The genetic architecture of post-zygotic reproductive isolation between *Anopheles coluzzii* and *An. quadriannulatus*

Kevin C. Deitz[1*], Willem Takken[2], Michel A. Slotman[1]

[1]Department of Entomology, Texas A&M University, College Station, Texas, USA
[2]Laboratory of Entomology, Wageningen University and Research, Wageningen, The Netherlands

*** Correspondence:**
Kevin C. Deitz, Department of Entomology, Texas A&M University, 2475 TAMU, 370 Olsen Blvd, College Station, TX 77843, USA, e-mail: kcdeitz@gmail.com

Keywords: *Anopheles gambiae* complex, speciation, hybrid, sterility, inviability, Dobzhansky–Muller incompatibilities, post-zygotic reproductive isolation

Abstract

The *Anopheles gambiae* complex is comprised of eight morphologically indistinguishable species and has emerged as a model system for the study of speciation genetics due to the rapid radiation of its member species over the past two million years. Male hybrids between most *An. gambiae* complex species pairs are sterile, and some genotype combinations in hybrid males cause inviability. We investigated the genetic basis of hybrid male inviability and sterility between *An. coluzzii* and *An. quadriannulatus* by measuring segregation distortion and performing a QTL analysis of sterility in a backcross population. Hybrid males were inviable if they inherited the *An. coluzzii* X chromosome and were homozygous at one or more loci in 18.9 Mb region of chromosome 3. The *An. coluzzii* X chromosome has a disproportionately large effect on hybrid sterility when introgressed into an *An. quadriannulatus* genetic background. Additionally, an epistatic interaction between the *An. coluzzii* X and a 1.12 Mb, pericentric region of the *An. quadriannulatus* 3L chromosome arm has a statistically significant contribution to the hybrid sterility phenotype. This same epistatic interaction occurs when the *An. coluzzii* X is introgressed into the genetic background of *An. arabiensis,* the sister species of *An. quadriannulatus*, suggesting that this may represent one of the first Dobzhansky–Muller

incompatibilities to evolve early in the radiation of the *Anopheles gambiae* species complex. We describe the additive effects of each sterility QTL, epistatic interactions between them, and genes within QTL with protein functions related to mating behavior, reproduction, spermatogenesis, and microtubule morphogenesis, whose divergence may contribute to post-zygotic reproductive isolation between *An. coluzzii* and *An. quadriannulatus.*

Introduction

The study of speciation genetics and genomics has largely focused on the evolution of hybrid dysfunction (Coyne and Orr, 2004; Dobzhansky, 1937) because it represents the first step in the formation of post-zygotic reproductive isolation between diverging populations. The evolution of post-zygotic isolation mechanisms is positively correlated with the time of divergence between two putative species, and is impacted by the number of mutations separating two allopatric lineages at any given time, the number of incompatibilities per mutation, and the fitness of each incompatibility (Orr and Turelli, 2001).

Post-zygotic isolation occurs as Dobzhansky-Müller incompatibilities (DMI) arise between populations evolving in allopatry (Dobzhansky, 1936; Müller, 1940). Upon secondary contact of genetically divergent populations, epistatic loci on parental chromosomes are unable to function effectively in hybrids. The resulting gene products in hybrids can have unpredictable, deleterious epistatic interactions (Maheshwari and Barbash, 2011). Orr (1995) and Orr and Turelli (2001) proposed the "snowball effect" of DMI evolution, which predicts that the number of DMIs accumulating between lineages "snowballs", or increases faster than linearly (see Kalirad and Azevedo, 2017 for explicit testing of the underlying assumptions of this model). The snowball effect has been supported by empirical evidence of the rate of DMI evolution between *Drosophila* (Matute et al., 2010; Matute and Gavin-Smyth, 2014) and *Solanum* sibling species pairs (Moyle and Nakazato, 2010).

DMI can evolve as a result of various mechanisms in either an ecological speciation (divergent evolution due to selection pressures imposed by novel environments) or mutation-order speciation scenarios (fixation of different, incompatible mutations in allopatric populations experiencing the same environmental conditions) (Mani and Clarke, 1990; Nosil and Flaxman, 2011). These mechanisms include mutation-driven co-evolution (e.g. copper tolerance in *Mimulus guttatus*, MacNair, 1983; MacNair and Christie, 1983), gene duplication (e.g. the *OdsH* gene in *Drosophila simulans* × *D. mauritiania* backcross males, Ting et al., 2004; Sun et al., 2004), gene transposition (e.g. the *JYAlpha* gene in D. *melanogaster* × *D. simulans* backcross males, Masley et al., 2006), or a molecular arms race (e.g. divergent NB-LRR alleles in *Arabidopsis thaliana* strains, Bomblies et al., 2007; Bomblies and Weigal, 2007; Bakker et al., 2006). Furthermore, divergence in gene regulatory regions between parental species can result in in allelic imbalance in hybrids due to incompatible *cis*- and/or *trans*-acting factors (Graze et al., 2012), and segregation distortion of parental loci can skew sex ratios of hybrid progeny (Orr and Irving, 2005; Phadnis and Orr, 2009).

The *Anopheles* (*An.*) *gambiae* complex is comprised of eight morphologically indistinguishable species and has emerged as a model system for the study of speciation genetics due to the rapid radiation of its member species over the past two million years (Fontaine et al., 2015). All members of the *An. gambiae* complex are competent vectors of human malaria parasites, though they differ in their host specificity, ranging from highly anthropophilic (*An. gambiae s.s.* and *An. coluzzii*) (Takken and Knols, 1990; Dekker and Takken, 1998; Pates et al., 2001a) to almost entirely zoophilic (*An. quadriannulatus* A) (White, 1974; Gibson, 1996; Dekker and Takken, 1998; but see Pates et al., 2001b). From a vector control stand point, recent advances in the development of transgenic mosquito control methods have made understanding the permeability of species boundaries even more urgent (e.g. Kyrou et al., 2018) because it is important to know whether transgenic alleles have the potential to introgress between target and non-target species.

F1 hybrid female progeny of crosses between species of the *Anopheles gambiae* complex are fertile, but, with the exception of the *An. gambiae* - *An. coluzzii* species pair, males range from semi- to completely sterile (White, 1974; Hunt, 1998). A previous QTL analyses of sterility in bi-directional backcross hybrids identified regions in the *An. coluzzii* genome that cause male and female sterility when introgressed into an *An. arabiensis* genetic background, and vice versa (Slotman et al., 2004; Slotman et al., 2005). One or more loci on the X chromosome contribute to the effect of the X chromosome on the hybrid sterility phenotype in both directions of the cross. Unfortunately, recombination on the X chromosomes is suppressed in hybrids by the *Xag* inversion, which is fixed within *An. coluzzii* but absent in *An. arabiensis.* The *Xag* inversion comprises ~75% of the *An. coluzzii* X chromosome. This limits our ability to map genes to the X chromosome between species in the complex that harbor the *Xag* inversion (*An. gambiae, An. coluzzii,* and *An. merus*), and those that have the standard arrangement, including *An. quadriannulatus*.

Slotman et al. (2004) also identified both partially dominant and partially recessive male sterility QTL on the 2nd and 3rd chromosomes in both backcrosses. Consistent with the Dobzhansky-Muller model (Dobzhansky, 1936; Muller, 1940), and previous results from *Drosophila* (Wu and Beckenback, 1983; Orr and Coyne, 1989) the position of autosomal sterility QTL differs between each backcross. Complete sterility is achieved through the inheritance of at least three to four autosomal sterility loci (Slotman et al., 2004), including a pericentromeric QTL on the 3L chromosome arm. Hybrid male inviability is caused by an incompatibility between the *An. coluzzii* X chromosome and at least one locus on each autosome of *An. arabiensis.*

Female hybrid sterility QTL are present on the third chromosomes of both *An. coluzzii* and *An. arabiensis* (Slotman et al., 2005). A QTL on the second chromosome *An. coluzzii* causes female hybrid sterility when it is introgressed into an *An. arabiensis* genomic background. However, the female sterility phenotype is affected primarily by the interaction of a foreign X chromosome with the two autosomal QTL. Approximately 75% of observed phenotypic variance is due to the X

chromosome, which only represents 8.8% of the genome (Holt et al., 2002). Due to the lack of recombination between X chromosomes in the cross, the number of loci on the X chromosome that effect the hybrid sterility phenotype is unknown. Therefore, the impact of the X on the hybrid sterility phenotype could result from one locus of major effect, or the effects could be spread across multiple loci. A comparison of the data between the male (Slotman et al., 2004) and female (Slotman et al., 2005) *An. coluzzii* × *An. arabiensis* hybrid sterility data shows that there are a greater number of male than female hybrid sterility QTL, supporting previous studies that found male sterility factors to evolve faster than female sterility factors (Charlesworth et al., 1987; Kirkpatrick and Hall, 2004). This is consistent with faster male evolution observed in *Aedes* mosquitoes (Presgraves and Orr, 1998), and the higher numbers of male versus female sterility factors in *Drosophila* (e.g. Hollocher et al., 1996; True et al., 1996; Tao and Hartl, 2003).

In order to better understand how post-zygotic reproductive isolation evolved during the rapid radiation of the of the *An. gambiae* species complex we performed a QTL analysis of male hybrid sterility in a (*An. coluzzii* × *An. quadriannulatus*) F1 × *An. quadriannulatus* backcross, and identified regions of the *An. coluzzii* genome that contribute to hybrid male sterility and inviability when introgressed into an *An. quadriannulatus* genetic background. We discuss how these QTL relate to those identified in the (*An. coluzzii* × *An. arabiensis*) F1 x *An. arabiensis* backcross (Slotman et al., 2004), and the implications these findings have on our understanding of the evolution of post-zygotic reproductive isolation in the *Anopheles gambiae* species complex, and more broadly, in the context of rapid species radiations.

Methods

Backcross and Phenotype Scoring

We performed a backcross between the SUA2La strain of *An. coluzzii* and the SANGUA strain of *An. quadriannulatus* to identify regions of the *An. coluzzii* genome that exhibit DMI with

the *An. quadriannulatus* genome, resulting in F1 male sterility. The *An. coluzzii* strain was established from collections in Suakoko, Libera and the *An. quadriannulatus* strain was established in from populations in Sangwe, Zimbabwe. Both strains have been lab reared for hundreds of generations. We used a standard backcross scheme that has been used in studies to investigate the genetic basis of hybrid sterility and inviability going back to Dobzhansky (1936). F1 *An. coluzzii* × *An. quadriannulatus* (females × males, respectively) hybrids were reared and backcrossed to *An. quadriannulatus* males *en masse*. Consistent with crosses between *An. coluzzii* and *An. arabienesis*, the Xag inversion was expected to suppress recombination between the X chromosomes of *An. coluzzii* and *An. quadriannulatus*.

The *2La* inversion is also fixed between the mosquito strains used in this cross, limiting our mapping resolution in this region of the genome. While the *2La* is polymorphic within *An. coluzzii*, and *An. coluzzii* strains with the standard arrangement (co-linear with *An. quadriannulatus*) are available, we used the same *An. coluzzii* strain as Slotman et al. (2004) to allow a comparison between them.

The testes of male hybrids were dissected and squashed between a microscope slide and a cover slip. Each hybrid male mosquito was assigned a sterility phenotype following Slotman et al. (2004). Sterility scores were primarily based upon the proportion of normal to abnormal sperm. If no sperm was present, testes were often atrophied as well. In some instances, testes were missing completely. Phenotype scores were as follows: (1) normal testes and sperm, (2) slightly abnormal sperm, (3) 50/50 normal and abnormal sperm, (4) mostly abnormal sperm, (5) entirely abnormal sperm, (6) no sperm present and/or severely atrophied testes, (7) no testes present. Phenotyping was conducted by a single observer, without repeat observations. If unambiguous scoring was not possible, or if the reproductive tissues were damaged during dissection, the sample was discarded. Mosquitoes were preserved in 100% ethanol after phenotyping for DNA extraction an genotyping.

DNA sequencing

DNA extractions were performed on a Qiagen Biosprint DNA extraction machine (Qiagen Inc, Valencia, CA). Extracted DNA was suspended in 200µl Qiagen elution buffer and stored at -20ºC. Each mosquito was genotyped using the restriction enzyme-based, genotype-by-sequencing approach according to the multiplexed shotgun genotyping (MSG) approach published by Andolfatto et al. (2011).

Briefly, 10 ng of genomic DNA of each mosquito was digested with the restriction enzyme MseI, which cuts the *An. gambiae* genome on average every 183.2 bp. Samples were digested for three hours at 37˚C, followed by a 20 minute, 65˚C enzyme deactivation step. Non-ligated bar-coded adapters were removed using an isopropanol precipitation, and bar-coded samples with unique adapters were pooled (96 per plate). Pooled samples were precipitated via centrifugation and re-suspended in TE buffer. Next, DNA was bead-purified using an Agencourt AMPure PCR purification kit. Bar-coded, purified DNA was size selected to a range of 250-400 bp using a Pippen Prep and then amplified using a Phusion High-Fidelity PCR kit. Samples were amplified using a common primer, and an indexed, library-specific primer (one per plate) that also incorporates the Illumina flow-cell adapter. Thus, individuals were identifiable to a library prep by their index and to the individual level by their barcoded adapter. Library amplification (PCR) was performed at a range of 12-18 cycles to identify the minimum number necessary to amplify the library to a target of 10-30 ng/uL (post clean up) in an effort to mitigate the incorporation of PCR errors. Amplified libraries were bead-purified using an Agencourt AMPure PCR purification kit. Library size distribution, integrity, and DNA concentration was confirmed by running each library on a NanoDrop and BioAnalyzer prior to sequencing. We phenotyped and sequenced a total of 841 male backcross hybrids. Hybrids were sequenced at the Texas A&M Genomics and Bioinformatics core facility on an Illumina HiSeq 2500 using 125 bp, single-end chemistry to 0.5-2.0x depth per individual.

