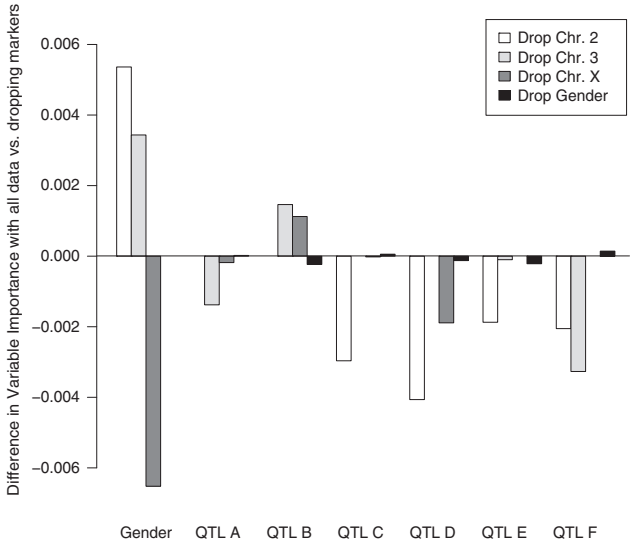

Figure 6 Change in VI at QTL peaks from dropping markers on a single chromosome, or the trait gender. Columns show the difference in the mean of 10 runs of random forests with all data, minus 10 runs with data omitted.

DISCUSSION

Multiple QTL associated with anopheline SW tolerance

Here, we have used QTL mapping with crosses of the FW species *An. coluzzii* and euryhaline *An. merus* to identify six genomic regions associated with inheritance of SW tolerance in anopheline mosquitoes. Bayesian interval mapping supported QTL from a random forests classification approach. Two QTL regions were identified per chromosome, consisting of one on arm 2R (QTL A), one spanning the centromere of chromosome 2 (QTL B), two on arm 3R (QTL C and D) and two on the X chromosome (QTL E and F). Evidence exists for a possible association of mosquito gender with SW tolerance, although this may reflect the importance of X-linked markers rather than identity as male or female.

Previous research showed survival of F_1 hybrids in SW is higher when female *An. merus* are mated to male *An. coluzzii* than with the reciprocal cross, with a LC_{50} (lethal concentration) of ~35.0% for hybrids with *An. merus* mothers and ~27.5% for those with *An. coluzzii* mothers (White *et al.*, 2013). Based on this, the authors hypothesized that maternal inheritance influences tolerance, proposing that epigenetics, a cytoplasmic factor, or X-linkage could be involved. Consistent with the importance of X-linked loci, here we report that most tolerant males were hemizygous for the *An. merus* genotype across the QTL peaks on the X chromosome. Yet only about half of the females were homozygous for *An. merus* at the X chromosome QTL peaks. Perhaps having a single allele (hemizygosity for *An. merus* in males or heterozygosity in females) with the *An. merus* sequence at these loci suffices to promote SW tolerance. Importantly, the prior study by White *et al.* (2013) indicated that 50% salinity, as used here, was a discriminating dose. Although hybrids exhibited intermediate tolerance relative to the FW species *An. coluzzii* and SW-tolerant *An. merus*, no individual hybrids and no parental *An. coluzzii* survived 24 h at 50% salinity (White *et al.*, 2013). Mortality was below 1% for parental *An. merus* at this concentration.

Epistasis and complexity of inheritance of SW tolerance

SW-tolerant individuals typically are homozygous for the *An. merus* parental genotype across markers within at least two of the six QTL regions. However, which QTL regions are of the *An. merus* genotype varies among the tolerant mosquitoes, and a small number of individuals are heterozygous at all QTL markers. These findings suggest a complex pattern of inheritance, and/or presence of undetected QTL. Notably, our design focused on SNP markers that were predicted to be fixed in either species, and it would be interesting to expand future studies to consider segregating polymorphisms. In a recent comparison of wild and domesticated rabbits, few fixed SNPs were detected in derived alleles (Carneiro *et al.*, 2014). The authors suggest that domestication evolved via polygenic shifts in allele frequency, rather than mutations at a few key loci.

