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

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

20 **Authors email addresses**

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

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

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

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

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

26 Muhammad M. Mukhtar: muhammadmahemukhtar@gmail.com

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

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

29 Edward I. Patterson: ipatterson@brocku.ca

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

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

32

33

34

35

36

37

38

Abstract

Information on common markers of metabolic resistance in malaria vectors from countries sharing similar eco-climatic characteristics can facilitate coordination of malaria control. Here, we characterized populations of the major malaria vector *Anopheles coluzzii* from Sahel 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 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 (fixed). High frequencies of the epidemiologically important chromosomal inversions, 2La, 2Rb and 2Rc were observed (~80% for 2Rb and 2Rc). 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 24 h). 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 were responsible for the high overexpression of *GSTe2* in the resistant mosquitoes. 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 observed 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.

82 **Background**

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

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

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

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

127 the resistance markers in the field, as recently done for 119F-*GSTe2* (23), *CYP6P9a*_R and
128 *CYP6P9b*_R markers (10, 11) in *An. funestus*.

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

138 To support malaria pre-elimination effort in sub-Saharan Africa we targeted the Sahelo-
139 Sudanian region, which represent northern-most limit of malaria endemicity in sub-Saharan
140 Africa, and where malaria is highly seasonal [offering excellent target for pre-elimination
141 effort through sustained seasonal vector control and seasonal malaria chemoprevention
142 (25)]. Focusing on the predominant malaria vector, *An. coluzzii* from Sudano-Sahelian
143 transects of four countries, Nigeria, Niger, Chad and Cameroon (7, 26-28) we identified the
144 major metabolic resistance genes mediating pyrethroid resistance and cross-resistance in this
145 region, established the genetic variants which explained the resistance. We also functionally
146 validated the roles of two major candidate genes in the resistance (*CYP6Z2* and *GSTe2*) using
147 transgenic *Drosophila melanogaster* flies (GAL4/UAS system), as well as identifying single
148 nucleotide polymorphisms in the 5′ regulatory elements of *GSTe2*, responsible for its
149 overexpression in the resistant population.

150 # Materials and Methods

151 ## Study Site and Mosquito Sampling

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

162

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

165 ***RNA extraction, library preparation and sequencing***

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

171 manufacturer's protocol with *Dnase* I-treatment to remove contaminating DNA. The
172 quantity and quality of RNA was measured using a NanoDrop spectrophotometer
173 (ThermoFisher, MA, USA) and Bioanalyzer (Agilent, CA, USA).

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

192

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

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

213

214

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

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

225

226 Quantitative PCR measurement of expression profiles of metabolic resistance 227 genes

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

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

241

242 ***Detection of Signatures of Selective Sweep***

243 To detect signature of selective sweeps in the major metabolic resistance genes of interest, a
244 Snakemake RNAseq population genetics pipeline, the RNAseq-Pop was utilised (40). The
245 workflow aligns RNA-Seq reads to the reference genome, and calls genomic variants with
246 *Freebayes*, at a user-provided level of ploidy, in our case 16 (8 diploid pooled mosquitoes). F_{st}
247 (41) per gene between population pairs and Tajima's D per gene within population were
248 estimated. This was performed against all SNPs passing quality and missingness filters.
249 Population branch statistic (PBS) scans was performed with the Snakemake, conditional on
250 the presence of three suitable populations (42). Hudson's F_{st} and PBS scans were ran, taking
251 the average for each protein-coding gene, as opposed to in windows. The population
252 genetics statistical analyses were calculated in scikit-allel v1.2.1 (43).

253

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

255 After genome alignment, RNA-Seq-Pop utilises samtools (44) to query specific positions of
256 the genome, calculating raw allele frequencies at those sites with a custom R script.

257

258 ***Detection of chromosomal inversion polymorphisms and metabolic genes within its***
259 ***breakpoint***

260 A modified version of the Python 3 program, compkaryo (45) was used to karyoptype the
261 major, *An. coluzzii/gambiae* phenotypically important inversion polymorphisms in
262 chromosome 2, and calculate its frequencies, *in silico*, using the previously identified tag
263 SNPs significantly associated with inversions. This allows to predict with high confidence
264 genotypes of the six common polymorphic inversions on chromosome 2, in the sequenced
265 field *An. coluzzii,* as well as in the Ngoussou. Compkaryo uses the Ag1000 database
266 (The *Anopheles gambiae* 1000 Genomes Consortium 2017) (15) by leveraging a subset of
267 cytologically karyotyped specimens to develop a computational approach for karyotyping
268 applicable to whole genome sequence. Modifications in the Snakemake pipeline allows for
269 variable ploidy (useful in the case of replicates from our pooled RNAsequencing samples)
270 here.

271

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

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

274 Two resistance-associated genes, *GSTe2* and *CYP6Z2* which feature most premoninently in
275 the resistant *An. coluzzii* populations across the Sahel were chosen for functional validation.
276 To establish presence of allelic variants which could impact catalytic activities, full length
277 coding sequences (cDNAs) of *GSTe2* and *CYP6Z2* were amplified and sequenced from alive
278 mosquitoes in the four Sahel countries, as well as from the Ngoussou. This was done using
279 total RNA extracted from 5 individual pools of 8 F_1 *An. coluzzii* females (2-4 day old) alive

280 after exposure to deltamethrin (resistant, R for *CYP6Z2*) or DDT (R for *GSTe2*). Protocol for
281 RNA extraction was as described in previous section, above. Amplification was done using
282 Phusion HotStart II Taq Polymerase (ThermoFisher SCIENTIFIC, MA, USA) and the Full
283 primers listed in Additional Table 4. The PCR mix comprised 5x Phusion HF Buffer
284 (containing 1.5 mM $MgCl_2$), 85.7 μM deoxynucleotides (dNTPs), 0.34 μM each of forward
285 and reverse primers, 0.015 U of Phusion HotStart II DNA Polymerase (Fermentas, MA,
286 USA), 10.71 μL of ddH_20 and 1 μL cDNA. Thermocycling conditions were 1 cycle at 95 °C
287 for 5 min; followed by 35 cycles each of 94 °C for 20 s, 60 °C for 30 s, 72 °C for 2 min (1 min
288 for *GSTe2*); and finally, one cycle at 72 °C for 5 min. PCR products were cleaned with a
289 QIAquick® PCR Purification Kit (QIAGEN, Hilden, Germany) and ligated into the
290 pJET1.2/blunt cloning vector using the CloneJET PCR Cloning Kit (ThermoFisher
291 SCIENTIFIC, MA, USA). These were then cloned into *E. coli DH5α*, plasmids miniprepped
292 with the QIAprep® Spin Miniprep Kit (QIAGEN) and sequenced on both strands using
293 pJET1.2 primers.

294 Polymorphisms were detected through examination and manual editing of sequence traces
295 using BioEdit version 7.2.3.0 (46) 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 maximum likelihood phylogenetic tree using MEGA X (47). Genetic
298 parameters of polymorphism including number of haplotypes (h) and its diversity (H_d),
299 number of polymorphic sites (S) and nucleotide diversity (π) were computed using DnaSP
300 v6.12.03 (48).

301

302

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

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

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

319 The 351 bp 5′-UTR fragments of *GSTe2* and 1078 bp fragment of the *CYP6Z2* were analysed
320 with the Gene Promoter Miner (http://gpminer.mbc.nctu.edu.tw/) and MatInspector (49) to
321 identify putative promoter elements, predict transcription start (TSS) and potential
322 transcription factor binding sites.

323

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

326 Following analysis of the above sequences the 351 bp intergenic fragments of *GSTe2* were
327 amplified from the most predominant sequences of DDT-alive, DDT-dead and Ngoussou.
328 Same was done for for *CYP6Z2* amplifying 1087 bp fragment from deltamethrin-alive and -
329 dead, and Ngoussou. Primers bearing *kpn*I and *BgL*II sites (Additional Table 4) allowed
330 incorporation into pGL3-Basic reporter vector containing luciferase gene from the firefly
331 *Photinus pyralis* (Promega, Wisconsin, USA). Amplification was carried out using Phusion
332 HotStart II Polymerase, with conditions as above, followed by purification of PCR
333 amplicons, and cloning into pJET1.2 vector. Positive colonies (sequencing primers for pGL3-
334 Basic provided in Additional Table 4) were miniprepped; the minipreps digested with the
335 above restriction enzymes, gel-purified and ligated upstream of luciferase gene in pGL3-
336 Basic vector already linearized with the same restriction enzymes. Positive colonies were
337 cloned and miniprepped and concentrations of the recombinant plasmids adjusted to 200
338 ng/µL before co-transfection. Protocols for maintenance of the *An. gambiae* cell line 4a-3B
339 (MRA-919, https://www.beiresources.org/), co-transfection of 200 ng of either *GSTe2* or
340 *CYP6Z2* promoter constructs, LRIM promoter in pGL3-Basic vector (50), or promoter-less
341 pGL3-Basic control, with 1 ng/µL of internal control, sea pansy *Renilla reniformis* luciferase
342 containing the *Drosophila* Actin *5C* promoter in pRL-null (50), are provided in Additional
343 File 1 (additional text for Methods and Results). Protocol for cell lysis and measurement of
344 the activities of the firefly luciferase (Dual-Luciferase Reporter Assay kit, Promega), with
345 normalisation using the Renilla luciferase activity, were as outlined in the Promega Quick
346 Protocol, and provided in detail in Additional File 1.

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

350 *Generation of GSTe2 promoter deletion constructs and assays*

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

369 primer for the 351 nucleotides intergenic region, with a newly designed primer,
370 GSTe2_minus_5′-UTR-R (Additional Table 4).

371

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

373 *Cloning and microinjection of GSTe2 and CYP6Z2 into Drosophila melanogaster*

374 Transgenic flies expressing recombinant GSTe2 and CYP6Z2 were created using GAL4/UAS
375 system, as outlined in a previous publication (51) and used in contact bioassays, to confirm if
376 over-expression of these genes alone can confer resistance to insecticides. Details of
377 amplification of full-length *GSTe2* and *CYP6Z2* from the predominant alleles of sequences of
378 cDNA, cloning into pUASattB vector, microinjection into germ-line *D. melanogaster* lines,
379 crossing of the UAS-lines with the GAL4-Actin driver strain are outlined in Additional File
380 1.

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

387 *Insecticide Susceptibility Contact Bioassay*

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

391 (BHD/Merck, Hesse, Germany). Flies overexpressing *CYP6Z2* were also exposed to the
392 primary product of hydrolysis of pyrethroids: 4 % and 20 % (5x) 3-phenoxybenzaldehyde
393 (PBAld) and 3-phenoxybenzylalcohol (PBAlc). Transgenic flies expressing *GSTe2* were
394 exposed to 2 % permethrin, 0.15 % deltamethrin, 0.05 % α-cypermethrin and 4 % DDT only.
395 Impregnated papers were rolled and introduced into 45 cc plastic vials to cover the entire
396 wall and the vials plugged with cotton soaked in 10 % sucrose (38). 20–25 flies were placed
397 in each vial, and the mortality plus knockdown scored at 1 h, 3 h, 6 h, 12 h and 24 h of
398 exposure to the insecticides. For each insecticide, assays were performed in 6 replicates and
399 Student's t-test used to compare the mortality plus knockdown between the experimental
400 groups and the control. Controls were adult F_1 progeny with the same genetic background
401 as the experimental group but without the *GSTe2* or *CYP6Z2* inserts. They were obtained by
402 crossing virgin females from the driver strain Act5C-GAL4 and the UAS recipient male lines
403 with white eyes (not carrying the pUASattB-GSTe2 or pUASattB-CYP6Z2 insertions).

404

405 # Results

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

407 A three-way pairwise comparison was conducted for the data from each country: resistant
408 *vs* susceptible (R-S), resistant *vs* unexposed control (R-C) and unexposed control *vs*
409 susceptible (C-S). This captures background variations due to geographical differences in the
410 resistant vs susceptible (R-S) comparison, accounts for genes overexpressed due to induction
411 (R-C comparison), as well as genes that are constitutively overexpressed (C-S comparison).
412 A total of 1384 genes were significantly differentially expressed (FDR-adjusted $p < 0.05$ and

413 $\log_2$ fold change threshold of 1/FC ≥ 2) in R-S comparison in Nigeria (1077 upregulated and
414 307 down-regulated); 1185 genes were differentially expressed in C-S (1002 upregulated and
415 183 downregulated); and 295 genes in R-C (129 upregulated and 166 downregulated). Of
416 these, 52 genes were commonly differentially expressed in all 3 comparisons (Figure S3,
417 panel a), including the upregulated genes, *COEAE80* (AGAP006700), *CYP4H18*
418 (AGAP028019), *CYP4H17* (AGAP008358) and cuticular proteins, *CPLCX3* (AGAP006149),
419 *CPR59* (AGAP006829), *CPR76* (AGAP009874) and *CPR75* (AGAP009871). The Additional
420 Figure S3, panels a-d depicts the differentially expressed genes for the four countries. For
421 Niger, 881 genes were differentially expressed in R-S comparison (619 upregulated and 262
422 down-regulated) (Additional Figure S3b); 1256 genes were differentially expressed in C-S
423 (986 upregulated and 270 downregulated); and 196 genes in R-C (81 upregulated and 115
424 downregulated). Of these, 22 genes were commonly differentially expressed in all 3
425 comparisons [including the upregulated *aminopeptidase N1* (AGAP012757), *CYP6Z2*
426 (AGAP008218), *chymotrypsin-3* (AGAP006711) and *chymotrypsin-2* (AGAP006710), and an
427 acid trehalase (AGAP008547)]. For Chad, 1392 genes were differentially expressed in R-S
428 comparison (975 upregulated and 417 down-regulated) (Additonal Figure S3c); 1284 in C-S
429 (105 upregulated and 269 downregulated); and 526 genes in R-C (270 upregulated and 256
430 downregulated). Of these, 97 genes were commonly differentially expressed in all 3
431 comparisons [including *CYP4C27* (AGAP009246), *SULTD1* (AGAP012672), *aminopeptidase*
432 *N1*, and diverse cuticular proteins, e.g. chitinase (*Cht24*, AGAP006191, *CPFL1*
433 (AGAP010902), *CPCFC1* (AGAP007980), *CPLCX3*, *CPR24* (AGAP005999), *CPR106*
434 (AGAP006095) and *CPR130* (AGAP000047). Finally, for Cameroon, 376 genes were
435 differentially expressed in R-S comparison (204 upregulated and 172 down-regulated)
436 (Additional Figure S3d); 932 in C-S (778 upregulated and 154 downregulated); and 116

437 genes in R-C [only 9 upregulated and 107 downregulated (probably due to a single, low
438 quality replicate in the raw data from Cameroon (Additional Figure S2). Not surprising,
439 only 7 genes were commonly differentially expressed in all 3 comparisons. These include the
440 highly upregulated gene, *GSTe2* (AGAP009194) and *chymotrypsin-1* (AGAP006709).

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

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

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

453 The most differentially expressed genes were presented in Figure 1, a volcano plot of fold
454 change *vs* significance levels for Nigeria and Niger, and Figure 2, for Chad and Cameroon.
455 Detailed lists of these genes of interest is provided in Additional File 2. Comparisons of
456 genes commonly, upregulated and/or downregulated in R-S/R-C/C-S, from the four
457 countries revealed similar transcriptomic profiles between R and C compared with the S.
458 The most commonly and consistently overexpressed genes across the Sahel (taking account

459 mean expressions) are the chymotrypsins, 3, -2 and -1 (*CHYM3*/AGAP006711 and
460 *CHYM2*/AGAP006710 and *CHYM1*/AGAP006709) (Figures 1 and 2, Additional File 2), the
461 glutathione S-transferase, *GSTe2* (AGAP009194), an aquaporin, *AQP3* (AGAP010326),
462 *CYP6Z2* (AGAP008218), *CYP6Z3* (AGAP008217), *CYP4C27* (AGAP009246), a chitinase, *Cht24*
463 (AGAP006191), a thioester-containing protein-1, *TEP-1* (AGAP010815),), a trehalose 6-
464 phosphate synthase/phosphatase, *TPS*1/2 (AGAP008227), a lipase (AGAP002353), and
465 AGAP012818 (V-type protein ATPase subunit A). More on these genes is provided in the
466 sections, below.

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

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

479 Several genes were significantly downregulated across the four countries, particularly the
480 cytochrome P450s and the ATP-binding cassette transporters. The most consistently down-
481 regulated P450s were *CYP6AK1* (AGAP010961), downregulated in R-S and C-S comparisons

482 in Nigeria, Niger and Cameroon, while down-regulated in all comparisons in Chad;
483 *CYP4H24* (AGAP013490), downregulated in R-S and C-S comparisons for Nigeria and
484 Cameroon, and downregulated in all comparisons in Niger and Chad; *CYP6M4*
485 (AGAP008214), downregulated in R-S and C-S comparisons in Nigeria, Niger and
486 Cameroon, while downregulated in all comparisons in Chad; and finally *CYP6P5*
487 (AGAP002866), downregulated in R-S and C-S comparisons in Niger and Chad, but only in
488 C-S in Nigeria.

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

490 Special attention was given to the known metabolic resistance genes, implicated in
491 insecticide resistance in *Anopheles* and/or other insects, including, CYP450s, GSTs,
492 carboxylesterases, cuticular proteins, chemosensory proteins/SAPs, uridine
493 diphosphoglucuronosyltransferases, etc, in addition to other important genes, e.g., the
494 immune proteins. Analysis of the data from the list of genes significantly, differentially
495 expressed (FDR-adjusted $p < 0.05$ and $\log_2 FC = 1/ FC \geq 2$) revealed the major, commonly
496 upregulated metabolic resistance genes across the Sahel. Top 50 genes from these lists
497 (Additional File 3) from each country are displayed as heatmaps of fold changes in
498 Additional Figure S5a, -b, -c and -d). The most common genes linked with resistance and/or
499 other physiologically important phenotypes are tabulated in Table 1, showing fold changes
500 of 55 genes (R-S and C-S comparisons), common to all countries, or three or two countries at
501 least. The most commonly and consistently overexpressed genes across the Sahel (taking
502 account mean expressions) are the chymotrypsins, -1, -2 and -3 (*CHYM1*/AGAP006709,
503 *CHYM3*/AGAP006711 and *CHYM2*/AGAP006710) (Table 1, Additional Figure S5, Additional
504 File 2, Results section), with fold change for *CHYM1* in R-S and C-S comparisons of 31.62

505 and 54.41 for Nigeria, 12.31 and 41.59 for Niger, and 18.00 and 18.08 for Chad, and 5.69 and
506 11.77 for Cameroon; for *CHYM3*, 32.45 and 58.43 for Nigeria R-S and C-S, respectively, 7.30
507 and 43.27 for Niger, 13.40 and 7.24 for Chad, and 2.59 and 8.15 for Cameroon. Of *Anopheles*
508 metabolic resistance genes, *GSTe2* (AGAP009194) was the most consistently upregulated
509 gene across the Sahel, FC = 9.61 and 7.16, respectively for Nigeria R-S and C-S comparisons;
510 6.55 and 11.32 for Niger; 9.09 and 10.02 for Chad, 9.68 and 20.67 for Cameroon. Other
511 common GSTs were *GSTZ1*, upregulated in three countries (Nigeria, Niger and Chad),
512 *GSTe4*, *GSTU1*, *GSTD1-4* and *GSTD3*.

513 Other most consistently and highly upregulated genes include a lipase, AGAP002353: FC of
514 26.73 and 20.80 for Nigeria R-S and C-S comparisons, 16.53 and 29.76 for Niger, 14.95 and
515 9.95 for Chad, 14.58 and 14.53 for Cameroon; a chitinase, *Cht24* (AGAP006191), FC = 21.12
516 and 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
517 and 40.53 for Cameroon); a V-type protein ATPase subunit a (AGAP012818, FC = 22.12 and
518 14.61 for Nigeria, 31.05 and 25.43 for Nigeri, 6.00 and 4.31 for Chad, 44.92 and 43.4 for
519 Cameroon), and an aquaporin, *AQP3* (AGAP010326, FC = 14.81 and 11.03 for Nigeria, 12.09
520 and 21.94 for Niger, 14.86 and 1.21 for Chad, 14.17 and 17.3 for Cameroon).