We also performed pool-seq of ten *An. coluzzii* females and ten *An. quadriannulatus* females in order to identify fixed genetic difference between parental strains that would be diagnostic of the origin of genomic regions introgressed from *An. coluzzii* into *An. quadriannulatus* in the backcross. Libraries were sequenced on a single lane of Illumina HiSeq 2500 using 150 bp, paired-end chemistry to an average depth of 110.5x (*An. coluzzii*) and 66.9x (*An. quadriannulatus*). Additionally, we prepared and sequenced MSG libraries for 192 *An. coluzzii* and 192 *An. quadriannulatus* females to further validate ancestry informative markers and tune the MSG software error model.

Parental pseudo-reference Construction

An. coluzzii pool-seq reads were aligned to the *An. gambiae* AgamP4.4 reference genome using the BWA mem alignment algorithm (Li and Durban, 2010) and SNPs were called using bcftools mpileup (v1.3) (Li et al., 2009). SNPs within three bases of indels were excluded. A pseudo-reference for *An. coluzzii* was created by updating the AgamP4.4 reference genome (vectorbase.org) with SNPs that had read depth between greater than 10 and less than 1000, map quality greater than 20, and a phred-scaled quality score for the alternative allele greater than 26. The *An. quadriannulatus* pseudo-reference was created in a similar fashion, but by aligning it to the *An. coluzzii* pseudo-reference and updating this with fixed SNPs (*An. coluzzii* – *An. quadriannulatus* divergent) that met the thresholds above.

Hybrid Genotyping

We used Trimmomatic version 0.30 (Bolger et al., 2014) to trim MSG sequencing reads by quality by simultaneously soft clipping the reads from both 5' and 3' ends to an average phred quality score of 20, with no single bp in a four bp window below a phred quality score of 20. Only reads $\geq$ 50 bp were retained for alignment and subsequent analyses.

We used the MSG software package (Andolfatto et al., 2011) to align hybrid MSG data to parental pseudo-references and estimate the posterior probabilities of ancestry at genome-wide

SNPs that were informative of parental ancestry. We removed 33 hybrids from this analysis due to low sequencing coverage (<10K reads). Posterior probabilities of ancestry were thinned using pull_thin_tsv.py (https://github.com/dstern/pull_thin). We considered a marker to be informative if its conditional probability of being homozygous differed by more than 0.10 from its neighboring marker. Missing genotypes were imputed using the hidden Markov model implemented in the sim.geno() function of R/qtl (Broman et al., 2003)

Sterility QTL Mapping

We used R/qtl (Broman et al., 2003) to identify redundant, adjacent markers and further thinned our marker set to include only informative markers. It is important to note that this experiment is not marker, but recombination limited. Therefore, many markers are redundant, and we removed these to reduce computational time. Next, we used R/qtl to estimate the genetic map of 808 hybrid males using the Kosambi map function (Kosambi, 1943). We performed quantitative trait locus mapping with the composite interval mapping (CIM) method using 15 markers as co-factors, and also assessed the significance of QTL models using multiple imputation mapping. This method drops one QTL at a time to calculate the probability of QTL locations and their interactions using an ANOVA (Sen and Churchill, 2001). In all analyses, sterility was treated as a normal trait, and we used 1000 permutation replicates to assess the significance of log of odds (LOD) scores.

Ultimately, we took three approaches to identify sterility QTL in the dataset. The first was to scan for QTL using the composite interval mapping and calculating the logarithm of these odds (LOD) score every 1 cM. When calculating LOD scores we controlled for the genetic background of adjacent markers by using 15 marker covariates across the genome. Due to the skew of the phenotype scores toward “normal”, and severe segregation distortion of the X chromosome, we hypothesized that the QTL identified through CIM of the full dataset were in fact comprised of two types: (1) autosomal *An. coluzzii* loci that are incompatible with the *An. quadriannulatus* X, and (2) homozygous *An. quadriannulatus* autosomal loci that interact epistatically with the *An. coluzzii* X in individuals that are

225 otherwise primarily heterozygous at autosomal loci (recovering viability despite the inheritance of a
226 non-recombinant *An. coluzzii* X chromosome). Thus, we also split the dataset into individuals that had
227 inherited the *An. quadriannulatus* X (N = 568) and *An. coluzzii* X (N = 211), respectively, and
228 performed CIM (as above) on these datasets separately.

229 Next, we tested all peaks identified by CIM, regardless of significance or LOD score, in the full
230 dataset, and any peaks that were unique to either the *An. quadriannulatus* X or *An. coluzzii* X datasets,
231 respectively, by adding each peak independently to the multiple imputation QTL model that also
232 contained the highly significant X chromosome QTL. After identifying all peaks that had a significant
233 additive effect in the multiple imputation QTL model, we tested epistatic interactions between each, in
234 a pairwise fashion, by adding each of these interactions to the model independently. This approach
235 allowed us to identify peaks (and their interactions) which, despite having non-significant LOD scores
236 in the CIM permutation test, were found to have a significant impact on sterility when incorporated
237 into the multiple imputation QTL model.

238 We identified candidate sterility genes within QTL by determining the 1.5 LOD interval of all
239 QTL peaks in base pairs, and used these coordinates to download the list of annotated genes within
240 QTL from the AgamP4 genome using the VectorBase (vectorbase.org) genome browser using the
241 "Export Data" function. Next we used the PANTHER classification system (Mi et al., 2019) gene list
242 analysis search function (pantherdb.org) to search for the protein annotations and functions of these
243 genes. Finally, we performed a protein network analysis on the entire gene list using STRING
244 (Szklarczyket al., 2019) in order to identify direct interactions between putative sterility genes.

245 *Identification of Inviability Loci*

246 We identified regions of the *An. coluzzii* genome that result in hybrid inviability when
247 introgressed into an *An. quadriannulatus* genetic background by analyzing segregation distortion
248 of autosomal loci in the *An. quadriannulatus* X chromosome and *An. coluzzii* X chromosome
249 datasets. We used the mean, genome-wide proportion of *An. quadriannulatus* homozygosity in the

full *An. coluzzii* × *An. quadriannulatus* backcross population as the expected homozygosity proportion, assuming Hardy-Weinberg equilibrium. We used a chi-square test to identify loci that deviate significantly (p-value < 6.01e-05 after Bonferroni correction) from expected homozygosity.

Results

Pseudo-reference Construction, Genotyping, and Genetic Map

After whole genome alignment, SNP calling, and variant filtering of the *An. coluzzii* pool-seq data, we generated the *An. coluzzii* pseuo-reference by incorporating 4,391,200 filtered SNPs into the *An. gambiae s.s.* AgamP4 reference genome. To construct the *An. quadriannulatus* pseuo-reference, we updated the *An. coluzzii* pseudo-reference with 4,242,426 *An. quadriannulatus* SNPs that were found to be fixed between the *An. quadriannulatus* and *An. coluzzii* pool-seq samples. MSG data is inherently low coverage, so not all fixed differences between the pool-seq datasets were subsequently used for hybrid genotyping.

We calculated genotype probabilities at 239,420 ancestry informative SNPs for 808 *An. coluzzii* × *An. quadriannulatus* backcross hybrid males. Ultimately, we trimmed the dataset to 860 non-redundant markers (Table 1). The mean proportion of missing data per individual was 0.021 (sd = 0.060, range = 0.0 – 0.445). These markers were used to calculate a genetic map (Figure 1) with Kosambi distances of 206.1 cM for the second chromosome (110.9 Mb), 100.7 cM for the third (95.2 Mb), and 3.7 cM for the X (24.4 Mb) (Table 1, Figure 1). The larger relative genetic map of the second chromosome is likely due to the presence of the *2La* inversion, which is fixed between these species and can significantly increase recombination frequency on the 2R (i.e. Schultz-Redfield Effect, Schultz and Redfield, 1951; Lucchesi and Suzuki, 1968; Stevison et al., 2011), and segregation distortion resulting from male inviability for some genotype combinations (Zuo et al., 2019; Hackett and Broadfoot, 2003).

The genetic length of the X chromosome was expected to be low due to the presence of the *Xag* inversion, which comprises 80% of the X and is fixed for alternate arrangements between these species.

The majority of hybrids (70.3%, N = 568) inherited a non-recombinant *An. quadriannulatus* X chromosome, 26.1% (N = 211) inherited a non-recombinant *An. coluzzii* X, and 3.5% (N = 29) a recombinant X. The mean sterility score (ranging from 1-7, normal to completely sterile) was 2.62 (sd = 2.34). The majority of backcross mosquitoes (64.23%, N =519) were fully fertile (sterility score = 1).

Segregation distortion and inviability

Segregation distortion was previously documented in male and female hybrids between member species of this complex (Slotman et al., 2004; Slotman et al., 2005; Smith et al., 2015). Furthermore, the complete absence of specific genotypes can indicate their inviability. We tested for significant segregation distortion of autosomal loci in the *An. coluzzii* X and *An. quadriannulatus* X datasets, respectively, be comparing the proportion of *An. quadriannulatus* homozygosity at each locus to the genome-wide average in the full dataset. It is important to note that this analysis was performed on a per locus basis, and is agnostic of interactions between loci.

In the *An. quadriannulatus* X dataset, 67.3% (N=296) of chromosome 2 loci and 18.4% (N=72) of chromosome 3 loci showed significant segregation distortion (Figure 2). Loci on the chromosome 2 exhibit both homozygote excess and deficit (heterosis), while only homozygote excess is exhibited on chromosome 3. In the *An. coluzzii* X dataset, 23.9% (N=105) of chromosome 2 loci and 69.6% (N=273) of chromosome 3 loci showed significant segregation distortion. Interestingly, 40.8% of chromosome 3 loci (spanning a 18.9 Mb region) lack homozygotes in all genotyped individuals, indicating individuals that inherited the *An. coluzzii* X and were *An. quadriannulatus* homozygous at loci inside this region were completely inviable (Figure 2). No autosomal heterozygote excess was observed in the *An. coluzzii* X dataset.

Composite Interval Mapping

Due to minimal recombination events, the X chromosome is effectively one large QTL with a LOD score of 203.4 (Table 2). By comparison, none of the autosomal QTL reach a LOD score of 4.0.

No autosomal QTL fall above the CIM 5% significance threshold (LOD=4.31, 1,000 permutations). Despite this, we annotated three QTL peaks on the second chromosome (1.63 cM, 20.87 cM, and 134.26 cM) (Figure 3, Table 2) that contributed significantly to the sterility phenotype when included in the multiple imputation QTL model (see *Multiple Imputation*, below, and Table 4). The second chromosome peak at 1.63 cM (QTL "2R-A") was present in the full and *An. quadriannulatus* X datasets, but not the *An. coluzzii* X dataset (Figure 3). QTL "2L" at 134.26 cM was present in the full and *An. coluzzii* X datasets, but not the *An. quadriannulatus* X dataset. Finally, QTL "2R-B" (20.87 cM) was found in the full and both X chromosome split datasets, suggesting that it contains loci involved in autosomal-autosomal incompatibilities (Figure 3). While additional LOD peaks are present in the *An. coluzzii* X dataset, these were non-significant when included in the multiple imputation QTL model (see *Multiple Imputation*, below) (Figure 3). None of the second chromosome QTL overlap with those found in the *An. coluzzii* × *An. arabiensis* cross (Slotman et al., 2004) (Figure 3, Table 2).

Two QTL were identified on the third chromosome at 53.95 (QTL "3L-A") and 67.46 cM (QTL "3L-B") (Figure 3, Table 2). These are both present in the full and *An. quadriannulatus* X dataset, but not the *An. coluzzii* X dataset, indicating that *An. coluzzii* introgressions into a predominantly *An. quadriannulatus* genetic background are driving this signal. QTL "3L-A" in the *An. coluzzii* × *An. quadriannulatus* dataset overlaps with a QTL in the *An. coluzzii* × *An. arabiensis* dataset (Slotman *et al*, 2004) (Figure 3, Table 2). The 3L-A and 3L-B QTL have seemingly additive-only effects on sterility (see *Multiple Imputation*, below); these QTL were not identified by CIM in the *An. coluzzii* X dataset, suggesting that they result only from *An. coluzzii* introgressions into predominantly *An. quadriannulatus* genetic backgrounds. However, only three hybrid males were sampled that are *An. quadriannulatus* 3L-B homozygous and inherited the *An. coluzzii* X (Figure 4), and each had a sterility score of five or six. No hybrid males were sampled that are *An. quadriannulatus* 3L-A homozygous and inherited the *An. coluzzii* X. The reduction (or absence) of homozygous genotypes at the 3L-A and 3L-B QTL differs significantly from expectation (the percentage of homozygous genotypes across all

autosomal loci, assuming HWE) (Table 3). Thus, these regions of the *An. quadriannulatus* genome interact epistatically with the X and result in almost complete male inviability when introgresesed into an *An. coluzzii* genetic background.

Multiple Imputation

CIM detects QTL by calculating LOD scores at each marker or map location along each chromosome and does this agnostic of other QTL contributions to the phenotype and potential epistatic interactions between QTL. Due to this limitation, we built multiple imputation QTL models to detect the significance of all QTL LOD peaks found during CIM of the full (i.e., not split by X chromosome) dataset. The QTL discussed above, and those reported in Table 2, are only those that contributed significantly to the variance in sterility in the multiple imputation models described herein. We report two QTL models, one that considers only additive effects of QTL on sterility, and one which includes significant epistatic effects between QTL. We took an iterative process during testing QTL models and added QTL locations identified through CIM in a stepwise fashion. After we had identified those that had significant additive effects, we tested all pairwise epistatic effects between these loci. The final "additive only" and "epistasis allowed" models explain 71.66% and 72.29% of the variance in sterility, respectively (Table 4). The vast majority of this variance is explained by the inheritance of the X chromosome. As expected, individuals harboring the *An. coluzzii* X chromosome have on average higher sterility scores because in these backcrosses there are more opportunities for *An. quadriannulatus* autosomal – *An. coluzzii* X incompatibilities than vice versa (Figure 4).

