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 sterility and inviability between *An. coluzzii* and *An. quadriannulatus* by performing a QTL analysis of sterility in a backcross population. The *An. coluzzii* X chromosome has a disproportionately large effect on hybrid sterility when introgressed into an *An. quadriannulatus* genetic background, and interacts epistatically with a 1.12 Mb, pericentric region of the 3L chromosome arm to cause complete inviability. 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,

26 spermatogenesis, and microtubule morphogenesis, whose divergence may contribute to post-zygotic
27 reproductive isolation between *An. coluzzii* and *An. quadriannulatus.*
28
29 **Introduction**
30 The study of speciation genetics and genomics has largely focused on the evolution hybrid
31 dysfunction (Coyne and Orr, 2004; Dobzhansky, 1937) because it represents the first step in the
32 formation of post-zygotic reproductive isolation between diverging populations. The evolution of
33 post-zygotic isolation mechanisms is positively correlated with the time of divergence between two
34 putative species, and is impacted by the number of mutations separating two allopatric lineages at
35 any given time, the number of incompatibilities per mutation, and the fitness of each
36 incompatibility (Orr and Turelli, 2001).
37 Post-zygotic isolation occurs as Dobzhansky-Müller incompatibilities (DMI) arise between
38 populations evolving in allopatry (Dobzhansky, 1936; Müller, 1940). Upon secondary contact of
39 genetically divergent populations, epistatic loci on parental chromosomes are unable to function
40 effectively in hybrids. The resulting gene products in hybrids can have unpredictable, deleterious
41 epistatic interactions (Maheshwari and Barbash, 2011). Orr (1995) and Orr and Turelli (2001)
42 proposed the "snowball effect" of DMI evolution, which predicts that the number of DMIs
43 accumulating between lineages "snowballs", or increases faster than linearly (but see Kalirad and
44 Azevedo, 2017 for explicit testing of the underlying assumptions of this model). The snowball
45 effect has been supported by empirical evidence of the rate of DMI evolution between *Drosophila*
46 (Matute et al., 2010; Matute and Gavin-Smyth, 2014) and *Solanum* sibling species pairs (Moyle
47 and Nakazato, 2010).
48 DMI can evolve as a result of various mechanisms in either an ecological speciation
49 (divergent evolution due to selection pressures imposed by novel environments) or mutation-order
50 speciation scenarios (fixation of different, incompatible mutations in allopatric populations

51 experiencing the same environmental conditions) (Mani and Clarke, 1990; Nosil and Flaxman,
52 2011). These mechanisms include mutation-driven co-evolution (e.g. copper tolerance in *Mimulus*
53 *guttatus*, MacNair, 1983; MacNair and Christie, 1983), gene duplication (e.g. the *OdsH* gene in
54 *Drosophila simulans* x *D. mauritiania* backcross males, Ting et al., 2004; Sun et al., 2004), gene
55 transposition (e.g. the *JYAlpha* gene in D. *melanogaster* x *D. simulans* backcross males, Masley et
56 al., 2006), or a molecular arms race (e.g. divergent NB-LRR alleles in *Arabidopsis thaliana* strains,
57 Bomblies et al., 2007; Bomblies and Weigal, 2007; Bakker et al., 2006). Furthermore, divergence
58 in gene regulatory regions between parental species can result in in allelic imbalance in hybrids
59 due to incompatible *cis*- and/or *trans*-acting factors (Graze et al., 2012), and segregation distortion
60 of parental loci can skew sex ratios of hybrid progeny (Orr and Irving, 2005; Phadnis and Orr,
61 2009).

62 The *Anopheles* (*An.*) *gambiae* complex is comprised of eight morphologically
63 indistinguishable species and has emerged as a model system for the study of speciation genetics
64 due to the rapid radiation of its member species over the past two million years (Fontaine et al.,
65 2015). All members of the *An. gambiae* complex are competent vectors of human malaria parasites,
66 though they differ in their host specificity, ranging from highly anthropophilic (*An. gambiae s.s.*
67 and *An. coluzzii*) (Takken and Knols, 1990; Dekker and Takken, 1998; Pates et al., 2001a) to almost
68 entirely zoophilic (*An. quadriannulatus* A) (White, 1974; Gibson, 1996; Dekker and Takken, 1998;
69 but see Pates et al., 2001b). From a vector control stand point, it is important to know which
70 genomic regions introgress between these species because of the potential for insecticide resistance
71 genes to pass between species (e.g. the movement of *knock down resistance* (*kdr*) between
72 chromosomal forms of *Anopheles gambiae*) (Weill et al., 2000). Indoor residual spraying (IRS) of
73 insecticides and the distribution of long-lasting insecticide-treated bed nets (LLINs) remain the
74 primary methods used to limit contact between humans and malaria vectors, thus the spread of
75 insecticide resistant alleles between species can severely undermine malaria control efforts.

76 Furthermore, recent advances in the development of transgenic mosquito control methods have
77 made our understanding the permeability of species boundaries even more urgent (e.g. Kyrou et
78 al., 2018).

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

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

99 Female hybrid sterility QTL are present on the third chromosomes of both *An. coluzzii* and
100 *An. arabiensis* (Slotman et al., 2005). A QTL on the second chromosome *An. coluzzii* causes female

101 hybrid sterility when it is introgressed into an *An. arabiensis* genomic background. However, the
102 female sterility phenotype is affected primarily by the interaction of a foreign X chromosome with
103 the two autosomal QTL. Approximately 75% of observed phenotypic variance is due to the X
104 chromosome, which only represents 8.8% of the genome (Holt et al., 2002). Due to the lack of
105 recombination between X chromosomes in the cross, the number of loci on the X chromosome that
106 effect the hybrid sterility phenotype is unknown. A comparison of the data between the male
107 (Slotman et al., 2004) and female (Slotman et al., 2005) *An. coluzzii* x *An. arabiensis* hybrid sterility
108 data shows that there are a greater number of male than female hybrid sterility QTL, supporting
109 previous studies that found male sterility factors to evolve faster than female sterility factors
110 (Charlesworth et al., 1987; Kirkpatrick and Hall, 2004). This is consistent with faster male
111 evolution observed in *Aedes* mosquitoes (Presgraves and Orr, 1998), and the higher numbers of
112 male versus female sterility factors in *Drosophila* (e.g. Hollocher et al., 1996; True et al., 1996;
113 Tao and Hartl, 2003).

