Cuticular hydrocarbon biosynthesis in malaria vectors: insights from the adult oenocyte transcriptome.

Linda Grigoraki[1], Xavier Grau-Bove[1], Henrietta Carrington-Yates[1], Gareth J Lycett[1] and Hilary Ranson[1]

Liverpool School of Tropical Medicine, Vector Biology Department, L3 5QA Liverpool, United Kingdom

Corresponding Authors: Linda Grigoraki, email: Linta.Grigoraki@lstmed.ac.uk and Hilary Ranson, email: Hilary.Ranson@lstmed.ac.uk

Keywords: Anopheles gambiae, fatty acid synthase, desaturase, cuticle, RNAseq, Fluorescent Activated Cell Sorting

Abstract

The surface of insects is coated in cuticular hydrocarbons (CHCs); variations in the composition of this layer affect a range of traits including adaptation to arid environments and defence against pathogens and toxins. In the African malaria vector, *Anopheles gambiae* quantitative and qualitative variance in CHC composition have been associated with speciation, ecological habitat and insecticide resistance. Understanding how these modifications arise will inform us of how mosquitoes are responding to climate change and vector control interventions. CHCs are synthesised in sub-epidermal cells called oenocytes that are very difficult to isolate from surrounding tissue. Here we utilise a transgenic line with fluorescent oenocytes to purify these cells for the first time. Comparative transcriptomics revealed the enrichment of biological processes related to long chain fatty acyl-CoA biosynthesis and elongation of mono-, poly-unsaturated and saturated fatty acids and enabled us to delineate, and partially validate, the hydrocarbon biosynthetic pathway in *An gambiae*.

Introduction

The cuticle, also known as the exoskeleton, is the outermost part of the insect body and plays a pivotal role in its physiology and ability to adapt and survive in terrestrial environments. The cuticle consists of multiple layers with different composition and properties. The thickest layer, the procuticle, is divided into the endo- and exo-cuticle, both of which are rich in chitin and cuticular proteins. The outer layer, or epi-cuticle, is mainly composed of lipids and hydrocarbons (Lockey, 1988). Cuticular hydrocarbons (CHCs) are relatively simple molecules but form complex and varied mixtures of n-alkanes, unsaturated hydrocarbons (alkenes), and terminally and internally methyl-branched alkanes/alkenes. These mixtures of CHCs protect insects from desiccation, are the first barrier to infections from microorganisms and can act as mating recognition signals (pheromones)(Blomquist et al., 2010). The cuticle composition

has also been associated with resistance to insecticides, via reduced penetration, in several insect species (reviewed in (Balabanidou et al., 2018)).

Anopheles mosquitoes are intensively studied because of their importance as vectors of malaria and lymphatic filariasis that together affect millions of people every year causing intolerable levels of mortality and morbidity. Recently it was shown that populations of the major African malaria vector *Anopheles gambiae* have developed a thicker cuticle with elevated amounts of hydrocarbons and this is associated with a reduction in the penetration rate of pyrethroid insecticides contributing to the high levels of resistance observed (Balabanidou et al., 2016). The emergence of pyrethroid resistance is a major concern for vector control strategies as it threatens the efficiency of the insecticide treated nets, all of which contain this insecticide class, that have proven so successful in reducing the malaria burden in Africa (Bhatt et al., 2015). CHCs are also important in conferring desiccation tolerance in *An. gambiae*, which may be vital in adaptation to arid conditions and survival during the dry season. (Reidenbach et al., 2014, Arcaz et al., 2016).

Cuticular hydrocarbons are synthesized in oenocytes which are secretory cells of ectodermal origin found in most, if not all, pterygote insects (Makki et al., 2014). In adult mosquitoes oenocytes are found in characteristic, predominantly ventral, subcuticular clumps that form rows in each segment, while in larval stages they are located in small groups underneath each of the abdominal appendages (Lycett et al., 2006).

The biosynthesis of hydrocarbons has been studied using radiolabelled precursors (Dillwith et al., 1981) and the biochemical steps of their biosynthetic pathway have been established (Blomquist et al., 2010, Chung and Carroll, 2015). The pathway starts with a fatty acid synthase (FAS) that uses malonyl-CoA to generate a fatty acyl-CoA. In the case of methyl-branched hydrocarbons propionyl-CoA groups (as methyl-malonyl-CoA) are also incorporated in the growing fatty acyl-CoA chain. The fatty acyl-CoA chain is further extended by elongases, which extend the chain to different lengths depending on their specificity. Desaturases introduce double bonds, contributing to the generation of unsaturated hydrocarbons, and reductases convert the generated acyl-CoA to aldehydes. These aldehydes serve as substrates for the final step of the pathway, which involves a single carbon chain-shortening conversion to hydrocarbons catalysed by P450 enzymes (Qiu et al., 2012). Only this latter step has been delineated in *Anopheles* mosquitoes with two P450 decarbonylases identified, Cyp4G16 and Cyp4G17 (Balabanidou et al., 2016, Kefi et al., 2019).

Only a subset of the large number of lipid metabolic enzymes encoded in the genome are likely to be significant players in CHC synthesis, but we hypothesised that transcripts from these genes will be specifically enriched in oenocytes to enable this function. Here we report the isolation of oenocytes from adult *An. gambiae* mosquitoes using a transgenic line with fluorescently tagged oenocytes (Lynd et al., 2019). RNAseq of the isolated oenocytes identified the key biological processes enriched in these cells and revealed candidate genes for each step of the CHC biosynthetic pathway. A member of the putative pathway was

validated by perturbing expression of the AGAP001899 fatty acid synthase (hereafter called FAS1899). The elucidation of this pathway is a major milestone in delineating the role of variable hydrocarbon composition on key traits that impact vectorial capacity of these important vectors of human disease.

Results

FACS isolation of fluorescent oenocytes from transgenic *An. gambiae* mosquitoes

To tag adult *An. gambiae* oenocytes, we expressed the red fluorescent marker m-cherry specifically in these cells using the GAL4/UAS system (Lynd and Lycett, 2012). Two transgenic lines were crossed: 1) a homozygous UAS-mCD8: mCherry responder line (Adolfi et al., 2018) with 2) a homozygous oenocyte enhancer-GAL4 driver line (Oeno-Gal4) (Lynd et al., 2019). Progeny of this cross had the expected m-cherry fluorescent oenocytes throughout their development (Lynd et al., 2019). To purify adult oenocytes, mosquitoes were dissected to expose the oenocytes that are dispersed throughout tissues attached to the ventral abdominal integument. Their release was facilitated using trypsin and mechanical homogenization of the tissue (Figure 1A) and subsequent isolation with Fluorescent Activated Cell Sorting (FACS) (Figure 1B). Tagged cells corresponded to 1-5% of the total events counted during the FACS sorting and their morphology was consistent with oenocytes by microscopic inspection of sorted cells (Supplementary Figure 1).

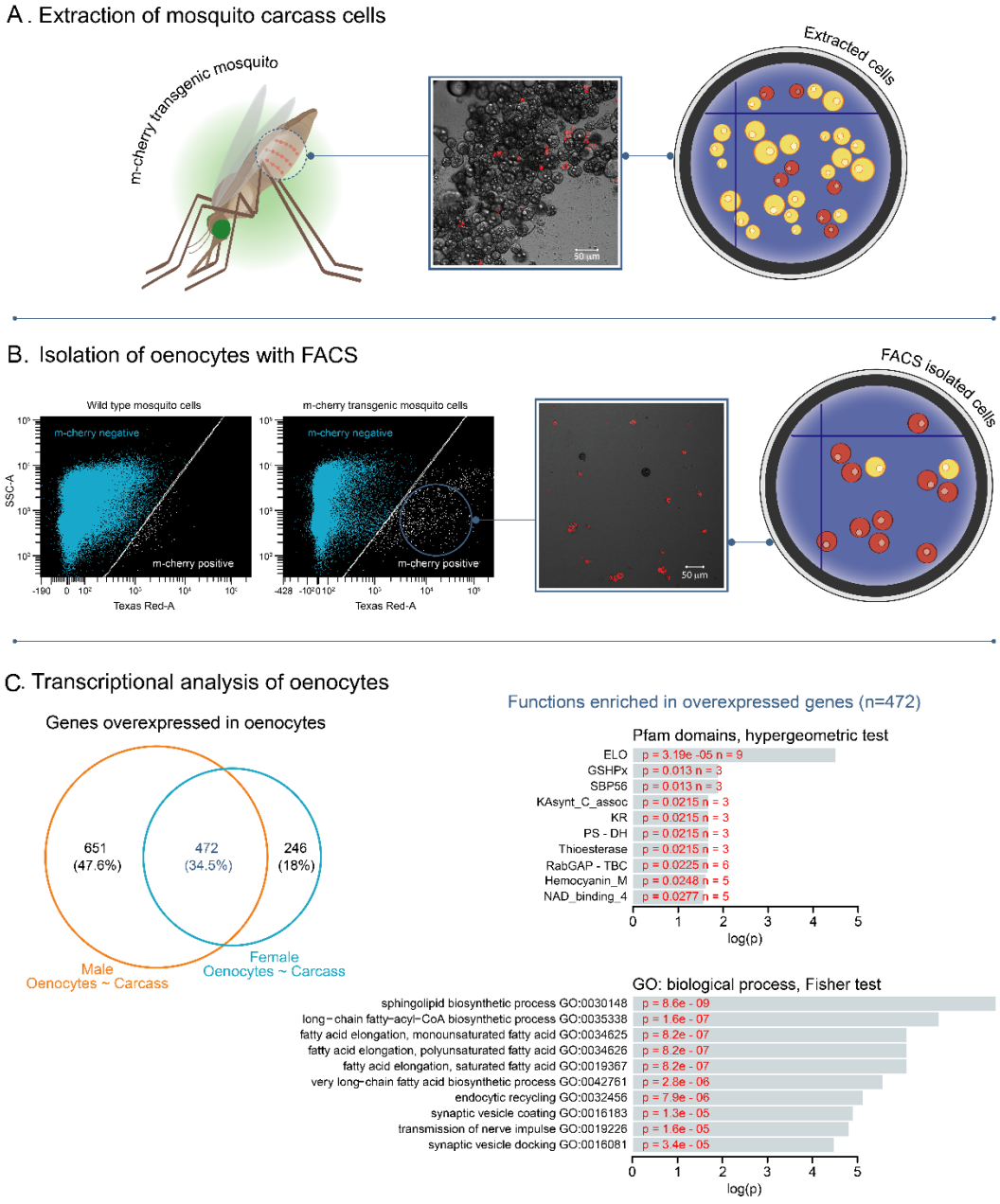


Figure 1: Isolation of fluorescently tagged oenocytes and transcriptomic analysis with RNAseq. A) Schematic image of total carcass cells extracted from transgenic *An. gambiae* mosquitoes (progeny of UAS-mCD8: mCherry line and Oeno-Gal4 driver line) expressing the m-cherry fluorescent marker in oenocytes (red cells). **B)** FACS dot plots. Side-scatter intensity (vertical axis) is plotted against fluorescence intensity (horizontal axis). The sample on the left is from wild type G3 mosquitoes, the sample on the right is from transgenic mosquitoes with fluorescent oenocytes. The white line crossing the plots represents the threshold used for sorting mCherry positive cells. **C)** Transcriptomic analysis for isolated oenocytes and total carcass cells. Venn diagram for genes over-expressed in both female and male oenocytes vs total carcass cells. Go term (biological process) and Pfam domain enrichment analysis is shown for the 472 genes commonly over-expressed in female and male oenocytes. (ELO: fatty acid elongation, GSHPx: Glutathione Peroxidase, SBP56: Selenium Binding Protein, KAsynt-C: Ketoacyl – synthetase C-terminal extension, KR: KR domain found in polyketide and fatty-acid synthases, PS-DH: Polyketide synthase dehydratase, RabGAP-TBC: RabGTPase-TBC domain).

Transcriptome analysis of isolated oenocytes and total carcass cells using RNAseq

Triplicate RNAseq libraries were generated using mRNA from isolated tagged cells and total cell populations (cell preparation before FACS, referred herein as carcass cells) from female and male mosquitoes, barcoded and run on the same lane of an Illumina HiSeq sequencer (CGR University of Liverpool). Paired end reads were processed to remove Illumina adapter sequences and low-quality reads. 97.12% of reads passed the quality control and generated a total of 425 million reads, of which 58.1% (+/- 0.89% standard error) were successfully mapped to the annotated transcripts of *An. gambiae* (Vector Base AgamP4.9).

To visualize how gene expression varied in the different samples we performed a principal component analysis (PCA) using the normalised gene counts of each sample. The first component accounted for 30.1% of the variance in gene expression and separated oenocyte from carcass samples, whereas the second component accounted for 25.9% of variance and reflected differences between females and males. All three replicates of each condition (total female carcass cells, total male carcass cells, female oenocytes, male oenocytes) clustered together (Supplementary Figure 2) providing support for robustness of replication between samples.

Differential expression analysis reveals genes and biological processes enriched in oenocytes

We next identified transcripts significantly over-expressed [$\log_2$(Fold Change)>1, Benjamini-Hochberg adjusted pvalue<0.001, from a Wald test) in oenocytes compared to total (pre-sorted) cells. Our analysis of differential expression identified 1,123 genes over-expressed in male oenocytes compared to male carcass cells and 718 genes over-expressed in female oenocytes compared to female carcass cells. From all over-expressed genes 472 were commonly over-expressed in both female and male oenocytes (Figure 1C and Supplementary File 1). Gene Ontology enrichment analysis for these 472 genes showed an enrichment in biological processes related to sphingolipid biosynthesis, long chain fatty acyl-CoA biosynthesis and elongation of mono-, poly- unsaturated and saturated fatty acids (Figure 1C), supporting the role of oenocytes in lipid and hydrocarbon biosynthesis. Other biological processes enriched in the oenocyte samples included endocytic recycling, synaptic vesicle coating and docking, and transmission of nerve impulses. Enrichment analysis of Pfam protein domains showed the over-representation of the ELO family that consists of integral membrane proteins involved in the elongation of fatty acids (Figure 1C).

We also investigated whether specific gene isoforms are differentially expressed in oenocytes (at p<0.05, obtained from an empirical cumulative distribution of isoform frequency changes). 672 genes had at least one isoform differentially expressed in female oenocytes compared to female total carcass cells and 752 have at least one isoform differentially expressed in male oenocytes compare to male total carcass cells. The same analysis was performed for female and male oenocytes showing 578 genes to have at least one isoform differentially expressed between sexes (Supplementary Document, Supplementary File 2 and Supplementary Figure 3).

Identification of key candidate genes in the CHC biosynthetic pathway

We next examined which transcripts from members of the six gene families (propionyl-CoA synthases, fatty acid synthetases, elongases, desaturases, reductases and P450 decarbonylases) having roles in the hydrocarbon biosynthetic pathway (Figure 2) are differentially expressed in oenocytes. The two P450s, Cyp4G16 (AGAP001076) and Cyp4G17 (AGAP000877), that catalyse the last step in the production of cuticular hydrocarbons, plus the P450 reductase (CPR) that supplies electrons to all P450 monooxygenation reactions, were among the significantly enriched genes (Supplementary File 1). Immunolocalization experiments have previously shown these genes to be highly expressed in *An. gambiae* oenocytes (Balabanidou et al., 2016, Lycett et al., 2006), lending confidence that our experimental design detects oenocyte enriched genes.

The single propionyl-CoA synthase, AGAP001473, likely responsible for the generation of precursor molecules for the synthesis of methyl-branched hydrocarbons (Blomquist et al., 2010) was enriched in oenocytes. Of the four remaining gene families, specific members were found to be oenocyte enriched; these consisted of three of the four fatty acid synthases (AGAP001899, AGAP08468, AGAP028049), nine of the 20 elongases (AGAP013219, AGAP004372, AGAP001097, AGAP003196, AGAP005512, AGAP007264, AGAP013094, AGAP003195, AGAP003197), one desaturase (AGAP003050 out of nine in the genome) and five of the 17 reductases (AGAP005986, AGAP004787, AGAP005984, AGAP004784, AGAP005985) (Figure 3). In addition, the fatty acid transporter AGAP001763, the ortholog of the *Drosophila melanogaster Fatp* (CG7400) functionally implicated in CHC biosynthesis, (Chiang et al., 2016) was also enriched in the *An. gambiae* oenocyte transcriptome. The majority of these genes were highly expressed in oenocytes (among the top 200 most highly expressed), with Cyp4G16, Cyp4G17 and FAS1899 (AGAP001899) being in the top ten, followed by the elongase AGAP007264 (Supplementary tables S2, S3 and Supplementary File 1).

Interestingly, several of these genes have highly correlated expression. A meta-analysis of 48 transcriptomic datasets from insecticide resistant and susceptible *Anopheles* populations (Ingham et al., 2018) identified 44 transcripts co-regulated with Cyp4G16, eight of which were predicted to be part of the CHC pathway. All these eight transcripts, with at least one from each of the six gene families, were present in our experimentally determined CHC synthesizing candidate gene list (Figure 2).

Notably expression of four genes with a lipid synthesizing role was significantly reduced in oenocytes (depicted on Figure 4). These include the fatty acid synthase AGAP009176, the desaturase AGAP004572, the reductase AGAP003606 and the elongase AGAP003600. Thus, these genes may be involved in the synthesis of Long Chain Fatty Acids (LCFA) in other tissues, most likely in the fat body, and not specific to the CHC biosynthetic pathway.