Each of the QTL described had a significant additive effect in both the "additive only" and "epistasis allowed" models (F-test p-value < 0.01) (Table 4). After testing all possible pair-wise interactions between these loci, we found that only the X and 2L (134.26 cM) QTL pair had a significant interaction (F-test p-value = 5.60E-06). In both models, additive effects of autosomal loci, and the epistatic interaction between the X and 2L QTL, each explained less than 1% of the variance

in hybrid sterility. The "epistasis allowed" model explained a higher proportion of the variance and had a higher model LOD (225.78) when compared to the "additive only" model (221.25) (Table 4).

To better understand the direction of the effect of each QTL in the full dataset, we compared the phenotypes of each individual given their genotype (*An. quadriannulatus* homozygous (QQ) vs. *An. quadriannulatus/An. coluzzii* heterozygous (QC)) at autosomal QTL and the X chromosome (Figure 5). Each autosomal locus increases sterility in the heterozygous state, and individuals with the *An. coluzzii* X are significantly more sterile. The interaction plot (Figure 5, "X vs 2L", right panel) between the X and 2L QTL demonstrates that heterozygosity at the 2L QTL alone does not impact sterility when an individual has the *An. quadriannulatus* X, but reduces sterility when inherited with the *An. coluzzii* X. The evidence suggests that each autosomal QTL indicate the location of a (partially) dominant *An. coluzzii* sterility gene that causes incompatibilities with the *An. quadriannulatus* background.

We determined the 1.5 LOD interval of each QTL and, using the base pair positions of our SNP markers, determined the physical map span of each (Table 2). Next, we compared QTL locations to male sterility QTL found in a *An. coluzzii* × *An. arabiensis* cross (Figure 3) (Slotman et al., 2004). We re-analyzed this data using CIM with three marker covariates. No significant QTL on the second chromosome overlap between these crosses. However, the 3L-A QTL in our *An. coluzzii* × *An. quadriannulatus* cross falls within the 3L QTL from the *An. coluzzii* × *An. arabiensis* cross, and the 3L-B QTL overlaps it. In the Slotman et al. (2004) data the third chromosome QTLs are also only present when the hybrid has the backcross (*An. arabiensis* in this case) parent X chromosome (Figure 3). Thus, interestingly this represents a region of the *An. coluzzii* that contributes to sterility when introgressed into both *An. quadriannulatus* and *An. arabiensis* genetic backgrounds.

Candidate Speciation Genes

We identified all annotated genes in the AgamP4 genome within the 1.5 LOD intervals of each QTL using VectorBase (www.vectorbase.org) and found the protein family (if available) of each gene

using PANTHER. Due to the lack of resolution on the X chromosome, we excluded this chromosome from this analysis. We identified 349 genes within all autosomal QTL (2.47 Mb total, Table 2). We identified ten genes with protein functions related to mating, reproduction, testes, and microtubule morphogenesis and function (Table 5) in the full gene set (Table S1).

We performed a protein network analysis on the entire gene list using STRING (Szklarczyk et al. 2019). The full STRING protein interaction results can be found in Supplementary File 1. Many protein products of genes in autosomal QTL regions have direct interactions, and four proteins with functions related to reproduction (bold in Table 5) interact directly in an insular protein network: dynein light chain Tctex-type, intra-flagellar transport protein 74 homolog, tubulin-specific chaperone D, and cytoplasmic dynein 2 heavy chain 1.

Discussion

One or more loci on the X chromosomes of both species have a large impact on hybrid male sterility in the *An. coluzzii* × *An. quadriannulatus* hybrids. In both multiple imputation QTL models presented (with and without X by 2L epistasis) the cumulative effect of the X chromosome accounted for greater than 60 % of the phenotypic variation. These results support the theory of large X effect on hybrid male sterility due to its hemizygosity in the heterogametic sex (*Anopheles* males) where recessive sterility genes are expressed, but cannot discount the role of faster X evolution, which we did not test. The large proportion of variance explained by the X in our *An. coluzzii* × *An. quadriannulatus* backcross contrasts with the *An. coluzzii* × *An. arabiensis* backcross (to *An. arabiensis*) performed by Slotman et al. (2004). Slotman et al. (2004) found that the X chromosome accounted for only 4.7% of the variance in the hybrid male sterility phenotype while each of five autosomal QTL accounted for 3.4% to 20.3% of the variance (Table 4, Slotman et al., 2004). However, the relatively small proportion of the variance explained by the X in the *An. coluzzii* × *An. arabiensis* backcross resulted, at least in

part, from the inviability observed when hybrids inherited the *An. coluzzii* X (only 9.7% of hybrid males inherited the *An. coluzzii* X, Table 1 of Slotman et al., 2004).

When *An. coluzzii* × *An. arabiensis* F1 hybrids were backcrossed to *An. coluzzii,* the X accounted for 39.5% of the variance and additive effects of autosomal QTL accounted for 4.2-7.9% of the variance (Slotman et al., 2004). The differences in the effect size of the X between crosses, and directions of a back cross between two species (Slotman et al., 2004) is not unexpected. First, the X chromosome has different effects on inviability in each direction of a cross. Second, the genetic architecture of DMIs are unique to each species pair. Additionally, DMIs accumulate over time faster than linearly, and the genetic architecture of individual DMIs becomes more complex as additional mutations accumulate (Matute et al., 2010; Matute and Gavin-Smyth, 2014, Moyle and Nakazato, 2010, Orr and Turelli, 2001).

Unfortunately, due to the *Xag* inversion, the entire X chromosome is effectively one QTL, and we have limited resolution to identify candidate genes on this chromosome. While the 1.5 LOD interval of the X chromosome QTL is 8.53 Mb, the entire span of the QTL is 14.2 Mb, or 58% of the X chromosome. However, we determined which autosomal *An. coluzzii* loci contributed to sterility when introgressed into a predominantly *An. quadriannulatus* genomic background, and vice versa, by dividing the dataset by the parental origin of the X chromosome. This analysis revealed that the epistatic interaction between the 2L and the X QTL is driven by the interaction between *An. quadriannulatus* QTL 2L homozygotes and the hemizygous *An. coluzzii* X chromosome (Figures 3, 4). Individuals with this genotype combination are among the most sterile in the cross (Figure 4), and the 2L QTL is absent in the *An. quadriannulatus* X chromosome CIM results (Figure 3).

The 3L-A QTL sterility QTL in our *An. coluzii* × *An. quadriannulatus* cross overlaps entirely with the 3L QTL identified in the *An. coluzzii* × *An. arabiensis* cross (backcrossed to *An. arabiensis,* Figure 2.D of Slotman et al., 2004). No *An. arabiensis* homozygotes were sampled in this region when the hybrid inherited the *An. coluzzii* X chromosome. Thus, the *An. coluzzii* X chromosome interacts

epistatically with this same region of 3L in both *An.* quadriannulatus and *An. arabiensis* to contribute to hybrid male sterility. We hypothesize that the 3L-A QTL harbors an ancestral allele that is shared between *An.* quadriannulatus and *An. arabiensis* that has not been subjected to introgression between *An. coluzzii* and *An. arabiensis*.

Due to the rapid radiation of the *Anopheles gambiae* species complex, incomplete lineage sorting, and ongoing introgression between member species via fertile F1 hybrid females, the topology of the species complex phylogeny, which is based upon the X chromosome, is discordant with the whole-genome topology (Figure 1 of Fontaine et al., 2015). Bi-directional introgression between *An. arabiensis* and the ancestor of *An. gambiae s.s.* + *An. coluzzii* has reduced the genetic distance between the autosomes of these species relative to the X chromosome. As a result, *An. arabiensis* is sister to *An. quadriannulatus* in the X chromosome phylogeny but is sister to *An. gambiae s.s.* + *An. coluzzii* in the whole genome phylogeny.

While gene tree phylogenies vary across the genome due to the reasons outlines above, on the 3L, where we find overlap of sterility QTL between *An. quadriannulatus* and *An. arabiensis* crosses, the most commonly found gene phylogenies are those that group *An. arabiensis* closer to *An. coluzzii* than *An. quadriannulatus* (Figure 3 of Fontaine et al., 2015). If we assume that genes involved in post-zygotic reproductive isolation between this species trio evolved first on the X chromosome between the ancestor of *An. gambiae s.s.* + *An. coluzzii* and the ancestor of *An. arabiensis* + *An. quadriannulatus*, we can hypothesize that the 3L-A QTL shared between these crosses may have an X chromosome-like topology despite the homogenizing effect of introgression between the *An. coluzzii* and *An. arabiensis* autosomes. This topology is found in ~12.6% of 50-kb genomics windows on the 3L (Figure 3 of Fontaine et al., 2015; see "other, A+Q trees").

Our hypothesis that the 3L-A by X chromosome DMI may have been the first to arise between the ancestor of *An. gambiae s.s.* + *An. coluzzii* and the ancestor of *An. arabiensis* + *An. quadriannulatus* can be tested in the following ways. First, this sterility QTL should also exist in a cross between *An.*

448 *gambiae s.s.* and *An. arabiensis* or *An. quadriannulatus,* and second, it should be absent in a cross
449 between *An. arabiensis* and *An. quadriannulatus*. No hybrid sterility or inviability exists between *An.*
450 *coluzzii* and its sister species *An. gambiae s.s.*, so we expect that the same loci will be responsible for
451 reproductive isolation between both of these species and either *An. arabiensis* or *An. quadriannulatus*.
452 A cross between *An. arabiensis* and *An. quadriannulatus* may yield more interesting results; it would
453 help us to understand the order by which individual DMI arose between these species pairs, and how
454 this effects our understanding of the "snowball effect" of DMI evolution in the context of rapidly
455 radiating species. For example, while the 3L-A by X chromosome DMI has been maintained between
456 *An. coluzzii* and both *An. arabiensis* and *An. quadriannulatus*, additional DMI may have evolved in a
457 "snowball effect" type scenario between *An. coluzzii* and *An. quadriannulatus,* where there is no
458 evidence historical autosomal introgression (Fontaine et al., 2015), whereas DMI may have been lost
459 as a result of autosomal introgression between *An. coluzzii* and *An. arabiensis*, or their relative rate of
460 evolution may have been slower.

461 Our candidate sterility gene analysis identified seven genes (out of 349) within autosomal *An.*
462 *coluzzii* × *An. quadriannulatus* hybrid male sterility QTL that had protein functions related to tubulin
463 or microtubule biogenesis, structure, or function. Four of these (AGAP004782, AGAP001219,
464 AGAP010790, and AGAP001229) have experimentally determined, direct interactions with each
465 other, while another interaction is predicted due to co-expression (AGAP004724 and AGAP010790)
466 (Table 5, Supplementary File 1). While these genes are found within autosomal QTL that contribute to
467 a small proportion of variation in the sterility phenotype, one or more could have an interaction with
468 an X chromosome locus that is disrupted during hybrid testes morphogenesis and/or spermatogenesis.
469 In *Drosophila,* spermatogenesis progresses through mitosis of germ cells, meiosis, and finally the
470 differentiation of spermatids into mature sperm (White-Cooper et al., 2009). Infertility in *Drosophila*
471 hybrid males is associated with spermiogenic (post- rather than pre-meiotic) failure associated with
472 sperm individualization and maturation (Wu et al., 1992). Recent work in the *Anopheles gambiae*

473 complex demonstrates that hybrid sterility results in part from pre-meiotic arrest in degenerate testes
474 (Liang et al, 2019).

475 Sperm flagellum elongate from the spermatid head after meiosis. The axoneme is the central
476 component of the sperm flagellum and has a 9+2 structure of a central pair of microtubules surrounded
477 by nine doublet microtubules. Doublet microtubules have an inner and outer dynein arm that serve to
478 attach and detach to neighboring doublet microtubules, and function in the motor activity of the
479 flagellum. The axoneme is surrounded by a mitochondrial sheath that serves to provide ATP and power
480 motor activity. Tektins are located near the junction points of dynein arms and neighboring
481 microtubules (Gagon, 1995; Linck et al., 2016). Seven of the ten reproduction genes have PANTHER
482 protein functions related to tubulin or microtubule biogenesis, structure, or function: gamma-tubulin
483 complex component 5, microtubule-associated protein futsch, tubulin-like protein alpha-4B-related,
484 dynein light chain Tctex-type, intra-flagellar transport protein 74 homolog, tubulin-specific chaperone
485 D, and cytoplasmic dynein 2 heavy chain 1. The latter five genes (bold, Table 5) in this list have direct
486 interactions in the STRING protein interaction network, and may represent a candidate genetic pathway
487 whose disruption results in hybrid male sterility in the *An. coluzzii* × *An. quadriannulatus* cross.
488 Interestingly, another tubulin-specific chaperone, E-like, is required for sperm individualization and
489 male fertility in *Drosophila* (Nuwal et al., 2012).