It is likely that additional, undetected loci such as small effect variants contribute to anophelines' SW tolerance. Although random forests correctly classified most individuals as SW tolerant or susceptible, the mean (±1 s.e.) out-of-bag error rate from 10 randomly seeded runs was 28.86% (±0.10), consistent with the existence of other variables not captured by random forests classification. The presence of interacting SNPs can influence the ability for detecting QTL, with even random forests exhibiting low detection rates for interacting SNPs that lack marginal effect contributions (Winham *et al.*, 2012). Advanced crossing designs with greater numbers of generations, and hence more recombination, may improve classification accuracy and power for detecting QTL. Moreover, we mapped SW tolerance as a binary trait, and continuous data typically are more informative than binary data. An inevitable fact of using mortality as our endpoint is that we cannot rule out the possibility that some larvae scored as SW susceptible may have died from lethal hybrid sequence combinations, which could introduce some noise into the QTL mapping. However, preliminary results suggest hybrid inviability loci do not overlap with the SW QTL peaks, at least in males (White, personal observation).

The large widths of some of our QTL intervals may be caused by low resolution from insufficient crossover events in our backcross, and for QTL A, by the low recombination in the 2Rop inversion. Alternatively, the large interval size may reflect the presence of multiple loci within the QTL regions that affect osmoregulation. Another unavoidable limitation in our study is the presence of segregation distortion. Although to be expected to some extent in an interspecific cross with male sterility in hybrids (Davidson, 1964), we cannot exclude the potential for distortion to influence our findings. However, Zhang *et al.* (2010) suggest segregation distortion typically will not increase discovery of false QTL, or substantially impact assessment of QTL effect and position, particularly in large mapping populations. Not only did we use a large mapping population (>1000 individuals), but also we note that distortion was relatively uniform across each chromosome arm.

Both Bayesian interval mapping and random forests indicate epistatic interactions contribute to SW tolerance. The epistatic effect on chromosome 2 at QTL B detected by Bayesian interval mapping is supported by random forests analysis. Random forests suggest disruptive epistasis between QTL B and chromosomes 3 and X. All other interactions between a QTL peak and loci on chromosomes not containing the QTL appear to represent synergistic epistasis. In particular, lower VI scores from omission of the second chromosome suggest synergistic epistasis of chromosome 2 with QTL C, D, E and F. Interactions could occur among genic or non-genic regions, including trans-acting transcription factors. For instance, the single gene within QTL D is a potential ortholog of the *Slowboke binding protein.* In *Drosophila,* this gene is associated with the insulin signaling pathway (Sheldon *et al.*, 2011). Although this gene could be implicated in SW tolerance, an alternative explanation for detection of QTL D is an osmoregulatory role from non-coding regions. Future research and

experimental tests will be needed, but overall our data indicate that SW tolerance involves multiple loci, and epistatic interactions among loci.

The location of QTL A on chromosomal arm 2R overlaps the fixed 2Rop inversion of *An. merus*. *Anopheles coluzzii* possesses the standard chromosomal arrangement in this region, albeit additional inversions within 2Rop bring part of 2Rop into collinear arrangement between the species (Coluzzi *et al.*, 2002). The 2Rop inversion contains several genes that could have a role in osmoregulation, such as sodium/potassium/chloride transporters. Lower recombination rates found within inversions such as 2Rop may serve to limit recombination between suites of co-adapted alleles for SW tolerance, or between regulators of transcription such as cis-acting promoters and their targets. For example, FW and marine sticklebacks have diverged with respect to chromosomal inversions that are proposed to be involved in producing different isoforms of the voltage-gated potassium channel KCNH4, and in general contribute to maintenance of distinct ecotypes (Jones *et al.*, 2012). Future work is needed to test the potential role of inversions in larval ecotype differences. Yet it is an interesting possibility first suggested by Coluzzi. While documenting the multiple and often overlapping inversions on arm 2R that distinguish members of the *An. gambiae* complex, he speculated this region could have a role in oviposition site preference and the different larval habitats of species (Coluzzi *et al.*, 2002).

