Molecular drivers of insecticide resistance in the Sahelo-Sudanian populations of a major malaria vector

Sulaiman S. Ibrahim[1,2,3,*], Abdullahi Muhammad[1,4], Jack Hearn[1], Gareth D. Weedall[5], Sanjay C. Nagi[1], Muhammad M. Mukhtar[2], Amen N. Fadel[3], Leon J. Mugenzi[3], Edward I. Patterson[6], Helen Irving[1] and Charles S. Wondji[1,3]

[1] Vector Biology Department, Liverpool School of Tropical Medicine (LSTM), L3 5QA, UK

[2] Department of Biochemistry, Bayero University, PMB 3011, Kano, Nigeria

[3] LSTM Research Unit, Centre for Research in Infectious Diseases (CRID), P.O. Box 13591, Yaoundé, Cameroon

[4] Centre for Biotechnology Research, Bayero University, PMB 3011, Kano, Nigeria

[5] School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK

[6] Department of Biological Sciences, Brock University, St. Catharines, Ontario, L2S 3A1, Canada.

* Correspondence: suleiman.ibrahim@lstmed.ac.uk ; Tel.: +44-74-4044-3871; ORCID: 0000-0003-0744-802X

Authors email addresses

Sulaiman S. Ibrahim: suleiman.ibrahim@lstmed.ac.uk

Abdullahi Muhammad: abdullahi.muhammad@lstmed.ac.uk

Jack Hearn: jack.hearn@lstmed.ac.uk

Gareth D. Weedall: g.d.weedall@ljmu.ac.uk

Sanjay C. Nagi: sanjay.nagi@lstmed.ac.uk

Muhammad M. Mukhtar: muhammadmahemukhtar@gmail.com

Amen N. Fadel: amen.fadel@crid-cam.net

Leon J. Mugenzi: leon.mugenzi@crid-cam.net

Edward I. Patterson: ipatterson@brocku.ca

Helen Irving: helen.irving@lstmed.ac.uk

Charles S. Wondji: charles.wondji@lstmed.ac.uk

Abstract

Information on common markers of metabolic resistance in malaria vectors from countries/regions sharing similar eco-climatic characteristics can facilitate coordination of malaria control. Here, we characterized populations of the major malaria vector *Anopheles coluzzii* from Sahelian transect/region, spanning four sub-Saharan African countries: Nigeria, Niger, Chad and Cameroon. Genome-wide transcriptional analysis identified major genes previously implicated in pyrethroid and/or cross resistance to other insecticides, overexpressed across the Sahel, including CYP450s, glutathione S-transferases, carboxylesterases, and cuticular proteins. Several, well-known variants/markers of insecticide resistance were found in high frequencies - including in the voltage-gated sodium channel (V402L, I940T, L995F, I1527T and N1570Y), the *acetylcholinesterase*-1 gene (G280S) and the *CYP4J5*-L43F, which is fixed. High frequencies of the phenotypically important chromosomal inversions, 2La, 2Rb and 2Rc were observed (~80% for 2La and 2Rb). The 2La alternative arrangement is fixed across the Sahel. Low frequencies of these inversions (<10%) were observed in the fully insecticide susceptible laboratory colony of *An. coluzzii* (Ngoussou). Several of the most commonly overexpressed metabolic resistance genes sit in these three inversions. Two commonly overexpressed genes, *GSTe2* and *CYP6Z2* were functionally validated. Transgenic *Drosophila melanogaster* expressing *GSTe2* exhibited extremely high DDT and permethrin resistance (mortalities < 10% in 24h). Serial deletion of the 5′ intergenic region, to identify putative nucleotide(s) associated with *GSTe2* overexpression, revealed that simultaneous insertion of adenine nucleotide and a transition (T->C), between Fork-head box L1 and c-EST putative binding sites are responsible for the high overexpression of *GSTe2* in the resistant mosquitoes, across the Sahel. Transgenic flies expressing *CYP6Z2* exhibited

marginal resistance towards 3-phenoxybenzylalcohol (a primary product of pyrethroid hydrolysis by carboxylesterases) and a type II pyrethroid, α-cypermethrin. However, significantly higher mortalities were obtained in *CYP6Z2* transgenic flies compared with controls, on exposure to the neonicotinoid, clothianidin. This suggests a possible bioactivation of clothianidin into a toxic intermediate, which if true make it an ideal insecticide against populations of *An. coluzzii* overexpressing this P450. These findings will facilitate regional collaborations within the Sahel region, and refine implementation strategies through re-focusing interventions, improving evidence-based, cross-border policy towards local and regional malaria pre-elimination.

Keywords: *Anopheles coluzzii*, Sahel, insecticides, pyrethroid, DDT, metabolic, resistance, genes, inversions.

Background

Since the year 2000 the massive scale-up in vector control interventions and treatment with antimalarial drugs had cut malaria incidence by ~40 % across Africa (1). Bolstered by this progress the World Health Organization (WHO) has been pushing to eliminate malaria, as proposed in the Global Technical Strategy (GTS) 2016-2030 (2). Unfortunately, the GTS, an ambitious framework with targets to reduce global malaria burden by 90 % in 15 years was dealt an immediate blow by a rebound in malaria transmission, with increased cases between 2016 and 2019 (3, 4). This stark warning of the risk posed to control and elimination efforts was a reflection of the lack of progress in the primary regions of interest in sub-Saharan Africa, which constitute 96 % of the 627,000 malaria-related deaths in 2020 alone (4, 5). Indeed, as the WHO widens the net of malaria elimination (E-2025) its important to acknowledge that no meaningful progress will be made without progress in sub-Saharan Africa, where none of the six countries having the highest global burden of malaria (e.g. Nigeria alone contributing ~27 % of all cases) is on the path to elimination (6).

The recent escalation of insecticide resistance in the major malaria vectors (7-9) makes the development of molecular tools to anticipate emergence and predict spread of resistance across the African continent imperative, in order to achieve malaria control and elimination. For decades no simple molecular assays were available to track metabolic resistance and assess its impact on malaria control and transmission, until the recent discovery of two DNA markers in the *cis*-regulatory region of two cytochrome P450s, *CYP6P9a* (10) and *CYP6P9b* (11) in the major malaria vector *Anopheles funestus*; findings which allowed the design of simple PCR assays to detect and track metabolic resistance in the field. Unfortunately, these markers explain resistance only in southern Africa for *An. funestus*, since genetic basis of pyrethroid

106 resistance, and cross-resistance with other insecticides is complex in other African regions
107 with different fronts, driven by distinct genes in the presence of barriers to gene flow (10, 12,
108 13). No major metabolic resistance markers for P450s and/or GSTs (functionally validated and
109 characterised for epidemiological impact in the natural populations) exist for the major
110 malaria vectors of the *Anopheles gambiae* Complex, though it is the omnipresent vector species
111 with widespread presence across Africa (14). This is hindering early detection and tracking of
112 molecular drivers of resistance in this species, slowing down evidence-based control measures
113 and resistance management.

114 Major genomic regions associated with metabolic resistance to pyrethroids in *An. gambiae*
115 sensu lato include the *CYP6* P450 clusters on the 2R and 3R chromosomes, and a *GST* epsilon
116 (*GSTe*) cluster on chromosome 3R (15), with some key resistance-associated genes functionally
117 validated. These genes include *CYP6P3* shown to confer cross resistance to pyrethroids and
118 organophosphates (16, 17), *CYP6M2* conferring cross resistance to pyrethroids (18, 19) and
119 DDT (19), as well as *GSTe2* shown to confer resistance to DDT (20). For *GSTe2*, in addition to
120 the I114T marker (20), a recent study has found a novel mutation (*Gste2*-119V) associated with
121 resistance (21) using a high-throughput genotypic panel for markers. Functional validation
122 and fild data is required establish the empirical evidence of the role of this mutation in
123 resistance. What has been missing in the case of *Anopheles gambiae/coluzzii* is reliable molecular
124 markers of resistance for major metabolic gene families, e.g. P450s (with field and laboratory
125 validated data) to aid creation of DNA-based diagnostic assays which will allow (i) easy
126 tracking of resistance in the field (e.g. the case of 119F-*GSTe2* mutation in *An. funestus* (22),
127 and (ii) determination of the operational impact of the resistance markers in the field, as

128 recently done for 119F-*GSTe2* (23) and *CYP6P9a*_R/*CYP6P9b*_R markers (10, 11) in *An.*
129 *funestus*.

130 The *An. gambiae s.l.*, especially populations in the semi-arid steppe, exhibit high frequency of
131 paracentric chromosomal inversions [one of the most effective instruments for speciation and
132 local adaptations (Ayala et al., 2014, Dobzhansky, 1971, Kirkpatrick, 2010)], maintained in
133 spatially and temporally heterogenous environment, and which segregate along climatic
134 gradients of increasing aridity (24)]. The 2La inversions are associated with resistance to
135 desiccation in adults (Fouet et al., 2012, Gray et al., 2009) and thermal stress in larvae (Cassone
136 et al., 2011). It was also shown that inversion 2La assorts with insecticide resistance, e.g.,
137 dieldrin plus fipronil (Brooke et al., 2000), and is associated with thermotolerance and
138 permethrin resistance in the Sahelain *An. coluzzii* (Ibrahim et al., 2021).

139 To support malaria pre-elimination effort in sub-Saharan Africa we targeted the Sahelo-
140 Sudanian region, which represent northern-most limit of malaria endemicity in sub-Saharan
141 Africa, and where malaria is highly seasonal [offering excellent target for pre-elimination
142 effort through sustained seasonal vector control and seasonal malaria chemoprevention (25)].
143 Focusing on the predominant malaria vector, *An. coluzzii* from Sudan/Sahel transects of four
144 countries, Nigeria, Niger, Chad and Cameroon (7, 26-28) we identified the major metabolic
145 resistance genes mediating pyrethroid resistance and cross-resistance in this region,
146 established the genetic variants which explained the resistance. We also functionally validated
147 the roles of two major candidate genes in the resistance (*CYP6Z2* and *GSTe2*), as well as single
148 nucleotide polymorphism in the 5′ regulatory elements of *GSTe2*, responsible for its
149 overexpression in the resistant population.

150

151 # Materials and Methods

152 ## Study Site and Mosquito Sampling

153 Blood fed female *An. coluzzii* mosquitoes, resting indoor were collected at one locality each in
154 Sudan/Sahel (Additional Figure S1): Hadiyau (HAD: 12°21′38″N, 9°59′15″E), a Sudan/sub-
155 Sahel village in northern Nigeria; Takatsaba (TAK: 13°44'01.8"N 7°59'05.2"E), a Sahel village
156 in southern Niger; Simatou (SIMAT: 10°50'40.7"N 14°56'40.9"E), a Sudan/sub-Sahel village in
157 Maga Department, far north of Cameroon; and Massakory (CHAD: 12° 6′ N, 15° 02′ E), a Sahel
158 town in Chad Republic. Details of sampling approaches and resistance profiles of mosquitoes
159 collected from Nigeria, Niger, and Chad are available in previously published articles (7, 26,
160 28). As in the above countries the Simatou F_1 females were also highly pyrethroid resistant,
161 with a mortality of only 3.7% from WHO tube bioassays using 0.05% deltamethrin, and no
162 mortality at all with 0.75% permethrin (data not published).

163

164 ## Genome-wide transcriptional analysis for common insecticide resistance genes in
165 the Sahel/sub-Sahel regions

166 ### *RNA extraction, library preparation and sequencing*

167 The RNA was extracted using the Arcturus PicoPure RNA isolation Kit (Applied Biosystems,
168 CA, USA) from three pools of 8 F_1 *An. coluzzii* females (2-4 day old) alive after exposure to
169 deltamethrin (resistant, R), unexposed (control, C), and also from unexposed females of the
170 fully susceptible laboratory colony of *An. coluzzii*, Ngoussou (susceptible, S) (29). RNA
171 isolation was carried out following the manufacturer's protocol with *Dnase* I-treatment to

172 remove contaminating DNA. The quantity and quality of RNA was measured using a
173 NanoDrop spectrophotometer (ThermoFisher, MA, USA) and Bioanalyzer (Agilent, CA,
174 USA).

175 Library preparation, sequencing and data quality control were carried out by the Centre for
176 Genomic Research (CGR), University of Liverpool, UK. RNA samples were subjected to
177 poly(A) mRNA enrichment and libraries prepared from the poly(A) mRNA-enriched
178 materials (dual-indexed, strand-specific RNAseq libraries were prepared using the NEBNext
179 polyA selection and Ultra Directional RNA library preparation kits). Libraries were
180 sequenced on a single lane of an Illumina HiSeq 4000 (paired-end, 2x150 bp sequencing,
181 generating data from >280 M clusters per lane). Basecalling and de-multiplexing of indexed
182 reads were performed by CASAVA version 1.8.2 (Illumina). De-multiplexed fastq files were
183 trimmed to remove Illumina adapter sequences using Cutadapt version 1.2.1 (30). Option -O
184 3 was used, so that the 3' end of any reads which matched the adapter sequence for 3 bp or
185 more were trimmed. Reads were further trimmed to remove low quality bases
186 using Sickle version 1.200 (31) with a minimum window quality score of 20. Reads shorter
187 than 20 bp after trimming were removed. If both reads from a pair passed this filter, each was
188 included in either the R1 (forward reads) or R2 (reverse reads) file. If only one of a read pair
189 passed this filter, it is included in the R0 (unpaired) file. Statistics were generated using fastq-
190 stats from EAUtils (32). Summary of total number of reads for each sample and distribution
191 of trimmed read length for forward (R1) and reverse (R2) reads and reads unpaired after
192 trimming (R0) are provided in Additional Figure S2.

193

194 ***Data analysis and estimation of transcript abundance by tag counting and differential gene***
195 ***expression***

196 Paired data for each replicate per country was aligned to the *An. gambiae* reference transcript
197 ome AgamP4.10 downloaded from VectorBase (https://vectorbase.org/) in salmon (0.11.4),
198 using 'validate mappings', 'seqBias', 'gcBias' and 'rangeFactorizationBins 4' flags. Read
199 mapping results (pre-alignment and post-alignment descriptive statistics (flagstat output
200 files) showing sequencing depth and coverage are given in Additional Tables 1 and -2,
201 respectively. Salmon results were converted into a gene expression matrix using the
202 Bioconductor package 'tximport' for input to DESeq2 1.26.0 (33). Differential gene expression
203 was tested for the three possible combinations of Exposed (R, deltamethrin resistant),
204 Unexposed (C, control) and Susceptible replicates (S). For results interpretation $\log_2$-fold
205 change thresholds of 1 was imposed, with false discovery rate adjusted p-values of 0.05
206 applied to accept significance. Principal components analysis implemented in DESeq2 was
207 used to examine relationships between respective replicates and treatments. This was carried
208 out based on the 500 most variable genes, with data transformation (normalisation/scaling) as
209 implemented in VST (DESeq2). For visualisation of expression results, volcano plots were
210 created using the Enhanced Volcano package (34) using the top most overexpressed genes
211 from lists which were prepared with $\log_2$FC cut off of 1 and p value of 0.01. Heatmaps were
212 generated (35) using the list of the top 50 most overexpressed metabolic resistance genes with
213 $\log_2$FC cut off of 1 and p value of 0.01.

214

215

216 ***Gene Ontology Enrichment and Mapping/Functional annotations***

217 The enrichment analysis for GO terms was carried out using a topGO package (36) against
218 *An. gambiae* (AgamP4.10), with data reannotated using EggNOG v5.0
219 (http://eggnog5.embl.de/#/app/home). Gene lists used in topGO are from the $\log_2$FC = 0, p
220 <0.05 analysis (the default in DESeq2). For the data from each site (country) six set of results
221 were generated (R, C and S contrasts) for genes either up- or down-regulated in each contrast,
222 for molecular function (MF) and biological process (BP) ontologies. For analysis the GO terms
223 lists for MF were used as input into the Revigo (37) for interpretation and visualisation,
224 querying the Whole Uniprot database, and SimRel setting for semantic similarity
225 measurements.

226

227 **Quantitative PCR measurement of expression profiles of metabolic resistance**
228 **genes**

229 The level of expression of 12 resistance-associated genes was validated by qRT-PCR using the
230 primers provided in Additional Table 3. These include the *GSTe2* (AGAP009194), *GSTZ1*
231 (AGAP002898), *CYP6Z2* (AGAP008218), *CYP6Z3* (AGAP008217), *CYP4C27* (AGAP009246),
232 *CYP4G16* (AGAP001076), *CYP4G17* (AGAP000877), *CYP6P3* (AGAP002865), *CYP6M2*
233 (AGAP008212), *CYP9K1* (AGAP000818), a *UGT-B19* (AGAP007920) and *COEBE3C*
234 (AGAP005372). The qRT-PCR was carried out using three technical replicates each of cDNA
235 extracted from 1 µg of total RNA of three biological replicates each from the Resistant (R),
236 Control (C) and Ngoussou (S). Protocol followed was as established in previous studies (38),
237 with relative expression level and fold change (FC) of each target gene in R and C relative to

238 S calculated according to the $2^{-\Delta\Delta CT}$ method incorporating the PCR efficiency (39), after
239 normalization with the housekeeping genes ribosomal protein S7, *RPS7* (AGAP010592) and
240 glycerol-3-phosphate dehydrogenase, *GPDH* (AGAP007593). Significant differences were
241 calculated using ANOVA with Dunnett's post hoc test.