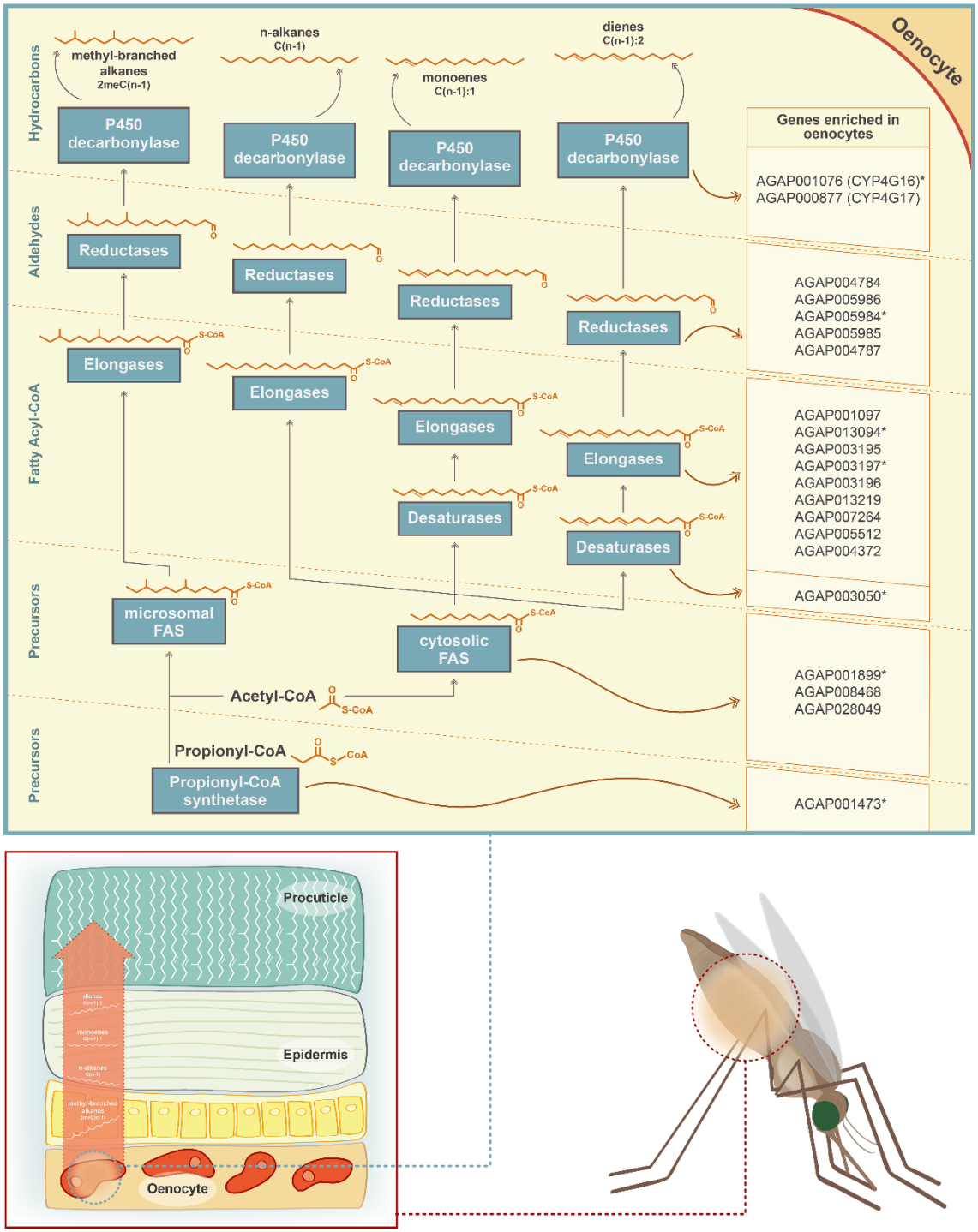


Figure 2: Schematic representation of the CHC biosynthetic pathway (adapted from (Chung and Carroll, 2015)). Gene families implicated in the pathway (propionyl-synthetases, fatty acid synthases FAS, elongases, desaturases, fatty acid reductases and decarbonylases) are depicted in blue boxes. The chemical structure of the two precursor molecules of the pathway (Acetyl-CoA and Propionyl-CoA) is shown, as well as the chemical structure of the product of each step of the pathway. Candidate genes for each step of the pathway, with enriched expression in *An. gambiae* oenocytes, are listed on the left. Genes with an asterisk are members of the Cyp4G16 correlation network (Ingham et al., 2018)

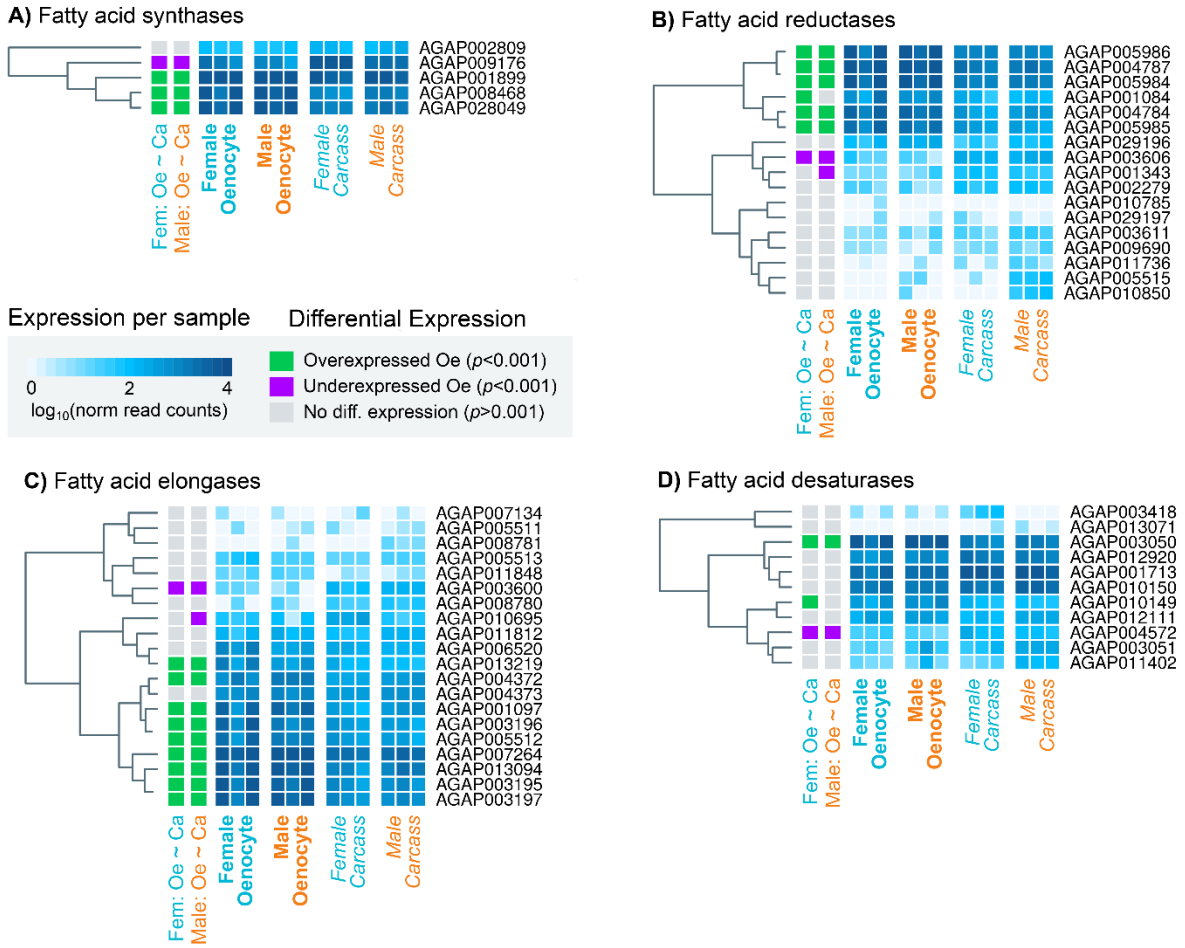


Figure 3: Heat maps showing the expression levels of all *An. gambiae* genes belonging to the four gene families (fatty acid synthetases, elongases, desaturases, reductases) implicated in CHC biosynthesis. Expression levels (presented as different intensities of blue and using the $\log_{10}$ of the normalized read counts) are shown for all 12 samples used in the RNAseq experiment. The differential expression status in female and male oenocytes vs female and male total carcass cells is shown on the left of each panel. Trees on the left of each map are based on similarities in gene expression. Source data: Supplementary Files 1 and 3.

Sex specific differential expression analysis of oenocyte expressed genes.

We next compared the transcription profile of isolated oenocytes from female versus male mosquitoes. Out of 216 genes that were differentially expressed, 72 were over-expressed in female oenocytes and 144 in male oenocytes. Three genes expressing cuticular proteins (CPR130, CPR25 and CPR26) were significantly and highly ($\log_2$FC>3.9) over-expressed in female oenocytes. However, with the strict criteria we used for the differential expression analysis ($\log_2$FC>1, BH adjusted *p*-value <0.001 in all three replicates) we did not find any gene belonging to the hydrocarbon biosynthetic gene families to be differentially expressed between sexes (Supplementary file 3).

215

216 **Phylogenetic relationships of *An. gambiae* genes implicated in CHC biosynthesis**

217 Phylogenetic trees were constructed for the *An. gambiae*, *Ae. aegypti* and *D. melanogaster*
218 gene families of fatty acid synthases, elongases, desaturases and reductases to provide
219 further insights into gene function, in cases where *Drosophila* orthologs have been
220 characterised, and to identify priority candidates for further study in all three species.

221 The fatty acid synthases AGAP001899 and AGAP009176 cluster closely with three *Drosophila*
222 FAS genes (Figure 4A and Supplementary Figure 4), two of which have been shown by *in situ*
223 hydridization to be expressed in oenocytes (Chung et al., 2014). AGAP001899 is
224 phylogenetically closest to CG17374 (FASN3) known to be expressed in *Drosophila* oenocytes
225 whereas AGAP009176, the only *An. gambiae* FAS down-regulated in oenocytes, is most
226 closely related to CG3523 (FASN1), which is expressed in the *Drosophila* fat body.

227 AGAP003050 is the only desaturase enriched in both female and male oenocytes and is a
228 clear 1:1 ortholog of *D. melanogaster* CG15531 (Figure 4B and Supplementary Figure 5) with
229 a predicted stearoyl-CoA 9-desaturase activity, and AAEL003611 (also found expressed in *Ae.*
230 *aegypti* oenocytes (Martins et al., 2011)). AGAP001713 and AGAP012920, the paralog of the
231 three *Drosophila* desaturases Desat1 (CG5887), Desat2 (CG5925) and Fad 2 (CG7923) involved
232 in the production of unsaturated hydrocarbons (Dallerac et al., 2000, Chertemps et al., 2006),
233 some of which act as pheromones, were not among the oenocyte enriched genes.

234 The elongase family appears to have radiated further after evolutionary separation of
235 *Drosophila* and mosquitoes. Five out of the nine *An. gambiae* elongases that are enriched in
236 oenocytes (Figure 4C and Supplementary Figure 6), (AGAP001097, AGAP003195,
237 AGAP003196, AGAP003197, AGAP013219) form a cluster of paralogs phylogenetically related
238 to a single *Drosophila* elongase, CG6660, a gene over-expressed in adult oenocytes (Huang et
239 al., 2019). Two of these paralogs (AGAP003196 and AGAP013219) are closely related to the
240 *Ae. aegypti* AAEL013542 elongase, which is also expressed in pupae oenocytes (Martins et al.,
241 2011). AGAP013094, another oenocyte enriched elongase is the single *An. gambiae* gene in a
242 cluster of *D. melanogaster* paralogs with known functions in CHC biosynthesis, such as eloF
243 (CG16905) (Chertemps et al., 2007), CG30008, CG18609 and CG9458 (Dembeck et al., 2015).

244 Contrary to the other gene families, most fatty acid reductases enriched in oenocytes did not
245 have clear orthology relationships with functionally characterised *D. melanogaster* genes
246 (Figure 4D and Supplementary Figure 7). For example, AGA005984, AGA005985 and
247 AGA005986 clustered in a culicine-specific group of paralogs with no *Drosophila* orthologs.
248 Similarly, most of the fly genes that are functionally linked to CHC profiles (Dembeck et al.,
249 2015), such as CG13091, CG10097, CG17562 and CG18031, form a cluster of paralogs with no
250 one-to-one orthologs in *An. gambiae*.

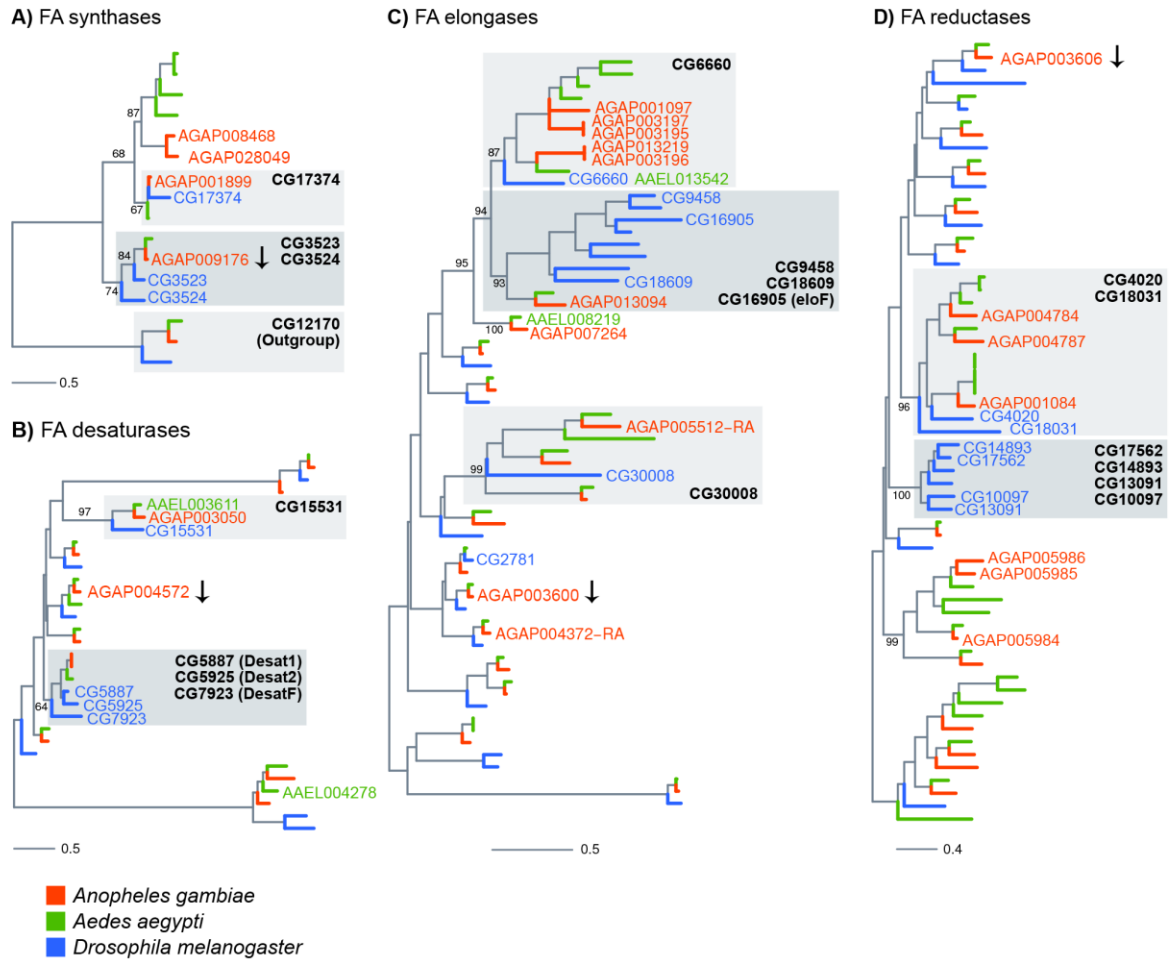


Figure 4: Phylogenetic trees constructed for *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* genes using the protein domains of Fatty acid synthases (PF00109), desaturases (PF00487), elongases (PF01151) and Fatty acyl-CoA reductases (PF07993). Genes named on trees are all *An. gambiae* genes found enriched in adult oenocytes (with the exception of genes followed by an arrow, which were significantly down-regulated), *Ae. aegypti* genes found expressed in pupae oenocytes and *D. melanogaster* genes expressed in oenocytes and/or functionally validated (based on provided references). Grey boxes have been added to clades that are discussed in the text and named based on the *D. melanogaster* members. Scale bars show the number of aminoacid substitutions per alignment position. Trees with all gene names are provided in Supplementary figures 4-7.

Functional validation of candidate genes

The fatty acid synthase, FAS1899 and the desaturase, Desat3050, both of which were significantly enriched in oenocytes, were selected for functional validation. We knocked-down their expression through oenocyte specific RNAi and examined the effect on the CHC profile.

Firstly UAS-regulated responder lines carrying *FAS1899* and *Desat3050* hairpin RNAi constructs were established. Crossing the responder lines with the oenocyte specific-Gal4 promoter line (Oeno-Gal4) (Lynd et al., 2019) resulted in ~80% knock down for the FAS1899 and ~26% for the Desat3050 (in L2 larvae). In both cases oenocyte specific RNAi suppression

was lethal at the L2/L3 larvae stages. Subsequently we crossed the two responder lines with the Ubi-A10 Gal4 line (marked by CFP) (Adolfi et al., 2018) which directs widespread tissue expression, but at lower levels in oenocytes compared to the oeno-Gal4 line. The majority of progeny from these crosses expressing dsRNA for FAS1899 and Desat3050 reached the pupae stage, but 70-80% died either as mid to late pupae or during adult emergence (Supplementary Figure 8). QPCR analysis in whole adults indicated a ~26% knock down of FAS1899 transcripts, but no significant difference in Desat3050 knockdown.