Osmoregulatory signaling genes, pathways and ion transport

Fine-scale mapping will be a challenging but important step to narrow our QTL intervals for identifying causal variants. Attempts at identifying candidate osmoregulatory genes within the present intervals are highly speculative. Yet as this is the first genome-wide investigation of larval mosquito SW tolerance, we mention a few genes of interest. Two voltage-driven Na^+/H^+ antiporters with proposed roles in anopheline sodium transport, *AgNHA1* and *AgNHA2* (Xiang *et al.*, 2012), are found in QTL A. The gene *Na^+/K^+-ATPase* also occurs in QTL A. This gene is particularly intriguing, because downregulation of *Na^+/K^+-ATPase* in non-DAR cells is hypothesized to lead to sodium excretion by tolerant anophelines placed in SW (Smith *et al.*, 2008). The diuretic peptide *Anoga-DH44* (AGAP003269) (Coast *et al.*, 2005) is yet another candidate in QTL A. Ion channels including two sodium/potassium/chloride channels occur in QTL A, and both sodium and chloride channels are found in QTL F. Scattering of channels across QTL could help explain the synergistic epistasis detected, with multiple channels and transporters contributing to osmoregulation. Osmotic stress responses for diverse organisms from fish (Whitehead *et al.*, 2012) to yeast (Li *et al.*, 2012) involve mitogen-activated protein kinase signaling, and we did detect a handful of mitogen-activated protein kinase genes distributed across QTL A, E and F. Novel candidates exist as well. The QTL A peak contains transcription factor *HNF-4*, and *Atrial natriuretic peptide receptor A* (AGAP003283). In killifish under hypo-osmotic stress, *HNF-4-alpha* emerged as the hub within a network of genes characterized by salinity- and population-dependent expression (Whitehead *et al.*, 2012). Atrial natriuretic peptides are hormones that lower blood pressure by dilating blood vessels, and can initiate sodium excretion and urine production, in part via interaction with Na^+-ATPase (Misono *et al.*, 2011).

Although we have noted some intriguing genes within the QTL peaks, we emphasize that any inferences of candidate osmoregulatory genes (or quantitative trait nucleotides) from our mapping strategy are highly speculative. The single recombination event per chromosome expected between *An. coluzzii* and *An. merus* genotypes within the F_1 mothers of our backcross progeny does not provide sufficient resolution for fine-scale mapping. Several difficulties will make fine-scale QTL mapping a challenge. *Anopheles* naturally mate in swarms, and both this and hybrid male sterility (Davidson, 1964) present difficulties for intercross designs of QTL mapping between *An. coluzzii* and *An. merus*. We attempted crosses with single-pair matings to allow tracking of parental genotypes, but most matings were unsuccessfully and the few F_1 female progeny obtained did not produce viable eggs (HAS and NJB, unpublished). Conducting single-pair matings with anophelines is possible at least within a species (Zheng *et al.*, 1996), but labor intensive. The overlap of QTL peak A with the 2Rop inversion will make it particularly difficult to improve resolution of this region because of the reduced recombination frequency within inversions. Notwithstanding, here we provide insight into the genetic architecture of SW tolerance, and particularly the role of multiple, likely interacting loci across the autosomes and X chromosome.

Besides fine-scale mapping, an important avenue for future research will be investigations of additional species. Expression QTL or other approaches also could be applied to test for cis- vs trans-acting elements. Ultimately, it will be important to validate results from the laboratory colonies with studies of field populations as well. The potential influence of chemoreception on oviposition site choice and salinity preference also will be an interesting area for follow-up research. Meanwhile, we are applying complementary approaches to investigate differential gene expression associated with osmoregulation in *An. merus* and *An. coluzzii*. Not only are kinases within the QTL peaks, but also we have preliminary evidence suggesting genes within a stress-associated mitogen-activated protein kinase pathway are differentially expressed in response to water salinity (HAS, CC and NJB, unpublished data).

CONCLUSION

Considering the mainly coastal distribution of both *An. melas* and *An. merus* despite their ability to survive in both FW and brackish water, SW tolerance may be important to niche diversification and the biogeography of anophelines. With climate change anticipated to cause a rise in sea levels and potentially expand the extent of brackish and coastal habitats (Ramasamy and Surendran, 2011), understanding the ecology of SW-tolerant vector species may become increasingly important to projects aimed at malaria control. Increased salinization of inland waters via changes in climate or anthropogenic influences (for example, removal of water for irrigation) similarly may increase habitat for SW species. The fact that tolerance appears to have arisen multiple times within mosquitoes (Albers and Bradley, 2011; Surendran *et al.*, 2011; Ramasamy *et al.*, 2014), coupled with climate change perhaps altering selection pressure and ecotype distributions, warrants close observation of SW tolerance and its evolutionary origins.

