Mechanisms of transcriptional regulation in *Anopheles gambiae* revealed by allele specific expression

Authors

Naomi A. Dyer[1], Eric R. Lucas[1], Sanjay C. Nagi[1], Daniel McDermott[1], Jon H. Brenas[2], Alistair Miles[2], Chris S. Clarkson[2], Henry D. Mawejje[3], Craig S. Wilding[4], Marc S. Halfon[5], Hasiba Asma[5], Eva Heinz[1,6], Martin J. Donnelly[1]

Short title

Allele specific expression in *Anopheles gambiae*

Affiliations

1 Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

2 Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

3 Infectious Diseases Research Collaboration (IDRC), Plot 2C Nakasero Hill Road, P.O.Box 7475, Kampala, Uganda

4 School of Biological and Environmental Sciences, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK

21 5 Department of Biochemistry, Jacobs School of Medicine & Biomedical Sciences, University
22 at Buffalo-State University of New York, 955 Main Street, Buffalo, New York 14203, USA
23 6 Department of Clinical Sciences, Liverpool School of Tropical Medicine, Pembroke Place,
24 Liverpool, L3 5QA, UK
25

26 **Keywords**

27 Allele, cis-regulation, transcript, insecticide, resistance

28 **Correspondence**

29 Naomi.Dyer@lstmed.ac.uk
30

31 **Abstract**

32 Malaria control relies on insecticides targeting the mosquito vector, but this is increasingly
33 compromised by insecticide resistance, which can be achieved by elevated expression of
34 detoxifying enzymes that metabolize the insecticide. In diploid organisms, gene expression
35 is regulated both in *cis*, by regulatory sequences on the same chromosome, and by *trans*
36 acting factors, affecting both alleles equally. Differing levels of transcription can be caused
37 by mutations in *cis*-regulatory modules (CRM), but few of these have been identified in
38 mosquitoes. We crossed bendiocarb resistant and susceptible *Anopheles gambiae* strains to
39 identify *cis*-regulated genes that might be responsible for the resistant phenotype using
40 RNAseq, and *cis*-regulatory module sequences controlling gene expression in insecticide
41 resistance relevant tissues were predicted using machine learning. We found 115 genes
42 showing allele specific expression in hybrids of insecticide susceptible and resistant strains,
43 suggesting *cis* regulation is an important mechanism of gene expression regulation in

Anopheles gambiae. The genes showing allele specific expression included a higher proportion of *Anopheles* specific genes on average younger than genes those with balanced allelic expression.

Introduction

Malaria prevalence in Sub-Saharan Africa has reduced by 50% since 2000, primarily due to insecticide-based control of mosquito vectors (1). Recently, progress has stagnated (2), partly due to increasing levels of resistance against insecticides in mosquito populations (3). *Anopheles gambiae* is one of the dominant malaria vectors in Sub-Saharan Africa, the primary vector across most of Uganda (4) and the vector for which the largest resource of genome data is available, with 7275 genomes sequenced (5-8). A common cause of insecticide resistance is increased degradation of insecticides (termed metabolic resistance) (9) with overexpression of insecticide-metabolizing P450s repeatedly implicated (10-12). This can be caused by mutations in *cis*-regulatory regions regulating expression of metabolic resistance genes. In diploid organisms, such mutations only affect expression of the allele of the gene located on the same chromosome. Although some *trans* factors involved in metabolic resistance gene regulation in *Anopheles* are known (13, 14), few studies have identified genetic variation causing metabolic resistance (15-18). The multiallelic nature of metabolic insecticide resistance which can involve different mutations affecting the same gene in different populations, as well as the involvement of multiple genes, makes marker identification challenging as it limits the power of association studies unless very large sample sizes are used (15).

66 Despite the primary role of gene overexpression in metabolic resistance, only one *cis*-
67 regulatory variant for resistance-linked differential expression has been identified in
68 *Anopheles gambiae* (19), and markers for such variants are therefore absent in the current
69 genetic marker panel for resistance [18]. Copy number variants (CNV) have been observed
70 in *Anopheles gambiae* metabolic resistance gene clusters (20). For example, in *Anopheles*
71 *coluzzi* copy number of *Cyp6AA1* is associated with deltamethrin resistance (21). The
72 relative contribution of CNV and *cis*-regulation on *Anopheles* gene expression has not yet
73 been determined.
74 Uganda sees a high burden of malaria; comprising 7.8% of all global cases in 2021 (22). To
75 address this public health burden, the insecticide bendiocarb has been used for indoor
76 residual spraying (IRS) to complement the distribution of long-lasting insecticidal nets (LLIN).
77 Some resistance to bendiocarb in *An*. *gambiae* was observed in Nagongera (southeast
78 Uganda) and Kihihi (southwest Uganda), with 83% and 70% mortality, respectively, to WHO
79 bioassays with a diagnostic dose of 0.1% bendiocarb in 2014 prior to the IRS (23). The IRS
80 campaign starting in December 2014 succeeded in reducing human biting rate, cases of
81 malaria and test positivity rate in Nagongera (23), but the potential for further increases in
82 resistance puts the long-term usefulness of bendiocarb into question.
83 Mosquitoes collected from Nagongera in 2014, which have moderate resistance to
84 bendiocarb, showed significant differential expression of many genes compared to the
85 susceptible Kisumu strain, including salivary gland protein encoding *D7r2* and *D7r4* genes as
86 well as the detoxification associated genes *Gstd3* and *Cyp6m2* (24). Expression of *D7r4* was
87 associated with a single-nucleotide polymorphism (SNP) in a non-coding transcript
88 downstream of the D7 cassette (24).

89 In diploid organisms, allele specific expression (ASE) provides strong evidence that genes
90 may be under differential *cis*-regulatory control [19] (Figure 1). Using a method that has
91 been applied to a variety of taxa [19-24] but not mosquitoes, we describe the identification
92 of genes showing ASE in *An. gambiae* which potentially confer metabolic resistance. The
93 challenges of applying this method that arise from mosquito biology and genome structure
94 are discussed. In complementary work we predicted the sequences of some of the *cis*
95 regulatory modules that may underlie the expression of genes involved in insecticide
96 resistance using machine learning. Predictions included potential CRMs proximal to the
97 genes showing ASE and genes that show consistent differential expression patterns in
98 multiple resistant *Anopheles* strains, providing a starting point for future investigations into
99 CRM variants during the evolution of insecticide resistance.
100

101 # Materials and Methods

102 ## Strains and Crosses

103 Resting female mosquitoes were collected in Siwa Village, Nagongera, Tororo District in
104 Eastern Uganda (0°46'12.0"N, 34°01'34.0"E) in March 2013 and allowed to lay eggs. 65
105 *Anopheles gambiae* (as ascertained by species ID PCR (25)) egg batches laid by the collected
106 females were reared to larvae and combined to establish the Nagongera colony. All the
107 colony founding mothers were screened for a known bendiocarb target site resistance
108 mutation G119S in the *Ace1* gene using the TaqMan assay described in (26). No resistance
109 associated variants were observed and full sequencing of the *Ace1* gene did not reveal any
110 other potential target site resistance mutations. The colony was assayed using WHO
111 diagnostic doses at F2 and was found to be highly resistant to DDT and deltamethrin,

112 intermediate resistance to the carbamate insecticide bendiocarb and was highly resistant to
113 pyrethroids and DDT. Bioassays using the WHO standard protocol (27, 28) with preexposure
114 to 4% piperonyl butoxide (which inhibits P450 mediated metabolism) for one hour prior to
115 bendiocarb exposure increased mortality compared to bendiocarb exposure alone, implying
116 that bendiocarb resistance is metabolic rather than target site mediated.
117 A schematic of the experimental design is shown in figure 1. In brief, reciprocal crosses
118 between Nagongera mosquitoes and an inbred insecticide susceptible colony (origin
119 Kisumu, Kenya) were performed in cages as described previously (6). Reciprocal crosses
120 each involved a total of 13 males and multiple females; since *Anopheles gambiae* are swarm
121 maters, multiple males are required to induce mating. Following mating, females were
122 transferred to individual cups for egg laying. F1 progeny of three females from each
123 reciprocal cross were raised to adults under standard insectary conditions of 12 hours light
124 12 hours dark cycle, 26°C +/-2°C and 70% relative humidity, and fed on 10% sucrose
125 solution.

126 **Sequencing**

127 RNA was extracted from pools of ten female F1 progeny from each of the six crosses 3-5
128 days after eclosion using RNAqueous4PCR total RNA isolation kit (Invitrogen). RNA quality
129 and quantity was checked to be adequate for library preparation (Agilent Bioanalyser and
130 Qubit® 2.0 Fluorometer). Total RNA libraries were prepared for Illumina paired-end indexed
131 sequencing according to the Illumina TruSeq RNA sample preparation v2 guide (29). cDNA
132 libraries were bar coded, pooled and sequenced using the Illumina HiSeq1500 platform,
133 with 100bp paired end reads.
134 DNA was extracted using the Qiagen DNeasy Kit from the 6 individual mothers that laid
135 viable eggs, all individual potential fathers that were alive following the cross, and individual

136 male F1 siblings (Figure 2) and sequenced using the Illumina HiSeq 2000 as described
137 previously (6).

138 Genome analysis

139 Quality control of genomic sequences for mothers, potential fathers and F1 progeny, and
140 matching of fathers to the progeny using Mendelian Error analysis was performed as
141 described previously [6]. Briefly, for the autosomes for each cross where the mother's
142 genome sequence was of sufficient quality (K2, K4, K6 and B5), the per progeny sample
143 Mendelian error was calculated for each possible parental pair of the known mother with all
144 potential fathers using scikit-allel (30). The father in the parental pair that had the smallest
145 median progeny sample Mendelian error on at least 3 autosomal arms was deemed to be
146 the true father.
147 Scikit-allel [32] was used to analyze the genome of potential fathers including generating a
148 consensus sequence of homozygous sites for Kisumu and Nagongera colonies, and
149 calculating single nucleotide polymorphisms (SNPs) differing between the colonies.

150 RNAseq analysis

151 Samples were compared using the rna-seq-pop snakemake workflow (31) which includes
152 quality control of RNAseq reads using with FastQC (32) and MultiQC (33), read alignment to
153 the *Anopheles gambiae* PEST reference genome (AgamP4, INSDC Assembly
154 GCA_000005575.1, Feb 2006) using HISAT2 (34), principal components analysis to group
155 samples by gene expression, differential expression analysis using Kallisto (35) with DESeq2
156 (36), variant calling using Freebayes version 1.3.2 (37) with ploidy 20 (as 10 diploid
157 individuals were pooled in each sample), calculation of population diversity statistics with
158 scikit-allel (30), and gene set enrichment analysis using GSEA (38). RNAseq data for the

159 Kisumu parental colony, and Busia G28 deltamethrin selected colony (31) was compared
160 with the cross progeny RNAseq data in this study.