521 The most overrepresented gene family, and most consistently upregulated across the Sahel
522 were the cytochrome P450s, with *CYP6Z2*, *CYP6Z3*, and *CYP4C27*, in addition to the two
523 P450s linked to cuticular resistance, *CYP4G16* (upregulated in Nigeria and Chad) and
524 *CYP4G17* (upregulated in Nigeria, Niger and Chad). The three P450s, *CYP6Z2, CYP6Z3 and*
525 *CYP4C27* are consistently overexpressed: FC for *CYP6Z2* is 6.55 and 2.10 for Nigeria R-S
526 and C-S comparisons, 6.73 and 2.32 for Niger, 13.96 and 1.48 for Chad, 11.84 and 8.39
527 for Cameroon; FC for *CYP6Z3* is 4.73 and 1.82 for Nigeria, 6.09 and 1.42 for Niger, 3.51 and

528 1.11 for Chad, 7.11 and 3.79 for Cameroon; and FC for *CYP4C27* is 7.36 and 4.6 for Nigeria,
529 4.03 and 5.64 for Niger, 6.04 and 2.62 for Chad, 2.01 and 2.63 for Cameroon. The two well-
530 known insecticide resistance genes *CYP6P3* and *CYP6M2* were not
531 overexpressedprominently across the Sahel of these countries.

532 The second most commonly overrepresented gene families belong to cuticular proteins,
533 including cuticular proteins Rebers and Riddiford (RR), several cuticular proteins of low
534 complexity (CPLC) and chitin-binding cuticular proteins (CPAP), with the commonly
535 upregulated ones across all four countries being *CPR75* (AGAP009871) and *CPR76*
536 (AGAP009874).

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

543 Other genes in the top 55 metabolic genes include the malate dehydrogenase, *TPS* (*TPS1/2*),
544 *AQP2*, thioester-containing proteins, *TEP1*, *-12* and *-14*, a cyanogenic-beta-glucosidase
545 (AGAP006422) and three chemosensory proteins, *CSP5*, sensory appendage proteins, -1 and
546 -2.

547
548
549
550
551

552 **Table 1: The common detoxification and metabolic genes differentially upregulated in**
553 **Sahel *An. coluzzii* (log_2FC values and 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

554 *Gene Ontology Enrichment Analysis*

555 Gene ontology enrichment analysis, for genes significant in comparisons ($log_2FC = 1$, $p <$
556 0.05) revealed crucial differences in overrepresented GO terms between the up- and -
557 downregulated genes, in data from the four countries. For example, the most over-
558 represented, semantically similar GO terms associated with xenobiotics metabolism is
559 oxidoreductase activity ((highest frequency of 12.9 %, Additional Figure S6a, REVIGO Table
560 View), which cluster together in Nigerian R vs S comparison (for genes upregulated in R).
561 Other over-represented GO terms in this comparison include glutathione S-transferase
562 activity (the most enriched/specific term), peroxidase activity, odorant binding and chitin
563 binding. In contrast, GO terms over-represented in R vs S (down-regulated in R) were
564 mostly involved in neurotransmission, metal ions binding and receptor channelling
565 activities (Additional Figure S6b).

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

574

575

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

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

590

591 **Detection of Signatures of Selective Sweeps**

592 Signatures of selection were investigated in the major metabolic resistance genes, by
593 estimating Tajima's D per gene within populations, and *F*st per gene between population
594 pairs. These tests of neutrality revealed several genes exhibiting genetic differentiation, or
595 possibly undergoing expansion. Among the top 100 most overexpressed metabolic genes
596 across the four countries (Additional File 2), 6 genes were possibly undergoing genetic
597 differentiation. These include *TEP1,* and *TEP3*, with average Tajima's D of -1.4 and -1.6

598 respectively, in populations from Chad, and *F*st values of ≥ 0.85 for *TEP1*, for Nigeria, Chad
599 and Cameroon samples, compared to the average *F*st of 3L chromosomal, which was
600 calculated as 0.24 (Additional File 3), *CYP9K1* (average Tajima's D = -1.00 for Cameroon,
601 and *F*st within ranges of 0.86 - 1 for Nigeria, Cameroon and Niger, compared with 0.21
602 calculated for the chromosome X), *CPR15* (average Tajima's D = -0.94); *CPAP3-A1b* (Tajima's
603 D = -0.70 in Chad, -0.95 in Nigeria, and -1.32 in Cameroon, with *F*st values in range of 0.7 -
604 0.9 for the four countries, compared with the average *F*st of 0.21 for chromosome X) and
605 *GSTU1*. Several genes from the GST family exhibited strong genetic differentiation, with
606 reduced variations prominently in *GSTe2* (< 3 SNPs in Chad and Cameroon samples), with
607 Tajima's D of -1.45 and -1.15 for Nigeria and Niger, and combined. Information on more
608 genes is provided in Additional File 1, and detailed information on all genes can be found
609 in the Additional File 3.

610

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

612 The RNA-Seq-Pop workflow calculated allele frequencies of variants of interest found in the
613 raw RNA-sequencing read data. Figure 4 displays allele frequencies of each variant in
614 respective treatment replicates used for the RNAseq, as well as Ngoussou. Several mutaions
615 were found within the voltage-gated sodium channel. In addition to the L995F *kdr*
616 mutation, the recently identified V402L and I1527T mutations exist in all the field
617 populations across the Sahel, at frequencies ranging from 29 % to 67 %. We also detected the
618 N1570Y mutation, which shares the same haplotype as L995F. The G280S *acetylcholinesterase-*
619 1 mutation was also observed in low to moderate frequencies in all populations except for

620 Ngoussou and samples from Niger. The L43F pyrethroid resistance marker of *CYP4J5* was
621 also seen, fixed in most populations.

622

623 **Detection of chromosomal inversion polymorphisms and metabolic genes within**
624 **its breakpoints**

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

643 are among the top 50 most overexpressed genes in all countries. Other genes include:
644 *COEJHE5E*, one of the 14 carboxylesterases in the 2La region, and which was among the top
645 100 most overexpressed genes across the Sahel; the most overexpressed chymotrypsin genes,
646 *CHYM1* and *CHYM2*; three ionotropic receptors, *IR136* (AGAP006440), *IR139* (AGAP006691)
647 and *IR142* (AGAP006407); as well as five P450s, *CYP301A1*, *CYP302A1*, *CYP4J5*, *CYP4J9* and
648 *CYP4J10*. The most overrepresented genes within 2La inversions are the CPR cuticular
649 proteins (73 in total), with *CPR21*, *-26*, *-30*, *-59* and *-140* from the list of the top 100 resistance
650 genes from the Sahel.
651 Carbonic anhydrase I (the most overexpressed metabolic resistance in Nigeria,
652 AGAP013402) sits within 2Rb inversion. Four chitinases (*Cht4*, *Cht5-1*, *Cht5-3*, and *Cht5-5*)
653 are located within 2Rb inversion; *Cht5-1* and *Cht5-5* were among the observed top 50 most
654 overexpressed metabolic genes in the Sahel. Lipase (*AGAP002353*) which is the among the
655 top 6 overexpressed metabolic gene in the four countries sit within the 2Rb inversion. Eight
656 CYP450s reside in the 2Rb inversion, including *CYP4D15*, *CYP4D17* and *CYP4AR1*, which
657 were among the top 100 metabolic resistance genes. Three cuticular proteins, *CPR7*, *-8* and *-9*
658 were also within the 2Rb inversion.

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

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

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

682

683 **Investigation of the role of intergenic region elements in overexpression of *GSTe2***
684 **and *CYP6Z2***

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

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

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

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

711 a C -> G transition, 2 nucleotides downstream the arthropod initiator (Inr consensus)
712 sequence, and finally (x) a C -> G transition, within the 5′-UTR of the *GSTe2*, 28 nucleotides
713 downstream the transcriptional start site (49 nucleotides upstream the *GSTe2* start codon).

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

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

724

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

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

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

741

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

743 Sequential deletion of the intergenic region of the *GSTe2* resulted in progressive reduction in
744 luciferase activity. Deletion of the 43 bp 3′-UTR of the *GSTe1* (-308 from the start codon of
745 *GSTe2*) reduced activity of the Sahel-alive construct by only 18.6 % suggesting that the c-
746 Myb transcriptional factor binding site may not be critical for overexpression (Figure 5).
747 Deletion of an additional 38 nucleotides from the flanking region (-270 fragment,
748 obliterating the δEF1 binding site) had comparable impact as above, with activities reducing
749 by 15.3 % only. But removing the *GSTe-1* 3′-UTR and an additional 46 nucleotides from the
750 flanking region (-262 fragment, which obliterated in addition the FOX-L1 putative binding
751 site) reduced activity by 64.1 % (Q = 6.18, p = 0.01) indicating the importance of this binding
752 site. Deletion of *GSTe1* 3′-UTR plus 75 nucleotides of the flanking region (-232 fragment,
753 located 10 nucleotides from the c-EST binding site) significantly reduced the activity by
754 ~70% (Q = 6.14, p = 0.01), suggesting the importance of the simultaneous insertion of adenine

755 and transition T -> C, in positions 113 and 114 respectively, between the FOX-L1 and c-EST
756 binding sites. Removing the fragment of the 5'UTR of *GSTe2* (43 nucleotides preceding the
757 AUG codon obliterated the promoter activity, reducing the luciferase expression by 98 % (Q
758 = 9.41, p = 0.002). This is despite presence of all the above binding sites and transcriptional
759 start site in this fragment. Deletion of the 5'-UTR from Ngoussou significantly reduced
760 activities as well (reduction by 88 % compared with the full Ngoussou intergenic region
761 construct, Q = 13.75, p = 0.001). Overall, these findings suggest that the essential binding sites
762 for overexpression of *GSTe2* span the FOX-L1 and c-EST binding sites, with the 5'-UTR
763 essential for activity.

764

765 **Investigating the role of *GSTe2* and *CYP6Z2* in insecticides resistance using**
766 **transgenic analysis**

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

772 Contact bioassays carried out using 0.05 % α-cypermethrin revealed a high susceptibility in
773 the transgenic flies expressing *GSTe2* (Act5C-GAL4-UAS-GSTe2) and controls (Figure 6a),
774 with mortalities increasing from 62 % and 70 % respectively in 1 h, to 94 % and 99 % in 24 h.
775 However, significantly reduced mortalities were observed in transgenic flies expressing
776 *CYP6Z2* (Act5C-GAL4-UAS-CYP6Z2) compared to control flies, at 1 h (mortality = 35 % *vs* 70

777 % in control, $p < 0.001$) and 3 h (mortality = 52 % vs 89 % in the control, $p < 0.001$). High
778 susceptibilities were also seen in all the experimental flies exposed to 0.05 % deltamethrin
779 (both for *GSTe2* and *CYP6Z2* flies), except for 1 h with flies expressing *CYP6Z2* (mortality =
780 47 % compared with 58 % in the control flies, $p < 0.010$).

781 The *GSTe2* transgenic flies were highly resistant to 2 % permethrin, with no mortality at all
782 in 1 h (Figure 6b), and average mortalities of only 8.5 % at 24 h ($p < 0.0001$), compared with
783 85 % at 1 h for control flies, which increased to 100 % from 3 h.

784 Initial exposure to 4 % of 3-phenoxybenzaldehyde (PBAld) and 3-phenoxybenzyl alcohol
785 (PBAlc) had no toxic effect on *CYP6Z2* transgenic flies (Figure not shown). However, 5x
786 concentration of these primary products of pyrethroid hydrolysis induced mortalities, albeit
787 low in all flies (Figure 6c). For PBAld, mortalities ranged from 15.5 % and 17.5 % at 1 h, for
788 *CYP6Z2* transgenic flies and control flies respectively, to 30 % and 27.5 % at 24 h. Although
789 low mortalities were observed with PBAlc, but at 1 h and 3 h exposure times the mortalities
790 in the transgenic flies expressing *CYP6Z2* were significantly lower compared with
791 mortalities from the control flies (1 h mortality = 2.5 % *vs* 12.5 %, $p < 0.001$; 3 h mortality = 9.5
792 % *vs* 17.5 %, $p < 0.001$).

793 Marginal tolerance towards DDT was observed in the *CYP6Z2* transgenic flies when
794 compared with the control flies, but only at 3 h ($p < 0.001$) and 6 h ($p < 0.01$) (Figure 6d). This
795 is in contrast with the *GSTe2* transgenic flies, with very low mortalities, in ranges of 0 % to
796 9.52 % for 1 h to 24 h, when compared with control flies (mortality = 34 % in 1 h and 100 % in
797 6 – 24 h, $p < 0.0001$).

798 Susceptibility to clothianidin was very high (Figure 6d), but surprisingly the *CYP6Z2*
799 transgenic flies exhibited a contrasting phenotype, with significantly higher mortalities
800 compared with the control flies, at 1 h (mortality = 67 % vs 42 %, $p < 0.001$), 3 h (mortality =
801 71 % vs 53 %, $p < 0.001$) and 6 h (mortality = 78 % vs 62 %, $p < 0.01$).

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

805

806 **Discussion**

807 Escalating insecticide resistance across Africa (7-9, 52), if not tackled will compromise the
808 malaria control and elimination efforts. Molecular markers of metabolic resistance e.g., (10,
809 11, 21), will support evidence-based control and resistance management. Identification and
810 validation of resistance markers in malaria vectors across regions of sub-Saharan Africa can
811 promote communication, cooperation and coordination among malaria control/elimination
812 programs, and allow control efforts to be tailored to the vector species involved in
813 transmission across borders (53), and tracking of the evolution and spread of resistance
814 markers across regions (22). To support malaria pre-elimination efforts in Africa, in this
815 study we targeted the Sudan savannah and Sahel (regions of Africa sharing similar eco-
816 climatic conditions, and characterised by high seasonal transmission), which are ideal for
817 control and elimination of malaria using seasonal vector control (25) and chemoprevention.

818

819

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

821 Contrary to the previous observations that *An. arabiensis* tends to predominate in arid
822 savannas, while *An. gambiae* is the dominant species in humid forest zones (54-56), *An.*
823 *coluzzii* has repeatedly been identified as the major malaria vector in the Sudan savannah
824 and Sahel of several, neighbouring countries (including in northern Nigeria, southern Niger,
825 central Chad, and northern Cameroon) (7, 26-28, 57) in recent years, suggesting that this
826 vector has adapted well in drier regions of the Sahel transects, and is probably
827 predominating over *An. arabiensis* and *An. gambiae s.s*. This is not surprising as this species
828 exhibits higher exploitation of breeding sites associated with anthropogenic activities, and
829 behavioural plasticity to avoid predators (58), and is known to survive a long dry season *in*
830 *situ*/aestivation, which allows it to predominate, becoming the primary force of malaria
831 transmission (59, 60).

832

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

834 Several genes shown to confer metabolic resistance in *Anopheles* mosquitoes and other
835 insects were found constitutively overexpressed and/or induced in this study. *GSTe2*
836 (AGAP009194) is one of the most regularly encountered metabolic genes in resistant
837 populations of the major malaria vectors *An. gambiae*, *An. coluzzii* and *An. funestus* (20-23).
838 *Anopheles gambiae GSTe2* has been validated, using transgenic flies to confer DDT (20) and
839 fenitrothion (61) resistance. It was extensively studied in *An. funestus*, in which it was shown
840 to confer cross-resistance to DDT and permethrin (22), reduce efficacy of the LLINs,
841 PermaNet 2.0 and PermaNet 3.0 (side panels) (23), and even increase the longevity of the

842 resistant populations carrying the 119F mutation (62). These and our findings suggest that
843 the overexpression of this GST alone can confer resistance to three insecticides from three
844 different classes (DDT, permethrin and fenitrothion). The absence of mutations in the cDNA
845 coding sequences of *GSTe2* in *An. coluzzii* from these four countries, suggest that
846 overexpression of this GST is enough to confer resistance.

847 Several cytochrome P450s previously linked with insecticide resistance were found
848 overexpressed across the Sahel. For example, *An. gambiae CYP6Z2* (AGAP008218) known to
849 metabolize carbaryl (63), the insect juvenile hormone analogue insecticide, pyriproxyfen (64)
850 and mitochondrial complex I inhibitors, fenazaquin, pyridaben and tolfenpyrad (65). This
851 P450 also plays a pivotal role in the clearance of pyrethroid insecticides via further
852 catabolism of pyrethroid derivatives (PBAld and PBAlc) obtained by the action of
853 carboxylesterases (66), in line with our findings of this gene conferring marginal tolerance to
854 high concentration of PBAlc, and α-cypermethrin. However, our findings suggest that
855 overexpression of this P450 may enhance the efficacy of clothianidin, which will be
856 epidemiologically advantageous in terms of control. Indeed, bioactivation by P450s is
857 known to be a requirement for insecticidal toxicity of several classes of insecticides, e.g., the
858 organophosphates and chlorpenafyr. In contrast to *GSTe2*, the *CYP6Z2* from across Sahel
859 contain three cDNA mutations, which makes it different from Ngoussou coding sequences.
860 These are $K^{211}N$ and $T^{218}S$ mutations, both within the substrate recognitions site 2, and an
861 $A^{282}E$. Other important CYP450s found to be commonly overexpressed across the Sahel
862 include the *CYP4C27* (AGAP009246), *CYP6Z3* (AGAP008217) and *CYP9K1* (AGAP000818),
863 all three shown to be consistently overexpressed in field populations of *An. gambiae/coluzzii*
864 and *An. funestus* across Africa (9, 67, 68). *CYP9K1* has been shown to be epidemiologically

important pyrethroid-metabolising P450 linked with metabolism of deltamethrin and pyriproxyfen in *An. gambiae* (69). We also have recently shown that his P450 is involved in pyrethroids in *An. funestus* (70).

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

Common metabolic resistance markers probably exacerbate resistance across the Sahel

Several well-known genetic variants implicated in resistance, as well as the recently discovered ones exist in high frequencies in the Sahelian *An. coluzzii,* compared with the Ngoussou. For example, the pyrethroid resistance *CYP4J5*-L43F marker (74) was found fixed across the Sahel. The G280S/G119S *ace*-1 mutation, found in high frequencies across Sahel confers organophosphate an carbamates resistance (75), and is shown recently to confer resistance to pirimiphos-methyl, in *An. coluzzii/gambiae* (76). Several mutations found within

887 the VGSC have recently been described/validated. For example, the resistance mutation,
888 L995F (77) and the V402L/I15227T haplotypes have been observed across Africa (78).
889 Recently, the two mutations (V402L/I15227T) are described to be in tight linkage and
890 mutually exclusive to the classical L995F/S mutations (77). Our results suggest haplotypes
891 carrying the V402L/I1527T combination plus the L995 replacement do exist in the Sahel *An.*
892 *coluzzii*. Not only that, it is also in addition to the N1570Y replacement. However, isolation
893 of Ngoussou colony over generations in the artificial insectary conditions could have led to
894 genetic isolation and drift, from inbreeding, leading to strong population differentiation in
895 comparison to the natural populations studied here.

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

903

904 **Cuticular resistance mechanism probably playing a key role in Sahelian *An.***
905 ***coluzzii***

906 Our results suggest cuticular mechanism play a role in pyrethroid resistance in these
907 populations. For example, the findings of *CYP4G16* (AGAP001076) and *CYP4G17*

908 (AGAP000877) overexpressed across the Sahel. The former P450 was previously shown to be
909 involved in epicuticular hydrocarbon biosynthesis associated with resistance (80).

910 The three major classes of the insect cuticular proteins - the CPR, CPLC and CPAP, were
911 found overrepresented in the top overexpressed genes across the Sahel, with the commonly
912 upregulated ones being *CPR76*, *CPR15* and *CPR30*, a chitin-binding cuticular protein,
913 *CPAP3-A1b*, and cuticular proteins of low complexity, *CPLCX3* and *CPLCA1*. Indeed,
914 *CPAP3-A1b* (AGAP000987) have been shown to be highly overexpressed in deltamethrin-
915 resistant Sahelian population of *An. coluzzii* from Burkina Faso (9) and induced by blood
916 feeding in *An. gambiae* (81). The CPR and CPLC cuticular proteins have been described to
917 potentially play a crucial role in insecticide resistance through leg cuticle
918 remodelling/thickening, regulating penetration rate of insecticides in *An. coluzzii* (71).
919 Furthermore, a recent study has found most of the cuticular proteins we have described
920 here, as highly overexpressed in permethrin/malathion resistant populations of Ethiopian
921 *An. arabiensis* (82), e.g., *CPR30*, *CPR75*, *CPR81*, and *CPLCP11*. Out of the several chitinases
922 overexpressed across the Sahel, four were amongst the top 50 most overexpressed metabolic
923 genes. These include the *Cht24* and *Cht6* that have been shown to be overexpressed in the
924 *An. arabiensis* from the above study (82), with the ortholog of *Cht24*, AARA007329 among the
925 top 10 most overexpressed genes in *An. arabiensis*, in line with our observations in *An.*
926 *coluzzii* across the Sahel. There is an overwhelming need to functionally investigate the
927 role/contribution towards insecticide resistance of these cuticular proteins, chitinases and a
928 chitin synthase (AGAP001748) significantly overexpressed in the field *An. coluzzii* from
929 Nigeria and Chad.

