Expansion of the fatty acyl reductase gene family shaped pheromone communication in Hymenoptera

Michal Tupec[1,$], Aleš Buček[1,2,$,*], Heiko Vogel[3], Václav Janoušek[4], Darina Prchalová[1], Jiří Kindl[1], Tereza Pavlíčková[1], Petra Wenzelová[1], Ullrich Jahn[1], Irena Valterová[1] and Iva Pichová[1,*]

[1]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic

[2]Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa, Japan

[3]Department of Entomology, Max Planck Institute for Chemical Ecology, Jena, Germany

[4]Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic

[$]These authors contributed equally.

*Corresponding authors. E-mail: IP: iva.pichova@uochb.cas.cz; AB: ales.bucek@oist.jp

Abstract

The conserved fatty acyl reductase (FAR) family is involved in biosynthesis of fatty alcohols that serve a range of biological roles. In moths, butterflies (Lepidoptera), and bees (Hymenoptera), FARs biosynthesize fatty alcohol pheromones participating in mate-finding strategies. Using a combination of next-generation sequencing, analysis of transposable elements (TE) in the genomic environment of FAR genes, and functional characterization of FARs from *Bombus lucorum*, *B. lapidarius*, and *B. terrestris*, we uncovered a massive expansion of the FAR gene family in Hymenoptera, presumably facilitated by TEs. Expansion occurred in the common ancestor of bumblebees (Bombini) and stingless bees (Meliponini) after their divergence from the honeybee lineage. We found that FARs from the expanded FAR-A orthology group contributed to the species-specific male marking pheromone composition. Our results indicate that TE-mediated expansion and functional diversification of the FAR gene family played a key role in the evolution of pheromone communication in the crown group of Hymenoptera.

Abbreviations: MMP: male marking pheromone, FA: fatty acid, FAME: fatty acid methyl ester, FAR: fatty acyl reductase, LG: labial gland, FB: fat body, TE: transposable element.

Introduction

Accumulation of DNA sequencing data is greatly outpacing our ability to experimentally assess the function of the sequenced genes, and most of these genes are expected to never be functionally characterized[1]. Important insights into the evolutionary processes shaping the genomes of individual species or lineages can be gathered from predictions of gene families, gene orthology groups and gene function. However, direct experimental evidence of the function of gene family members is often unavailable or limited[2–6]. Gene duplication is hypothesized to be among the major genetic mechanisms of evolution[7,8]. Although the most probable evolutionary fate of duplicated genes is the loss of one copy, the temporary redundancy accelerates gene sequence divergence and can result in gene subfunctionalization or neofunctionalization—acquisition of slightly different or completely novel functions in one copy of the gene[9,10].

The alcohol-forming fatty acyl-CoA reductases (FARs, EC 1.2.1.84) belong to a multigene family that underwent a series of gene duplications and subsequent gene losses, pseudogenizations and possibly functional diversification of some of the maintained copies, following the birth-and-death model of gene family evolution[11]. FARs exhibit notable trends in gene numbers across organism lineages; there are very few FAR genes in fungi, vertebrates and non-insect invertebrates such as *Caenorhabditis elegans*, whereas plant and insect genomes typically harbour extensive FAR gene family members[11]. FARs are critical for production of primary fatty alcohols, which are naturally abundant fatty acid (FA) derivatives with a wide variety of biological roles. Fatty alcohols are precursors of waxes and other lipids that serve as surface-protective or hydrophobic coatings in plants, insects and other animals[12–14]; precursors of energy-storing waxes[15–17]; and components of ether lipids abundant in the cell membranes of cardiac, nervous and immunological tissues[18].

Additionally, in some insect lineages, FARs were recruited for yet another task—biosynthesis of fatty alcohols that serve as pheromones or pheromone precursors. Moths

(Lepidoptera) are the most well-studied model of insect pheromone biosynthesis and have been the subject of substantial research effort related to FARs. Variation in FAR enzymatic specificities is a source of sex pheromone signal diversity among moths in the genus *Ostrinia*[19] and is also responsible for the distinct pheromone composition in two reproductively isolated races of the European corn borer *Ostrinia nubilalis*[20]. Divergence in pheromone biosynthesis can potentially install or strengthen reproductive barriers, ultimately leading to speciation[21]. However, the biological significance of a large number of insect FAR paralogs remains unclear, as all FARs implicated in moth and butterfly sex pheromone biosynthesis are restricted to a single clade, indicating that one FAR group was exclusively recruited for pheromone biosynthesis[20,22–24]. While more than 20 FARs have been experimentally characterized from 23 moth and butterfly (Lepidoptera) species[25], FARs from other insect orders have received far less attention. Single FAR genes have been isolated and experimentally characterized from *Drosophila* (Diptera)[14], the European honeybee (Hymenoptera) [26] and the scale insect *Ericeus pela* (Hemiptera)[27]. Our limited knowledge about FAR function prevents us from drawing inferences about the biological significance of the FAR gene family expansion in insects.

Bumblebees (Hymenoptera: Apidae) are a convenient experimental model to study insect FAR evolution because the majority of bumblebee species produces fatty alcohols as species-specific components of male marking pheromones (MMPs)[28], which are presumed to be biosynthesized by some of the numerous bumblebee FAR gene family members[29]. Bumblebee males employ MMPs to attract conspecific virgin queens[30]. In addition to fatty alcohols, MMPs generally contain other FA derivatives and terpenoid compounds. The MMP composition serves as a phylogenetic signal and can be used as a taxonomic tool to discriminate bumblebee species and subspecies[31–33]. Pheromones from three common European bumblebee species, *B. terrestris*, *B. lucorum* and *B. lapidarius*, represent the diversity of fatty alcohol MMP components. Fatty alcohols are the major compounds in MMPs of *B. lapidarius* (hexadecanol and *Z*9-hexadecenol)

and accompanying electroantennogram-active compounds in *B. terrestris* (hexadecanol, octadecatrienol, octadecenol) and *B. lucorum* (hexadecanol, *Z*9,*Z*12-octadecadienol, *Z*9,*Z*12,*Z*15-octadecatrienol, octadecanol)[34–40].