161 **Analysis of allele specific expression (ASE)**

162 For parent-based mapping, SNPs that were homozygous for different alleles in the parents
163 were used to distinguish whether reads originated from a maternal or paternal allele: due to
164 the pooling of 10 females for RNA sequencing it was not possible to use parental
165 heterozygous SNPs. For sibling-based mapping, SNPs that distinguished maternal and
166 paternal alleles on the autosomes were inferred from the genome sequences of the male
167 siblings, using the following criteria: SNPs must be heterozygous and biallelic in all siblings
168 where data was available. To eliminate potential low-quality SNPs, if 5 or more siblings had
169 missing data at a SNP it was not used. Since F1 genome data was unphased, it was not
170 possible to assign SNPs to either the Kisumu or Nagongera parent using this method.
171 Analysis of F1 RNAseq data for ASE was performed using ASEReadCounter* (39), which is
172 based on ASEReadCounter (40), using individual parent genomes where available for each
173 cross to assign reads to parental alleles (options –vcf_mat, -vcf_pat). Where only sibling
174 inferred SNPs were known, option –vcf_joind was used on the unphased inferred F1 VCF
175 file.
176 The properties of SNPs inferred from parents and siblings were compared by inferring
177 parental pseudogenomes and running ASEReadCounter* with both sets of SNPs. SNP level
178 ASE was calculated for all SNPs with >=10 total counts. SNPs which were inferred from
179 siblings and not from parents, which had >100 counts and ASE < .25 or > .75 were removed
180 from the sibling inferred SNPs for subsequent analysis.
181 Due to overdispersion, RNAseq data is not binomially distributed but can be better
182 described by a beta binomial distribution. The distribution of counts data and

183 overdispersion parameter rho for the beta binomial distribution was calculated using the
184 counts data for each cross, including the counts for the SNP with the maximum counts in
185 each gene (with minimum 10 counts), using the R package glmmTMB (41), which uses
186 maximum likelihood estimation to fit binomial and beta binomial model to the data for each
187 cross. Models were compared using ANOVA.
188 To obtain a gene level measurement of ASE, the R package MBASED (42) was used to
189 aggregate SNP level counts to gene level ASE by pseudo phasing SNPs in each gene such that
190 the SNPs with the higher read count at each variable site are combined into the major allele,
191 with frequency between 0.5 and 1. Only sites with at least ten counts were used. For SNPs
192 inferred from the siblings of the RNAseq pool, MBASED was run assuming SNPs were not
193 phased. Where phase was known from the parents, MBASED was run in both phased and
194 non-phased mode to measure the impact on power to detect loci with ASE. In all cases the
195 same seed (988482) was set prior to running 10^6 simulations in which the total read count at
196 each SNP position in each gene are kept constant, but reference allele counts are drawn
197 from a null distribution beta binomial (mean 0.5 x total count, rho 0.038). The P value is the
198 proportion of simulations in which the major allele frequency is greater than or equal to the
199 observed major allele frequency. False discovery correction was applied with a nominal rate
200 of 5% (43).
201 For SNPs inferred from parents, allelic count tables were additionally analyzed using Qleelic
202 (39) with a binomial test of the null hypothesis of equal numbers of transcripts arising from
203 maternal and paternal alleles. ASE was considered significant at p=0.05 following Bonferroni
204 correction. Genes with fewer than ten read counts were excluded. As technical replicates
205 were unavailable, an exploratory solution for correcting overdispersion was splitting reads
206 into groups (pseudoreplicates K4 14 groups, K2 three groups, K6 nine groups, B5 nine

207 groups). This produced a consistent quality correction constant (QCC) value of 1.56 for QCC
208 correction. R packages UpSetR (44), stats and ggplot2 (45) were used for regressions and
209 plots.

210 Enrichment analysis

211 Hypergeometric tests for enrichment of gene ontology terms, Pfam protein domain
212 enrichment, and KEGG pathways were implemented using python (46). The test set for each
213 cross was the set of genes showing significant ASE following FDR correction, and the
214 "universe set" was the set of genes for each cross that had at least ten reads containing
215 SNPs distinguishing maternal and paternal alleles (for these genes ASE would have been
216 detectable if present).
217 The evolutionary age of genes was taken from the Princeton Protein Orthology Database
218 PPODv4_PTHR7-OrthoMCL (47), with age estimates based on Wagner parsimony. A gene
219 age enrichment test was implemented using the age_enrichment.py script in
220 ProteinHistorian (48), using option '-a ignore' to ignore the proteins not present in the
221 database. P values were obtained using a Mann Whitney U test for difference in overall age
222 distribution between sets and Fisher's exact test for difference in fraction of proteins in each
223 age in the set.
224 The null hypothesis, the proportion of genes showing significant allelic imbalance of
225 expression does not differ between autosome arm, was tested using a two-sided Fisher's
226 exact test (for those crosses where the counts were too high for efficient computation P
227 value was simulated using 2000 replicates).
228 To assess selective sweeps, lists of genes showing ASE were compared to a database of
229 regions under selection as defined by H_{12} signal (49) both across the range of *Anopheles*

230 *gambiae* and restricted to Uganda using a custom Python script

231 (https://github.com/azurillandfriend/sweeps_ASE.git).

232 Prediction of *cis*-regulatory modules

233 *Anopheles* cis-regulatory modules (CRM) were predicted using SCRMshaw s (50-53). Briefly,

234 *Drosophila* CRM which drive gene expression in the tissue of interest were downloaded

235 from the Redfly database (v9.6.0, database updated 02/01/2023) (54, 55),with max size

236 2000. Only non-overlapping sequences (>100bp) were included. Syntenous regions from

237 related *Drosophila* species (putatively containing the equivalent CRM) were added to the

238 training data set using liftOver at https://genome.ucsc.edu/cgi-bin/hgLiftOver (56) as

239 described in (51). SCRMshaw was trained using this augmented training set and a 10x bigger

240 set of non-CRM non-exonic regions. Repeats in CRM, non-CRM and the target *Anopheles*

241 *gambiae* PEST genome were masked using repeat masker (57). Existing training sets for

242 adult peripheral nervous system, embryonic and larval excretory, and embryonic/ larval

243 Malpighian tubules were downloaded from GitHub

244 (https://github.com/HalfonLab/dmel_training_sets). These were compared with the 2272

245 Transposase Tn5 hypersensitive sites identified by Ruiz et al (58).

246 Results and Discussion

247 Crosses

248 Three females from each reciprocal cross between the Kisumu and Nagongera strains laid

249 viable eggs (Figure 1, Table 1). For four crosses the father was identified by minimizing

250 median Mendelian error (Supplementary Table 1, but for the other two crosses the mother

251 failed sequencing quality control and none of the sequenced putative fathers were a good

252 match. We presume the true fathers for these crosses died during the experiment,
253 precluding extraction of good quality DNA.
254 RNAseq statistics for the pooled F1 females from each of these six crosses are shown in
255 Table 1. The number of reads recommended for 60% power to detect ASE at 1.5 fold is 500
256 per gene [37], which for the 13,796 annotated genes in *Anopheles gambiae* would require
257 around 6.9×10^6 reads. Five crosses exceeded this value with one cross, K2, having slightly
258 fewer (6.4×10^6reads). The proportion of reads mapping to the *Anopheles gambiae* genome
259 was similar for all crosses (mean 87.0%, standard deviation 1.1%).

260 Between pools gene expression comparison

261 Comparison between reciprocal crosses revealed similar gene expression in the F1 progeny,
262 with just 38 genes significantly downregulated and 38 genes significantly upregulated in the
263 F1 between reciprocal crosses at $P_{adj} \leq 0.05$ (Figure 2a). When comparing only F1 crosses in
264 PCA, B1 seemed to be an outlier in terms of expression (Supplementary figure 1a). However,
265 when compared with other colonies, the F1 progeny formed a distinct group
266 (Supplementary figure 1b).
267 F1 gene expression differed from the Kisumu (parental) colony, with a total of 1362 genes
268 upregulated and 1265 genes downregulated in the F1 compared to the Kisumu colony (Padj
269 ≤ 0.001) (Figure 2b). RNAseq data was unavailable for the Nagongera colony, which no
270 longer exists. Gene set enrichment analysis of the differentially expressed genes indicated
271 that genes with Gene Ontology (GO) terms associated with odorant binding, olfactory
272 receptor activity, sensory perception of smell, response to stimulus, detection of chemical
273 stimulus involved in sensory perception of smell, structural constituent of cuticle and signal
274 transduction were upregulated in the F1 compared to Kisumu, whereas those with GO
275 terms associated with translation, the ribosome, mitochondrion, structural component of

276 cuticle, and serine type endopeptidase activity were downregulated (all at P adjusted 0.05).
277 KEGG pathway analysis showed significant upregulation of ribosome, citrate cycle (TCA
278 cycle) and oxidative phosphorylation in the F1 compared to Kisumu.
279 Although we did not perform RNAseq on the Nagongera colony, a previous study compared
280 gene expression in mosquitoes collected from the same location in 2014 (one year later
281 than the colony founders) with the susceptible Kisumu colony using microarrays (24). We
282 compared the genes that show significant differential expression between Kisumu and F1 in
283 RNAseq (4848 genes) and the earlier microarrays' 3363 genes. A total of 1320 genes were
284 on both these lists, including *D7r4* which was upregulated in both Nagongera and the cross
285 compared to Kisumu. However, only in 645/1320 genes was the direction of fold change the
286 same in both comparisons highlighting the difficulty of comparing microarray and RNAseq
287 expression studies, that gene expression may have changed both during the 2013-2014
288 period as well as during selection of the Nagongera colony, and possible over and
289 underdominance of expression in the F1 .

290 Selection of SNPs to detect ASE

291 Detection of ASE relies on sufficient SNP differing between the parents (59). The number of
292 SNPs differing between maternal and paternal genomes varied (Table 2). For crosses B1 and
293 B3, with parent genomes unavailable, we instead attempted to use parental colony
294 consensus sequences for mapping. However, the number of SNPs differing in Nagongera
295 and Kisumu consensus sequences (Table 3) was insufficient for ASE analysis. The low
296 number of differing SNPs may be due to barriers to recombination imposed by the genome
297 structure of *Anopheles gambiae*, which limit the extent to which colonies become inbred,
298 with heterozygosity persisting even after many generations of inbreeding (60). This
299 contrasts with *Drosophila* where colony genotypes have been used effectively for inference

300 of ASE (61, 62). Since the mosquitoes were pooled prior to RNA extraction, inferring SNPs
301 from the RNAseq data to detect ASE is problematic, since most SNPs are heterozygous in
302 one or both parents, resulting in a mixture of homozygous and heterozygous SNPs in the
303 progeny, and we would not know whether to test for deviation from 1:1 expression or 3:1.
304 To increase the number of SNPs to detect ASE in these pooled samples we therefore used
305 the siblings of the RNA sequenced F1 to infer which SNPs were homozygous but different
306 between the parents (opposite homozygous SNPs). For a Mendelian inherited SNP, opposite
307 homozygous SNP in parents should result in all progeny having heterozygous biallelic SNP. If
308 one or both parents were heterozygous (biallelic) then 50% of the progeny are expected to
309 be heterozygous, and the probability of all the progeny being heterozygous in this situation
310 is given by the binomial probability density function where p=0.5, the number of trials (per
311 SNP) is the number of progeny and the number of "successes" is also the number of
312 progeny. The number of progeny and SNPs per chromosome arm where all progeny were
313 heterozygous is given in Supplementary Table 2. Almost all sequenced siblings were male,
314 the heterogametic sex in *An. gambiae*, so it was not possible to infer SNPs on the X
315 chromosome using this method.
316 The suitability of these SNPs to detect ASE was first tested by comparing the SNPs inferred
317 from the siblings to those inferred from the parents, where parental genotypes were known
318 (Supplementary table 3). Most SNPs were shared between the parents and siblings, with
319 only a small fraction (<7.5%) unique to the siblings. The total number of usable SNPs from
320 the siblings was lower than those in the parents: a reduction of between 6% and 16%. Per
321 SNP ASE was calculated as counts at reference SNP/counts at reference SNP plus counts at
322 alternative SNP for the sibling SNPs, and as counts at maternal SNP/ counts at maternal plus
323 paternal SNP for parent SNPs. The mean per SNP ASE for SNPs with at least ten reads

324 counted remained between 0.49 and 0.51 confirming that the SNPs are likely to be
325 heterozygous in the F1 and that mapping bias is not a problem for this dataset. The standard
326 deviation of per SNP ASE increased very slightly (from 0.12-0.14 to 0.13 to 0.15). For crosses
327 B1 and B3, the per SNP ASE was in line with the other crosses with mean 0.5 and standard
328 deviation 0.13 and 0.15 respectively. However, when the per SNP ASE was plotted against
329 the total counts at each SNP, a cluster of SNPs a with high ASE and expression was seen for
330 the sibling but not parent inferred SNPs (Figure 3). SNPs in crosses K2, K4 and B5 that were
331 unique to siblings and with high ASE and high expression were therefore removed from the
332 analysis for crosses B1 and B3 before inferring per gene ASE.