930

931 **Insecticide resistance- and thermotolerance-associated genes sit within**
932 **chromosomal inversions**

933 In this study, the findings of high frequency of 2La, 2Rb and 2Rc inversion polymorphisms
934 in the populations of *An. coluzzii*, compared with the Ngoussou, suggested strong
935 phenotypic adaptations in this species, across the Sahel. Most importantly, in addition to
936 several of the cuticular protein genes associated with resistance (chitinases, chitin synthase,
937 CPR, CPLC and CPAP proteins), several other genes previously implicated in
938 thermotolerance and/or desiccation resistance in *An. gambiae/coluzzii,* and which were highly
939 overexpressed in this study sit within these inversions. For example, the heat shock proteins,
940 *hsp83* (AGAP006958) and *hsp90* hptG (AGAP006961), both of which are known to be heat-
941 and insecticide-stress inducible (83) and were among the core set of hsp genes involved in a
942 common and immediate response to thermal stress in *An. gambiae* populations (84), sit
943 within the 2La inversion. These two genes were among the overexpressed genes in both
944 heat-hardened and permethrin-resistant *An. coluzzii* populations from northern Nigeria (57).
945 Several ionotropic glutamate receptors were found within the 2La inversion breakpoints:
946 *IR136* (AGAP006440), *IR139* (AGAP006691) and *IR142* (AGAP006407). This is not surprising
947 as ionotropic receptors are commonly associated with chemosensation, thermosensation,
948 and hygrosensation (85, 86), characteristics
949 which can confer adaptive advantages in xeric environs. The *IR25a* (AGAP010272) and *IR21*
950 *a* (AGAP008511) are known to mediate both humidity and temperature preference in the
951 fruit fly, *D. melanogaster* (86, 87), in addition to *IR21a* driving heat seeking and heat-
952 stimulated blood feeding in *An. gambiae* (87). These two genes have been shown to be

953 overexpressed/induced in thermotolerant/permethrin-resistant populations of *An. coluzzii*
954 (57).

955

956 ## Conclusions

957 Information on molecular basis of resistance and/or resistance genes and its markers
958 facilitates evidence-based control measures. In this study we characterised a major malaria
959 vector, *An. coluzzii* from the Sahel region of four countries, with findings which could
960 promote evidence-based, cross-border policy towards local and regional malaria control.
961 The study found that across Sahel (where malaria is highly seasonal, reaching its peak in the
962 rainy season), *An. coluzzii* is a dominant vector. And that a handful of common cross-
963 resistance genes are responsible for multiple insecticide resistance in this species. Findings
964 from this study suggest pleotropic role of some key genes – able to confer insecticide
965 resistance and/or stabilize the insecticide resistance gene, at the same time conferring
966 environmental adaptations, such as the ability to survive thermal stress (thermotolerance),
967 as expected in this Sahelian region. From operational vector control perspective this study
968 provided evidence of the role of key insecticide metabolism gene, *CYP6Z2* in increasing
969 insecticidal potency of clothianidin, which could increase the efficacy of the ingredients in
970 malaria control tools, when targeting field populations overexpressing this key P450.

971

972

973

Declarations

Ethics approval and consent to participate

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

Consent for publication

Not Applicable.

Availability of data and materials

The dataset(s) supporting the conclusions of this article are available in the European Nucleotide Archive, with accession PRJEB51644, and secondary accession of ERP136291, for the RNA-seqe raw sequence reads. cDNA sequences of *GSTe2* and *CYP6Z2* were deposited in GenBank (accession numbers: ON169006 - ON169030 for *GSTe2* and ON169031 - ON169050 for *CYP6Z2*) and 5′-UTR DNA fragment sequences were deposited in GenBank (accession numbers: ON169051 - ON169140).

Competing interests

The authors declare that they have no competing interests.

994 **Funding**

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

1000

1001 **Authors' contributions**

1002 Conceived and designed by SSI and CSW. SSI carried out the molecular analyses, with
1003 support from AM, LMJM, EIP and HI. ANF and MMM, participated in field collection of
1004 mosquitoes in Nigeria, Chad and Cameroon. SSI carried out data analysis with the support
1005 of JH, GDW, SCN for the RNAseq component. SSI wrote the manuscript with inputs from
1006 CSW, JH and SCN. All authors contributed to corrections of the final draft and approved
1007 final version of the manuscript.

1008

1009 **Acknowledgements**

1010 Not Applicable

1011

1012

1013

Reference

1. Bhatt S, Weiss D, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526(7572):207.
2. WHO. Global Technical Strategy for Malaria 2016-2030. World Health Organisation. . Geneva, Switzerland; 2015 May 2015. Report No.: ISBN 978 92 4 156499 1.
3. WHO. World Malaria Report. World Health Organization. . Geneva Switzerland; 2018. Report No.: Licence: CC BY-NC-SA 3.0 IGO Contract No.: ISBN 978-92-4-156565-3.
4. WHO. World Malaria Report. Geneva: Switzerland; 2021. Report No.: 978-92-4-004049-6 Contract No.: 978-92-4-004049-6
5. WHO. World Malaria Report 2019. Geneva, Switzerland; 2019. Report No.: 978-92-4-156572-1.
6. WHO. Update on the E-2020 initiative of 21 malaria-eliminating countries: report and country briefs. Geneva. Switzerland: World Health Organization; 2018. Report No.: WHO/CDS/GMP/2018.10.
7. Ibrahim SS, Mukhtar MM, Datti JA, Irving H, Kusimo MO, Tchapga W, et al. Temporal escalation of Pyrethroid Resistance in the major malaria vector Anopheles coluzzii from Sahelo-Sudanian Region of northern Nigeria. Scientific reports. 2019;9(1):7395.
8. Riveron JM, Huijben S, Tchapga W, Tchouakui M, Wondji MJ, Tchoupo M, et al. Escalation of Pyrethroid Resistance in the Malaria Vector Anopheles funestus Induces a Loss of Efficacy of Piperonyl Butoxide-Based Insecticide-Treated Nets in Mozambique. J Infect Dis. 2019;220(3):467-75.
9. Toe KH, N'Fale S, Dabire RK, Ranson H, Jones CM. The recent escalation in strength of pyrethroid resistance in Anopheles coluzzi in West Africa is linked to increased expression of multiple gene families. BMC Genomics. 2015;16:146.
10. Weedall GD, Mugenzi LMJ, Menze BD, Tchouakui M, Ibrahim SS, Amvongo-Adjia N, et al. A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy. Sci Transl Med. 2019;11(484).
11. Mugenzi LMJ, Menze BD, Tchouakui M, Wondji MJ, Irving H, Tchoupo M, et al. Cis-regulatory CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles funestus. Nat Commun. 2019;10(1):4652.
12. Barnes KG, Weedall GD, Ndula M, Irving H, Mzihalowa T, Hemingway J, et al. Genomic Footprints of Selective Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors Are Driven by Scale up of Insecticide-Based Vector Control. PLoS Genet. 2017;13(2):e1006539.
13. Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, et al. An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. PLoS Genet. 2020;16(6):e1008822.
14. Bruce-Chwatt LJ. Malaria in Nigeria. Bulletin of the World Health Organization. 1951;4(3):301-27.
15. Anopheles gambiae Genomes C, Data analysis g, Partner working g, Sample c-A, Burkina F, Cameroon, et al. Genetic diversity of the African malaria vector Anopheles gambiae. Nature. 2017;552(7683):96-100.
16. Muller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-caught permethrin-resistant Anopheles gambiae overexpress CYP6P3, a P450 that metabolises pyrethroids. PLoS Genet. 2008;4(11):e1000286.
17. Yunta C, Hemmings K, Stevenson B, Koekemoer LL, Matambo T, Pignatelli P, et al. Cross-resistance profiles of malaria mosquito P450s associated with pyrethroid resistance against WHO insecticides. Pestic Biochem Physiol. 2019;161:61-7.
18. Stevenson BJ, Bibby J, Pignatelli P, Muangnoicharoen S, O'Neill PM, Lian LY, et al. Cytochrome P450 6M2 from the malaria vector Anopheles gambiae metabolizes pyrethroids:

Sequential metabolism of deltamethrin revealed. Insect biochemistry and molecular biology. 2011;41(7):492-502.

19. Mitchell SN, Stevenson BJ, Müller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae from Ghana. Proceedings of the National Academy of Sciences. 2012;109(16):6147-52.

20. Mitchell SN, Rigden DJ, Dowd AJ, Lu F, Wilding CS, Weetman D, et al. Metabolic and target-site mechanisms combine to confer strong DDT resistance in Anopheles gambiae. PLoS One. 2014;9(3):e92662.

21. Lucas ER, Rockett KA, Lynd A, Essandoh J, Grisales N, Kemei B, et al. A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in Anopheles gambiae. Sci Rep. 2019;9(1):13335.

22. Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol. 2014;15(2):R27.

23. Menze BD, Kouamo MF, Wondji MJ, Tchapga W, Tchoupo M, Kusimo MO, et al. An Experimental Hut Evaluation of PBO-Based and Pyrethroid-Only Nets against the Malaria Vector Anopheles funestus Reveals a Loss of Bed Nets Efficacy Associated with GSTe2 Metabolic Resistance. Genes (Basel). 2020;11(2).

24. Coluzzi M, Sabatini A, Petrarca V, Di Deco MA. Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans R Soc Trop Med Hyg. 1979;73(5):483-97.

25. Cairns M, Roca-Feltrer A, Garske T, Wilson AL, Diallo D, Milligan PJ, et al. Estimating the potential public health impact of seasonal malaria chemoprevention in African children. Nat Commun. 2012;3:881.

26. Ibrahim SS, Mukhtar MM, Irving H, Labbo R, Kusimo MO, Mahamadou I, et al. High Plasmodium infection and multiple insecticide resistance in a major malaria vector Anopheles coluzzii from Sahel of Niger Republic. Malaria journal. 2019;18(1):181.

27. Fadel AN, Ibrahim SS, Tchouakui M, Terence E, Wondji MJ, Tchoupo M, et al. A combination of metabolic resistance and high frequency of the 1014F kdr mutation is driving pyrethroid resistance in Anopheles coluzzii population from Guinea savanna of Cameroon. Parasit Vectors. 2019;12(1):263.

28. Ibrahim SS, Fadel AN, Tchouakui M, Terence E, Wondji MJ, Tchoupo M, et al. High insecticide resistance in the major malaria vector Anopheles coluzzii in Chad Republic. Infect Dis Poverty. 2019;8(1):100.

29. Mitchell SN, Stevenson BJ, Muller P, Wilding CS, Egyir-Yawson A, Field SG, et al. Identification and validation of a gene causing cross-resistance between insecticide classes in Anopheles gambiae from Ghana. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(16):6147-52.

30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-2.

31. Joshi N, Fass J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software]. 2011.

32. Aronesty E. ea-utils: Command-line tools for processing biological sequencing data. Durham, NC; 2011.

33. Love MI, Huber, W., & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(550).

34. Blighe K, S. Rana, and M. Lewis. . “EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling.". 2018.

35. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.

36. Alexa A. RJ. topGO: Enrichment Analysis for Gene Ontology. R package version 2.46.0. 2021 [

1114 37. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene
1115 ontology terms. PLoS One. 2011;6(7):e21800.
1116 38. Riveron JM, Ibrahim SS, Chanda E, Mzilahowa T, Cuamba N, Irving H, et al. The highly
1117 polymorphic CYP6M7 cytochrome P450 gene partners with the directionally selected CYP6P9a and
1118 CYP6P9b genes to expand the pyrethroid resistance front in the malaria vector Anopheles funestus
1119 in Africa. BMC Genomics. 2014;15(1):817.
1120 39. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method.
1121 Nature protocols. 2008;3(6):1101-8.
1122 40. Nagi CS, Oruni A, Weetman D, Donnelly J. M. RNA-Seq-Pop: Exploiting the sequence in RNA-
1123 Seq - a Snakemake workflow reveals patterns of insecticide resistance in the malaria vector
1124 *Anopheles gambiae*. 2022;2022.06.17.493894.
1125 41. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence
1126 data. Genetics. 1992;132(2):583-9.
1127 42. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, et al. Sequencing of 50 human exomes
1128 reveals adaptation to high altitude. Science. 2010;329(5987):75-8.
1129 43. Alistair Miles MFR, Peter Ralph, Nick Harding, Rahul Pisupati, and Summer Rae. cggh/scikit-
1130 allel: A Python package for exploring and analysing genetic variation data 2020 [cggh/scikit-allel:
1131 v1.3.1 (v.3.1). Zenodo.].
1132 44. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
1133 Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
1134 45. Love RR, Redmond SN, Pombi M, Caputo B, Petrarca V, Della Torre A, et al. In Silico
1135 Karyotyping of Chromosomally Polymorphic Malaria Mosquitoes in the Anopheles gambiae Complex.
1136 G3 (Bethesda). 2019;9(10):3249-62.
1137 46. Hall TA, editor BioEdit: a user-friendly biological sequence alignment editor and analysis
1138 program for Windows 95/98/NT. Nucleic acids symposium series; 1999.
1139 47. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics
1140 Analysis across Computing Platforms. Molecular biology and evolution. 2018;35(6):1547-9.
1141 48. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et
1142 al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular biology and
1143 evolution. 2017;34(12):3299-302.
1144 49. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, et al. MatInspector and
1145 beyond: promoter analysis based on transcription factor binding sites. Bioinformatics.
1146 2005;21(13):2933-42.
1147 50. Lynd A, Lycett GJ. Optimization of the Gal4-UAS system in an Anopheles gambiae cell line.
1148 Insect Mol Biol. 2011;20(5):599-608.
1149 51. Ibrahim SS, Riveron JM, Bibby J, Irving H, Yunta C, Paine MJI, et al. Allelic Variation of
1150 Cytochrome P450s Drives Resistance to Bednet Insecticides in a Major Malaria Vector. PLoS Genet.
1151 2015;11(10):e1005618.
1152 52. Riveron JM, Tchouakui M., Mugenzi L., Menze B. D., Mu-Chun C. and Wondji C. S. Insecticide
1153 Resistance in Malaria Vectors: An Update at a Global Scale. In: S. Manguin VD, editor. Towards
1154 Malaria Elimination - A Leap Forward. London: IntechOpen; 2018. p. 452.
1155 53. Shiff C. Integrated approach to malaria control. Clinical microbiology reviews.
1156 2002;15(2):278-93.
1157 54. Lindsay S, Parson, L. and Thomas, C. . Mapping the range and relative abundance of the two
1158 principal African malaria vectors, Anopheles gambiae sensu stricto and An. arabiensis, using climate
1159 data. Proceedings of the Royal Society of London B: Biological Science. 1998; 265((1399)): 847–54.
1160 55. Coetzee M, Craig, M. and Le Sueur, D. . Distribution of African malaria mosquitoes belonging
1161 to the Anopheles gambiae complex. Parasitology today. 2000;16(2):74-7.
1162 56. Ranson H, Abdallah H, Badolo A, Guelbeogo WM, Kerah-Hinzoumbe C, Yangalbe-Kalnone E,
1163 et al. Insecticide resistance in Anopheles gambiae: data from the first year of a multi-country study
1164 highlight the extent of the problem. Malar J. 2009;8:299.

1165 57. Ibrahim SS, Mukhtar MM, Muhammad A, Wondji CS. 2La Paracentric Chromosomal Inversion
1166 and Overexpressed Metabolic Genes Enhance Thermotolerance and Pyrethroid Resistance in the
1167 Major Malaria Vector Anopheles gambiae. Biology (Basel). 2021;10(6).
1168 58. Gimonneau G, Bouyer J, Morand S, Besansky NJ, Diabate A, Simard F. A behavioral
1169 mechanism underlying ecological divergence in the malaria mosquito Anopheles gambiae. Behav
1170 Ecol. 2010;21(5):1087-92.
1171 59. Dao A, Yaro AS, Diallo M, Timbine S, Huestis DL, Kassogue Y, et al. Signatures of aestivation
1172 and migration in Sahelian malaria mosquito populations. Nature. 2014;516(7531):387-90.
1173 60. Lehmann T, Weetman D, Huestis DL, Yaro AS, Kassogue Y, Diallo M, et al. Tracing the origin
1174 of the early wet-season Anopheles coluzzii in the Sahel. Evol Appl. 2017;10(7):704-17.
1175 61. Adolfi A, Poulton B, Anthousi A, Macilwee S, Ranson H, Lycett GJ. Functional genetic
1176 validation of key genes conferring insecticide resistance in the major African malaria vector,
1177 Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of
1178 America. 2019;116(51):25764-72.
1179 62. Tchouakui M, Riveron JM, Djonabaye D, Tchapga W, Irving H, Soh Takam P, et al. Fitness
1180 Costs of the Glutathione S-Transferase Epsilon 2 (L119F-GSTe2) Mediated Metabolic Resistance to
1181 Insecticides in the Major African Malaria Vector Anopheles Funestus. Genes (Basel). 2018;9(12).
1182 63. Chiu TL, Wen Z, Rupasinghe SG, Schuler MA. Comparative molecular modeling of Anopheles
1183 gambiae CYP6Z1, a mosquito P450 capable of metabolizing DDT. Proceedings of the National
1184 Academy of Sciences of the United States of America. 2008;105(26):8855-60.
1185 64. Yunta C, Grisales N, Nasz S, Hemmings K, Pignatelli P, Voice M, et al. Pyriproxyfen is
1186 metabolized by P450s associated with pyrethroid resistance in An. gambiae. Insect biochemistry and
1187 molecular biology. 2016;78:50-7.
1188 65. Lees RS, Ismail HM, Logan RAE, Malone D, Davies R, Anthousi A, et al. New insecticide
1189 screening platforms indicate that Mitochondrial Complex I inhibitors are susceptible to cross-
1190 resistance by mosquito P450s that metabolise pyrethroids. Sci Rep. 2020;10(1):16232.
1191 66. Chandor-Proust A, Bibby J, Regent-Kloeckner M, Roux J, Guittard-Crilat E, Poupardin R, et al.
1192 The central role of mosquito cytochrome P450 CYP6Zs in insecticide detoxification revealed by
1193 functional expression and structural modelling. Biochem J. 2013;455(1):75-85.
1194 67. Antonio-Nkondjio C, Poupardin R, Tene BF, Kopya E, Costantini C, Awono-Ambene P, et al.
1195 Investigation of mechanisms of bendiocarb resistance in Anopheles gambiae populations from the
1196 city of Yaounde, Cameroon. Malar J. 2016;15(1):424.
1197 68. Riveron JM, Ibrahim SS, Mulamba C, Djouaka R, Irving H, Wondji MJ, et al. Genome-Wide
1198 Transcription and Functional Analyses Reveal Heterogeneous Molecular Mechanisms Driving
1199 Pyrethroids Resistance in the Major Malaria Vector Anopheles funestus Across Africa. G3 (Bethesda).
1200 2017;7(6):1819-32.
1201 69. Vontas J, Grigoraki L, Morgan J, Tsakireli D, Fuseini G, Segura L, et al. Rapid selection of a
1202 pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. Proceedings of the
1203 National Academy of Sciences of the United States of America. 2018;115(18):4619-24.
1204 70. Hearn J, Tagne CD, Ibrahim SS, Tene-Fossog B, Mugenzi LJ, Irving H, et al. Multi-omics
1205 analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector
1206 Anopheles funestus in East Africa. bioRxiv 2022;2021.10.21.465247.
1207 71. Kefi M, Charamis J, Balabanidou V, Ioannidis P, Ranson H, Ingham VA, et al. Transcriptomic
1208 analysis of resistance and short-term induction response to pyrethroids, in Anopheles coluzzii legs.
1209 BMC Genomics. 2021;22(1):891.
1210 72. Dunse KM, Kaas, Q., Guarino, R.F., Barton, P.A., Craik, D.J. and Anderson, M.A. Molecular
1211 basis for the resistance of an insect chymotrypsin to a potato type II proteinase inhibitor.
1212 Proceedings of the National Academy of Sciences. 2010;107(34):15016-21.
1213 73. Hu HX, Zhou D, Ma L, Shen B, Sun Y, Zhu CL. Lipase is associated with deltamethrin
1214 resistance in Culex pipiens pallens. Parasitol Res. 2020;119(1):23-30.