DATA ARCHIVING

Sequence files, and associated data files with genotype and SW tolerance phenotype information, have been deposited at NCBI SRA under accession SRP050537.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank D Ayala for helpful discussions, S Emrich for assistance with pseudo-reference assemblies and M Kern for aid with library construction. Comments from reviewers improved the manuscript. Anopheline eggs were obtained

through the MR4 as part of the BEI Resources Repository, NIAID, NIH: *Anopheles gambiae* Mali-NIH M form, MRA-860, deposited by T Lehmann—NJB, and *Anopheles merus* MAF, MRA-1156, deposited by M Coetzee. This research was funded by National Institutes of Health award 1R21AI101459 to NJB.

Albers MA, Bradley TJ (2011). On the evolution of saline tolerance in the larvae of mosquitoes in the genus *Ochlerotatus*. *Physiol Biochem Zool* **84**: 258–267.

Andolfatto P, Davison D, Erezyilmaz D, Hu TT, Mast J, Sunayama-Morita T *et al.* (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res* **21**: 610–617.

Arribas P, Andújar C, Abellán P, Velasco J, Millán A, Ribera I (2014). Tempo and mode of the multiple origins of salinity tolerance in a water beetle lineage. *Mol Ecol* **23**: 360–373.

Broman KW, Sen Ś (2009). *A Guide to QTL Mapping with R/qtl*. Springer: New York.

Cande J, Andolfatto P, Prud'homme B, Stern DL, Gompel N (2012). Evolution of multiple additive loci caused divergence between *Drosophila yakuba* and *D. santomea* in wing rowing during male courtship. *PLoS One* **7**: e43888.

Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alföldi J, Barrio AM *et al.* (2014). Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**: 1074–1079.

Chen X, Wang M, Zhang H (2011). The use of classification trees for bioinformatics. *Wiley Interdiscip Rev Data Min Knowl Discov* **1**: 55–63.

Coast GM, Garside CS, Webster SG, Schegg KM, Schooley DA (2005). Mosquito natriuretic peptide identified as a calcitonin-like diuretic hormone in Anopheles gambiae (Giles). *J Exp Biol* **208**: 3281–3291.

Cohen E (2013). Chapter one - water homeostasis and osmoregulation as targets in the control of insect pests. In: Ephraim C (ed). *Target Receptors in the Control of Insect Pests: Part 1*. Academic Press: Oxford, Vol 44, pp 1–61.

Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415–1418.

Davidson G (1964). *Anopheles gambiae*, a complex of species. *Bull WHO* **31**: 625–634.

Dow JAT (2009). Insights into the Malpighian tubule from functional genomics. *J Exp Biol* **212**: 435–445.

Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov IV *et al.* (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **437**: 1258524.

Holliday JA, Wang T, Aitken S (2012). Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (Picea sitchensis) using random forest. *G3* **2**: 1085–1093.

Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR *et al.* (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J *et al.* (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Ketterson ED, Nolan V Jr (1999). Adaptation, exaptation, and constraint: a hormonal perspective. *Am Nat* **154**: S4–S25.

Li SC, Diakov TT, Rizzo JM, Kane PM (2012). Vacuolar H+-ATPase works in parallel with the HOG pathway to adapt *Saccharomyces cerevisiae* cells to osmotic stress. *Eukaryot Cell* **11**: 282–291.

Liaw A, Wiener M (2002). Classification and regression by random forest. *R News* **2**: 18–22.

Megy K, Emrich SJ, Lawson D, Campbell D, Dialynas E, Hughes DST *et al.* (2012). VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. *Nucleic Acids Res* **40**: D729–D734.

Misono KS, Philo JS, Arakawa T, Ogata CM, Qiu Y, Ogawa H *et al.* (2011). Structure, signaling mechanism and regulation of the natriuretic peptide receptor guanylate cyclase. *FEBS J* **278**: 1818–1829.