333 **ASE inference**

334 For the four crosses where the parent was known, per gene ASE was calculated as reads
335 mapping to maternal genome/total reads mapped at that gene. In addition the major allele
336 frequency at each gene was calculated for all crosses. Read count data were fitted to
337 binomial and beta binomial models. ANOVA indicated that the beta binomial model was a
338 better fit to the data ($P < 2.2e-16$). The dispersion parameter phi(disp) for the beta binomial
339 distribution of counts data for crosses B1, B3, B5, K2, and K4 was estimated at 25.8, giving a
340 rho 1/(1+phi(disp)) of 0.038. Per gene ASE statistics for all crosses are shown in Table 4. Due
341 to variations in parental genome and sequencing depth the power to detect ASE varies
342 between crosses, so it is not possible to infer whether the proportion of genes with ASE is
343 different between the crosses. Three crosses showed similar numbers of genes with
344 maternal or paternal ASE, whereas cross K6 surprisingly seemed to show extreme paternal
345 ASE. Using SNPs from F1 male siblings to infer ASE also showed a larger number of genes
346 with ASE in cross K6 than the other crosses. To investigate whether K6 shows a very
347 different pattern of ASE to the other crosses, or a sampling error occurred so that the RNA

348 sequenced samples were not from the same cross as the parents and siblings. We therefore
349 checked the effect of having selected the wrong SNPs to infer allelic imbalance. SNPs
350 inferred from parents of the other crosses were used to infer allelic imbalance for cross K6.
351 If incorrect SNPs are used ASE is artificially inflated due to counting reads at sites where
352 some or all the progeny are homozygous. When the SNPs from the parents of cross K4 were
353 used to infer allelic counts for cross K6, the mean imbalance was shifted back to 0.5,
354 suggesting that these SNPs are more suitable than the erroneously inferred parents (figure
355 3, Supplementary table 4). Furthermore, when SNPs called from the RNAseq data were
356 compared, the progeny of cross K4 and K6 were extremely genetically similar and clustered
357 closely in PCA for all chromosomal arms, suggesting they may have hatched from two egg
358 batches from the same cross. The analysis for cross K6 was therefore repeated using SNPs
359 from K4 parents.
360 Genes showing ASE showed some overlap between the different crosses (Figure 4). At the
361 most conservative estimate, 13 genes showed ASE in all crosses. This exceeds the overlap
362 expected by random genes showing allelic imbalance, as in 100,000 simulations of randomly
363 drawing the observed number of significant genes from the set genes where SNPs were
364 available to detect ASE in all crosses, the maximum expected overlap was 1. Most genes
365 showing ASE were unique to individual crosses (Figure 4), but there were many genes
366 showing ASE in combinations of multiple crosses that may also be under consistent
367 differential *cis*-regulation between Tororo and Nagongera colonies, Table 5 shows the 115
368 genes with significant ASE in at least 4 out of the 6 crosses. Directionality of ASE was
369 inferred for crosses B5, K2, K4 and K6 (Supplementary Table 7).
370

371 Analysis of gene ages that showed or did not show ASE revealed an enrichment of younger,
372 *Anopheles* specific genes showing significant ASE (Figure 5 and Supplementary Table 5). The
373 same trend was seen in all crosses.
374 Genes showing ASE showed an unequal distribution along chromosomal arms compared to
375 genes not showing ASE (Supplementary table 6, Figure 6, Supplementary figure 2). This
376 trend was significant for all crosses except for K2. The trend was for slight enrichment on
377 3L. For the 4 crosses where parents were available, this was also compared with the X
378 chromosome, which had a lower proportion of ASE/total detectable genes (Figure 6,
379 Supplementary figure 2). ANOVA indicated both chromosome and cross explained the
380 variance in the proportion of ASE/total detectable genes per chromosome arm ($P < 0.05$).

381 **Copy number variation**

382 True ASE is caused by *cis*-regulation, but copy number variation of the expressed gene could
383 lead to apparent ASE if there are different numbers of gene copies containing the SNPs used
384 to count reads; e.g. if two copies of a duplicated gene with total of three copies bear one
385 SNP and the third the other SNP, apparent ASE would be inferred without differential *cis*-
386 regulation. We therefore checked for CNV in the parental and F1 sibling genomes at the
387 genes that showed ASE in at least four out of six crosses, and in the genes which contained
388 sufficient SNPs to infer allelic imbalance but expression appeared to be in balance. Copy
389 number variation was deemed possible if any of the parents and F1 siblings from crosses B5,
390 K2, K4 or K6 (total 63 individuals) had a copy number call other than 2 on the autosomes. As
391 ASE could not be inferred on the X chromosome for crosses B1 and B3 we excluded it from
392 the analysis. A total of 60/114 (53%) autosomal genes showing ASE contained possible CNV.
393 For genes that did not show ASE and had available SNPs to infer ASE in all crosses, 481/
394 1333 (36%) had possible CNV. A two-sided Fisher's exact test rejected the hypothesis that

395 the odds ratio is 1 (odds ratio 2.0, P=0.0006). This suggests that for some genes CNV may be
396 contributing to the observed ASE. The number of individuals with a possible CNV in each
397 gene varied between the genes (Table 5)
398

399 **Overlaps**

400 The list of 115 genes showing ASE in at least four crosses was compared with published
401 gene expression data. For all other available data, it is not known to what extent the
402 expression is *cis* or *trans* regulated, meaning that lack of consistency between the datasets
403 may be due to the absence of *trans* regulated genes in the ASE data.
404 Genes showing ASE in at least four crosses were compared with genes with consistently high
405 median fold change in a meta-analysis of 35 experiments comparing RNAseq data between
406 *Anopheles gambiae* s.l. and *Anopheles funestus* strains (46), and with genes showing
407 significant fold change with consistent directionality in microarray data comparing
408 susceptible and resistant populations of *Anopheles coluzzi* (63). For the RNAseq metadata
409 set, the genes with the top 5% of median fold changes between susceptible and resistant
410 populations (429/8599 total genes in the dataset), six genes were also present in the ASE
411 gene set. These were AGAP001251 (*Eupolytin*), AGAP008218 (*Cyp6Z2*) and AGAP012296
412 (*Cyp9J5*), AGAP011068 (A*ldose reductase*), AGAP003583 (*L-iditol 2-dehydrogenase*) and
413 AGAP008331 (*WD repeat-containing protein 59*). There was no overlap with the 40 genes
414 showing in microarray.
415

416 Of the gene families previously implicated in metabolic resistance, we observed four P450s
417 with ASE in four of the six crosses: *Cyp12F3* (AGAP008019), *Cyp12F2* (AGAP008020), *Cyp6Z2*
418 (AGAP008212) and *Cyp9J5* (AGAP012296).

419

420 The D7 protein family has previously been implicated in bendiocarb resistance in Uganda
421 (24). AGAP008282 (*D7r2*) showed significant ASE in two crosses (detectable in 4) and *D7r4* in
422 none (detectable in 3). Other *D7r* genes, *D7r1* (AGAP008284), *D7r5* (AGAP008280) also
423 showed significant ASE in two crosses, *D7L1*, *D7L2* and *D7r3* in one cross.

424

425 Gene set enrichment analysis did not reveal consistent enrichment of any particular gene
426 ontology term amongst genes showing ASE across the progeny of the crosses.

427

428 We finally asked whether genes showing ASE were more likely to be in a genomic region
429 that has undergone a recent selective sweep. Out of 115 genes showing ASE in at least 4/6
430 crosses, 13 were in a swept region, whereas 103 of the 1333 genes showing no evidence of
431 ASE in any of the 6 crosses were in swept regions. A two sided Fisher's exact test did not
432 reject the hypothesis that the odds ratio is 1 (odds ratio 1.5, P=0.2), suggesting that genes
433 showing ASE are no more likely to be in a swept region than genes that do not.

434

435 CRM prediction

436 *Drosophila* CRM for adult midgut, adult Malpighian tubules, larval midgut and legs were
437 used as training data from SCRMshawHD to predict *Anopheles* CRM operating in the same
438 tissues, together with previously developed training sets for the adult peripheral nervous
439 system, embryonic and larval excretory system and embryonic/ larval Malpighian tubules
440 (https://github.com/HalfonLab/dmel_training_sets). These tissues were selected based on
441 previous studies examining the tissue specific expression of genes involved in insecticide
442 resistance with roles including detoxification and cuticular resistance (64-69). Training CRM

443 sets used for the first time in this study are shown in table 6, with full sequences at
444 https://github.com/azurillandfriend/traning_sets_IR.git. Oenocyte (FBbt:00004995), cuticle
445 (FBbt:00004970) and adult epidermis (FBbt:00005401) CRM could not be used as training
446 data due to insufficient known CRM, highlighting the need for more research to identify
447 CRM in these tissues. The top scoring 250 predictions for each training set and method were
448 combined, producing a total of 4122 unique CRM predictions. 62 of these predicted CRMs
449 were flanked by a gene showing ASE in at least 4/6 crosses (Table 7). In total, CRM were
450 predicted for 33 of the 115 genes showing ASE in at least 4/6 crosses. 211 predicted CRM
451 were flanked by a gene showing consistently high median fold change between resistant
452 and susceptible strains (Supplementary table 8). CRM were predicted for a total of 141 of
453 the 429 genes in this set. The predicted CRM were also compared with a published dataset
454 of Tn5 transposase sensitive sites in the adult midgut (58, 70). 50 of the predicted CRMs
455 overlapped with Tn5 transposase sensitive sites identified as *cis*-regulatory elements by Ruiz
456 *et al* (58), flanking a total of 48 genes. The vast majority of these CRMs (47) were previously
457 predicted by Kazemian (71) despite the different training sets. The CRMs identified in these
458 predictions provide a starting point for future studies to examine genetic variation in
459 *Anopheles* populations with different insecticide resistance phenotypes, and for validation
460 using reporter assays.
461

462 # Conclusions

463 Sample pooling, while a cost-effective solution for bulk RNAseq of small samples, does limit
464 the power of experiments targeting the study of ASE; in future, to maximize the power to
465 detect ASE, individual samples should be used. RNAseq on whole mosquitoes may have

466 masked tissue specific ASE. This is likely to increase the number of genes showing ASE, since
467 pleiotropic effects may limit the potential for genes to up or down regulated in all tissues
468 simultaneously but permit the evolution of tissue specific regulation (72). Future studies
469 should target specific tissues of interest. Despite these limitations, we were able to detect
470 genes showing ASE in pooled RNA from whole mosquitoes in crosses between *Anopheles*
471 strains with different carbamate resistance phenotypes, indicating different *cis*-regulation
472 patterns between the strains. Whilst we detected some genes previously implicated in
473 insecticide resistance, there was no consistent enrichment of these or any particular gene
474 ontology term amongst genes showing ASE. This probably indicates that the strains have
475 undergone extensive *cis*-regulatory divergence, affecting both genes involved in insecticide
476 resistance but also genes involved in many other functions. Comparing ASE in progeny of
477 crosses between multiple insecticide resistance and susceptible strains, together with
478 examining gene expression in the parental strains, would enable a more comprehensive
479 survey of the *cis* versus *trans* regulation of insecticide resistance genes. The bias towards
480 younger, *Anopheles* specific genes showing ASE in the F1 suggests there may be a higher
481 degree of *cis*-regulatory divergence between the parental strains for younger genes.
482 It was possible to computationally predict some CRM involved in tissue specific expression,
483 including potential CRM for genes showing ASE and genes previously implicated in
484 insecticide resistance. This was hampered by the lack of good quality training data for the
485 tissues thought to be most relevant to insecticide resistance in adult mosquitoes,
486 highlighting the need for future experimental CRM discovery.

487 **Data availability statement**

488 1. RNAseq data for all crosses and read counts per SNP are available at Gene Expression
489 Omnibus accession GSE241768.
490 2. Potential fathers sequencing data are available at European nucleotide archive
491 (accession numbers in Supplementary Table 9).
492 3. Sequencing data from siblings from crosses B1 and B3 are at European nucleotide
493 (accession numbers in supplementary Table 10)
494 4. All other nucleotide sequences used are already publicly available in the Phase 3
495 release of the *Anopheles gambiae* 1000 genomes project, at accession PRJEB42254.

496 Acknowledgements

497 This project was funded by Daphne Jackson Fellowship to N.A.D. sponsored by the
498 Biotechnology and Biosciences Research Council, with RNA sequencing funded by Infravec
499 (www.infravec.eu). Additional support to this work was provided by the National Institute of
500 Allergy and Infectious Diseases (NIAID R01-AI116811) and the Medical Research Council
501 (MR/T001070/1). M.J.D. is supported by a Royal Society Wolfson Fellowship (RSWF\FT
502 \180003). We thank Luciene Salas Jennings and Andrew Carey for providing administrative
503 support to the project. The authors are grateful for discussions with Louise Cerdeira, Laura
504 Brettell and Shannon Quek.