1215 74. Weetman D, Wilding CS, Neafsey DE, Muller P, Ochomo E, Isaacs AT, et al. Candidate-gene
1216 based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing
1217 genetic variation in East African Anopheles gambiae. Sci Rep. 2018;8(1):2920.
1218 75. Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique
1219 mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. Insect
1220 Mol Biol. 2004;13(1):1-7.
1221 76. Grau-Bove X, Lucas E, Pipini D, Rippon E, van 't Hof AE, Constant E, et al. Resistance to
1222 pirimiphos-methyl in West African Anopheles is spreading via duplication and introgression of the
1223 Ace1 locus. PLoS Genet. 2021;17(1):e1009253.
1224 77. Williams J, Cowlishaw R, Sanou A, Ranson H, Grigoraki L. In vivo functional validation of the
1225 V402L voltage gated sodium channel mutation in the malaria vector An. gambiae. Pest Manag Sci.
1226 2022;78(3):1155-63.
1227 78. Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, et al. The genetic
1228 architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors
1229 Anopheles gambiae and Anopheles coluzzii. Mol Ecol. 2021;30(21):5303-17.
1230 79. Ding Y, Hawkes N, Meredith J, Eggleston P, Hemingway J, Ranson H. Characterization of the
1231 promoters of Epsilon glutathione transferases in the mosquito Anopheles gambiae and their
1232 response to oxidative stress. Biochem J. 2005;387(Pt 3):879-88.
1233 80. Balabanidou V, Kampouraki A, MacLean M, Blomquist GJ, Tittiger C, Juarez MP, et al.
1234 Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production
1235 in Anopheles gambiae. Proceedings of the National Academy of Sciences of the United States of
1236 America. 2016;113(33):9268-73.
1237 81. Vannini L, Augustine Dunn W, Reed TW, Willis JH. Changes in transcript abundance for
1238 cuticular proteins and other genes three hours after a blood meal in Anopheles gambiae. Insect
1239 biochemistry and molecular biology. 2014;44:33-43.
1240 82. Messenger LA, Impoinvil LM, Derilus D, Yewhalaw D, Irish S, Lenhart A. A whole
1241 transcriptomic approach provides novel insights into the molecular basis of organophosphate and
1242 pyrethroid resistance in Anopheles arabiensis from Ethiopia. Insect biochemistry and molecular
1243 biology. 2021;139:103655.
1244 83. Zhao LaJ, W.A. . Expression of heat shock protein genes in insect stress responses. .
1245 Invertebrate Survival Journal. 2012;9(1):93-101.
1246 84. Cassone BJ, Molloy MJ, Cheng C, Tan JC, Hahn MW, Besansky NJ. Divergent transcriptional
1247 response to thermal stress by Anopheles gambiae larvae carrying alternative arrangements of
1248 inversion 2La. Mol Ecol. 2011;20(12):2567-80.
1249 85. VanHook AM. Understanding hygrosensation: How flies sense changes in humidity. Science
1250 Signaling. 2016;9(430):ec127-ec.
1251 86. Enjin A, Zaharieva EE, Frank DD, Mansourian S, Suh GS, Gallio M, et al. Humidity Sensing in
1252 Drosophila. Curr Biol. 2016;26(10):1352-8.
1253 87. Greppi C, Laursen WJ, Budelli G, Chang EC, Daniels AM, van Giesen L, et al. Mosquito heat
1254 seeking is driven by an ancestral cooling receptor. Science. 2020;367(6478):681-4.

1255

1256 **Figure Titles**

1257 **Figure 1: A volcano plot of differentially expressed genes showing fold changes and**
1258 **levels of significance in R-S, C-S and R-C comparisons, for Nigeria and Niger populations**
1259 **of *An. coluzzii*.** The plot depicts the top upregulated and top downregulated genes for each

1260 comparison, with several genes including chymotrypsins, a malate dehudrogenrase,
1261 CYP450s, a glutathione S-transferase (*GSTe2*), and a lipase, commonly upregulated in R vs S
1262 and C vs S comparisons. R stands for the female, resistant mosquitoes (survivors of 0.05%
1263 deltamethrin, 24 h after exposure), C stands for control (mosquitoes from the same
1264 population and of the same age not exposed to deltamethrin, and S stands for susceptible
1265 (females fromfully insecticide susceptible colony, Ngoussou).

1266 **Figure 2: A volcano plot of differentially expressed genes showing fold changes and**
1267 **significance in R-S, C-S and R-C comparisons, for Chad and Cameroon populations of *An.***
1268 ***coluzzii*.** The plot depicts the top upregulated and the top downregulated genes for each
1269 comparison with several genes including chymotrypsins, a malate dehudrogenrase, an
1270 aquaporn, CYP450s (particulary *CYP6Z2*), a glutathione S-transferase (*GSTe2*) and a lipase,
1271 commonly upregulated in R vs S and C vs S comparisons. R stands for the female, resistant
1272 mosquitoes (survivors of 0.05% deltamethrin, 24 h after exposure), C stands for control
1273 (mosquitoes from the same population and of the same age not exposed to deltamethrin,
1274 and S stands for susceptible (females from fully insecticide susceptible colony, Ngoussou).

1275

1276 **Figure 3: Validation of candidate resistance genes.** qRT-PCR profile of twelve
1277 metabolic resistance genes (in R-S and C-S comparisons) from the Sahelian *An. coluzzii*. Each
1278 bar (data point) is analysis from three biological replicates and three technical replicates for
1279 each gene, with error bars representing standard deviations. Significant differences are also
1280 indicated with the symbols ϯ, †, † and ‡, for $p < 0.0001$, $p < 0.001$, $p < 0.01$ and $p < 0.05$,
1281 respectively.

1282 **Figure 4: Identification of resistance variants of interest.** A heatmap showing frequencies
1283 of the resistance variants (haplotypes) in some key genes of interest.

1284 **Figure 5: Characterization of intergenic region (5′ regulatory element) of *GSTe2*.** Results of
1285 dual-luciferase reporter assays of the promoter (intergenic region) constructs, showing
1286 progressive loss of activity following sequential deletion of the constructs.

1287 **Figure 6: Validation of the role of metabolic resistance genes in insecticide resistance.**
1288 Results of insecticide susceptibility bioassays with transgenic flies expressing *GSTe2* and
1289 *CYP6Z2*. **a.** α-cypermethrin and deltamethrin, **b.** permethrin, **c.** 3-phenoxybenzaldehyde and
1290 3-phenoxybenzyl alcohol, **d.** DDT and clothianidin.

1291

1292 **Additional Files**

1293 **Additional File 1**: Supplementary text for methods and results.

1294 **Additional File 2:** Common differentially expressed genes across the Sahel.

1295 **Additional File 3:** Population genetics analyses.

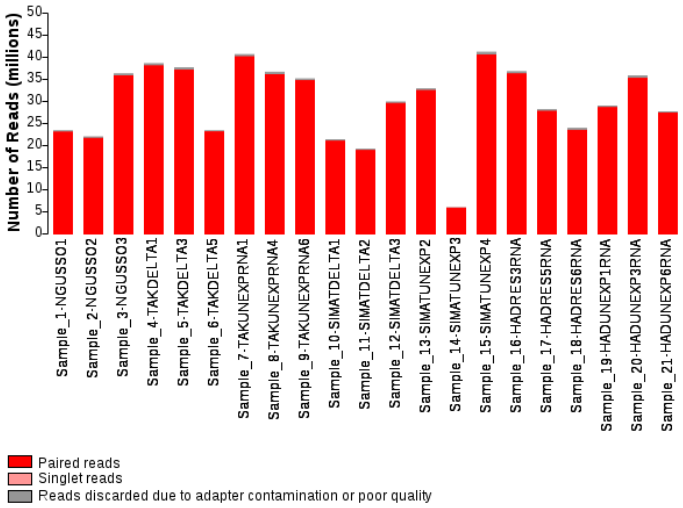
1296 **Additional File 4:** Genes within 2La, 2Rb and 2Rc inversion polymorphisms.

1297 **Additional Tables:** Additional tables, S1-S6.

1298 **Additional Figures:** Additional figures, S1-S11.

Supplementary Figures

Figure S1: Sampling sites in the Sahel region of the four sampled countries.



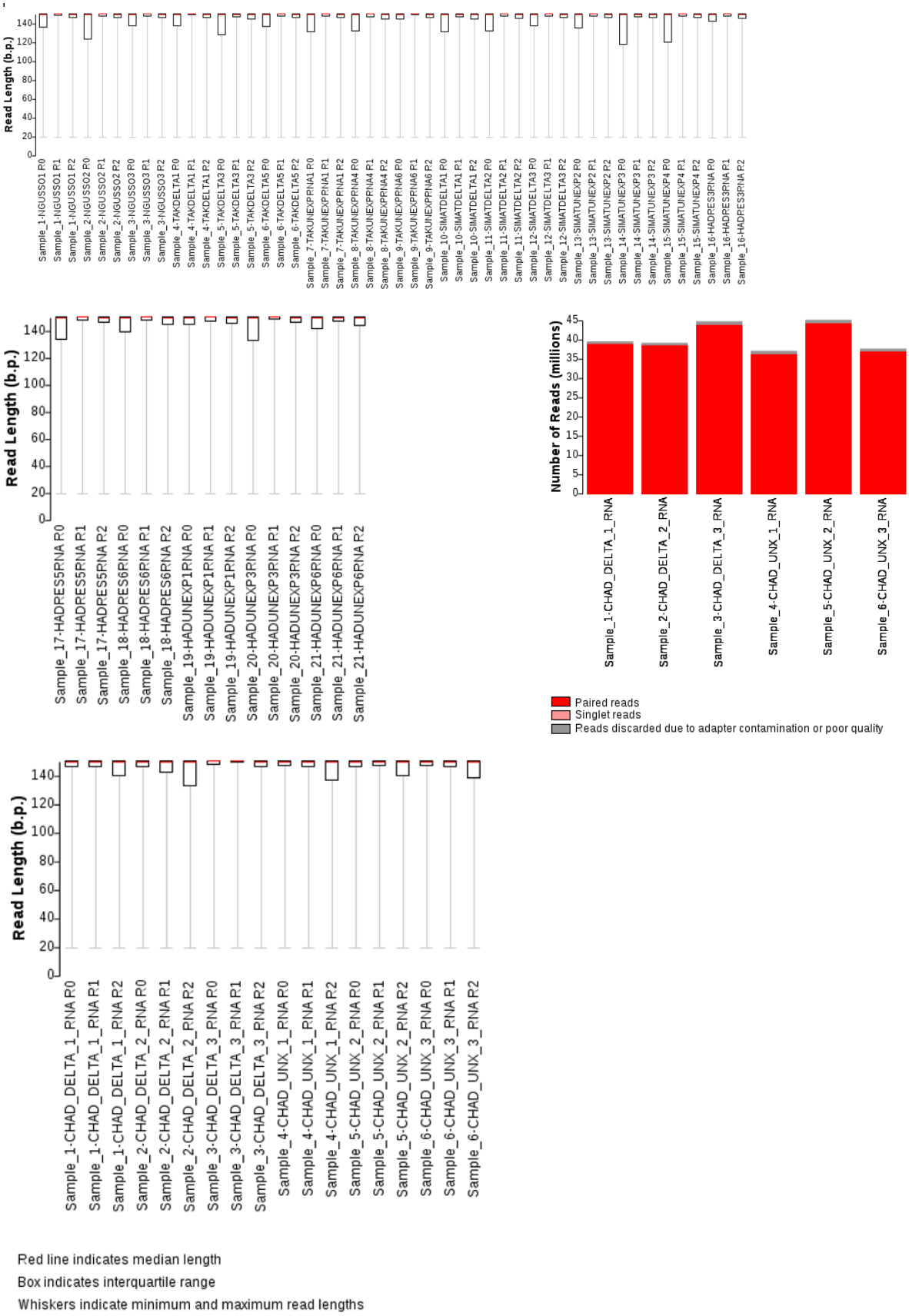


Figure S2: Summary of total number of reads for each sample and distribution of trimmed read length for forward (R1) and reverse (R2) reads and reads unpaired after trimming (R0).

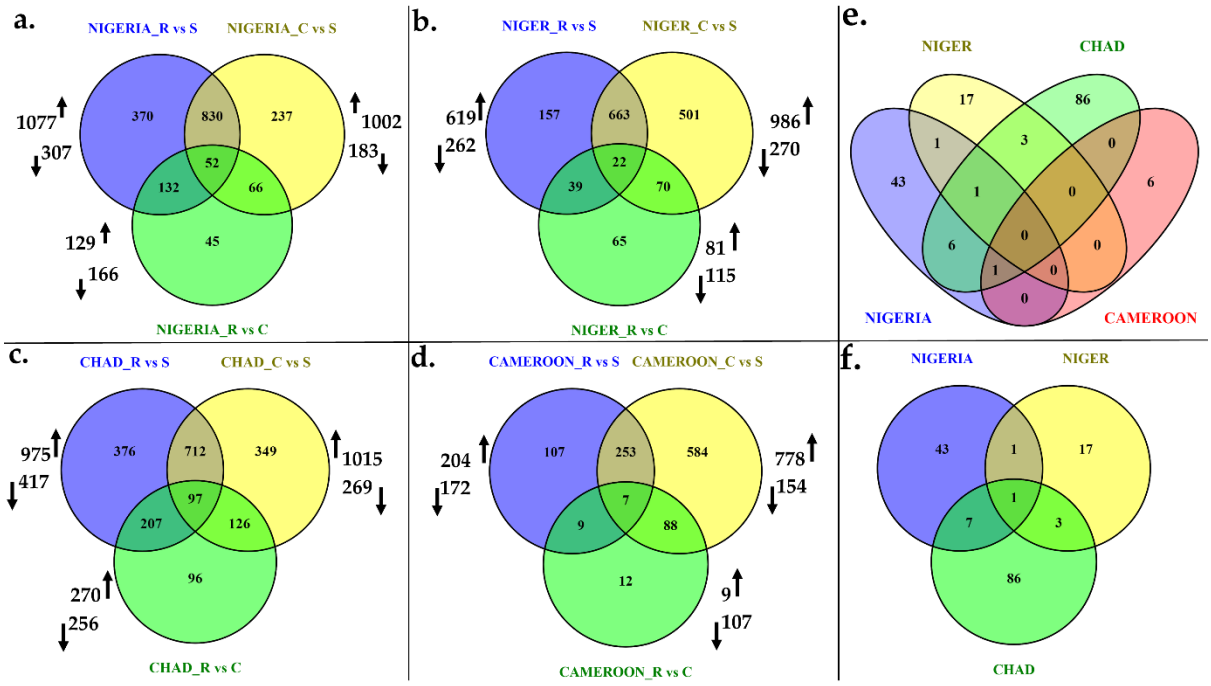


Figure S3: Venn diagrams summarizing the differentially expressed (DE) genes between resistant (R), unexposed (C) and susceptible (S) samples with a transcription ratio $\log_2 FC \geq 1$ in either direction, and a corrected p value < 0.05.

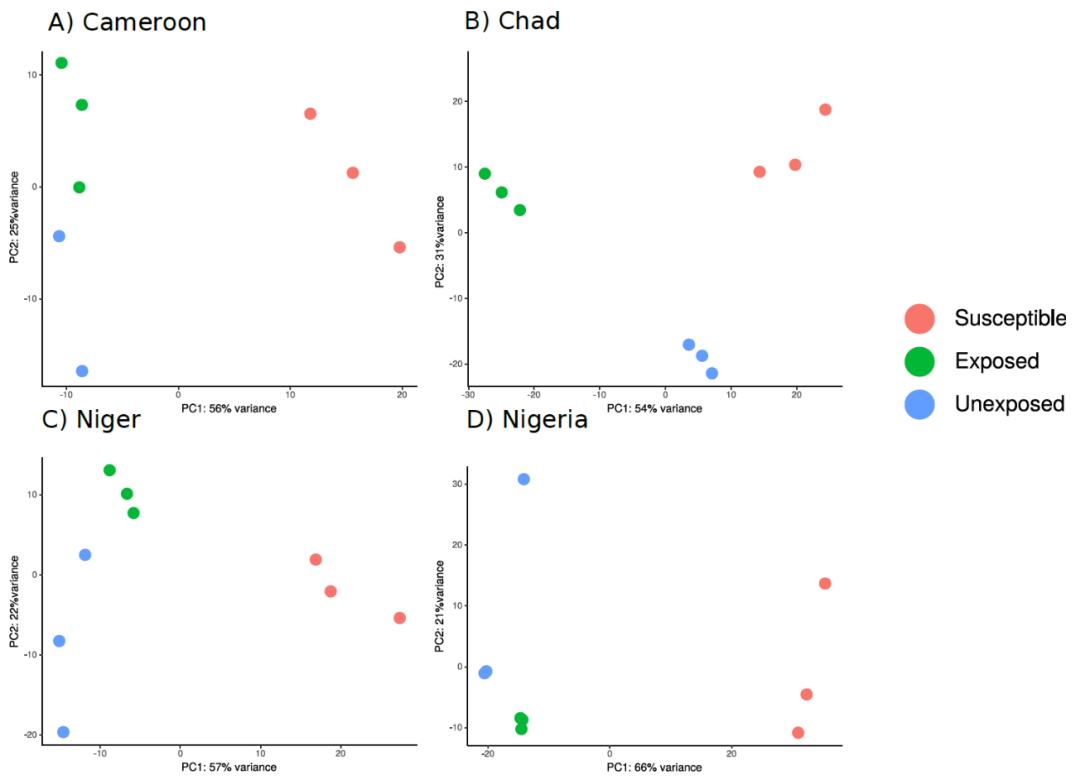


Figure S4: Principal component analysis of the 500 most variable genes from all experimental arms in the data from the four countries.