242

243 ***Detection of Signatures of Selective Sweep***

244 To detect signature of select sweeps in the major metabolic resistance genes of interest, a
245 RNAseq population genetics pipeline, the Snakemake workflow
246 (https://zenodo.org/record/6078337) was utilised. The workflow aligns RNA-Seq reads to the
247 reference genome, and calls genomic variants with *Freebayes*, at a user-provided level of
248 ploidy, in our case 16 (8 diploid pooled mosquitoes). F_{st} (40) per gene between population
249 pairs and Tajima's D per gene within population were estimated. This was performed against
250 all SNPs passing quality and missingness filters. Population branch statistic (PBS) scans may
251 also be performed with the Snakemake, conditional on the presence of three suitable
252 populations (41). It is also possible to run Hudson's F_{st} and PBS scans, taking the average for
253 each protein-coding gene, as opposed to in windows. The population genetics statistical
254 analyses were calculated in scikit-allel v1.2.1 (42).

255

256 ***Establishment of allele frequencies of variants in genes of interest***

257 After genome alignment, RNA-Seq-Pop utilises samtools (43) to query specific positions of
258 the genome, calculating raw allele frequencies at those sites with a custom R script.

259

260 ***Detection of chromosomal inversion polymorphisms and metabolic genes within its***
261 ***breakpoint***

262 A modified version of the Python 3 program, compkaryo (44) was used to karyoptype the
263 major, *An. coluzzii/gambiae* phenotypically important inversion polymorphisms in
264 chromosome 2, and calculate its frequencies, *in silico*, using the previously identified tag SNPs
265 significantly associated with inversions. This allows to predict with high confidence
266 genotypes of the six common polymorphic inversions on chromosome 2 in individually
267 sequenced genomes of the field *An. coluzzii,* as well as in the Ngoussou. Compkaryo uses the
268 Ag1000 database (The *Anopheles gambiae* 1000 Genomes Consortium 2017) (15) by leveraging
269 a subset of cytologically karyotyped specimens to develop a computational approach for
270 karyotyping applicable to whole genome sequence. Modifications in the Snakemake pipeline
271 allows for variable ploidy (useful in the case of replicates from our pooled RNAsequencing
272 samples) here.

273

274 **Functional validation of the commonly overexpressed resistance-associated genes**

275 ***Comparative analysis of coding sequences of major resistance genes***

276 To establish presence of allelic variants which could impact catalytic activities, full length
277 coding sequences (cDNAs) of the two resistance-associated genes, *GSTe2* and *CYP6Z2* were
278 amplified and sequenced from alive mosquitoes in the four Sahel countries, as well as from
279 Ngoussou. This was done using total RNA extracted from 5 individual pools of 8 F_1 *An.*
280 *coluzzii* females (2-4 day old) alive after exposure to deltamethrin (resistant, R for *CYP6Z2*) or
281 DDT (R for *GSTe2*). Protocol for RNA extraction was as described in previous section, above.

282 Amplification was done using Phusion HotStart II Taq Polymerase (ThermoFisher
283 SCIENTIFIC, MA, USA) and the Full primers listed in Additional Table 4. The PCR mix
284 comprised 5x Phusion HF Buffer (containing 1.5 mM $MgCl_2$), 85.7 μM deoxynucleotides
285 (dNTPs), 0.34 μM each of forward and reverse primers, 0.015 U of Phusion HotStart II DNA
286 Polymerase (Fermentas, MA, USA), 10.71 μL of ddH_20 and 1 μL cDNA. Thermocycling
287 conditions were 1 cycle at 95 °C for 5 min; followed by 35 cycles each of 94 °C for 20 s, 60 °C
288 for 30 s, 72 °C for 2 min (1 min for *GSTe2*); and finally, one cycle at 72 °C for 5 min. PCR
289 products were cleaned with a QIAquick® PCR Purification Kit (QIAGEN, Hilden, Germany)
290 and ligated into the pJET1.2/blunt cloning vector using the CloneJET PCR Cloning Kit
291 (ThermoFisher SCIENTIFIC, MA, USA). These were then cloned into *E. coli DH5α*, plasmids
292 miniprepped with the QIAprep® Spin Miniprep Kit (QIAGEN) and sequenced on both strands
293 using pJET1.2 primers.

294 Polymorphisms were detected through examination and manual editing of sequence traces
295 using BioEdit version 7.2.3.0 (45) and nucleotide differences in sequences aligned using CLC
296 Sequence Viewer 7.0 (http://www.clcbio.com/). Different haplotypes were compared by
297 constructing a phylogenetic maximum likelihood tree using MEGA X (46). Genetic parameters
298 of polymorphism including number of haplotypes (h) and its diversity (H_d), number of
299 polymorphic sites (S) and nucleotide diversity (π) were computed using DnaSP v6.12.03 (47).

300

301

302

303

304 ***Characterization of the 5′ regulatory regions of GSTe2 and CYP6Z2***

305 *Amplification, cloning and sequence characterisation of 5′ regulatory element*

306 To investigate presence of genetic variants in the regulatory elements, which could be
307 responsible for overexpression of *GSTe2*, 351 bp intergenic regions (spanning the 43 bp 3′ UTR
308 of *GSTe1*, 248 bp flanking sequence and 60 bp 5′UTR of *GSTe2*) preceding the start codon were
309 amplified from 10 each of DDT-alive and -dead females from the 4 Sahel countries, as well as
310 from the Ngoussou females (primers provided in Additional Table 4). For *CYP6Z2*, a 1078 bp
311 intergenic region was retrieved from the VectorBase and used for amplification of the putative
312 5′-regulatory elements. Primers spanning 38 bp 3′ UTR of *CYP6Z1*, a 937 bp flanking sequence
313 and 103 bp 5′UTR of *CYP6Z2*, preceding the start codon of *CYP6Z2* were used to amplify
314 fragments from 10 each of deltamethrin-alive and -dead females from Nigeria and Niger, as
315 well as from the Ngoussou females. Amplification was carried out using HotStart II
316 Polymerase (ThermoFisher SCIENTIFIC, MA, USA) with similar thermocycling conditions as
317 outlined above for coding region of *GSTe2* and *CYP6Z2*, respectively. Purification of PCR
318 amplicons, cloning into pJET1.2 vector, sequencing and polymorphism analysis were done as
319 outlined above.

320 The 351 bp 5′-UTR fragments of *GSTe2* and 1078 bp fragment of the *CYP6Z2* were analysed
321 with the Gene Promoter Miner (http://gpminer.mbc.nctu.edu.tw/) and MatInspector (48) to
322 identify putative promoter elements, predict transcription start (TSS) and potential
323 transcription factor binding sites.

324

325 *Cloning of GSTe2 and CYP6Z2 5' regulatory elements in PGL3-Basic vector and dual luciferase*
326 *reporter assay*

327 Following analysis of the above sequences the 351 bp intergenic fragments of *GSTe2* were
328 amplified from the most predominant sequences of DDT-alive, DDT-dead and Ngoussou.
329 Same was done for for *CYP6Z2* amplifying 1087 bp fragment from deltamethrin-alive and -
330 dead, and Ngoussou. Primers bearing *kpn*I and *BgL*II sites (Additional Table 4) allowed
331 incorporation into pGL3-Basic reporter vector containing luciferase gene from the firefly
332 *Photinus pyralis* (Promega, Wisconsin, USA). Amplification was carried out using Phusion
333 HotStart II Polymerase, with conditions as above, followed by purification of PCR amplicons,
334 and cloning into pJET1.2 vector. Positive colonies (sequencing primers for pGL3-Basic
335 provided in Additional Table 4) were miniprepped; the minipreps digested with the above
336 restriction enzymes, gel-purified and ligated upstream of luciferase gene in pGL3-Basic vector
337 already linearized with the same restriction enzymes. Positive colonies were cloned and
338 miniprepped and concentrations of the recombinant plasmids adjusted to 200 ng/µL before
339 co-transfection. Protocols for maintenance of the *An. gambiae* cell line 4a-3B (MRA-919,
340 https://www.beiresources.org/), co-transfection of 200 ng of either *GSTe2* or *CYP6Z2* promoter
341 constructs, LRIM promoter in pGL3-Basic vector (49), or promoter-less pGL3-Basic control,
342 with 1 ng/µL of internal control, sea pansy *Renilla reniformis* luciferase containing the
343 *Drosophila* Actin *5C* promoter in pRL-null (49), are provided in Additional File 1 (additional
344 text for Methods and Results). Protocol for cell lysis and measurement of the activities of the
345 firefly luciferase (Dual-Luciferase Reporter Assay kit, Promega), with normalisation using the
346 Renilla luciferase activity, were as outlined in the Promega Quick Protocol, and provided in
347 detail in Additional File 1.

348 Assay background was also measured using lysate from non-transfected control cells. Results
349 were compared using a two-tailed Chi-Square test of independence using GraphPad Prism
350 7.02 (GraphPad Inc., La Jolla, CA, USA).

351 *Generation of GSTe2 promoter deletion constructs and assays*

352 The intergenic region separating *GSTe1* from *GSTe2* was progressively delineated. Partial
353 fragments of the 5′ regions (encompassing the 3′-UTR of *GSTe1*, the *GSTe1*/*GSTe2* flanking
354 region and the 5′-UTR of the *GSTe2*) were created by sequential deletion using forward
355 primers (provided in Additional Table 4, with numbers in primer names referring to distances
356 from the AUG start codon of *GSTe2*). Reverse primers were those initially used for
357 amplification of the full 351 bp nucleotide fragments. These fragments include (i) 308
358 nucleotides fragment generated from deletion of the 43 bp 3′-UTR of the *GSTe1* (-308 primers)
359 obliterating cellular-Myb (c-Myb) transcriptional factor binding site; (ii) 291 nucleotide
360 fragments generated from deletion of *GSTe1* 3′-UTR plus 38 nucleotides of the flanking region
361 (-270 primers), which removed in addition δ elongation factor 1 (δEF1) binding site; (iii) 262
362 nucleotide fragments produced from deletion of *GSTe1* 3′-UTR plus 46 nucleotides of the
363 flanking region (-262 primers), which obliterated in addition Forkhead box L1 (FOX-L1)
364 putative binding site; as well as (iv) 232 nucleotide fragments produced from deletion of
365 *GSTe1* 3′-UTR plus 75 nucleotides of the flanking region (-232 primers), located 11 nucleotides
366 from the c-EST/grainy head transcription factor binding site. Also, to assess the importance of
367 the *GSTe2* 5′UTR, 43 nucleotides upstream the AUG codon were deleted (from the 60
368 nucleotide 5′-UTR of both the Sahel-alive and Ngoussou), leaving the putative transcription
369 start site motif untouched. This was done using the original forward primer for the 351

370 nucleotides intergenic region, with a newly designed primer, GSTe2_minus_5'-UTR-R
371 (Additional Table 4).

372

373 ***Characterisation of major resistance genes using transgenic analysis***

374 *Cloning and microinjection of GSTe2 and CYP6Z2 in Drosophila melanogaster*

375 Transgenic flies expressing recombinant GSTe2 and CYP6Z2 were generated using
376 GAL4/UAS system and used in contact bioassays, to confirm if over-expression of these genes
377 alone can confer resistance to insecticides. Transgenic flies were created as outlined in a
378 previous publication (50). Details of amplification of full-length *GSTe2* and *CYP6Z2* from the
379 predominant alleles of sequences of cDNA, cloning into pUASattB vector, microinjection into
380 germ-line *D. melanogaster* lines, crossing of the UAS-lines with the GAL4-Actin driver strain
381 are outlined in Additional File 1.

382 To confirm overexpression of *GSTe2* and *CYP6Z2* in the transgenic flies three replicates each
383 of 6 females (both experimental and control group) were used for qRT-PCR using a previously
384 established protocol (38). Total RNA and cDNA were extracted as described above, and the
385 relative expression levels of transgenes were assessed, with normalization using the *RPL11*
386 housekeeping gene. The qtrg primers used for the two genes and the *RPL11* primers are
387 provided in Additional Table 4.

388 *Insecticide Susceptibility Contact Bioassay*

389 For insecticide bioassays, 3- to 4-day old experimental and control F_1 females were exposed to
390 0.15 % deltamethrin, 2 % permethrin, 0.05 % α-cypermethrin, 4 % DDT and 2 % clothianidin-
391 impregnated papers prepared in acetone and Dow Corning 556 Silicone Fluid (BHD/Merck,

392 Hesse, Germany). Flies overexpressing *CYP6Z2* were also exposed to the primary product of
393 hydrolysis of pyrethroids: 4 % and 20 % (5x) 3-phenoxybenzaldehyde (PBAld) and 3-
394 phenoxybenzylalcohol (PBAlc). Transgenic flies expressing *GSTe2* were exposed to 2 %
395 permethrin, 0.15 % deltamethrin, 0.05 % α-cypermethrin and 4 % DDT only. Impregnated
396 papers were rolled and introduced into 45 cc plastic vials to cover the entire wall and the vials
397 plugged with cotton soaked in 10 % sucrose (38). 20–25 flies were placed in each vial, and the
398 mortality plus knockdown scored at 1 h, 3 h, 6 h, 12 h and 24 h of exposure to the insecticides.
399 For each insecticide, assays were performed in 6 replicates and Student's t-test used to
400 compare the mortality plus knockdown between the experimental groups and the control.

401

402 # Results

403 ## Genome-wide transcriptional profile of the Sahelian *An. coluzzii* populations

404 A three-way pairwise comparison was conducted for the data from each country: resistant *vs*
405 susceptible (R-S), resistant *vs* unexposed control (R-C) and unexposed control *vs* susceptible
406 (C-S). This captures background variations due to geographical differences in the resistant vs
407 susceptible (R-S) comparison, accounts for genes overexpressed due to induction (R-C
408 comparison), as well as genes that are constitutively overexpressed (C-S comparison). A total
409 of 1384 genes were significantly differentially expressed (FDR-adjusted $p < 0.05$ and $\log_2$ fold
410 change threshold of $1/FC \geq 2$) in R-S comparison in Nigeria (1077 upregulated and 307 down-
411 regulated); 1185 genes were differentially expressed in C-S (1002 upregulated and 183
412 downregulated); and 295 genes in R-C (129 upregulated and 166 downregulated). Of these,
413 52 genes were commonly differentially expressed in all 3 comparisons (Figure S3, panel a),

414 including the upregulated genes, *COEAE80* (AGAP006700), *CYP4H18* (AGAP028019),
415 *CYP4H17* (AGAP008358) and cuticular proteins, *CPLCX3* (AGAP006149), *CPR59*
416 (AGAP006829), *CPR76* (AGAP009874) and *CPR75* (AGAP009871). The Additional Figure S3,
417 panels a-d depicts the differentially expressed genes for the four countries. For Niger, 881
418 genes were differentially expressed in R-S comparison (619 upregulated and 262 down-
419 regulated) (Additional Figure S3b); 1256 genes were differentially expressed in C-S (986
420 upregulated and 270 downregulated); and 196 genes in R-C (81 upregulated and 115
421 downregulated). Of these, 22 genes were commonly differentially expressed in all 3
422 comparisons [including the upregulated *aminopeptidase N1* (AGAP012757), *CYP6Z2*
423 (AGAP008218), *chymotrypsin-3* (AGAP006711) and *chymotrypsin-2* (AGAP006710), and an acid
424 trehalase (AGAP008547)]. For Chad, 1392 genes were differentially expressed in R-S
425 comparison (975 upregulated and 417 down-regulated) (Additonal Figure S3c); 1284 in C-S
426 (105 upregulated and 269 downregulated); and 526 genes in R-C (270 upregulated and 256
427 downregulated). Of these, 97 genes were commonly differentially expressed in all 3
428 comparisons [including *CYP4C27* (AGAP009246), *SULTD1* (AGAP012672), *aminopeptidase N1*,
429 and diverse cuticular proteins, e.g. chitinase (*Cht24*, AGAP006191, *CPFL1* (AGAP010902),
430 *CPCFC1* (AGAP007980), *CPLCX3*, *CPR24* (AGAP005999), *CPR106* (AGAP006095) and *CPR130*
431 (AGAP000047). Finally, for Cameroon, 376 genes were differentially expressed in R-S
432 comparison (204 upregulated and 172 down-regulated) (Additional Figure S3d); 932 in C-S
433 (778 upregulated and 154 downregulated); and 116 genes in R-C [only 9 upregulated and 107
434 downregulated (probably due to a single, low quality replicate in the raw data from
435 Cameroon (Additional Figure S2). Not surprising, only 7 genes were commonly differentially
436 expressed in all 3 comparisons. These include the highly upregulated gene, *GSTe2*
437 (AGAP009194) and *chymotrypsin-1* (AGAP006709).

438 All data analysed together (Additional Figure S3e) revealed no single gene differentially
439 expressed in common; possibly due to the low quality with the Chad unexposed (C) data.
440 Analysis of data from Nigeria, Niger and Chad, revealed a single gene (AGAP000046,
441 transporter major facilitator superfamily) differentially expressed across all countries
442 (Additional Figure S3f). Niger and Chad shared *hexamerin* (AGAP010658), *aminopeptidase N1*,
443 and unknown protein, AGAP0290967; Nigeria and Niger share only a single gene,
444 AGAP003248; while seven genes were common to Nigeria and Chad, including *CPLCX3*.