505

506 References

507

1. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207-+.
2. World Health Organisation. Malaria Report 2019. 1019.
3. Ranson H, Lissenden N. Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. Trends in parasitology. 2016;32(3):187-96.
4. Lynd A, Gonahasa S, Staedke SG, Oruni A, Maiteki-Sebuguzi C, Dorsey G, et al. LLIN Evaluation in Uganda Project (LLINEUP): a cross-sectional survey of species diversity and insecticide resistance in 48 districts of Uganda. Parasites & vectors. 2019;12(1):94.
5. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. The genome sequence of the malaria mosquito Anopheles gambiae. Science. 2002;298(5591):129-49.
6. The Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector Anopheles gambiae. Nature. 2017;552(7683):96-100.
7. The Anopheles gambiae 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. Genome research. 2020;30(10):1533-46.
8. MalariaGen. Anopheles gambiae genomic surveillance project 2023 [updated 17/11/2023. Available from: https://www.malariagen.net/anopheles-gambiae-genomic-surveillance-project.
9. Liu N. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. Annu Rev Entom. 2015;60:537-59.
10. Mitchell SN, Stevenson BJ, Muller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae from Ghana. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(16):6147-52.
11. Edi CV, Djogbenou L, Jenkins AM, Regna K, Muskavitch MA, Poupardin R, et al. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito Anopheles gambiae. PLoS genetics. 2014;10(3):e1004236.
12. Muller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-caught permethrin-resistant Anopheles gambiae overexpress CYP6P3, a P450 that metabolises pyrethroids. PLoS genetics. 2008;4(11):e1000286.
13. Ingham VA, Pignatelli P, Moore JD, Wagstaff S, Ranson H. The transcription factor Maf-S regulates metabolic resistance to insecticides in the malaria vector Anopheles gambiae. BMC genomics. 2017;18(1):669.
14. Wilding CS. Regulating resistance: CncC:Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance. Current opinion in insect science. 2018;27:89-96.
15. Lucas ER, Nagi SC, Egyir-Yawson A, Essandoh J, Dadzie S, Chabi J, et al. Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae s.l. bioRxiv. 2023:2023.01.13.523889.
16. Weetman D, Wilding CS, Neafsey DE, Muller P, Ochomo E, Isaacs AT, et al. Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African Anopheles gambiae. Scientific reports. 2018;8(1):2920.

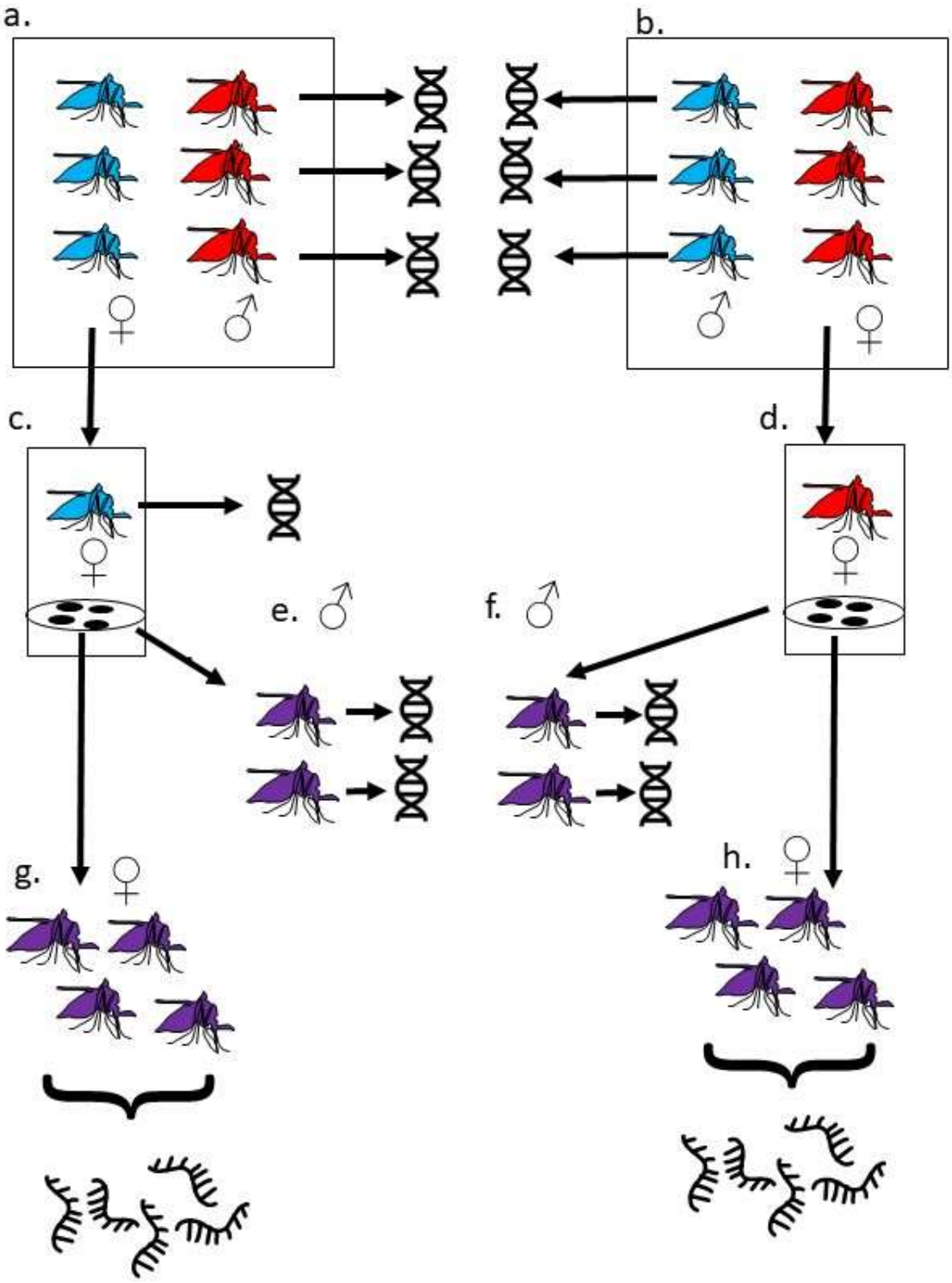
554 17. Ibrahim SS, Riveron JM, Bibby J, Irving H, Yunta C, Paine MJ, et al. Allelic Variation of
555 Cytochrome P450s Drives Resistance to Bednet Insecticides in a Major Malaria Vector. PLoS
556 genetics. 2015;11(10):e1005618.
557 18. Main BJ, Everitt A, Cornel AJ, Hormozdiari F, Lanzaro GC. Genetic variation associated
558 with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. Parasites
559 & vectors. 2018;11(1):225.
560 19. Ibrahim SS, Muhammad A, Hearn J, Weedall GD, Nagi SC, Mukhtar MM, et al.
561 Molecular drivers of insecticide resistance in the Sahelo-Sudanian populations of a major
562 malaria vector. bioRxiv. 2022:2022.03.21.485146.
563 20. Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, et al.
564 Whole-genome sequencing reveals high complexity of copy number variation at insecticide
565 resistance loci in malaria mosquitoes. Genome research. 2019;29(8):1250-61.
566 21. Lucas ER, Nagi SC, Egyir-Yawson A, Essandoh J, Dadzie S, Chabi J, et al. Genome-wide
567 association studies reveal novel loci associated with pyrethroid and organophosphate
568 resistance in Anopheles gambiae and Anopheles coluzzii. Nature communications.
569 2023;14(1):4946.
570 22. World Health Organisation. Malaria Report 2022. Geneva; 2022.
571 23. Katureebe A, Zinszer K, Arinaitwe E, Rek J, Kakande E, Charland K, et al. Measures of
572 Malaria Burden after Long-Lasting Insecticidal Net Distribution and Indoor Residual Spraying
573 at Three Sites in Uganda: A Prospective Observational Study. PLoS medicine.
574 2016;13(11):e1002167.
575 24. Isaacs AT, Mawejje HD, Tomlinson S, Rigden DJ, Donnelly MJ. Genome-wide
576 transcriptional analyses in Anopheles mosquitoes reveal an unexpected association
577 between salivary gland gene expression and insecticide resistance. BMC genomics.
578 2018;19(1):225.
579 25. Scott JA, Brogdon WG, Collins FH. Identification of single specimens of the Anopheles
580 gambiae complex by the polymerase chain reaction. The American journal of tropical
581 medicine and hygiene. 1993;49(4):520-9.
582 26. Bass C, Nikou D, Vontas J, Williamson MS, Field LM. Development of high-throughput
583 real-time PCR assays for the identification of insensitive acetylcholinesterase (ace-1R) in
584 Anopheles gambiae. Pesticide biochemistry and physiology. 2010;96(2):80-5.
585 27. World Health Organisation. Standard operating procedure for determining the ability
586 of PBO to restore susceptibility of adult mosquitoes to pyrethroid insecticides in WHO tube
587 tests2022. Available from: https://www.who.int/publications/i/item/9789240043855.
588 28. World Health Organisation. Manual for monitoring insecticide resistance in mosquito
589 vectors and selecting appropriate interventions2022. Available from:
590 https://www.who.int/publications/i/item/9789240051089.
591 29. illumina. TruSeq® RNA Sample Preparation v2 Guide2012. Available from:
592 https://support.illumina.com/content/dam/illumina-
593 support/documents/documentation/chemistry_documentation/samplepreps_truseq/truse
594 qrna/truseq-rna-sample-prep-v2-guide-15026495-f.pdf.
595 30. Miles A, bot pi, Murillo R, Ralph P, Harding N, Pisupati R, et al. scikit-allel: v1.3.3
596 (v1.3.3). 2021.
597 31. Nagi SC, Oruni A, Weetman D, Donnelly MJ. RNA-Seq-Pop: Exploiting the sequence in
598 RNA sequencing-A Snakemake workflow reveals patterns of insecticide resistance in the
599 malaria vector Anopheles gambiae. Molecular ecology resources. 2023;23(4):946-61.

32. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010.
33. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047-8.
34. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology. 2019;37(8):907-15.
35. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nature biotechnology. 2016;34(5):525-7.
36. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.
37. Garrison E, G. M. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907 [q-bioGN]. 2012.
38. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.
39. Mendelevich A, Vinogradova S, Gupta S, Mironov AA, Sunyaev SR, Gimelbrant AA. Replicate sequencing libraries are important for quantification of allelic imbalance. Nature communications. 2021;12(1):3370.
40. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome biology. 2015;16:195.
41. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal. 2017;9(2):378-400.
42. Mayba O, Gilbert HN, Liu J, Haverty PM, Jhunjhunwala S, Jiang Z, et al. MBASED: allele-specific expression detection in cancer tissues and cell lines. Genome biology. 2014;15(8):405.
43. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289-300.
44. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33(18):2938-40.
45. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.
46. Nagi SC. AnoExpress: v0.1.7: Zenodo; 2023 [Available from: sanjaynagi/AnoExpress: v0.1.7 (v0.1.7). Zenodo.
47. Heinicke S, Livstone MS, Lu C, Oughtred R, Kang F, Angiuoli SV, et al. The Princeton Protein Orthology Database (P-POD): a comparative genomics analysis tool for biologists. PloS one. 2007;2(8):e766.
48. Capra JA, Williams AG, Pollard KS. ProteinHistorian: tools for the comparative analysis of eukaryote protein origin. PLoS computational biology. 2012;8(6):e1002567.
49. Garud NR, Messer PW, Buzbas EO, Petrov DA. Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. PLoS genetics. 2015;11(2):e1005004.
50. Asma H, Halfon MS. Computational enhancer prediction: evaluation and improvements. BMC bioinformatics. 2019;20(1):174.