GC-MS analysis of the hexane extracted hydrocarbons revealed the presence of at least 60 CHC peaks in all samples; 15 of which were alkanes, 5 unsaturated alkanes and 40 methyl-branched alkanes. While 19 of the CHC peaks had an abundance of ≥1%, the alkanes C29, C27 and C31, and the methyl-branched methyl-C31 were consistently among the most abundant accounting for approximately half of the total CHCs (Supplementary File 4).

The CHC profile of surviving FAS1899 and Desat3050 knock down adults was compared to control siblings. A significant (Student's t-test p-value ≤ 0.05) 25% reduction in the total amount of hydrocarbons was observed in both female and male FAS1899i mosquitoes. The proportion of the different CHC categories also changed significantly (Student's t-test p-value ≤ 0.05) in the FAS1899i individuals, as the relative abundance of methyl-branched hydrocarbons decreased while the relative abundance of unsaturated and n-alkanes increased (Figure 5). No difference in the total amount of CHCs, nor of the % of unsaturated CHCs, was observed for the surviving Desat3050 knock down adults.

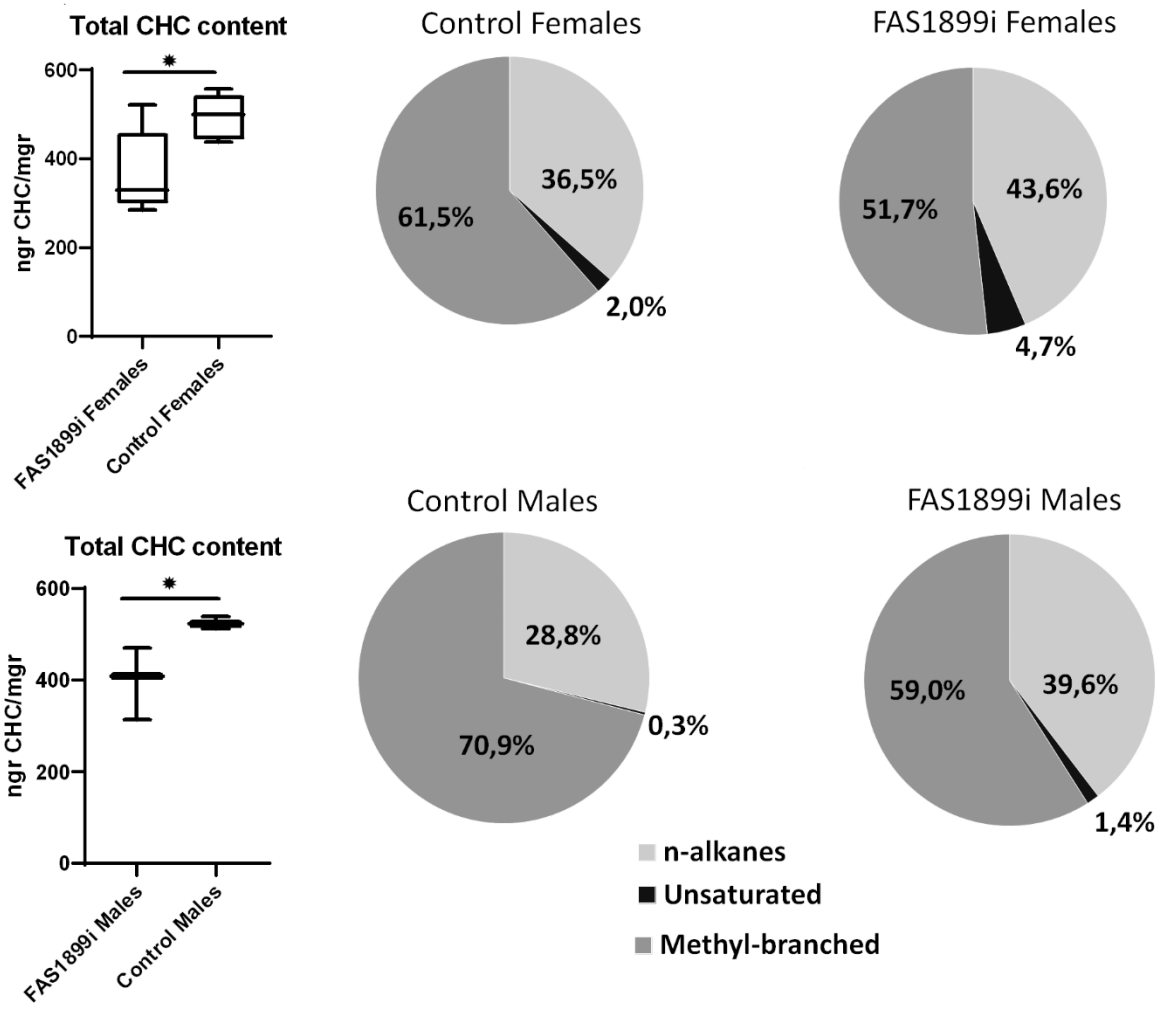


Figure 5: Comparison of the total CHC content quantified with GC-MS in female and male adults with knock down of FAS1899 (Ubi-A10 Gal4/UAS-FAS1899i) and control siblings (heterozygous Ubi-A10 Gal4/+). The mean values of total CHC/mgr (±SEM) are: for FAS1899i females 368ngr/mgr, for control females 494ngr/mgr, for FAS1899i males 398ngr/mgr and for control males 525ngr/mgr (5 biological replicates for females and 3 for males). The box plots show the 25th and 75th percentile; the mean is shown as a black line within the box; error bars correspond to the minimum and maximum values. Pie charts represent the relative abundance of the three CHC categories (n-alkanes, unsaturated alkanes and methyl-branched alkanes) in FAS1899i and control individuals. Statistical test performed: Student's t-test (p-value ≤ 0.05), Source data and p-values provided in Supplementary File 4.

Discussion

CHCs affect key traits in *Anopheles* mosquitoes that determine their fitness and thus vectorial capacity. The difficulties in isolating the CHC synthesizing cells in adult mosquitoes, due to their close association with fat body cells within the abdomen, and the absence of clear one to one orthologs with *Drosophila* in some families (Figure 4), has hindered the identification of genes involved in mosquito CHC production. In this study we describe the FACS purification of fluorescently tagged oenocytes from adult *An. gambiae* mosquitoes, and the subsequent

transcriptomic analysis of the purified cells which enabled us to identify key candidate genes in the CHC biosynthetic pathway.

The samples analysed consisted of total cells recovered from dissected abdomen integument, containing ~12% of tagged oenocyte cells, which were then compared to purified oenocyte cells isolated by passage through the FACS. The abdomen tissue is mainly composed of fat body and epithelial cells, neither of which are expected to synthesize hydrocarbons. Fat bodies do however have a primary role in lipid biosynthesis, which has several steps in common with the CHC biosynthetic pathway, both utilising fatty acid synthases, elongases and desaturases. The analysis pathway was purposively designed to reveal genes and isoforms that are predominantly enriched in oenocytes and thus likely to be involved in CHC biosynthesis but a limitation, in our goal to delineate the entire CHC pathway, is that it will likely fail to detect genes that are expressed at similar levels in fat bodies and oenocytes and are involved in both CHC and lipid biosynthesis (Wicker-Thomas et al., 2015).

Our data set is the first transcriptomic data for adult mosquito oenocytes. Limited depth transcriptional analysis of larval *Ae. aegypti* oenocytes that persist during early pupal development, and are relatively easily dissected in pure form due to their distinct large size and loose attachment as clumps of cells to the integument (Makki et al., 2014), has previously been performed (Martins et al., 2011). Comparison of the partial oenocyte *Aedes* transcriptome with our adult *Anopheles* oenocyte data set provides insights into key genes potentially involved in CHC synthesis throughout development. Seven genes involved in lipid biosynthesis were detected in *Aedes* larval oenocytes, including one acetyl-coA synthetase (AAEL007283), two elongases (AAEL008219 and AAEL013542), two desaturases (AAEL003611 and AAE004278) and the two orthologs of Cyp4G16 and Cyp4G17 (AAEL004054 and AAEL006824). Clear orthologs for five of these larval oenocyte expressed genes (except for the desaturase AAEL004278) were present in our *An. gambiae* adult oenocyte transcriptome (Figure 4). Further work to characterise *Anopheles* oenocyte transcriptomes at earlier life stages will be facilitated by this FACS approach to enable functional analysis of these cells during mosquito development.

In addition to genes involved in lipid and hydrocarbon biosynthesis, genes associated with the biological processes of synaptic vesicle coating and docking, and nerve impulse transmission were found enriched in the oenocyte transcriptome. The Oeno-Gal4 driver line used to generate the mosquito population with fluorescent oenocytes has a red fluorescent marker (dsRed) under the control of the 3xP3 promoter that drives expression in the eyes and nerve cord. A small contamination of the FACS isolated oenocytes with cells of the nerve cord could be speculated, although nerve cells were not observed when visually observing the isolated cells with confocal microscopy. Moreover, oenocytes have been reported to play a role in the neuronal processes during *D. melanogaster* embryogenesis through the secretion of semaphorin (Sema2a), a peptide that drives axon elongation; ablation of oenocytes results in sensory axon defects similar to the *sema2a* mutant phenotype (Bates and Whitington, 2007).

We functionally validated the role of the fatty acid synthase FAS1899 in CHC biosynthesis, by stably knocking down its expression during mosquito development. Oenocyte specific knock down of FAS1899 was lethal at the L2/L3 larvae stages, showing its important role for the

normal mosquito development, possibly by synthesizing Very Long Chain Fatty Acids (VLCFA) that are utilized at the larvae stage either for waterproofing the respiratory system (Parvy et al., 2012) or for other metabolic purposes. Lethality was also reported for the RNAi-mediated knock down of its ortholog (CG17374) in *D. melanogaster* before adult eclosion (Chung et al., 2014). Silencing of the FAS1899 expression using the polyubiquitin (Ubi) promoter also resulted in high levels of mortality (70-80%), but this time at the pupae stage and during adult emergence. This milder phenotype could be explained by the fact that the Ubi promoter drives lower levels of expression in oenocytes, which is supported also by the quantitative real time PCR data (26% knock down of the FAS1899 in adult progeny of the UAS-FAS1899i x Ubi-A10 Gal4 cross compared to the 70% of knock down seen in L2/L3 progeny of the UAS-FAS1899i x Oeno-Gal4 cross).

The relative expression levels of FAS1899 affect both the quantity and composition of CHCs produced in adult oenocytes. A 25% reduction in the total amount of hydrocarbons was observed for adults surviving knock down of FAS1899 and the CHC profile showed a decrease in the total proportion of methyl branched CHCs and an increase in saturated chains. Silencing Cyp4G16 or Cyp4G17 transcript levels in *An. gambiae* oenocytes by approximately 90 % resulted in high mortality in late pupae, pharate adults and during adult emergence and, in surviving adults, a 50% reduction in the total amount of CHCs was observed (Lynd et al., 2019). The Cyp4G16 and Cyp4G17 P450s catalyse the final decarbonylation step in the cuticular hydrocarbon synthetic pathway, while FAS1899 is thought to catalyse the first step using acetyl-CoA to generate and elongate a fatty acyl-CoA chain. Thus, perturbing both extremes of the pathway can influence the final amount of synthesized hydrocarbons.

Partial knock down of the Desat3050 transcripts in larval oenocytes was correlated with larval lethality, similar to FAS1899 knock down. High levels of mortality were also observed when using the weaker oenocyte (but more widespread driver line). However, no qualitative or quantitative differences in the CHC profile were observed in surviving adults. Further work is required, but it may indicate that Desat3050 catalyses the formation of unsaturated lipids that are not converted to hydrocarbons but are important in development, such that even a slight perturbance in the expression levels of this gene can have severe developmental effect.

Variations in the relative abundance of CHCs on the cuticular surface have been correlated in *Anopheles* mosquitoes with species, karyotype, age and mating status (Caputo et al., 2005, Polerstock et al., 2002). Sex specific differences in the relative abundance of some CHC compounds have also been reported in *An. gambiae* (Caputo et al., 2005), but in contrast to other insects like *Drosophila melanogaster* (Coyne and Oyama, 1995), sexual dimorphism in CHCs in mosquitoes has not been reported. This lack of sex specificity is reflected in the absence of sex specific expression of CHC synthesizing genes in our analysis. However, interestingly we did identify some splice variants of Cyp4G16, encoding for a different C-terminus, to be differentially expressed between male and female oenocytes, but further work is needed to validate this observation. A change in C- terminus is likely to alter the intracellular location of proteins through removal of the ER retention signal. Previous work on females has demonstrated enriched localisation of CYP4G16 on the oenocyte plasma

membrane surface (Balabanidou et al., 2016). It would be interesting to examine males in comparison.

Variation in the abundance of CHCs has been associated in *An. coluzzii* with insecticide resistance; a 30% increase in CHC content has been correlated with a decrease in the penetration rate of pyrethroid insecticides (Balabanidou et al., 2016). Several of the genes implicated in CHC biosynthesis from the results of the current study are expressed at elevated levels in pyrethroid resistant mosquitoes and may provide useful genetic markers for detecting this emerging resistance phenotype. For example FAS1899 is a member of the Cyp4G16 correlation network and is over-expressed in pyrethroid insecticide resistant *An. gambiae* and *An. coluzzii* populations from Burkina Faso and Côte d'Ivoire (data from the IR-TEx web-based application (Ingham et al., 2018)). Thus, this gene could be implicated in cuticular resistance, through the production of a thicker cuticle with more hydrocarbons.

In addition to insecticide exposure, environmental factors can also select for changes in the CHC profile; relative proportions of unsaturated and methyl-branched CHCs altered following exposure to arid conditions in the insectary (Reidenbach et al., 2014). The pleiotropic effect of alterations in CHC composition has important implications. Selection pressures that alter the CHC composition, for example the extensive use of insecticides, or an increase in aridity due to climate change, could have multiple effects on mosquito fitness and impacts on disease transmission. Investigating how the different traits influence one another and how this is regulated by the CHC composition is a key next step to understand how mosquitoes adapt and survive in a changing environment and in response to disease control interventions.

Materials and Methods:

Mosquito rearing and preparation of samples for FACS.

An. gambiae mosquitoes were reared at 28 °C under 80% humidity and at a 12/12-h day/night cycle. Larvae were fed with fish food (TetraMin, Tetra GmbH), and adult mosquitoes were fed ad libitum with 10% sugar. To generate mosquitoes with fluorescent oenocytes we crossed males from the UAS-mCD8: mCherry responder line (Adolfi et al., 2018) with virgin females of the oeno-Gal4 driver line (Lynd et al., 2019). Adult progeny (2-4 days old) were collected, anesthetised on ice and dissected in 1X PBS. The head, thorax and internal tissues (midgut, malpigian tubules and reproductive tissues) were removed and the remaining integument (carcass) was cut open. Each sample (N=12 in total, Supplementary Table 1) consisted of 30 carcasses. Samples were washed twice with 1X PBS and incubated for 30min at 37°C with 0,25% trypsin in 1X PBS. After incubation tissues were washed twice with 1X PBS and homogenised by pipetting up and down in 1X PBS containing 1% fetal bovine serum. Dissociated cells were filtered through a plastic filter mesh (ThermoFisher 70µm Nylon Mesh). For samples used to isolate oenocytes (N=6), cells were immediately used for FACS sorting. In the case of total carcass cells (N=6) total RNA was extracted after filtering using the Arcturus PicoPure RNA extraction kit.

FACS and RNA sequencing

For oenocyte isolation the BD ARIA III Cell Sorter (BD Biosciences) equipped with lasers at 405 and 561nm was used. Cells were gated based on the m-Cherry fluorescence. A sample of cells from wild type G3 mosquitoes with no fluorescence was used as control to define the threshold of fluorescence for isolation. All samples were acquired in Facsdiva software version 8.1 (BD Biosciences). All debris doublets were removed from the analysis. The purity of isolation was initially assessed by visualization of isolated cells. Oenocytes were directly sorted in the extraction buffer of the Arcturus PicoPure RNA extraction kit. Total RNA was extracted based on the manufacturer's instructions, including treatment with DNAse. Generation and amplification (11 cycles) of c-DNA from all samples was done in the Center for Genome Research (University of Liverpool) using the SMART-Seq® v4 Ultra® Low Input RNA Kit, according to manufacturer's instructions. The cDNA samples were purified using AMPure XP beads (Beckman Coulter) and their concentration and quality determined using the Agilent 2100 Bioanalyzer and Agilent's High Sensitivity DNA Kit. Libraries were constructed with a total of 1ng of Smarter amplified material and amplified using 12 cycles of PCR. Quality control was performed by running 1 µl undiluted library on an Agilent Technology 2100 Bioanalyzer using a High Sensitivity DNA kit. Samples were run on a Illumina HiSeq 4000.

Pre-processing of transcriptome data

Illumina adapter sequences were removed from the read files (24 *fastq* files in total: 12 RNA-seq runs with right and left reads) using *cutadapt* 1.2.1 (Martin, 2011) (flag *-O 3*). Low-quality reads were removed using *Sickle* 1.200 (minimum window quality score of Phread = 20, removing reads shorter than 20bp)(Joshi and Fass, 2011), retaining only read pairs in which both left and right reads passed quality filters. These steps were performed by the Liverpool University CGR sequencing facility. Each read file was analysed with *fastqc* 0.11.5 (Andrews, 2014) to confirm the absence of adapters sequences. Overall, 97.12% of reads passed the quality control process (Supplementary Table 1).

Genome data download

The reference gene annotation and assembly of *An. gambiae* was obtained from VectorBase (Giraldo-Calderon et al., 2015) (GFF and FASTA formats, version AgamP4.9).