In our previous investigation of the molecular basis of pheromone diversity in bumblebees, we found that the substrate specificities of fatty acyl desaturases (FADs), enzymes presumably acting upstream of FARs in pheromone biosynthesis[41], are conserved across species despite differences in the compositions of their unsaturated FA-derived pheromone components[42]. These findings suggest that the substrate specificity of FADs expressed in the male bumblebee MMP-producing labial gland (LG) contributes only partially to the species-specific composition of FA-derived MMPs[42]. The fatty alcohol content in bumblebee MMPs is therefore presumably co-determined by the enzymatic specificity of other pheromone biosynthetic steps, such as FA biosynthesis/transport or FA reduction. Analysis of the *B. terrestris* male LG transcriptome uncovered a remarkably high number of putative FAR paralogs, including apparently expressed pseudogenes, strongly indicating dynamic evolution of the FAR gene family[29].

Here, we aimed to determine how the members of the large FAR gene family in the bumblebee lineage contribute to MMP biosynthesis. We sequenced and analysed *B. lucorum* and *B. lapidarius* male LG and FB transcriptomes and functionally characterized the FAR enzymes, along with FAR candidates from *B. terrestris*, in a yeast expression system. We combined experimental information about FAR enzymatic specificities with quantitative information about bumblebee FAR expression patterns, as well as comprehensive GC analysis of MMPs and their FA precursors in the bumblebee male LG, with inference of the hymenopteran FAR gene tree. In addition, we investigated the content of transposable elements (TE) in the genomic environment of FAR genes in *B. terrestris*. We concluded that a dramatic TE-mediated expansion of the FAR gene family started in the common ancestor of the bumblebee (Bombini: *Bombus*) and stingless

bee (Meliponini) lineages, which presumably shaped the pheromone communication in these lineages.

Results

Identification of FARs in bumblebee transcriptomes

We sequenced, assembled and annotated male LG and male fat body (FB) transcriptomes of two bumblebee species, *B. lucorum* and *B. lapidarius*. LG is the MMP-producing organ and is markedly enlarged in males, while FB was used as a reference tissue not directly involved in MMP biosynthesis[43]. Searches of the LG and FB transcriptomes of *B. lucorum* and *B. lapidarius* and the previously sequenced FB and LG transcriptomes of *B. terrestris*[29] yielded 12, 26, and 16 expressed FAR homologs in *B. lapidarius*, *B. terrestris*, and *B. lucorum*, respectively (Supplementary Fig. 1).

FAR gene family evolution in Hymenoptera

To gain insight into the evolution of FAR gene family in Hymenoptera, we reconstructed a FAR gene tree using predicted FARs from species representing ants, wasps, parasitic wasps and several bee lineages (Fig. 1). We assigned the names FAR-A to FAR-K to 11 FAR orthology groups that were retrieved as branches with high bootstrap support in the FAR gene tree. These orthology groups typically encompass one or more orthologs from each of the hymenopteran species used in the tree inference, with the exception of apparent species-specific FAR duplications or losses (Fig. 1). Notably, we identified a massive expansion of the FAR-A orthology group in the bumblebee and stingless bee (subfamily Meliponini) lineages, the two most closely related lineages in our dataset. The number of FAR homologs is inflated by a large number of FAR transcripts with incomplete protein coding sequences lacking catalytically critical regions such as the putative active site, $NAD(P)^+$ binding site or substrate binding site (Fig. 1,

Supplementary Table 1). We also inferred a FAR gene tree encompassing FARs from three representatives of non-hymenopteran insect orders—the beetle *Tribolium castaneum*, the moth *Bombyx mori* and the fly *Drosophila melanogaster* (Supplementary Fig. 2). The only functionally characterized FAR from *D. melanogaster*—Waterproof (NP_651652.2), which is involved in biosynthesis of a protective wax layer[14]—was placed in the FAR-J orthology group (Supplementary Fig. 2). The FAR-G orthology group includes a FAR gene from *Apis mellifera* with unclear biological function[26] and a sex pheromone-biosynthetic FAR from *B. mori*[44] (Supplementary Fig. 2). In the gene tree, the majority of FAR orthology groups contain predicted FARs from both hymenopteran and non-hymenopteran insect species. These orthology groups are presumably ancestral to insects. Only FAR-D and FAR-K do not include any non-hymenopteran FARs from our dataset and thus presumably represent hymenoptera-specific FAR gene family expansions (Supplementary Fig. 2).

Genomic organization and TE content

To uncover the details of genetic organization of FAR-A genes, we attempted to analyse the shared synteny of FAR genes in the genomes of *B. terrestris* and *A. mellifera*[45]. We aligned the *A. melifera* and *B. terrestris* genomes, but we were not able to identify any positional *A. mellifera* homologs of *B. terrestris* FAR-A genes (data not shown). While the majority of FAR genes belonging to the non-FAR-A gene orthology group localize to the *B. terrestris* genome assembled to linkage groups, most of the *B. terrestris* FAR-A genes localize to unlinked short scaffolds (Supplementary Table 2). Some of the FAR-A genes in the *B. terrestris* genome are arranged in clusters (Supplementary Fig. 3).