646 51. Kazemian M, Halfon MS. CRM Discovery Beyond Model Insects. Methods Mol Biol.
647 2019;1858:117-39.
648 52. Kazemian M, Zhu Q, Halfon MS, Sinha S. Improved accuracy of supervised CRM
649 discovery with interpolated Markov models and cross-species comparison. Nucleic acids
650 research. 2011;39(22):9463-72.
651 53. Kantorovitz MR, Kazemian M, Kinston S, Miranda-Saavedra D, Zhu Q, Robinson GE, et
652 al. Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse.
653 Developmental cell. 2009;17(4):568-79.
654 54. Rivera J, Keranen SVE, Gallo SM, Halfon MS. REDfly: the transcriptional regulatory
655 element database for Drosophila. Nucleic acids research. 2019;47(D1):D828-D34.
656 55. Keranen SVE, Villahoz-Baleta A, Bruno AE, Halfon MS. REDfly: An Integrated
657 Knowledgebase for Insect Regulatory Genomics. Insects. 2022;13(7).
658 56. WJ K, CW S, TS F, KM R, TH P, AM Z, et al. The human genome browser at UCSC.
659 Genome research. 2002;12(6):996-1006.
660 57. Smit A, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015.
661 58. Ruiz JL, Ranford-Cartwright LC, Gomez-Diaz E. The regulatory genome of the malaria
662 vector Anopheles gambiae: integrating chromatin accessibility and gene expression. NAR
663 Genom Bioinform. 2021;3(1):lqaa113.
664 59. Fontanillas P, Landry CR, Wittkopp PJ, Russ C, Gruber JD, Nusbaum C, et al. Key
665 considerations for measuring allelic expression on a genomic scale using high-throughput
666 sequencing. Molecular ecology. 2010;19 Suppl 1:212-27.
667 60. Turissini DA, Gamez S, White BJ. Genome-wide patterns of polymorphism in an
668 inbred line of the African malaria mosquito Anopheles gambiae. Genome biology and
669 evolution. 2014;6(11):3094-104.
670 61. Frochaux MV, Bou Sleiman M, Gardeux V, Dainese R, Hollis B, Litovchenko M, et al.
671 cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila
672 genetic reference panel. Genome biology. 2020;21(1):6.
673 62. Floc'hlay S, Wong E, Zhao B, Viales RR, Thomas-Chollier M, Thieffry D, et al. Cis-acting
674 variation is common across regulatory layers but is often buffered during embryonic
675 development. Genome research. 2020.
676 63. Ingham VA, Wagstaff S, Ranson H. Transcriptomic meta-signatures identified in
677 Anopheles gambiae populations reveal previously undetected insecticide resistance
678 mechanisms. Nature communications. 2018;9(1):5282.
679 64. Ingham VA, Jones CM, Pignatelli P, Balabanidou V, Vontas J, Wagstaff SC, et al.
680 Dissecting the organ specificity of insecticide resistance candidate genes in Anopheles
681 gambiae: known and novel candidate genes. BMC genomics. 2014;15:1018.
682 65. Kefi M, Charamis J, Balabanidou V, Ioannidis P, Ranson H, Ingham VA, et al.
683 Transcriptomic analysis of resistance and short-term induction response to pyrethroids, in
684 Anopheles coluzzii legs. BMC genomics. 2021;22(1):891.
685 66. Pignatelli P, Ingham VA, Balabanidou V, Vontas J, Lycett G, Ranson H. The Anopheles
686 gambiae ATP-binding cassette transporter family: phylogenetic analysis and tissue
687 localization provide clues on function and role in insecticide resistance. Insect molecular
688 biology. 2018;27(1):110-22.
689 67. Mohanty AK, Dey G, Kumar M, Sreenivasamurthy SK, Garg S, Prasad TSK, et al.
690 Proteome data of female Anopheles stephensi antennae. Data in brief. 2019;24:103911.
691 68. Balabanidou V, Kampouraki A, MacLean M, Blomquist GJ, Tittiger C, Juarez MP, et al.
692 Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon

693 production in Anopheles gambiae. Proceedings of the National Academy of Sciences of the
694 United States of America. 2016;113(33):9268-73.
695 69. Grigoraki L, Grau-Bove X, Carrington Yates H, Lycett GJ, Ranson H. Isolation and
696 transcriptomic analysis of Anopheles gambiae oenocytes enables the delineation of
697 hydrocarbon biosynthesis. eLife. 2020;9.
698 70. Ruiz JL, Yerbanga RS, Lefevre T, Ouedraogo JB, Corces VG, Gomez-Diaz E. Chromatin
699 changes in Anopheles gambiae induced by Plasmodium falciparum infection. Epigenetics &
700 chromatin. 2019;12(1):5.
701 71. Kazemian M, Suryamohan K, Chen JY, Zhang Y, Samee MA, Halfon MS, et al. Evidence
702 for deep regulatory similarities in early developmental programs across highly diverged
703 insects. Genome Biol and Evol. 2014;6(9):2301-20.
704 72. Quiver MH, Lachance J. Adaptive eQTLs reveal the evolutionary impacts of pleiotropy
705 and tissue-specificity while contributing to health and disease. HGG Adv. 2022;3(1):100083.
706

707 **Figure 1: Crossing, DNA and RNA extraction schema**



708

709

710

711 Figure 1 legend:

712 a. 13 Kisumu females (blue) were crossed to 13 Nagongera males (red). Females mate
713 only once. Following mating, genomic DNA from Nagongera males was extracted and
714 sequenced.

715 b. In the reciprocal cross, 13 Nagongera females (red) were crossed to 13 Kisumu males
716 (blue). Following mating, genomic DNA was extracted and sequenced from Kisumu males.

717 c. Individual mated Kisumu females were transferred to laying cups. Following egg
718 laying, genomic DNA was extracted and sequenced from three of these

719 d. Individual mated Nagongera females were transferred to laying cups. Following egg
720 laying, genomic DNA was extracted and sequenced from three of these.

721 e. The F1 progeny (purple) from the three Kisumu mothers sequenced at step c. were
722 raised to adulthood. Genomic DNA was extracted and sequenced from individual males

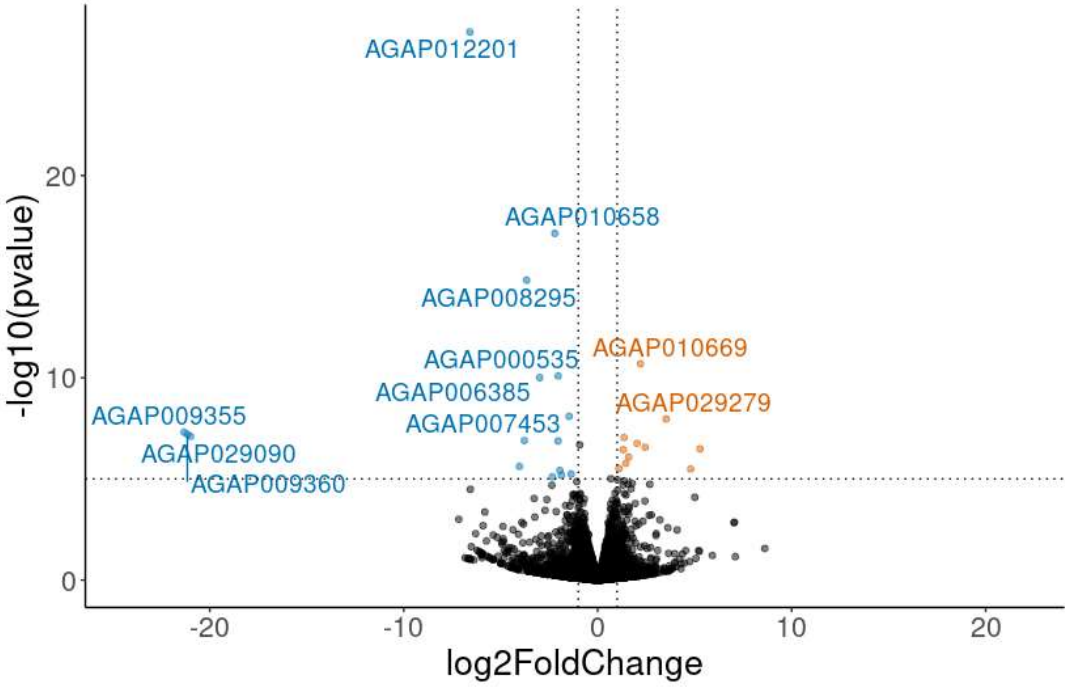
723 f. The F1 progeny (purple) from the three Nagongera mothers sequenced at step d.
724 were raised to adulthood. Genomic DNA was extracted and sequenced from
725 individual males

726 g. and h. Female F1 from each of the six mothers were raised to adulthood. RNA was
727 extracted and sequenced from a pool of ten F1 females three to five days after eclosion.

728

729 **Figure 2: Reciprocal crosses show similar overall gene expression**

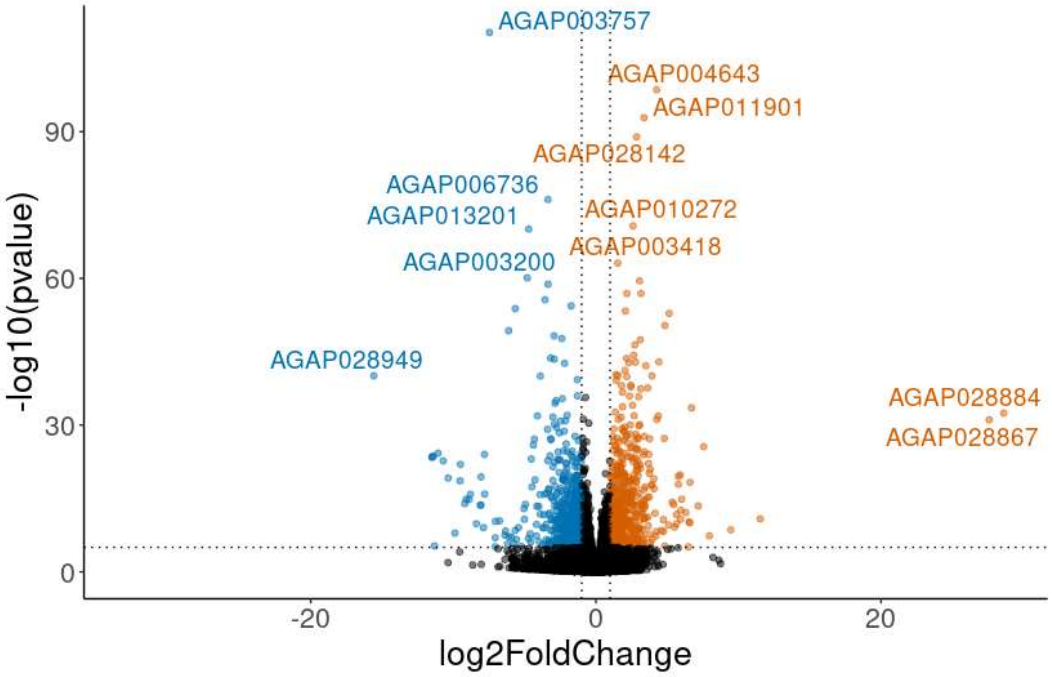
730 a



731

732

733 b.