445 Principal component analysis for the top 500 most variable genes in all experimental arms
446 revealed data from field samples (R and C) from all four countries clustering closer in PC1
447 and PC2 axes, away from the data from the susceptible females, Ngoussou (Additional Figure
448 S4).

449 *Analysis of the common differentially expressed genes across the Sahel*

450 The most differentially expressed genes were presented in Figure 1, a volcano plot of fold
451 change *vs* significance levels for Nigeria and Niger, and Figure 2, for Chad and Cameroon.
452 Exhaustive lits of these genes of interest is provided in Additional File 2. Comparisons of
453 genes commonly, upregulated and/or downregulated in R-S/R-C/C-S, from the four countries
454 revealed similar transcriptomic profiles between R and C compared with the S. The most
455 commonly and consistently overexpressed genes across the Sahel (taking account mean
456 expressions) are the chymotrypsins, 3, -2 and -1 (*CHYM3*/AGAP006711 and
457 *CHYM2*/AGAP006710 and *CHYM1*/AGAP006709) (Figures 1 and 2, Additional File 2), the
458 glutathione S-transferase, *GSTe2* (AGAP009194), an aquaporin, *AQP3* (AGAP010326),
459 *CYP6Z2* (AGAP008218), *CYP6Z3* (AGAP008217), *CYP4C27* (AGAP009246), a chitinase, *Cht24*
460 (AGAP006191), a thioester-containing protein-1, *TEP-1* (AGAP010815),), a trehalose 6-

461 phosphate synthase/phosphatase, *TPS*1/2 (AGAP008227), a lipase (AGAP002353), and
462 AGAP012818 (V-type protein ATPase subunit A). More on these genes is provided in the
463 sections, below.

464 Other genes commonly overexpressed in data from two or three countries include
465 AGAP005501 (dehydrogenase/reductase SDR family 11) upregulated in Nigeria and Niger (R-
466 S and C-S comparisons), and in R-S for Chad and Cameroon; AGAP008091 (CLIP-domain
467 serine protease, *CLIPE1*), upregulated in all countries in R-S and C-S comparisons; a chitinase,
468 *Cht5-5*, AGAP013260) upregulated in Nigeria, Nigeria and Chad R-S and C-S, with both
469 induced upregulated in R-C in Chad, as heme peroxidase (*HPX15*), upregulated in R-S and C-
470 S comparisons in Nigeria, Chad and Cameroon, as well as malate dehydrogenase
471 (AGAP000184), upregulated in Nigeria, Chad and Cameroon, R-S and C-S.

472 However, the most overexpressed genes in Nigeria are a cubilin, a histone (H2B), carbonic
473 anhydrase I, and a cuticular protein, *CPR131* (Additional File 2), while for Cameroon its H2B,
474 Serpin 9 inhibitory serine protease inhibitor, galectin 4 and a Protease m1 zinc
475 metalloprotease.

476 Several genes were significantly downregulated across the four countries, particularly the
477 cytochrome P450s and the ATP-binding cassette transporters. The most consistently down-
478 regulated P450s were *CYP6AK1* (AGAP010961), downregulated in R-S and C-S comparisons
479 in Nigeria, Niger and Cameroon, while down-regulated in all comparisons in Chad; *CYP4H24*
480 (AGAP013490), downregulated in R-S and C-S comparisons for Nigeria and Cameroon, and
481 downregulated in all comparisons in Niger and Chad; *CYP6M4* (AGAP008214),
482 downregulated in R-S and C-S comparisons in Nigeria, Niger and Cameroon, while

483 downregulated in all comparisons in Chad; and finally *CYP6P5* (AGAP002866),
484 downregulated in R-S and C-S comparisons in Niger and Chad, but only in C-S in Nigeria.

485 *Analysis of commonly overexpressed metabolic resistance genes across the Sahel*

486 Special attention was given to the known metabolic resistance genes, implicated in insecticide
487 resistance in *Anopheles* and/or other insects, including, CYP450s, GSTs, carboxylesterases,
488 cuticular proteins, chemosensory proteins/SAPs, uridine diphosphoglucuronosyltransferases,
489 etc, in addition to other important genes, e.g., the immune proteins. Analysis of the data from
490 the list of genes significantly, differentially expressed (FDR-adjusted $p < 0.05$ and $\log_2 FC = 1/$
491 $FC \geq 2$) revealed the major, commonly upregulated metabolic resistance genes in the four
492 Sahel countries. Top 50 genes from these lists (Additional File 3) from each country are
493 displayed as heatmaps of fold changes in Additional Figure S5a, -b, -c and -d). The most
494 common genes linked with resistance and/or other physiologically important phenotypes are
495 tabulated in Table 1, showing fold changes of 55 genes (R-S and C-S comparisons), common
496 to all countries, or three or two countries at least. The most commonly and consistently
497 overexpressed genes across the Sahel (taking account mean expressions) are the
498 chymotrypsins, -1, -2 and -3 (*CHYM1*/AGAP006709, *CHYM3*/AGAP006711 and
499 *CHYM2*/AGAP006710) (Table 1, Additional Figure S5, Additional File 2, Results section), with
500 fold change for *CHYM1* in R-S and C-S comparisons of 31.62 and 54.41 for Nigeria, 12.31 and
501 41.59 for Niger, and 18.00 and 18.08 for Chad, and 5.69 and 11.77 for Cameroon; for *CHYM3*,
502 32.45 and 58.43 for Nigeria R-S and C-S, respectively, 7.30 and 43.27 for Niger, 13.40 and 7.24
503 for Chad, and 2.59 and 8.15 for Cameroon. Of *Anopheles* metabolic resistance genes, *GSTe2*
504 (AGAP009194) was the most consistently upregulated gene across the Sahel, FC = 9.61 and
505 7.16, respectively for Nigeria R-S and C-S comparisons; 6.55 and 11.32 for Niger; 9.09 and 10.02

506 for Chad, 9.68 and 20.67 for Cameroon. Other common GSTs were *GSTZ1*, upregulated in
507 three countries (Nigeria, Niger and Chad), *GSTe4*, *GSTU1*, *GSTD1-4* and *GSTD3*.

508 Other most consistently and highly upregulated genes include a lipase, AGAP002353: FC of
509 26.73 and 20.80 for Nigeria R-S and C-S comparisons, 16.53 and 29.76 for Niger, 14.95 and 9.95
510 for Chad, 14.58 and 14.53 for Cameroon; a chitinase, *Cht24* (AGAP006191), FC = 21.12 and
511 38.55 for Nigeria R-S and C-S, 40.08 and 67.18 for Niger, 91.42 and 5.71 for Chad, 44.26 and
512 40.53 for Cameroon); a V-type protein ATPase subunit a (AGAP012818, FC = 22.12 and 14.61
513 for Nigeria, 31.05 and 25.43 for Nigeri, 6.00 and 4.31 for Chad, 44.92 and 43.4 for Cameroon),
514 and an aquaporin, *AQP3* (AGAP010326, FC = 14.81 and 11.03 for Nigeria, 12.09 and 21.94 for
515 Niger, 14.86 and 1.21 for Chad, 14.17 and 17.3 for Cameroon).

516 The most overrepresented gene family, and most consistently upregulated across the Sahel
517 were the cytochrome P450s, with *CYP6Z2*, *CYP6Z3*, and *CYP4C27* taking the front seat, in
518 addition to the two P450s linked to cuticular resistance, *CYP4G16* (upregulated in Nigeria and
519 Chad) and *CYP4G17* (upregulated in Nigeria, Niger and Chad). The three P450s, *CYP6Z2,*
520 *CYP6Z3 and CYP4C27* are consistently overexpressed: FC for *CYP6Z2* is 6.55 and 2.10 for
521 Nigeria R-S and C-S comparisons, 6.73 and 2.32 for Niger, 13.96 and 1.48 for Chad, 11.84 and
522 8.39 for Cameroon; FC for *CYP6Z3* is 4.73 and 1.82 for Nigeria, 6.09 and 1.42 for Niger, 3.51
523 and 1.11 for Chad, 7.11 and 3.79 for Cameroon; and FC for *CYP4C27* is 7.36 and 4.6 for Nigeria,
524 4.03 and 5.64 for Niger, 6.04 and 2.62 for Chad, 2.01 and 2.63 for Cameroon. The two well-
525 known insecticide resistance genes *CYP6P3* and *CYP6M2* did not future prominently across
526 the Sahel of these countries.

527 The second most commonly overrepresented gene families belong to cuticular proteins,
528 including cuticular proteins Rebers and Riddiford (RR), several cuticular proteins of low

529 complexity (CPLC) and chitin-binding cuticular proteins (CPAP), with the commonly
530 upregulated ones across all four countries being *CPR75* (AGAP009871) and *CPR76*
531 (AGAP009874).

532 Two carboxylesterases picture prominently across the Sahel – the beta esterase, *COEBE3C*
533 (AGAP005372) upregulated in all four countries, while *COEJHE5E* was common to Nigeria,
534 Niger and Chad. For Phase II metabolism enzymes, several uridine-diphospho-
535 glucuronosyltransferases, UGTs were upregulated, including the *UGT-B19*, which is
536 upregulated in Nigeria and Niger (Table 1), and unknown UGT (AGAP006222) upregulated
537 across all the countries (Additional File 2).

538 Other genes in the top 55 metabolic genes include the malate dehydrogenase, *TPS* (*TPS1/2*),
539 *AQP2*, thioester-containing proteins, *TEP1, -12* and *-14,* a cyanogenic-beta-glucosidase
540 (AGAP006422) and three chemosensory proteins, *CSP5*, sensory appendage proteins, -1 and -
541 2.

542

543

544

545

546

547

548

549 **Table 1: Common detoxification and metabolic genes differentially upregulated in Sahel**
550 ***An. coluzzii* (FDR-adjusted p < 0.05).**

	NIGERIA		NIGER		CHAD		CAMEROON		
Gene	**R-S**	**C-S**	**R-S**	**C-S**	**R-S**	**C-S**	**R-S**	**C-S**	**Gene Description**
AGAP006400	2.2	4.2	1.7	4.5	-	-	2.1	2.8	alkaline phosphatase 2
AGAP028491	4.3	2.8	1.4	2.7	2.9	2.6	-	-	aquaporin, AQP2
AGAP010326	14.8	11.0	12.1	21.9	14.9	1.2	14.2	17.3	Aquaporin, AQP3
AGAP005372	9.5	12.2	3.3	4.3	6.6	3.6	1.4	2.6	carboxylesterase, COEBE3C
AGAP005837	4.7	5.0	1.4	3.3	1.9	2.6			carboxylesterase, COEJHE5E
AGAP004904	3.9	2.2	-	-	2.3	1.7	-	-	catalase, CAT1
AGAP029127	4.1	3.2	-	-	1.7	2.2	-	-	chemosensory protein 5, CSP5
AGAP006191	21.1	38.6	40.1	67.2	91.4	5.7	44.3	40.5	chitinase, Cht24
AGAP013260	4.4	11.5	1.3	3.9	3.1	0.8	-	-	chitinase, Cht5-5
AGAP006898	8.9	5.5	-	-	2.2	3.9	-	-	chitinase, Cht6
AGAP006709	31.6	54.4	12.3	41.6	18.0	18.1	5.7	11.8	chymotrypsin-1, CHYM1
AGAP006710	9.9	19.4	2.1	8.8	4.2	3.9	-	-	chymotrypsin-2, CHYM2
AGAP006711	32.5	58.4	7.3	43.3	13.4	7.2	2.6	8.2	chymotrypsin-3, CHYM3
AGAP000987	4.3	2.7	-	-	1.3	2.6	-	-	cuticular protein, CPAP3-A1b
AGAP000988	5.3	1.6	-	-	0.7	5.4	-	-	cuticular protein, CPAP3-A1c
AGAP000986	7.3	4.6	-	-	1.7	3.5	-	-	cuticular protein, CPAP3-D
AGAP006149	3.8	0.8	-	-	0.7	5.8	-	-	cuticular protein, CPLCX3
AGAP010123	33.3	54.9	2.4	13.2	-	-	-	-	cuticular protein, CPR131
AGAP005456	5.2	1.8	-	-	1.2	2.2	1.1	2.1	cuticular protein, CPR15
AGAP006009	-	-	2.9	1.3	-	-	3.9	1.6	cuticular protein, CPR30
AGAP009871	8.7	2.7	1.1	2.6	2.2	4.2	1.7	2.6	cuticular protein, CPR75
AGAP009874	10.6	2.2	1.1	2.6	3.7	5.6	1.1	2.0	cuticular protein, CPR76
AGAP009879	7.4	2.6	-	-	1.9	4.6	-	-	cuticular protein, CPR81
AGAP006422	7.5	9.4	-	-	8.4	7.3	2.1	2.0	cyanogenic beta-glucosidase
AGAP002417	5.5	4.2	0.9	1.2	1.9	1.9	-	-	cytochrome P450, CYP4AR1
AGAP009246	7.4	4.6	4.0	5.6	6.0	2.6	2.0	2.6	cytochrome P450, CYP4C27
AGAP012957	5.0	2.8	-	-	2.9	2.5	2.0	2.6	cytochrome P450, CYP4D17
AGAP001076	5.8	4.7	1.5	1.8	3.8	2.4	0.9	1.5	cytochrome P450, CYP4G16
AGAP000877	3.8	2.0	2.3	1.9	3.8	2.1	1.2	2.1	cytochrome P450, CYP4G17
AGAP008358	-	-	2.9	4.1	3.4	3.3	3.6	2.0	cytochrome P450, CYP4H17
AGAP007480	5.4	3.9	1.3	2.7	2.5	2.0	-	-	cytochrome P450, CYP6AH1
AGAP002865	1.11	0.5	2.3	1.9	0.7	12.8	2.1	2.3	cytochrome P450, CYP6P3
AGAP008207	3.6	1.9	-	-	2.1	1.6	1.7	2.4	cytochrome P450, CYP6Y2
AGAP008218	6.6	2.1	6.7	2.3	14.0	1.5	11.8	8.4	cytochrome P450, CYP6Z2
AGAP008217	4.7	1.8	6.1	1.4	3.5	1.1	7.1	3.8	cytochrome P450, CYP6Z3
AGAP000818	3.0	1.2	0.8	0.5	0.9	1.6	3.4	2.2	cytochrome P450, CYP9K1
AGAP010400	3.6	2.6	1.6	2.8	3.3	1.4	-	-	flavin-containing monooxygenase
AGAP007920	5.7	2.4	2.8	1.8	-	-	-	-	glucuronosyltransferases, UGT-B19
AGAP004164	2.2	1.9	1.7	2.3	1.4	1.3	-	-	glutathione S-transferase, GSTD1-4
AGAP009194	9.6	7.2	6.6	11.3	9.1	10.0	9.7	20.7	glutathione S-transferase, GSTe2
AGAP009193	3.3	2.3	1.1	2.1	1.5	1.6	-	-	glutathione S-transferase, GSTe4
AGAP000947	4.4	2.3	1.2	2.2	1.9	3.2	-	-	glutathione S-transferase, GSTU1
AGAP002898	5.5	3.5	1.7	2.6	2.8	2.4	-	-	glutathione S-transferase, GSTZ1
AGAP002198	-	-	2.0	5.8	4.8	6.6	-	-	glycine-N-methyltransferase
AGAP013327	-	-	1.2	7.3	7.9	20.1	5.1	14.8	heme peroxidase, HXP15
AGAP002353	26.7	20.8	16.5	29.8	15.0	10	14.6	14.5	lipase
AGAP000184	5.2	6.3	-	-	4.9	4.3	1.2	1.9	malate dehydrogenase
AGAP008051	3.7	4.0	1.1	1.3	1.7	1.7	-	-	sensory appendage protein 1, SAP1
AGAP008052	5.5	10.8	1.0	1.5	3.6	1.1	-	-	sensory appendage protein 2, SAP2
AGAP010815	5.7	14.4	4.1	8.6	7.8	3.8	1.7	3.1	thioester-containing protein, TEP1
AGAP008368	3.3	6.1	1.4	3.9	3.4	2.0	-	-	thioester-containing protein, TEP14
AGAP008654	4.6	17.4	2.1	10.0	-	-	-	-	thioester-containing protein, TEP12
AGAP008227	4.1	4.0	4.0	2.0	4.6	1.5	2.6	4.1	trehalose 6-phosphate synthase/phosphatase
AGAP012818	22.1	14.6	31.1	25.4	6.0	4.3	44.9	43.4	V-type proton ATPase, subunit a

551 ***Gene Ontology Enrichment Analysis***

552 Gene ontology enrichment analysis, for genes significant in comparisons ($\log_2 FC = 1$, $p < 0.05$)
553 revealed crucial differences in overrepresented GO terms between the up- and -
554 downregulated genes, in data from the four countries. For example, the most over-
555 represented, semantically similar GO terms associated with xenobiotics metabolism is
556 oxidoreductase activity ((highest frequency of 12.9 %, Additional Figure S6a, REVIGO Table
557 View), which cluster together in Nigerian R vs S comparison (for genes upregulated in R).
558 Other over-represented GO terms in this comparison include glutathione S-transferase
559 activity (the most enriched/specific term), peroxidase activity, odorant binding and chitin
560 binding. In contrast, GO terms over-represented in R vs S (down-regulated in R) were mostly
561 involved in neurotransmission, metal ions binding and receptor channelling activities
562 (Additional Figure S6b).