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

123

124 **Methods**

125 *Backcross and Phenotype Scoring*

126 We performed a backcross between the SUA2La strain of *An. coluzzii* and the SANGUA
127 strain of *An. quadriannulatus* (*An. quadriannulatus*) to identify regions of the *An. coluzzii* genome
128 that exhibit DMI with the *An. quadriannulatus* genome, resulting in F1 male sterility. The *An.*
129 *coluzzii* strain was established from collections in Suakoko, Libera and the *An. quadriannulatus*
130 strain was established in from populations in Sangwe, Zimbabwe. Both strains have been lab reared
131 for hundreds of generations. We used a standard backcross scheme that has been used in studies to
132 investigate the genetic basis of hybrid sterility and inviability going back to Dobzhansky (1936).
133 F1 *An. coluzzii* x *An. quadriannulatus* (females x males, respectively) hybrids were reared and
134 backcrossed to *An. quadriannulatus* males *en masse* as the goal of the study was to examine the
135 effect of *An. coluzzii* sterility factors when introgressed into an *An. quadriannulatus* background.
136 Consistent with crosses between *An. coluzzii* and *An. arabienesis*, the Xag inversion was expected
137 to suppress recombination between the X chromosomes of *An. coluzzii* and *An. quadriannulatus*.
138 The testes of male hybrids were dissected and squashed between a microscope slide and a
139 cover slip. Each mosquito was assigned a sterility phenotype based upon sperm abundance and
140 morphology and testes morphology modified from Slotman et al. (2004). Phenotype scores were
141 assigned based on a one to seven scale: (1) normal testes and sperm, (2) slightly abnormal sperm,
142 (3) 50/50 normal and abnormal sperm, (4) mostly abnormal sperm, (5) entirely abnormal sperm,
143 (6) no sperm present and/or severely atrophied testes, (7) no testes present. Mosquitoes were
144 preserved in 100% ethanol after phenotyping for DNA extraction.
145 *DNA sequencing*
146 DNA extractions were performed on a Qiagen Biosprint DNA extraction machine (Qiagen
147 Inc, Valencia, CA). Extracted DNA was suspended in 200µl elution buffer and stored at -20ºC.
148 Each mosquito was genotyped using the restriction enzyme-based, genotype-by-sequencing
149 approach according to the multiplexed shotgun genotyping (MSG) approach published by
150 Andolfatto et al. (2011).

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

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

176 *quadriannulatus* females to further validate ancestry informative markers and tune the MSG
177 software error model.

178 *Parental pseudo-reference Construction*

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

188 *Hybrid Genotyping*

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

193 We used the MSG software package (Andolfatto et al., 2011) to align hybrid MSG data to
194 parental pseudo-references and estimate the posterior probabilities of ancestry at genome-wide
195 SNPs that were informative of parental ancestry. We removed 33 hybrids from this analysis due to
196 low sequencing coverage. Posterior probabilities of ancestry were thinned to informative markers
197 using pull_thin_tsv.py (https://github.com/dstern/pull_thin). We considered a marker to be
198 informative if its conditional probability of being homozygous differed by more than 0.10 from its
199 neighboring marker.

200 *QTL Mapping*

201 We used R/qtl (Broman et al., 2003) to estimate the genetic map of 808 hybrid males using
202 the Kosambi map function (Kosambi, 1943). We used this package to identify adjacent redundant,
203 adjacent markers and further thinned our marker set to include only informative markers. We
204 performed quantitative trait locus mapping with the composite interval mapping (CIM) method
205 using 15 markers as co-factors. We assessed the significance of QTL models using multiple interval
206 mapping, which drops one QTL at a time to calculate the probability of QTL locations and their
207 interactions using an ANOVA (Sen and Churchill, 2001). In all analyses, sterility was treated as a
208 normal trait, and we used 1000 permutation replicates to assess the significance of log of odds
209 (LOD) scores.
210
211 **Results**
212 *Mapping and Genetic Map*
213 We calculated genotype probabilities at 239,420 ancestry informative markers for 808 *An.*
214 *coluzzii* x *An. quadriannulatus* backcross hybrid males. Ultimately, we trimmed the dataset to 860 non-
215 redundant markers, which had 3.1% missing genotypes (Table 1). These markers were used to calculate
216 a genetic map (Figure 1) with Kosambi distances of 206.1 cM for the second chromosome (110.9 Mb),
217 100.7 cM for the third (95.2 Mb), and 3.7 cM for the X (24.4 Mb) (Table 1, Figure 1). The larger
218 relative genetic map of the second chromosome is likely due to the presence of the *2La* inversion,
219 which is fixed between these species and can significantly increase recombination frequency on the 2R
220 (i.e. Schultz-Redfield Effect, Schultz and Redfield, 1951; Lucchesi and Suzuki, 1968; Stevison et al.,
221 2011), and segregation distortion resulting from male inviability for some genotype combinations (Zuo
222 et al., 2019; Hackett and Broadfoot, 2003).
223 Segregation distortion has been previously documented in male and female hybrids between
224 member species of this complex (Slotman et al., 2004; Slotman et al., 2005; Smith et al., 2015). The
225 genetic length of the X chromosome was expected to be low due to the presence of the *Xag* inversion,

226 which comprises 80% of the X and is fixed for alternate arrangements between these species. The
227 majority of hybrids (70.3%) inherited a non-recombinant *An. quadriannulatus* X chromosome, 26.1%
228 inherited a non-recombinant *An. coluzzii* X, and 3.5% a recombinant X. The mean sterility score
229 (ranging from 1-7, normal to completely sterile) was 2.62 (sd=2.34). The majority of backcross
230 mosquitoes (64.23%) were fully fertile (sterility score = 1).

231 *Composite Interval Mapping*

232 We took three approaches to identify sterility QTL in the dataset. The first was to scan for QTL
233 using the composite interval mapping and calculating the logarithm of these odds (LOD) score every
234 1 cM. When calculating LOD scores at each map location we controlled for the genetic background of
235 adjacent markers by using 15 marker covariates across the genome. Due to the skew of the phenotype
236 scores toward "normal", and severe segregation distortion of the X chromosome, we hypothesized that
237 the QTL identified through CIM of the full dataset were in fact comprised of two types: (1) autosomal
238 *An. coluzzii* loci that are incompatible with the *An. quadriannulatus* X, and (2) homozygous *An.*
239 *quadriannulatus* autosomal loci that interact epistatically with the *An. coluzzii* X in individuals that are
240 otherwise primarily heterozygous at autosomal loci (recovering viability despite the inheritance of a
241 non-recombinant *An. coluzzii* X chromosome). Thus, we also split the dataset into individuals that had
242 inherited the *An. quadriannulatus* and *An. coluzzii* X, respectively, and performed CIM (as above) on
243 these datasets separately.

244 Due to minimal recombination events, the X chromosome is effectively one large QTL with a
245 LOD score of 203.4 (Table 2). By comparison, none of the autosomal QTL reach a LOD score of 4.0.
246 No autosomal QTL fall above the CIM 5% significance threshold (LOD=4.31, 1,000 permutations).
247 Despite this, we annotated three QTL peaks on the second chromosome (1.63 cM, 20.87 cM, and
248 134.26 cM) (Figure 2.A, Table 2) that contributed significantly to the sterility phenotype when included
249 in the multiple imputation QTL model (see *Multiple Imputation*, below, and Table 4). The second
250 chromosome peak at 1.63 cM (QTL "2R-A") was present in the full and *An. quadriannulatus* X

251 datasets, but not the *An. coluzzii* X dataset (Figure 2.A). QTL "2L" at 134.26 cM was present in the
252 full and *An. coluzzii* X datasets, but not the *An. quadriannulatus* X dataset. Finally, QTL "2R-B" (20.87
253 cM) was found in the full and both X chromosome split datasets, suggesting that it contains loci
254 involved in autosomal-autosomal incompatibilities (Figure 2.A). While additional LOD peaks are
255 present in the *An. coluzzii* X dataset, these were non-significant when included in the multiple
256 imputation QTL model (see *Multiple Imputation*, below) (Figure 2.A). None of the second
257 chromosome QTL overlap with those found in the *An. coluzzii* x *An. arabiensis* cross (Slotman et al.,
258 2004) (Figure 2.B, Table 2).