Gene functional annotations

We obtained the predicted peptides of each gene using *gffread* (Geo, 2019). Then, we annotated their Gene Ontology functional annotations using *eggNOG emapper* 1.0.3 (Huerta-Cepas et al., 2017) (HMM mode, which uses *hmmscan* from *HMMER* 3.2.1 (HMMER 2015)) with the euNOG database of eukaryotic protein annotations (Huerta-Cepas et al., 2016) (eggNOG version 4.5) as a reference. In parallel, we annotated the protein domains using Pfamscan, based on version 31 of the Pfam database (Punta et al., 2012).

Analysis of differential expression

We quantified gene expression using the trimmed, clean reads. Specifically, we used *Salmon* 0.10.2 (Patro et al., 2017) to build an index of transcripts (*salmon index*; using the longest

isoform per gene as a reference), using the quasi-mapping procedure (*--type quasi* flag) and k-mers of length 31 (*-k 31*); and then quantified transcript abundance (*salmon quant*) in each sample using the paired-end read files (using automated library type inference, *-l A* flag), in order to obtain TPM (transcripts per million) values for each gene.

Then, we performed a differential expression analysis between sample groups (female oenocytes vs female carcass cells, male oenocytes vs male carcass cells and female oenocytes vs male oenocytes) using the R *DESeq2* library 1.24.0 (Love et al., 2014). First, we imported the transcript quantification values from *Salmon* (see above) using the *tximport* library 1.12.0 (Soneson et al., 2015). Then, we performed targeted differential expression analyses between groups of samples using the *DESeq* function from *DESeq2* (using the Wald procedure for significance testing), produced a table of normalised gene counts per sample using the *counts* function (using *DESeq2* normalisation factors), and obtained the fold changes and *p*-values from a Wald test for each gene, using the *results* command (using a Benjamini-Hochberg [FDR] *p*-value correction (Benjamini and Hochberg, 1995) and an alpha threshold = 0.001, and all combinations of samples from Supplementary Table 1 to define the *contrast* parameter). The log-fold change values were corrected (shrunken) with *lfcShrink* and the *apeglm* algorithm (Zhu et al., 2019). We defined a gene as being differentially expressed in a given comparison if the adjusted $p < 0.001$ and the absolute shrunken log-fold change > 1 (i.e. absolute fold change > 2).

We explored the variation in gene expression across samples using the normalised gene counts (log-transformed, and standardized to mean = 0 and standard deviation = 1 using the *scale* R function). First, we performed a Principal Components analysis (PCA) using the normalised gene counts of each sample (*prcomp* function of the R *stats* library).

Heatmaps of gene expression for selected genes

To visualise changes in expression for genes involved in CHC biosynthesis, we produced heatmaps of gene expression by plotting the normalised gene counts of each gene in each sample (*pheatmap* function from the *pheatmap* 1.012 R library (Kolde 2019), using Pearson correlation values to set the order of genes).

Analysis of alternative splicing

We used *SUPPPA2* (Trincado et al., 2018) to generate a set of alternative splicing events from the annotated isoforms in the *An. gambiae* genome (GFF file from Vectorbase, AgamP4.9), using the *generateEvents* mode to detect retained introns, skipped exons, and alternative first or last exons, and mutually exclusive exons (*-e SE MX RI SS FL*), with 10bp as the minimum exon length (*-l 10*). We also calculated the expression levels at the isoform level using *Salmon* 0.10.2 (Patro et al., 2017) (output in TPM). Then, we used *SUPPA2 psiPerIsoform* mode to calculate the inclusion rates of each isoform (PSI: percentage spliced-in) in each sample, using the expression levels of each isoform (obtained from *Salmon*) as a reference. Differential splicing was quantified by calculating the calculating the average difference in PSI values between each sample group (male/female oenocytes and carcasses), and *p*-values were obtained using the empirical significance calculation method described in *SUPPA2 (Trincado et al., 2018).*

The PSI values of selected differentially spliced genes ($p<0.05$) belonging to the biosynthesis pathway were reported using a heatmap table (*pheatmap* function from the *pheatmap* 1.012 R library).

Gene functional enrichment analysis

Gene Ontology enrichments based on the GOs annotated with *eggNOG* mapper (see above) were computed using the *topGO* R library (2.34) (Alexa and Rahnenfuhrer 2018). Specifically, we computed the functional enrichments based on the counts of genes belonging to the group of interest relative to all annotated genes, using Fisher's exact test and the *elim* algorithm for GO graph weighting (Alexa et al., 2006).

Functional enrichment tests of Pfam domain annotations were performed using hypergeometric tests as implemented in the R stats 3.6 library (*phyper*) (R Core Team 2017), comparing the frequencies of presence of Pfam domains in a list of genes of interest to the same frequencies in the whole gene set (using unique domains per gene). We adjusted *p* values using the Benjamini-Hochberg procedure.

Construction of phylogenetic trees

We retrieved genes belonging to gene family-members of the fatty acid biosynthesis pathway from the proteomes of *An. gambiae* (Vectorbase, AgamP4.9 annotation), *Ae. aegypti* (Vectorbase LVP_AGWG AaegL5.1 annotation) and *D. melanogaster* (Flybase r6.21 annotation). Specifically, we defined the list of candidate genes for phylogenetic analysis according to the presence of the following catalytic Pfam domains: FA_desaturase (PF00487) for desaturases (totalling 29 individual domains), ELO (PF01151) for elongases (62), NAD_binding_4 (PF07993) for reductases (61), ketoacyl-synt (PF00109) for synthases (16). Pfam annotations were obtained from Pfamscan as described above. Functional domain sequence sets were aligned using *MAFFT* 7.310 (1,000 rounds of iterative refinement, L-INS-i algorithm)(Katoh and Standley, 2013), and later trimmed position-wise using trimAL 1.4 (*automated1* procedure)(Capella-Gutierrez et al., 2009). The trimmed alignments were used to build maximum-likelihood phylogenetic trees for each gene family, using *IQ-TREE* 1.6.10(Nguyen et al., 2015). The best-fitting evolutionary model (LG substitution matrix (Le and Gascuel, 2008) with four Γ categories and accounting for invariant sites, or LG+I+G4) was selected for each gene family according to the BIC criterion. Phylogenetic statistical supports were calculated using the UF bootstrap procedure (1,000 replicates) (Hoang et al., 2018).

The resulting phylogenetic trees were mid-point rooted using the *R* phangorn 2.53 library (Schliep, 2011), and visualisations were produced using the *phytools* 0.6-60 (Revell, 2012) and *ape* 5.3 libraries (plot.phylo)(Paradis and Schliep, 2019).

Plasmid construction and generation of the UAS FAS1899RNAi responder line by PhiC31-Mediated Cassette Exchange

A UAS responder plasmid was generated for the expression of dsRNA targeting the third exon of the AGAP001899 gene. Specifically 200bp inverted repeats separated by the 203bp fourth

intron of the *Drosophila melanogaster* white eye gene (CG2759) were synthesized by GeneScript and cloned into the YFP-marked responder plasmid pSL*attB:YFP:Gyp:UAS14i:Gyp:attB (Lynd et al., 2019) downstream of the UAS using EcoRI/NheI restriction enzymes. The intron of the *Drosophila* white eye gene was used because all internal introns of the AGAP001899 gene were shorter than 100bp, making the synthesis of the 200bp inverted repeats impossible. Embryo injections were performed using the A11 docking line (Lynd et al., 2019), which carries 2 inverted attP sites and is marked with 3xP3-driven CFP. 350 ng/µL of the responder plasmid and 150 ng/µL of the integrase helper plasmid pKC40 encoding the phiC31 integrase (Ringrose, 2009) were injected as described in Pondeville et al.(Pondeville et al., 2014). Emerging F0 individuals were outcrossed with wild type G3 individuals of the opposite sex. The F1 generation was screened for the expression of the YFP marker in the eyes and nerve cord and the absence of the CFP marker, indicating the successful cassette exchange. The direction of the cassette exchange was determined as described in Adolfi et.al (Adolfi et al., 2019) and shown to be of the A orientation. The FAS1899 RNAi and Desat3050 RNAi responder lines that were established were kept as a mix of homozygous and heterozygous individuals so as to obtain Gal4/+ progeny after crossing with the Gal4 driver lines and obtain siblings that serve as transgenic blank controls.

Crosses of transgenic lines and qRT PCR for gene expression analysis

Crosses were performed between the responder lines UAS-FAS1899i , UAS-Desat3050 and the two Gal4 lines: oenocyte specific-GAL4 (Oeno-Gal4) (Lynd et al., 2019) and Ubi-A10 Gal4 line (Adolfi et al., 2018). Progeny (at least 10 individuals for each group, pooled in 2-3 biological replicates) of these crosses, YFP marked and blank (control siblings), were collected either at the L2-L3 stage (for the cross with the oeno-Gal4 line) or at the adult stage (for the cross with the Ubi-A10 line) and used to extract RNA with the PicoPure RNA isolation kit (Thermo Fisher Scientific) and treated with DNase using the Qiagen RNase-free DNase kit. 2ugr of RNA were reverse transcribed using SuperScript III (Invitrogen) and oligo(dT)20 primers to produce cDNA. Expression of AGAP001899 (FAS1899) and AGAP003050 (Desat3050) was validated by qPCR using the following primers: (FAS1899 Forward: 5'-AGCGATCTGCGTGATGTACC-3', FAS1899 Reverse: 5'-GCCTTCCTCCTTAAACCCGTC-3', Desat3050 Forward: 5' CCGTACTACAGCGACAAGGAC-3', Desat3050 Reverse 5'- GAACATCACAATACCGTCCGC-3') and reference gene for normalization the Ribosomal S7 (AGAP010592) (Forward: 5'-AGAACCAGCAGACCACCATC-3' Reverse: 5'-GCTGCAAACTTCGGCTATTC-3'). Expression analysis was performed according to Pfaffl,2001 (Pfaffl, 2001).

Extraction of Cuticular Hydrocarbons and analysis with GC-MS

CHCs were extracted from pools of adult (3-5 days old) mosquitoes (each pool consisted of 2-5 mosquitoes depending on availability, at least 3 pools per condition) by immersing them and gently agitating them, for 10 min at room temperature, in 200 µl of hexane (Sigma-Aldrich) spiked with 1ng/ml of octadecane (Sigma-Aldrich) as internal standard. Hexane extracts were concentrated under a N_2 stream and 2µl injected in a Waters GCT gas

chromatograph-mass spectrometer. The GC column was a 30 m long, 0.25 mm internal diameter, 0.25 µm film thickness BPX5 (SGE). The oven temperature gradient was 50°C to 370°C at 10°C/minute and the carrier gas was helium (BOC) at a flow rate of 1 ml/minute. The scan range was *m/z* 40 to 450 Da in scan time 0.9 s. Compounds were identified based on their mass spectra in comparison to those of an alkane standard mixture (C10-C40, Merck 68281-2ML-F), by comparison of their retention times and fragmentation patterns to published *Anopheles gambiae* CHC mass spectra (Balabanidou et al., 2016) and searches of the NIST mass spectrum library supplied with Waters MassLynx software. Peak areas were measured manually using the peak integration tool in the Waters MassLynx software. The total amount of hydrocarbon present was calculated by summing all the peak areas measured relative to the area of the internal standard. Student's t-test was performed for the statistical analysis of differences in total CHC amount and relative abundance of CHC categories.

Availability of data and materials

Transcriptome sequencing has been deposited in the European Nucleotide Archive (ENA), under **PRJEB37240** project. All transgenic lines produced in this study will be provided by L.G upon request.

All data and code (in R) required to perform the differential expression, alternative splicing and phylogenetic analyses in this paper is available in the following Github repository: https://github.com/xgrau/oenocytes-agam

Acknowledgements

We would like to thank Jesus Reiner (LSTM) for performing the FACS; Mark Prescott and Rob Beynon (University of Liverpool) who performed the CHC analysis at the Centre for Proteome Research, University of Liverpool; Simon Wagstaff (LSTM) for his advice on the RNAseq ; Andriana Adolfi for providing the Ubi-A10 line; Rachel Davies (LSTM) for her assistance with mosquito rearing, Amalia Anthousi and Fraser Colman (LSTM) for long term maintenance and provision of stock driver and docking lines and Manuela Bernardi for preparation of figures. Flow cytometric sorting was performed on a BD FacsAria III funded by a Welcome Trust Multi-User Equipment Grant (104936/Z/14/Z). This study was funded by a Director Catalyst Fund (pump priming award from LSTM) and a Sir Henry Wellcome Postdoctoral fellowship (Grant reference number: 215894/Z/19/Z) to L.G.

References

ADOLFI, A., PONDEVILLE, E., LYND, A., BOURGOUIN, C. & LYCETT, G. J. 2018. Multi-tissue GAL4-mediated gene expression in all Anopheles gambiae life stages using an endogenous polyubiquitin promoter. *Insect Biochem Mol Biol,* 96**,** 1-9.

ADOLFI, A., POULTON, B., ANTHOUSI, A., MACILWEE, S., RANSON, H. & LYCETT, G. J. 2019. Functional genetic validation of key genes conferring insecticide resistance in the major African malaria vector, Anopheles gambiae. *Proc Natl Acad Sci U S A,* 116**,** 25764-25772.

ALEXA, A., RAHNENFUHRER, J. & LENGAUER, T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics,* 22**,** 1600-7.

ANDREWS, S. 2014. FastQC. http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc

ARCAZ, A. C., HUESTIS, D. L., DAO, A., YARO, A. S., DIALLO, M., ANDERSEN, J., BLOMQUIST, G. J. & LEHMANN, T. 2016. Desiccation tolerance in Anopheles coluzzii: the effects of spiracle size and cuticular hydrocarbons. *J Exp Biol,* 219**,** 1675-88.

BALABANIDOU, V., GRIGORAKI, L. & VONTAS, J. 2018. Insect cuticle: a critical determinant of insecticide resistance. *Curr Opin Insect Sci,* 27**,** 68-74.

BALABANIDOU, V., KAMPOURAKI, A., MACLEAN, M., BLOMQUIST, G. J., TITTIGER, C., JUAREZ, M. P., MIJAILOVSKY, S. J., CHALEPAKIS, G., ANTHOUSI, A., LYND, A., ANTOINE, S., HEMINGWAY, J., RANSON, H., LYCETT, G. J. & VONTAS, J. 2016. Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in Anopheles gambiae. *Proc Natl Acad Sci U S A,* 113**,** 9268-73.

BATES, K. E. & WHITINGTON, P. M. 2007. Semaphorin 2a secreted by oenocytes signals through plexin B and plexin A to guide sensory axons in the Drosophila embryo. *Dev Biol,* 302**,** 522-35.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B,* 57**,** 289-300.

BHATT, S., WEISS, D. J., CAMERON, E., BISANZIO, D., MAPPIN, B., DALRYMPLE, U., BATTLE, K., MOYES, C. L., HENRY, A., ECKHOFF, P. A., WENGER, E. A., BRIET, O., PENNY, M. A., SMITH, T. A., BENNETT, A., YUKICH, J., EISELE, T. P., GRIFFIN, J. T., FERGUS, C. A., LYNCH, M., LINDGREN, F., COHEN, J. M., MURRAY, C. L. J., SMITH, D. L., HAY, S. I., CIBULSKIS, R. E. & GETHING, P. W. 2015. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature,* 526**,** 207-211.

BLOMQUIST, G. J., BAGNÈRES, A. G., WICKER-THOMAS, C., CHERTEMPS, T., GIBBS, A., RAJPUROHIT, S., MILLAR, J., BUCKNER, J. S., OZAKI, M., WADA-KATSUMATA, A., VAN ZWEDEN, J. S., D'ETTORRE, P., GREEN, M., LIEBIG, J., LORENZI, C., FERVEUR, J. F., COBB, M., TRABALON, M., GINZEL, M. D., BARTELT, R. & HEFETZ, A. 2010. Insect Hydrocarbons Biology, Biochemistry, and Chemical Ecology.

CAPELLA-GUTIERREZ, S., SILLA-MARTINEZ, J. M. & GABALDON, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics,* 25**,** 1972-1973.

CAPUTO, B., DANI, F. R., HORNE, G. L., PETRARCA, V., TURILLAZZI, S., COLUZZI, M., PRIESTMAN, A. A. & DELLA TORRE, A. 2005. Identification and composition of cuticular hydrocarbons of the major Afrotropical malaria vector Anopheles gambiae s.s. (Diptera: Culicidae): analysis of sexual dimorphism and age-related changes. *J Mass Spectrom,* 40**,** 1595-604.

CHERTEMPS, T., DUPORTETS, L., LABEUR, C., UEDA, R., TAKAHASHI, K., SAIGO, K. & WICKER-THOMAS, C. 2007. A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in Drosophila melanogaster. *Proc Natl Acad Sci U S A,* 104**,** 4273-8.

CHERTEMPS, T., DUPORTETS, L., LABEUR, C., UEYAMA, M. & WICKER-THOMAS, C. 2006. A female-specific desaturase gene responsible for diene hydrocarbon biosynthesis and courtship behaviour in Drosophila melanogaster. *Insect Mol Biol,* 15**,** 465-73.

CHIANG, Y. N., TAN, K. J., CHUNG, H., LAVRYNENKO, O., SHEVCHENKO, A. & YEW, J. Y. 2016. Steroid Hormone Signaling Is Essential for Pheromone Production and Oenocyte Survival. *PLoS Genet,* 12**,** e1006126.

CHUNG, H. & CARROLL, S. B. 2015. Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. *Bioessays,* 37**,** 822-30.