563 For R vs C comparison (upregulated in C) the over-represented GO terms include
564 oxidoreductase (frequency = 12.88 %), glutathione S-transferase activity (the most
565 enriched/specific term), peroxidase activity, glucuronosyltransferase activity, aldehyde
566 oxidase activity, chitin binding, carbohydrate binding activities, etc, (Additional Figure S6c).
567 In contrast the most enriched GO terms, downregulated in C were endopeptidase activities,
568 Toll binding, oxidoreductase (frequency = 1.21 %) and monooxygenase activities (Additional
569 Figure S6d). For the rest of the three countries similar contrasts were also observed, between
570 phenotypes, and detailed in Additional File 1, Results section.

571

572

573 **Quantitative PCR validation of expression profiles of metabolic resistance genes**

574 The relative expression levels of 12 metabolic resistance genes were validated. The qRT-PCR
575 results support the transcriptomic patterns obtained from the RNAseq analysis, with *GSTe2*
576 as the most overexpressed gene, followed by *CYP6Z2* (Figure 3). For example, for R-S and C-
577 S comparisons *GSTe2* had fold changes of 49.25 ± 5.05 and 28.45 ± 4.45, for Nigeria ($p \leq 0.0001$);
578 37.01 ± 2.22 and 33.34 ± 6.61 for Niger ($p \leq 0.0001$); 29.97 ± 5.39 and 30.16 ± 6.16 for Chad
579 ($p \leq 0.0001$); and 28.30 ± 3.30 ($p \leq 0.0001$) and 17.03 ± 7.03 for Cameroon ($p \leq 0.001$). *CYP6Z2* has
580 R-S and C-S fold changes of 15.93 ± 2.54 ($p \leq 0.001$) and 7.35 ± 1.52 ($p \leq 0.05$), for Nigeria; 24.21
581 ± 2.17 ($p \leq 0.001$) and 20.83 ± 3.02 for Niger ($p \leq 0.001$); 15.02 ($p \leq 0.001$) ± 2.01 and 6.09 ± 2.08 for
582 Chad ($p \leq 0.05$); and 10.58 ± 3.00 ($p \leq 0.01$) and 3.43 ± 1.43 for Cameroon. Correlation analyses
583 for R-S comparisons for all genes revealed positive and significant association (Additional
584 Figure S7) between the RNA-seq and qRT-PCR data in data from Niger ($R^2 = 0.58$, p = 0.03)
585 and Chad ($R^2 = 0.408$, p = 0.025), with positive, but non-significance seen in Cameron ($R^2 = 0.58$
586 p = 0.08) and Nigerian ($R^2 = 0.503$, p = 0.06).

587

588 **Detection of Signatures of Selective Sweeps**

589 Signatures of selection were investigated in the major metabolic resistance genes, by
590 estimating Tajima's D per gene within populations, and *F*st per gene between population
591 pairs. These tests of neutrality revealed several genes exhibiting high genetic differentiation,
592 or possibly undergoing expansion. Among the top 100 most overexpressed metabolic genes
593 across the four countries (Additional File 2), 13 genes were possibly undergoing genetic
594 differentiation. These include *TEP1,* and *TEP3*, with average Tajima's D of -1.4 and -1.6

595 respectively, in populations from Chad, and average *F*st values of 0.58 and 0.49 for all
596 countries (Additional File 3, provide the number of SNPs and chromosomal location for each
597 gene), *CYP9K1* (average Tajima's D = -1.00 for Cameroon, and combined *F*st = 0.55), *CPR15*
598 (average Tajima's D = -0.94, and *F*st = 0.17); *CPAP3-A1b* (Tajima's D = -0.70 in Chad, -0.95 in
599 Nigeria, and -1.32 in Cameroon, with combined *F*st of 0.55) and *CPLCX3*. Several genes from
600 the GST family exhibited strong genetic differentiation, with reduced variations prominently
601 in *GSTe2* (< 3 SNPs in Chad and Cameroon samples), with Tajima's D of -1.45 and -1.15 for
602 Nigeria and Niger, and combined *F*st of 0.02. Detailed analysis is provided in Additional File
603 1, and Additional File 3 contain the respective frequencies of these variables.

604

605 **Identificaton of genetic variants associated with insecticide resistance**

606 The RNA-Seq-Pop workflow calculated allele frequencies of variants of interest found in the
607 raw RNA-sequencing read data. Figure 4 displays allele frequencies of each variant in
608 respective treatment replicates used for the RNAseq, as well as Ngoussou. Several mutaions
609 were found within the voltage-gated sodium channel. In addition to the L995F *kdr* mutation,
610 the recently identified V402L and I1527T mutations exist in all the field populations across the
611 Sahel, at frequencies ranging from 29 % to 67 %. We also detected the N1570Y mutation, which
612 shares the same haplotype as L995F. The G280S *acetylcholinesterase*-1 mutation was also
613 observed in low to moderate frequencies in all populations except for Ngoussou and samples
614 from Niger. The L43F pyrethroid resistance marker of *CYP4J5* was also seen, fixed in most
615 populations.

616

617 **Detection of chromosomal inversion polymorphisms and metabolic genes within**
618 **its breakpoints**

619 The frequency of the major inversion polymorphisms in chromosome 2 were calculated,
620 considering ploidy (8 individual female mosquitoes were pooled for RNA extraction, for each
621 replicate). Additional Table 5 provided the frequencies of the respective inversions for the
622 populations from each country, as well as for Ngoussou. High frequency of the 2La, 2Rb and
623 2Rc were observed, in contrast to the 2Rd, 2Rj and 2Ru which were in lower frequencies
624 [except for the 2Rj in Nigeria (30.38 %) and 2Ru in Cameroon (24.29 %)]. The 2La inversion
625 was found fixed in the field populations (100 % in Nigeria and Cameroon, 99.92 % in Niger
626 and 99.21 % in Chad), in contrast with Ngoussou, with frequency of only 6.26 %. Similar
627 pattern was observed with the 2Rb inversion, with high frequencies in three field populations
628 (79.01 %, 84.85 %, 80.22 % in Nigeria, Niger and Chad), but lower in Cameroon (29.08 %) and
629 Ngoussou (5.35 %). Frequencies of the 2Rc inversion were 88.54 %, 85.52 %, 88.82 %, 46.99 %
630 and 9.89 % for Nigeria, Niger, Chad, Cameroon and Ngoussou respectively. Some genes
631 among the top 100 most overexpressed metabolic resistance genes (above) and which were
632 likely undergoing recent population expansion were located within the 2La, 2Rb and 2Rc
633 inversions (Additional File 4). These include the *hsp83* (AGAP006958) and molecular
634 chaperone hptG (AGAP006961), *CPLCX3* (AGAP006149), *CPLCA1* (AGAP006145,) within the
635 inverted 2La chromosomal arm. Other genes within 2La inversion include a group of
636 chitinase enzymes (*Cht6*, *-8*, *-11* and *-24*), most notably the *Cht24* and *Cht6*, which are among
637 the top 50 most overexpressed genes in all countries. Other genes include: *COEJHE5E*, one of
638 the 14 carboxylesterases in the 2La region, and which was among the top 100 most
639 overexpressed genes across the Sahel; the most overexpressed chymotrypsin genes, *CHYM1*

640 and *CHYM2*; three ionotropic receptors, *IR136* (AGAP006440), *IR139* (AGAP006691) and
641 *IR142* (AGAP006407); as well as five P450s, *CYP301A1*, *CYP302A1*, *CYP4J5*, *CYP4J9* and
642 *CYP4J10*. The most overrepresented genes within 2La inversions are the CPR cuticular
643 proteins (73 in total), with *CPR21*, *-26*, *-30*, *-59* and *-140* from the list of the top 100 resistance
644 genes from the Sahel.

645 Carbonic anhydrase I (the most overexpressed metabolic resistance in Nigeria, AGAP013402)
646 sits within 2Rb inversion. Four chitinases (*Cht4*, *Cht5-1*, *Cht5-3*, and *Cht5-5*) are located within
647 2Rb inversion; *Cht5-1* and *Cht5-5* were among the observed top 50 most overexpressed
648 metabolic genes in the Sahel. Lipase (*AGAP002353*) which is the among the top 6
649 overexpressed metabolic gene in the four countries sit within the 2Rb inversion. Eight
650 CYP450s reside in the 2Rb inversion, including *CYP4D15*, *CYP4D17* and *CYP4AR1*, which
651 were among the top 100 metabolic resistance genes. Three cuticular proteins, *CPR7*, *-8* and *-9*
652 were also within the 2Rb inversion.

653 Interesting genes sitting within 2Rc inversion include *GSTZ1*; 5 ionotropic receptor genes, *IR7i*
654 (AGAP013363), *IR7u* (AGAP013285), *IR7t* (AGAP002763), *IR7w* (AGAP013416) and *IR41a*
655 (AGAP002904); a carboxylesterase, *COEAE6O* (AGAP002863); and 11 CYP450s, including
656 *CYP6AA1* (AGAP002862), *CYP6AA2* (AGAP013128), *CYP6P15P* (AGAP002864), *CYP6P3*
657 (AGAP002865), *CYP6P5* (AGAP002866), *CYP6P4* (AGAP002867), *CYP6P1* (AGAP002868),
658 *CYP6P2* (AGAP002869), *CYP6AD1* (AGAP002870) and *CYP6Z4* (AGAP002894)

659

660

661

662 **Investigation of the polymorphism in the coding sequences of *GSTe2* and *CYP6Z2***

663 Analysis of the polymorphism patterns of full-length cDNA sequences of *GSTe2* (666 bp) and
664 *CYP6Z2* (1479 bp) from the Sahel region of Africa revealed complete homogeneity for *GSTe2*,
665 with no polymorphism detected in the field populations (all sequences were identical to those
666 from Ngoussou and the AGAP009194 reference). This suggests a fixed allele, consistent with
667 the observation from the analyses from the fixation index (Additional File 3). For *CYP6Z2*
668 (1479 bp), homogeneity was observed within each country and Ngoussou, except for Niger,
669 characterised by an unusually high polymorphism (Additional Figure S8a, -b). *CYP6Z2* is
670 polymorphic with 10 haplotypes across the Sahel, with 75 polymorphic sites of which 65 were
671 synonymous, and 12 led to amino acids substitutions. The bulk of the polymorphisms were
672 contributed from larger variations in the Niger and Cameroon (S = 42 and 23 respectively),
673 while highest homogeneity was observed in Chad, with a single haplotype. Haplotype
674 diversity is high (H_d = 0.921), from 10 haplotypes out of only 20 sequences, with the lowest
675 H_d in the Chad sequences, suggesting a directional selection/fixed allele. The haplotypes
676 cluster according to origin on the maximum likelihood phylogenetic tree, except for Niger
677 (Additional Figure S8c).

678

679 **Investigation of the role of intergenic region elements in overexpression of *GSTe2***
680 **and *CYP6Z2***

681 ***Investigation of polymorphism in the intergenic region/regulatory elements***

682 To investigate polymorphisms in the regulatory elements of the above genes, the 351
683 intergenic regions of *GSTe2* (spanning the 43 bp 3'-UTR of *GSTe1*, 248 bp flanking sequence

684 and 60 bp 5′-UTR of *GSTe2*) preceding the start codon were amplified from 10 each of DDT-
685 alive and -dead females from the 4 Sahel countries, as well as from Ngoussou females,
686 successfully. For *CYP6Z2,* 1078 bp (spanning 38 bp 3′ UTR of *CYP6Z1,* 937 bp flanking
687 sequence and 103 bp 5′UTR of *CYP6Z2,* preceding the start codon of *CYP6Z2*) were used to
688 amplify fragments from 10 each of deltamethrin-alive and -dead females from Nigeria, and
689 from Ngoussou females.

690 Out of the 90 *GSTe2* 5′-UTR sequences analysed differences were observed between the alive
691 and dead mosquitoes, with a total of 65 sequences from alive and dead females (regardless of
692 country of origin) identical to the 10 sequences of Ngoussou. From the 40 sequences of the
693 dead females, 39 were identical to Ngoussou (regardless of the country). Twelve sequences,
694 all from the alive females (from across the four countries) were similar, with several mutations
695 in putative transcriptional factors binding sites, which may impact overexpression of the
696 *GSTe2.* In short, 8 mutations were shared in common between these 12 sequences of DDT-
697 alive females from across the countries (3 each from Nigeria, Chad and Cameroon) and
698 additional mutations in the Niger samples (4 sequences). These mutations include (i) T->A
699 transition withing the cellular myb-DNA (c-myb) binding domain (Additional Figure S9), (ii)
700 a T->C transversion in the zinc-finger homeodomain, δEF1 (δ elongation factor 1) binding site,
701 (iii) T->A transition in the nuclear matric protein 4 (NMP4), (iv) simultaneous insertion of
702 adenine and a transition T -> C, in positions 113 and 114 respectively, between the Fork-head
703 box L1 and c-EST binding sites, (v) an A -> C transversion in a second NMP4, (vi) a T -> C
704 transition, six nucleotides downstream the nuclear factor κB (NF-κB), (vii) followed by an A -
705 > G transversion 3 nucleotides downstream the NF-κB, (viii) a G -> A transversion, 7
706 nucleotides upstream the GC box, (ix) a C -> G transition, 2 nucleotides downstream the

707 arthropod initiator (Inr consensus) sequence, and finally (x) a C -> G transition, within the 5′-
708 UTR of the *GSTe2*, 28 nucleotides downstream the transcriptional start site/49 nucleotides
709 upstream the *GSTe2*, start codon.

710 Analysis of the 90 sequences revealed a very low polymorphism in the dead mosquitoes (S =
711 0, for Niger-dead, Chad-dead and Ngoussou), but high polymorphism in the alive (S = 16 for
712 Nigeria-alive, and 13 each for Chad-alive and Cameroon-alive) (Additional Table 6). All
713 sequences produced 17 polymorphic sites, with 6 haplotypes (Additional Figure S10a and -b).
714 High haplotype diversities were obtained from the alive mosquitoes (for example, Nigeria
715 alive, Hd = 0.71, Niger-, Hd = 0.53). Regardless of country of origin, the haplotypes cluster
716 according to phenotype on the maximum likelihood phylogenetic tree, with the alive
717 haplotypes forming a distinct/separate clade (Additional Figure S10c).

718 With regard to *CYP6Z2* no major differences were observed when comparing the
719 deltamethrin-alive and -dead sequences with the Ngoussou.

720

721 ***Measurement of activities of the 5′ regulatory elements of the GSTe2 and CYP6Z2***

722 Initial promoter analyses were conducted with the 5′- regulatory element sequences of *GSTe2*,
723 for the DDT-alive females (representative sequence with the 8 common polymorphic
724 positions, designated, Sahel-alive), comparing it with the sequence from the DDT-
725 dead/Ngoussou (designated Ngoussou/Sahel-dead). For *CYP6Z2*, a predominant sequence
726 from the Nigeria field sample (with 8 nucleotide insertion) was compared with the Ngoussou
727 sequence.

728 The ability to drive heterologous expression of the firefly luciferase was determined, with the
729 Sahel-alive construct and Ngoussou/Sahel-dead producing increased luciferase activity
730 [~3090× (normalized luciferase activity = 2.58) and 479× (normalized activity = 0.400),
731 respectively] compared with the promoterless pGL3-Basic vector. But the Sahel-alive
732 construct was significantly more active than the Ngoussou/Sahel dead counterpart [promoter
733 activity ~6-fold higher (Tukey HSD Q = 8.11, p = 0.004)]. In contrast, for *CYP6Z2* no significant
734 differences were observed when comparing the field construct (normalized luciferase activity
735 = 5.95 for alive/dead field construct), compared with 5.84 for the Ngoussou construct.

736

737 ***GSTe2 promoter delineation and measurement of activity***

738 Sequential deletion of the intergenic region of the *GSTe2* resulted in progressive reduction in
739 luciferase activity. Deletion of the 43 bp 3′-UTR of the *GSTe1* (-308 from the start codon of
740 *GSTe2*) reduced activity of the Sahel-alive construct by only 18.6 % suggesting that the c-Myb
741 transcriptional factor binding site may not be critical for overexpression (Figure 5). Deletion
742 of an additional 38 nucleotides from the flanking region (-270 fragment, obliterating the δEF1
743 binding site) had comparable impact as above, with activities reducing by 15.3 % only. But
744 removing the *GSTe-1* 3′-UTR and an additional 46 nucleotides from the flanking region (-262
745 fragment, which obliterated in addition the FOX-L1 putative binding site) reduced activity by
746 64.1 % (Q = 6.18, p = 0.01) indicating the importance of this binding site. Deletion of *GSTe1* 3′-
747 UTR plus 75 nucleotides of the flanking region (-232 fragment, located 10 nucleotides from
748 the c-EST binding site) significantly reduced the activity by ~70% (Q = 6.14, p = 0.01),
749 suggesting the importance of the simultaneous insertion of adenine and transition T -> C, in
750 positions 113 and 114 respectively, between the FOX-L1 and c-EST binding sites. Removing

751 the fragment of the 5'UTR of *GSTe2* (43 nucleotides preceding the AUG codon obliterated the
752 promoter activity, reducing the luciferase expression by 98 % (Q = 9.41, p = 0.002). This is
753 despite presence of all the above binding sites and transcriptional start site in this fragment.
754 Deletion of the 5'-UTR from Ngoussou significantly reduced activities as well (reduction by
755 88 % compared with the full Ngoussou intergenic region construct, Q = 13.75, p = 0.001).
756 Overall, these findings suggest that the essential binding sites for overexpression of *GSTe2*
757 span the FOX-L1 and c-EST binding sites, with the 5'-UTR essential for activity.