259 Two QTL were identified on the third chromosome at 53.95 (QTL "3L-A") and 67.46 cM (QTL
260 "3L-B") (Figure 2.C, Table 2). These are both present in the full and *An. quadriannulatus* X dataset,
261 but not the *An. coluzzii* X dataset, indicating that *An. coluzzii* introgressions into a predominantly *An.*
262 *quadriannulatus* genetic background are driving this signal. QTL "3L-A" in the *An. coluzzii* x *An.*
263 *quadriannulatus* dataset overlaps with a QTL in the *An. coluzzii* x *An. arabiensis* dataset (Slotman *et*
264 *al*, 2004) (Figure 2.D, Table 2). The 3L-A and 3L-B QTL have seemingly additive-only effects on
265 sterility (see *Multiple Imputation*, below); these QTL were not identified by CIM in the *An. coluzzii* X
266 dataset, suggesting that they result only from *An. coluzzii* introgressions into predominantly *An.*
267 *quadriannulatus* genetic backgrounds. However, only three hybrid males were sampled that are *An.*
268 *quadriannulatus* 3L-B homozygous and inherited the *An. coluzzii* X (Figure 3), and each had a sterility
269 score of five or six. No hybrid males were sampled that are *An. quadriannulatus* 3L-A homozygous
270 and inherited the *An. coluzzii* X. The reduction (or absence) of homozygous genotypes at the 3L-A and
271 3L-B QTL differs significantly from expectation (the percentage of homozygous genotypes across all
272 autosomal loci, assuming HWE) (Table 3). Thus, these regions of the *An. quadriannulatus* genome
273 interact epistatically with the X and result in almost complete male inviability when introgresesed into
274 an *An. coluzzii* genetic background.

275 *Multiple Imputation*

276 CIM detects QTL by calculating LOD scores at each marker or map location along each
277 chromosome and does this agnostic of other QTL contributions to the phenotype and potential epistatic
278 interactions between QTL. Due to this limitation, we built multiple imputation QTL models to detect
279 the significance of all QTL LOD peaks found during CIM of the full (i.e., not split by X chromosome)
280 dataset. The QTL discussed above, and those reported in Table 2, are only those that contributed
281 significantly to the variance in sterility in the multiple imputation models described herein. We report
282 two QTL models, one that considers only additive effects of QTL on sterility, and one which includes
283 significant epistatic effects between QTL. We took an iterative process during testing QTL models and
284 added QTL locations identified through CIM in a stepwise fashion. After we had identified those that
285 had significant additive effects, we tested all pairwise epistatic effects between these loci. The final
286 "additive only" and "epistasis allowed" models explain 71.66% and 72.29% of the variance in sterility,
287 respectively (Table 4). The vast majority of this variance is explained by the inheritance of the X
288 chromosome. As expected, individuals harboring the *An. coluzzii* X chromosome have on average
289 higher sterility scores because in these backcrosses there are more opportunities for *An.*
290 *quadriannulatus* autosomal – *An. coluzzii* X incompatibilities than vice versa (Figure 3).

291 Each of the QTL described had a significant additive effect in both the "additive only" and
292 "epistasis allowed" models (F-test p-value < 0.01) (Table 4). After testing all possible pair-wise
293 interactions between these loci, we found that only the X and 2L (134.26 cM) QTL pair had a
294 significant interaction (F-test p-value = 5.60E-06). In both models, additive effects of autosomal loci,
295 and the epistatic interaction between the X and 2L QTL, each explained less than 1% of the variance
296 in hybrid sterility. The "epistasis allowed" model explained a higher proportion of the variance and
297 had a higher model LOD (225.78) when compared to the "additive only" model (221.25) (Table 4).

298 To better understand the direction of the effect of each QTL in the full dataset, we compared
299 the phenotypes of each individual given their genotype (*An. quadriannulatus* homozygous (QQ) vs.
300 *An. quadriannulatus/An. coluzzii* heterozygous (QC)) at autosomal QTL and the X chromosome

301 (Figure 4). Each autosomal locus increases sterility in the heterozygous state, and individuals with the
302 *An. coluzzii* X are significantly more sterile. The interaction plot (Figure 4, "X vs 2L", right panel)
303 between the X and 2L QTL demonstrates that heterozygosity at the 2L QTL alone does not impact
304 sterility when an individual has the *An. quadriannulatus* X, but reduces sterility when inherited with
305 the *An. coluzzii* X. The evidence suggests that each autosomal QTL indicate the location of a (partially)
306 dominant *An. coluzzii* sterility gene that causes incompatibilities with the *An. quadriannulatus*
307 background.

308 We determined the 1.5 LOD interval of each QTL and, using the base pair positions of our SNP
309 markers, determined the physical map span of each (Table 2). Next, we compared QTL locations to
310 male sterility QTL found in a *An. coluzzii* x *An. arabiensis* cross (Figure 2 B, D) (Slotman et al., 2004).
311 We re-analyzed this data using CIM with three marker covariates. No significant QTL on the second
312 chromosome overlap between these crosses. However, the 3L-A QTL in our *An. coluzzii* x *An.*
313 *quadriannulatus* cross falls within the 3L QTL from the *An. coluzzii* x *An. arabiensis* cross, and the
314 3L-B QTL overlaps it. In the Slotman et al. (2004) data the third chromosome QTLs are also only
315 present when the hybrid has the backcross (*An. arabiensis* in this case) parent X chromosome (Figure
316 2.D). Thus, interestingly this represents a region of the *An. coluzzii* that contributes to sterility when
317 introgressed into both *An. quadriannulatus* and *An. arabiensis* genetic backgrounds.

318 *Candidate Speciation Genes*

319 We identified all annotated genes in the AgamP4 genome within the 1.5 LOD intervals of each
320 QTL using VectorBase (www.vectorbase.org) and found the protein family (if available) of each gene
321 using PANTHER. Due to the lack of resolution on the X chromosome, we excluded this chromosome
322 from this analysis. We identified 349 genes within all autosomal QTL (2.47 Mb total, Table 2). No
323 gene ontology terms were over-represented in the dataset. However, we identified ten genes with
324 protein functions related to mating, reproduction, testes, and microtubule morphogenesis and function
325 (Table 5) in the full gene set (Table S1).

326 We performed a protein network analysis on the entire gene list using STRING (Szklarcyk et
327 al. 2019). The full STRING protein interaction results can found in Supplementary File 1. Many protein
328 products of genes in autosomal QTL regions have direct interactions, and four proteins with functions
329 related to reproduction (bold in Table 5) interact directly in an insular protein network: dynein light
330 chain Tctex-type, intra-flagellar transport protein 74 homolog, tubulin-specific chaperone D, and
331 cytoplasmic dynein 2 heavy chain 1.
332
333 **Discussion**
334 Similar to the *An. coluzzii* x *An. arabiensis* cross (Slotman et al., 2004), the X chromosome in
335 the *An. coluzzii* x *An. quadriannulatus* cross plays a major role in the hybrid male sterility phenotype.
336 In all multiple imputation QTL models tested, the X chromosome accounted for >60 % of the variation
337 in the sterility phenotype. Unfortunately, due to the *Xag* inversion, the entire X chromosome is
338 effectively one QTL, and we have limited resolution to identify candidate genes. While the 1.5 LOD
339 interval of the X chromosome QTL is 8.53 Mb, the entire span of the QTL is 14.2 Mb, or 58% of the
340 X chromosome. We were able to determine which autosomal *An. coluzzii* loci contributed to sterility
341 when introgressed into a predominantly *An. quadriannulatus* genomic background, and vice versa, by
342 dividing the dataset by the parental origin of the X chromosome.
343 This analysis also revealed that the epistatic interaction between the 2L and the X QTL is driven
344 by the interaction between *An. quadriannulatus* QTL 2L homozygotes and the hemizygous *An. coluzzii*
345 X chromosome (Figures 2, 3). Individuals with this genotype combination are among the most sterile
346 in the cross (Figure 3), and the 2L QTL is absent in the *An. quadriannulatus* X chromosome CIM
347 results (Figure 2.A).
348 The 3L-A QTL sterility/inviability QTL in our *An. coluzii* x *An. quadriannulatus* cross overlaps
349 entirely with the 3L QTL identified in the *An. coluzzii* x *An. arabiensis* cross (Slotman et al., 2004).
350 No *An. arabiensis* homozygotes were sampled in this region when the hybrid inherited the *An. coluzzii*