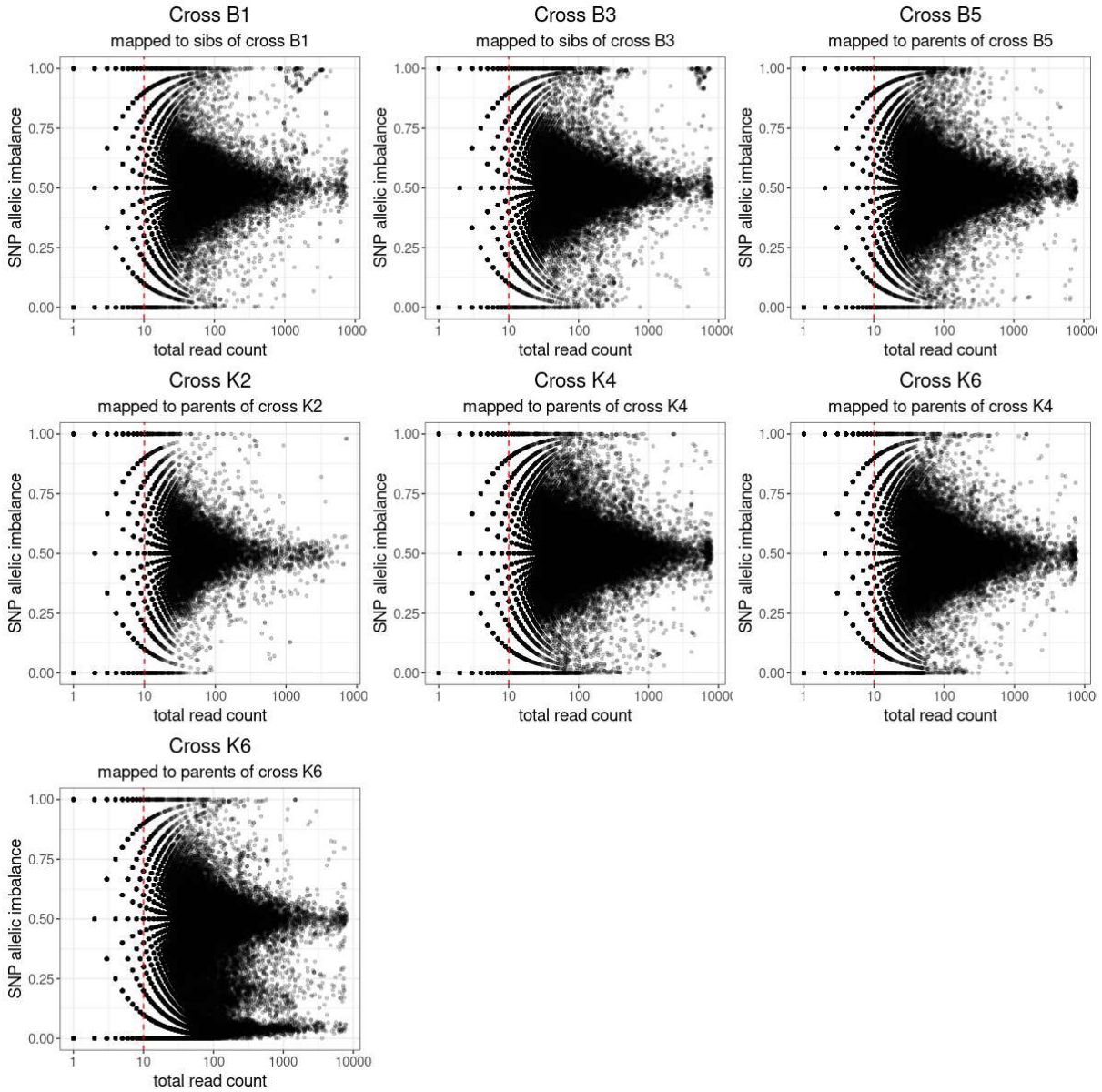
734

735

Figure 2 Legend

a. Volcano plot of log2 fold change against -log10 P value comparing gene expression in the F1 progeny of reciprocal crosses between Kisumu and Nagongera strains. Blue points indicate genes downregulated in progeny of Nagongera mothers compared to progeny of Kisumu mothers, orange points indicate genes upregulated in progeny of Nagongera mothers compared to progeny of Kisumu mothers.
b. Volcano plot of log2 fold change against -log10 P value comparing gene expression between F1 progeny of Nagongera and Kisumu with the Kisumu parental strain. Blue points indicate genes downregulated in the cross progeny compared to Kisumu, orange points indicate genes upregulated in cross progeny of compared to Kisumu.

749 **Figure 3: Allele balance in progeny of crosses between strains**



750

751 Figure 3 legend:

752 Plot of total read count against allelic imbalance at each SNP. Cross and source for the SNPs

753 used to map and count reads are indicated at the top of each plot. Cross K6 is shown with

754 allelic imbalance inferred using SNPs from both the parents of cross K4 and for the initially

755 assumed parents of cross K6.

Figure 4: Intersection of genes showing ASE between crosses

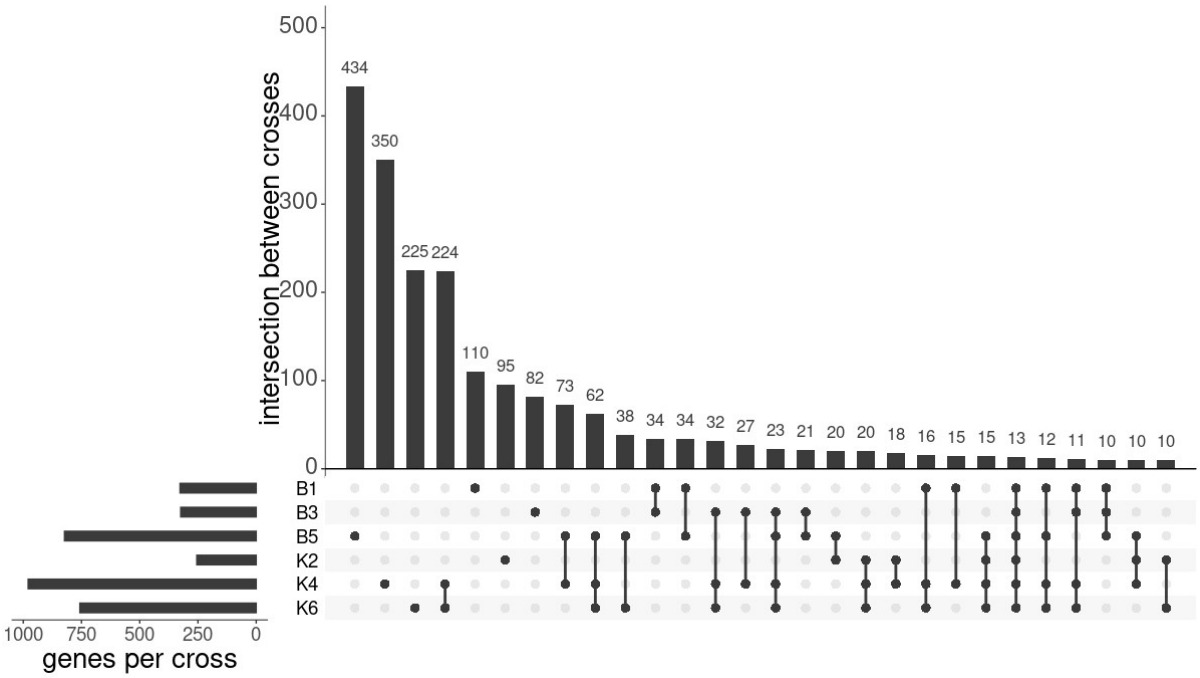
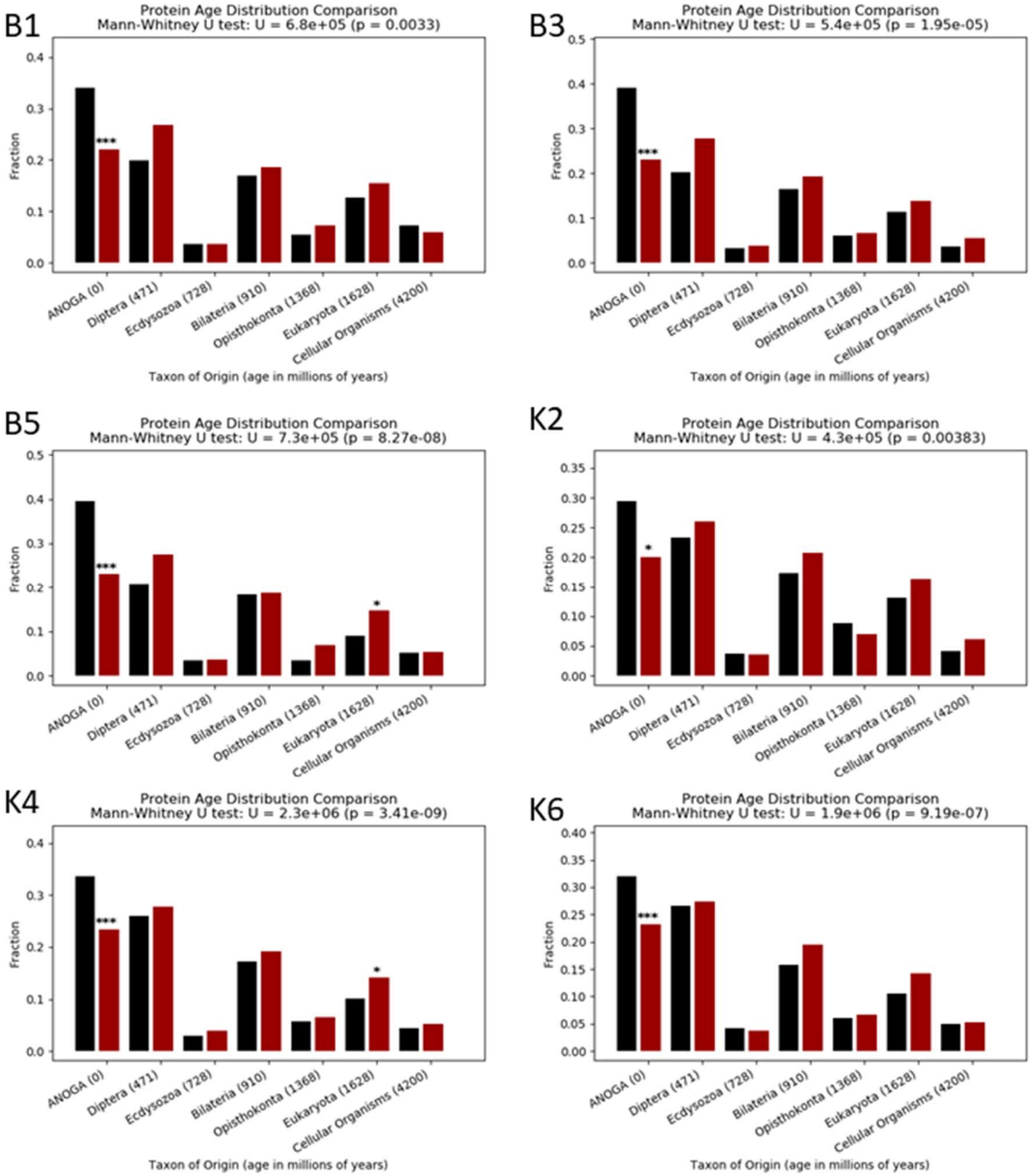


Figure 4 legend

UpSet plot with the number of genes showing ASE for each cross and the intersection of these genes between crosses. Only the first 28 sets of overlaps are shown.

762 **Figure 5: Ages of genes showing ASE or not showing ASE**



763

764 Figure 5 legend, bar plots comparing the age of genes which showed ASE or did not in the

765 progeny of six crosses between Nagongera and Kisumu strains, using the Wagner parsimony

766 method. Fisher's exact test P values for the difference in fraction of genes in each age

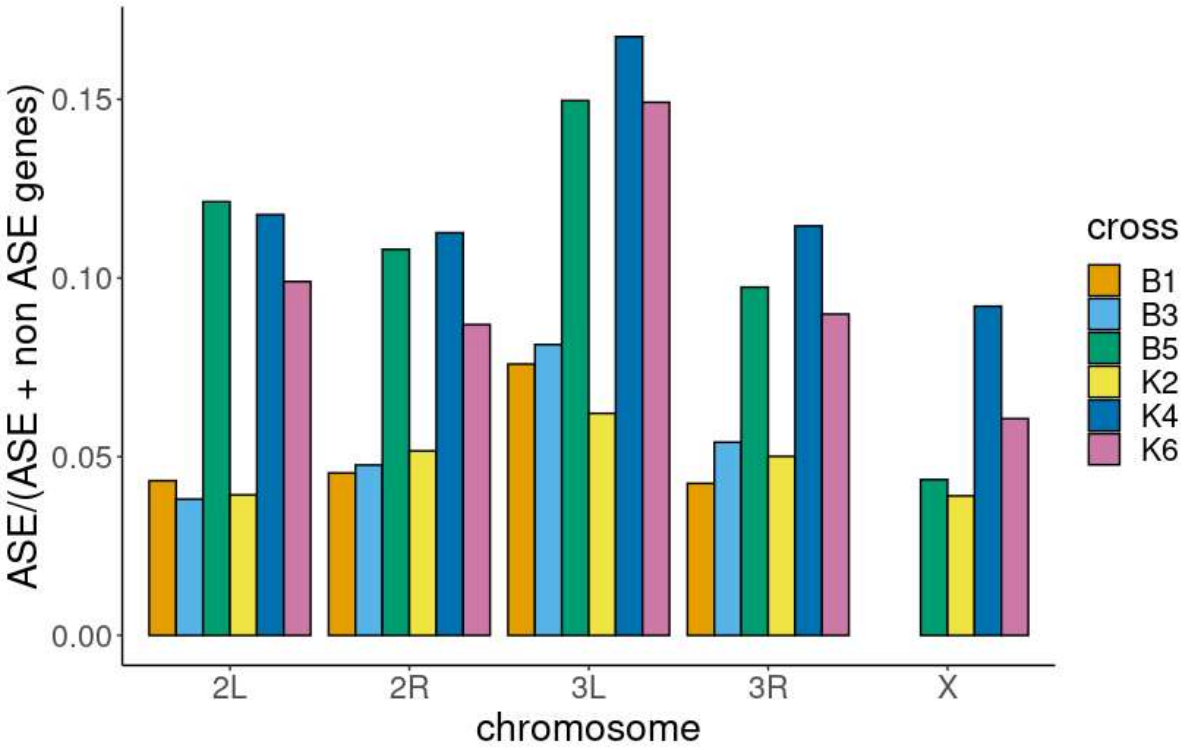
767 between genes showing ASE (red bars) and genes not showing ASE (black bars) are

768 displayed above the bars; *: P<0.05, ***:P<=0.001.