758

759 **Investigating the role of *GSTe2* and *CYP6Z2* in insecticides resistance using**
760 **transgenic analysis**

761 A qRT-PCR was conducted using the transgenic flies to first establish overexpression of the
762 above genes. Relative fold changes (FC) of 32.7 ± 4.6 and 25.30 ± 2.8 were obtained in flies
763 overexpressing the *GSTe2* and *CYP6Z2* respectively, compared with control flies (progenies
764 of crosses between the parental line flies with no gene insertion, crossed with GAL4/UAS
765 driver line) (Additional Figure S11).

766 Contact bioassays carried out using 0.05 % α-cypermethrin revealed a high susceptibility in
767 the transgenic flies expressing *GSTe2* (Act5C-GAL4-UAS-GSTe2) and controls (Figure 6a),
768 with mortalities increasing from 62 % and 70 % respectively in 1 h, to 94 % and 99 % in 24 h.
769 However, significantly reduced mortalities were observed in transgenic flies expressing
770 *CYP6Z2* (Act5C-GAL4-UAS-CYP6Z2) compared to control flies, at 1 h (mortality = 35 % *vs* 70
771 % in control, p < 0.001) and 3 h (mortality = 52 % vs 89 % in the control, p < 0.001). High
772 susceptibilities were also seen in all the experimental flies exposed to 0.05 % deltamethrin

773 (both for *GSTe2* and *CYP6Z2* flies), except for 1 h with flies expressing *CYP6Z2* (mortality = 47
774 % compared with 58 % in the control flies, $p < 0.010$).

775 The *GSTe2* transgenic flies were highly resistant to 2 % permethrin, with no mortality at all in
776 1 h (Figure 6b), and average mortalities of only 8.5 % at 24 h ($p < 0.0001$), compared with 85 %
777 at 1 h for control flies, which increased to 100 % from 3 h.

778 Initial exposure to 4 % of 3-phenoxybenzaldehyde (PBAld) and 3-phenoxybenzyl alcohol
779 (PBAlc) had no toxic effect on *CYP6Z2* transgenic flies (Figure not shown). However, 5x
780 concentration of these primary products of pyrethroid hydrolysis induced mortalities, albeit
781 low in all flies (Figure 6c). For PBAld, mortalities ranged from 15.5 % and 17.5 % at 1 h, for
782 *CYP6Z2* transgenic flies and control flies respectively, to 30 % and 27.5 % at 24 h. Although
783 low mortalities were observed with PBAlc, but at 1 h and 3 h exposure times the mortalities
784 in the transgenic flies expressing *CYP6Z2* were significantly lower compared with mortalities
785 from the control flies (1 h mortality = 2.5 % *vs* 12.5 %, $p < 0.001$; 3 h mortality = 9.5 % *vs* 17.5 %,
786 $p < 0.001$).

787 Marginal tolerance towards DDT was observed in the *CYP6Z2* transgenic flies when
788 compared with the control flies, but only at 3 h ($p < 0.001$) and 6 h ($p < 0.01$) (Figure 6d). This
789 is in contrast with the *GSTe2* transgenic flies, with very low mortalities, in ranges of 0 % to
790 9.52 % for 1 h to 24 h, when compared with control flies (mortality = 34 % in 1 h and 100 % in
791 6 – 24 h, $p < 0.0001$).

792 Susceptibility to clothianidin was very high (Figure 6d), but surprisingly the *CYP6Z2*
793 transgenic flies exhibited a contrasting phenotype, with significantly higher mortalities

794 compared with the control flies, at 1 h (mortality = 67 % vs 42 %, $p < 0.001$), 3 h (mortality = 71
795 % vs 53 %, $p < 0.001$) and 6 h (mortality = 78 % vs 62 %, $p < 0.01$).

796 Taken together, these results confirmed that over-expression of *GSTe2* alone is sufficient to
797 confer resistance to type I pyrethroid (permethrin) and DDT, while overexpression of *CYP6Z2*
798 alone, may confer marginal resistance to the α-cypermethrin and PBAlc.

799

800 Discussion

801 Escalating insecticide resistance across Africa (7-9, 51), if not tackled, will probably
802 compromise the malaria control and elimination efforts. Molecular markers of metabolic
803 resistance e.g., (10, 11, 21), will support evidence-based control and resistance management.
804 Identification and validation of resistance markers in malaria vectors across regions of sub-
805 Saharan Africa can promote communication, cooperation and coordination among malaria
806 control/elimination programs, and allow control efforts to be tailored to the vector species
807 involved in transmission across borders (52), and tracking of the evolution and spread of
808 resistance markers across regions (22). To support malaria pre-elimination efforts in Africa, in
809 this study we targeted the Sudan savannah and Sahel (regions of Africa sharing similar eco-
810 climatic conditions, and characterised by high seasonal transmission), which are ideal for
811 control and elimination of malaria using seasonal vector control (25) and chemoprevention.

812 *Anopheles coluzzii* is a major malaria vector across the Sudano-Sahelian transects

813 Contrary to the previous observations that *An. arabiensis* tends to predominate in arid
814 savannas, while *An. gambiae* is the dominant species in humid forest zones (53-55), *An. coluzzii*
815 has repeatedly been identified as the major malaria vector in the Sudan savannah and Sahel

816 of several, neighbouring countries (including in northern Nigeria, southern Niger, central
817 Chad, and northern Cameroon) (7, 26-28, 56) in recent years, suggesting that this vector has
818 adapted well in drier regions of the Sahel transects, and is probably predominating over *An.*
819 *arabiensis* and *An. gambiae s.s*. This is not surprising as this species exhibits higher exploitation
820 of breeding sites associated with anthropogenic activities, and behavioural plasticity to avoid
821 predators (57), and is known to survive a long dry season *in situ*/aestivation, which allows it
822 to predominate, becoming the primary force of malaria transmission (58, 59). Also,
823 photoperiod and lower nightly temperature have been shown to significantly increase the
824 longevity of the *An. coluzzii*; mechanisms which allow it to diapause in the dry season and re-
825 establish first in the early rainy season (60).

826

827 **Common metabolic genes mediate multiple resistance in the Sahelian *An. coluzzii***

828 Several genes shown to confer metabolic resistance in *Anopheles* mosquitoes and other insects
829 were found constitutively overexpressed and/or induced in this study. *GSTe2* (AGAP009194)
830 is one of the most regularly encountered metabolic genes in resistant populations of the major
831 malaria vectors *An. gambiae*, *An. coluzzii* and *An. funestus* (20-23). *Anopheles gambiae GSTe2* has
832 been validated, using transgenic flies to confer DDT (20) and fenitrothion (61) resistance. It
833 was extensively studied in *An. funestus*, in which it was shown to confer cross-resistance to
834 DDT and permethrin (22), reduce efficacy of the LLINs, PermaNet 2.0 and PermaNet 3.0 (side
835 panels) (23), and even increase the longevity of the resistant populations carrying the 119F
836 mutation (62). These and our findings suggest that the overexpression of this GST alone can
837 confer resistance to three insecticides from three different classes (DDT, permethrin and
838 fenitrothion). The absence of mutations in the cDNA coding sequences of *GSTe2* in *An. coluzzii*

839 from these four countries, suggest that overexpression of this GST is enough to confer
840 resistance.

841 Several cytochrome P450s previously linked with insecticide resistance were found
842 overexpressed across the Sahel. For example, *An. gambiae CYP6Z2* (AGAP008218) known to
843 metabolize carbaryl (63), the insect juvenile hormone analogue insecticide, pyriproxyfen (64)
844 and mitochondrial complex I inhibitors, fenazaquin, pyridaben and tolfenpyrad (65). This
845 P450 also plays a pivotal role in the clearance of pyrethroid insecticides via further catabolism
846 of pyrethroid derivatives (PBAld and PBAlc) obtained by the action of carboxylesterases (66),
847 in line with our findings of this gene conferring marginal tolerance to high concentration of
848 PBAlc, and α-cypermethrin. However, our findings suggest that overexpression of this P450
849 may enhance the efficacy of clothianidin, which will be epidemiologically advantageous in
850 terms of control. Indeed, bioactivation by P450s is known to be a requirement for insecticidal
851 toxicity of several classes of insecticides, e.g., the organophosphates and chlorpenafyr. In
852 contrast to *GSTe2*, the *CYP6Z2* from across Sahel contain three cDNA mutations, which makes
853 it different from Ngoussou coding sequences. These are $K^{211}N$ and $T^{218}S$ mutations, both
854 within the substrate recognitions site 2, and an $A^{282}E$. Other important CYP450s found to be
855 commonly overexpressed across the Sahel include the *CYP4C27* (AGAP009246), *CYP6Z3*
856 (AGAP008217) and *CYP9K1* (AGAP000818), all three shown to be consistently overexpressed
857 in field populations of *An. gambiae/coluzzii* and *An. funestus* across Africa (9, 67, 68). *CYP9K1*
858 has been shown to be epidemiologically important pyrethroid-metabolising P450 linked with
859 metabolism of deltamethrin and pyriproxyfen in *An. gambiae* (69). We also have recently
860 shown that his P450 is involved in pyrethroids in *An. funestus* (70).

861 Several carboxylesterases were also upregulated/induced across the Sahel, with the *COEBE3C*
862 (AGAP005372) upregulated in all four countries. This beta esterase is enriched in the legs
863 (where xenobiotic detoxification probably occur) of pyrethroid-resistant *An. coluzzii* (71).
864 Other genes consistently overexpressed across the Sahel, include the chymotrypsins
865 (*CHYM3*/AGAP006711 and *CHYM1*/AGAP006709) and a lipase (AGAP002353).
866 Chymotrypsins are known to defend insects against plants' proteinase inhibitors (72); and
867 previous transcriptional studies have shown *CHYM1* and *CHYM3* overexpressed in
868 insecticide resistant populations of *An. gambiae* and *An. coluzzii,* respectively (9, 67). Using *in*
869 *vitro* and *in vivo* tools, lipases have been linked with deltamethrin resistance in *Culex pipiens*
870 pallens (73).

871 **Common metabolic resistance markers probably exacerbate resistance across the**
872 **Sahel**

873 Several well-known genetic variants implicated in resistance, as well as the recently
874 discovered ones exist in high frequencies in the Sahelian *An. coluzzii*. For example, the
875 pyrethroid resistance *CYP4J5*-L43F marker (74) was found fixed across the Sahel. The
876 G280S/G119S *ace*-1 mutation, found in high frequencies across Sahel confers organophosphate
877 an carbamates resistance (75), and is shown recently to confer resistance to pirimiphos-methyl,
878 in *An. coluzzii/gambiae* (76). Several mutations found within the VGSC have recently been
879 described/validated. For example, the resistance mutation, L995F (77) and the V402L/I15227T
880 haplotypes have been observed across Africa (78). Recently, the two mutations
881 (V402L/I15227T) are described to be in tight linkage and mutually exclusive to the classical
882 L995F/S mutations (77). Our results suggest haplotypes carrying the V402L/I1527T

883 combination plus the L995 replacement do exist in the Sahel *An. coluzzii*. Not only that, it is
884 also in addition to the N1570Y replacement.

885 In contrast to *An. funestus GSTe2*, where overexpression and 119F mutation combined to
886 confer extreme DDT resistance (22), the absence of amino acid replacements in the *An. coluzzii*
887 *GSTe2* from the Sahel suggests that overexpression alone is the key mediator of DDT and
888 permethrin resistance. This is supported by the higher activity in the regulatory regions,
889 harbouring an insertion and nucleotide substitutions in the alive mosquitoes. Indeed, some of
890 the mutations we have found within the intergenic region of *GSTe2* are similar to those
891 observed in a previous study (79).

892

893 **Cuticular resistance mechanism probably playing a key role in Sahelian *An.***
894 ***coluzzii***

895 Our results suggest cuticular mechanism play a role in pyrethroid resistance in these
896 populations. For example, the findings of *CYP4G16* (AGAP001076) and *CYP4G17*
897 (AGAP000877) overexpressed across the Sahel. The former P450 was previously shown to be
898 involved in epicuticular hydrocarbon biosynthesis associated with resistance (80).

899 The three major classes of the insect cuticular proteins - the CPR, CPLC and CPAP, were found
900 overrepresented in the top overexpressed genes across the Sahel, with the commonly
901 upregulated ones being *CPR76*, *CPR15* and *CPR30*, a chitin-binding cuticular protein, *CPAP3-*
902 *A1b*, and cuticular proteins of low complexity, *CPLCX3* and *CPLCA1*. Indeed, *CPAP3-A1b*
903 (AGAP000987) have been shown to be highly overexpressed in deltamethrin-resistant
904 Sahelian population of *An. coluzzii* from Burkina Faso (9) and induced by blood feeding in *An.*

905 *gambiae* (81). The CPR and CPLC cuticular proteins have been described to potentially play a
906 crucial role in insecticide resistance through leg cuticle remodelling/thickening, regulating
907 penetration rate of insecticides in *An. coluzzii* (71). Furthermore, a recent study has found most
908 of the cuticular proteins we have described here, as highly overexpressed in
909 permethrin/malathion resistant populations of Ethiopian *An. arabiensis* (82), e.g., *CPR30*,
910 *CPR75*, *CPR81*, and *CPLCP11*. Out of the several chitinases overexpressed across the Sahel,
911 four were amongst the top 50 most overexpressed metabolic genes. These include the *Cht24*
912 and *Cht6* that have been shown to be overexpressed in the *An. arabiensis* from the above study
913 (82), with the ortholog of *Cht24*, AARA007329 among the top 10 most overexpressed genes in
914 *An. arabiensis*, in line with our observations in *An. coluzzii* across the Sahel. There is an
915 overwhelming need to functionally investigate the role/contribution towards insecticide
916 resistance of these cuticular proteins, chitinases and a chitin synthase (AGAP001748)
917 significantly overexpressed in the field *An. coluzzii* from Nigeria and Chad.

918

919 **Insecticide resistance- and thermotolerance-associated genes sit within**
920 **chromosomal inversions**

921 In this study, the findings of high frequency of 2La, 2Rb and 2Rc inversion polymorphisms in
922 the populations of *An. coluzzii*, compared with the Ngoussou, suggested strong phenotypic
923 adaptations in this species, across the Sahel. Most importantly, in addition to several of the
924 cuticular protein genes associated with resistance (chitinases, chitin synthase, CPR, CPLC and
925 CPAP proteins), several other genes previously implicated in thermotolerance and/or
926 desiccation resistance in *An. gambiae/coluzzii*, and which were highly overexpressed in this
927 study sit within these inversions. For example, the heat shock proteins, *hsp83* (AGAP006958)

928 and *hsp90* hptG (AGAP006961), both of which are known to be heat- and insecticide-stress
929 inducible (83) and were among the core set of hsp genes involved in a common and immediate
930 response to thermal stress in *An. gambiae* populations (84), sit within the 2La inversion. These
931 two genes were among the overexpressed genes in both heat-hardened and permethrin-
932 resistant *An. coluzzii* populations from northern Nigeria (56). Several ionotropic glutamate
933 receptors were found within the 2La inversion breakpoints: *IR136* (AGAP006440), *IR139*
934 (AGAP006691) and *IR142* (AGAP006407). This is not surprising as ionotropic receptors are
935 commonly associated with chemosensation, thermosensation, and hygrosensation (85, 86),
936 characteristics
937 which can confer adaptive advantages in xeric environs. The *IR25a* (AGAP010272) and *IR21*
938 *a* (AGAP008511) are known to mediate both humidity and temperature preference in the fruit
939 fly, *D. melanogaster* (86, 87), in addition to *IR21a* driving heat seeking and heat-stimulated
940 blood feeding in *An. gambiae* (87). These two genes have been shown to be
941 overexpressed/induced in thermotolerant/permethrin-resistant populations of *An. coluzzii*
942 (56).

943

944 ## Conclusions

945 Information on molecular basis of resistance and/or resistance genes and its markers facilitates
946 evidence-based control measures. In this study we characterised a major malaria vector, *An.*
947 *coluzzii* from the Sahel region of four countries, with findings which could promote evidence-
948 based, cross-border policy towards local and regional malaria control. The study found that
949 across Sahel (where malaria is highly seasonal, reaching its peak in the rainy season), *An.*

950 *coluzzii* is a dominant vector. And that a handful of common cross-resistance genes are
951 responsible for multiple insecticide resistance in this species. Findings from this study suggest
952 pleotropic role of some key genes – able to confer insecticide resistance and/or stabilize the
953 insecticide resistance gene, at the same time conferring environmental adaptations, such as
954 the ability to survive thermal stress (thermotolerance), as expected in this Sahelian region.
955 From operational vector control perspective this study provided evidence of the role of key
956 insecticide metabolism gene, *CYP6Z2* in increasing insecticidal potency of clothianidin, which
957 could increase the efficacy of the ingredients in malaria control tools, when targeting field
958 populations overexpressing this key P450.