351 X chromosome. Thus, the *An. coluzzii* X chromosome interacts epistatically with this same region of
352 3L in both *An.* quadriannulatus and *An. arabiensis* to contribute to to hybrid male sterility and causes
353 almost complete hybrid male inviability. We hypothesize that the 3L-A QTL harbors an ancestral allele
354 that is shared between *An.* quadriannulatus and *An. arabiensis* that has not been subjected to
355 introgression between *An. coluzzii* and *An. arabiensis*.

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

364 While gene tree phylogenies vary across the genome due to the reasons outlines above, on the
365 3L, where we find overlap of sterility QTL between *An. quadriannulatus* and *An. arabiensis* crosses,
366 the most commonly found gene phylogenies are those that group *An. arabiensis* closer to *An. coluzzii*
367 than *An. quadriannulatus* (Figure 2 of Fontaine et al., 2015). If we assume that genes involved in post-
368 zygotic reproductive isolation between this species trio evolved first on the X chromosome between
369 the ancestor of *An. gambiae s.s.* + *An. coluzzii* and the ancestor of *An. arabiensis* + *An.*
370 *quadriannulatus*, we can hypothesize that the 3L-A sterility/inviability QTL shared between these
371 crosses may have an X chromosome-like topology despite the homogenizing effect of introgression
372 between the *An. coluzzii* and *An. arabiensis* autosomes. This topology is found in ~12.6% of 50-kb
373 genomics windows on the 3L (Figure 2 of Fontaine et al., 2015; see "other, A+Q trees").

374 Our hypothesis that the 3L-A by X chromosome DMI may have been the first to arise between
375 the ancestor of *An. gambiae s.s.* + *An. coluzzii* and the ancestor of *An. arabiensis* + *An. quadriannulatus*

376 can be tested in the following ways. First, this sterility/inviability QTL should also exist in a cross
377 between *An. gambiae s.s.* and *An. arabiensis* or *An. quadriannulatus,* and second, it should be absent
378 in a cross between *An. arabiensis* and *An. quadriannulatus*. No hybrid sterility/inviability exists
379 between *An. coluzzii* and its sister species *An. gambiae s.s.*, so we expect that the same loci will be
380 responsible for reproductive isolation between both of these species and either *An. arabiensis* or *An.*
381 *quadriannulatus*. A cross between *An. arabiensis* and *An. quadriannulatus* may yield more interesting
382 results; it would help us to understand the order by which individual DMI arose between these species
383 pairs, and how this effects our understanding of the "snowball effect" of DMI evolution in the context
384 of rapidly radiating species. For example, while the 3L-A by X chromosome DMI has been maintained
385 between *An. coluzzii* and both *An. arabiensis* and *An. quadriannulatus*, additional DMI may have
386 evolved in a "snowball effect" type scenario between *An. coluzzii* and *An. quadriannulatus,* where
387 there is no evidence historical autosomal introgression (Fontaine et al., 2015), whereas DMI may have
388 been lost as a result of autosomal introgression between *An. coluzzii* and *An. arabiensis* (or their relative
389 rate of evolution/accumulation may have been lower).

390 Our candidate sterility/inviability gene analysis identified seven genes (out of 349) within
391 autosomal *An. coluzzii* x *An. quadriannulatus* hybrid male sterility QTL that had protein functions
392 related to tubulin or microtubule biogenesis, structure, or function, four of which interact directly with
393 each other (Table 5). Thus, we identified multiple genes whose interactions could be disrupted during
394 this critical stage of spermatogenesis in hybrids as a result of their divergence between the parental
395 species. In *Drosophila,* spermatogenesis progresses through mitosis of germ cells, meiosis, and finally
396 the differentiation of spermatids into mature sperm (White-Cooper et al., 2009). Infertility in
397 *Drosophila* hybrid males is associated with spermiogenic (post- rather than pre-meiotic) failure
398 associated with sperm individualization and maturation (Wu et al., 1992).

399 Sperm flagellum elongate from the spermatid head after meiosis. The axoneme is the central
400 component of the sperm flagellum and has a 9+2 structure of a central pair of microtubules surrounded

401 by nine doublet microtubules. Doublet microtubules have an inner and outer dynein arm that serve to
402 attach and detach to neighboring doublet microtubules, and function in the motor activity of the
403 flagellum. The axoneme is surrounded by a mitochondrial sheath that serves to provide ATP and power
404 motor activity. Tektins are located near the junction points of dynein arms and neighboring
405 microtubules (Gagon, 1995; Linck et al., 2016). Seven of the ten reproduction genes have PANTHER
406 protein functions related to tubulin or microtubule biogenesis, structure, or function: gamma-tubulin
407 complex component 5, microtubule-associated protein futsch, tubulin-like protein alpha-4B-related,
408 dynein light chain Tctex-type, intra-flagellar transport protein 74 homolog, tubulin-specific chaperone
409 D, and cytoplasmic dynein 2 heavy chain 1. The latter four genes (bold, Table 5) in this list have direct
410 interactions in the STRING protein interaction network, and may represent a candidate genetic pathway
411 whose disruption results in hybrid male sterility in the *An. coluzzii* x *An. quadriannulatus* cross.
412 Interestingly, another tubulin-specific chaperone, E-like, is required for sperm individualization and
413 male fertility in *Drosophila* (Nuwal et al., 2012).

414 The three genes without tubulin/microtubule functions are: transcription factor ken, protein
415 lingerer, and a checkpoint protein Hus1-like. Mutations in the ken and barbie transcription factor result
416 in malformed genitalia in *Drosophila* (Lukacsovich et al., 2003). lingerer encodes a putative RNA
417 binding protein and *Drosophila* lingerer mutant males fail to withdraw their genitalia after copulation
418 (Kuniyoshi et al., 2002). Finally, Hus1-like is a component of the 9-1-1 checkpoint protein complex,
419 which coordinates DNA damage sensing, cell cycle progression, and DNA repair (Cotta-Ramusino et
420 al., 2010). Future analyses of the molecular evolution of these genes and their regulatory regions
421 between member species of the *An. gambiae* complex may yield insight into the specific mutations
422 responsible for hybrid male sterility between *An. coluzzii*, *An. quadriannulatus*, and *An. arabiensis.*

423 In this study we have narrowly defined autosomal QTL that, in addition to the X chromosome,
424 contribute to a large portion of the hybrid sterility phenotype in a cross between *An. coluzzii* and *An.*
425 *quadriannulatus*. The shared sterility/inviability QTL regions between the *An. coluzzii* x *An.*

426 *quadriannulatus* and *An. coluzzii* x *An. arabiensis* crosses suggest that these are important autosomal
427 regions involved in the post-zygotic reproductive isolation of *An. coluzzii* from its sister species that
428 diverged early in the radiation of the *An. gambiae* species complex. Future analyses into factors
429 contributing to female hybrid sterility in the *An. coluzzii* x *An. quadriannulatus* cross will add to our
430 knowledge of the genetics of speciation in this evolutionarily and medically important group of human
431 malaria mosquitoes.
432

433 **References**

434 Andolfatto, P., Davison, D., Erezyilmaz, D., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011).
435 Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21,
436 610-617.