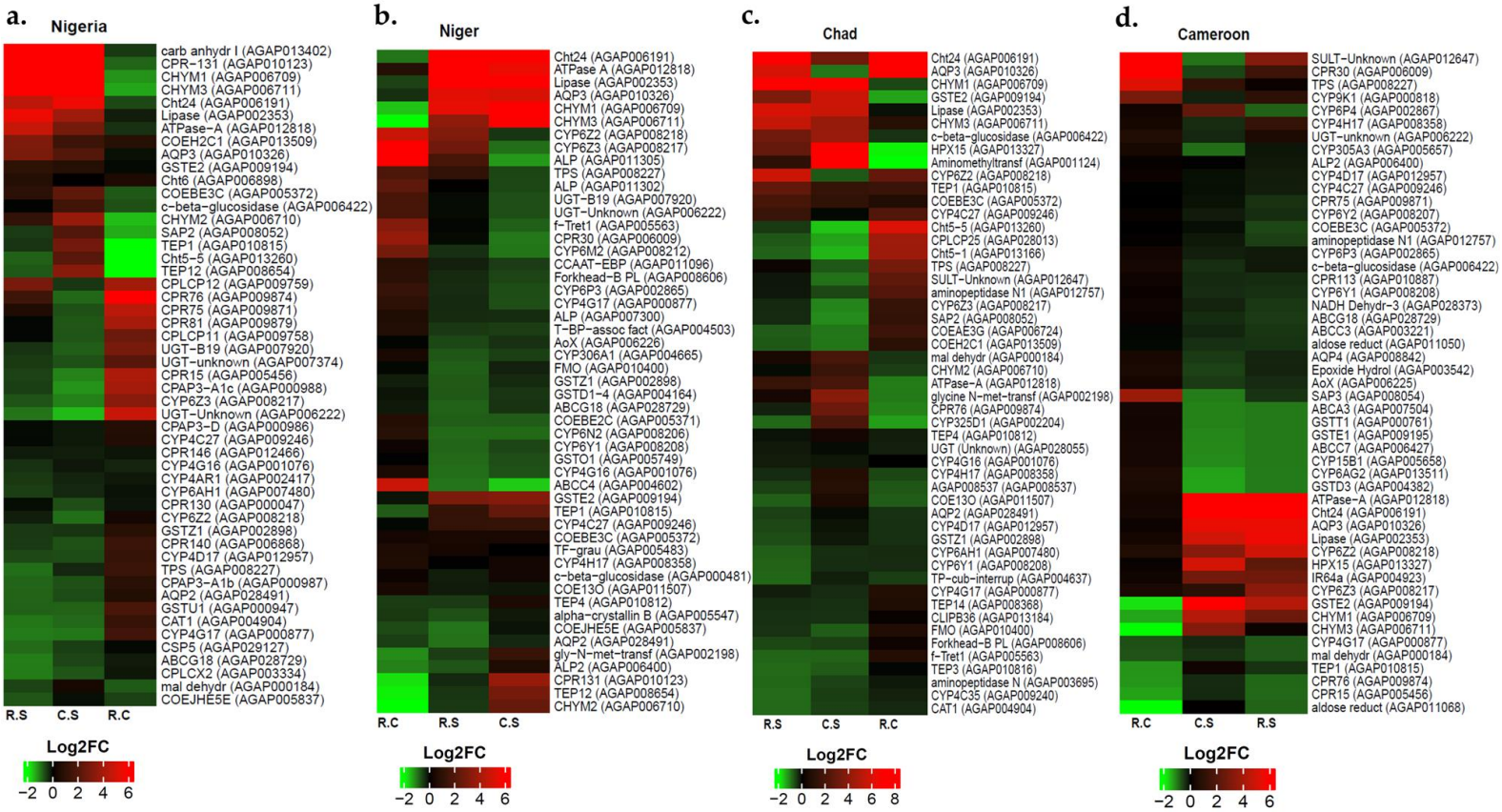
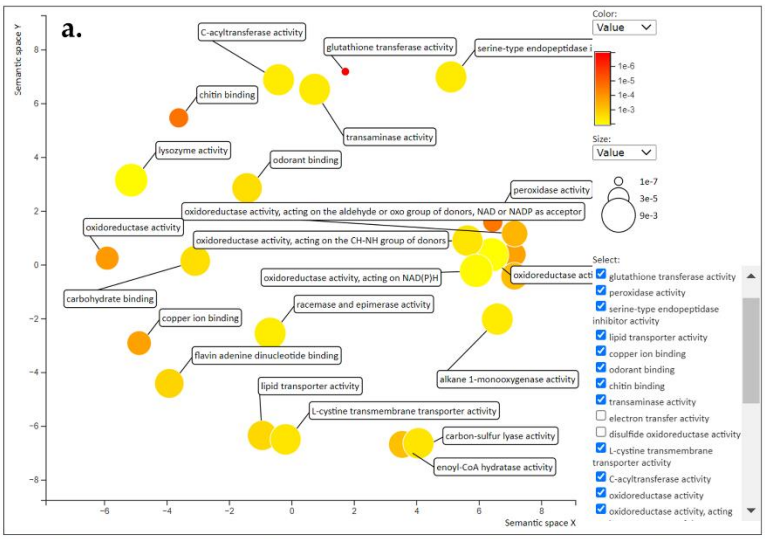
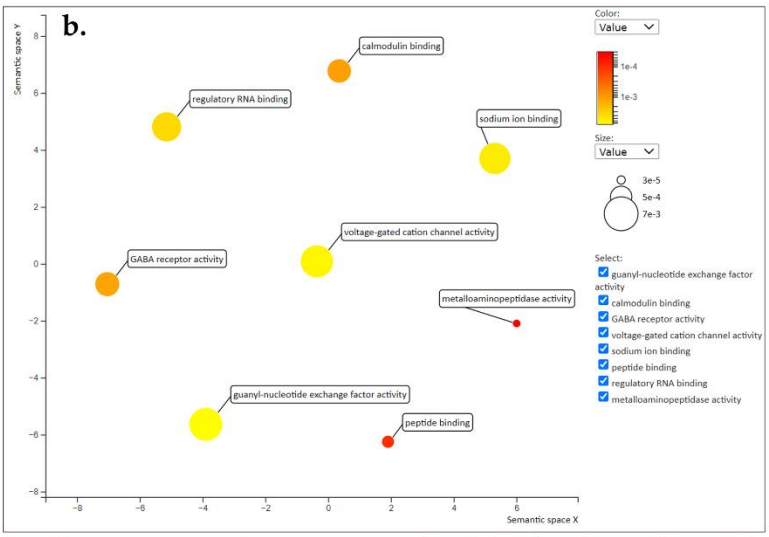


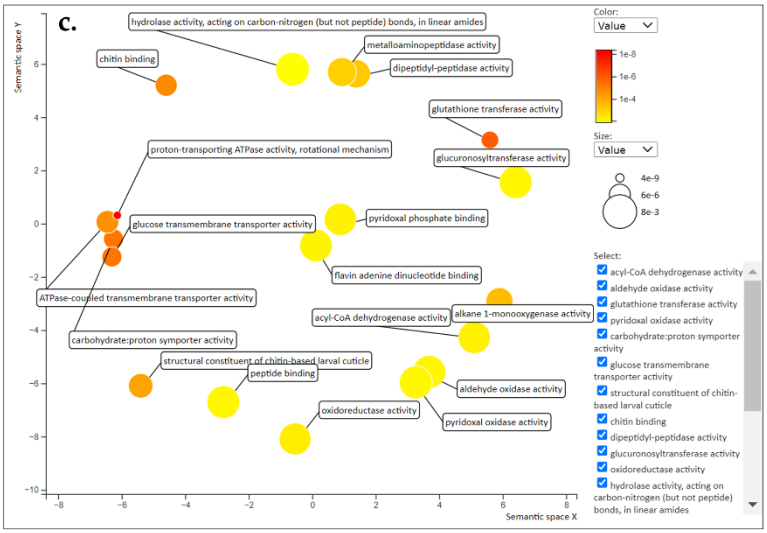
Figure S5: A heatmap showing the top 50 overexpressed genes in *An. coluzzii* populations from the Sahel region of each country. a. for Nigeria **b.**, **c.** and **d.** for Niger, Chad and Cameroon. Several genes including chymotrypsins, CYP450s, aquaporins, glutathione S-transferases and carboxylesterases are commonly upregulated across the Sahel of the four countries.

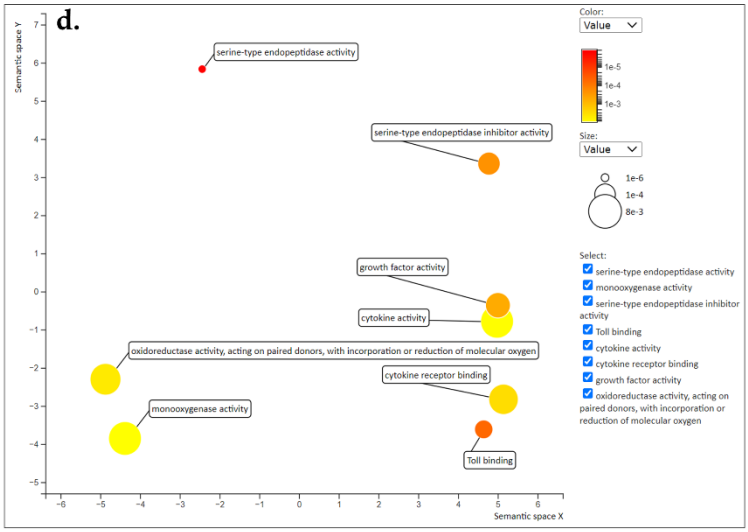




Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004867	serine-type endopeptidase inhibitor activity	0.121%		0.0039	0.91	0.00
GO:0015194	L-cystine transmembrane transporter activity	0.001%		0.0030	0.98	0.00
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1.855%		0.0095	0.80	0.00
GO:0030246	carbohydrate binding	0.797%		0.0023	0.98	0.00
GO:0003796	lysozyme activity	0.043%		0.0078	0.93	0.03
GO:0016408	C-acyltransferase activity	0.080%		0.0040	0.95	0.03
GO:0016846	carbon-sulfur lyase activity	0.134%		0.0030	0.94	0.03
GO:0016854	racemase and epimerase activity	0.375%		0.0042	0.93	0.03
GO:0005549	odorant binding	0.051%		0.0023	0.99	0.04
GO:0008061	chitin binding	0.116%		0.0000	0.98	0.05
GO:0016491	oxidoreductase activity	12.887%		0.0001	0.95	0.05
GO:0050660	flavin adenine dinucleotide binding	1.631%		0.0014	0.97	0.06
GO:0005507	copper ion binding	0.292%		0.0002	0.98	0.14
GO:0008483	transaminase activity	0.728%		0.0039	0.94	0.20
GO:0004364	glutathione transferase activity	0.086%		0.0000	0.95	0.20
GO:0018685	alkane 1-monooxygenase activity	0.001%		0.0040	0.88	0.23
GO:0005319	lipid transporter activity	0.146%		0.0017	0.98	0.30
GO:0015036	disulfide oxidoreductase activity	0.326%		0.0005	0.83	0.37
GO:0016645	oxidoreductase activity, acting on the CH-NH group of donors	0.350%		0.0030	0.83	0.37
GO:0004601	peroxidase activity	0.427%		0.0000	0.77	0.38

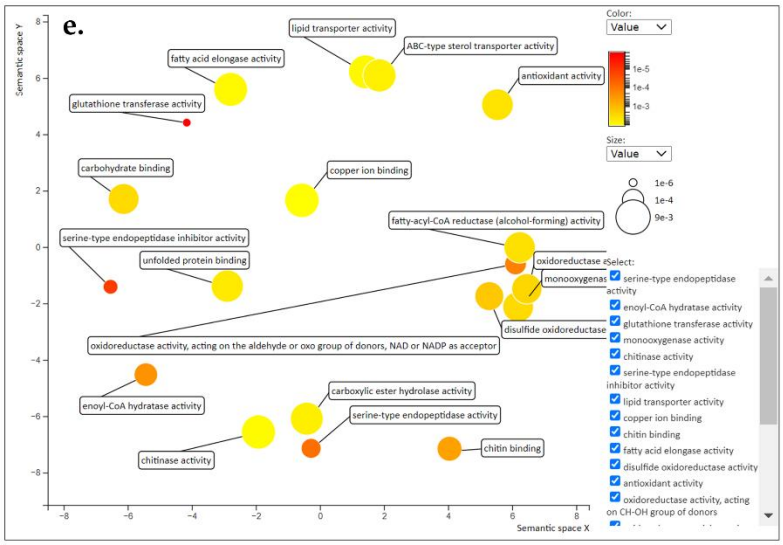
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0005085	guanyl-nucleotide exchange factor activity	0.309%		0.0070	0.90	0.00
GO:0016917	GABA receptor activity	0.027%		0.0011	0.95	0.00
GO:0022843	voltage-gated cation channel activity	0.155%		0.0060	0.43	0.00
GO:0031402	sodium ion binding	0.009%		0.0048	0.91	0.00
GO:0070006	metalloaminopeptidase activity	0.153%		0.0000	1.00	0.00
GO:0061980	regulatory RNA binding	0.013%		0.0031	0.99	0.03
GO:0005516	calmodulin binding	0.135%		0.0009	0.99	0.03
GO:0042277	peptide binding	0.100%		0.0001	0.99	0.04
GO:0008273	calcium, potassium:sodium antiporter activity	0.006%		0.0048	0.49	0.47
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0023	0.53	0.57
GO:0099095	ligand-gated anion channel activity	0.006%		0.0008	0.49	0.64
GO:0022848	acetylcholine-gated cation-selective channel activity	0.024%		0.0000	0.43	0.70
GO:0005283	amino acid:sodium symporter activity	0.045%		0.0031	0.45	0.72
GO:0005328	neurotransmitter:sodium symporter activity	0.001%		0.0000	0.58	0.73
GO:0008028	monocarboxylic acid transmembrane transporter activity	0.052%		0.0013	0.54	0.74
GO:0022839	ion gated channel activity	0.054%		0.0000	0.53	0.74
GO:0005254	chloride channel activity	0.123%		0.0000	0.45	0.80
GO:0005245	voltage-gated calcium channel activity	0.031%		0.0011	0.44	0.81
GO:0015171	amino acid transmembrane transporter activity	0.268%		0.0000	0.59	0.83
GO:0030955	potassium ion binding	0.068%		0.0048	0.91	0.86
GO:0005096	GTPase activator activity	0.247%		0.0003	0.90	0.87
GO:0005249	voltage-gated potassium channel activity	0.119%		0.0022	0.41	0.90

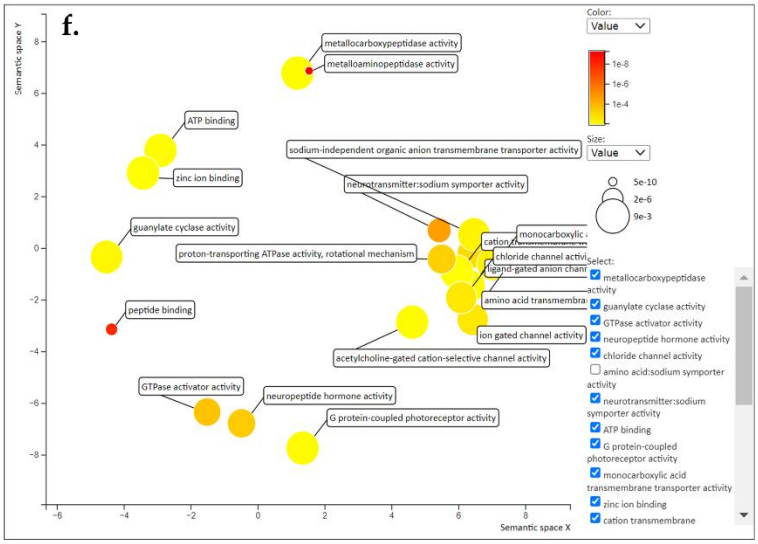




Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008010	structural constituent of chitin-based larval cuticle	0.002%		0.0000	1.00	0.00
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	0.737%		0.0084	0.93	0.00
GO:0042277	peptide binding	0.100%		0.0051	0.98	0.00
GO:0042626	ATPase-coupled transmembrane transporter activity	1.389%		0.0000	0.83	0.00
GO:0004732	pyridoxal oxidase activity	0.000%		0.0055	0.87	0.02
GO:0015020	glucuronosyltransferase activity	0.043%		0.0051	0.96	0.03
GO:0016491	oxidoreductase activity	12.887%		0.0032	0.97	0.05
GO:0050660	flavin adenine dinucleotide binding	1.631%		0.0044	0.96	0.05
GO:0008061	chitin binding	0.116%		0.0000	0.98	0.05
GO:0004364	glutathione transferase activity	0.086%		0.0000	0.96	0.16
GO:0003995	acyl-CoA dehydrogenase activity	0.152%		0.0038	0.92	0.16
GO:0018685	alkane 1-monooxygenase activity	0.001%		0.0002	0.93	0.20
GO:0070006	metalloaminopeptidase activity	0.153%		0.0006	0.89	0.23
GO:0030170	pyridoxal phosphate binding	1.138%		0.0038	0.96	0.31
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0000	0.86	0.33
GO:0008239	dipeptidyl-peptidase activity	0.046%		0.0004	0.89	0.65
GO:0005351	carbohydrate:proton symporter activity	0.026%		0.0000	0.80	0.71
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0000	0.83	0.71
GO:0004031	aldehyde oxidase activity	0.002%		0.0055	0.86	0.80

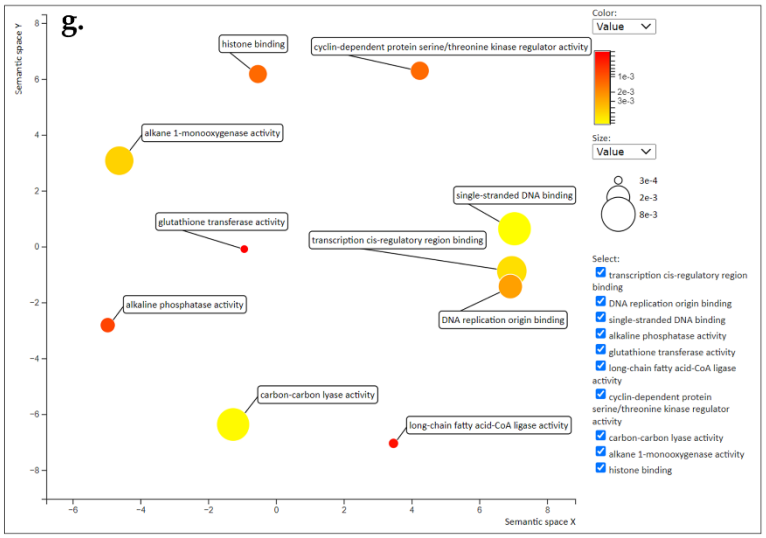
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004497	monooxygenase activity	1.211%		0.0081	0.88	0.00
GO:0005125	cytokine activity	0.080%		0.0079	0.38	0.00
GO:0004252	serine-type endopeptidase activity	0.893%		0.0000	0.98	0.04
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	1.285%		0.0040	0.88	0.41
GO:0005121	Toll binding	0.000%		0.0000	0.59	0.53
GO:0004867	serine-type endopeptidase inhibitor activity	0.121%		0.0002	0.69	0.58
GO:0005126	cytokine receptor binding	0.094%		0.0025	0.50	0.74
GO:0008083	growth factor activity	0.106%		0.0005	0.37	0.85

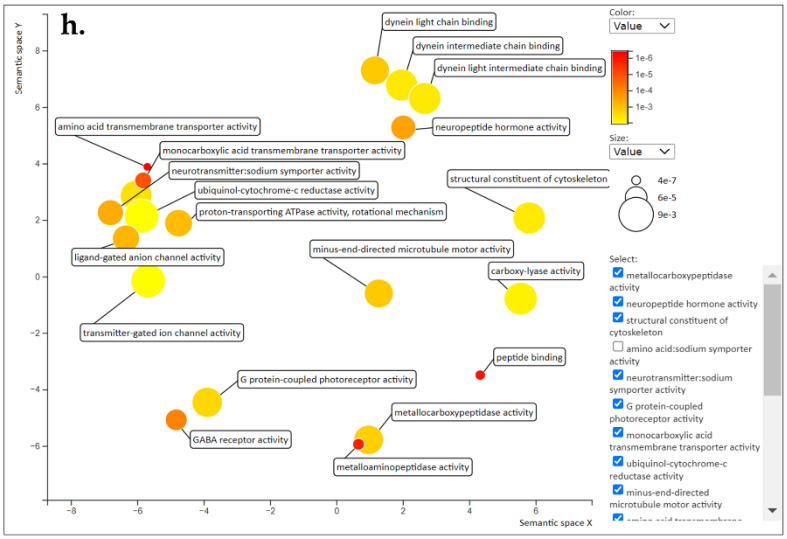




Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004568	chitinase activity	0.057%		0.0084	0.93	0.00
GO:0004867	serine-type endopeptidase inhibitor activity	0.121%		0.0000	1.00	0.00
GO:0005319	lipid transporter activity	0.146%		0.0076	0.94	0.00
GO:0005507	copper ion binding	0.292%		0.0093	0.99	0.00
GO:0016209	antioxidant activity	0.725%		0.0036	0.91	0.00
GO:0004601	peroxidase activity	0.427%		0.0008	0.73	0.91
GO:0080019	fatty-acyl-CoA reductase (alcohol-forming) activity	0.018%		0.0033	0.81	0.02
GO:0009922	fatty acid elongase activity	0.028%		0.0078	0.96	0.02
GO:0004300	enoyl-CoA hydratase activity	0.051%		0.0002	0.97	0.02
GO:0008061	chitin binding	0.116%		0.0004	0.99	0.04
GO:0051082	unfolded protein binding	0.424%		0.0045	0.98	0.05
GO:0030246	carbohydrate binding	0.797%		0.0027	0.98	0.05
GO:0004364	glutathione transferase activity	0.086%		0.0000	0.96	0.16
GO:0052689	carboxylic ester hydrolase activity	0.694%		0.0055	0.92	0.21
GO:0004252	serine-type endopeptidase activity	0.893%		0.0001	0.92	0.27
GO:0004497	monooxygenase activity	1.211%		0.0026	0.80	0.27
GO:0015036	disulfide oxidoreductase activity	0.326%		0.0014	0.82	0.35
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	2.116%		0.0023	0.79	0.43
GO:0034041	ABC-type sterol transporter activity	0.000%		0.0058	0.94	0.57
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	0.593%		0.0001	0.77	0.70