A genome assembly consisting of short scaffolds is often indicative of a repetitive structure in the assembled genomic region. Our analysis of the distribution of TEs in the vicinity of FAR genes in the *B. terrestris* genome confirmed that TEs are significantly enriched around FAR-A genes compared to the genome-wide average around randomly selected genes

($p < 0.0001$). FAR-A genes have on average more than 50% of their 10-kb surrounding region formed by TEs compared to an average 10% around randomly selected *B. terrestris* genes. In contrast, the densities of TEs in the vicinity of FAR genes not belonging to the FAR-A group do not differ from the genome-wide average ($p = 0.1041$; Fig. 2). Although all major known TE families are statistically enriched in the neighbourhood of the FAR-A genes (Fig. 2), the Class I comprising retroid elements contributes considerably to the elevated repeat content around FAR-A genes.

Tissue-specific expression

We selected 10 promising MMP-biosynthetic FAR candidates that were 1) among the 100 most abundant transcripts in the LG and were substantially more abundant in LG than in FB based on RNA-Seq-derived normalized expression values (Supplementary Fig. 1 and ref. [29]) and 2) included in the protein coding sequence all the predicted catalytically critical regions of FARs—the putative active site, $NAD(P)^+$ binding site and substrate binding site (Supplementary Table 1).

By employing RT-qPCR in an expanded set of bumblebee tissues, we confirmed that the FAR candidates follow a general trend of overexpression in male LG compared to FB, flight muscle and gut (all from male bumblebees) and virgin queen LG (Fig. 3, Supplementary Fig. 4, $p < 0.05$, one-way ANOVA followed by *post-hoc* Tukey's HSD test). Notably, *B. lapidarius* FAR-A1 (*Blap*FAR-A1) and *B. terrestris* FAR-J (*Bter*FAR-J) transcripts are also abundant in virgin queen LG, where they are expressed at levels comparable to those in male LG (Fig. 3, Supplementary Fig. 4).

Cloning and functional characterization

The full-length coding regions of the FAR candidates were isolated from male LG cDNA libraries using gene-specific PCR primers (Supplementary Table 3). In general, the FAR

187 candidates share high to very high protein sequence similarity within each orthology group. FARs
188 from three bumblebee species belonging to the FAR-J orthology group are nearly identical,
189 sharing 97.2–99.7% protein sequence identity; *Bluc*FAR-A1 and *Bter*FAR-A1 share 99.4%
190 protein sequence identity with each other and 60.9–61.1% with *Blap*FAR-A1. *Bluc*FAR-A2 and
191 *Bter*FAR-A2 share 94.8% protein sequence identity (Supplementary Table 4). *Bluc*FAR-J was not
192 cloned because of its very high similarity to *Bter*FAR-J (99.7% sequence identity, two amino acid
193 differences). We cloned two versions of *Blap*FAR-A1: one that was custom-synthesized based on
194 the predicted full-length coding sequence assembled from RNA-Seq data and one called
195 *Blap*FAR-A1-short that we consistently PCR-amplified from *B. lapidarius* male LG cDNA.
196 *Blap*FAR-A1-short has an in-frame internal 66 bp deletion in the coding region that does not
197 disrupt the predicted active site, putative $NAD(P)^+$ binding site or putative substrate binding site.
198 Using RT-qPCR with specific primers for each variant, we confirmed that both *Blap*FAR-A1 and
199 *Blap*FAR-A1-short are expressed in the *B. lapidarius* male LG and virgin queen LG
200 (Supplementary Fig. 5).

201 To test whether the MMP-biosynthetic FAR candidates code for enzymes with fatty acyl
202 reductase activity and to uncover their substrate specificities, we cloned the candidate FAR
203 coding regions into yeast expression plasmids, heterologously expressed the FARs in
204 *Saccharomyces cerevisiae* and assayed the fatty alcohol production by GC (Supplementary Fig. 8
205 and Supplementary Fig. 9). His-tagged FARs were detected in all yeast strains transformed with
206 plasmids bearing FARs (Fig. 4, Supplementary Fig. 11a), while no His-tagged proteins were
207 detected in the negative control (yeast strain transformed with an empty plasmid). In addition to
208 the major protein bands corresponding to the theoretical FAR molecular weight, we typically
209 observed protein bands with lower and/or higher molecular weight (Fig. 4a). The synthetic
210 *Bluc*FAR-A1-opt and *Bluc*FAR-A2-opt coding regions with codon usage optimized for *S.*
211 *cerevisiae* showed a single major Western blot signal corresponding to the position of the

212 predicted full-length protein (Fig. 4a). The shortened heterologously expressed proteins thus
213 presumably represent incompletely transcribed versions of full-length FARs resulting from
214 ribosome stalling[46], while the higher molecular weight bands might correspond to aggregates of
215 full-length and incompletely translated FARs. Because the codon-optimized *Bluc*FAR-A1-opt and
216 *Bluc*FAR-A2-opt exhibit the same overall specificity in yeast expression system as the respective
217 non-codon-optimized FARs (Supplementary Fig. 6), we employed non-codon-optimized FARs
218 for further functional characterization. We only used the codon-optimized versions of *Bluc*FAR-
219 A1 and *Bluc*FAR-A2 in experiments with exogenously supplemented substrates to increase the
220 possibility of product detection, as the optimized FARs produce overall higher quantities of fatty
221 alcohols ($p < 0.05$, two-tailed *t*-test).

222 Characterization of FAR enzymatic activities involved identification of numerous
223 individual FA derivatives, denoted using the position/configuration of the double bond if present
224 (e.g., *Z*9), the length of the carbon chain (e.g., 20), the number of double bonds (e.g., ":" or ":1"
225 for saturated and monounsaturated FAs, respectively) and the C1 moiety (COOH for acid, OH for
226 alcohol, Me for methyl ester, CoA for CoA-thioester).