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

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

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

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

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

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

450 Cotta-Ramusino, C., McDonald, E.R., Hurov, K., Sowa, M.E., Harper, J.W., and Elledge, S.J.
451 (2010). A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting
452 protein required for ATR signaling. *Science* 332, 1313-1317.

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

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

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

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

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

460 Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., *et al.*
461 (2015). Extensive introgression in a malaria vector species complex revealed by
462 phylogenomics. *Science* 347, 1258524-1-6.

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

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

468 Graze, R.M., Novelo, L.L., Amin, V., Fear, J.M., Casella, G., Nuzhdin, S.V., and McIntyre, L.M.
469 (2012). Allelic imbalance in *Drosophila* hybrid heads: exons, isoforms, and evolution.
470 *Molecular Biology and Evolution* 29, 1521-1532.

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

473 Hollocher, H., and Wu, C.I. (1996). The genetics of reproductive isolation in the *Drosophila*
474 *simulans* clade: X vs. autosomal effects and male vs. female effects. *Genetics* 143, 1243-
475 1255.

476 Holt, R.A., *et al.* (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science*
477 298, 129-149.

478 Hunt, R.H., Coetzee, M., and Fettene, M. (1998). The *Anopheles gambiae* complex: a new species
479 from Ethiopia. *Transactions of The Royal Society of Tropical Medicine and Hygiene* 92, 231-
480 235.

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

483 Kirkpatrick, M., and Hall, D.W. (2004). Male-biased mutation, sex linkage, and the rate of adaptive
484 evolution. *Evolution* 58, 437-440.
485 Kosambi, D.D. (1943). The estimation of map distances from recombination values. *Annals of*
486 *Eugenics* 12, 172-175.
487 Kuniyoshi, H., Baba, K., Ueda, R., Kondo, S., Awano, W., Juni, N., and Yamamoto, D. (2002).
488 lingerer, a *Drosophila* gene involved in initiation and termination of copulation, encodes a set
489 of novel cytoplasmic proteins. *Genetics* 162, 1775-1789.
490 Kyrou, K., Hammond, A., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T., and Crisanti,
491 A. (2018). A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population
492 suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnol* 36, 1062–1066.
493 Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
494 transform. *Bioinformatics* 25, 1754-1760.
495 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,
496 Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The sequence
497 alignment/map (SAM). format and SAMtools. *Bioinformatics* 25, 2078-2079.
498 Linck, R.Q., Chemes, H., and Albertini D. (2016). The axoneme: the propultive engine of spermatozoa
499 and cilia and associated ciliopathies leading to infertility. *Journal of Assisted Reproduction and*
500 *Genetics* 33, 141-156.
501 Lucchesi, J.C., and Suzuki, D.T. (1968). The interchromosomal control of recombination. *Annual*
502 *Review of Genetics* 2, 53-86.
503 Lukacsovich, T., Yuge K., Awano W., Asztalos Z., Kondo S., Juni N., and Yamamoto, D. (2003).
504 The *ken and barbie* gene encoding a putative transcription factor with a BTB domain and
505 three zinc finger motifs functions in terminalia development of *Drosophila*. *Archives of Insect*
506 *Biochemistry and Physiology* 54, 77-94.
507 MacNair, M.R. (1983). The genetic control of copper tolerance in the yellow monkey flower,

508 *Mimulus guttatus. Heredity* 50, 283-293.
509 Macnair, M.R., and Christie, P. (1983). Reproductive isolation as a pleiotropic effect of copper
510 tolerance in *Mimulus guttatus*? *Heredity* 50, 295-302.
511 Maheshwari, S., and Barbash, D.A. (2011). The genetics of hybrid incompatabilities. *Annual Review*
512 *Genetics* 45, 331-355.
513 Mani, G.S., and Clarke, B.C. (1990). Mutational order: a major stochastic process in evolution.
514 *Proceedings of the Royal Society of London, Biology* 240, 29-37.
515 Masley, J.P., Jones, C.D., Noor, M.A.F., Locke, J., and Orr, H.A. (2006). Gene transposition as a
516 cause of hybrid sterility in Drosophila. *Science* 313, 1448-1450.
517 Matute, D.R., Butler, I.A., Turissini, D.A., and Coyne, J.A. (2010). A test of the snowball theory for
518 the rate of evolution of hybrid incompatibilities. *Science* 329, 1518-1521.
519 Matute, D.R., and Gavin-Smyth, J. (2014). Fine mapping of dominant X-linked incompatibility
520 alleles in *Drosophila* hybrids. *PLoS Genetics* 11, e1005558.
521 Moyle, L.C., and Nakazato,T. (2010). Hybrid incompatibility "snowballs" between *Solanum* species.
522 *Science* 329, 1521-1523.
523 Müller, H.J. (1940). Bearing of the *Drosophila* work on systematics. In *The New Systematics*, ed. JS
524 Huxley, pp. 185-268. Oxford, UK. Clarendon Press.
525 Nosil, P., and Flaxman, S.M. (2011). Conditions for mutation-order speciation. *Proceedings of the*
526 *Royal Society of London, Biology* 278, 399-407.
527 Nuwal, T., Kropp, M,. Wegener, S., Racic, S., Montalban, I., and Buchner, E. (2012). The *Drosophila*
528 homologue of tubulin-specific chaperon E-like protein is required for synchronous sperm
529 individualization and normal male fertility. *Journal of Neurogenetics* 26, 374-381.
530 Orr, H.A., and Coyne ,J.A. (1989). The genetics of postzygotic isolation in the *Drosophila virilis*
531 group. *Genetics* 121, 527-537.
532 Orr, H.A., and Irving, S. (2005). Segregation distortion in hybrids between the Bogota and USA

533 subspecies of *Drosophila pseudoobscura*. *Genetics* 169, 671-682.

534 Orr, H.A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities.
535 *Genetics* 139, 1805–1813.

536 Orr, H.A., and Turelli, M. (2001). The evolution of postzygotic isolation: accumulating Dobzhansky-
537 Muller incompatibilities. *Evolution* 55, 1085-1094.

538 Pates, H.V., Takken, W., Curtis, C.F., Huisman, W., Akinpelu, O., and Gill, G.S. (2001b).
539 Unexpected anthropophagic behavior in *Anopheles quadriannulatus. Medical and Veterinary*
540 *Entomology* 15, 293-298.

541 Pates, H.V., Takken, W., Stuke, K., and Curtis, C.F. (2001a). Differential behaviour of *Anopheles*
542 *gambiae sensu stricto* (Diptera: Culicidae) to human and cow odours in the laboratory.
543 *Bulletin of Entomological Research* 4, 289-296.

544 Phadnis, N., and Orr, H.A. (2009). A single gene causes both male sterility and segregation distortion
545 in *Drosophila* hybrids. *Science* 323, 376-379.

546 Presgraves, D.C., and Orr, H.A. (1998). Haldane's rule in taxa lacking a hemizygous X. *Science* 282,
547 952-954.

548 Schultz, J., and Redfield, H. (1951). Interchromosomal effects on crossing over in *Drosophila*. *Cold*
549 *Spring Harbor Symposium on Quantitative Biology* 16, 175-197.