769

770 **Figure 6: proportion of genes showing ASE versus genes with**

771 **sufficient SNPs to detect ASE along each chromosome arm**



772

773

774 Figure 6 legend:

775 Bar plot of the proportion of genes showing allele specific expression/ total genes

776 containing suitable SNPs to detect ASE per chromosome arm. Colours indicate the different

777 crosses. X chromosomal SNPs were not available for crosses B1 and B3

778

779

780 **Table 1: Crosses between Kisumu and Nagongera and RNAseq summary**

781 **statistics**

RNAseq sample name	Cross name	Mother origin	Mother-Sanger code	Father origin	Father – Sanger code	Total reads counted (F1 RNAseq)	Percent of reads mapping to *Anopheles gambiae* PEST genome
Wilding_1	B1	Nagongera	failed QC	Kisumu	unknown	18,496,263	88.6%
Wilding_2	B3	Nagongera	failed QC	Kisumu	unknown	38,526,717	87.1%
Wilding_3	B5	Nagongera	AC0382-C	Kisumu	AC0416-C	28,969,694	87.7%
Wilding_4	K2	Kisumu	AC0300-C	Nagongera	AC0406-C	6,400,460	85.5%
Wilding_5	K4	Kisumu	AC0317-C	Nagongera	AC0398-C	46,229,505	86.3%
Wilding_6	K6	Kisumu	AC0334-C	Nagongera	AC0398-C	29,491,773	86.9%

782

783 **Table 2: Number of SNPs in exonic regions differing between the parents that**

784 **could be used to detect ASE**

Cross	2L	2R	3L	3R	X	Y / other	total
K2	13849	30134	19375	20877	15543	0	99778
K4	33139	35048	19074	22405	15250	0	124916
K6	30303	30811	14534	17801	15601	2	109052
B5	29627	36203	20865	25414	10539	1	122649

785

786 2L, 2R, 3L, 3R and X are the chromosome arms. The numbers are the number of SNPs in

787 exonic regions that are homozygous in the parents of each cross but differ between the

788 parents. These sites should be heterozygous in all cross progeny.

789

790 **Table 3: Number of SNPs differing between Kisumu and Nagongera colony**

791 **consensus sequences**

Chromosome	SNPs
2L	2575
2R	13507
3L	6216
3R	2949
X	10552
Total	35799

792

793 "SNPs" shows the number of SNPs differing across all sites in the consensus genome

794 sequences of Kisumu and Nagongera colonies

795

796

797 **Table 4: Number of genes showing ASE in each cross**

Cross	Genes showing ASE	Total genes with SNP(s) and sufficient counts to detect ASE
B1	327	6486
B3	324	6065
B5	823	7537
K2	255	5185
K4	979	8166
K6	757	7852
All (intersection)	13	2468

798

799 ASE detected with FDR of 5%

800

801

802

803

804

805

806 **Table 5: Genes showing ASE in at least four crosses**

ID	chr	Gene Name	Description	overlap	het ASE	Freq CNV	Gene ASE in cross			
							B5	K2	K4	K6
AGAP001002	X		Toll protein	4	0	NA	0.00	1.00	0.96	0.98
AGAP001190	2R		female reproductive tract protease GLEANR_896	6	0	18	0.92	0.06	0.22	0.12
AGAP001243	2R			4	4	NA	0.38	0.03	0.28	0.28
AGAP001251	2R		eupolytin	5	0	3	0.24	NA	0.83	0.92
AGAP001356	2R	ACE1	Acetylcholinesterase	5	2	15	0.52	0.35	0.22	0.27
AGAP001365	2R			4	0	15	0.57	0.44	0.38	0.32
AGAP001594	2R			4	4	1	0.45	0.63	0.71	0.75
AGAP001726	2R			4	0	4	0.75	0.47	0.23	0.31
AGAP001971	2R		Polyubiquitin-C	4	5	11	0.55	0.59	0.64	0.69
AGAP002008	2R			4	0	NA	0.17	0.80	0.76	0.79
AGAP002058	2R		beta-galactosidase	4	0	3	0.14	0.83	0.82	0.93
AGAP002728	2R		Isocitrate dehydrogenase [NAD] subunit, mitochondrial	4	0	5	0.60	0.38	0.33	0.36
AGAP003364	2R			4	0	NA	0.61	0.44	0.33	0.36
AGAP003417	2R			4	0	NA	0.41	0.53	0.67	0.62
AGAP003444	2R			4	2	22	0.27	0.62	0.84	0.78
AGAP003474	2R			6	0	NA	0.49	0.63	0.61	0.63
AGAP003514	2R			4	3	NA	0.33	0.84	0.72	0.66
AGAP003525	2R		CCR4-NOT transcription complex, subunit 2	4	2	2	0.49	0.72	0.66	0.64
AGAP003583	2R		L-iditol 2-dehydrogenase	4	0	8	0.30	NA	0.32	0.06
AGAP003657	2R			4	0	NA	0.08	0.76	0.67	0.64
AGAP003663	2R		ATP-dependent RNA helicase DBP2	4	3	NA	0.43	0.62	0.60	0.58
AGAP003669	2R		homeobox protein MSX	4	2	NA	0.33	0.58	0.68	0.64
AGAP003686	2R			4	0	1	0.40	0.58	0.62	0.66
AGAP003701	2R		zinc finger protein 593 homolog	4	0	62	NA	NA	0.14	0.34
AGAP003722	2R	ANXB10C	annexin B10C	4	3	6	NA	NA	0.73	0.71
AGAP003728	2R		Uncharacterized methyltransferase WBSCR22	4	0	1	0.34	0.53	0.72	0.69
AGAP003731	2R		vacuolar protein sorting-associated protein 37B	4	4	1	0.49	0.57	0.74	0.77
AGAP003802	2R			4	3	NA	0.37	0.62	0.60	0.53
AGAP003861	2R		Peptide deformylase	4	0	1	0.34	0.61	0.73	0.73
AGAP004102	2R			4	0	1	0.36	0.49	0.65	0.66

AGAP004109	2R		Xaa-Pro aminopeptidase	6	6	NA	0.74	0.28	0.37	0.40
AGAP004110	2R			6	3	NA	0.79	0.23	0.23	0.25
AGAP004113	2R	mRpL55	39S ribosomal protein L55, mitochondrial	5	0	4	0.75	0.34	0.36	0.36
AGAP004133	2R			4	1	35	0.35	0.97	0.63	0.75
AGAP004233	2R		transcription factor IIIB 90 kDa subunit	6	1	NA	0.19	0.76	0.74	0.76
AGAP004534	2R		cathepsin B precursor	4	0	NA	0.87	NA	0.21	0.23
AGAP004562	2R		MFS transporter, PCFT/HCP family, solute carrier family 46 (folate transporter)	4	0	NA	0.86	0.49	0.14	0.27
AGAP004790	2L		Up-regulated during skeletal muscle growth 5 homolog	4	0	4	0.05	NA	0.79	0.79
AGAP004891	2L		metallophosphoesterase domain-containing protein 1	4	2	NA	NA	NA	0.70	0.65
AGAP005012	2L		galactokinase	4	0	30	0.92	NA	0.21	0.18
AGAP005243	2L			4	2	NA	0.37	0.67	0.47	0.44
AGAP005548	2L		alpha-crystallin B chain	4	1	1	0.67	0.96	0.75	0.82
AGAP006690	2L		BTB (POZ) domain containing 9	4	0	NA	0.68	NA	0.62	0.70
AGAP006756	2L			4	0	NA	0.57	NA	0.74	0.68
AGAP006775	2L		glucosyl/glucuronosyl transferases	4	0	NA	0.87	NA	0.80	0.86
AGAP007159	2L		alpha-crystallin B chain	4	3	4	0.27	NA	0.17	0.13
AGAP007402	2L			4	1	33	0.61	0.90	0.87	0.86
AGAP007407	2L	CTLMA4	C-type lectin (CTL) - mannose binding	4	0	2	0.47	0.07	0.27	0.26
AGAP007413	2L			4	1	NA	0.56	0.76	0.64	0.60
AGAP007484	2L		Sugar transporter ERD6-like 4	5	0	NA	0.86	0.58	0.17	0.30
AGAP008019	3R	CYP12F3	cytochrome P450	4	0	14	0.78	NA	0.12	0.09
AGAP008020	3R	CYP12F2	cytochrome P450	4	0	46	0.87	0.47	0.21	0.27
AGAP008122	3R		U6 snRNA-associated Sm-like protein LSm7	4	2	NA	0.61	NA	0.40	0.40
AGAP008218	3R	CYP6Z2	cytochrome P450	4	1	1	0.89	NA	0.66	0.47
AGAP008325	3R		U3 small nucleolar RNA-associated protein 23	4	0	NA	0.58	0.29	0.27	0.30
AGAP008331	3R		WD repeat-containing protein 59	4	0	56	0.70	0.77	0.03	0.03
AGAP008551	3R		mothers against decapentaplegic homolog 1	4	0	NA	0.55	0.53	0.90	0.80

AGAP008849	3R		D-3-phosphoglycerate dehydrogenase	4	5	NA	0.54	0.48	0.48	0.46
AGAP009251	3R			4	2	26	0.86	NA	0.20	0.13
AGAP009418	3R			4	2	63	0.17	0.91	0.48	0.77
AGAP009518	3R			4	1	NA	0.47	0.41	0.45	0.41
AGAP009527	3R			5	0	6	NA	0.38	0.37	0.34
AGAP009585	3R			5	0	41	0.23	NA	0.07	0.28
AGAP009782	3R		methylosome protein 50	4	3	10	0.28	0.55	1.00	1.00
AGAP009791	3R	Nep1	neprilysin, neutral endopeptidase 1	4	0	NA	0.32	0.66	0.69	0.63
AGAP010658	3L			5	0	1	0.30	NA	0.94	0.85
AGAP010814	3L	TEP6	thioester-containing protein 6	5	5	14	0.73	0.30	0.68	0.64
AGAP010873	3L		Cytosolic Fe-S cluster assembly factor NUBP2 homolog	5	3	NA	0.56	0.06	0.19	0.15
AGAP011065	3L			4	3	63	0.65	0.51	0.76	0.65
AGAP011091	3L			4	0	1	0.51	NA	0.70	0.72
AGAP011155	3L		dual specificity phosphatase	4	0	NA	0.77	0.27	0.32	0.32
AGAP011202	3L			4	0	14	0.48	0.58	0.60	0.60
AGAP011379	3L	GPRFZ1	frizzled receptor	4	4	NA	0.75	0.51	0.37	0.38
AGAP011499	3L			4	4	63	0.49	0.56	0.66	0.69
AGAP011616	3L			4	0	NA	0.40	0.55	0.69	0.78
AGAP011750	3L		alpha-N-acetylglucosaminidase	4	0	NA	0.70	0.36	0.28	0.35
AGAP011783	3L	CLIPA13	CLIP-domain serine protease	5	0	NA	0.38	0.91	0.73	0.74
AGAP011784	3L			4	0	NA	0.83	0.53	0.16	0.53
AGAP011792	3L	CLIPA7	CLIP-domain serine protease	4	0	NA	0.63	0.41	0.36	0.38
AGAP012008	3L			5	3	NA	0.83	0.23	0.78	0.86
AGAP012053	3L		alpha,alpha-trehalase	6	5	NA	0.68	0.33	0.42	0.38
AGAP012100	3L	RpS26	40S ribosomal protein S26	4	0	5	0.20	0.79	0.77	0.71
AGAP012118	3L		FGFR1 oncogene partner 2-like protein	5	4	NA	0.54	0.47	0.63	0.59
AGAP012126	3L		lysophospholipase II	5	4	NA	0.95	0.26	0.09	0.11
AGAP012130	3L	ATN	allatotropin	6	4	NA	0.89	0.03	0.05	0.07
AGAP012135	3L		26S proteasome regulatory subunit N8	5	3	8	0.35	0.55	0.65	0.64
AGAP012148	3L		Mitogen-activated protein kinase	6	6	NA	0.71	0.18	0.15	0.18
AGAP012152	3L			4	2	NA	0.72	0.29	0.24	0.27
AGAP012154	3L		solute carrier family 15 member 1	6	5	NA	0.75	0.18	0.29	0.25