679 CHUNG, H., LOEHLIN, D. W., DUFOUR, H. D., VACCARRO, K., MILLAR, J. G. & CARROLL, S. B. 2014. A
680 single gene affects both ecological divergence and mate choice in Drosophila. *Science,* 343**,**
681 1148-51.
682 COYNE, J. A. & OYAMA, R. 1995. Localization of pheromonal sexual dimorphism in Drosophila
683 melanogaster and its effect on sexual isolation. *Proc Natl Acad Sci U S A,* 92**,** 9505-9.
684 DALLERAC, R., LABEUR, C., JALLON, J. M., KNIPPLE, D. C., ROELOFS, W. L. & WICKER-THOMAS, C.
685 2000. A delta 9 desaturase gene with a different substrate specificity is responsible for the
686 cuticular diene hydrocarbon polymorphism in Drosophila melanogaster. *Proc Natl Acad Sci U*
687 *S A,* 97**,** 9449-54.
688 DEMBECK, L. M., BOROCZKY, K., HUANG, W., SCHAL, C., ANHOLT, R. R. & MACKAY, T. F. 2015.
689 Genetic architecture of natural variation in cuticular hydrocarbon composition in Drosophila
690 melanogaster. *Elife,* 4.
691 DILLWITH, J. W., BLOMQUIST, G. J. & NELSON, D. R. 1981. Biosynthesis of the Hydrocarbon
692 Components of the Sex-Pheromone of the Housefly, Musca-Domestica L. *Insect*
693 *Biochemistry,* 11**,** 247-253.
694 GEO, P. 2019. gffread: GFF/GTF utility providing format conversions, region filtering, FASTA sequence
695 extraction and more. https://github.com/gpertea/gffread
696 GIRALDO-CALDERON, G. I., EMRICH, S. J., MACCALLUM, R. M., MASLEN, G., DIALYNAS, E., TOPALIS,
697 P., HO, N., GESING, S., VECTORBASE, C., MADEY, G., COLLINS, F. H. & LAWSON, D. 2015.
698 VectorBase: an updated bioinformatics resource for invertebrate vectors and other
699 organisms related with human diseases. *Nucleic Acids Res,* 43**,** D707-13.
700 HOANG, D. T., CHERNOMOR, O., VON HAESELER, A., MINH, B. Q. & VINH, L. S. 2018. UFBoot2:
701 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol,* 35**,** 518-522.
702 HUANG, K., CHEN, W., ZHU, F., LI, P. W., KAPAHI, P. & BAI, H. 2019. RiboTag translatomic profiling of
703 Drosophila oenocytes under aging and induced oxidative stress. *BMC Genomics,* 20**,** 50.
704 HUERTA-CEPAS, J., FORSLUND, K., COELHO, L. P., SZKLARCZYK, D., JENSEN, L. J., VON MERING, C. &
705 BORK, P. 2017. Fast Genome-Wide Functional Annotation through Orthology Assignment by
706 eggNOG-Mapper. *Mol Biol Evol,* 34**,** 2115-2122.
707 HUERTA-CEPAS, J., SZKLARCZYK, D., FORSLUND, K., COOK, H., HELLER, D., WALTER, M. C., RATTEI, T.,
708 MENDE, D. R., SUNAGAWA, S., KUHN, M., JENSEN, L. J., VON MERING, C. & BORK, P. 2016.
709 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for
710 eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res,* 44**,** D286-93.
711 INGHAM, V. A., WAGSTAFF, S. & RANSON, H. 2018. Transcriptomic meta-signatures identified in
712 Anopheles gambiae populations reveal previously undetected insecticide resistance
713 mechanisms. *Nat Commun,* 9**,** 5282.
714 JOSHI, N. & FASS, J. 2011. Windowed Adaptive Trimming for fastq files using quality.
715 *https://github.com/najoshi/sickle*
716 KATOH, K. & STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7:
717 improvements in performance and usability. *Mol Biol Evol,* 30**,** 772-80.
718 KEFI, M., BALABANIDOU, V., DOURIS, V., LYCETT, G., FEYEREISEN, R. & VONTAS, J. 2019. Two
719 functionally distinct CYP4G genes of Anopheles gambiae contribute to cuticular hydrocarbon
720 biosynthesis. *Insect Biochem Mol Biol,* 110**,** 52-59.
721 LE, S. Q. & GASCUEL, O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol,*
722 25**,** 1307-20.
723 LOCKEY, K. H. 1988. Lipids of the Insect Cuticle - Origin, Composition and Function. *Comparative*
724 *Biochemistry and Physiology B-Biochemistry & Molecular Biology,* 89**,** 595-645.
725 LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for
726 RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.
727 LYCETT, G. J., MCLAUGHLIN, L. A., RANSON, H., HEMINGWAY, J., KAFATOS, F. C., LOUKERIS, T. G. &
728 PAINE, M. J. 2006. Anopheles gambiae P450 reductase is highly expressed in oenocytes and
729 in vivo knockdown increases permethrin susceptibility. *Insect Mol Biol,* 15**,** 321-7.

LYND, A., BALABANIDOU, V., GROSMAN, R., MAAS, J., LIAN, L. Y., VONTAS, J. & LYCETT, G. 2019. Development of a functional genetic tool for Anopheles gambiae oenocyte characterisation: application to cuticular hydrocarbon synthesis. *bioRxiv,* doi: https://doi.org/10.1101/742619.

LYND, A. & LYCETT, G. J. 2012. Development of the bi-partite Gal4-UAS system in the African malaria mosquito, Anopheles gambiae. *PLoS One,* 7**,** e31552.

MAKKI, R., CINNAMON, E. & GOULD, A. P. 2014. The development and functions of oenocytes. *Annu Rev Entomol,* 59**,** 405-25.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. . *EMBnet.journal* 17.

MARTINS, G. F., RAMALHO-ORTIGAO, J. M., LOBO, N. F., SEVERSON, D. W., MCDOWELL, M. A. & PIMENTA, P. F. 2011. Insights into the transcriptome of oenocytes from Aedes aegypti pupae. *Mem Inst Oswaldo Cruz,* 106**,** 308-15.

NGUYEN, L. T., SCHMIDT, H. A., VON HAESELER, A. & MINH, B. Q. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol,* 32**,** 268-74.

PARADIS, E. & SCHLIEP, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics,* 35**,** 526-528.

PARVY, J. P., NAPAL, L., RUBIN, T., POIDEVIN, M., PERRIN, L., WICKER-THOMAS, C. & MONTAGNE, J. 2012. Drosophila melanogaster Acetyl-CoA-carboxylase sustains a fatty acid-dependent remote signal to waterproof the respiratory system. *PLoS Genet,* 8**,** e1002925.

PATRO, R., DUGGAL, G., LOVE, M. I., IRIZARRY, R. A. & KINGSFORD, C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods,* 14**,** 417-419.

PFAFFL, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res,* 29**,** e45.

POLERSTOCK, A. R., EIGENBRODE, S. D. & KLOWDEN, M. J. 2002. Mating alters the cuticular hydrocarbons of female Anopheles gambiae sensu stricto and aedes Aegypti (Diptera: Culicidae). *J Med Entomol,* 39**,** 545-52.

PONDEVILLE, E., PUCHOT, N., MEREDITH, J. M., LYND, A., VERNICK, K. D., LYCETT, G. J., EGGLESTON, P. & BOURGOUIN, C. 2014. Efficient PhiC31 integrase-mediated site-specific germline transformation of Anopheles gambiae. *Nat Protoc,* 9**,** 1698-712.

PUNTA, M., COGGILL, P. C., EBERHARDT, R. Y., MISTRY, J., TATE, J., BOURSNELL, C., PANG, N., FORSLUND, K., CERIC, G., CLEMENTS, J., HEGER, A., HOLM, L., SONNHAMMER, E. L., EDDY, S. R., BATEMAN, A. & FINN, R. D. 2012. The Pfam protein families database. *Nucleic Acids Res,* 40**,** D290-301.

QIU, Y., TITTIGER, C., WICKER-THOMAS, C., LE GOFF, G., YOUNG, S., WAJNBERG, E., FRICAUX, T., TAQUET, N., BLOMQUIST, G. J. & FEYEREISEN, R. 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A,* 109**,** 14858-63.

REIDENBACH, K. R., CHENG, C., LIU, F., LIU, C., BESANSKY, N. J. & SYED, Z. 2014. Cuticular differences associated with aridity acclimation in African malaria vectors carrying alternative arrangements of inversion 2La. *Parasit Vectors,* 7**,** 176.

REVELL, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution,* 3**,** 217-223.

RINGROSE, L. 2009. Transgenesis in Drosophila melanogaster. *Methods Mol Biol,* 561**,** 3-19.

SCHLIEP, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics,* 27**,** 592-3.

SONESON, C., LOVE, M. I. & ROBINSON, M. D. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res,* 4**,** 1521.

TRINCADO, J. L., ENTIZNE, J. C., HYSENAJ, G., SINGH, B., SKALIC, M., ELLIOTT, D. J. & EYRAS, E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol,* 19**,** 40.

781 WICKER-THOMAS, C., GARRIDO, D., BONTONOU, G., NAPAL, L., MAZURAS, N., DENIS, B., RUBIN, T.,
782 PARVY, J. P. & MONTAGNE, J. 2015. Flexible origin of hydrocarbon/pheromone precursors in
783 Drosophila melanogaster. *J Lipid Res,* 56**,** 2094-101.
784 ZHU, A., IBRAHIM, J. G. & LOVE, M. I. 2019. Heavy-tailed prior distributions for sequence count data:
785 removing the noise and preserving large differences. *Bioinformatics,* 35**,** 2084-2092.

786

787

788

789

790

791

792

Supplementary Information for:

Cuticular hydrocarbon biosynthesis in malaria vectors: insights from the adult oenocyte transcriptome.

Linda Grigoraki, Xavier Grau-Bove, Henrietta Carrington-Yates, Gareth J Lycett and Hilary Ranson

Corresponding authors: Linda Grigoraki and Hilary Ranson

Linta.Grigoraki@lstmed.ac.uk, Hilary.Ranson@lstmed.ac.uk

This PDF file includes:

Supplementary text: Differential expression of splice isoforms in oenocytes

Figures S1-S8

Tables S1-S3

Legends for Datasets S1-S4

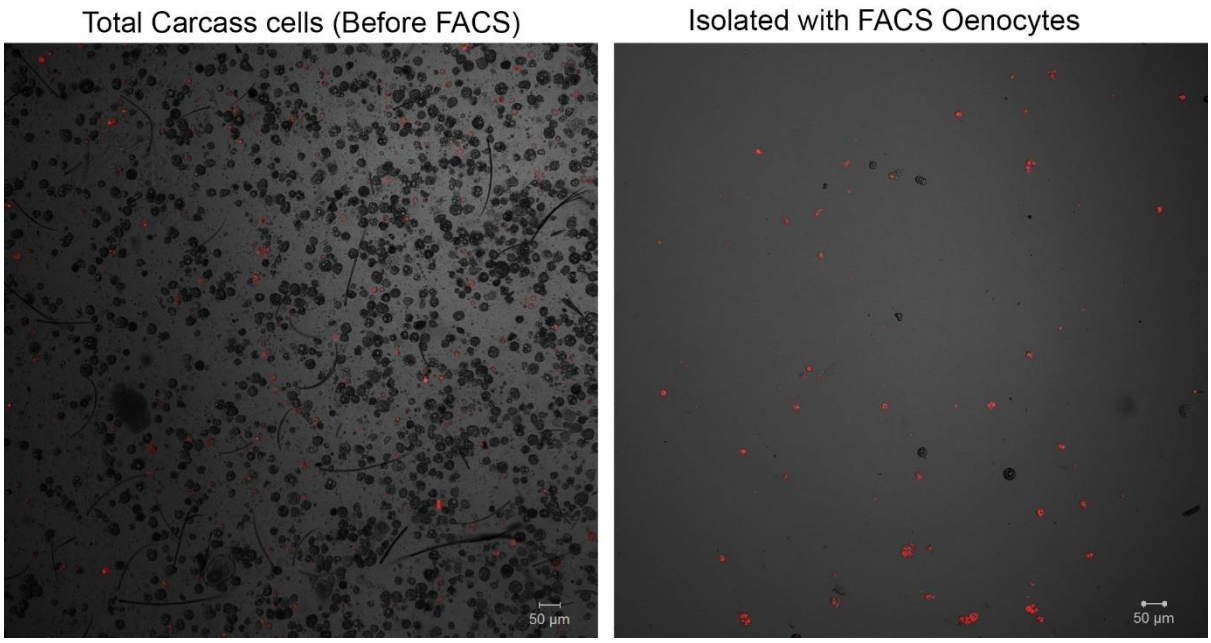
Supplementary text:

Differential expression of splice isoforms in oenocytes

We investigated whether specific gene isoforms are differentially expressed in oenocytes (at a $p<0.05$). 672 genes were found to have at least one isoform differentially expressed in female oenocytes compared to female carcass cells and 752 to have at least one isoform differentially expressed in male oenocytes compare to male carcass cells. The same analysis was performed for female and male oenocytes showing 578 genes to have at least one isoform differentially expressed between sexes (Supplementary File 2 and Supplementary Figure 3A). Five genes belonging to one of the six gene families implicated in the hydrocarbon biosynthetic pathway (the elongase AGAP004373, the desaturases AGAP003051, AGAP004572 and AGAP01713 and the decarbonylase P450 Cyp4G16) had at least one isoform differentially expressed in at least two of the three comparisons (female oenocytes vs female total carcass cells, male oenocytes vs male total carcass cells and female vs male oenocytes) (Supplementary Figure 3B). Most of the isoforms for these genes differ solely in the untranslated regions. Exceptions are the RD isoform of Cyp4G16 that encodes for a slightly truncated protein with a different C-terminus (last 19 a.a) compared to the other isoforms, and isoforms RA and RB of the desaturase AGAP003051, which encode proteins with highly diverged C-termini. We need however to point out that the two predicted isoforms for AGAP003051 might be affected by some annotation error, as the AGAP003051-RB isoform is identical with the adjacent AGAP003050 transcript after nucleotide 407 (total length of 1038 nt). The AGAP003051-RB isoform was more abundant in oenocytes compared to total Carcass cells and more abundant in female oenocytes compared to male oenocytes, although this latter difference was clearly driven by one of the male oenocyte replicates. The Cyp4G16-RD isoform was enriched in female oenocytes in comparison to both female carcass cells and male oenocytes (Supplementary Figure 3).

831

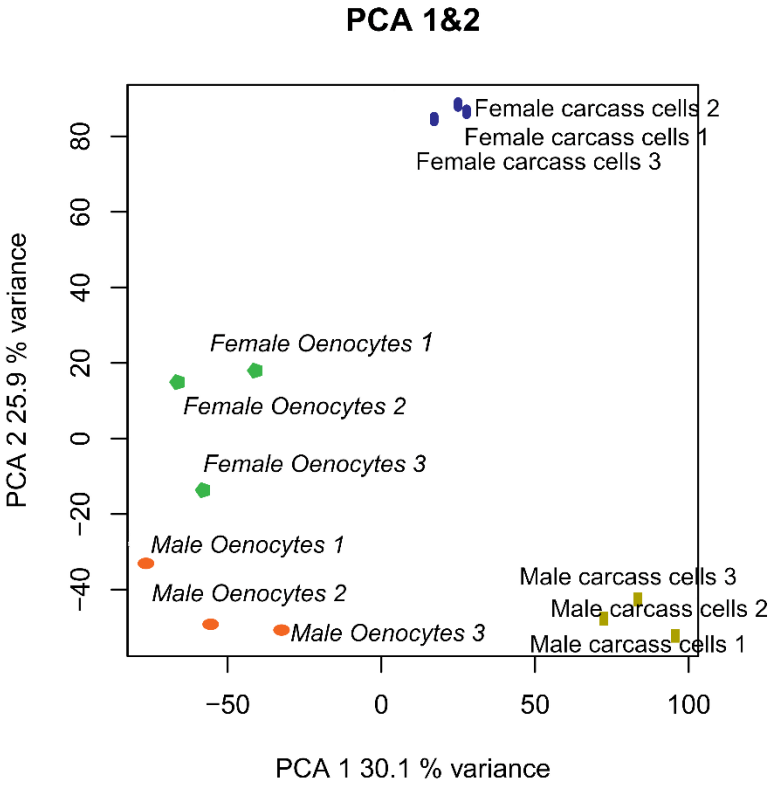
832



833

834 **Fig S1**: Representative confocal microscopy image for isolated oenocytes. A) carcass cells dissociated
835 from transgenic mosquitoes (progeny of UAS-mCD8: mCherry line and Oeno-Gal4 driver line) with
836 fluorescent oenocytes (pre-sorted total carcass cells sample) (objective 10X) and B) confocal
837 microscopy image for cells isolated with FACS (sample of isolated oenocytes) (objective 10X).

838



839

840 **Fig S2**: PCA analysis for the twelve samples used in RNAseq.