490 The three genes without tubulin/microtubule functions are: transcription factor ken, protein
491 lingerer, and a checkpoint protein Hus1-like. Mutations in the ken and barbie transcription factor result
492 in malformed genitalia in *Drosophila* (Lukacsovich et al., 2003). lingerer encodes a putative RNA
493 binding protein and *Drosophila* lingerer mutant males fail to withdraw their genitalia after copulation
494 (Kuniyoshi et al., 2002). Finally, Hus1-like is a component of the 9-1-1 checkpoint protein complex,
495 which coordinates DNA damage sensing, cell cycle progression, and DNA repair (Cotta-Ramusino et
496 al., 2010). Future analyses of the molecular evolution of these genes and their regulatory regions

497 between member species of the *An. gambiae* complex may yield insight into the specific alleles
498 responsible for hybrid male sterility between *An. coluzzii*, *An. quadriannulatus*, and *An. arabiensis.*

499 In this study we have determined that the X chromosome accounts for that majority of the
500 variance in the hybrid male sterility phenotype in a cross between *An. coluzzii* and *An. quadriannulatus*.
501 We have also narrowly defined autosomal QTL which have a statistically significant contribution to
502 sterility despite their smaller effect sizes in comparison to the X. The shared, autosomal sterility QTL
503 between the *An. coluzzii* × *An. quadriannulatus* and *An. coluzzii* × *An. arabiensis* backcrosses suggest
504 that these may represent important autosomal regions involved in the post-zygotic reproductive
505 isolation of *An. coluzzii* from its sister species that diverged early in the radiation of the *An. gambiae*
506 species complex. Future analyses into factors contributing to female hybrid sterility in the *An. coluzzii*
507 × *An. quadriannulatus* cross will add to our knowledge of the genetics of speciation in this
508 evolutionarily and medically important group of human malaria mosquitoes.
509

References

Andolfatto, P., Davison, D., Erezyilmaz, D., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21, 610-617.

Bakker, E.G., Toomajian, C., Kreitman, M., and Bergelson, J. (2006). A genome-wide survey of *R* gene polymorphisms in *Arabidopsis. The Plant Cell* 18, 1803-1818.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangle, J.L., and Weigel, D. (2007). Autoimmune response as a mechanism for Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biology* 5, e236.

Bomblies, K., and Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics* 8, 382-393.

Broman, K.W., Wu, H., Sen, S., and Churchill, G.A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889-890.

Charlesworth, B., Coyne, J.A., and Barton, N.H. (1987). The relative raes of evolution of sex chromosomes and autosomes. *The American Naturalist* 130, 113-146.

Cotta-Ramusino, C., McDonald, E.R., Hurov, K., Sowa, M.E., Harper, J.W., and Elledge, S.J. (2010). A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting protein required for ATR signaling. *Science* 332, 1313-1317.

Coyne, J.A., and Orr, H.A. (2004). Speciation. Sinauer Associates, Sunderland, MA. 545 pp.

Dekker, T., and Takken, W. (1998). Responses of *Anopheles arabiensis* Patton and *Anopheles quadriannulatus* Theobald to carbon dioxide, a man and a calf. *Medical and Veterinary Entomology* 12, 136–140.

Dobzhansky, T. (1936). Studies on hybrid sterility, II. Localization of sterility factors in *Drosophila*

Pseudoobscura hybrids. *Genetics* 146, 239-244.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. New York: Columbia University Press.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., *et al.* (2015). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* 347, 1258524-1-6.

Gagon, C. (1995). Regulation of Sperm Motility at the Axonemal Level. *Reproductive Fertility and Development* 7, 847-855.

Gibson, G. (1996). Genetics, ecology and behaviour of anophelines. *Olfaction in Mosquito–Host Interactions* (ed. by G. R.Bock and G.Cardew)., pp. 22–45. Ciba Foundation Symposium 200. John Wiley & Sons, Chichester.

Graze, R.M., Novelo, L.L., Amin, V., Fear, J.M., Casella, G., Nuzhdin, S.V., and McIntyre, L.M. (2012). Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms, and evolution. *Molecular Biology and Evolution* 29, 1521-1532.

Hackett, C., and Broadfoot, L. (2003). Effects of genotyping errors, missing values, and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90, 33-38.

Hollocher, H., and Wu, C.I. (1996). The genetics of reproductive isolation in the *Drosophila simulans* clade: X vs. autosomal effects and male vs. female effects. *Genetics* 143, 1243-1255.

Holt, R.A., *et al.* (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129-149.

Hunt, R.H., Coetzee, M., and Fettene, M. (1998). The *Anopheles gambiae* complex: a new species from Ethiopia. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 92, 231-235.

Kalirad, A., and Azevedo, R.B.R. (2017). Spiraling complexity: A test of the snowball effect in a computational model of RNA folding. *Genetics* 206, 377-388.

560 Kirkpatrick, M., and Hall, D.W. (2004). Male-biased mutation, sex linkage, and the rate of adaptive
561 evolution. *Evolution* 58, 437-440.
562 Kosambi, D.D. (1943). The estimation of map distances from recombination values. *Annals of*
563 *Eugenics* 12, 172-175.
564 Kuniyoshi, H., Baba, K., Ueda, R., Kondo, S., Awano, W., Juni, N., and Yamamoto, D. (2002).
565 lingerer, a *Drosophila* gene involved in initiation and termination of copulation, encodes a set
566 of novel cytoplasmic proteins. *Genetics* 162, 1775-1789.
567 Kyrou, K., Hammond, A., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T., and Crisanti,
568 A. (2018). A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population
569 suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnol* 36, 1062–1066.
570 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
571 transform. *Bioinformatics* 25, 1754-1760.
572 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
573 Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The sequence
574 alignment/map (SAM). format and SAMtools. *Bioinformatics* 25, 2078-2079.
575 Liang, J., and Sharakhov, I.V. (2019) Premeiotic and meiotic failures lead to hybrid male sterility in
576 the *Anopheles gambiae* complex. *Proceedings B* 286, 20191080.
577 Linck, R.Q., Chemes, H., and Albertini D. (2016). The axoneme: the propultive engine of spermatozoa
578 and cilia and associated ciliopathies leading to infertility. *Journal of Assisted Reproduction and*
579 *Genetics* 33, 141-156.
580 Lucchesi, J.C., and Suzuki, D.T. (1968). The interchromosomal control of recombination. *Annual*
581 *Review of Genetics* 2, 53-86.
582 Lukacsovich, T., Yuge K., Awano W., Asztalos Z., Kondo S., Juni N., and Yamamoto, D. (2003).
583 The *ken and barbie* gene encoding a putative transcription factor with a BTB domain and

584 three zinc finger motifs functions in terminalia development of *Drosophila*. *Archives of Insect*
585 *Biochemistry and Physiology* 54, 77-94.
586 MacNair, M.R. (1983). The genetic control of copper tolerance in the yellow monkey flower,
587 *Mimulus guttatus. Heredity* 50, 283-293.
588 Macnair, M.R., and Christie, P. (1983). Reproductive isolation as a pleiotropic effect of copper
589 tolerance in *Mimulus guttatus*? *Heredity* 50, 295-302.
590 Maheshwari, S., and Barbash, D.A. (2011). The genetics of hybrid incompatabilities. *Annual Review*
591 *Genetics* 45, 331-355.
592 Mani, G.S., and Clarke, B.C. (1990). Mutational order: a major stochastic process in evolution.
593 *Proceedings of the Royal Society of London, Biology* 240, 29-37.
594 Masley, J.P., Jones, C.D., Noor, M.A.F., Locke, J., and Orr, H.A. (2006). Gene transposition as a
595 cause of hybrid sterility in Drosophila. *Science* 313, 1448-1450.
596 Matute, D.R., Butler, I.A., Turissini, D.A., and Coyne, J.A. (2010). A test of the snowball theory for
597 the rate of evolution of hybrid incompatibilities. *Science* 329, 1518-1521.
598 Matute, D.R., and Gavin-Smyth, J. (2014). Fine mapping of dominant X-linked incompatibility
599 alleles in *Drosophila* hybrids. *PLoS Genetics* 11, e1005558.
600 Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019)
601 Protocol Update for large-scale genome and gene function analysis with the PANTHER
602 classification system (v.14.0). *Nature Protocols* 14, 703-721.
603 Moyle, L.C., and Nakazato,T. (2010). Hybrid incompatibility "snowballs" between *Solanum* species.
604 *Science* 329, 1521-1523.
605 Müller, H.J. (1940). Bearing of the *Drosophila* work on systematics. In *The New Systematics*, ed. JS
606 Huxley, pp. 185-268. Oxford, UK. Clarendon Press.
607 Nosil, P., and Flaxman, S.M. (2011). Conditions for mutation-order speciation. *Proceedings of the*
608 *Royal Society of London, Biology* 278, 399-407.

609 Nuwal, T., Kropp, M,. Wegener, S., Racic, S., Montalban, I., and Buchner, E. (2012). The *Drosophila*
610 homologue of tubulin-specific chaperon E-like protein is required for synchronous sperm
611 individualization and normal male fertility. *Journal of Neurogenetics* 26, 374-381.
612 Orr, H.A., and Coyne ,J.A. (1989). The genetics of postzygotic isolation in the *Drosophila virilis*
613 group. *Genetics* 121, 527-537.
614 Orr, H.A., and Irving, S. (2005). Segregation distortion in hybrids between the Bogota and USA
615 subspecies of *Drosophila pseudoobscura*. *Genetics* 169, 671-682.
616 Orr, H.A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities.
617 *Genetics* 139, 1805–1813.
618 Orr, H.A., and Turelli, M. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-
619 Muller incompatibilities. *Evolution* 55, 1085-1094.
620 Pates, H.V., Takken, W., Curtis, C.F., Huisman, W., Akinpelu, O., and Gill, G.S. (2001b).
621 Unexpected anthropophagic behavior in *Anopheles quadriannulatus. Medical and Veterinary*
622 *Entomology* 15, 293-298.
623 Pates, H.V., Takken, W., Stuke, K., and Curtis, C.F. (2001a). Differential behaviour of *Anopheles*
624 *gambiae sensu stricto* (Diptera: Culicidae) to human and cow odours in the laboratory.
625 *Bulletin of Entomological Research* 4, 289-296.
626 Phadnis, N., and Orr, H.A. (2009). A single gene causes both male sterility and segregation distortion
627 in *Drosophila* hybrids. *Science* 323, 376-379.
628 Presgraves, D.C., and Orr, H.A. (1998). Haldane's rule in taxa lacking a hemizygous X. *Science* 282,
629 952-954.
630 Schultz, J., and Redfield, H. (1951). Interchromosomal effects on crossing over in *Drosophila*. *Cold*
631 *Spring Harbor Symposium on Quantitative Biology* 16, 175-197.
632 Sen, S., Churchhill, G.A. (2001). A statistical framework for quantitative trait mapping. *Genetics*
633 159, 371-387.

634 Slotman, M.A., Della Torre, A., and Powell, J.R. (2004). The genetics of inviability and male sterility
635 in hybrids between *Anopheles gambiae* and *Anopheles arabiensis* (2004). *Evolution* 59, 1016-
636 1026.
637 Slotman, M.A., della Torre, A., and Powell, J.R. (2005). Female sterility in hybrids between
638 *Anopheles gambiae* and *A. arabiensis*, and the cause of Haldane's Rule. *Evolution* 59, 1016-
639 1026.
640 Smith, H.A., White, B.J., Kundert, P., Cheng, C., Romero-Severson, J., Andolfatto, P., and Besansky,
641 N.J. (2015). Genome-wide QTL mapping of saltwater tolerance in sibling species of
642 *Anopheles* (malaria vector). mosquitoes. *Heredity* 115, 471-479.
643 Stevison, L.S., Hoehn, K.B., and Noor, M.A. (2011). Effects of inversions on within- and between-
644 species recombination divergence. *Genome Biology and Evolution* 3, 830-841.
645 Sun, S., Ting, C.T., and Wu, C.I. (2004). The normal function of a speciation gene, *Odysseus*, and its
646 hybrid sterility effect. *Science* 305, 81-83.
647 Szklarczyk, D., Gable, A.L,. Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M.,
648 Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., and von Mering, C. (2019) STRING
649 v11: protein-protein association networks with increased coverage, supporting functional
650 discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, 607-613.
651 Takken, W., and Knols, B.G.J. (1990). Flight behaviour of *Anopheles gambiae* Giles (Diptera:
652 Culicidae) in response to host stimuli: a windtunnel study. *Proceedings of Experimental and*
653 *Applied Entomology* 1, 121–128.
654 Tao, Y., and Hartl, D.L. (2003). Genetic dissection of hybrid incompatibilities between *Drosophila*
655 *simulans* and *D. mauritiana.* III Heterogeneous accumulation of hybrid incompatibilities,
656 degree of dominance, and implications for Haldane's rule. *Evolution* 57, 2580-2598.
657 Ting, C.T., Tsaur, S.C., Sun, S., Browne, W.E., Chen, Y.C., Patel, N.H., and Wu, C.I. (2004). Gene
658 duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. *Proceedings of*

659 *the National Academy of Sciences, USA* 101, 12232-12235.

660 True, J.R., Weir, B.S., and Laurie, C.C. (1996). A genome-wide survey of hybrid incompatibility
661 factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into
662 *Drosophila simulans. Genetics* 142, 819-837.

663 White, G.B. (1974). Anopheles gambiae complex disease transmission in Africa. *Transactions of the*
664 *Royal Society of Tropical Medicine & Hygiene* 68, 278-298.

665 White-Cooper, H., Doggett, K., and Ellis, R.E. (2009). The evolution of spermatogenesis. In: Sperm
666 Biology: An Evolutionary Perspective (TR Birkhead, DJ Hosken & S Pitnick, eds)., 151–183.
667 Academic Press, Burlington, MA.

668 Wu, C.I., Perez, D.E., Davis, A.W., Johnson, N.A., Cabot, E.L., Palopoli, M.F., and Wu, M.L. (1992).
669 Molecular genetic studies of post-mating reproductive isolation in *Drosophila*. In: Takahata N,
670 Clark AG (eds). Molecular Paleo-population Biology. Springer-Verlag, Berlin, 191–212.

671 Wu, C.I., and Beckenback, A.T. (1983). Evidence for extensive genetic differentiation between the
672 sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *Drosophila*
673 *persimilis* and identification of hybrid sterility factors. *Genetics* 105, 71-86.

674 Zuo, J., Niu, Y., Cheng, P., Feng, J.Y., Han, S.F., Zhang, Y.H., Shu, G., Wang, Y., Zhang, and Y.M.
675 (2019). Effect of marker segregation distortion on high density linkage map construction and
676 QTL mapping in Soybean (*Glycine max* L.). *Heredity* 123, 579–592.