550 Sen, S., Churchhill, G.A. (2001). A statistical framework for quantitative trait mapping. *Genetics*
551 159, 371-387.

552 Slotman, M.A., Della Torre, A., and Powell, J.R. (2004). The genetics of inviability and male sterility
553 in hybrids between *Anopheles gambiae* and *Anopheles arabiensis* (2004). *Evolution* 59, 1016-
554 1026.

555 Slotman, M.A., della Torre, A., and Powell, J.R. (2005). Female sterility in hybrids between
556 *Anopheles gambiae* and *A. arabiensis*, and the cause of Haldane's Rule. *Evolution* 59, 1016-
557 1026.

558 Smith, H.A., White, B.J., Kundert, P., Cheng, C., Romero-Severson, J., Andolfatto, P., and Besansky,
559 N.J. (2015). Genome-wide QTL mapping of saltwater tolerance in sibling species of
560 *Anopheles* (malaria vector). mosquitoes. *Heredity* 115, 471-479.
561 Stevison, L.S., Hoehn, K.B., and Noor, M.A. (2011). Effects of inversions on within- and between-
562 species recombination divergence. *Genome Biology and Evolution* 3, 830-841.
563 Sun, S., Ting, C.T., and Wu, C.I. (2004). The normal function of a speciation gene, *Odysseus*, and its
564 hybrid sterility effect. *Science* 305, 81-83.
565 Szklarczyk, D., Gable, A.L,. Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M.,
566 Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., and von Mering, C. (2019) STRING
567 v11: protein-protein association networks with increased coverage, supporting functional
568 discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47, 607-613.
569 Takken, W., and Knols, B.G.J. (1990). Flight behaviour of *Anopheles gambiae* Giles (Diptera:
570 Culicidae) in response to host stimuli: a windtunnel study. *Proceedings of Experimental and*
571 *Applied Entomology* 1, 121–128.
572 Tao, Y., and Hartl, D.L. (2003). Genetic dissection of hybrid incompatibilities between *Drosophila*
573 *simulans* and *D. mauritiana.* III Heterogeneous accumulation of hybrid incompatibilities,
574 degree of dominance, and implications for Haldane's rule. *Evolution* 57, 2580-2598.
575 Ting, C.T., Tsaur, S.C., Sun, S., Browne, W.E., Chen, Y.C., Patel, N.H., and Wu, C.I. (2004). Gene
576 duplication and speciation in *Drosophila*: evidence from the *Odysseus* locus. *Proceedings of*
577 *the National Academy of Sciences, USA* 101, 12232-12235.
578 True, J.R., Weir, B.S., and Laurie, C.C. (1996). A genome-wide survey of hybrid incompatibility
579 factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into
580 *Drosophila simulans. Genetics* 142, 819-837.

581 Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pastuer, N., Guillet, P., and
582 Raymond, M. (2000). The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae s.s.*
583 through introgression. *Insect Molecular Biology* 9, 451-455.
584 White, G.B. (1974). Anopheles gambiae complex disease transmission in Africa. *Transactions of the*
585 *Royal Society of Tropical Medicine & Hygiene* 68, 278-298.
586 White-Cooper, H., Doggett, K., and Ellis, R.E. (2009). The evolution of spermatogenesis. In: Sperm
587 Biology: An Evolutionary Perspective (TR Birkhead, DJ Hosken & S Pitnick, eds)., 151–183.
588 Academic Press, Burlington, MA.
589 Wu, C.I., Perez, D.E., Davis, A.W., Johnson, N.A., Cabot, E.L., Palopoli, M.F., and Wu, M.L. (1992).
590 Molecular genetic studies of post-mating reproductive isolation in *Drosophila*. In: Takahata N,
591 Clark AG (eds). Molecular Paleo-population Biology. Springer-Verlag, Berlin, 191–212.
592 Wu, C.I., and Beckenback, A.T. (1983). Evidence for extensive genetic differentiation between the
593 sex-ratio and the standard arrangement of *Drosophila pseudoobscura* and *Drosophila*
594 *persimilis* and identification of hybrid sterility factors. *Genetics* 105, 71-86.
595 **Zuo, J., Niu, Y., Cheng, P., Feng, J.Y., Han, S.F., Zhang, Y.H., Shu, G., Wang, Y., Zhang, and**
596 **Y.M. (2019). Effect of marker segregation distortion on high density linkage map**
597 **construction and QTL mapping in Soybean (*Glycine max* L.). *Heredity* 123, 579–592.**
598

599 **Conflict of Interest**

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

602 **Author Contributions**

603 K.C. Deitz, W. Takken and M.A. Slotman designed the experiment. K.C. Deitz and M.A.
604 Slotman, performed the experiment. K.C. Deitz analysed the data. K.C. Deitz and M.A. Slotman
605 wrote the manuscript. W Takken edited the manuscript.

606 **Funding**

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

611 **Acknowledgments**

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

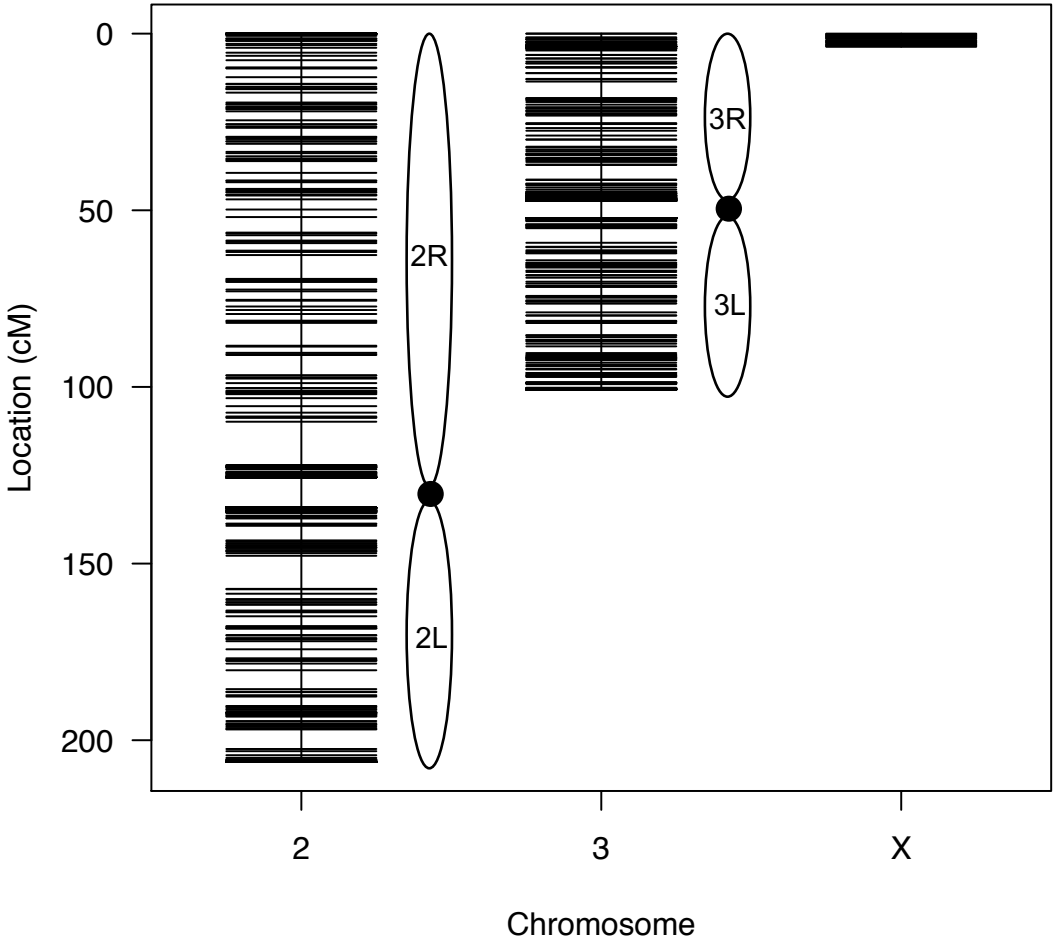
615 **Supplementary Material**

616 Supplementary tables can be found in Deitz_et_al_2020_Supplementary_Tables.docx.

617 **Data Availability Statement**

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

623 **Figures**



624
625 **Figure 1.** Genetic map of the *Acolz* x *Aquad* backcross showing centromere locations and marker density. The centromere of the telocentric
626 X chromosome is opposite zero cM.