841

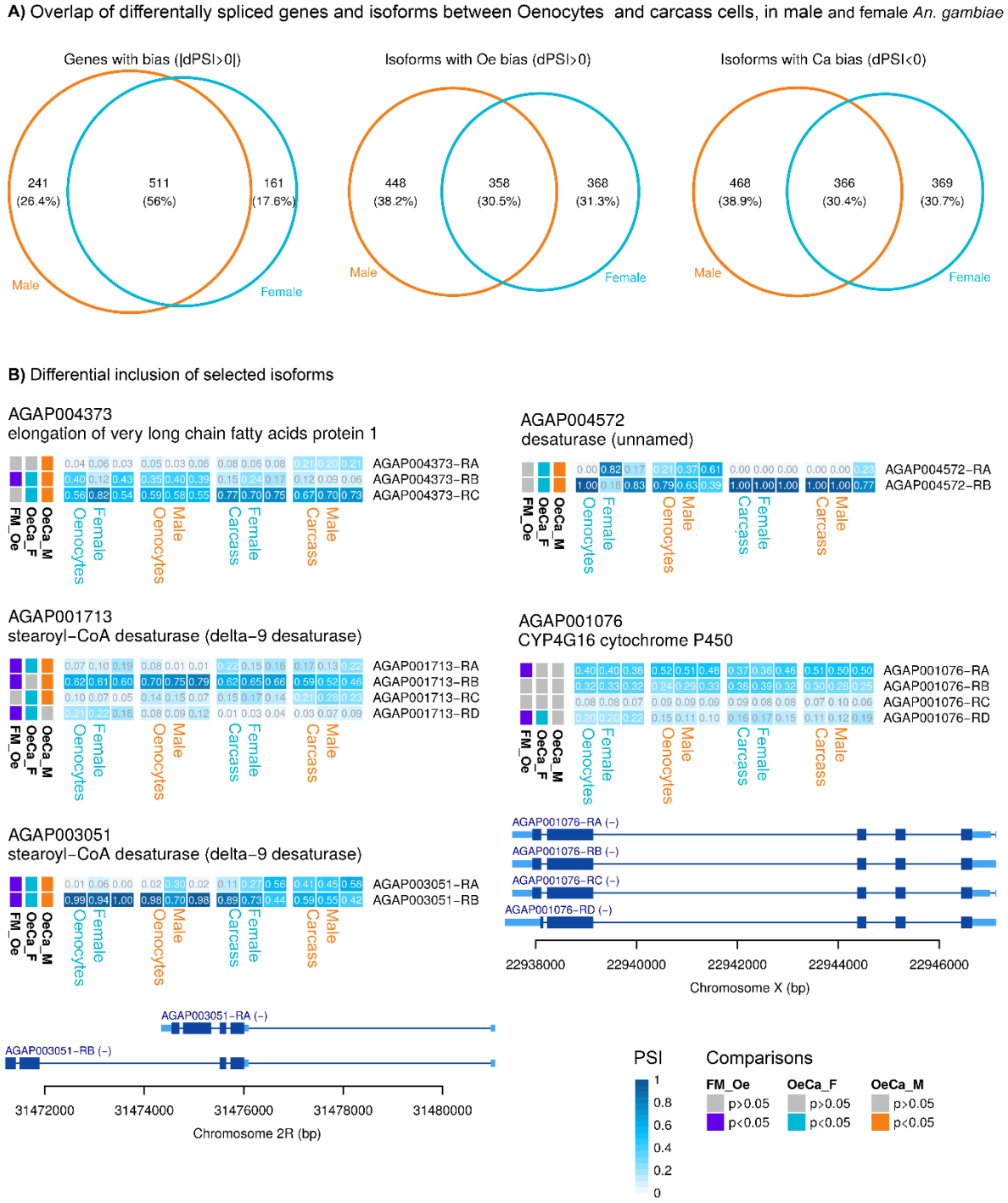
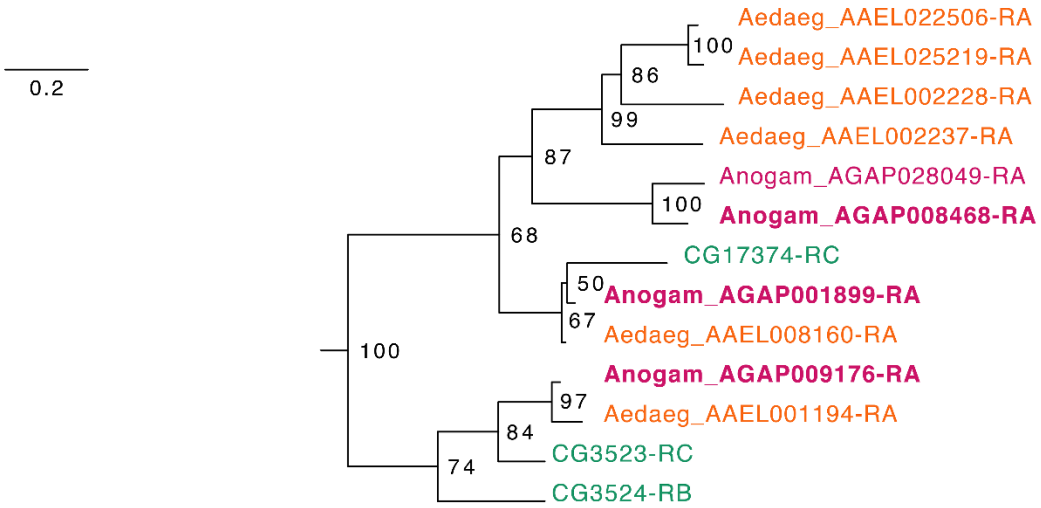


Fig S3: Splice variant analysis. A) Venn diagram (on the left) showing the number of genes with differential expression of at least one isoform in female and/or male oenocytes compared to total carcass cells. Venn diagrams showing the number of gene isoforms with enriched (middle diagram) or reduced (right diagram) expression in female and/or male oenocytes compared to total carcass cells. B) Heat maps showing the frequency (PSI) of isoforms (in each sample used for RNAseq) for genes belonging to gene families implicated in CHC biosynthesis. Isoforms that encode for different proteins are depicted. Comparisons performed: Female Oenocytes vs Female total carcass cells (OeCa_F), Male Oenocytes vs Male total carcass cells (OeCa_M), Female Oenocytes vs Male Oenocytes (FM_Oe). Source data: Supplementary File 2.

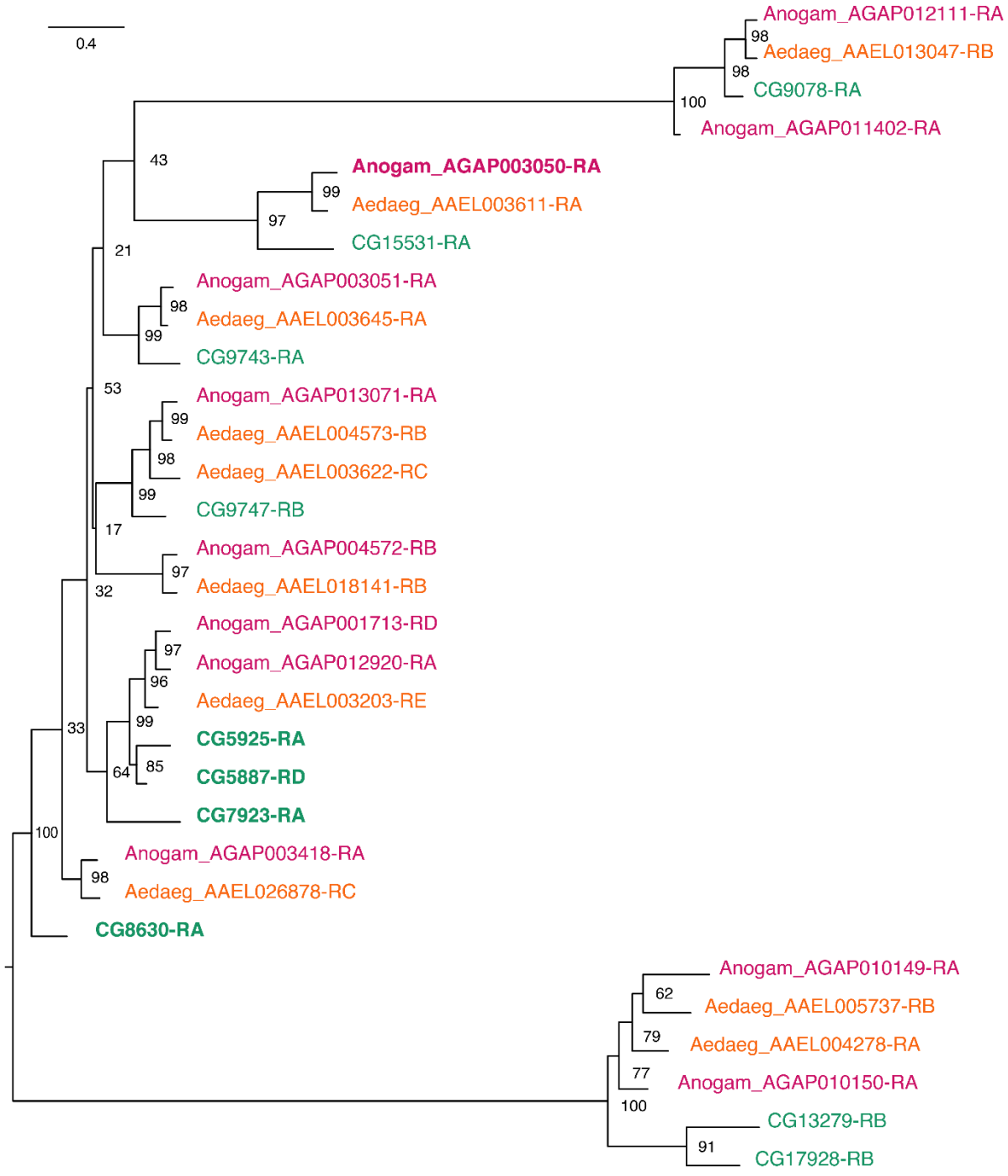


852

853 **Fig S4**: Phylogenetic tree for *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* Fatty
854 acid synthases. Scale bar shows the number of aminoacid substitutions per alignment position. Node
855 supports are % of bootstrap supports on 1,000 replications.

856

857



858

859 **Fig S5**: Phylogenetic tree for *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster*
860 Desaturases. Scale bar shows the number of aminoacid substitutions per alignment position. Node
861 supports are % of bootstrap supports on 1,000 replications.

862

863

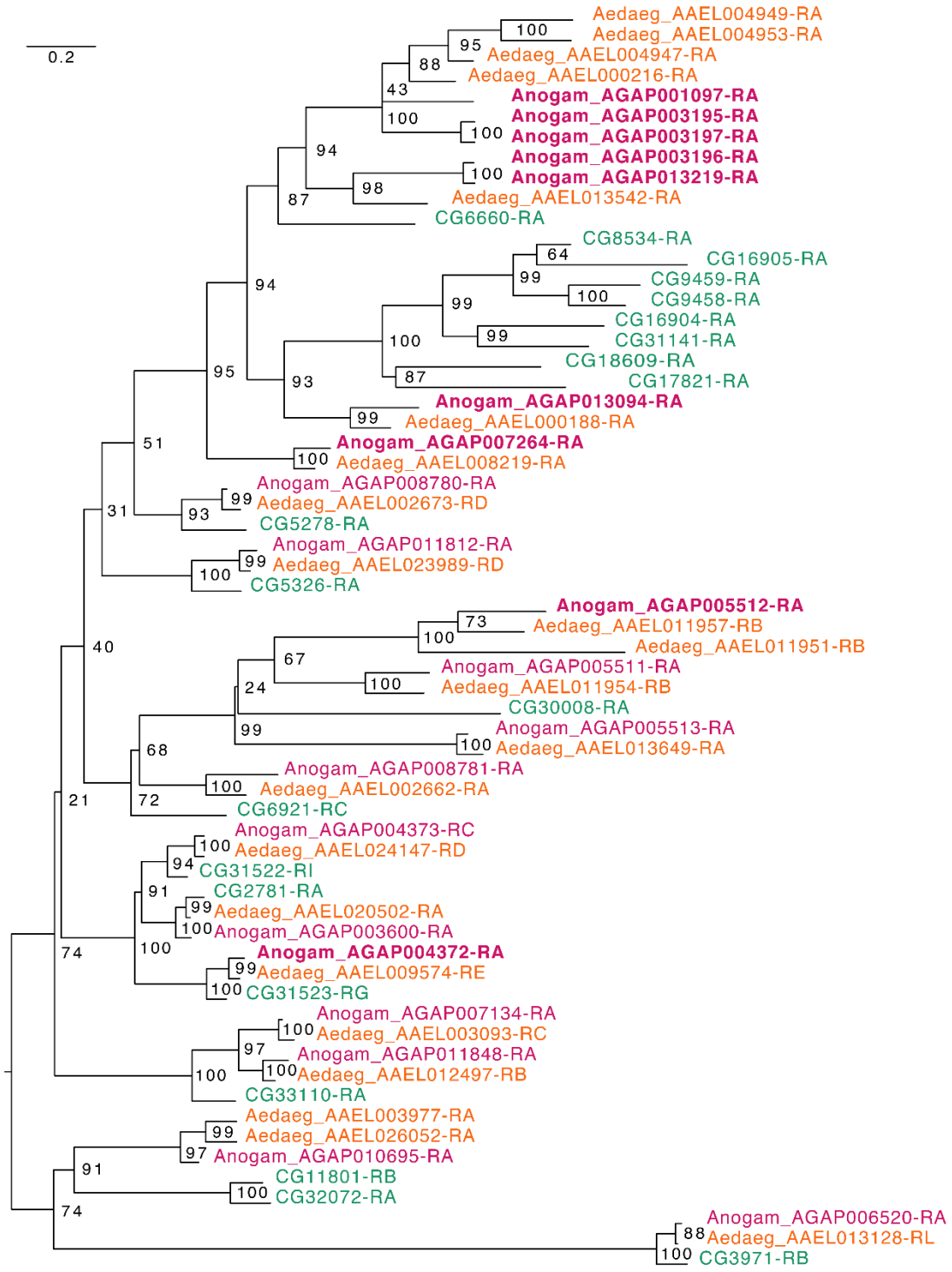
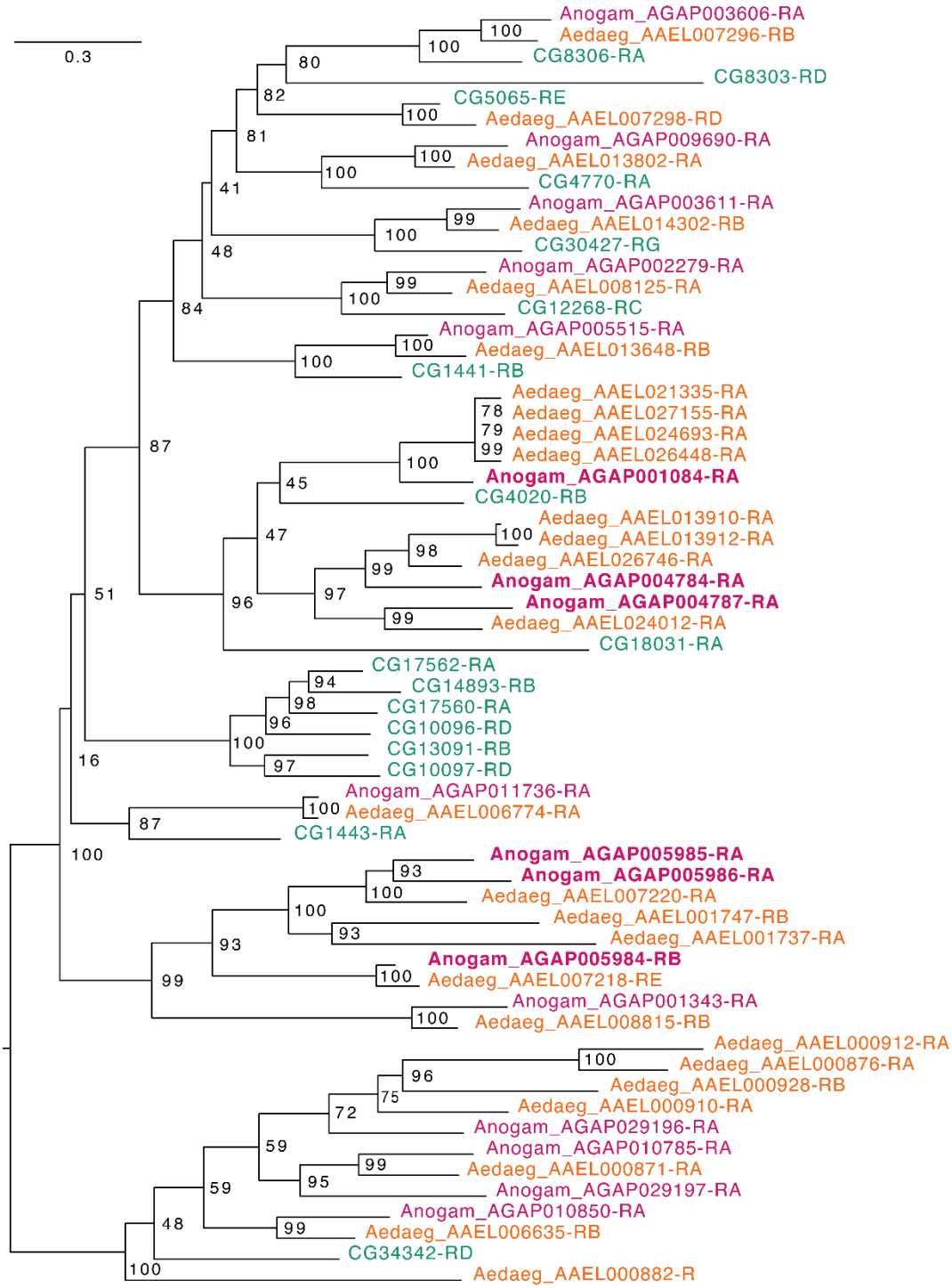


Fig S6: Phylogenetic tree for *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* Elongases. Scale bar shows the number of aminoacid substitutions per alignment position. Node supports are % of bootstrap supports on 1,000 replications.

871



872

873 **Fig S7**: Phylogenetic tree for *Anopheles gambiae*, *Aedes aegypti* and *Drosophila melanogaster* Fatty
874 acyl-CoA reductases. Scale bar shows the number of aminoacid substitutions per alignment
875 position. Node supports are % of bootstrap supports on 1,000 replications.