677

678 **Conflict of Interest**

679 The authors declare that the research was conducted in the absence of any commercial or
680 financial relationships that could be construed as a potential conflict of interest.

681 **Author Contributions**

682 K.C. Deitz, W. Takken and M.A. Slotman designed the experiment. K.C. Deitz and M.A.
683 Slotman, performed the experiment. K.C. Deitz analysed the data. K.C. Deitz and M.A. Slotman
684 wrote the manuscript. W Takken edited the manuscript.

685 **Funding**

686 This study was funded by a National Science Foundation Doctoral Dissertation Improvement
687 Grant (award # 1601675) to K.C. Deitz and M.A. Slotman, a Texas A&M University Genomics Seed
688 Grant to K.C. Deitz and M.A. Slotman, and a Texas A&M University Dissertation Fellowship to
689 K.C. Deitz.

690 **Acknowledgments**

691 We thank Patrick Reilly and Andrew Taverner for helpful discussions regarding the analysis
692 of the dataset and Leon Westerd for assistance with mosquito rearing. We are also thankful to BEI
693 Resources for providing mosquito strains.

694 **Supplementary Material**

695 Supplementary tables can be found in Deitz_et_al_2020_Supplementary_Tables.docx.

696 **Data Availability Statement**

697 Raw, whole-genome sequencing data of mosquito colonies used in this study to generate
698 pseudo-references are available on the NCBI Sequence Read Archive at project accession number
699 PRJNAXXXXX (data will be uploaded upon acceptance of the manuscript). Genotype data for
700 hybrids can be accessed on Dryad (https://doi.org/XX.XXXX/dryad.XXXXX) (data will be uploaded
701 upon acceptance of the manuscript).

Figures

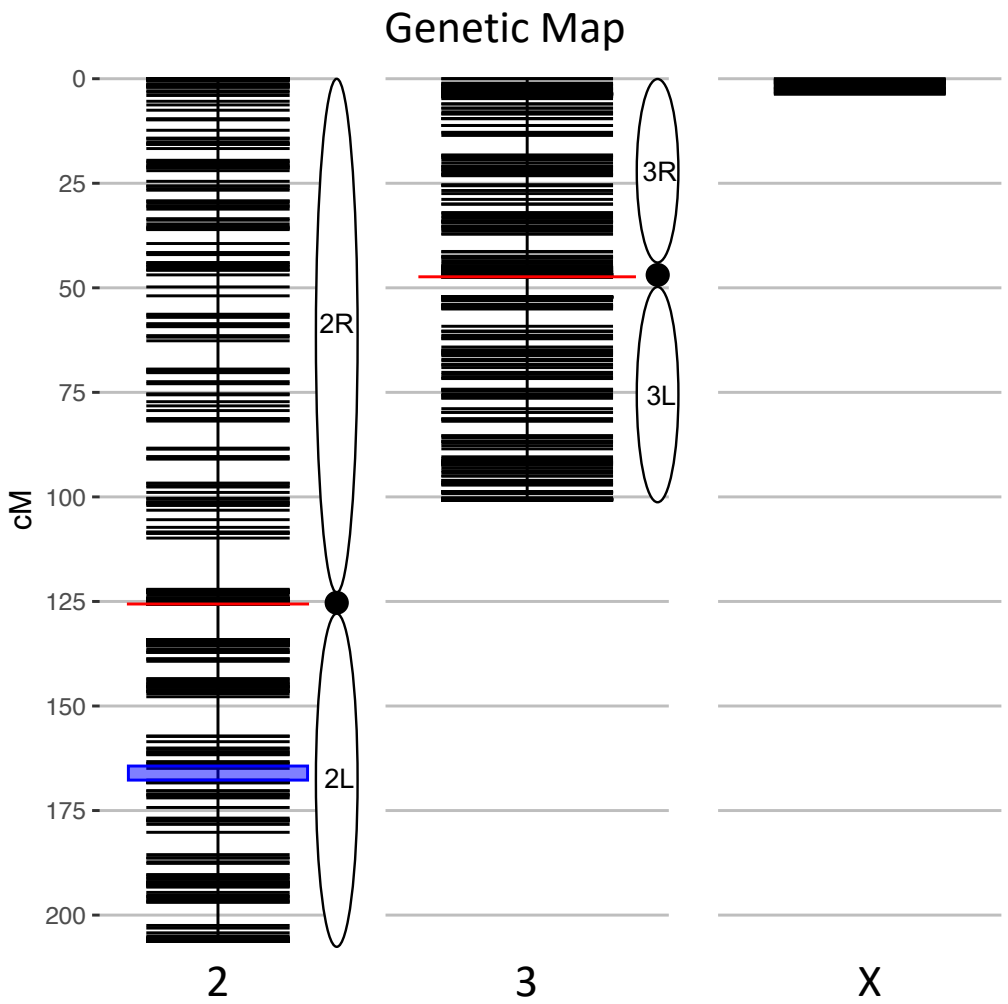

Figure 1. Genetic map of the *An. coluzzii* × *An. quadriannulatus* backcross showing centromere locations (red horizontal lines) and marker locations (black horizontal lines). The centromere of the telocentric X chromosome is opposite zero cM. The blue highlighted area on chromosome 2 indicates the location of the 2L*a* inversion in *An. coluzzii*.

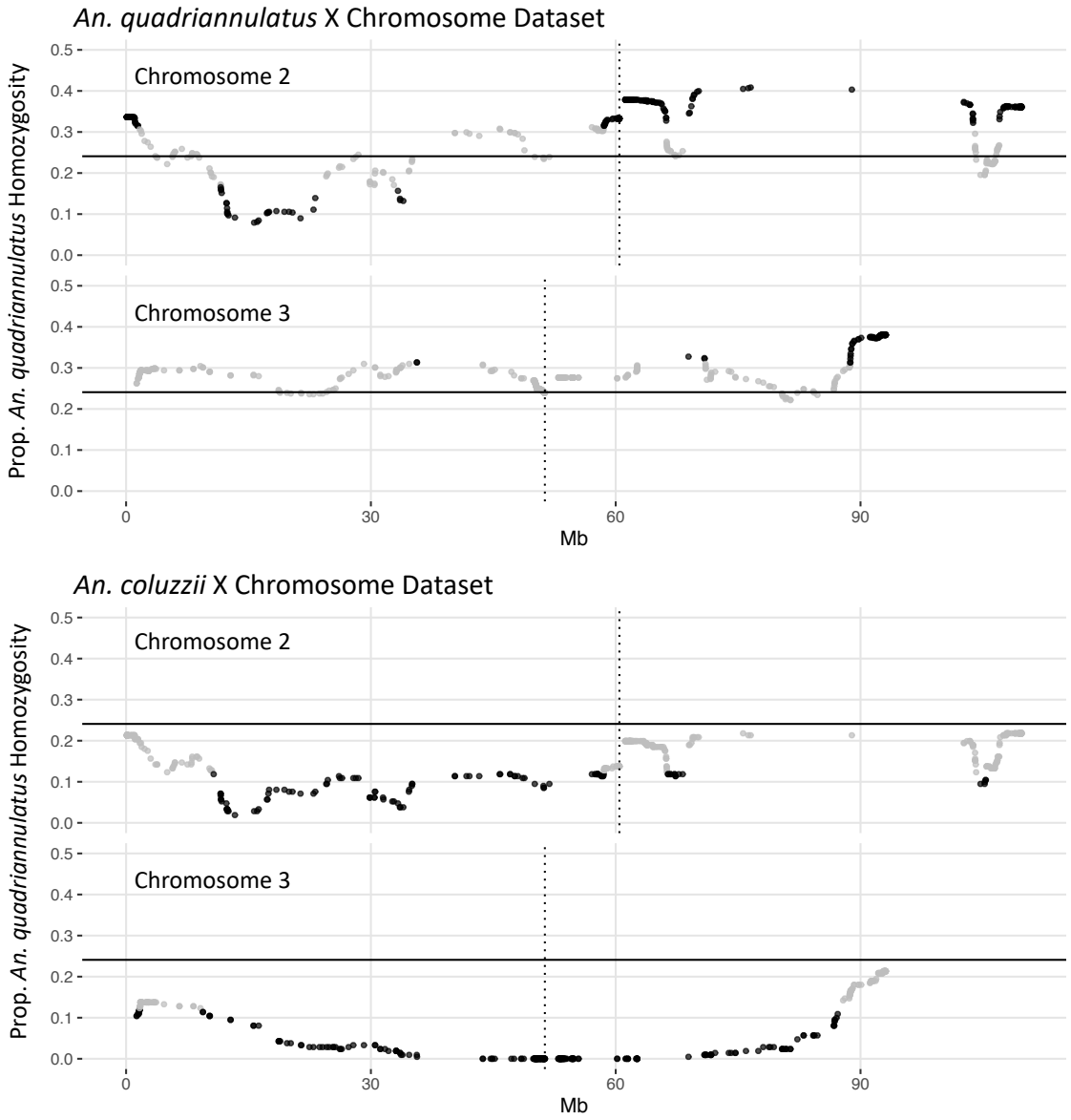

Figure 2. Segregation distortion at chromosome 2 and 3 loci for the *An. quadriannulatus* X chromosome (top) and *An. coluzii* X chromosome dataset (bottom). Grey dots indicate SNPs that do not deviate significantly from the expected proportion of homozygosity (solid horizontal line), while black dots deviate significantly (chi-square p-val < 0.05). Vertical dotted lines indicate centromere locations.

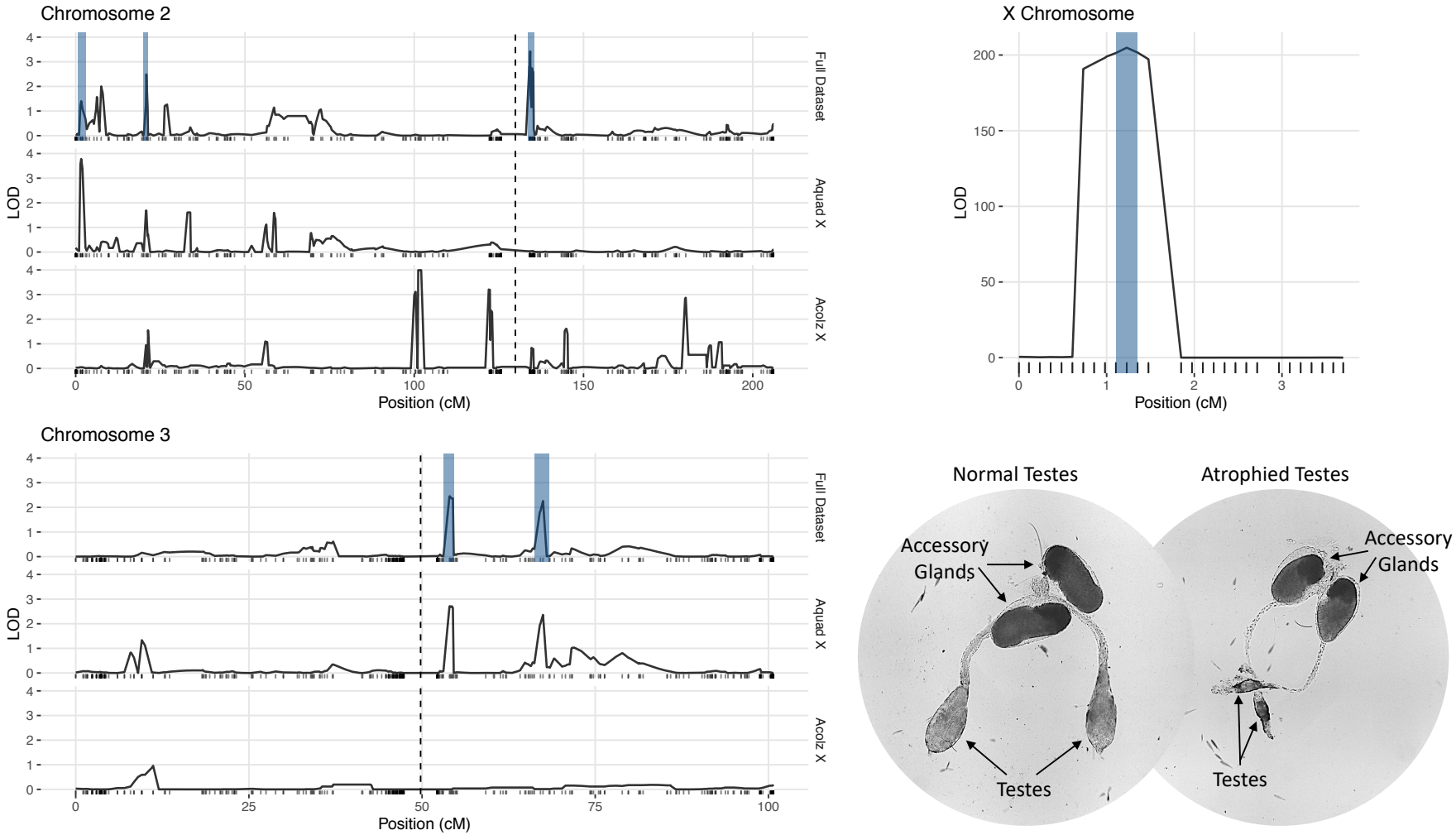

Figure 3. LOD plots for hybrid male sterility QTL identified with composite interval mapping in the *An. coluzzii* × *An. quadriannulatus* cross. Note the differences in the y-axis scales between the chromosome 2 (top left) and 3 (bottom left) panels, and the X chromosome panel (top right). Vertical, dashed lines indicate centromere locations on chromosomes 2 and 3. Second and third chromosome right arms are left of the centromere, and left arms are to the right. Vertical blue bars indicate 1.5 LOD intervals of QTL that explain a significant proportion of phenotypic variation in multiple imputation QTL model of the full dataset. The bottom right panel illustrates the gross morphological differences between hybrid testes with a normal sterility score (1) and a atrophied hybrid testes (sterility score = 6) which contain no sperm or only a few sperm with abnormal morphology.