227 Functional characterization of FARs from *B. terrestris* and *B. lucorum* in yeast indicated
228 that saturated C16 to C26 fatty alcohols are produced by both *Bter*/*Bluc*FAR-A1 and *Bter*FAR-J
229 enzymes (Fig. 5ad, Supplementary Fig. 9); *Bter*/*Bluc*FAR1 prefers C22 substrates, whereas
230 *Bter*FAR-J has an optimal substrate preference slightly shifted to C24. Unlike any of the other
231 characterized FARs, *Bter*/*Bluc*FAR-A1 are also capable of reducing supplemented
232 monounsaturated *Z*15-20:1 acyl to the corresponding alcohol (Fig. 5a and Supplementary Fig.
233 10c). Both *Bter*FAR-A2 and *Bluc*FAR-A2 reduce only 16: and 18: acyls (Fig. 5c, Supplementary
234 Fig. 8).

235 Characterization of *B. lapidarius* FARs showed that *Blap*FAR-A1, in contrast to
236 *Bter*/*Bluc*FAR-A1, produces *Z*9-16:1OH and *Z*9-18:1OH (Fig. 5A and Supplementary Fig. 8).

237 *Blap*FAR-A4 produces 16:OH and *Z*9-16:1OH, together with lower quantities of 14:OH and
238 *Z*9-18:1OH (Fig. 5b, Supplementary Fig. 7, Supplementary Fig. 8). *Blap*FAR-A5 produces
239 16:OH as a major product and lower amounts of 14:OH, *Z*9-16:1OH, 18:OH and Z9-18:1OH
240 (Fig. 5b, Supplementary Fig. 7, Supplementary Fig. 8). In addition, both *Blap*FAR-A1 and
241 *Blap*FAR-A4 are capable of reducing supplemented polyunsaturated fatty acyls (*Z*9,*Z*12-18:2 and
242 *Z*9,*Z*12,*Z*15-18:3) to their respective alcohols (Fig. 5ab and Supplementary Fig. 10ab). Similarly
243 to *Bter*FAR-J, *Blap*FAR-J also reduces saturated C16 to C26 acyls (Fig. 5d and Supplementary
244 Fig. 9). No fatty alcohols were detected in the negative control (Supplementary Fig. 8). We did
245 not detect the formation of fatty aldehydes in any of the yeast cultures (data not shown),
246 confirming that the studied FARs are strictly alcohol-forming fatty acyl-CoA reductases. In
247 contrast to *Blap*FAR-A1, *Blap*FAR-A1-short does not produce detectable amounts of any fatty
248 alcohol (Supplementary Fig. 11b), suggesting that the missing 22-amino acid region is crucial for
249 the retention of FAR activity (Supplementary Fig. 8).
250
251 **Quantification of fatty alcohols and fatty acyls in bumblebee male LG and FB**
252 In addition to functional characterization of FARs in a heterologous host, we performed a detailed
253 analysis of transesterifiable fatty acyls (free FAs and fatty acyls bound in esters) and fatty
254 alcohols in LGs and FBs of 3-day-old *B. lapidarius*, *B. terrestris* and *B. lucorum* males (Fig. 6,
255 Supplementary Table 5 and Supplementary Table 6).
256 A limited number of fatty alcohols (mainly 16:OH, *Z*9,*Z*12-18:2OH and
257 *Z*9,*Z*12,*Z*15-18:3OH) were detected in FBs of *B. lucorum* and *B. terrestris* (Supplementary Table
258 6), but at substantially lower abundance than in LGs. In the LGs, 4, 14, and 19 individual fatty
259 alcohol compounds were detected in *B. lapidarius*, *B. lucorum* and *B. terrestris*, respectively
260 (Supplementary Table 5). To assess the apparent *in-vivo* specificity of FARs in LGs and FBs, we
261 calculated the ratios of amounts of each fatty alcohol to the amount of its hypothetical fatty acyl

precursors (Supplementary Fig. 12). The fatty alcohol ratios are greater than 50% for most of the fatty alcohols in LGs and even approach 100% for some of the monounsaturated C20+ fatty alcohols (Supplementary Fig. 12ab), as the corresponding fatty acyls could not be quantified due to low abundance, suggesting that the FARs acting on these fatty chains convert almost all of the acyl substrate to alcohol. The specificities of the characterized FARs determined in yeast correlate well with the composition of LG fatty alcohols and fatty acyls (Fig. 6), except for *Z*9,*Z*12-18:2OH and *Z*9,*Z*12,*Z*15-18:3OH, as none of the studied FARs from *B. lucorum* or *B. terrestris* reduce the corresponding acyls.

Discussion

Since the first genome-scale surveys of gene families, gene duplications and lineage-specific gene family expansions have been considered major mechanisms of diversification and adaptation in eukaryotes[2] and prokaryotes[47]. Tracing the evolution of gene families and correlating them with the evolution of phenotypic traits has been facilitated by the growing number of next-generation genomes and transcriptomes from organisms spanning the entire tree of life. However, obtaining experimental evidence of the function of numerous gene family members across multiple species or lineages is laborious. Thus, such data are scarce, and researchers have mostly relied on computational inference of gene function[48]. Here, we aimed to combine computational inference with experimental characterization of gene function to understand the evolution of the FAR family, for which we predicted a notable gene number expansion in our initial transcriptome analysis of the buff-tailed bumblebee *B. terrestris*[29]. We specifically sought to determine whether the FARs that emerged through expansion of the FAR gene family substantially contribute to MMP biosynthesis in bumblebees.