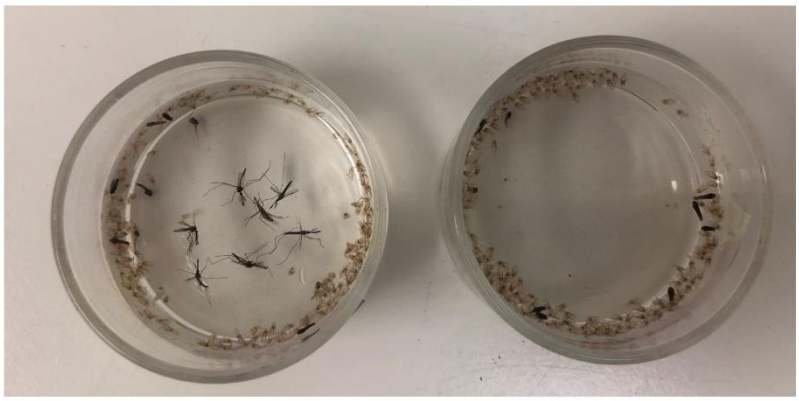
876

877

A.

FAS1899i

Control siblings



B.

Desat3050i

Control siblings

879 **Fig S8**: Lethality phenotype of progeny from crosses A) UAS-FAS1899i x Ubi-A10 Gal4 and B) UAS-
880 Desat3050i x Ubi-A10 Gal4. In each panel the upper photograph corresponds to individuals with
881 FAS1899 or Desat3050 knockdown. The lower photograph corresponds to control siblings, showing
882 the pupae cases left after successful adult emergence.

883

884

885 Table S1: Samples used for Illumina RNAseq. Number of raw reads produced for each sample and
886 number of reads after quality control.

887

Sample	Sex	Tissue	# raw reads	# reads postQC	% reads postQC
Sample_1-1F	Female	Oenocytes	33819432	32551857	96.25%
Sample_2-2F	Female	Oenocytes	33113392	31896623	96.33%
Sample_3-3F	Female	Oenocytes	39251164	38036798	96.91%
Sample_4-4M	Male	Oenocytes	32539976	31437568	96.61%
Sample_5-5M	Male	Oenocytes	25424014	24436348	96.12%
Sample_6-6M	Male	Oenocytes	40072886	39086446	97.54%
Sample_7-7BCF	Female	Carcass	33445880	32562341	97.36%
Sample_8-8BCF	Female	Carcass	53419358	52054378	97.44%
Sample_9-9BCF	Female	Carcass	38663816	37832559	97.85%
Sample_10-10BCM	Male	Carcass	41441984	40404352	97.50%
Sample_11-11BCM	Male	Carcass	38099604	37047825	97.24%
Sample_12-12BCM	Male	Carcass	42194602	41132109	97.48%
TOTAL	-	-	4,51E+08	4,38E+08	97.12%

888

889

890 Table S2: Genes, members of gene families implicated in CHC biosynthesis and over-expressed in
891 oenocytes, ranked in order of highest to lowest expression in Female Oenocytes. Their differential
892 expression (Log_2Fold change) compared to female carcass cells is also shown. Genes above the
893 double line are within the 200 most highly expressed genes.

894

895

Gene	RNAseq Normalized counts in Female Oenocytes	Log_2Fold Change (all at p-value<0.001)
Cyp4G16	266337	3,29
Fatty acid synthase AGAP001899	105057	3,44
Cyp4G17	93677	3,26
Elongase AGAP007264	34216	3,16
Fatty acid synthase AGAP028049	22062	3,37
Fatty acid Reductase AGAP004787	21437	2,87
Elongase AGAP013094	17688	3,73
Propionyl-CoA synthetase AGAP001473	15219	3,31
Desaturase AGAP003050	15068	3,03
Fatty acid Reductase AGAP005984	13904	2,60
Fatty acid Synthase AGAP008468	13811	3,33
Fatty acid Reductase AGAP005986	13041	3,02
Elongase AGAP003197	11810	3,52
Elongase AGAP003195	9596	3,49
Elongase AGAP003196	7688	3,36
Fatty acid Reductase AGAP004784	7504	3,41
Fatty acid Reductase AGAP005985	7160	1,85
Elongase AGAP001097	6026	3,36
Elongase AGAP005512	5523	3,05
Elongase AGAP013219	5099	3,84
Elongase AGAP004372	4785	2,92

896

897

Table S3: Genes, members of gene families implicated in CHC biosynthesis and over-expressed in oenocytes, ranked in order of highest to lowest expression in Male Oenocytes. Their differential expression (Log_2Fold change) compared to male carcass cells is also shown. Genes above the double line are within the 200 most highly expressed genes.

Gene	RNAseq Normalized counts in Male Oenocytes	Log_2Fold Change (all at p-value<0.001)
Cyp4G16	364580	2,83
Cyp4G17	137776	3,23
Fatty acid synthase AGAP001899	97839	3,08
Elongase AGAP007264	49263	3,08
Fatty acid synthase AGAP028049	25170	2,95
Propionyl-CoA synthetase AGAP001473	23730	3,22
Fatty acid Reductase AGAP004787	22177	3,02
Fatty acid Synthase AGAP008468	21699	3,08
Elongase AGAP013094	21692	3,00
Desaturase AGAP003050	17972	2,99
Fatty acid Reductase AGAP005986	14351	3,35
Fatty acid Reductase AGAP005984	14128	2,48
Elongase AGAP003197	10167	2,89
Elongase AGAP003195	7320	2,74
Elongase AGAP001097	6301	2,78
Fatty acid Reductase AGAP004784	5009	2,73
Fatty acid Reductase AGAP005985	3459	3,07
Elongase AGAP004372	2932	1,91
Elongase AGAP005512	2762	2,42
Elongase AGAP003196	2248	2,34
Elongase AGAP013219	753	2,45

Supplementary File 1: Genes differentially expressed in female oenocytes vs female total carcass cells (sheet 1); in male oenocytes vs male total carcass cells (sheet 2) and genes

908 commonly over-expressed in female and male oenocytes compared to female and male
909 total carcass cells (sheet 3).

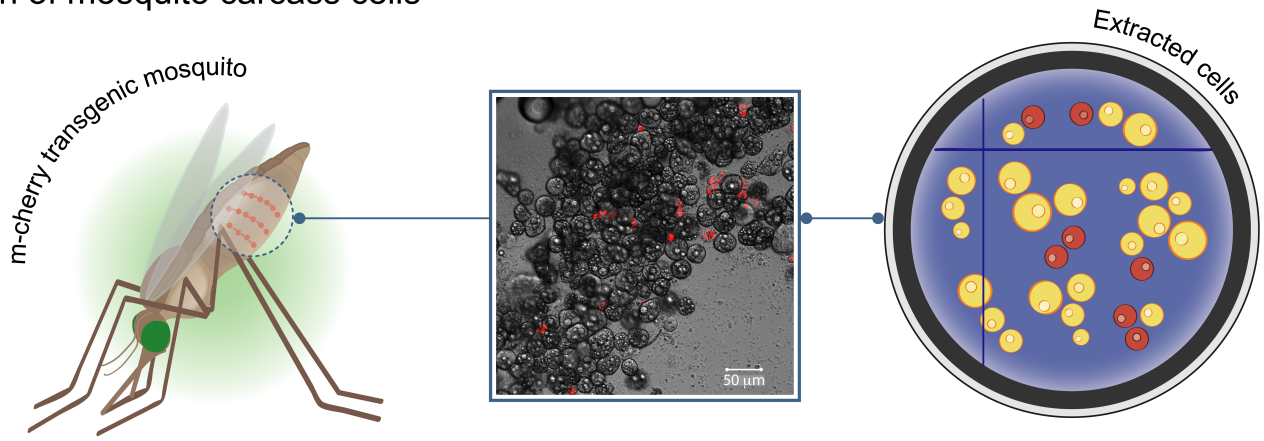
910 **Supplementary File 2**: Genes showing isoform specific differential expression in oenocytes.
911 Comparisons performed are: Female Oenocytes vs Female total Carcass cells (sheet 1), Male
912 Oenocytes vs Male total Carcass cells (sheet 2) and Female Oenocytes vs Male Oenocytes
913 (sheet 3).

914 **Supplementary File 3**: Genes differentially expressed in female oenocytes vs male
915 oenocytes

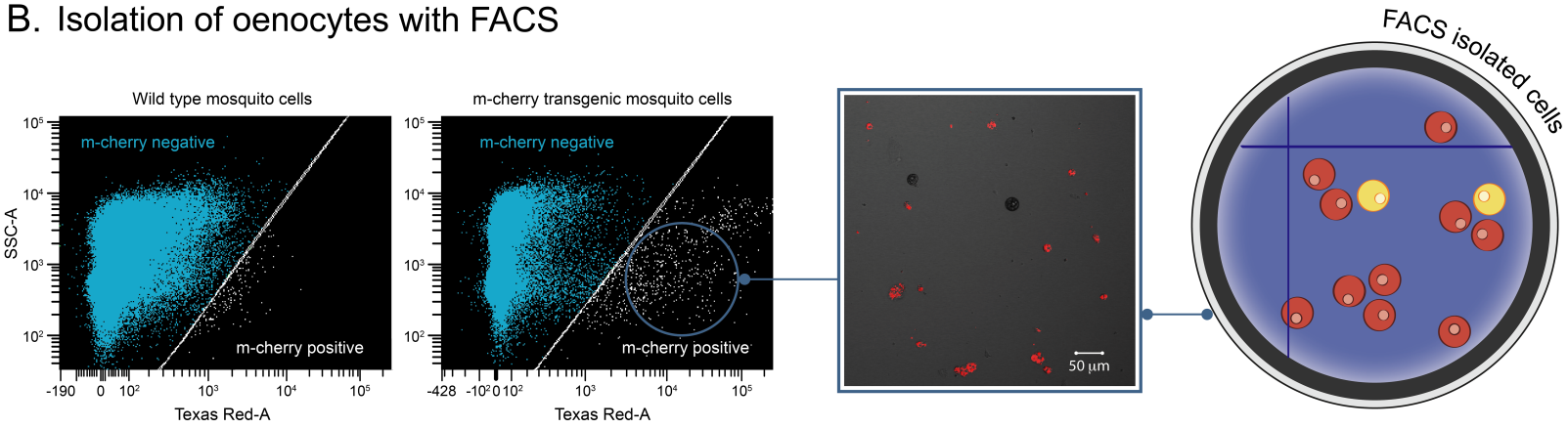
916 **Supplementary File 4**: GC-MS analysis of CHCs. The CHC peaks identified in each sample are
917 shown as well as their amount (ngr) (in 2µl of hexane extract) normalized to the internal
918 standard. A summary of the total ngr of CHCs/mgr is provided for all samples. In the last
919 sheet the FAS1899i females and control females are used to show the relative abundance
920 (in % to the total) of each CHC peak.

921

A. Extraction of mosquito carcass cells

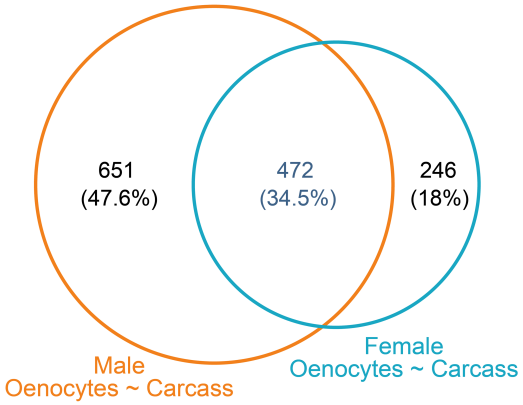


B. Isolation of oenocytes with FACS

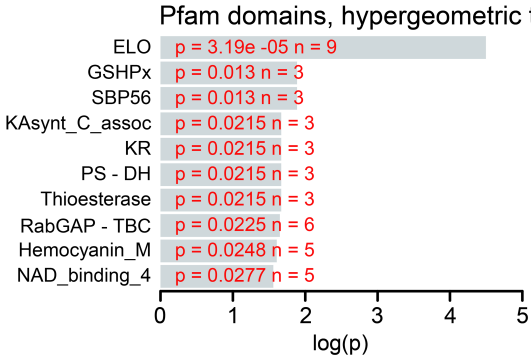


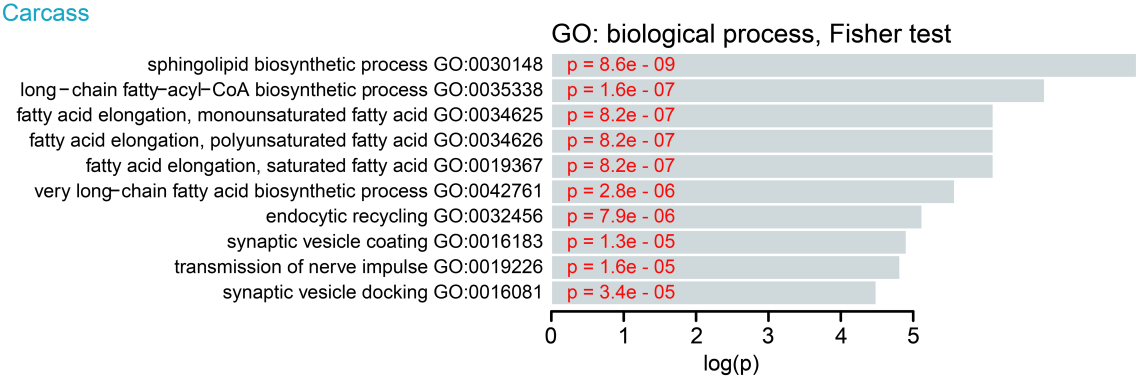
C. Transcriptional analysis of oenocytes

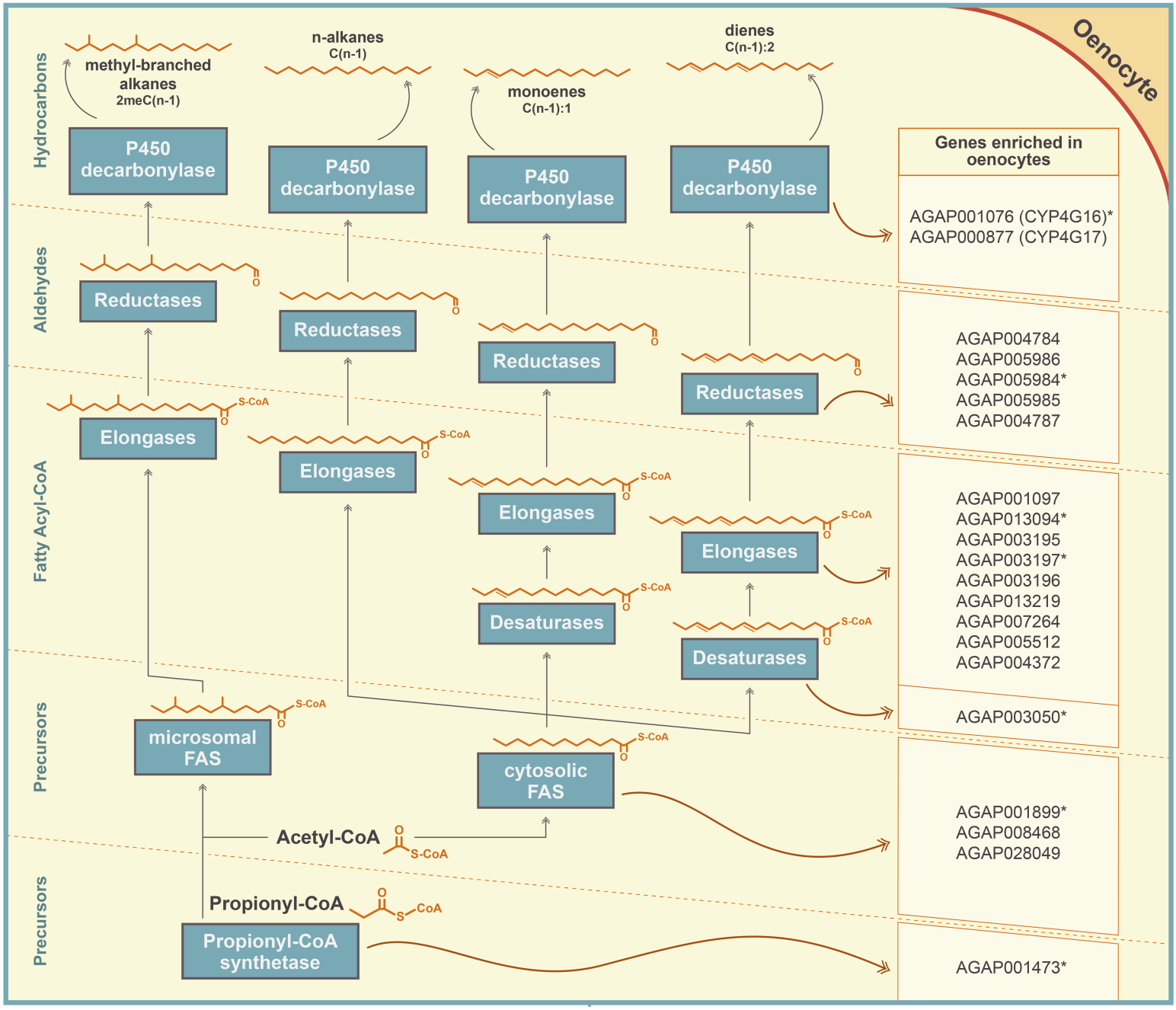
Genes overexpressed in oenocytes

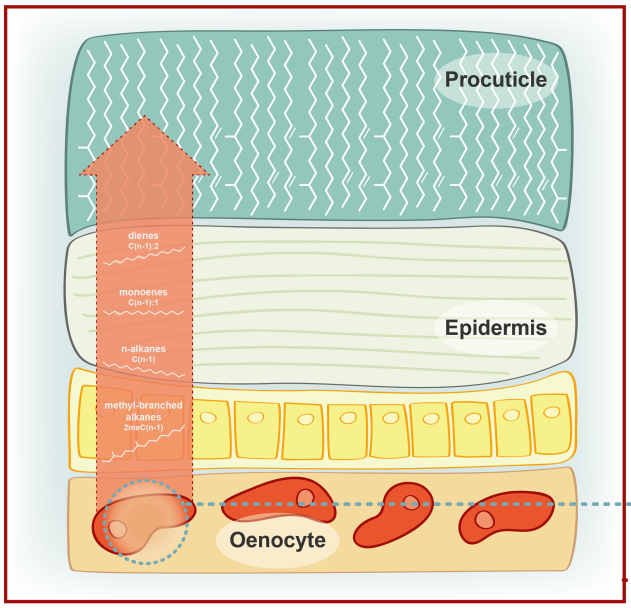


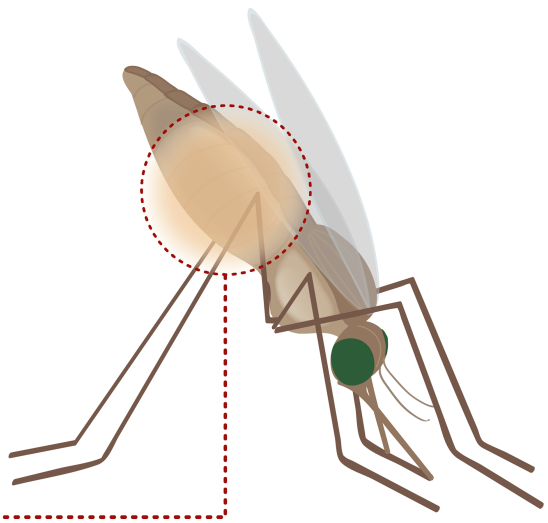
Functions enriched in overexpressed genes (n=472)