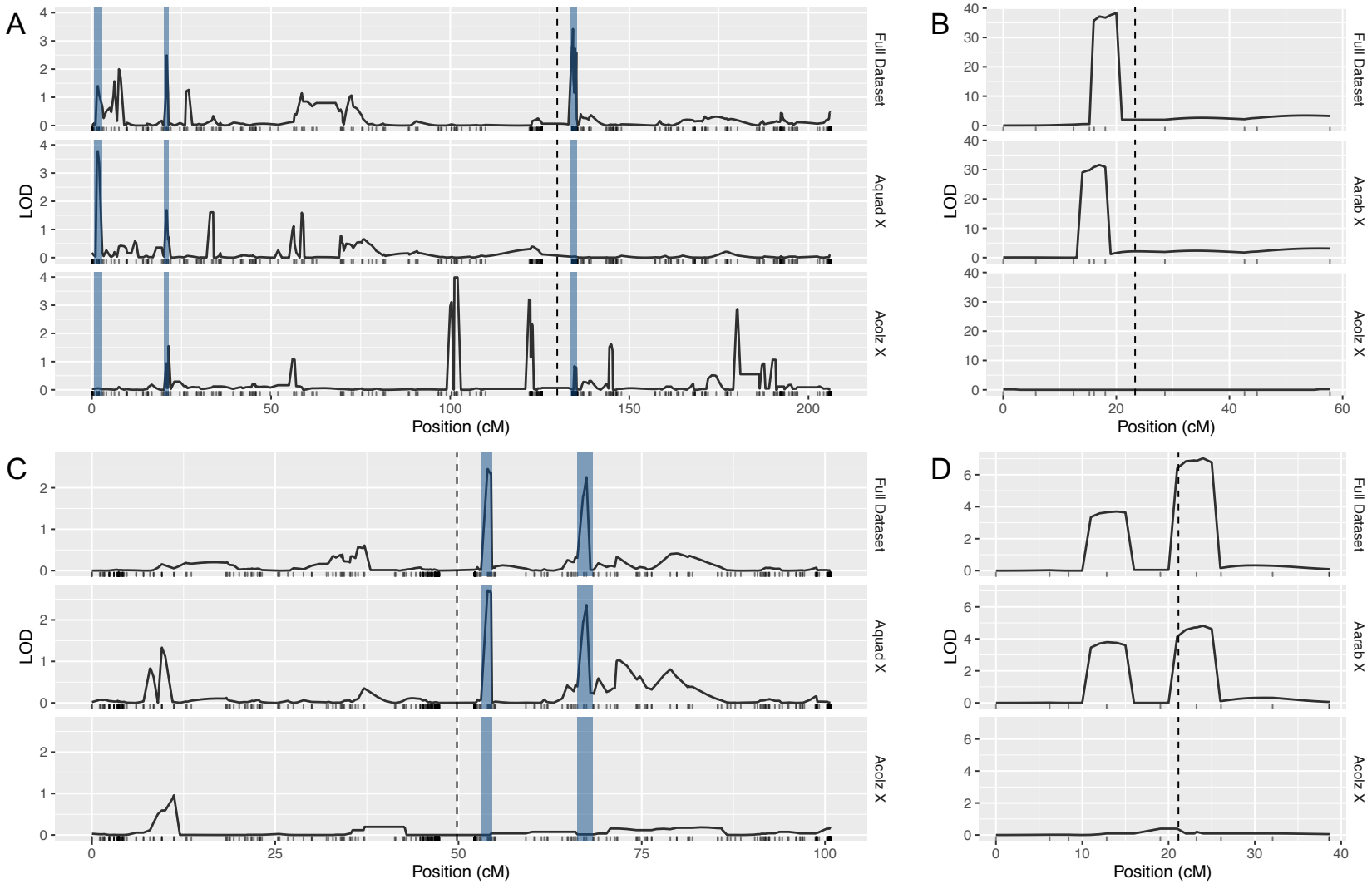


Figure 2. LOD plots for QTL identified with composite interval mapping. Vertical, dashed lines indicate centromere locations. Second and third chromosome right arms are left of the centromere, and left arms are to the right. Horizontal blue bars (A,C) indicate QTL that explain a significant proportion of phenotypic variation in the *An. coluzzii* x *An. quadriannulatus* male sterility multiple imputation QTL model. (A) 2nd chromosome LOD plot of the *An. coluzzii* x *An. quadriannulatus* cross, (C) 3rd chromosome LOD plot of the *An. coluzzii* x *An. quadriannulatus* cross, (B) 2nd chromosome LOD plot of the *An. coluzzii* x *An. arabiensis* cross (Slotman et al., 2004), (D) 3rd chromosome LOD plot of the *An. coluzzii* x *An. arabiensis* cross (Slotman et al., 2004).