We identified a massive expansion of the FAR-A orthologous gene group in stingless bees and bumblebees. The sister taxonomic relationship of these taxa and their position as a

crown group within the bee (Anthophila) clade[49–51] indicates that the FAR duplication process occurred or started in the common ancestor of bumblebees and stingless bees. According to estimated lineage divergence times, FAR duplication events started 76–85 million years ago after their divergence from *Apis* 82–93 million years ago[52]. The number of inferred FAR-A orthologs is inflated by predicted pseudogenes—FARs with fragmented coding sequences that lack some of the catalytically essential domains and motifs (Supplementary Fig. 2). These predicted catalytically inactive yet highly expressed FAR-A pseudogenes might play a role in regulating the FAR-catalysed reduction[53]. The number of predicted FAR-A pseudogenes indicate that the FAR-A orthology group expansion in this lineage was a highly dynamic process (Fig. 1, Supplementary Fig. 2). The high number of species-specific FAR-A duplications or losses between the closely related species *B. lucorum* and *B. terrestris*, which diverged approximately 5 million years ago[54], further indicates the dynamic evolutionary processes acting on the FAR gene.

Strikingly, stingless bees also employ LG secretion in scent marking[55]. In worker stingless bees, LG secretion is used as a trail pheromone to recruit nestmates to food resources and generally contains fatty alcohols such as hexanol, octanol, and decanol in the form of their fatty acyl esters[56,57]. The correlation between FAR-A gene orthologous group expansion and use of LG-produced fatty alcohols as marking pheromones suggests a critical role for FAR-A gene group expansion in the evolution of scent marking. In the future, identification and characterization of FAR candidates involved in production of stingless bee worker LG-secretion could corroborate this hypothesis.

Bumblebee orthologs (FAR-G) of *B. mori* pheromone-biosynthetic FARs[44] are not abundantly or specifically expressed in male bumblebee LGs, as evidenced by RNA-Seq RPKM values (Supplementary Fig. 1). MMP-biosynthetic FARs in bumblebees and female sex pheromone-biosynthetic FARs in moths (Lepidoptera) were most likely recruited independently for the tasks of pheromone biosynthesis.

Various models have attempted to describe the evolutionary mechanisms leading to the emergence and maintenance of gene duplicates[10]. The fragmented state of the *B. terrestris* genome and its limited synteny with the *A. mellifera* genome restricts our ability to reconstruct the genetic events accompanying the FAR duplications resulting in the FAR-A orthology group expansion. Taking into consideration the large quantities of MMPs in bumblebee males, we speculate that gene dosage benefits could substantially contribute to the duplications and duplicate fixation of MMP-biosynthetic FARs. Under this model, selection for increased amounts of fatty alcohols could fix the duplicated FARs in a population[2,10]. The MMP quantities in bees are substantially higher than quantities of pheromones in other insects of comparable size. For example, *B. terrestris*, *B. lucorum*, and *B. lapidarius* bumblebee males can produce several milligrams of MMPs (Kindl, personal communication), while the sphingid moth *Manduca sexta* produces tens of nanograms of sex pheromone[58]. Sexual selection favouring bumblebee males capable of producing large quantities of MMPs thus might have served as the evolutionary driver for repeated FAR-A duplication.

Mechanistically, gene duplications can be facilitated by associated TEs[59–61]. The content of repetitive DNA in the *B. terrestris* and *B. impatiens* genome assemblies is 14.8% and 17.9%, respectively [62], which is lower than in other insects such as the beetle *Tribolium castaneum* (30%), *Drosophila* (more than 20%) or the wasp *Nasonia vitripenis* (more than 30%) but substantially higher than in the honeybee *Apis mellifera* (9.5%)[63,64]. Our finding that TEs are enriched in the vicinity of FAR-A genes in the *B. terrestris* genome indicates that TEs presumably contributed to the massive expansion of the FAR-A orthology group (Fig. 2). One possible scenario is that the FAR-A gene in the common ancestor of bumblebees and stingless bees translocated to a TE-rich region, which subsequently facilitated expansions of this orthology group.

We have previously shown that the transcript levels of biosynthetic genes generally reflect the biosynthetic pathways most active in bumblebee LG[29,42]. For further experimental characterization, we therefore selected the FAR-A and FAR-J gene candidates, which exhibited high and preferential expression in male LG. The abundant expression of *Blap*FAR-A1 and *Bter*FAR-J in both virgin queen and male LG suggests that these FARs might also have been recruited for production of queen-specific signals[65].

The spectrum of fatty alcohols in *B. terrestris* and *B. lucorum* male LG differs substantially from that of *B. lapidarius*. In both *B. terrestris* and *B. lucorum*, the male LG extract contains a rich blend of C14–C26 fatty alcohols with zero to three double bonds (Fig. 6, Supplementary Table 5). In *B. lapidarius*, the male LG extract is less diverse and dominated by *Z*9-16:1OH and 16:OH (Fig. 6 and Supplementary Table 5). The functional characterization of LG-expressed FARs uncovered how the distinct repertoire of LG-expressed FAR orthologs, together with differences in FAR substrate specificities, contributes to the biosynthesis of species-specific MMPs. We found that the highly similar *Bluc*/*Bter*FAR-A1 and *Blap*FAR-A1 orthologs exhibit distinct substrate preferences for longer fatty acyl chains (C18-C26) and shorter monounsaturated fatty acyl chains (*Z*9-16:1 and *Z*9-18:1), respectively. This substrate preference correlates with the abundance of *Z*9-16:1OH in *B. lapidarius* MMP and the almost complete absence of *Z*9-16:1OH in *B. lucorum* and *B. terrestris* (Supplementary Table 5). *Blap*FAR-A4 and to some extent *Blap*FAR-A5 likely further contribute to the biosynthesis of *Z*9-16:1OH in *B. lapidarius*. The ability of *Bluc*/*Bter*FAR-A1 (and not of *Blap*FAR-A1) to reduce long monounsaturated fatty acyls (*Z*15-20:1) also correlates with the absence of detectable amounts of *Z*15-20:1OH in *B. lapidarius* MMP.