959

960 # Declarations

961 ## Ethics approval and consent to participate

962 This study did not use human participants, human data or human tissue. Ethical approval for
963 collection of indoor resting female mosquitoes in Nigeria, Niger, Chad and Cameroon have
964 been provided in the references cited in the Study Site and Mosquito Sampling (Materials and
965 Methods).

966 ## Consent for publication

967 Not Applicable.

968

969

970

971 **Availability of data and materials**

972 The dataset(s) supporting the conclusions of this article are available in the European
973 Nucleotide Archive, with accession PRJEB51644, and secondary accession of ERP136291, for
974 the RNA-seqe raw sequence reads. cDNA sequences of *GSTe2* and *CYP6Z2* were deposited in
975 GenBank (accession numbers: Submission # 2567942 for *GSTe2* and Submission # 2567961 for
976 *CYP6Z2*) and 5′-UTR DNA fragment sequences were deposited in GenBank (accession
977 numbers: Submission # 2568060).

978

979 **Competing interests**

980 The authors declare that they have no competing interests.

981

982 **Funding**

983 This research was funded in whole, by the Wellcome Trust [Grant numbers:
984 WT201918/Z/16/Z to SSI and WT217188/Z/19/Z to CSW]. For the purpose of open access, the
985 author has applied a CC BY public copyright licence to any Author Accepted Manuscript
986 version arising from this submission. The Wellcome Trust had no role in the design of this
987 study and collection, analysis, and interpretation of data and in writing of this manuscript.

988

989

990

991 **Authors' contributions**

992 Conceived and designed by SSI and CSW. SSI carried out the molecular analyses, with
993 support from AM, LMJM, EIP and HI. ANF and MMM, participated in field collection of
994 mosquitoes in Nigeria, Chad and Cameroon. SSI carried out data analysis with the support of
995 JH, GDW, SCN for the RNAseq component. SSI wrote the manuscript with inputs from CSW,
996 JH and SCN. All authors contributed to corrections of the final draft and approved final
997 version of the manuscript.

998

999 **Acknowledgements**

1000 Not Applicable

1001

1002

1003

1004 **Reference**

1005 1. Bhatt S, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria
1006 control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207.
1007 2. WHO. Global Technical Strategy for Malaria 2016-2030. World Health Organisation. .
1008 Geneva, Switzerland; 2015 May 2015. Report No.: ISBN 978 92 4 156499 1.
1009 3. WHO. World Malaria Report. World Health Organization. . Geneva Switzerland; 2018. Report
1010 No.: Licence: CC BY-NC-SA 3.0 IGO Contract No.: ISBN 978-92-4-156565-3.
1011 4. WHO. World Malaria Report. Geneva: Switzerland; 2021. Report No.: 978-92-4-004049-6
1012 Contract No.: 978-92-4-004049-6
1013 5. WHO. World Malaria Report 2019. Geneva, Switzerland; 2019. Report No.: 978-92-4-
1014 156572-1.
1015 6. WHO. Update on the E-2020 initiative of 21 malaria-eliminating countries: report and
1016 country briefs. Geneva. Switzerland: World Health Organization; 2018. Report No.:
1017 WHO/CDS/GMP/2018.10.

1018 7. Ibrahim SS, Mukhtar MM, Datti JA, Irving H, Kusimo MO, Tchapga W, et al. Temporal
1019 escalation of Pyrethroid Resistance in the major malaria vector Anopheles coluzzii from Sahelo-
1020 Sudanian Region of northern Nigeria. Scientific reports. 2019;9(1):7395.
1021 8. Riveron JM, Huijben S, Tchapga W, Tchouakui M, Wondji MJ, Tchoupo M, et al. Escalation of
1022 Pyrethroid Resistance in the Malaria Vector Anopheles funestus Induces a Loss of Efficacy of
1023 Piperonyl Butoxide-Based Insecticide-Treated Nets in Mozambique. J Infect Dis. 2019;220(3):467-75.
1024 9. Toe KH, N'Fale S, Dabire RK, Ranson H, Jones CM. The recent escalation in strength of
1025 pyrethroid resistance in Anopheles coluzzi in West Africa is linked to increased expression of
1026 multiple gene families. BMC Genomics. 2015;16:146.
1027 10. Weedall GD, Mugenzi LMJ, Menze BD, Tchouakui M, Ibrahim SS, Amvongo-Adjia N, et al. A
1028 cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing
1029 insecticide-treated bednet efficacy. Sci Transl Med. 2019;11(484).
1030 11. Mugenzi LMJ, Menze BD, Tchouakui M, Wondji MJ, Irving H, Tchoupo M, et al. Cis-regulatory
1031 CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles
1032 funestus. Nat Commun. 2019;10(1):4652.
1033 12. Barnes KG, Weedall GD, Ndula M, Irving H, Mzihalowa T, Hemingway J, et al. Genomic
1034 Footprints of Selective Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors
1035 Are Driven by Scale up of Insecticide-Based Vector Control. PLoS Genet. 2017;13(2):e1006539.
1036 13. Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, et al. An Africa-wide genomic
1037 evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective
1038 sweeps, copy number variations, gene conversion and transposons. PLoS Genet.
1039 2020;16(6):e1008822.
1040 14. Bruce-Chwatt LJ. Malaria in Nigeria. Bulletin of the World Health Organization.
1041 1951;4(3):301-27.
1042 15. Anopheles gambiae Genomes C, Data analysis g, Partner working g, Sample c-A, Burkina F,
1043 Cameroon, et al. Genetic diversity of the African malaria vector Anopheles gambiae. Nature.
1044 2017;552(7683):96-100.
1045 16. Muller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-caught
1046 permethrin-resistant Anopheles gambiae overexpress CYP6P3, a P450 that metabolises pyrethroids.
1047 PLoS Genet. 2008;4(11):e1000286.
1048 17. Yunta C, Hemmings K, Stevenson B, Koekemoer LL, Matambo T, Pignatelli P, et al. Cross-
1049 resistance profiles of malaria mosquito P450s associated with pyrethroid resistance against WHO
1050 insecticides. Pestic Biochem Physiol. 2019;161:61-7.
1051 18. Stevenson BJ, Bibby J, Pignatelli P, Muangnoicharoen S, O'Neill PM, Lian LY, et al.
1052 Cytochrome P450 6M2 from the malaria vector Anopheles gambiae metabolizes pyrethroids:
1053 Sequential metabolism of deltamethrin revealed. Insect biochemistry and molecular biology.
1054 2011;41(7):492-502.
1055 19. Mitchell SN, Stevenson BJ, Müller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification
1056 and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae
1057 from Ghana. Proceedings of the National Academy of Sciences. 2012;109(16):6147-52.
1058 20. Mitchell SN, Rigden DJ, Dowd AJ, Lu F, Wilding CS, Weetman D, et al. Metabolic and target-
1059 site mechanisms combine to confer strong DDT resistance in Anopheles gambiae. PLoS One.
1060 2014;9(3):e92662.
1061 21. Lucas ER, Rockett KA, Lynd A, Essandoh J, Grisales N, Kemei B, et al. A high throughput multi-
1062 locus insecticide resistance marker panel for tracking resistance emergence and spread in Anopheles
1063 gambiae. Sci Rep. 2019;9(1):13335.
1064 22. Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in
1065 the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria
1066 vector. Genome Biol. 2014;15(2):R27.
1067 23. Menze BD, Kouamo MF, Wondji MJ, Tchapga W, Tchoupo M, Kusimo MO, et al. An
1068 Experimental Hut Evaluation of PBO-Based and Pyrethroid-Only Nets against the Malaria Vector

1069 Anopheles funestus Reveals a Loss of Bed Nets Efficacy Associated with GSTe2 Metabolic Resistance.
1070 Genes (Basel). 2020;11(2).
1071 24. Coluzzi M, Sabatini A, Petrarca V, Di Deco MA. Chromosomal differentiation and adaptation
1072 to human environments in the Anopheles gambiae complex. Trans R Soc Trop Med Hyg.
1073 1979;73(5):483-97.
1074 25. Cairns M, Roca-Feltrer A, Garske T, Wilson AL, Diallo D, Milligan PJ, et al. Estimating the
1075 potential public health impact of seasonal malaria chemoprevention in African children. Nat
1076 Commun. 2012;3:881.
1077 26. Ibrahim SS, Mukhtar MM, Irving H, Labbo R, Kusimo MO, Mahamadou I, et al. High
1078 Plasmodium infection and multiple insecticide resistance in a major malaria vector Anopheles
1079 coluzzii from Sahel of Niger Republic. Malaria journal. 2019;18(1):181.
1080 27. Fadel AN, Ibrahim SS, Tchouakui M, Terence E, Wondji MJ, Tchoupo M, et al. A combination
1081 of metabolic resistance and high frequency of the 1014F kdr mutation is driving pyrethroid
1082 resistance in Anopheles coluzzii population from Guinea savanna of Cameroon. Parasit Vectors.
1083 2019;12(1):263.
1084 28. Ibrahim SS, Fadel AN, Tchouakui M, Terence E, Wondji MJ, Tchoupo M, et al. High insecticide
1085 resistance in the major malaria vector Anopheles coluzzii in Chad Republic. Infect Dis Poverty.
1086 2019;8(1):100.
1087 29. Mitchell SN, Stevenson BJ, Muller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification
1088 and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae
1089 from Ghana. Proceedings of the National Academy of Sciences of the United States of America.
1090 2012;109(16):6147-52.
1091 30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
1092 EMBnet journal. 2011;17(1):10-2.
1093 31. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files
1094 (Version 1.33)[Software]. 2011.
1095 32. Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. Durham,
1096 NC; 2011.
1097 33. Love MI, Huber, W., & Anders, S. Moderated estimation of fold change and dispersion for
1098 RNA-seq data with DESeq2. Genome Biology. 2014;15(550).
1099 34. Blighe K, S. Rana, and M. Lewis. . “EnhancedVolcano: Publication-ready volcano plots with
1100 enhanced colouring and labeling.". 2018.
1101 35. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in
1102 multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.
1103 36. Alexa A, Rahnenfuhrer J. topGO: Enrichment Analysis for Gene Ontology. R package version
1104 2.46.0. 2021 [
1105 37. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene
1106 ontology terms. PLoS One. 2011;6(7):e21800.
1107 38. Riveron JM, Ibrahim SS, Chanda E, Mzilahowa T, Cuamba N, Irving H, et al. The highly
1108 polymorphic CYP6M7 cytochrome P450 gene partners with the directionally selected CYP6P9a and
1109 CYP6P9b genes to expand the pyrethroid resistance front in the malaria vector Anopheles funestus
1110 in Africa. BMC Genomics. 2014;15(1):817.
1111 39. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method.
1112 Nature protocols. 2008;3(6):1101-8.
1113 40. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence
1114 data. Genetics. 1992;132(2):583-9.
1115 41. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes
1116 reveals adaptation to high altitude. Science. 2010;329(5987):75-8.
1117 42. Miles A, Rodrigues MF, Ralph P, Harding N, Pisupati R., Rae S. cggh/scikit-allel: A Python
1118 package for exploring and analysing genetic variation data. 2020:cggh/scikit-allel: v1.3.1 (v.3.1).
1119 Zenodo.

1120 43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1121 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
1122 44. Love RR, Redmond SN, Pombi M, Caputo B, Petrarca V, Della Torre A, et al. In Silico
1123 Karyotyping of Chromosomally Polymorphic Malaria Mosquitoes in the Anopheles gambiae Complex.
1124 G3 (Bethesda). 2019;9(10):3249-62.
1125 45. Hall TA, editor BioEdit: a user-friendly biological sequence alignment editor and analysis
1126 program for Windows 95/98/NT. Nucleic acids symposium series; 1999.
1127 46. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics
1128 Analysis across Computing Platforms. Molecular biology and evolution. 2018;35(6):1547-9.
1129 47. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et
1130 al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular biology and
1131 evolution. 2017;34(12):3299-302.
1132 48. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and
1133 beyond: promoter analysis based on transcription factor binding sites. Bioinformatics.
1134 2005;21(13):2933-42.
1135 49. Lynd A, Lycett GJ. Optimization of the Gal4-UAS system in an Anopheles gambiae cell line.
1136 Insect Mol Biol. 2011;20(5):599-608.
1137 50. Ibrahim SS, Riveron JM, Bibby J, Irving H, Yunta C, Paine MJI, et al. Allelic Variation of
1138 Cytochrome P450s Drives Resistance to Bednet Insecticides in a Major Malaria Vector. PLoS Genet.
1139 2015;11(10):e1005618.
1140 51. Riveron JM, Tchouakui M., Mugenzi L., Menze B. D., Mu-Chun C. and Wondji C. S. Insecticide
1141 Resistance in Malaria Vectors: An Update at a Global Scale. In: S. Manguin VD, editor. Towards
1142 Malaria Elimination - A Leap Forward. London: IntechOpen; 2018. p. 452.
1143 52. Shiff C. Integrated approach to malaria control. Clinical microbiology reviews.
1144 2002;15(2):278-93.
1145 53. Lindsay S, Parson, L. and Thomas, C. . Mapping the range and relative abundance of the two
1146 principal African malaria vectors, Anopheles gambiae sensu stricto and An. arabiensis, using climate
1147 data. Proceedings of the Royal Society of London B: Biological Science. 1998; 265((1399)): 847–54.
1148 54. Coetzee M, Craig, M. and Le Sueur, D. . Distribution of African malaria mosquitoes belonging
1149 to the Anopheles gambiae complex. Parasitology today. 2000;16(2):74-7.
1150 55. Ranson H, Abdallah H, Badolo A, Guelbeogo WM, Kerah-Hinzoumbe C, Yangalbe-Kalnone E,
1151 et al. Insecticide resistance in Anopheles gambiae: data from the first year of a multi-country study
1152 highlight the extent of the problem. Malar J. 2009;8:299.
1153 56. Ibrahim SS, Mukhtar MM, Muhammad A, Wondji CS. 2La Paracentric Chromosomal Inversion
1154 and Overexpressed Metabolic Genes Enhance Thermotolerance and Pyrethroid Resistance in the
1155 Major Malaria Vector Anopheles gambiae. Biology (Basel). 2021;10(6).
1156 57. Gimonneau G, Bouyer J, Morand S, Besansky NJ, Diabate A, Simard F. A behavioral
1157 mechanism underlying ecological divergence in the malaria mosquito Anopheles gambiae. Behav
1158 Ecol. 2010;21(5):1087-92.
1159 58. Dao A, Yaro AS, Diallo M, Timbine S, Huestis DL, Kassogue Y, et al. Signatures of aestivation
1160 and migration in Sahelian malaria mosquito populations. Nature. 2014;516(7531):387-90.
1161 59. Lehmann T, Weetman D, Huestis DL, Yaro AS, Kassogue Y, Diallo M, et al. Tracing the origin
1162 of the early wet-season Anopheles coluzzii in the Sahel. Evol Appl. 2017;10(7):704-17.
1163 60. Huestis DL, Artis ML, Armbruster PA, Lehmann T. Photoperiodic responses of Sahelian
1164 malaria mosquitoes Anopheles coluzzii and An. arabiensis. Parasit Vectors. 2017;10(1):621.
1165 61. Adolfi A, Poulton B, Anthousi A, Macilwee S, Ranson H, Lycett GJ. Functional genetic
1166 validation of key genes conferring insecticide resistance in the major African malaria vector,
1167 Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of
1168 America. 2019;116(51):25764-72.