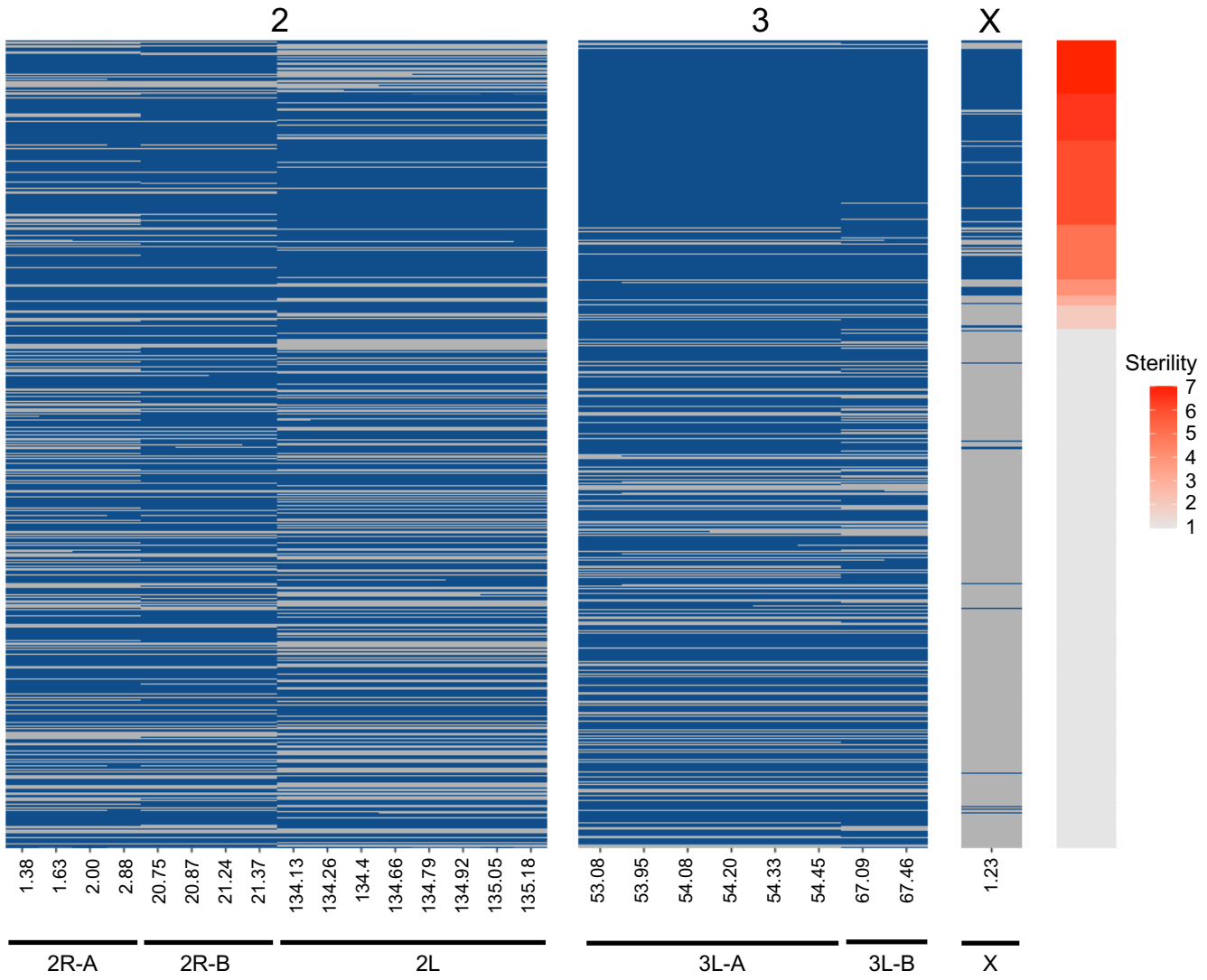

Figure 4. Genotype heat map for each *An. coluzzii* × *An. quadriannulatus* backcross male hybrid (y-axis) at the six QTL (x-axis) identified using CIM and multiple imputation. All markers within the 1.5 LOD interval of each QTL peak (Table 2) are included (x-axis, positions in cM). Individuals are sorted on the y-axis by sterility, with the individuals with the highest sterility score (7) at the top. Blue cells indicate *Aquad/Acolz* heterozygosity on the autosomes, and *Acolz* hemizygosity on the X chromosome. Grey cells indicate *Aquad/Aquad* homozygosity on the autosomes, and *Aquad* hemizygosity on the X chromosome.

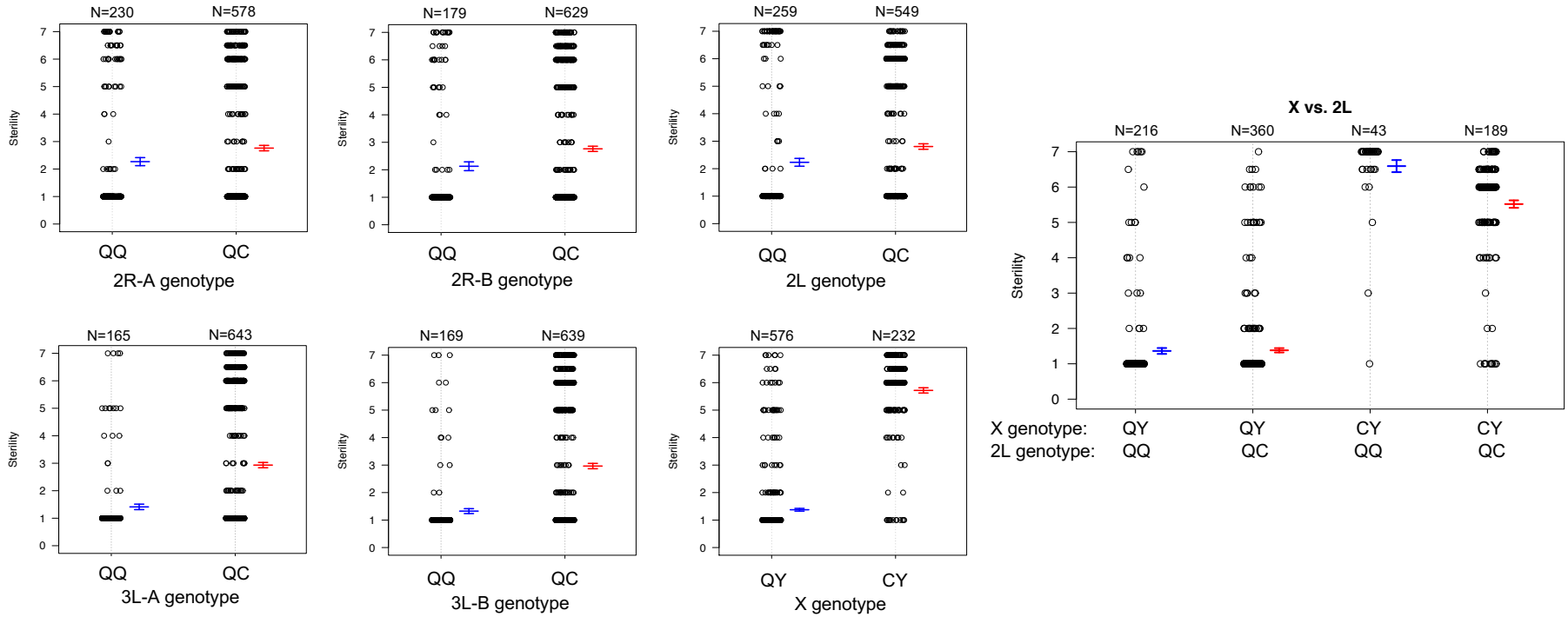

Figure 5. Genotype (x-axis) by phenotype (sterility score, y-axis) plots for the six QTL identified using CIM and multiple imputation, and an interaction plot (right) for the epistatic effect between the X chromosome and 2L QTL. Each individual is represented as a dot, which are jittered to avoid overlap. Box plots illustrate the mean (middle line) of each phenotype distribution and whiskers illustrate +/- 1 standard error of the mean. QTL names and locations correspond to those in Table 2. QQ: *An. quadriannulatus* homozygous, QC: *An quadriannulatus / An. coluzzii* heterozygous.

Tables

Table 1. The number of ancestry informative, thinned, and non-duplicate markers genotyped in hybrids, and the physical and map distances of each chromosome in the cross.

Chromosome	Informative Markers	Thinned Markers	Non-duplicate markers	Physical Length (Mb)	Genetic Length (cM)
2	120,717	710	440	110.9	206.1
3	97,706	629	392	95.2	100.7
X	20,997	217	28	24.4	3.7

Table 2. Genomic regions of QTL in physical (Mb) and map (cM) distances. Start an end positions indicate the 1.5 LOD intervals of QTL locations (LOD peak). Bold values indicate overlapping QTL between the crosses on the 3L, a region of the *An. coluzzii* genome that contributes to hybrid male sterility when introgressed into both *An. quadriannulatus* and *An. arabiensis* genetic backgrounds.

Cross	QTL	LOD	LOD peak (cM)	Start (cM)	End (cM)	Length (cM)	LOD peak (Mb)	Start (Mb)	End (Mb)	Length (Mb)	Num Genes
An. coluzzii × *An. quadriannulatus*	X	204.84	1.23	1.11	1.35	0.25	6.68	2.96	11.48	8.53	453
	2R-A	3.78	1.63	0.63	2.88	2.26	1.15	1.03	1.67	0.63	49
	2R-B	2.49	20.87	20.12	21.37	1.24	8.11	8.03	8.73	0.70	17
	2L	3.42	134.26	134.13	135.30	1.18	2.69	0.73	4.14	3.42	153
	3L-A	2.45	53.95	53.08	54.57	1.49	11.18	**10.18**	**11.30**	1.13	52
	3L-B	2.26	67.46	66.21	68.34	2.12	22.82	20.86	23.33	2.47	78
An. coluzzii × *An. arabiensis* (Slotman *et al., 2004*)	2R	38.30	16.07	15.25	18.05	2.80	40.44	33.18	48.30	14.83	-
	3R	3.69	12.82	8.41	19.04	10.63	14.83	8.79	44.24	23.58	-
	3L	7.03	23.23	19.04	26.08	7.04	16.86	**44.24 (3R)**	**19.77**	38.37	-

Table 3. Percentage of *An. quadriannulatus* homozygosity at hybrid male sterility autosomal QTL peaks in the full, *An. quadriannulatus* X chromosome, and *An. coluzzii* X chromosome datasets. p-values are reported for chi-square tests, which were used to test for significant deviation (p-value < 6.01e-05 after Bonferroni correction) from expected homozygosity. Expected homozygosity was calculated as the mean homozygosity percentage across all autosomal loci in the full dataset, assuming Hardy-Weinberg equilibrium (0.2409).

		Full dataset		***An. quadriannulatus* X**		***An. coluzzii* X**	
QTL	**cM**	**Homozygosity (%)**	**χ^2 p-val**	**Homozygosity (%)**	**χ^2 p-val**	**Homozygosity (%)**	**χ^2 p-val**
2R-A	1.63	28.5	3.7E-03	32.0	9.5E-06	20.4	2.1E-01
2R-B	20.87	22.2	2.0E-01	24.8	6.8E-01	16.1	6.7E-03
2L	134.26	32.1	1.2E-07	37.7	3.8E-14	19.9	1.5E-01
3L-A	53.95	20.4	1.5E-02	29.0	5.8E-03	0.0	2.7E-16
3L-B	67.46	20.9	3.5E-02	29.0	5.8E-03	1.4	1.3E-14

Table 4. Results of two multiple imputation QTL models (ANOVA). The top model includes only additive effects of each QTL. The bottom model includes all additive effects and also an epistatic effect between the X chromosome and the 2L QTL.

QTL	cM	LOD	% Variance	F value	p-value (F)
X	1.23	203.34	61.96	1751.36	2.00E-16
2R-A	1.63	2.47	0.40	11.36	7.85E-04
2R-B	20.87	2.89	0.47	13.30	2.82E-04
2L	134.26	1.49	0.24	6.85	9.04E-03
3L-A	53.95	1.66	0.27	7.62	5.90E-03
3L-B	67.46	2.46	0.40	11.30	8.12E-04
Model	y ~ X + 2R-A + 2R-B + 2L + 3L-A + 3L-B				
LOD	221.25				
% Variance	71.66				

QTL	cM	LOD	% Variance	F value	p-value (F)
X	1.23	207.86	62.68	907.89	2.00E-16
2R-A	1.63	2.73	0.43	12.52	4.25E-04
2R-B	20.87	3.02	0.48	13.87	2.10E-04
2L	134.26	6.02	0.96	13.96	1.10E-06
3L-A	53.95	1.62	0.26	7.40	6.66E-03
3L-B	67.46	2.15	0.34	9.87	1.74E-03
X × 2L epistatic		4.53	0.72	20.90	5.60E-06
Model	y ~ X + 2R-A + 2R-B + 2L + 3L-A + 3L-B + X:2L				
LOD	225.78				
% Variance	72.39				

Table 5. Annotated genes in the AgamP4 genome within autosomal QTL with functions related to mating, reproduction, and sperm motility and morphogenesis. Gene names in bold indicate those interacting in a regulatory network associated with sperm microtubule morphogenesis and function. STRING network analysis output can be found in Supplementary File 1: Deitz_et_al_2020_String_protein_interaction_network.xlsx.