AGAP012188	3L		ubiquinol-cytochrome c reductase subunit 7	5	3	4	0.68	0.23	0.30	0.42
AGAP012189	3L		abhydrolase domain-containing protein 12	6	3	NA	0.88	0.19	0.09	0.11
AGAP012217	3L		sterol O-acyltransferase	6	4	21	0.80	0.41	0.78	0.71
AGAP012219	3L	Med1	mediator of RNA polymerase II transcription subunit 1	5	4	1	0.41	0.40	0.32	0.30
AGAP012296	3L	CYP9J5	cytochrome P450	4	2	36	0.28	0.64	NA	NA
AGAP012305	3L		Muscle LIM protein at 84B	4	4	NA	0.37	0.66	NA	NA
AGAP012314	3L			4	4	2	0.23	NA	0.86	0.86
AGAP012984	2R		Protease m1 zinc metalloprotease	4	0	2	0.69	0.33	0.30	0.31
AGAP012996	2R		juvenile hormone-inducible protein	5	0	NA	0.38	0.63	0.59	0.60
AGAP013039	2R			4	0	NA	0.84	0.17	0.26	0.38
AGAP013188	2R	APN4	aminopeptidase 4	5	4	61	0.31	0.86	0.62	0.58
AGAP013255	2R	APN3	aminopeptidase N1	4	3	62	0.34	0.93	0.37	0.36
AGAP013359	2R			4	0	1	0.95	0.06	0.04	0.14
AGAP013393	2R	APN2	aminopeptidase N1	4	3	62	0.50	0.17	0.26	0.35
AGAP028064	3R			4	1	NA	0.25	NA	0.88	0.85
AGAP028157	3L			4	0	36	0.32	0.75	0.84	0.78
AGAP028209	3R			4	1	1	0.38	0.41	0.35	0.39
AGAP028542	3L			6	0	6	1.00	0.21	0.00	0.00
AGAP029122	3L			4	5	NA	0.40	0.49	0.68	0.62
AGAP029263	3L			4	4	4	0.76	0.17	0.25	0.29
AGAP029287	3L			5	4	NA	0.89	0.00	0.06	0.07
AGAP029362	2L			4	0	NA	0.19	0.71	0.80	0.63
AGAP029417	3R			4	1	37	0.91	0.06	0.12	0.12
AGAP029453	3R			5	4	NA	0.54	0.63	0.75	0.69
AGAP029526	2L			4	2	2	0.61	1.00	0.60	0.57
AGAP029597	3L			6	2	NA	0.96	0.15	0.03	0.03

Chrom: chromosome

ID: Vectorbase gene ID

Overlap: the number of crosses that showed significant allelic imbalance of expression of the gene, when using SNPs from the siblings to infer allelic expression.

Expression biased to the maternal allele is highlighted in red and expression biased to the paternal allele highlighted in blue.

Freq CNV is how many of the 63 parents and siblings with CNV calls (crosses K2, K4, K6 and B5) had a possible CNV covering that gene.

817 **Table 6: training CRM**

Tissue	Anatomy ontology term	Date of search	Total Number of CRM <2000bp in Redfly at time of search	Number of CRM meeting criteria (non overlapping)
Adult Midgut	FBbt:00003138	12/12/22	12	11
Larval Midgut	FBbt:00005825	12/12/22	16	15
Adult Malpighian Tubules	FBbt:00005725	12/12/22	10	9
Leg	FBbt:00004640	12/12/22	35	28
Antennae	FBbt:00004511	24/04/23	25	20
Oenocyte	FBbt:00004995	25/04/23	3	2

818

Table 7: CRM predictions for genes that showed ASE in at least 4/6 crosses

chrom	Flanking gene 1	Name flanking gene 1	flanking_gene_2	Name flanking gene 2	CRM length	models	score	CRM_start	CRM_end	training_sets
2L	AGAP005012		AGAP005013		500	imm	7.12	8246250	8246750	Adult Malpighian tubules
2L	AGAP005012		AGAP005013		500	imm	4.85	8288000	8288500	Adult midgut
2L	AGAP005244		AGAP005243		500	hexmcd	26.41	12797250	12797750	malpig.mapping1
2L	AGAP006690		AGAP006689		500	pac	0.39	36723500	36724000	Larval midgut
2L	AGAP006755	snRNP-U1-70K	AGAP006756		500	imm	17.54	37866500	37867000	emb-larv_excretory,malpig.mapping1
2R	AGAP001194		AGAP001190		500	imm	6.75	1190000	1190500	Adult Malpighian tubules
2R	AGAP001356	ACE1	AGAP001355		500	hexmcd	12.68	3467250	3467750	legs
2R	AGAP013039		AGAP001990		500	pac	0.35	13217500	13218000	Adult Malpighian tubules
2R	AGAP013039		AGAP001990		750	hexmcd, pac	11.86	13224000	13224750	Adult Malpighian tubules
2R	AGAP002007		AGAP002008		500	pac	0.36	13703500	13704000	Adult midgut
2R	AGAP002007		AGAP002008		500	pac	0.41	13710000	13710500	Adult pns
2R	AGAP002007		AGAP002008		500	pac	0.42	13711750	13712250	Adult pns
2R	AGAP002007		AGAP002008		500	hexmcd, imm, pac	16.10	13717000	13717500	Adult midgut, antenna
2R	AGAP002007		AGAP002008		500	pac	0.44	13721000	13721500	emb-larv excretory
2R	AGAP002007		AGAP002008		500	hexmcd, imm	14.42	13727000	13727500	legs
2R	AGAP002007		AGAP002008		500	hexmcd	13.10	13773750	13774250	Larval midgut
2R	AGAP002008		AGAP002007		500	pac	0.41	13804500	13805000	legs

2R	AGAP002008		AGAP002009		500	pac	0.40	13843500	13844000	legs
2R	AGAP003473		AGAP003474		500	hexmcd	10.65	38103750	38104250	Adult Malpighian tubules
2R	AGAP003663		AGAP013457		500	imm	5.17	41330500	41331000	Adult pns
2R	AGAP003669		AGAP003667		500	pac	0.39	41423000	41423500	Adult pns
2R	AGAP003669		AGAP003670		500	hexmcd	11.76	41493500	41494000	Adult pns
2R	AGAP003727		AGAP003728		500	hexmcd	12.09	42615500	42616000	Adult pns
2R	AGAP013081		AGAP004110		500	pac	0.33	50078500	50079000	Adult midgut
2R	AGAP004112		AGAP004113	mRpL55	500	pac	0.39	50141750	50142250	antenna
2R	AGAP004112		AGAP004113	mRpL55	500	pac	0.40	50151250	50151750	legs
2R	AGAP004112		AGAP004113	mRpL55	500	hexmcd	12.06	50160250	50160750	antenna
2R	AGAP004232		AGAP004233		500	pac	0.38	52891500	52892000	Larval midgut
2R	AGAP004232		AGAP004233		500	pac	0.34	52894750	52895250	Adult Malpighian tubules
2R	AGAP004232		AGAP004233		500	pac	0.34	52921250	52921750	Adult midgut
3L	AGAP010813	TEP18	AGAP010814	TEP6	500	hexmcd	13.38	11182250	11182750	Adult pns
3L	AGAP011064		AGAP011065		500	imm	5.20	16137250	16137750	Adult pns
3L	AGAP011379	GPRFZ1	AGAP011378		500	pac	0.36	22779500	22780000	Larval midgut
3L	AGAP011379	GPRFZ1	AGAP011378		500	pac	0.37	22796500	22797000	Larval midgut
3L	AGAP012008		AGAP012007		500	imm	10.24	36201750	36202250	Larval midgut
3L	AGAP012026		AGAP028157		500	imm	10.72	36626250	36626750	legs
3L	AGAP012026		AGAP028157		500	imm	10.90	36645500	36646000	antenna
3L	AGAP012026		AGAP028157		500	hexmcd	29.43	36652000	36652500	emb-larv excretory, legs, malpig.mapping1
3L	AGAP029616		AGAP012053		500	hexmcd	12.69	36968750	36969250	legs
3L	AGAP029616		AGAP012053		500	hexmcd, imm	29.45	37009250	37009750	emb-larv_excretory, legs, malpig.mapping1
3L	AGAP029122		AGAP012093		750	hexmcd, imm	35.98	37603750	37604500	emb-larv excretory, malpig.mapping1

3L	AGAP029122		AGAP012093		500	imm	11.19	37608750	37609250	antenna
3L	AGAP029612		AGAP012118		500	hexmcd	26.81	37824750	37825250	emb-larv_excretory ,malpig.mapping1
3L	AGAP029612		AGAP012118		500	imm	9.59	37830500	37831000	Larval midgut
3L	AGAP012313		AGAP012314		500	pac	0.45	40131000	40131500	malpig.mapping1
3L	AGAP012313		AGAP012314		500	hexmcd	11.57	40135250	40135750	antenna
3L	AGAP012313		AGAP012314		500	pac	0.44	40137750	40138250	malpig.mapping1
3L	AGAP012313		AGAP012314		500	pac	0.37	40140500	40141000	Larval midgut
3R	AGAP028064		AGAP028015		500	pac	0.38	333000	333500	Larval midgut
3R	AGAP008019	CYP12F3	AGAP008018	CYP12F4	500	imm	8.26	4322500	4323000	Adult midgut
3R	AGAP008551		AGAP008552	CYP4H27	500	imm	11.67	12539500	12540000	Adult midgut
3R	AGAP008553		AGAP029453		500	pac	0.40	12582000	12582500	Adult pns
3R	AGAP029453		AGAP008557		500	pac	0.37	12662750	12663250	Larval midgut
3R	AGAP029417		AGAP009416		500	pac	0.40	32426500	32427000	Adult pns
3R	AGAP009418		AGAP009417		500	pac	0.36	32603250	32603750	Adult midgut
3R	AGAP009518		AGAP029410		500	imm	17.77	35145750	35146250	emb-larv_excretory, malpig.mapping1
3R	AGAP009518		AGAP029410		500	pac	0.40	35148250	35148750	antenna
3R	AGAP009522		AGAP009527		500	hexmcd	20.04	35340750	35341250	Adult midgut
3R	AGAP009522		AGAP009527		500	hexmcd, imm	11.01	35348500	35349000	Larval midgut
3R	AGAP009527		AGAP009522		500	pac	0.37	35364000	35364500	Larval midgut
3R	AGAP009790	CPAP3-B	AGAP009791	Nep1	500	hexmcd	17.04	43310250	43310750	Adult midgut
3R	AGAP009790	CPAP3-B	AGAP009791	Nep1	500	imm	4.90	43317000	43317500	Adult midgut

The positions of the 62 predicted CRMs for genes that showed ASE in at least 4/6 crosses. The two closest genes flanking each CRM are shown, with the flanking gene that shoed ASE highlighted in green