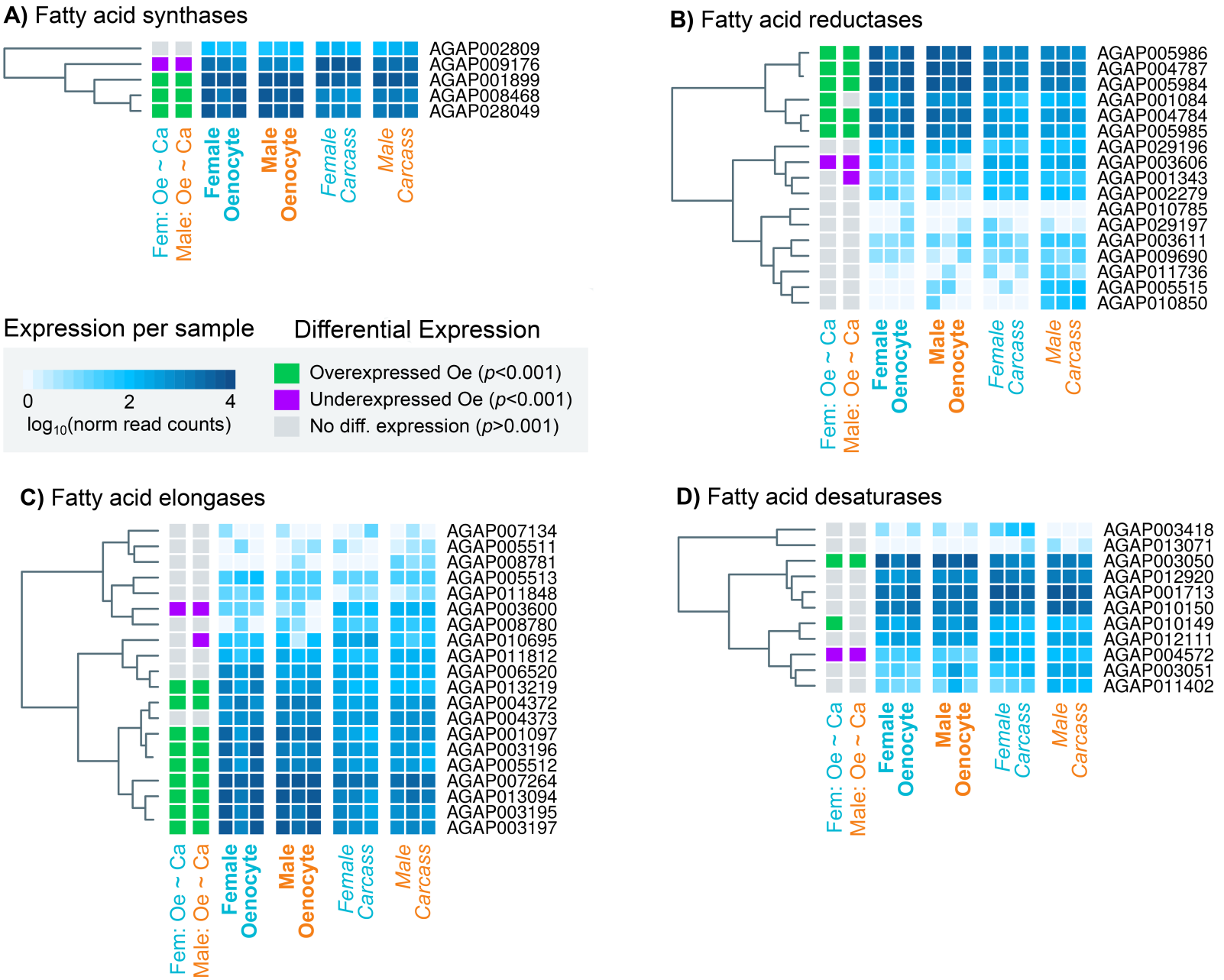


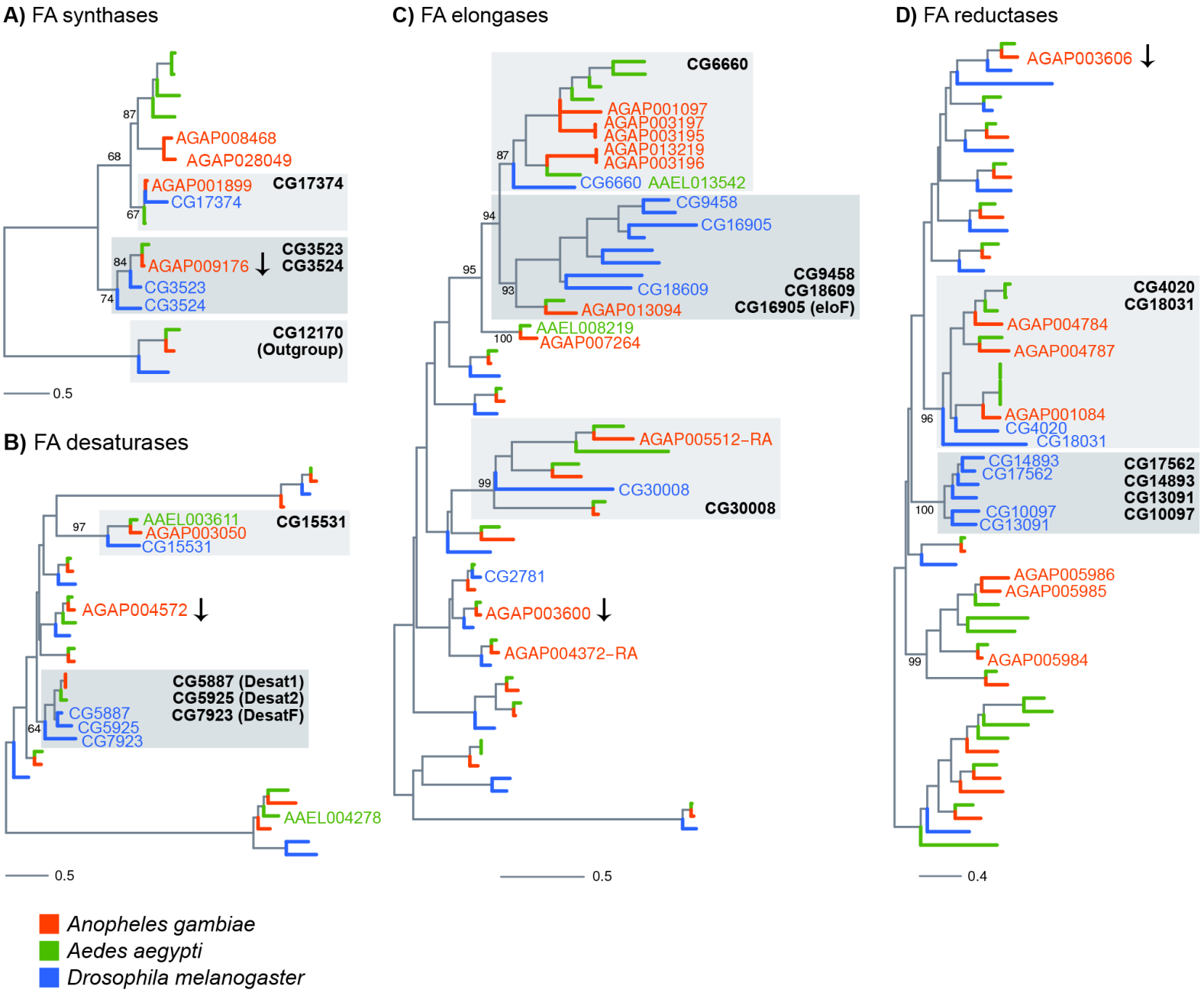


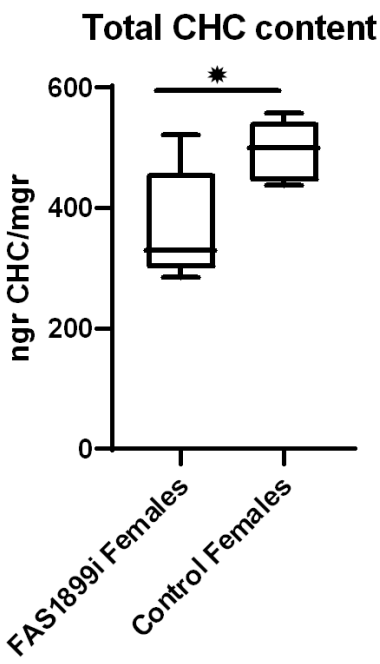

Oenocyte
Hydrocarbons
Aldehydes
Fatty Acyl-CoA
Precursors
Precursors
methyl-branched alkanes 2meC(n-1)
n-alkanes C(n-1)
monoenes C(n-1):1
dienes C(n-1):2
P450 decarbonylase
Reductases
Elongases
Desaturases
microsomal FAS
cytosolic FAS
Acetyl-CoA
Propionyl-CoA
Propionyl-CoA synthetase
S-CoA
Genes enriched in oenocytes
AGAP001076 (CYP4G16)*
AGAP000877 (CYP4G17)
AGAP004784
AGAP005986
AGAP005984*
AGAP005985
AGAP004787
AGAP001097
AGAP013094*
AGAP003195
AGAP003197*
AGAP003196
AGAP013219
AGAP007264
AGAP005512
AGAP004372
AGAP003050*
AGAP001899*
AGAP008468
AGAP028049
AGAP001473*

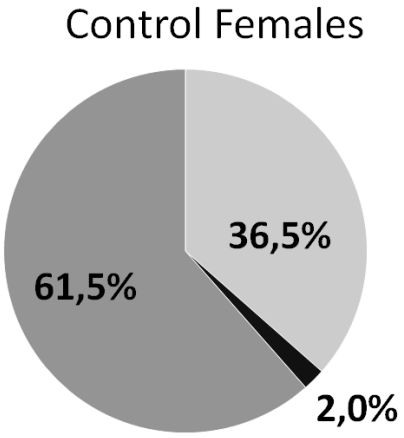

Procuticle
Epidermis
Oenocyte
dienes C(n-1):2
monoenes C(n-1):1
n-alkanes C(n-1)
methyl-branched alkanes 2meC(n-1)

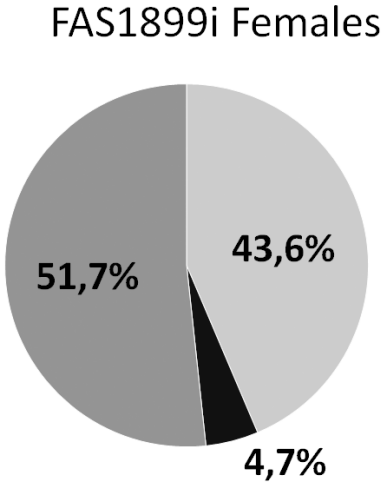


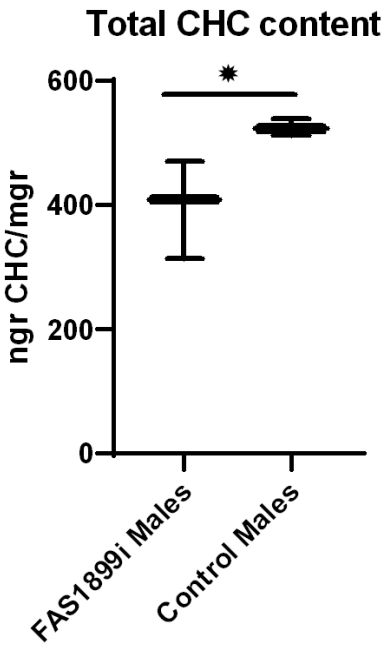

A) Fatty acid synthases
AGAP002809
AGAP009176
AGAP001899
AGAP008468
AGAP028049
Fem: Oe ~ Ca
Male: Oe ~ Ca
Female Oenocyte
Male Oenocyte
Female Carcass
Male Carcass
B) Fatty acid reductases
AGAP005986
AGAP004787
AGAP005984
AGAP001084
AGAP004784
AGAP005985
AGAP029196
AGAP003606
AGAP001343
AGAP002279
AGAP010785
AGAP029197
AGAP003611
AGAP009690
AGAP011736
AGAP005515
AGAP010850
Expression per sample
0
2
4
log10(norm read counts)
Differential Expression
Overexpressed Oe (p<0.001)
Underexpressed Oe (p<0.001)
No diff. expression (p>0.001)
C) Fatty acid elongases
AGAP007134
AGAP005511
AGAP008781
AGAP005513
AGAP011848
AGAP003600
AGAP008780
AGAP010695
AGAP011812
AGAP006520
AGAP013219
AGAP004372
AGAP004373
AGAP001097
AGAP003196
AGAP005512
AGAP007264
AGAP013094
AGAP003195
AGAP003197
D) Fatty acid desaturases
AGAP003418
AGAP013071
AGAP003050
AGAP012920
AGAP001713
AGAP010150
AGAP010149
AGAP012111
AGAP004572
AGAP003051
AGAP011402

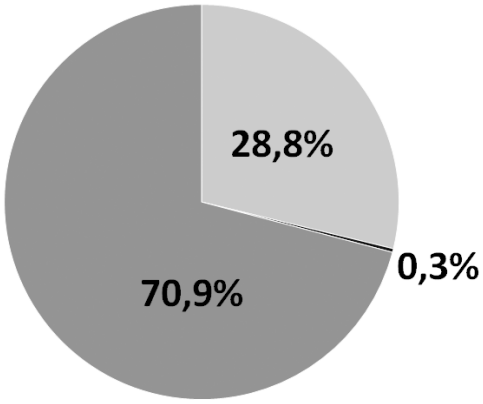

A) FA synthases
87
AGAP008468
AGAP028049
68
AGAP001899
CG17374
CG17374
67
84
AGAP009176
CG3523
CG3524
CG3523
CG3524
74
CG12170
(Outgroup)
0.5
B) FA desaturases
97
AAEL003611
AGAP003050
CG15531
CG15531
AGAP004572
CG5887 (Desat1)
CG5925 (Desat2)
CG7923 (DesatF)
64
CG5887
CG5925
CG7923
AAEL004278
0.5
C) FA elongases
CG6660
AGAP001097
AGAP003197
AGAP003195
AGAP013219
AGAP003196
87
CG6660
AAEL013542
94
CG9458
CG16905
95
93
CG18609
AGAP013094
CG9458
CG18609
CG16905 (eloF)
AAEL008219
100
AGAP007264
AGAP005512-RA
99
CG30008
CG30008
CG2781
AGAP003600
AGAP004372-RA
0.5
D) FA reductases
AGAP003606
CG4020
CG18031
AGAP004784
AGAP004787
96
AGAP001084
CG4020
CG18031
CG14893
CG17562
CG17562
CG14893
CG13091
CG10097
100
CG10097
CG13091
AGAP005986
AGAP005985
99
AGAP005984
0.4
Anopheles gambiae
Aedes aegypti
Drosophila melanogaster

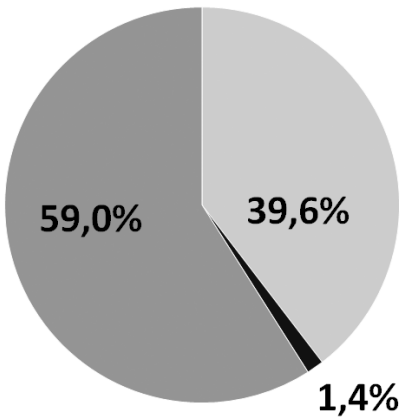

Total CHC content
*
600
400
200
0
ngr CHC/mgr
FAS1899i Females
Control Females

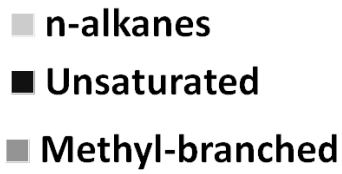

Control Females
36,5%
61,5%
2,0%


FAS1899i Females
43,6%
51,7%
4,7%


Total CHC content
*
600
400
200
0
ngr CHC/mgr
FAS1899i Males
Control Males


Control Males
28,8%
70,9%
0,3%


FAS1899i Males
39,6%
59,0%
1,4%


n-alkanes
Unsaturated
Methyl-branched