1169 62. Tchouakui M, Riveron JM, Djonabaye D, Tchapga W, Irving H, Soh Takam P, et al. Fitness
1170 Costs of the Glutathione S-Transferase Epsilon 2 (L119F-GSTe2) Mediated Metabolic Resistance to
1171 Insecticides in the Major African Malaria Vector Anopheles Funestus. Genes (Basel). 2018;9(12).
1172 63. Chiu TL, Wen Z, Rupasinghe SG, Schuler MA. Comparative molecular modeling of Anopheles
1173 gambiae CYP6Z1, a mosquito P450 capable of metabolizing DDT. Proceedings of the National
1174 Academy of Sciences of the United States of America. 2008;105(26):8855-60.
1175 64. Yunta C, Grisales N, Nasz S, Hemmings K, Pignatelli P, Voice M, et al. Pyriproxyfen is
1176 metabolized by P450s associated with pyrethroid resistance in An. gambiae. Insect biochemistry and
1177 molecular biology. 2016;78:50-7.
1178 65. Lees RS, Ismail HM, Logan RAE, Malone D, Davies R, Anthousi A, et al. New insecticide
1179 screening platforms indicate that Mitochondrial Complex I inhibitors are susceptible to cross-
1180 resistance by mosquito P450s that metabolise pyrethroids. Sci Rep. 2020;10(1):16232.
1181 66. Chandor-Proust A, Bibby J, Regent-Kloeckner M, Roux J, Guittard-Crilat E, Poupardin R, et al.
1182 The central role of mosquito cytochrome P450 CYP6Zs in insecticide detoxification revealed by
1183 functional expression and structural modelling. Biochem J. 2013;455(1):75-85.
1184 67. Antonio-Nkondjio C, Poupardin R, Tene BF, Kopya E, Costantini C, Awono-Ambene P, et al.
1185 Investigation of mechanisms of bendiocarb resistance in Anopheles gambiae populations from the
1186 city of Yaounde, Cameroon. Malar J. 2016;15(1):424.
1187 68. Riveron JM, Ibrahim SS, Mulamba C, Djouaka R, Irving H, Wondji MJ, et al. Genome-Wide
1188 Transcription and Functional Analyses Reveal Heterogeneous Molecular Mechanisms Driving
1189 Pyrethroids Resistance in the Major Malaria Vector Anopheles funestus Across Africa. G3 (Bethesda).
1190 2017;7(6):1819-32.
1191 69. Vontas J, Grigoraki L, Morgan J, Tsakireli D, Fuseini G, Segura L, et al. Rapid selection of a
1192 pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. Proceedings of the
1193 National Academy of Sciences of the United States of America. 2018;115(18):4619-24.
1194 70. Hearn J, Tagne CD, Ibrahim SS, Tene-Fossog B, Mugenzi LJ, Irving H, et al. Multi-omics
1195 analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector
1196 Anopheles funestus in East Africa. bioRxiv 2022;2021.10.21.465247.
1197 71. Kefi M, Charamis J, Balabanidou V, Ioannidis P, Ranson H, Ingham VA, et al. Transcriptomic
1198 analysis of resistance and short-term induction response to pyrethroids, in Anopheles coluzzii legs.
1199 BMC Genomics. 2021;22(1):891.
1200 72. Dunse KM, Kaas, Q., Guarino, R.F., Barton, P.A., Craik, D.J. and Anderson, M.A. Molecular
1201 basis for the resistance of an insect chymotrypsin to a potato type II proteinase inhibitor.
1202 Proceedings of the National Academy of Sciences. 2010;107(34):15016-21.
1203 73. Hu HX, Zhou D, Ma L, Shen B, Sun Y, Zhu CL. Lipase is associated with deltamethrin
1204 resistance in Culex pipiens pallens. Parasitol Res. 2020;119(1):23-30.
1205 74. Weetman D, Wilding CS, Neafsey DE, Muller P, Ochomo E, Isaacs AT, et al. Candidate-gene
1206 based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing
1207 genetic variation in East African Anopheles gambiae. Sci Rep. 2018;8(1):2920.
1208 75. Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique
1209 mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. Insect
1210 Mol Biol. 2004;13(1):1-7.
1211 76. Grau-Bove X, Lucas E, Pipini D, Rippon E, van 't Hof AE, Constant E, et al. Resistance to
1212 pirimiphos-methyl in West African Anopheles is spreading via duplication and introgression of the
1213 Ace1 locus. PLoS Genet. 2021;17(1):e1009253.
1214 77. Williams J, Cowlishaw R, Sanou A, Ranson H, Grigoraki L. In vivo functional validation of the
1215 V402L voltage gated sodium channel mutation in the malaria vector An. gambiae. Pest Manag Sci.
1216 2022;78(3):1155-63.
1217 78. Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, et al. The genetic
1218 architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors
1219 Anopheles gambiae and Anopheles coluzzii. Mol Ecol. 2021;30(21):5303-17.

1220 79. Ding Y, Hawkes N, Meredith J, Eggleston P, Hemingway J, Ranson H. Characterization of the
1221 promoters of Epsilon glutathione transferases in the mosquito Anopheles gambiae and their
1222 response to oxidative stress. Biochem J. 2005;387(Pt 3):879-88.
1223 80. Balabanidou V, Kampouraki A, MacLean M, Blomquist GJ, Tittiger C, Juarez MP, et al.
1224 Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production
1225 in Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of
1226 America. 2016;113(33):9268-73.
1227 81. Vannini L, Augustine Dunn W, Reed TW, Willis JH. Changes in transcript abundance for
1228 cuticular proteins and other genes three hours after a blood meal in Anopheles gambiae. Insect
1229 biochemistry and molecular biology. 2014;44:33-43.
1230 82. Messenger LA, Impoinvil LM, Derilus D, Yewhalaw D, Irish S, Lenhart A. A whole
1231 transcriptomic approach provides novel insights into the molecular basis of organophosphate and
1232 pyrethroid resistance in Anopheles arabiensis from Ethiopia. Insect biochemistry and molecular
1233 biology. 2021;139:103655.
1234 83. Zhao LaJ, W.A. . Expression of heat shock protein genes in insect stress responses. .
1235 Invertebrate Survival Journal. 2012;9(1):93-101.
1236 84. Cassone BJ, Molloy MJ, Cheng C, Tan JC, Hahn MW, Besansky NJ. Divergent transcriptional
1237 response to thermal stress by Anopheles gambiae larvae carrying alternative arrangements of
1238 inversion 2La. Mol Ecol. 2011;20(12):2567-80.
1239 85. VanHook AM. Understanding hygrosensation: How flies sense changes in humidity. Science
1240 Signaling. 2016;9(430):ec127-ec.
1241 86. Enjin A, Zaharieva EE, Frank DD, Mansourian S, Suh GS, Gallio M, et al. Humidity Sensing in
1242 Drosophila. Curr Biol. 2016;26(10):1352-8.
1243 87. Greppi C, Laursen WJ, Budelli G, Chang EC, Daniels AM, van Giesen L, et al. Mosquito heat
1244 seeking is driven by an ancestral cooling receptor. Science. 2020;367(6478):681-4.

1245

1246 **Figure Titles**

1247 **Figure 1: A volcano plot of differentially expressed genes showing fold changes and**
1248 **significance in R-S, C-S and R-C comparisons, for Nigeria and Niger populations of *An.***
1249 ***coluzzii*.**

1250 **Figure 2: A volcano plot of differentially expressed genes showing fold changes and**
1251 **significance in R-S, C-S and R-C comparisons, for Chad and Cameroon populations of *An.***
1252 ***coluzzii*.**

1253 **Figure 3: Validation of candidate resistance genes.** qRT-PCR of twelve metabolic resistance
1254 genes (R-S and C-S comparisons) from Sahelian *An. coluzzii*.

1255 **Figure 4: Identification of resistance variants of interest.** A heatmap showing frequencies of
1256 the resistance variants (haplotypes) in key genes of interest.

1257 **Figure 5: Characterization of intergenic region (5′ regulatory element) of *GSTe2*.** Results of
1258 dual-luciferase reporter assays of the promoter (intergenic region) constructs, showing
1259 progressive loss of activity following sequential deletion of the constructs.

1260 **Figure 6: Validation of the role of metabolic resistance genes in insecticide resistance.**
1261 Results of insecticide susceptibility bioassays with transgenic flies expressing *GSTe2* and
1262 *CYP6Z2*. **a.** α-cypermethrin and deltamethrin, **b.** permethrin, **c.** 3-phenoxybenzaldehyde and
1263 3-phenoxybenzyl alcohol, **d.** DDT and clothianidin.

1264

1265 **Additional Files**

1266 **Additional File 1**: Supplementary text for methods and results.

1267 **Additional File 2:** Common differentially expressed genes across the Sahel.

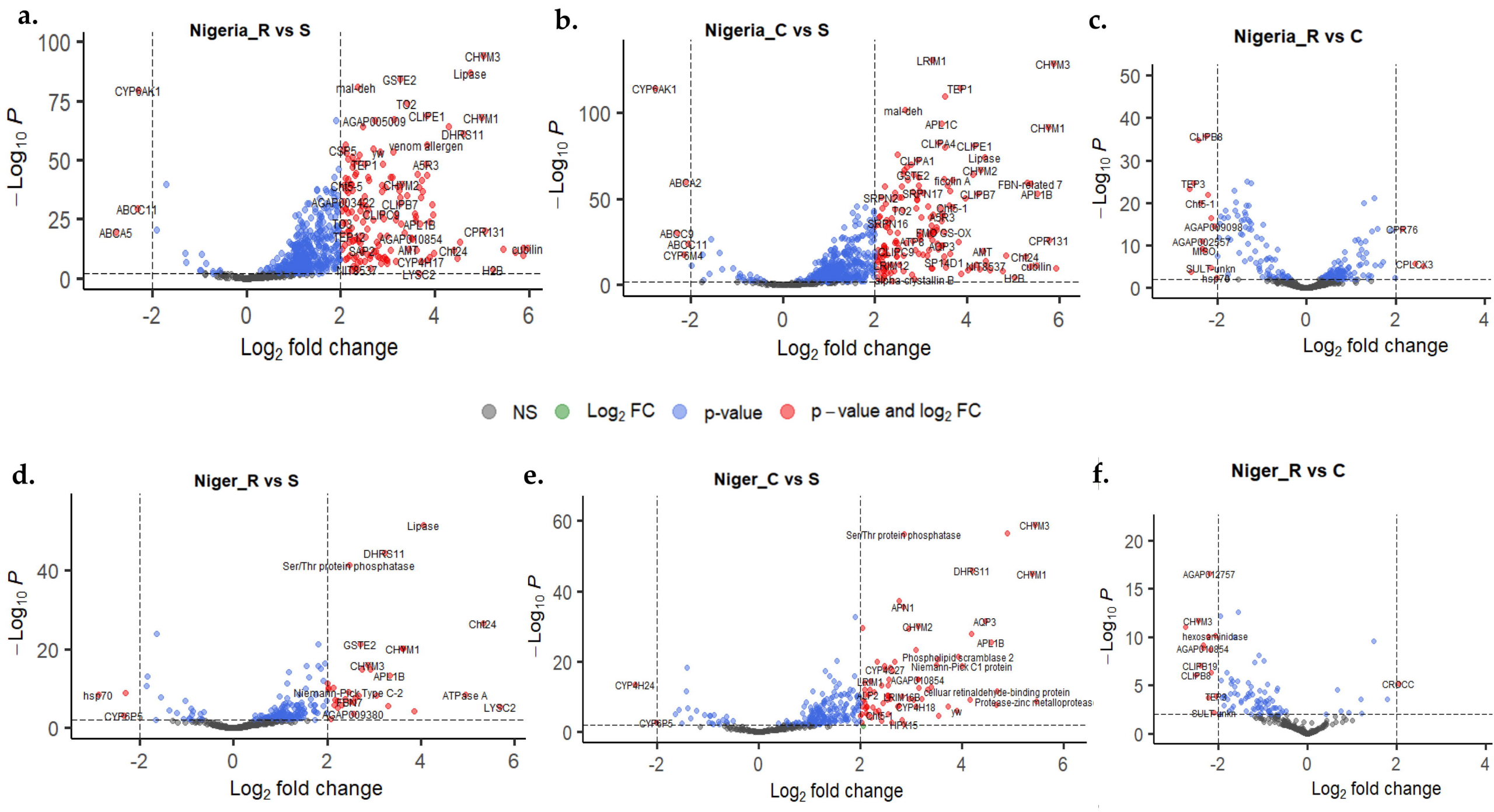
1268 **Additional File 3:** Population genetics analyses.

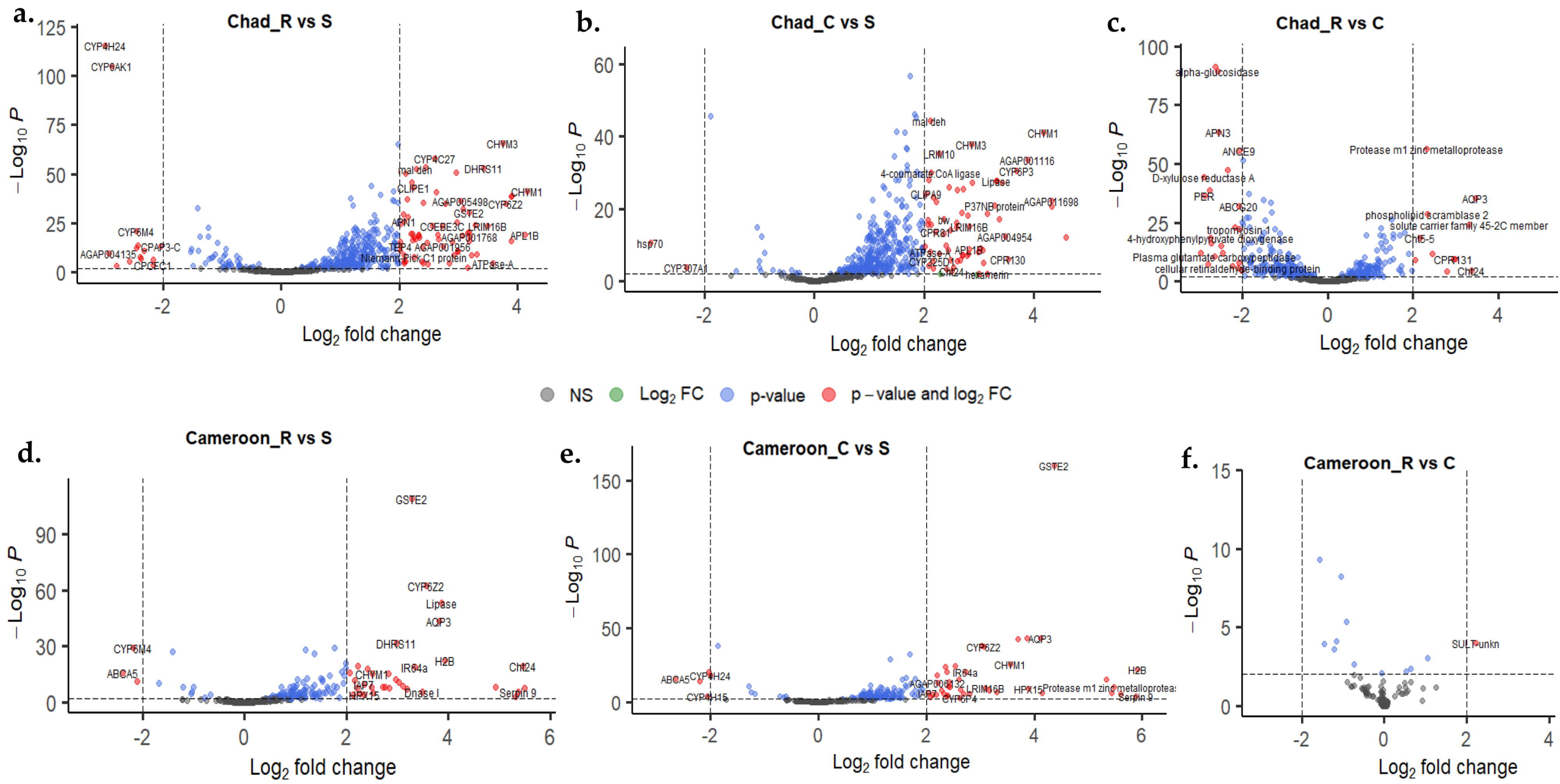
1269 **Additional File 4:** Genes within 2La, 2Rb and 2Rc inversion polymorphisms.

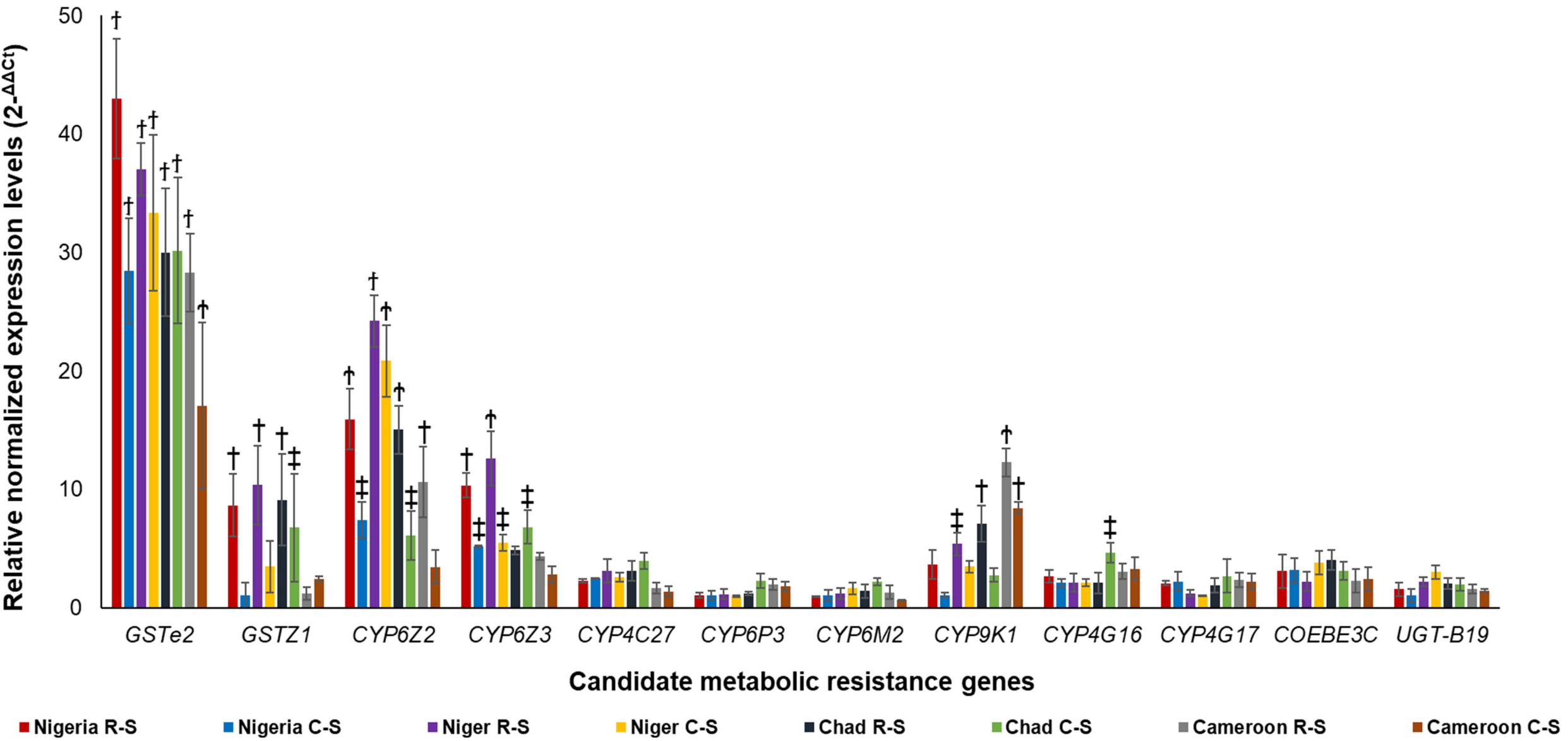
1270 **Additional Tables:** Additional tables, S1-S6.