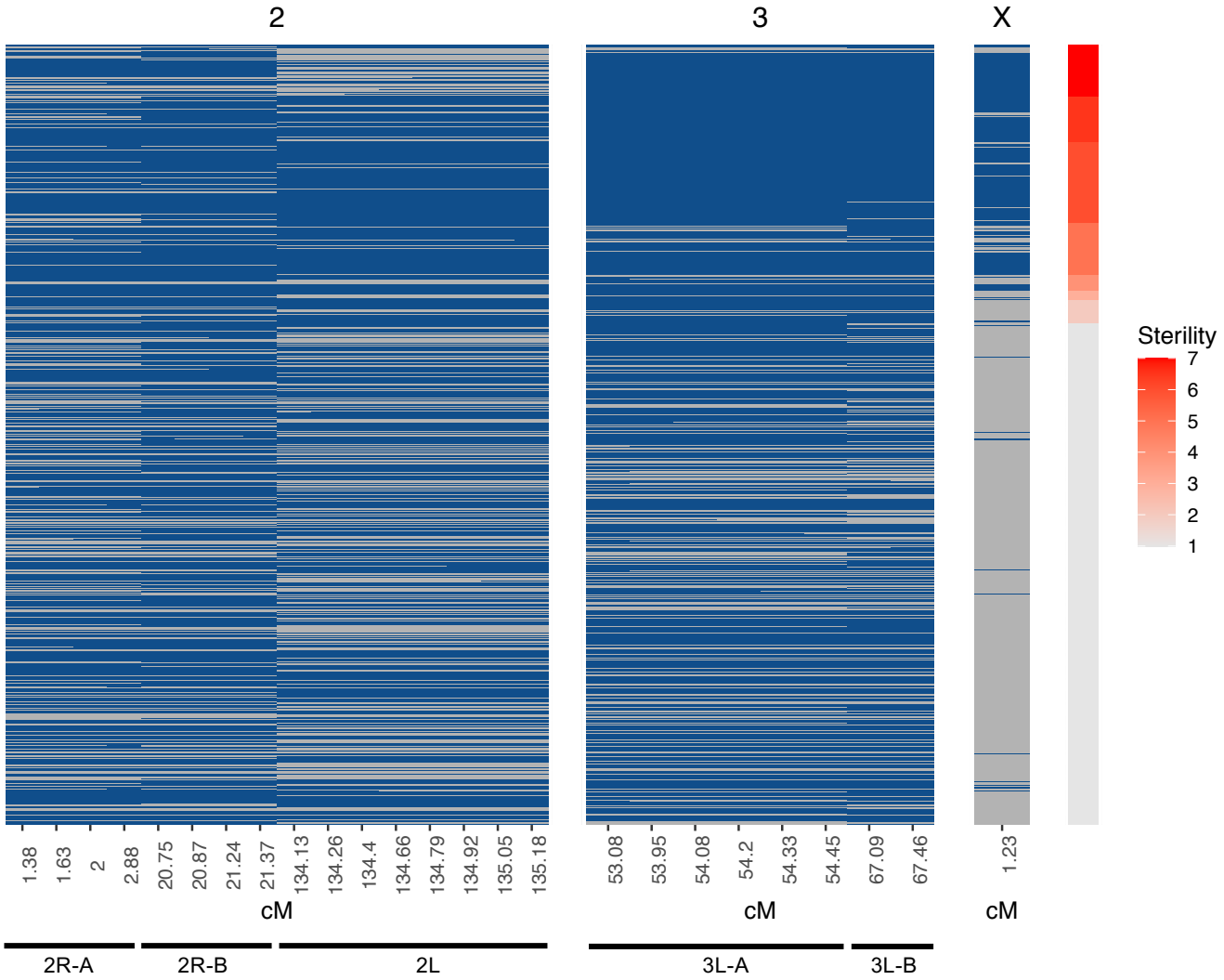


Figure 3. Genotype heat map for each *An. coluzzii* x *An. quadriannulatus* backcross male hybrid (y-axis) at the six QTL (x-axis) identified using CIM and multiple imputation. Individuals are sorted on the y-axis by sterility, with the individuals with the highest sterility score (7) at the top. Blue 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.

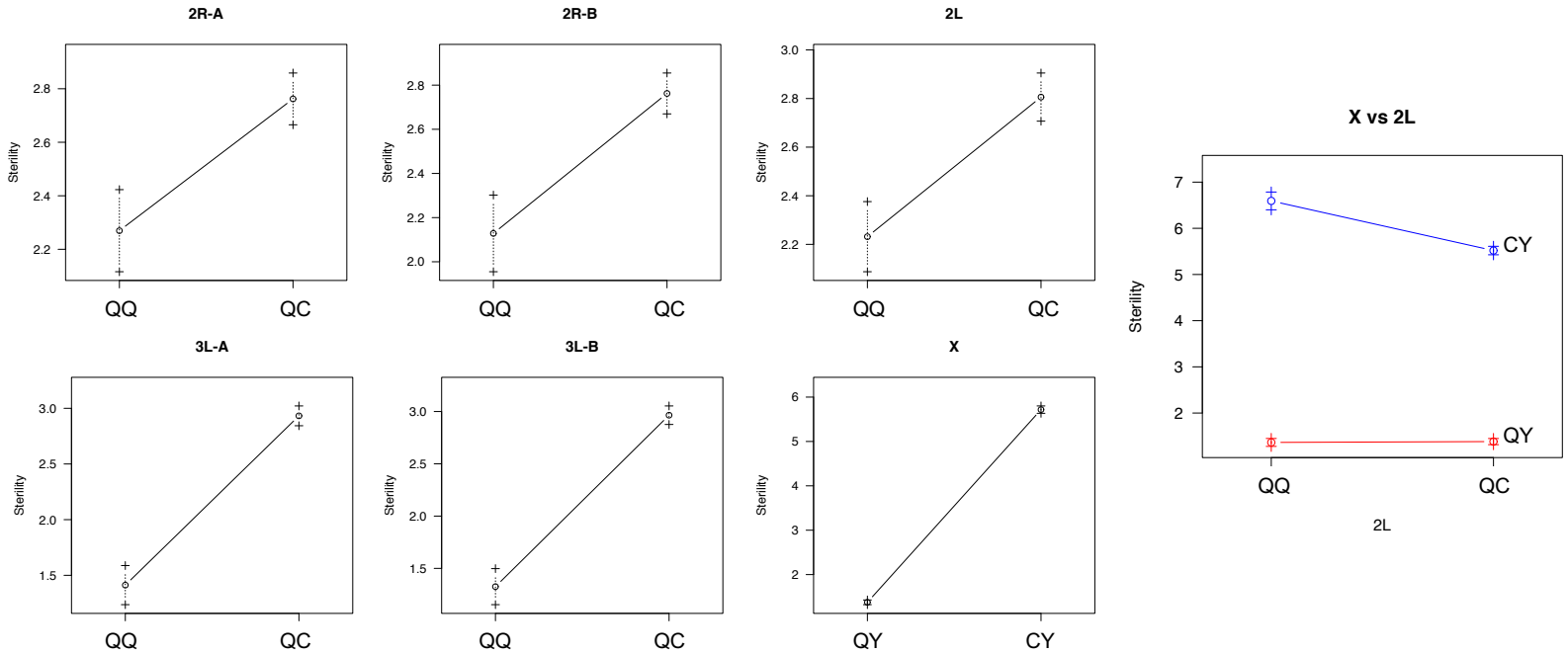


Figure 4. Effect plots for the six QTL identified using CIM and multiple imputation, and an interaction plot (right) for the epistatic effect between the X chromosome and 2L QTL. Blue (*An. coluzzii*) and red (*An. quadriannulatus*) lines in the interaction plot ("X vs 2L", right panel) denote the parental origin of the X chromosome. Marker 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

649 **Table 2**. Genomic regions of QTL in physical (Mb) and map (cM) distances. Bold values indicate overlapping QTL between the crosses on
650 the 3L, a region of the *An. coluzii* genome that contributes to hybrid male sterility when introgressed into both *An. quadriannulatus* and *An.*
651 *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 x *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 x *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	-

652

653 **Table 3.** Percentage of *An. quadriannulatus* homozygosity at hybrid male sterility autosomal QTL peaks in the full, *An. quadriannulatus* X
654 chromosome, and *An. coluzzii* X chromosome datasets. Expected homozygosity values are based on the allele frequency of the total *An.*
655 *coluzzii* x *An. quadriannulatus* backcross population assuming Hardy-Weinberg equilibrium. p-values are reported for Chi-square tests,
656 which were used to test for significant deviation from expected homozygosity.
657

		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

658

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 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				

662 **Table 5.** Annotated genes in the AgamP4 genome within autosomal QTL with functions related to mating, reproduction, and sperm motility
663 and morphogenesis. Gene names in bold indicate those interacting in a regulatory network associated with sperm microtubule
664 morphogenesis and function. STRING network analysis output can be found in Supplementary File 1:
665 Deitz_et_al_2020_String_protein_interaction_network.xlsx.
666

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

667

668 **Table S1**. The complete list of annotated genes within autosomal QTL and their PANTHER gene ontology families / sub-families (if
669 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)

670