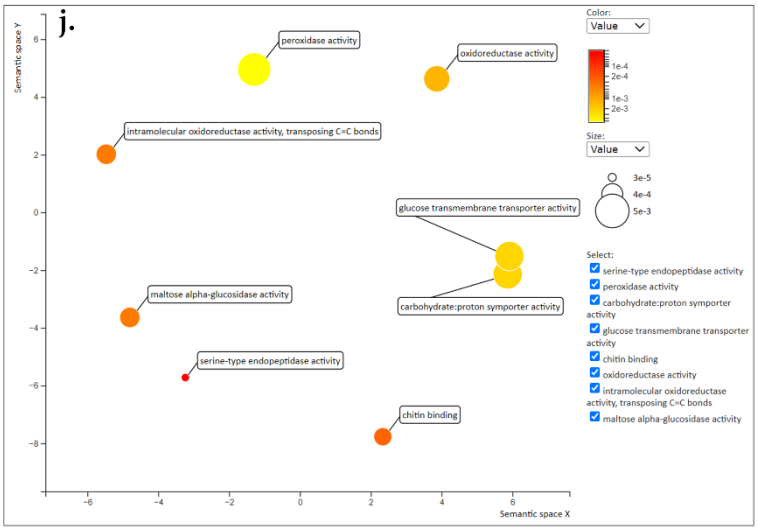
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004181	metallocarboxypeptidase activity	0.086%		0.0081	0.90	0.00
GO:0008270	zinc ion binding	3.638%		0.0092	0.96	0.00
GO:0008324	cation transmembrane transporter activity	2.405%		0.0090	0.53	0.00
GO:0008020	G protein-coupled photoreceptor activity	0.003%		0.0078	0.91	0.01
GO:0042277	peptide binding	0.100%		0.0000	0.98	0.05
GO:0004383	guanylate cyclase activity	0.024%		0.0059	0.97	0.10
GO:0005184	neuropeptide hormone activity	0.013%		0.0003	0.87	0.25
GO:0005524	ATP binding	13.644%		0.0081	0.96	0.27
GO:0005328	neurotransmitter:sodium symporter activity	0.001%		0.0000	0.63	0.43
GO:0015347	sodium-independent organic anion transmembrane transporter activity	0.001%		0.0048	0.60	0.44
GO:0005096	GTPase activator activity	0.247%		0.0002	0.91	0.54
GO:0022848	acetylcholine-gated cation-selective channel activity	0.024%		0.0078	0.49	0.56
GO:0008028	monocarboxylic acid transmembrane transporter activity	0.052%		0.0042	0.54	0.57
GO:0005254	chloride channel activity	0.123%		0.0073	0.51	0.63
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0005	0.56	0.65
GO:0022839	ion gated channel activity	0.054%		0.0022	0.63	0.65
GO:0005283	amino acid:sodium symporter activity	0.045%		0.0009	0.50	0.74
GO:0099095	ligand-gated anion channel activity	0.006%		0.0022	0.55	0.80
GO:0070006	metalloaminopeptidase activity	0.153%		0.0000	0.90	0.83
GO:0015171	amino acid transmembrane transporter activity	0.268%		0.0001	0.60	0.83



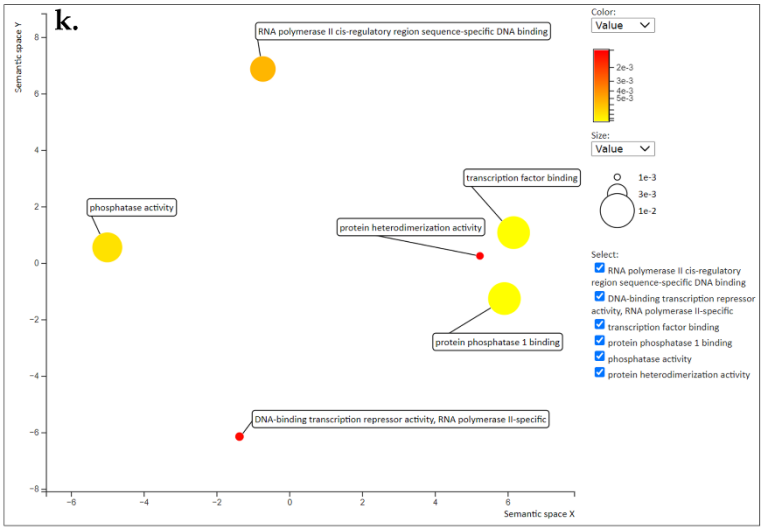


Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0003697	single-stranded DNA binding	0.255%		0.0079	0.86	0.00
GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	0.069%		0.0012	1.00	0.00
GO:0016830	carbon-carbon lyase activity	1.062%		0.0074	0.98	0.00
GO:0018685	alkane 1-monooxygenase activity	0.001%		0.0045	0.98	0.02
GO:0004035	alkaline phosphatase activity	0.021%		0.0008	0.98	0.03
GO:0004364	glutathione transferase activity	0.086%		0.0003	0.98	0.03
GO:0004467	long-chain fatty acid-CoA ligase activity	0.034%		0.0004	0.98	0.04
GO:0042393	histone binding	0.086%		0.0012	0.97	0.04
GO:0000976	transcription cis-regulatory region binding	0.373%		0.0053	0.75	0.33
GO:0003688	DNA replication origin binding	0.043%		0.0024	0.77	0.81

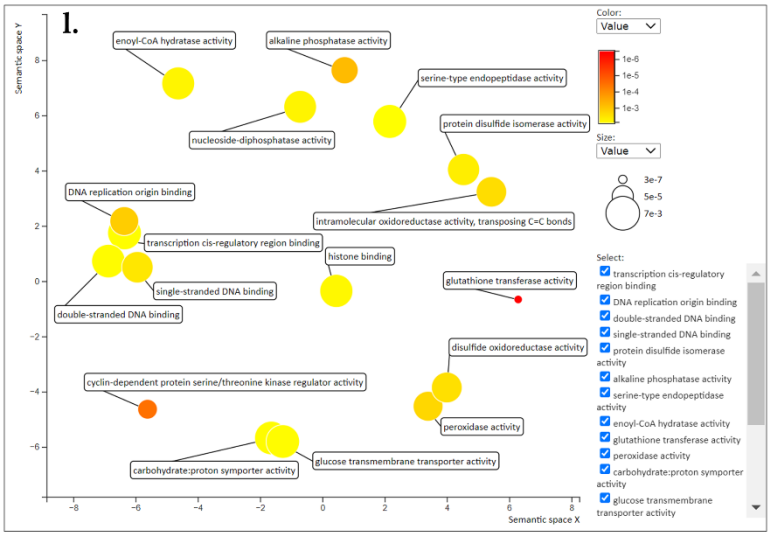
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008121	ubiquinol-cytochrome-c reductase activity	0.036%		0.0094	0.66	0.00
GO:0008569	minus-end-directed microtubule motor activity	0.040%		0.0011	0.96	0.00
GO:0051959	dynein light intermediate chain binding	0.026%		0.0039	0.90	0.00
GO:0005200	structural constituent of cytoskeleton	0.076%		0.0041	0.99	0.01
GO:0016831	carboxy-lyase activity	0.619%		0.0055	0.99	0.03
GO:0004181	metallocarboxypeptidase activity	0.086%		0.0014	0.91	0.03
GO:0042277	peptide binding	0.100%		0.0000	0.99	0.04
GO:0045503	dynein light chain binding	0.003%		0.0011	0.91	0.31
GO:0005184	neuropeptide hormone activity	0.013%		0.0002	0.84	0.34
GO:0045505	dynein intermediate chain binding	0.030%		0.0039	0.90	0.36
GO:0022824	transmitter-gated ion channel activity	0.062%		0.0092	0.55	0.45
GO:0022848	acetylcholine-gated cation-selective channel activity	0.024%		0.0002	0.54	0.94



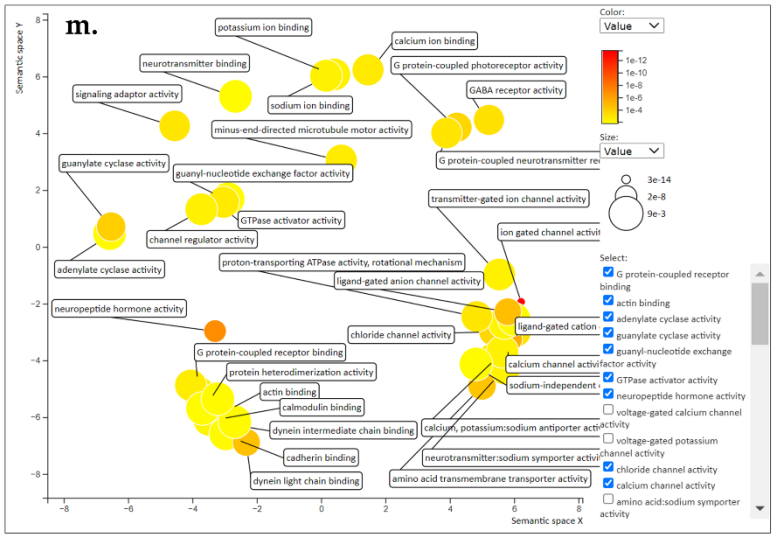
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004601	peroxidase activity	0.427%		0.0048	0.96	0.00
GO:0008061	chitin binding	0.116%		0.0002	1.00	0.00
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0021	0.80	0.01
GO:0032450	maltose alpha-glucosidase activity	0.006%		0.0004	0.93	0.02
GO:0016863	intramolecular oxidoreductase activity, transposing C=C bonds	0.050%		0.0004	0.97	0.03
GO:0016491	oxidoreductase activity	12.887%		0.0012	0.96	0.04
GO:0004252	serine-type endopeptidase activity	0.893%		0.0000	0.92	0.18
GO:0005351	carbohydrate:proton symporter activity	0.026%		0.0021	0.80	0.71



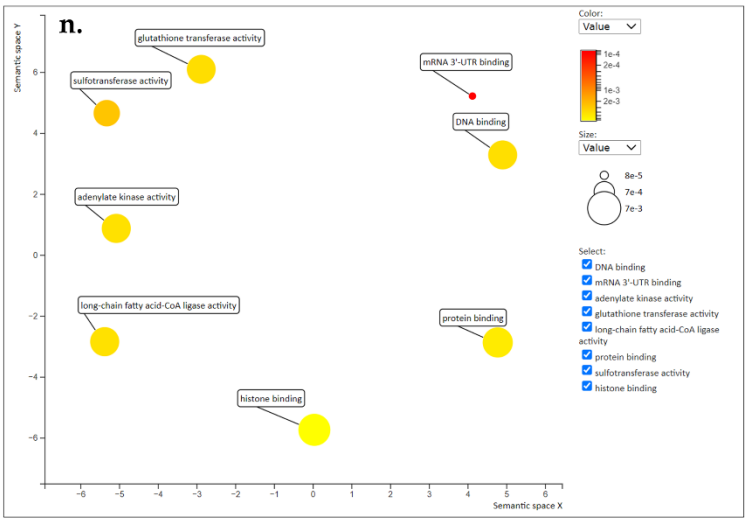
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0001227	DNA-binding transcription repressor activity, RNA polymerase II-specific	0.041%		0.0013	1.00	0.00
GO:0008157	protein phosphatase 1 binding	0.009%		0.0096	0.71	0.00
GO:0016791	phosphatase activity	1.292%		0.0076	1.00	0.00
GO:0000978	RNA polymerase II cis-regulatory region sequence-specific DNA binding	0.193%		0.0053	0.95	0.03
GO:0008134	transcription factor binding	0.111%		0.0096	0.68	0.36
GO:0046982	protein heterodimerization activity	0.275%		0.0012	0.67	0.46



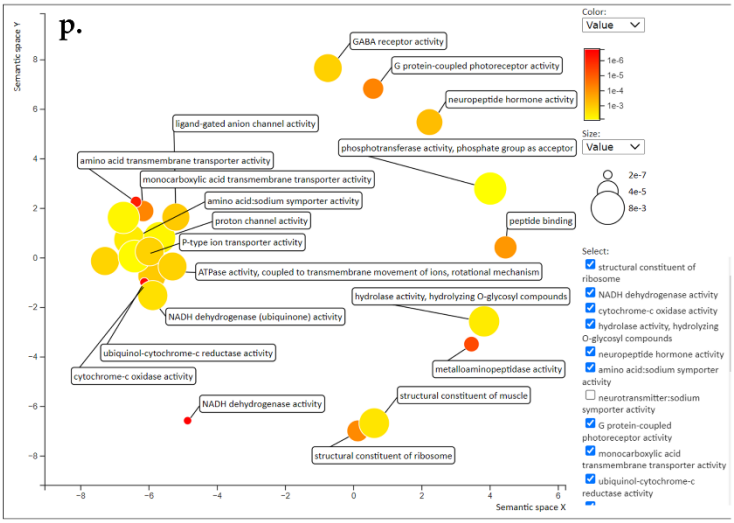
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0003690	double-stranded DNA binding	0.682%		0.0065	0.87	0.00
GO:0004252	serine-type endopeptidase activity	0.893%		0.0074	0.90	0.00
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0066	0.91	0.00
GO:0016538	cyclin-dependent protein serine/threonine kinase regulator activity	0.069%		0.0000	1.00	0.00
GO:0004300	enoyl-CoA hydratase activity	0.051%		0.0051	0.98	0.03
GO:0004364	glutathione transferase activity	0.086%		0.0000	0.98	0.03
GO:0015036	disulfide oxidoreductase activity	0.326%		0.0022	0.94	0.03
GO:0042393	histone binding	0.086%		0.0056	0.98	0.05
GO:0017110	nucleoside-diphosphatase activity	0.016%		0.0047	0.94	0.19
GO:0004035	alkaline phosphatase activity	0.021%		0.0005	0.94	0.20
GO:0003756	protein disulfide isomerase activity	0.028%		0.0038	0.86	0.31
GO:0004601	peroxidase activity	0.427%		0.0015	0.94	0.32
GO:0003697	single-stranded DNA binding	0.255%		0.0028	0.88	0.35
GO:0000976	transcription cis-regulatory region binding	0.373%		0.0071	0.82	0.36
GO:0016863	intramolecular oxidoreductase activity, transposing C=C bonds	0.050%		0.0019	0.90	0.69
GO:0005351	carbohydrate:proton symporter activity	0.026%		0.0066	0.91	0.71
GO:0003688	DNA replication origin binding	0.043%		0.0010	0.83	0.81



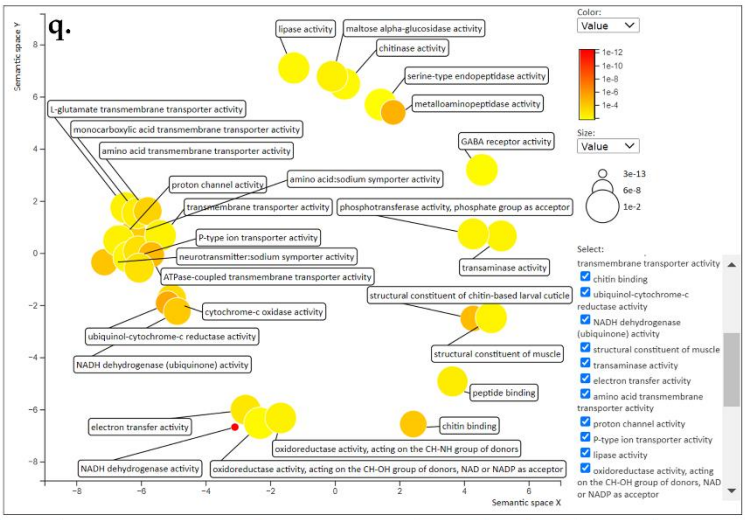
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0003779	actin binding	0.467%		0.0087	0.80	0.00
GO:0004016	adenylate cyclase activity	0.077%		0.0054	0.96	0.00
GO:0005085	guanyl-nucleotide exchange factor activity	0.309%		0.0057	0.88	0.00
GO:0008569	minus-end-directed microtubule motor activity	0.040%		0.0012	0.98	0.00
GO:0015347	sodium-independent organic anion transmembrane transporter activity	0.001%		0.0057	0.72	0.00
GO:0035591	signaling adaptor activity	0.023%		0.0006	1.00	0.00
GO:0042165	neurotransmitter binding	0.001%		0.0057	0.98	0.03
GO:0031402	sodium ion binding	0.009%		0.0021	0.93	0.04
GO:0099528	G protein-coupled neurotransmitter receptor activity	0.007%		0.0006	0.84	0.14
GO:0005509	calcium ion binding	1.117%		0.0010	0.95	0.21
GO:0099094	ligand-gated cation channel activity	0.080%		0.0050	0.58	0.36
GO:0045503	dynein light chain binding	0.003%		0.0000	0.85	0.37
GO:0005328	neurotransmitter:sodium symporter activity	0.001%		0.0025	0.70	0.40
GO:0045296	cadherin binding	0.018%		0.0052	0.83	0.42
GO:0051959	dynein light intermediate chain binding	0.026%		0.0052	0.83	0.43
GO:0045505	dynein intermediate chain binding	0.030%		0.0052	0.83	0.44
GO:0001664	G protein-coupled receptor binding	0.082%		0.0018	0.79	0.47
GO:0008020	G protein-coupled photoreceptor activity	0.003%		0.0001	0.88	0.48
GO:0016917	GABA receptor activity	0.027%		0.0004	0.90	0.48
GO:0005516	calmodulin binding	0.135%		0.0027	0.81	0.49
GO:0046982	protein heterodimerization activity	0.275%		0.0062	0.81	0.52
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0006	0.64	0.54



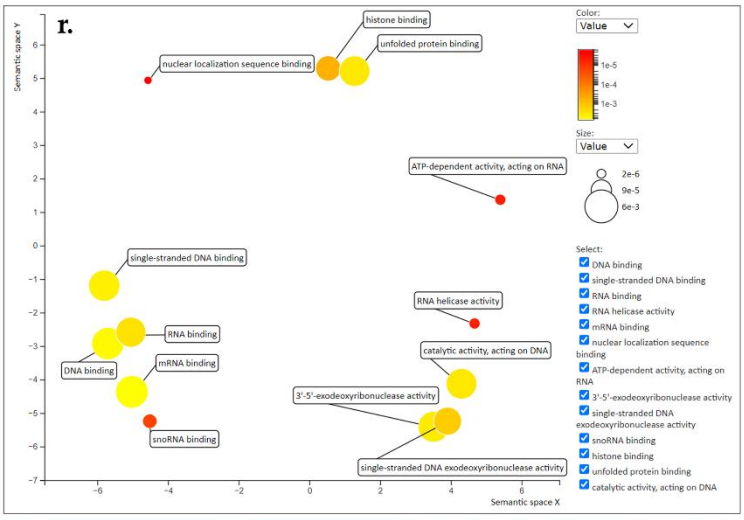
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004467	long-chain fatty acid-CoA ligase activity	0.034%		0.0039	0.93	0.00
GO:0042393	histone binding	0.086%		0.0066	0.96	0.00
GO:0004364	glutathione transferase activity	0.086%		0.0037	0.90	0.02
GO:0003730	mRNA 3'-UTR binding	0.026%		0.0001	0.91	0.04
GO:0005515	protein binding	4.557%		0.0047	0.94	0.05
GO:0008146	sulfotransferase activity	0.121%		0.0024	0.90	0.17
GO:0004017	adenylate kinase activity	0.059%		0.0039	0.85	0.20
GO:0003677	DNA binding	11.354%		0.0038	0.88	0.26