QTL	Gene Name	Panther Gene Ontology Family / Subfamily	Protein Description	Protein Description Source
2R-A	**AGAP001229**	DYNEIN LIGHT CHAIN TCTEX-TYPE (PTHR21255:SF4)	Tctex-1 (t-complex testis expressed-1) is a dynein light chain. Dynein translocates rhodopsin-bearing vesicles along microtubules. In mice, the chromosomal location and pattern of expression of Tctex-1 make it a candidate for involvement in male sterility.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR005334/
2R-A	AGAP001227	GAMMA-TUBULIN COMPLEX COMPONENT 5 (PTHR19302:SF33)	The microtubule organizing centers (MTOCs) of eukaryotic cells are the sites of nucleation of microtubules, and are known as the centrosome in animal cells and the spindle pole body in yeast. Gamma-tubulin, which is 30% identical to alpha and beta tubulins that form microtubules, appears to be a key protein involved in nucleation of microtubules.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR007259/
2R-A	AGAP001194	MICROTUBULE-ASSOCIATED PROTEIN FUTSCH (PTHR13843:SF12)	futsch (futsch) encodes a microtubule binding protein involved in the formation of synaptic buttons at the neuromuscular junctions. Its expression levels are regulated by the product of TBPH to prevent synaptic defects.	https://flybase.org/reports/FBgn0259108

2R-A	**AGAP001219**	TUBULIN-LIKE PROTEIN ALPHA-4B-RELATED (PTHR11588:SF239)	Microtubules are polymers of tubulin, a dimer of two 55kDa subunits, designated alpha and beta. Within the microtubule lattice, alpha-beta heterodimers associate in a head-to-tail fashion, giving rise to microtubule polarity. Fluorescent labelling studies have suggested that tubulin is oriented in microtubules with beta-tubulin toward the plus end.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR002452/
2L	AGAP004767	TRANSCRIPTION FACTOR KEN (PTHR45993:SF7)	Mutations in the ken and barbie locus are accompanied by the malformation of terminalia in adult Drosophila. Male and female genitalia often remain inside the body, and the same portions of genitalia and analia are missing in a fraction of homozygous flies. Rotated and/or duplicated terminalia are also observed. Terminalia phenotypes are enhanced by mutations in the gap gene tailless, the homeobox gene caudal, and the decapentaplegic gene that encodes a TGFbeta-like morphogen. The ken and barbie gene encodes a protein with three CCHH-type zinc finger motifs that are conserved in several transcription factors such as Krüppel and BCL-6.	https://flybase.org/reports/FBgn0011236
2L	**AGAP004724**	INTRAFLAGELLAR TRANSPORT PROTEIN 74 HOMOLOG (PTHR31432:SF0)	Intraflagellar transport of ciliary precursors such as tubulin from the cytoplasm to the ciliary tip is required for ciliogenesis. Intraflagellar transport protein 74 (IFT74) is a component of the intraflagellar transport (IFT) complex B, which contains at least 20 different protein subunits. Together with IFT81, it forms a tubulin-binding module that specifically mediates transport of tubulin within the cilium.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR029602/
2L	AGAP004817	PROTEIN LINGERER (PTHR16308:SF13)	lingerer (lig) encodes a putative RNA binding protein that forms a complex with the products of orb and orb2. Loss of lig results in defects in copulation and short-term memory in *Drosophila*.	https://flybase.org/reports/FBgn0020279

2L	**AGAP004782**	TUBULIN-SPECIFIC CHAPERONE D (PTHR12658:SF0)	The tubulin heterodimer consists of one alpha- and one beta-tubulin polypeptide. Tubulin-specific chaperones are essential for bring the alpha- and beta-tubulin subunits together into a tightly associated heterodimer.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR033162/
3L-A	**AGAP010790**	CYTOPLASMIC DYNEIN 2 HEAVY CHAIN 1 (PTHR10676:SF352)	Cytoplasmic dynein 2 heavy chain 1 (DYNC2H1) is a subunit of the cytoplasmic dynein complex 2. It may function as a motor for intraflagellar retrograde transport. It also plays a role in cilia biogenesis.	https://www.ebi.ac.uk/interpro/entry/InterPro/IPR026815/
3L-B	AGAP011361	CHECKPOINT PROTEIN (PTHR12900:SF0)	Hus1-like (Hus1-like) encodes a protein that together with proteins encoded by Rad1 and Rad9 form the 9-1-1 checkpoint protein complex. This complex plays a central role in the DNA damage-induced checkpoint response. Involved in G2/M transition of mitotic cell cycle, meiotic and mitotic nuclear division.	https://flybase.org/reports/FBgn0026417

754

Table S1. The complete list of annotated genes within autosomal QTL and their PANTHER gene ontology families / sub-families (if available).

QTL	Gene Name	Panther Gene Ontology Family / Subfamily
2R A	AGAP001195	NUCLEAR CAP-BINDING PROTEIN SUBUNIT 1 (PTHR12412:SF2)
2R A	AGAP001226	
2R A	AGAP001189	GEO07581P1-RELATED (PTHR11857:SF43)
2R A	AGAP001207	COMM DOMAIN-CONTAINING PROTEIN 4 (PTHR16231:SF4)
2R A	AGAP001202	FI24011P1-RELATED (PTHR24064:SF335)
2R A	AGAP001206	
2R A	AGAP013230	
2R A	AGAP013386	DNA POLYMERASE ZETA CATALYTIC SUBUNIT (PTHR45812:SF1)
2R A	AGAP001215	TETRATRICOPEPTIDE REPEAT PROTEIN 39C (PTHR31859:SF1)
2R A	AGAP001199	AT20289P-RELATED (PTHR24276:SF78)
2R A	AGAP001196	G KINASE-ANCHORING PROTEIN 1 (PTHR14899:SF0)
2R A	AGAP001200	GLYCOGEN DEBRANCHING ENZYME (PTHR10569:SF2)
2R A	AGAP013011	TRICHOPLEIN KERATIN FILAMENT-BINDING PROTEIN (PTHR31183:SF2)
2R A	AGAP013543	
2R A	AGAP001217	SOLUBLE CALCIUM-ACTIVATED NUCLEOTIDASE 1 (PTHR13023:SF3)
2R A	AGAP013056	
2R A	AGAP001225	
2R A	AGAP001223	ATP-DEPENDENT RNA HELICASE DDX55 (PTHR24031:SF2)
2R A	AGAP001208	ARGININE/SERINE-RICH COILED-COIL PROTEIN 2 (PTHR22426:SF2)
2R A	AGAP013523	
2R A	AGAP001214	SOLUTE CARRIER FAMILY 35 MEMBER G1 (PTHR22911:SF6)
2R A	AGAP001187	HOMEOBOX PROTEIN CUT (PTHR14043:SF2)
2R A	AGAP001205	CHITIN SYNTHASE 2, ISOFORM D (PTHR22914:SF14)
2R A	AGAP013012	
2R A	AGAP001201	DEFECTIVE PROBOSCIS EXTENSION RESPONSE 7, ISOFORM F (PTHR23279:SF6)
2R A	AGAP001211	LD44762P (PTHR47980:SF24)
2R A	AGAP001213	

2R A	AGAP001190	AT20289P-RELATED (PTHR24276:SF78)
2R A	AGAP001229	DYNEIN LIGHT CHAIN TCTEX-TYPE (PTHR21255:SF4)
2R A	AGAP001193	
2R A	AGAP001228	DNA-DIRECTED RNA POLYMERASE II SUBUNIT RPB4 (PTHR21297:SF0)
2R A	AGAP013518	
2R A	AGAP001198	
2R A	AGAP001227	GAMMA-TUBULIN COMPLEX COMPONENT 5 (PTHR19302:SF33)
2R A	AGAP001194	MICROTUBULE-ASSOCIATED PROTEIN FUTSCH (PTHR13843:SF12)
2R A	AGAP013177	
2R A	AGAP001224	CLEAVAGE AND POLYADENYLATION SPECIFICITY FACTOR SUBUNIT 3 (PTHR11203:SF11)
2R A	AGAP001192	
2R A	AGAP001209	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 32 (PTHR21646:SF46)
2R A	AGAP001212	PEPTIDOGLYCAN-RECOGNITION PROTEIN LB-RELATED (PTHR11022:SF67)
2R A	AGAP013048	GEO11133P1 (PTHR48024:SF12)
2R A	AGAP001219	TUBULIN-LIKE PROTEIN ALPHA-4B-RELATED (PTHR11588:SF239)
2R A	AGAP001220	NAD-DEPENDENT PROTEIN DEACETYLASE SIRTUIN-7 (PTHR45853:SF4)
2R A	AGAP001222	METHYLTRANSFERASE-LIKE PROTEIN 23 (PTHR14614:SF2)
2R A	AGAP001203	AGAP001203-PA (PTHR22933:SF19)
2R B	AGAP001694	EXOCYST COMPLEX COMPONENT 4 (PTHR14146:SF0)
2R B	AGAP001680	LD08718P (PTHR12876:SF35)
2R B	AGAP001687	RHO GTPASE-ACTIVATING PROTEIN 19 (PTHR14963:SF7)
2R B	AGAP001695	ZGC:112496 (PTHR21521:SF0)
2R B	AGAP001688	EXOSTOSIN-LIKE 3 (PTHR11062:SF73)
2R B	AGAP001690	RAB3 INTERACTING MOLECULE, ISOFORM F (PTHR12157:SF21)
2R B	AGAP001681	UBIQUITIN CONJUGATION FACTOR E4 A (PTHR13931:SF2)
2R B	AGAP001684	ALKALINE PHOSPHATASE-RELATED (PTHR11596:SF75)
2R B	AGAP001686	U6 SNRNA-ASSOCIATED SM-LIKE PROTEIN LSM1 (PTHR15588:SF8)
2R B	AGAP001677	LD24704P (PTHR10026:SF13)
2R B	AGAP001682	MITOCHONDRIAL INNER MEMBRANE PROTEASE SUBUNIT 1 (PTHR12383:SF16)

2R B	AGAP001683	PERIPHERAL PLASMA MEMBRANE PROTEIN CASK (PTHR23122:SF7)
2R B	AGAP013145	PROTEIN COUCH POTATO (PTHR10501:SF41)
2R B	AGAP001685	NUCLEAR PORE COMPLEX PROTEIN NUP107 (PTHR13003:SF2)
2L	AGAP004812	85/88 KDA CALCIUM-INDEPENDENT PHOSPHOLIPASE A2 (PTHR24139:SF34)
2L	AGAP004725	EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT C-RELATED (PTHR13937:SF0)
2L	AGAP004686	WW DOMAIN-CONTAINING ADAPTER PROTEIN WITH COILED-COIL (PTHR15911:SF6)
2L	AGAP004798	COATOMER SUBUNIT BETA' (PTHR19876:SF2)
2L	AGAP004757	PIRNA BIOGENESIS PROTEIN EXD1 (PTHR46628:SF1)
2L	AGAP004823	PROTEIN SPROUTY (PTHR12365:SF7)
2L	AGAP004767	TRANSCRIPTION FACTOR KEN (PTHR45993:SF7)
2L	AGAP004729	CHORION TRANSCRIPTION FACTOR CF2-RELATED (PTHR24388:SF77)
2L	AGAP004806	GALECTIN (PTHR11346:SF164)
2L	AGAP004745	TRANSCRIPTION ELONGATION REGULATOR HOMOLOG (PTHR15377:SF3)
2L	AGAP004799	
2L	AGAP004771	
2L	AGAP004718	RE21922P (PTHR11003:SF276)
2L	AGAP004723	FI06908P-RELATED (PTHR22812:SF151)
2L	AGAP004827	AGAP004827-PA (PTHR19871:SF28)
2L	AGAP004699	RAF HOMOLOG SERINE/THREONINE-PROTEIN KINASE RAF (PTHR44329:SF193)
2L	AGAP004719	EG:BACR7A4.3 PROTEIN-RELATED (PTHR24260:SF107)
2L	AGAP004763	FI02944P-RELATED (PTHR10174:SF216)
2L	AGAP004687	6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING (PTHR11811:SF25)
2L	AGAP004811	
2L	AGAP004739	H/ACA RIBONUCLEOPROTEIN COMPLEX SUBUNIT DKC1 (PTHR23127:SF0)
2L	AGAP004807	GALECTIN (PTHR11346:SF164)
2L	AGAP004797	AGAP004797-PA (PTHR22904:SF525)
2L	AGAP004781	BETA4GALNACTA (PTHR19300:SF52)
2L	AGAP004810	
2L	AGAP004819	LP05237P-RELATED (PTHR12300:SF22)

2L	AGAP004786	PYRUVATE DEHYDROGENASE E1 COMPONENT SUBUNIT ALPHA (PTHR11516:SF55)
2L	AGAP004747	PRESEQUENCE PROTEASE, MITOCHONDRIAL (PTHR43016:SF13)
2L	AGAP004724	INTRAFLAGELLAR TRANSPORT PROTEIN 74 HOMOLOG (PTHR31432:SF0)
2L	AGAP004780	CYCLIN-DEPENDENT KINASE 12 (PTHR24056:SF233)
2L	AGAP004791,AGAP004789	HIGH MOBILITY GROUP PROTEIN 2 (PTHR46040:SF3)
2L	AGAP004693	NUCLEAR RECEPTOR SUBFAMILY 6 GROUP A MEMBER 1 (PTHR48092:SF18)
2L	AGAP004762	FI02944P-RELATED (PTHR10174:SF216)
2L	AGAP004711	ATP-DEPENDENT RNA HELICASE DDX41-RELATED (PTHR47958:SF103)
2L	AGAP004733	AGAP004733-PA (PTHR46105:SF5)
2L	AGAP004758	PROTEASOMAL UBIQUITIN RECEPTOR ADRM1 (PTHR12225:SF0)
2L	AGAP004769	RE63021P (PTHR24070:SF406)
2L	AGAP004748	
2L	AGAP004717	FI03418P (PTHR11003:SF279)
2L	AGAP004800	
2L	AGAP004696	HOMEOBOX PROTEIN EXTRADENTICLE (PTHR11850:SF108)
2L	AGAP004749	OXYGEN-DEPENDENT COPROPORPHYRINOGEN-III OXIDASE, MITOCHONDRIAL (PTHR10755:SF0)
2L	AGAP004768	PERIODIC TRYPTOPHAN PROTEIN 1 HOMOLOG (PTHR14091:SF0)
2L	AGAP004813	
2L	AGAP004737	RHOMBOID-4 (PTHR45840:SF8)
2L	AGAP004772	SLIT-ROBO GAP HOMOLOG (PTHR14166:SF17)
2L	AGAP004742	PYRUVATE CARBOXYLASE, MITOCHONDRIAL (PTHR43778:SF2)
2L	AGAP004752	TRNA (GUANINE-N(7)-)-METHYLTRANSFERASE (PTHR23417:SF16)
2L	AGAP004744	SUCCINATE--COA LIGASE [ADP-FORMING] SUBUNIT BETA, MITOCHONDRIAL (PTHR11815:SF10)
2L	AGAP004776	PRE-MRNA-SPLICING FACTOR SLU7 (PTHR12942:SF2)
2L	AGAP004770	GH13245P2-RELATED (PTHR24276:SF83)
2L	AGAP004759	OLIGORIBONUCLEASE, MITOCHONDRIAL (PTHR11046:SF0)
2L	AGAP004820	RNA 3'-TERMINAL PHOSPHATE CYCLASE (PTHR11096:SF0)
2L	AGAP004753	
2L	AGAP004741	AT07769P-RELATED (PTHR24260:SF90)