Ramasamy R, Jude PJ, Veluppillai T, Eswaramohan T, Surendran SN (2014). Biological differences between brackish and fresh water-derived *Aedes aegypti* from two locations in the Jaffna Peninsula of Sri Lanka and the implications for arboviral disease transmission. *PLoS One* **9**: e104977.

Ramasamy R, Surendran SN (2011). Possible impact of rising sea levels on vector-borne infectious diseases. *BMC Infect Dis* **11**: 18.

Scott JA, Brogdon WG, Collins FH (1993). Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg* **49**: 520–529.

Sen Ś, Churchill GA (2001). A statistical framework for quantitative trait mapping. *Genetics* **159**: 371–387.

Sharakhov IV, White BJ, Sharakhova MV, Kayondo J, Lobo NF, Santolamazza F *et al.* (2006). Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. *Proc Natl Acad Sci USA* **103**: 6258–6262.

Sharakhova MV, George P, Brusentsova IV, Leman SC, Bailey JA, Smith CD *et al.* (2010). Genome mapping and characterization of the Anopheles gambiae heterochromatin. *BMC Genomics* **11**: 459.

Sheldon AL, Zhang J, Fei H, Levitan IB (2011). SLOB, a SLOWPOKE channel binding protein, regulates insulin pathway signaling and metabolism in Drosophila. *PLoS ONE* **6**: e23343.

Smith KE, Raymond SL, Valenti ML, Smith PJS, Linser PJ (2010). Physiological and pharmacological characterizations of the larval *Anopheles albimanus* rectum support a change in protein distribution and/or function in varying salinities. *Comp Biochem Physiol A Mol Integr Physiol* **157**: 55–62.

Smith KE, VanEkeris LA, Okech BA, Harvey WR, Linser PJ (2008). Larval anopheline mosquito recta exhibit a dramatic change in localization patterns of ion transport proteins in response to shifting salinity: a comparison between anopheline and culicine larvae. *J Exp Biol* **211**: 3067–3076.

Stergiopoulos K, Cabrero P, Davies S-A, Dow JAT (2009). *Salty dog*, an SLC5 symporter, modulates *Drosophila* response to salt stress. *Physiol Genomics* **37**: 1–11.

Surendran SN, Jude PJ, Ramasamy R (2011). Variations in salinity tolerance of malaria vectors of the *Anopheles subpictus* complex in Sri Lanka and the implications for malaria transmission. *Parasit Vectors* **4**: 117.

White BJ, Kundert PN, Turissini DA, Van Ekeris L, Linser PJ, Besansky NJ (2013). Dose and developmental responses of *Anopheles merus* larvae to salinity. *J Exp Biol* **216**: 3433–3441.

Whitehead A, Roach JL, Zhang S, Galvez F (2012). Salinity- and population-dependent genome regulatory response during osmotic acclimation in the killifish (*Fundulus heteroclitus*) gill. *J Exp Biol* **215**: 1293–1305.

Winham SJ, Colby CL, Freimuth RR, Wang X, de Andrade M, Huebner M *et al.* (2012). SNP interaction detection with random forests in high-dimensional genetic data. *BMC Bioinform* **13**: 164.

Xiang MA, Linser PJ, Price DA, Harvey WR (2012). Localization of two Na^+– or K^+–H^+ antiporters, AgNHA1 and AgNHA2, in Anopheles gambiae larval Malpighian tubules and the functional expression of AgNHA2 in yeast. *J Insect Physiol* **58**: 570–579.

Yandell BS, Mehta T, Banerjee S, Shriner D, Venkataraman R, Moon JY *et al.* (2007). R/qtlbim: QTL with Bayesian interval mapping in experimental crosses. *Bioinformatics* **23**: 641–643.

Yi N, Shriner D, Banerjee S, Mehta T, Pomp D, Yandell BS (2007). An efficient Bayesian model selection approach for interacting quantitative trait loci models with many effects. *Genetics* **176**: 1865–1877.

Zhang L, Wang S, Li H, Deng Q, Zheng A, Li S *et al.* (2010). Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theor Appl Genet* **121**: 1071–1082.

Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC (1996). An integrated genetic map of the African human malaria vector mosquito, Anopheles gambiae. *Genetics* **143**: 941–952.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)