Our comprehensive GC analysis of bumblebee male LGs, however, indicates that the composition of LG fatty acyls is another factor that contributes substantially to the final MMP composition. For example, the very low quantities of *Z*9-16:1OH in *B. terrestris* and *B. lucorum*

and of *Z*15-20:1OH in *B. lapidarius* can be ascribed to the absence of a FAR with the corresponding substrate specificity, but the very low amount of *Z*9-16:1 acyl in *B. lucorum* and *B. terrestris* male LG and the absence of detectable *Z*15-20:1 acyl in *B. lapidarius* LG also likely contribute (Supplementary Table 5).

We detected several fatty alcohols in FBs of *B. terrestris* and *B. lucorum*, 16:OH, *Z*9,*Z*12-18:2OH and *Z*9,*Z*12,*Z*15-18:3OH being the most abundant (Supplementary Table 6). Fatty alcohols are not expected to be transported from FB across haemolymph to LG[29]. However, the presence of *Z*9,*Z*12-18:2 and *Z*9,*Z*12,*Z*15-18:3 fatty alcohols in FB provides an explanation for why we did not find a FAR reducing *Z*9,*Z*12-18:2 and *Z*9,*Z*12,*Z*15-18:3 among the functionally characterized candidates from *B. terrestris* and *B. lucorum*. Our candidate selection criteria were based on the LG-specific FAR transcript abundance, and we might have disregarded a FAR that is capable of polyunsaturated fatty acyl reduction and is expressed at comparable levels in both LG and FB.

We noted several discrepancies between the FAR specificity in the yeast expression system and the apparent FAR specificity *in vivo* (i.e., the apparent specificity of fatty acyl reduction in bumblebee LG calculated from the fatty acyl and fatty alcohol content). We found that *Blap*FAR-A1 is capable of producing substantial amounts of *Z*9-16:1OH and *Z*9-18:1OH (Supplementary Fig. 7b, Supplementary Fig. 8) in the yeast system, while in *B. lapidarius* LG, only *Z*9-16:1 acyl is converted to *Z*9-16:1OH, as evidenced by the absence of detectable amounts of *Z*9-18:1OH (Fig. 6 and Supplementary Table 5). Additionally, *Blap*FAR-A1 and *Blap*FAR-A4 in the yeast expression system produce polyunsaturated fatty alcohols that are not present in *B. lapidarius* male LG, despite the presence of corresponding fatty acyls in the LG (Supplementary Table 5). A possible explanation for the differences between FAR specificities in the bumblebee LG and the yeast expression system is that the pool of LG fatty acyls that we assessed and used to evaluate the apparent FAR specificities has a different composition than the LG pool of fatty

acyl-CoAs, which are the form of fatty acyls accepted by FARs as substrates. The relatively low concentrations of *Z*9,*Z*12-18:2CoA, *Z*9,*Z*12,*Z*15-18:3CoA, and Z9-18:1CoA in the LG of male *B. lapidarius* compared to the concentrations of the respective fatty acyls could prevent detectable accumulation of the corresponding fatty alcohols. We therefore propose that the selectivity of enzymes and binding proteins that convert fatty acyls to fatty acyl-CoAs[66] and protect fatty acyl-CoAs from hydrolysis[67] represents an additional mechanism shaping the species-specific fatty alcohol composition in bumblebee male LGs.

In sum, the functional characterization of bumblebee FARs indicates that the combined action of FARs from the expanded FAR-A orthology group has the capability to biosynthesize the majority of bumblebee MMP fatty alcohols. The substrate specificity of FARs apparently contributes to the species-specific MMP composition, but other biosynthetic steps, namely the process of fatty acyl and fatty acyl-CoA accumulation, likely also contribute to the final fatty alcohol composition of bumblebee MMPs.

Conclusion

In the present work, we substantially broadened our limited knowledge of the function of FARs in Hymenoptera, one of the largest insect orders. The experimentally determined reductase specificity of FARs that are abundantly expressed in bumblebee male LGs is consistent with their role in MMP biosynthesis. The majority of these MMP-biosynthetic FARs belong to the FAR-A orthology group. We found that the FAR-A group expanded in the *Bombus* and Meliponini lineage. By conducting transcriptome- and genome-scale comparative studies of a FAR gene family across Hymenoptera, assaying tissue-specific FAR gene expression, and experimentally characterizing FAR enzymatic specificities, we provide evidence that lineage-specific gene family expansion shaped the genetic basis of pheromone production in the crown group of bees.

Our analysis of TE distribution in the *B. terrestris* genome indicates that TEs enriched in the vicinity of FAR-A genes might have substantially contributed to the dramatic expansion of the FAR-A gene group. In the future, the increasing availability of genomic and transcriptomic resources for Hymenoptera should enable us to more precisely delineate the taxonomic extent and evolutionary timing of the massive FAR gene family expansion and assess in detail the role of TEs in the process.

Methods

Insects

Specimens of *Bombus lucorum* and *Bombus lapidarius* were obtained from laboratory colonies established from naturally overwintering bumblebee queens. The *Bombus terrestris* specimens originated from laboratory colonies obtained from a bumblebee rearing facility in Troubsko, Czech Republic.

LG and FB samples used for transcriptome sequencing were prepared from 3-day-old bumblebee males by pooling tissues from multiple specimens from the same colony. The cephalic part of the LG and a section of the abdominal peripheral FB were dissected, transferred immediately to TRIzol (Invitrogen), then flash-frozen at −80 °C and stored at this temperature prior to RNA isolation.