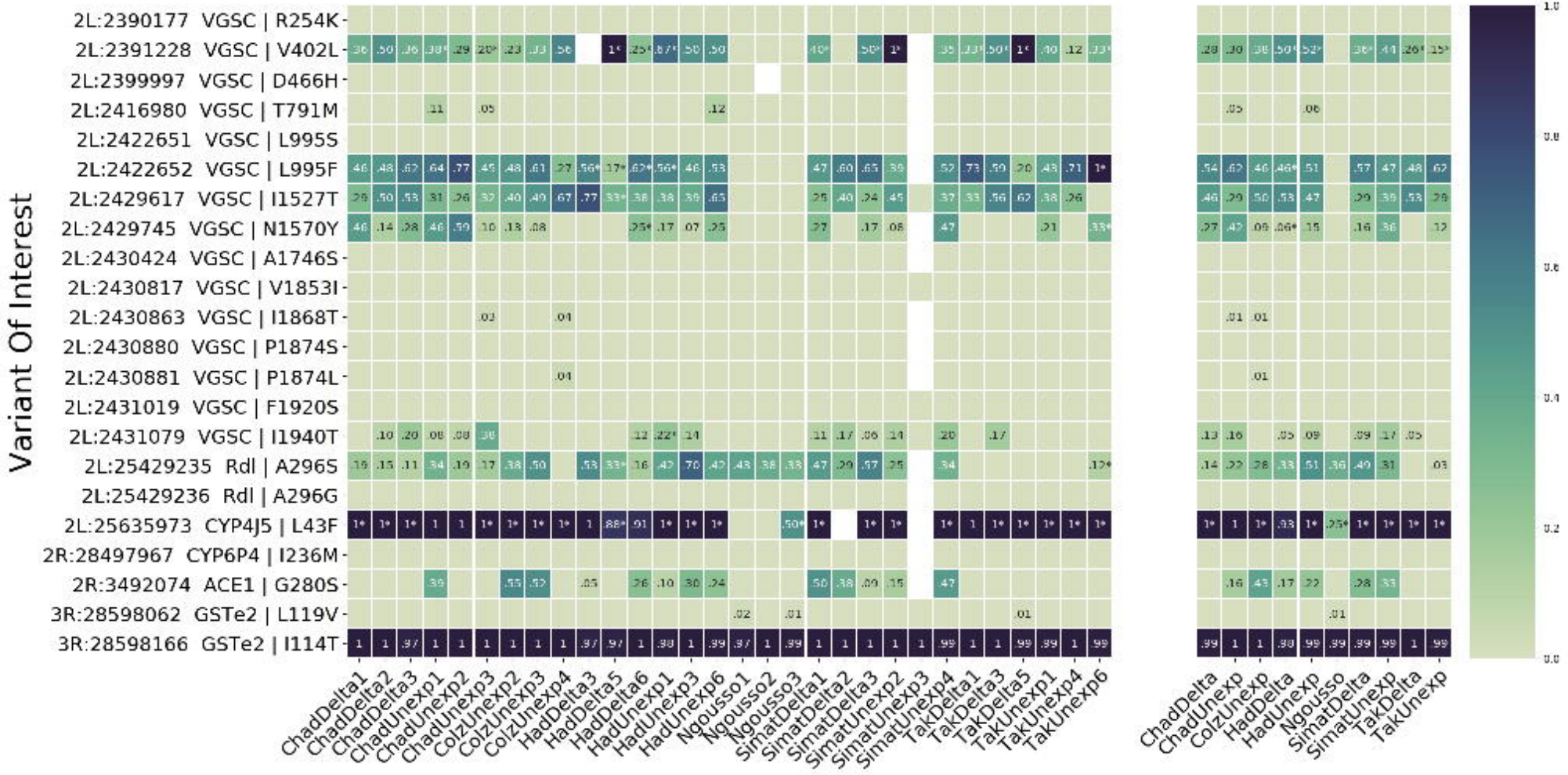
1271 **Additional Figures:** Additional figures, S1-S11.

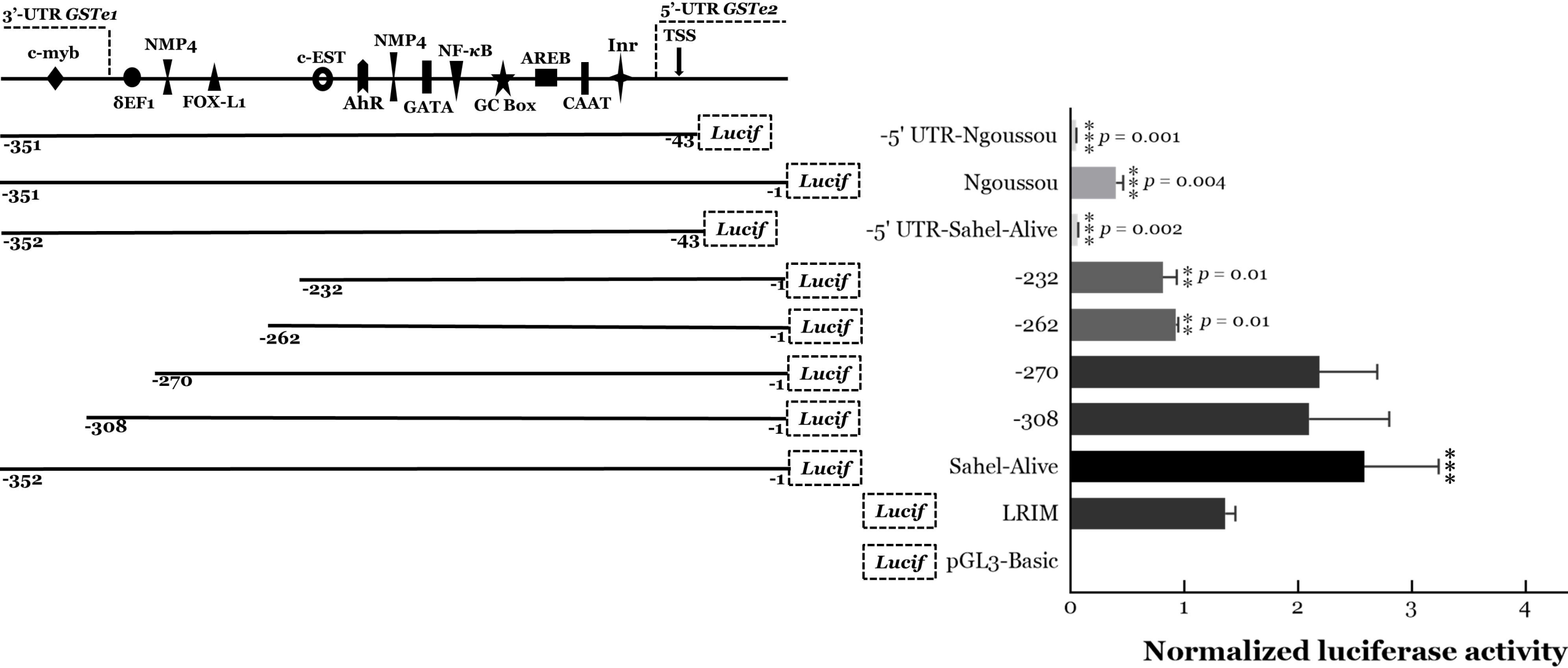
1272

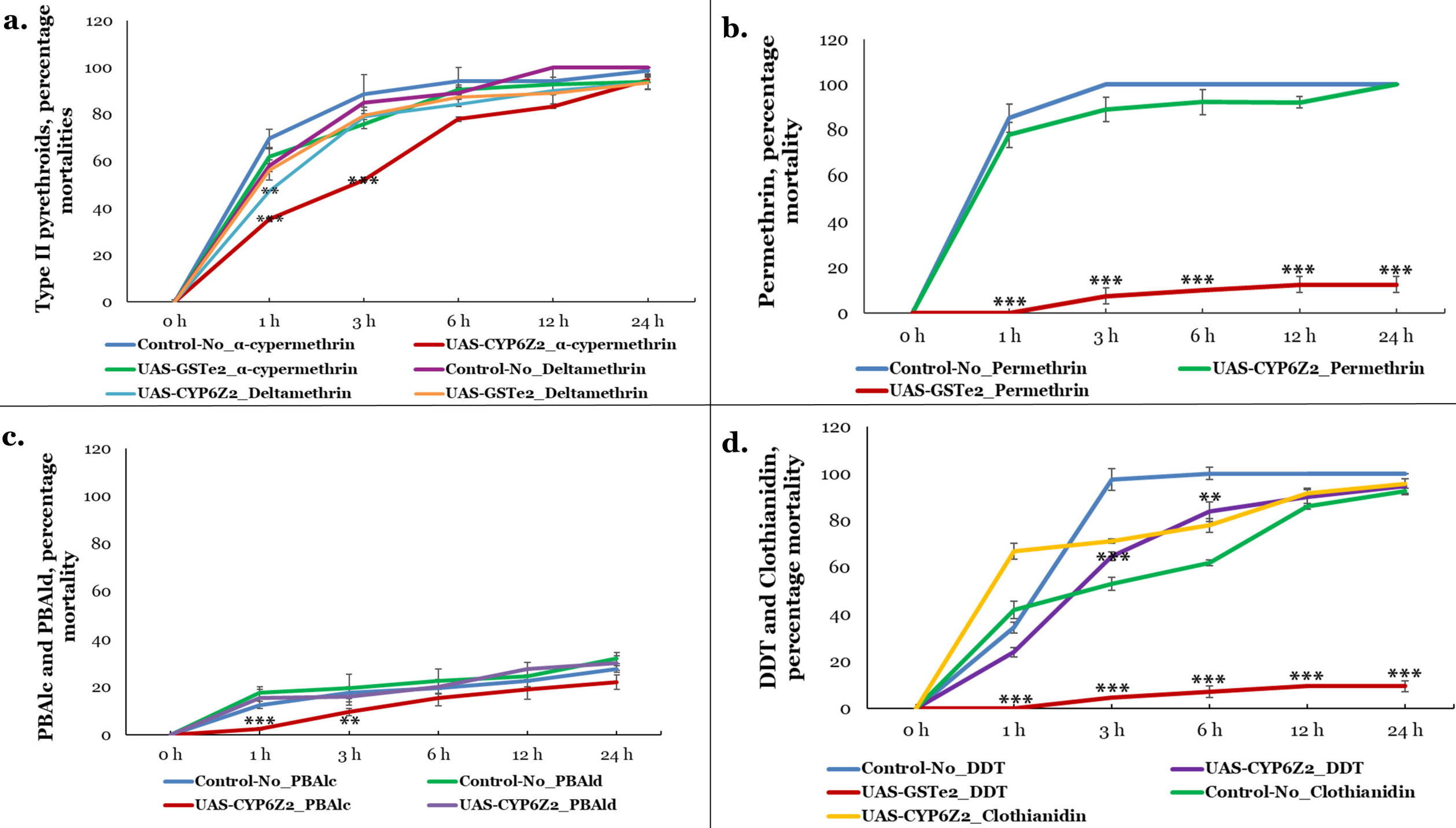

a.
Nigeria_R vs S
b.
Nigeria_C vs S
c.
Nigeria_R vs C
d.
Niger_R vs S
e.
Niger_C vs S
f.
Niger_R vs C
−Log10 P
Log2 fold change
NS
Log2 FC
p-value
p – value and log2 FC


a.
Chad_R vs S
b.
Chad_C vs S
c.
Chad_R vs C
d.
Cameroon_R vs S
e.
Cameroon_C vs S
f.
Cameroon_R vs C
−Log10 P
Log2 fold change
NS
Log2 FC
p-value
p – value and log2 FC


Relative normalized expression levels ($2^{-\Delta\Delta Ct}$)
50
40
30
20
10
0
GSTe2
GSTZ1
CYP6Z2
CYP6Z3
CYP4C27
CYP6P3
CYP6M2
CYP9K1
CYP4G16
CYP4G17
COEBE3C
UGT-B19
Candidate metabolic resistance genes
Nigeria R-S
Nigeria C-S
Niger R-S
Niger C-S
Chad R-S
Chad C-S
Cameroon R-S
Cameroon C-S


Variant Of Interest
2L:2390177 VGSC | R254K
2L:2391228 VGSC | V402L
2L:2399997 VGSC | D466H
2L:2416980 VGSC | T791M
2L:2422651 VGSC | L995S
2L:2422652 VGSC | L995F
2L:2429617 VGSC | I1527T
2L:2429745 VGSC | N1570Y
2L:2430424 VGSC | A1746S
2L:2430817 VGSC | V1853I
2L:2430863 VGSC | I1868T
2L:2430880 VGSC | P1874S
2L:2430881 VGSC | P1874L
2L:2431019 VGSC | F1920S
2L:2431079 VGSC | I1940T
2L:25429235 Rdl | A296S
2L:25429236 Rdl | A296G
2L:25635973 CYP4J5 | L43F
2R:28497967 CYP6P4 | I236M
2R:3492074 ACE1 | G280S
3R:28598062 GSTe2 | L119V
3R:28598166 GSTe2 | I114T
ChadDelta1
ChadDelta2
ChadDelta3
ChadUnexp1
ChadUnexp2
ChadUnexp3
ColzUnexp2
ColzUnexp3
ColzUnexp4
HadDelta3
HadDelta5
HadDelta6
HadUnexp1
HadUnexp3
HadUnexp6
Ngousso1
Ngousso2
Ngousso3
SimatDelta1
SimatDelta2
SimatDelta3
SimatUnexp2
SimatUnexp3
SimatUnexp4
TakDelta1
TakDelta3
TakDelta5
TakUnexp1
TakUnexp4
TakUnexp6
ChadDelta
ChadUnexp
ColzUnexp
HadDelta
HadUnexp
Ngousso
SimatDelta
SimatUnexp
TakDelta
TakUnexp
1.0
0.8
0.6
0.4
0.2
0.0


3'-UTR GSTe1
5'-UTR GSTe2
c-myb
NMP4
c-EST
NMP4
NF-κB
AREB
Inr
TSS
δEF1
FOX-L1
AhR
GATA
GC Box
CAAT
-351
-43
Lucif
-351
-1
Lucif
-352
-43
Lucif
-232
-1
Lucif
-262
-1
Lucif
-270
-1
Lucif
-308
-1
Lucif
-352
-1
Lucif
-5' UTR-Ngoussou
p = 0.001
Ngoussou
p = 0.004
-5' UTR-Sahel-Alive
p = 0.002
-232
p = 0.01
-262
p = 0.01
-270
-308
Sahel-Alive
Lucif
LRIM
Lucif
pGL3-Basic
0
1
2
3
4
Normalized luciferase activity


a.
Type II pyrethroids, percentage mortalities
120
100
80
60
40
20
0
0 h
1 h
3 h
6 h
12 h
24 h
**
***
***
Control-No_α-cypermethrin
UAS-CYP6Z2_α-cypermethrin
UAS-GSTe2_α-cypermethrin
Control-No_Deltamethrin
UAS-CYP6Z2_Deltamethrin
UAS-GSTe2_Deltamethrin
b.
Permethrin, percentage mortality
120
100
80
60
40
20
0
0 h
1 h
3 h
6 h
12 h
24 h
***
***
***
***
***
Control-No_Permethrin
UAS-CYP6Z2_Permethrin
UAS-GSTe2_Permethrin
c.
PBAlc and PBAld, percentage mortality
120
100
80
60
40
20
0
0 h
1 h
3 h
6 h
12 h
24 h
***
**
Control-No_PBAlc
Control-No_PBAld
UAS-CYP6Z2_PBAlc
UAS-CYP6Z2_PBAld
d.
DDT and Clothianidin, percentage mortality
120
100
80
60
40
20
0
0 h
1 h
3 h
6 h
12 h
24 h
**
***
***
***
***
***
***
Control-No_DDT
UAS-CYP6Z2_DDT
UAS-GSTe2_DDT
Control-No_Clothianidin
UAS-CYP6Z2_Clothianidin