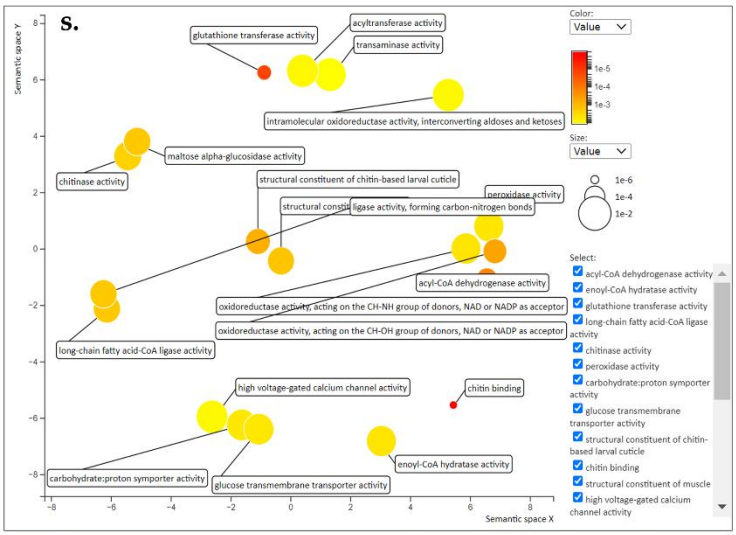
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008307	structural constituent of muscle	0.004%		0.0025	0.96	0.00
GO:0015662	P-type ion transporter activity	0.107%		0.0065	0.48	0.00
GO:0016776	phosphotransferase activity, phosphate group as acceptor	0.361%		0.0076	0.99	0.00
GO:0016917	GABA receptor activity	0.027%		0.0011	0.96	0.00
GO:0005184	neuropeptide hormone activity	0.013%		0.0005	0.97	0.01
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	1.613%		0.0033	0.97	0.03
GO:0042277	peptide binding	0.100%		0.0001	1.00	0.03
GO:0070006	metalloaminopeptidase activity	0.153%		0.0000	0.97	0.24
GO:0015217	ADP transmembrane transporter activity	0.020%		0.0051	0.57	0.43
GO:0008020	G protein-coupled photoreceptor activity	0.003%		0.0000	0.93	0.46
GO:0003735	structural constituent of ribosome	2.348%		0.0001	0.96	0.50
GO:0015252	proton channel activity	0.315%		0.0051	0.49	0.52
GO:0099095	ligand-gated anion channel activity	0.006%		0.0008	0.60	0.58
GO:0005283	amino acid:sodium symporter activity	0.045%		0.0034	0.44	0.61
GO:0008137	NADH dehydrogenase (ubiquinone) activity	0.260%		0.0021	0.45	0.61
GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	0.115%		0.0012	0.52	0.62
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0003	0.44	1.00



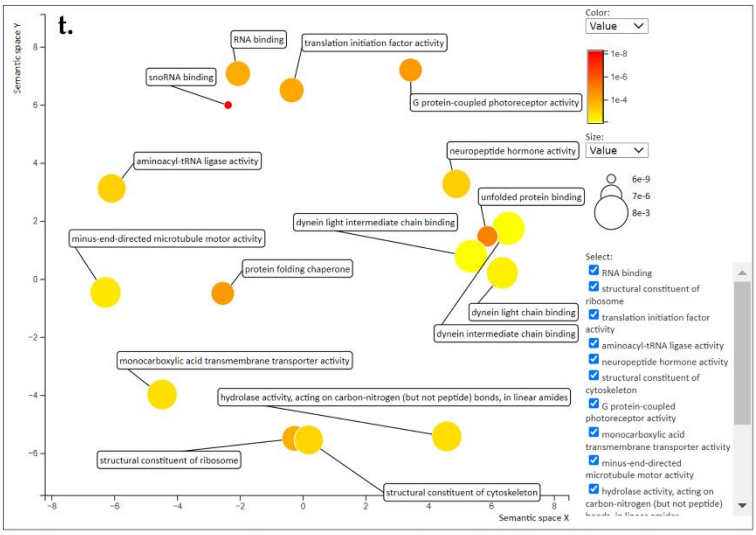
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004252	serine-type endopeptidase activity	0.893%		0.0096	0.90	0.00
GO:0008307	structural constituent of muscle	0.004%		0.0039	0.98	0.00
GO:0015662	P-type ion transporter activity	0.107%		0.0040	0.55	0.00
GO:0016917	GABA receptor activity	0.027%		0.0073	1.00	0.00
GO:0042277	peptide binding	0.100%		0.0018	1.00	0.00
GO:0016776	phosphotransferase activity, phosphate group as acceptor	0.361%		0.0037	0.96	0.03
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1.855%		0.0061	0.85	0.04
GO:0008061	chitin binding	0.116%		0.0001	1.00	0.04
GO:0004568	chitinase activity	0.057%		0.0058	0.92	0.21
GO:0008483	transaminase activity	0.728%		0.0036	0.96	0.23
GO:0016298	lipase activity	0.261%		0.0049	0.93	0.24
GO:0008010	structural constituent of chitin-based larval cuticle	0.002%		0.0000	0.98	0.33
GO:0016645	oxidoreductase activity, acting on the CH-NH group of donors	0.350%		0.0028	0.86	0.37
GO:0005313	L-glutamate transmembrane transporter activity	0.004%		0.0028	0.65	0.39
GO:0009055	electron transfer activity	1.194%		0.0011	0.77	0.42
GO:0032450	maltose alpha-glucosidase activity	0.006%		0.0028	0.91	0.47
GO:0005328	neurotransmitter:sodium symporter activity	0.001%		0.0000	0.65	0.47
GO:0015252	proton channel activity	0.315%		0.0028	0.57	0.52
GO:0022857	transmembrane transporter activity	9.001%		0.0033	0.61	0.53
GO:0070006	metalloaminopeptidase activity	0.153%		0.0000	0.91	0.54



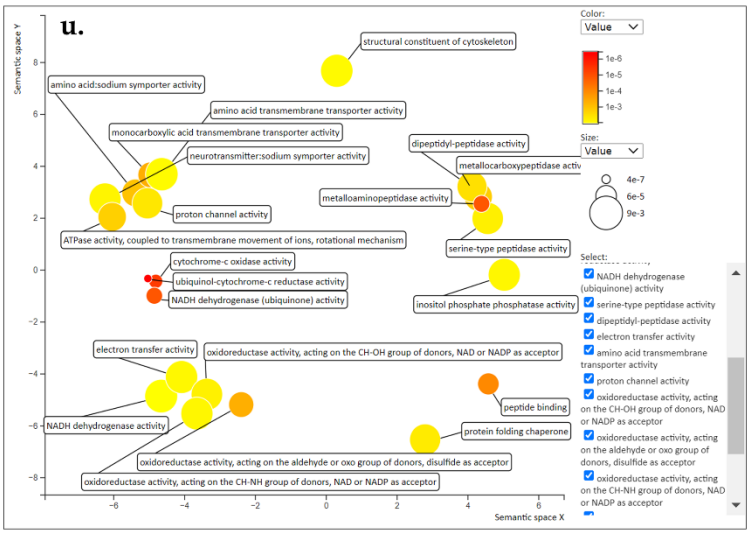
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0003729	mRNA binding	0.276%		0.0056	0.78	0.00
GO:0008186	ATP-dependent activity, acting on RNA	0.334%		0.0000	0.92	0.00
GO:0140097	catalytic activity, acting on DNA	2.626%		0.0028	0.79	0.00
GO:0008139	nuclear localization sequence binding	0.010%		0.0000	0.96	0.04
GO:0051082	unfolded protein binding	0.424%		0.0028	0.87	0.05
GO:0003697	single-stranded DNA binding	0.255%		0.0035	0.81	0.22
GO:0003677	DNA binding	11.354%		0.0048	0.75	0.33
GO:0008296	3'-5'-exodeoxyribonuclease activity	0.010%		0.0031	0.78	0.39
GO:0030515	snoRNA binding	0.048%		0.0000	0.80	0.40
GO:0042393	histone binding	0.086%		0.0005	0.88	0.47
GO:0003723	RNA binding	5.630%		0.0023	0.76	0.53
GO:0003724	RNA helicase activity	0.307%		0.0000	0.73	0.54
GO:0008297	single-stranded DNA exodeoxyribonuclease activity	0.006%		0.0011	0.78	0.70



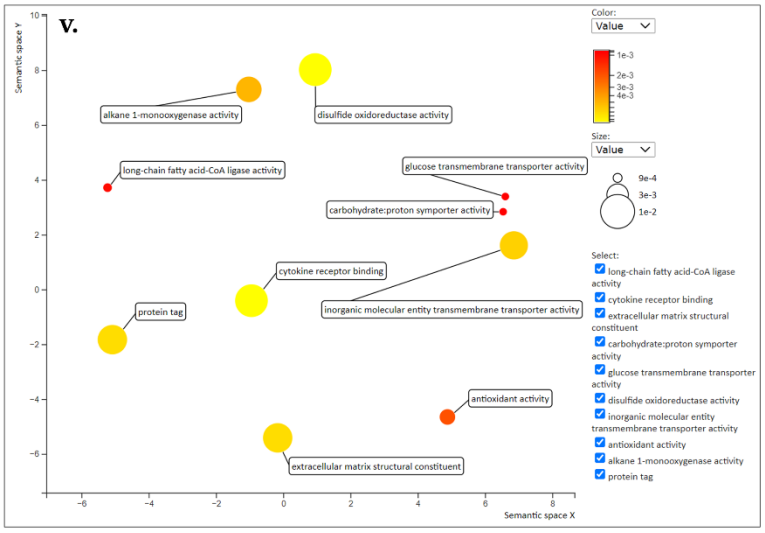
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008061	chitin binding	0.116%		0.0000	1.00	0.00
GO:0008307	structural constituent of muscle	0.004%		0.0013	0.96	0.00
GO:0008331	high voltage-gated calcium channel activity	0.001%		0.0081	0.93	0.00
GO:0008483	transaminase activity	0.728%		0.0100	0.92	0.00
GO:0004300	enoyl-CoA hydratase activity	0.051%		0.0039	0.97	0.03
GO:0004568	chitinase activity	0.057%		0.0020	0.92	0.03
GO:0016861	intramolecular oxidoreductase activity, interconverting aldoses and ketoses	0.309%		0.0081	0.96	0.03
GO:0004601	peroxidase activity	0.427%		0.0042	0.86	0.03
GO:0016879	ligase activity, forming carbon-nitrogen bonds	1.271%		0.0015	0.90	0.04
GO:0004364	glutathione transferase activity	0.086%		0.0000	0.93	0.20
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0044	0.90	0.22
GO:0016746	acyltransferase activity	3.062%		0.0079	0.91	0.29
GO:0003995	acyl-CoA dehydrogenase activity	0.152%		0.0001	0.87	0.30
GO:0016646	oxidoreductase activity, acting on the CH-NH group of donors, NAD or NADP as acceptor	0.269%		0.0039	0.87	0.31
GO:0008010	structural constituent of chitin-based larval cuticle	0.002%		0.0006	0.96	0.33
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1.855%		0.0004	0.85	0.38



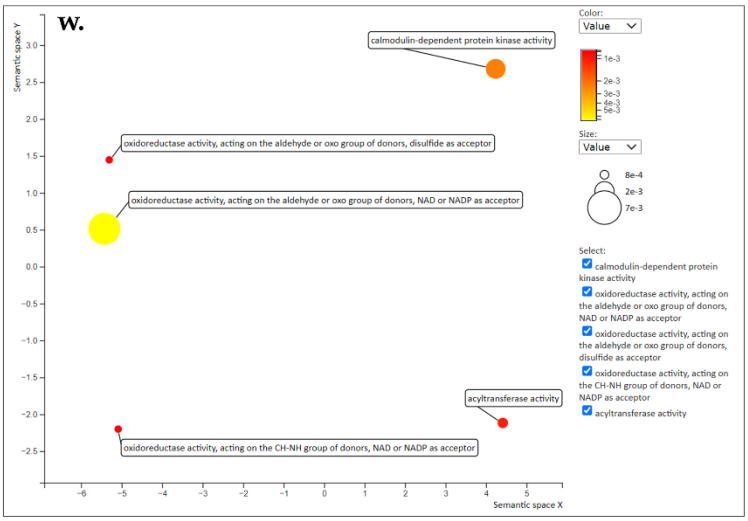
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008028	monocarboxylic acid transmembrane transporter activity	0.052%		0.0014	1.00	0.00
GO:0008569	minus-end-directed microtubule motor activity	0.040%		0.0023	1.00	0.00
GO:0016811	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides	0.737%		0.0014	1.00	0.00
GO:0051959	dynein light intermediate chain binding	0.026%		0.0080	0.80	0.00
GO:0044183	protein folding chaperone	0.019%		0.0000	1.00	0.01
GO:0005200	structural constituent of cytoskeleton	0.076%		0.0008	0.91	0.01
GO:0004812	aminoacyl-tRNA ligase activity	0.895%		0.0006	1.00	0.03
GO:0003743	translation initiation factor activity	0.394%		0.0001	0.84	0.04
GO:0008020	G protein-coupled photoreceptor activity	0.003%		0.0000	0.94	0.16
GO:0045503	dynein light chain binding	0.003%		0.0037	0.82	0.31
GO:0003723	RNA binding	5.630%		0.0001	0.89	0.32
GO:0005184	neuropeptide hormone activity	0.013%		0.0006	0.76	0.34
GO:0045505	dynein intermediate chain binding	0.030%		0.0080	0.80	0.36
GO:0030515	snoRNA binding	0.048%		0.0000	0.89	0.41
GO:0051082	unfolded protein binding	0.424%		0.0000	0.77	0.43
GO:0003735	structural constituent of ribosome	2.348%		0.0001	0.92	0.63



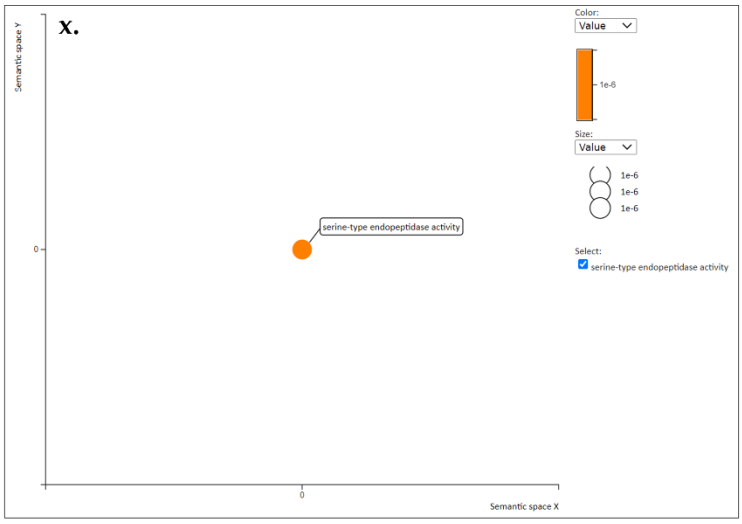
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0003954	NADH dehydrogenase activity	0.374%		0.0094	0.76	0.00
GO:0015171	amino acid transmembrane transporter activity	0.268%		0.0079	0.68	0.00
GO:0042277	peptide binding	0.100%		0.0001	1.00	0.00
GO:0044183	protein folding chaperone	0.019%		0.0043	0.99	0.01
GO:0005200	structural constituent of cytoskeleton	0.076%		0.0076	0.99	0.01
GO:0052745	inositol phosphate phosphatase activity	0.046%		0.0067	0.89	0.02
GO:0008236	serine-type peptidase activity	1.328%		0.0051	0.75	0.21
GO:0004252	serine-type endopeptidase activity	0.892%		0.0001	0.76	0.94
GO:0005328	neurotransmitter:sodium symporter activity	0.001%		0.0064	0.69	0.25
GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	0.123%		0.0004	0.82	0.29
GO:0016646	oxidoreductase activity, acting on the CH-NH group of donors, NAD or NADP as acceptor	0.269%		0.0067	0.81	0.31
GO:0009055	electron transfer activity	1.194%		0.0071	0.69	0.36
GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	0.115%		0.0015	0.64	0.37
GO:0046961	proton-transporting ATPase activity, rotational mechanism	0.115%		0.0001	0.57	1.00
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	1.855%		0.0053	0.78	0.42



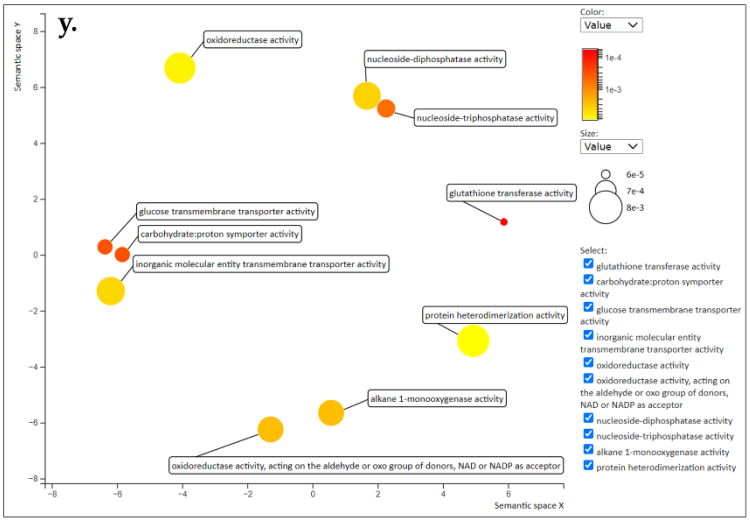
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0005126	cytokine receptor binding	0.094%		0.0096	1.00	0.00
GO:0005201	extracellular matrix structural constituent	0.037%		0.0070	1.00	0.00
GO:0015036	disulfide oxidoreductase activity	0.326%		0.0096	0.95	0.00
GO:0015318	inorganic molecular entity transmembrane transporter activity	2.924%		0.0062	0.84	0.00
GO:0031386	protein tag	0.011%		0.0070	1.00	0.00
GO:0016209	antioxidant activity	0.725%		0.0019	0.99	0.01
GO:0004467	long-chain fatty acid-CoA ligase activity	0.034%		0.0010	0.98	0.02
GO:0018685	alkane 1-monooxygenase activity	0.001%		0.0048	0.95	0.21
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0009	0.78	0.35
GO:0005351	carbohydrate:proton symporter activity	0.026%		0.0009	0.77	0.71



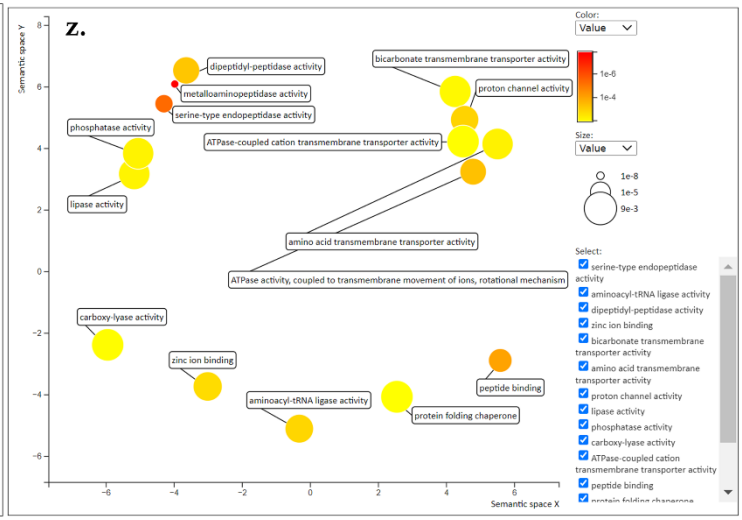
Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	0.593%		0.0067	0.49	0.00
GO:0004683	calmodulin-dependent protein kinase activity	0.022%		0.0023	0.87	0.02
GO:0016746	acyltransferase activity	3.062%		0.0010	0.85	0.20
GO:0016646	oxidoreductase activity, acting on the CH-NH group of donors, NAD or NADP as acceptor	0.269%		0.0008	0.70	0.32
GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	0.123%		0.0008	0.51	0.81



Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0004252	serine-type endopeptidase activity	0.893%		0.0000	1.00	0.00



Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0015318	inorganic molecular entity transmembrane transporter activity	2.924%		0.0038	0.85	0.00
GO:0017110	nucleoside-diphosphatase activity	0.016%		0.0034	0.87	0.00
GO:0046982	protein heterodimerization activity	0.275%		0.0083	1.00	0.00
GO:0018685	alkane 1-monooxygenase activity	0.001%		0.0023	0.93	0.02
GO:0004364	glutathione transferase activity	0.086%		0.0001	0.97	0.02
GO:0016491	oxidoreductase activity	12.887%		0.0067	0.96	0.04
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	0.593%		0.0023	0.92	0.22
GO:0005355	glucose transmembrane transporter activity	0.012%		0.0003	0.78	0.35
GO:0017111	nucleoside-triphosphatase activity	3.353%		0.0005	0.86	0.52
GO:0005351	carbohydrate:proton symporter activity	0.026%		0.0003	0.77	0.71



Term ID	Name	Frequency	Pin?	Value	Uniqueness	Dispensability
GO:0008270	zinc ion binding	3.638%		0.0015	0.99	0.00
GO:0016831	carboxy-lyase activity	0.619%		0.0080	0.97	0.00
GO:0044183	protein folding chaperone	0.019%		0.0091	0.99	0.00
GO:0019829	ATPase-coupled cation transmembrane transporter activity	0.352%		0.0088	0.72	0.01
GO:0016298	lipase activity	0.261%		0.0051	0.83	0.03
GO:0004812	aminoacyl-tRNA ligase activity	0.895%		0.0011	0.97	0.03
GO:0042277	peptide binding	0.100%		0.0001	0.99	0.05
GO:0008239	dipeptidyl-peptidase activity	0.046%		0.0005	0.79	0.19
GO:0015171	amino acid transmembrane transporter activity	0.268%		0.0047	0.80	0.38
GO:0015106	bicarbonate transmembrane transporter activity	0.003%		0.0066	0.81	0.41
GO:0004252	serine-type endopeptidase activity	0.893%		0.0000	0.78	0.49
GO:0016791	phosphatase activity	1.292%		0.0048	0.81	0.55
GO:0070006	metalloaminopeptidase activity	0.153%		0.0000	0.77	0.65
GO:0015252	proton channel activity	0.315%		0.0009	0.76	0.66
GO:0044769	ATPase activity, coupled to transmembrane movement of ions, rotational mechanism	0.115%		0.0004	0.78	0.68

Figure S6: Revigo Scatter and Table Views of the GO terms over-represented in all comparisons (R vs S, R vs C and C vs S) for data from Nigeria, Niger, Chad and Cameroon.