RNA isolation and cDNA library construction

For cloning of FARs and RT-qPCR analysis of tissue-specific gene expression, RNA was isolated from individual bumblebee tissues by guanidinium thiocyanate-phenol-chloroform extraction followed by RQ1 DNase (Promega) treatment and RNA purification using the RNeasy Mini Kit (Qiagen). The tissue sample for RNA isolation from virgin queen LGs consisted of pooled glands

433 from two specimens. For RT-qPCR analysis of age-specific expression in *B. lapidarius*, RNA
434 was isolated using the Direct-zol RNA MicroPrep Kit (Zymo Research). A nanodrop ND-1000
435 spectrophotometer (Thermo Fisher) was employed to determine the isolated RNA concentration.
436 The obtained RNA was kept at −80 °C until further use.
437 The cDNA libraries of LGs from 3-day-old bumblebee males were constructed from
438 0.50 µg total RNA using the SMART cDNA Library Construction Kit (Clontech) with either
439 Superscript III (Invitrogen) or M-MuLV (New England Biolabs) reverse transcriptase.
440

441 **Transcriptome sequencing, assembly and annotation**

442 The transcriptomes of male LGs and FBs of *B. lucorum* and *B. terrestris* were assembled as
443 previously described[42]. The male LG and FB transcriptomes of *B. lapidarius* were sequenced and
444 assembled as described[42]. Briefly, total RNA was isolated from the LGs and FBs of three 3-day-
445 old *B. lapidarius* males and pooled into a FB and LG sample. Total RNA (5 µg) from each of the
446 samples was used as starting material. Random primed cDNA libraries were prepared using
447 poly(A)$^+$ enriched mRNA and standard Illumina TrueSeq protocols (Illumina). The resulting
448 cDNA was fragmented to an average of 150 bp. RNA-Seq was carried out by Fasteris (Fasteris)
449 and was performed using an Illumina HiSeq 2500 Sequencing System. Quality control, including
450 filtering high-quality reads based on the fastq score and trimming the read lengths, was carried
451 out using CLC Genomics Workbench software v. 7.0.1 (http://www.clcbio.com). The complete
452 transcriptome libraries were assembled *de novo* using CLC Genomics Workbench software. FAR
453 expression values were calculated by mapping Illumina reads against the predicted coding
454 regions of FAR sequences using bowtie2 v2.2.6[68] and counting the mapped raw reads using ht-
455 seq v0.9.1[69]. The raw read counts were normalized for the FAR coding region length and the total
456 number of reads in the sequenced library, yielding reads per kilobase of transcript per million
457 mapped reads (RPKM) values[70]. A constant value of 1 was added to each RPKM value and

458 subsequently log2-transformed and visualized as heatmaps using the ggplot2 package in R.
459 Complete short read (Illumina HiSeq2500) data for FB and LG libraries from *B. lucorum* and *B.*
460 *lapidarius* were deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) with
461 BioSample accession numbers SAMN08625119, SAMN08625120, SAMN08625121, and
462 SAMN08625122 under BioProject ID PRJNA436452.
463
464 **FAR sequence prediction**
465 The FARs of *B. lucorum* and *B. lapidarius* were predicted based on Blast2GO transcriptome
466 annotation and their high protein sequence similarity to previously characterized FARs from the
467 European honeybee *Apis mellifera*[26] and the silk moth *Bombyx mori*[44].
468 FARs from annotated genomes or transcriptomes of other hymenopteran species (*Bombus*
469 *impatiens*[62], *Bombus terrestris*[62], *Melipona quadrifasciata*[71], *Apis mellifera*[64], *Megachile*
470 *rotundata*[71], *Dufourea novaeangliae*, *Camponotus floridanus*[72], *Acromyrmex echinatior*[73],
471 *Harpegnathos saltator*[72], *Nasoni vitripenis*[74] and *Polistes canadensis*[75]) were retrieved by blastp
472 searches (*E*-value cutoff 10^{-5}) of the species-specific NCBI RefSeq protein database or UniProt
473 protein database using predicted protein sequences of *B. lucorum*, *B. lapidarius* and *B. terrestris*
474 FARs (accessed February 2017). An additional round of blastp searches using FARs found in the
475 first blastp search round did not yield any additional significant (*E*-value < 10^{-5}) blastp hits,
476 indicating that all FAR homologs were found in the first round of blastp searches (data not
477 shown). For FARs with multiple predicted splice variants, only the longest protein was used for
478 phylogenetic tree reconstruction. FARs from non-annotated transcriptomes of *Bombus rupestris*
479 and *Tetragonula carbonaria* were retrieved via tblastn search (*E*-value cutoff 10^{-5}) of the publicly
480 available contig sequences (BioProject PRJNA252240 and PRJNA252285, respectively[76]) using
481 *Bombus* FARs as a query. The longest translated ORFs were used as a query in tblastn searches
482 against NCBI non-redundant nucleotide database (nr/nt) and ORFs not yielding highly-scoring

483 blast hits annotated as FARs were rejected. The final FAR proteins were used for gene tree
484 reconstruction.

485 The active site, conserved Rossmann fold NAD(P)$^{+}$ binding domain (NABD)[77] and a
486 putative substrate binding site in FAR coding sequences were predicted using Batch conserved
487 domain search[78]. The matrix of protein identities was calculated using Clustal Omega with
488 default parameters (https://www.ebi.ac.uk/Tools/msa/clustalo/ accessed February 2018).
489
490 **FAR gene tree reconstruction**
491 The protein sequences of predicted hymenopteran FARs were aligned using mafft v7.305. The
492 gene tree was inferred in IQTREE v1.5.5 with 1,000 ultrafast bootstrap approximation
493 replicates[79], and with a model of amino acid substitution determined by ModelFinder[80]
494 implemented in IQTREE. The tree was visualized and annotated using the ggtree package[81] in R
495 programming language[82].
496
497 **Genome alignment and TE-enrichment analysis**
498 The genomes of *A. mellifera* and *B. terrestris* were aligned using MAUVE 2.4.0[83]. The genomic
499 position of predicted *B. terrestris* FAR genes was visually inspected using the NCBI Graphical
500 sequence viewer (accessed January 2018 at Nucleotide Entrez Database).