2L	AGAP004814	
2L	AGAP004703	DNA-DIRECTED RNA POLYMERASE III SUBUNIT RPC1 (PTHR19376:SF32)
2L	AGAP004704	CXXC-TYPE ZINC FINGER PROTEIN 1 (PTHR46174:SF1)
2L	AGAP004755	VACUOLAR PROTEIN-SORTING-ASSOCIATED PROTEIN 36 (PTHR13128:SF12)
2L	AGAP004715	FI22513P1 (PTHR13847:SF257)
2L	AGAP004712	
2L	AGAP004691	CHIP, ISOFORM B (PTHR10378:SF19)
2L	AGAP004765	ATP-DEPENDENT RNA HELICASE DDX1 (PTHR24031:SF307)
2L	AGAP004722	IRON-SULFUR CLUSTER CO-CHAPERONE PROTEIN HSCB (PTHR14021:SF15)
2L	AGAP004692	RAS GTPASE-ACTIVATING PROTEIN-RELATED (PTHR10194:SF125)
2L	AGAP004764	SNAIL FAMILY ZINC FINGER 2 TRANSCRIPTION FACTOR HOMOLOG (PTHR24409:SF375)
2L	AGAP004802	4-HYDROXYPHENYLPYRUVATE DIOXYGENASE (PTHR11959:SF1)
2L	AGAP004804	
2L	AGAP004778	BOLA-LIKE PROTEIN 3 (PTHR46188:SF1)
2L	AGAP004790,AGAP004788	ATP SYNTHASE MEMBRANE SUBUNIT DAPIT, MITOCHONDRIAL (PTHR34038:SF1)
2L	AGAP004777	AGAP004777-PA (PTHR21228:SF62)
2L	AGAP004773	PYRUVATE DEHYDROGENASE E1 COMPONENT SUBUNIT ALPHA (PTHR11516:SF55)
2L	AGAP004720	AGAP004720-PA (PTHR11003:SF263)
2L	AGAP004815	
2L	AGAP004794	MIP10846P1-RELATED (PTHR11610:SF161)
2L	AGAP004793	ORNITHINE AMINOTRANSFERASE, MITOCHONDRIAL (PTHR11986:SF18)
2L	AGAP004710	CYTOCHROME B-C1 COMPLEX SUBUNIT 9 (PTHR12980:SF0)
2L	AGAP004808	AMINOPEPTIDASE (PTHR11533:SF283)
2L	AGAP004805	FORMIN-LIKE PROTEIN (PTHR45857:SF6)
2L	AGAP004738	PROTEIN RED (PTHR12765:SF5)
2L	AGAP004761	MULTIPLE WING HAIRS, ISOFORM C (PTHR45857:SF4)
2L	AGAP004824	EUKARYOTIC TRANSLATION INITIATION FACTOR 5B (PTHR43381:SF4)
2L	AGAP004740	AT07769P-RELATED (PTHR24260:SF90)
2L	AGAP004795	RNA POLYMERASE II ELONGATION FACTOR ELL (PTHR23288:SF17)

2L	AGAP004731	RH14732P (PTHR12253:SF34)
2L	AGAP004736	MITOCHONDRIAL RIBOSOME-ASSOCIATED GTPASE 1 (PTHR45782:SF4)
2L	AGAP004766	ENDOPHILIN-A (PTHR14167:SF89)
2L	AGAP004709	39S RIBOSOMAL PROTEIN L18, MITOCHONDRIAL (PTHR12899:SF3)
2L	AGAP004787	FATTY ACYL-COA REDUCTASE (PTHR11011:SF24)
2L	AGAP004784	FATTY ACYL-COA REDUCTASE (PTHR11011:SF24)
2L	AGAP004816	PHOSPHATIDYLINOSITOL-GLYCAN BIOSYNTHESIS CLASS F PROTEIN-RELATED (PTHR43157:SF31)
2L	AGAP004708	ARGININE--TRNA LIGASE, CYTOPLASMIC-RELATED (PTHR11956:SF5)
2L	AGAP004707	SODIUM CHANNEL PROTEIN PARA (PTHR10037:SF288)
2L	AGAP004756	F-BOX/WD REPEAT-CONTAINING PROTEIN 4 (PTHR14381:SF1)
2L	AGAP004706	
2L	AGAP004690	PUPAL CUTICLE PROTEIN C1B-LIKE PROTEIN (PTHR39068:SF5)
2L	AGAP004785	GH10083P-RELATED (PTHR10174:SF222)
2L	AGAP004801	HUNTINGTIN INTERACTING PROTEIN 1 (PTHR10407:SF15)
2L	AGAP004694	RE24423P (PTHR11161:SF22)
2L	AGAP004818	28S RIBOSOMAL PROTEIN S16, MITOCHONDRIAL (PTHR12919:SF27)
2L	AGAP004689	LD40707P (PTHR13055:SF12)
2L	AGAP004817	PROTEIN LINGERER (PTHR16308:SF13)
2L	AGAP004700	FI18411P1-RELATED (PTHR24260:SF91)
2L	AGAP004726	AGAP004726-PA (PTHR10559:SF18)
2L	AGAP004716	
2L	AGAP004774	HOST CELL FACTOR (PTHR46003:SF1)
2L	AGAP004713	
2L	AGAP004825	BCDNA.LD24702 (PTHR13030:SF14)
2L	AGAP004792	STAR-RELATED LIPID TRANSFER PROTEIN 7, MITOCHONDRIAL (PTHR19308:SF8)
2L	AGAP004775	XAA-PRO DIPEPTIDASE (PTHR43226:SF1)
2L	AGAP004754	CASPASE DRONC (PTHR10454:SF196)
2L	AGAP004714	CROL ALPHA (PTHR23226:SF176)
2L	AGAP013546	VACUOLAR ATPASE ASSEMBLY INTEGRAL MEMBRANE PROTEIN VMA21 HOMOLOG (PTHR31792:SF6)

2L	AGAP004782	TUBULIN-SPECIFIC CHAPERONE D (PTHR12658:SF0)
2L	AGAP004743	TRANSMEMBRANE EMP24 DOMAIN-CONTAINING PROTEIN 2 (PTHR22811:SF177)
2L	AGAP004728	ARMADILLO REPEAT-CONTAINING PROTEIN 2 (PTHR21356:SF1)
2L	AGAP004809	AMINOPEPTIDASE (PTHR11533:SF283)
2L	AGAP004695	MULTIVESICULAR BODY SUBUNIT 12A (PTHR31612:SF2)
3L A	AGAP010820	
3L A	AGAP010803	
3L A	AGAP010787	PROTEIN TIMELESS HOMOLOG (PTHR22940:SF4)
3L A	AGAP010790	CYTOPLASMIC DYNEIN 2 HEAVY CHAIN 1 (PTHR10676:SF352)
3L A	AGAP010793	SYNEMBRYN (PTHR12425:SF5)
3L A	AGAP010809	CHAOPTIN-LIKE PROTEIN (PTHR45617:SF82)
3L A	AGAP010814	GH01829P-RELATED (PTHR11412:SF136)
3L A	AGAP010789	LD24340P (PTHR39069:SF4)
3L A	AGAP010784	PROTEIN O-MANNOSYL-TRANSFERASE 1-RELATED (PTHR10050:SF51)
3L A	AGAP010812	GH01829P-RELATED (PTHR11412:SF136)
3L A	AGAP010798	GH13245P2-RELATED (PTHR24276:SF83)
3L A	AGAP010792	NADH DEHYDROGENASE [UBIQUINONE] 1 ALPHA SUBCOMPLEX SUBUNIT 9, MITOCHONDRIAL (PTHR12126:SF11)
3L A	AGAP010777	HUNCHBACK-LIKE (PTHR24392:SF33)
3L A	AGAP010811	GH05177P-RELATED (PTHR19143:SF383)
3L A	AGAP013327	PEROXIDASE (PTHR11475:SF86)
3L A	AGAP010817	AGAP006059-PA (PTHR47537:SF3)
3L A	AGAP010818	GH01829P-RELATED (PTHR11412:SF136)
3L A	AGAP010783	GEO02292P1-RELATED (PTHR23301:SF79)
3L A	AGAP010799	
3L A	AGAP010813	
3L A	AGAP010786	RNA (RNA) POLYMERASE II ASSOCIATED PROTEIN HOMOLOG (PTHR45984:SF1)
3L A	AGAP010816	GH01829P-RELATED (PTHR11412:SF136)
3L A	AGAP010802	COREPRESSOR OF PANGOLIN, ISOFORM A-RELATED (PTHR12243:SF47)
3L A	AGAP010780	

3L A	AGAP010782	RAL GUANINE NUCLEOTIDE DISSOCIATION STIMULATOR-LIKE, ISOFORM E (PTHR23113:SF312)
3L A	AGAP010800	ZINC FINGER CCCH DOMAIN-CONTAINING PROTEIN 18 (PTHR46582:SF1)
3L A	AGAP010807	TRNA PSEUDOURIDINE(38/39) SYNTHASE (PTHR11142:SF5)
3L A	AGAP010819	GH01829P-RELATED (PTHR11412:SF136)
3L A	AGAP010785	FATTY ACYL-COA REDUCTASE (PTHR11011:SF81)
3L A	AGAP010805	
3L A	AGAP010806	BCDNA.GH04802 (PTHR11003:SF158)
3L A	AGAP010808	ARC, ISOFORM A (PTHR19964:SF69)
3L A	AGAP010776	WHIRLIN (PTHR23116:SF37)
3L A	AGAP010791	POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE 1, MITOCHONDRIAL (PTHR11252:SF0)
3L A	AGAP010794	FI23527P1 (PTHR45820:SF4)
3L A	AGAP010801	PHOSPHATIDYLINOSITOL 4-KINASE ALPHA-RELATED (PTHR10048:SF15)
3L A	AGAP010804	SI:CH73-302O18.2 (PTHR45752:SF27)
3L B	AGAP011332	FINGER PUTATIVE TRANSCRIPTION FACTOR FAMILY-RELATED (PTHR24399:SF23)
3L B	AGAP011316	SENTRIN-SPECIFIC PROTEASE 8 (PTHR46468:SF1)
3L B	AGAP011345	GROWTH ARREST AND DNA DAMAGE-INDUCIBLE PROTEINS-INTERACTING PROTEIN 1 (PTHR31761:SF1)
3L B	AGAP011320	SUBFAMILY NOT NAMED (PTHR24230:SF141)
3L B	AGAP011387	MITOCHONDRIAL DIVISION PROTEIN 1-RELATED (PTHR19857:SF8)
3L B	AGAP011323	TCF3 FUSION PARTNER (PTHR35084:SF1)
3L B	AGAP011361	CHECKPOINT PROTEIN (PTHR12900:SF0)
3L B	AGAP011330	GLUTAREDOXIN-LIKE PROTEIN C5ORF63 HOMOLOG (PTHR33558:SF1)
3L B	AGAP011317	
3L B	AGAP011374	ADAPTOR PROTEIN COMPLEX 1, MU SUBUNIT (PTHR10529:SF262)
3L B	AGAP011365	CARBOXYLIC ESTER HYDROLASE (PTHR43142:SF1)
3L B	AGAP011377	
3L B	AGAP011389	PROTEIN CBG05349 (PTHR15192:SF8)
3L B	AGAP011376	
3L B	AGAP011362	GATA ZINC FINGER DOMAIN-CONTAINING PROTEIN 1 (PTHR13340:SF2)

3L B	AGAP011329	ACETYL-COA ACETYLTRANSFERASE, MITOCHONDRIAL (PTHR18919:SF156)
3L B	AGAP011356	AMYLOID PROTEIN-BINDING PROTEIN 2 (PTHR46575:SF1)
3L B	AGAP011368	AGAP011368-PA (PTHR21066:SF17)
3L B	AGAP011352	GH12731P-RELATED (PTHR10937:SF0)
3L B	AGAP011338	GDP-FUCOSE PROTEIN O-FUCOSYLTRANSFERASE 1 (PTHR21420:SF3)
3L B	AGAP011349	GAMMA-AMINOBUTYRIC ACID RECEPTOR ALPHA-LIKE (PTHR18945:SF835)
3L B	AGAP011363	RAS-RELATED PROTEIN RAB6 (PTHR24073:SF352)
3L B	AGAP011326	DEATH-ASSOCIATED INHIBITOR OF APOPTOSIS 2 (PTHR10044:SF139)
3L B	AGAP011372	
3L B	AGAP011367	
3L B	AGAP011321	PROTOPORPHYRINOGEN OXIDASE (PTHR42923:SF3)

757