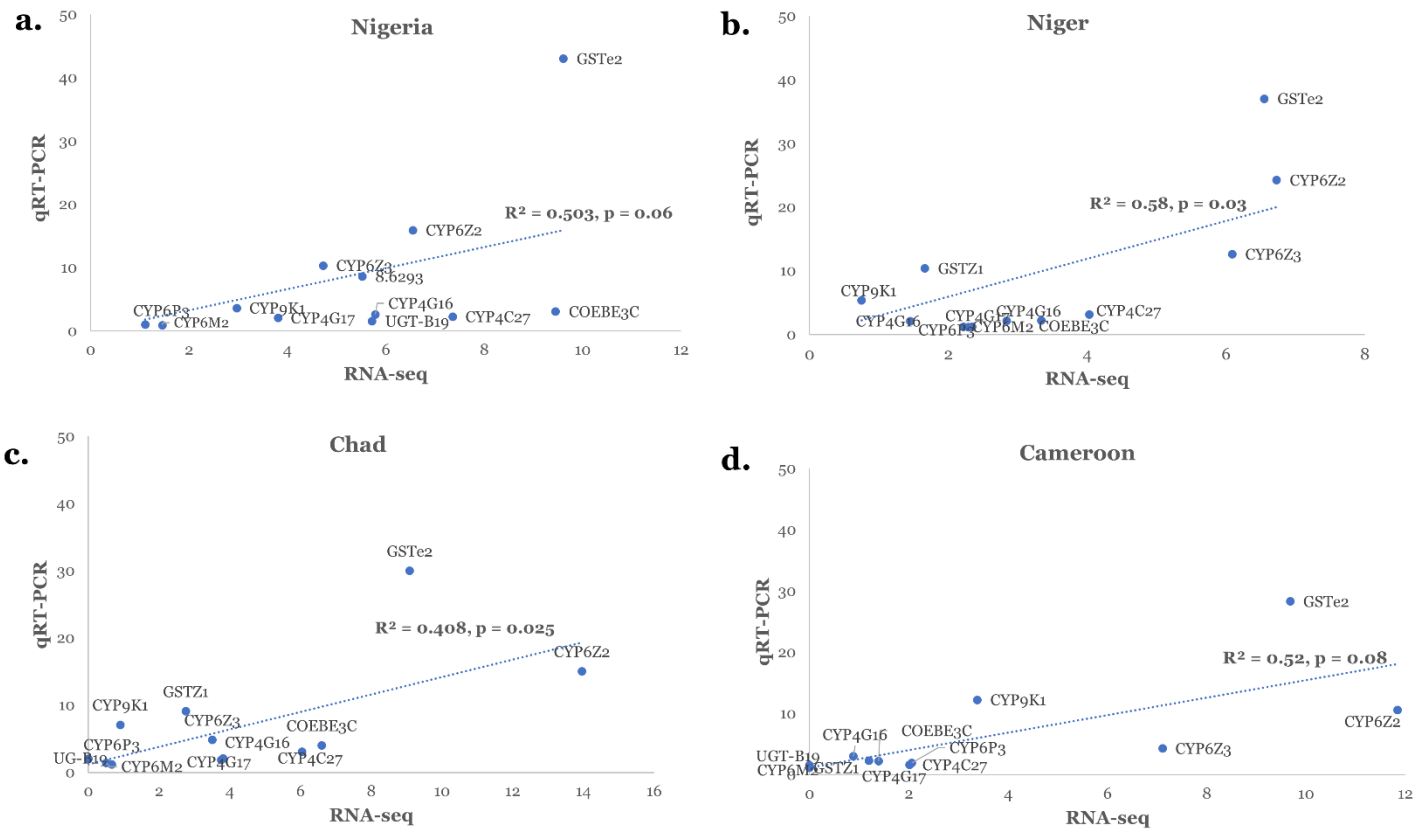


Figure S7: Validation of RNA-seq results using qRT-PCR analysis: correlation between RNA-seq fold changes and those obtained from qRT-PCR for 12 metabolic resistance genes.

a.

Population	n	S	Syn	Nonsyn	h	H_d	π (k)	D (Tajima)	D* (Fu and Li)
Nigeria	4	2	2	0	2	0.667	0.00090	1.893^{ns}	1.893^{ns}
Niger	4	42	36	6	3	0.833	0.01848	2.006^{ns}	2.006^{ns}
Chad	4	0	0	0	1	0.00	0.0000	0.000	0.000
Cameroon	4	23	20	3	2	0.500	0.0078	-0.857^{ns}	0.857^{ns}
Ngoussou	4	1	1	0	2	0.500	0.00034	-0.612^{ns}	-0.612^{ns}
All	20	75	65	12	10	0.921	0.01648	0.501^{ns}	1.298^{ns}

b.

Hap #	Freq.	Sequences
Hap_1:	2	>Nigeria2 >Nigeria3
Hap_2:	2	>Nigeria1 >Nigeria7
Hap_3:	1	>Niger1
Hap_4:	1	>Niger2
Hap_5:	2	>Niger3 >Niger4
Hap_6:	4	>Chad1 >Chad2 >Chad3 >Chad6
Hap_7:	3	>Cameroon1 >Cameroon2 >Cameroon3
Hap_8:	1	>Cameroon4
Hap_9:	3	>Ngoussou2 >Ngoussou4 >Ngoussou6
Hap_10:	1	>Ngoussou3

c.

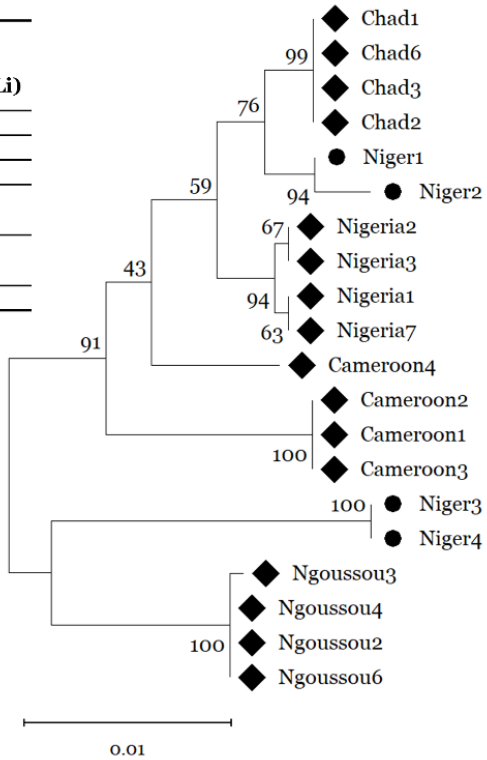


Figure S8: Polymorphism analysis of full length cDNA of *CYP6Z2*. a. Summary statistics for polymorphisms of *CYP6Z2* cDNA across the Sahel countries. n = number of sequences (n); S, number of polymorphic sites; h, haplotype; H_d, haplotype diversity Syn, Synonymous mutations; Nonsyn, Non-synonymous mutations; π, nucleotide diversity (k= mean number of nucleotide differences); Tajima's D and Fu and Li's D statistics, ns, not significant. **b.** *CYP6Z2* cDNA haplotypes and its frequencies, **c.** a maximum likelihood phylogenetic tree of *CYP6Z2* cDNA sequences.

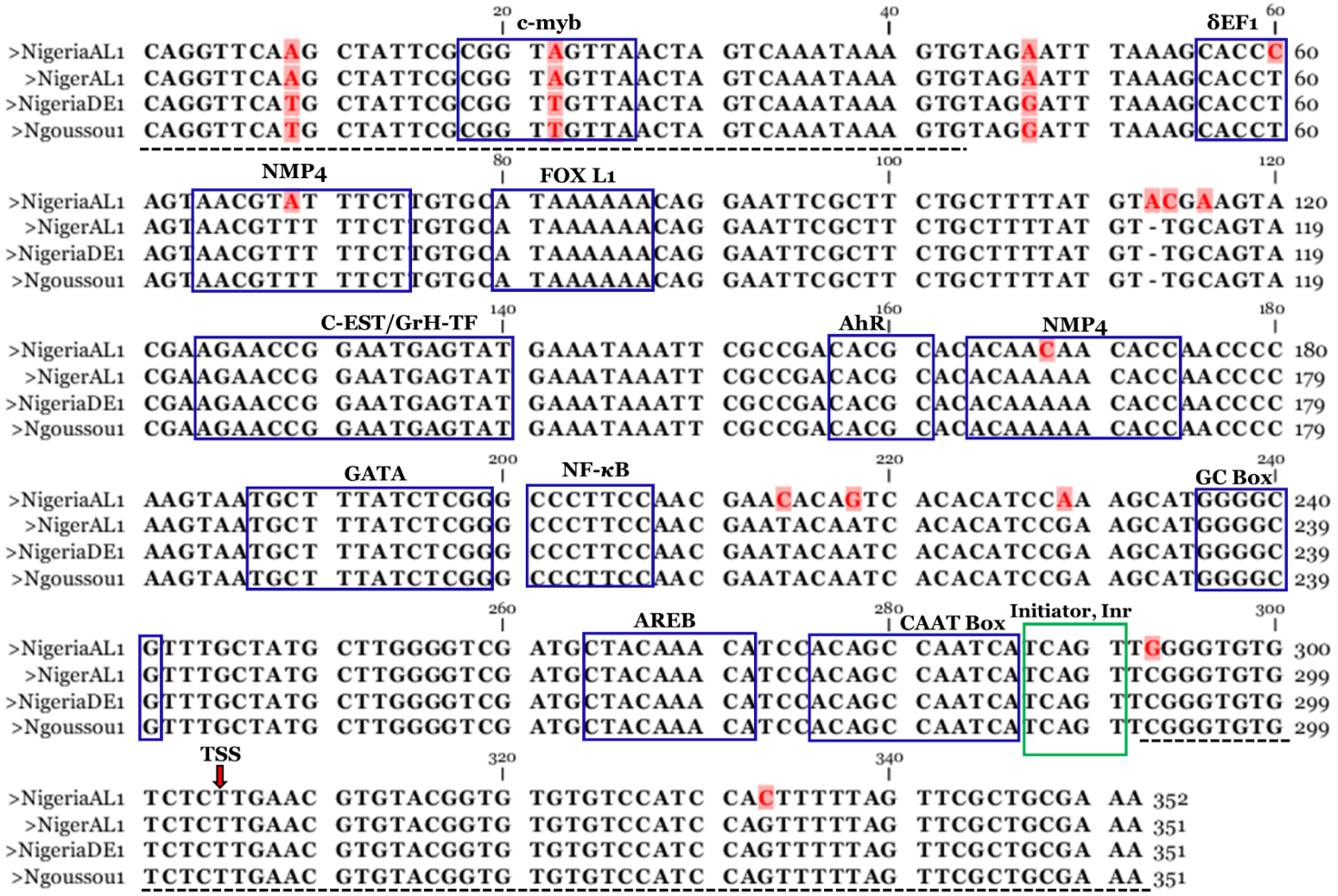


Figure S9: Comparative alignment of the *GSTe2* 5'-UTR fragments from the various haplotypes, showing putative transcription factors binding sites (blue boxes), and polymorphic positions (in red and highlighted in pink). The arthropod Inr consensus sequence is in green box, the transcription start site indicated with red arrow, and the 3'-UTR of *GSTe1* and 5'-UTR of *GSTe2* indicated with dashed lines.

a.

Hap# Freq. Sequences

Hap_1: 75 NigeriaDE1 NigeriaDE2 NigeriaDE3 NigeriaDE4 NigeriaDE11 NigeriaDE15 NigeriaDE16 NigeriaDE20 NigeriaDE26 Ngoussou1 Ngoussou2 Ngoussou4 Ngoussou5 Ngoussou7 Ngoussou9 Ngoussou11 Ngoussou13 Ngoussou15 Ngoussou16 NigerAL3 NigerAL4 NigerAL5 NigerAL6 NigerAL7 NigerAL8 NigerDE1 NigerDE2 NigerDE3 NigerDE4 NigerDE5 NigerDE6 NigerDE7 NigerDE9 NigerDE11 NigerDE12 ChadAL1 ChadAL3 ChadAL4 ChadAL5 ChadAL6 ChadAL7 ChadAL9 ChadAL10 ChadDE1 ChadDE2 ChadDE3 ChadDE4 ChadDE6 ChadDE7 ChadDE9 ChadDE11 ChadDE13 ChadDE24 CameroonAL1 CameroonAL2 CameroonAL3 CameroonAL4 CameroonAL8 CameroonAL9 CameroonAL11 CameroonDE1 CameroonDE2 CameroonDE3 CameroonDE4 CameroonDE14 CameroonDE12 CameroonDE11 CameroonDE8 CameroonDE10 CameroonDE6 NigeriaAL7 NigeriaAL4 NigeriaAL8 NigeriaAL17 NigeriaAL2

Hap_2: 1 NigeriaDE29

Hap_3: 1 NigeriaAL3

Hap_4: 1 NigeriaAL5

Hap_5: 4 NigerAL1 NigerAL9 NigerAL10 NigerAL11

Hap_6: 8 NigeriaAL1 NigeriaAL6 NigeriaAL9 ChadAL2 ChadAL13 CameroonAL5 CameroonAL25 CameroonAL7

b.

```
        0000001111 2222223
        0246680116 11 23993
        9270952468489 41 33
Hap_1   TTGTTATTCATAGATCG
Hap_2   ...........................G.....
Hap_3   ........... C......................
Hap_4   ..ACAG. CACCGA .AG.
Hap_5   AAA.............................
Hap_6   AAACA. .CACCGA...GC
```

c.

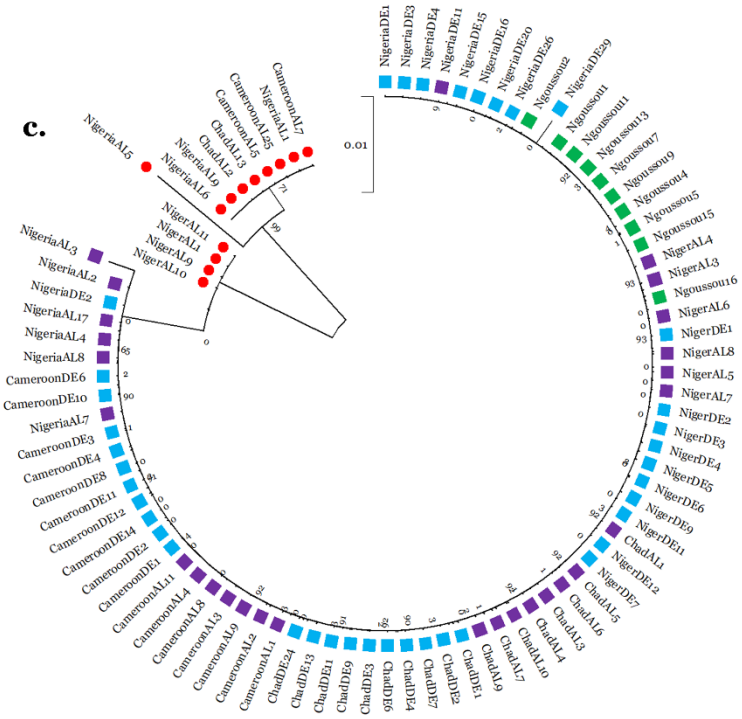


Figure S10: Genetic variability of the *GSTe2* 5'-UTR fragment from Nigeria, Niger, Chad, Cameroon and Ngoussou populations. a. the six haplotypes of the 5'-UTR with respective frequencies, **b.** polymorphic positions with numbers on top indicating the nucleotide position in the 351 bp fragment, **c.** maximum likelihood phylogenetic tree of the 90 sequences, showing the DDT-alive sequences with mutations forming a clade specific to its phenotype.

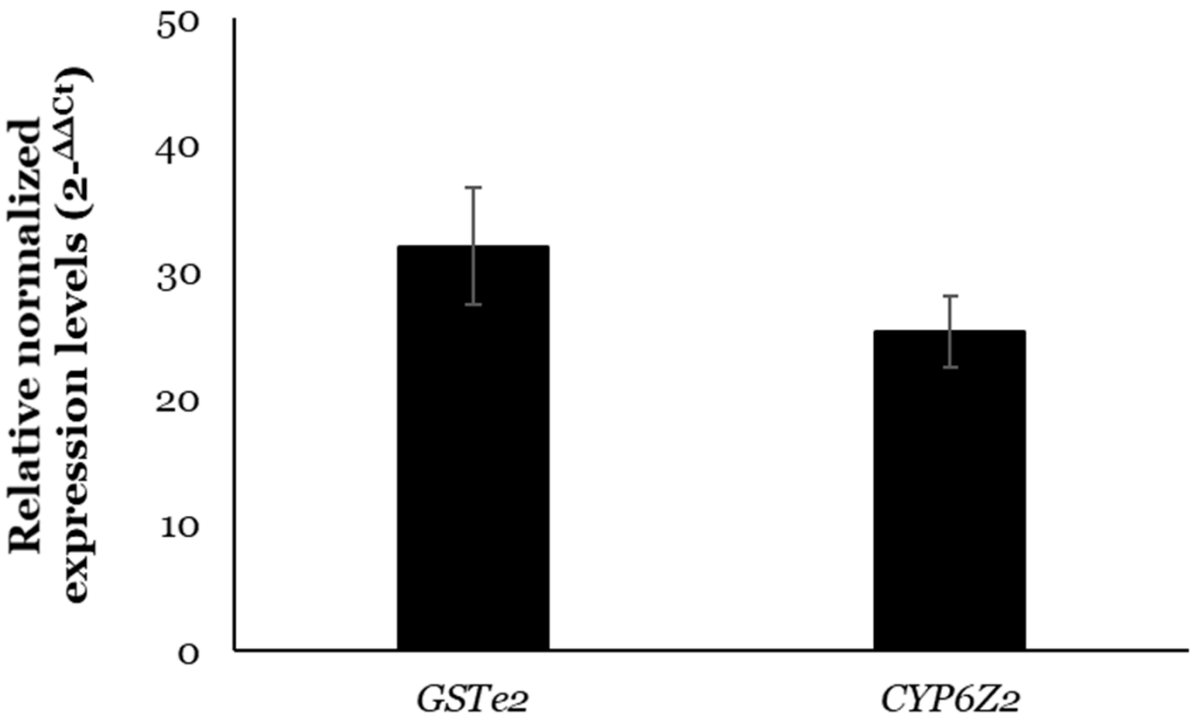


Figure S11: qRT-PCR validation of the overexpression of *GSTe2* and *CYP6Z2*. The relative expression of the *GSTe2* and *CYP6Z2* in the experimental transgenic *D. melanogaster* in relation to control flies. Data shown are the mean ± standard deviation (n = 6).