501 TE-enrichment analysis in the vicinity of FAR genes in the *B. terrestris* genome was
502 carried out to explore the impact of TEs in extensive expansion of FAR-A genes. The genomic
503 data were retrieved from the FTP server of the Bumble Bee Genome Project (accessed March
504 2018)[62]. TE density around FAR genes was calculated 10 kb upstream and downstream of each
505 FAR gene, separately for FAR-A genes and non-FAR-A genes. Statistical significances were
506 obtained by permutation test. We compared FAR-A/non-FAR-A gene set average TE density to
507 the null distribution of the average TE densities around *B. terrestris* genes built by 10,000

randomly sampling gene sets of a size corresponding to the size of the FAR-A/non-FAR-A gene set from publicly available Augustus gene predictions[62]. TE densities were analysed for pooled set of all TEs and separately for each TE class and major TE family (Class I: LINE, LTR, LARD; Class II: TIR, MITE, TRIM) using custom shell scripts and bedtools, a suite of Unix genomic tools[84]. R programming language was used for statistical analysis[82].

Quantitative PCR analysis of FAR expression

First-strand cDNA was synthesized from 0.30 µg total RNA using oligo(dT)$_{12-18}$ primers and Superscript III reverse transcriptase. The resulting cDNA samples were diluted 5-fold with water prior to RT-qPCR. The primers used for the assay (Supplementary Table 4) were designed with Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/)[85] and first tested for specificity by employing amplicon melting curve analysis on pooled cDNAs from each species.

The reaction mixtures were prepared in a total volume of 20 µL consisting of 2 µL sample and 500 nM of each primer using LightCycler 480 SYBR Green I Master kit (Roche). The reactions were run in technical duplicates for each sample. RT-qPCR was performed on a LightCycler 480 Instrument II (Roche) in 96-well plates under the following conditions: 95 °C for 60 s, then 45 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s followed by a final step at 72 °C for 2 min.

The acquired data were processed with LightCycler 480 Software 1.5 (Roche) and further analysed with MS Excel (Microsoft Corporation). FAR transcript abundances were normalized to the reference genes phospholipase A2 (PLA2) and elongation factor 1α (eEF1α) as described[86].

FAR gene isolation and cloning

The predicted coding regions of FARs from *B. lucorum*, *B. lapidarius* and *B. terrestris* were

amplified by PCR from LG cDNA libraries using gene-specific primers (Supplementary Table 4) and Phusion HF DNA polymerase (New England Biolabs). Parts of the full-length coding sequence of *Blap*FAR-A5 were obtained by RACE procedure. Briefly, the PCR-amplified sequences containing the 5' and 3' ends of *Blap*FAR-A5 were inserted into pCR2.1 TOPO vector using TOPO TA Cloning kit (Invitrogen) and sequenced by Sanger method. The resulting sequences overlapped with contig sequences retrieved from the *B. lapidarius* transcriptome. The full-length *Blap*FAR-A5-coding region was subsequently isolated using gene-specific PCR primers. The sequence of *Blap*FAR-A1 and yeast codon-optimized sequences of *Bluc*FAR-A1-opt and *Bluc*FAR-A2-opt were obtained by custom gene synthesis (Genscript); see Supplementary Data 1 for synthetic sequences. The individual FAR coding regions were then inserted into linearized pYEXTHS-BN vector[87] using the following restriction sites: *Bter*/*Bluc*FAR-A1 and *Blap*FAR-J at *Sph*I-*Not*I sites; *Bter*/*Bluc*FAR2, *Blap*FAR-A1, *Blap*FAR-A1-short and *Blap*FAR-A5 at *Bam*HI-*Not*I sites; and *Bluc*FAR-A1-opt/FAR-A2-opt and *Blap*FAR-A4 at *Bam*HI-*Eco*RI sites. In the case of *Bter*FAR-J, the *Taq* DNA polymerase (New England Biolabs)-amplified sequence was first inserted into pCR2.1 TOPO vector and then subcloned into pYEXTHS-BN via *Bam*HI-*Eco*RI sites using the In-Fusion HD Cloning kit (Clontech).

The resulting vectors containing FAR sequences *N*-terminally fused with 6×His-tag were subsequently transformed into *E. coli* DH5α cells (Invitrogen). The plasmids were isolated from bacteria with Zyppy Plasmid Miniprep kit (Zymo Research) and Sanger sequenced prior to transformation into yeast. The protein-coding sequences of all studied FARs were deposited to GenBank under accession numbers MG450697–MG450704 and MG930980–MG930983.

556 **Functional assay of FARs in yeast**

557 Expression vectors carrying FAR-coding sequences were transformed into *Saccharomyces*
558 *cerevisiae* strain BY4741 (*MAT*a *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*)[88] using S.c. EasyComp
559 Transformation Kit (Invitrogen). To test FAR specificity, yeasts were cultured for 3 days in
560 20 mL synthetic complete medium lacking uracil (SC−U) supplemented with 0.5 mM Cu^{2+}
561 (inducer of heterologous gene expression), 0.2% peptone and 0.1% yeast extract. The yeast
562 cultures were then washed with water and the cell pellets lyophilized before proceeding with lipid
563 extraction. FAR specificities were determined with the FARs acting on natural substrates present
564 in yeast cells and with individual fatty acyls added to the cultivation media, with the respective
565 fatty alcohols present in the LGs of studied bumblebees. Yeast cultures were supplemented with
566 the following fatty acyls: 0.1 mM *Z*9,*Z*12-18:2COOH (linoleic acid), *Z*9,*Z*12,*Z*15-18:3COOH
567 (*α*-linolenic acid) or *Z*15-20:1Me solubilized with 0.05% tergitol. We chose *Z*15:20:1 as a
568 representative monounsaturated C20+ fatty acyl substrate because *Z*15-20:1OH is the most
569 abundant monounsaturated fatty alcohol in *B. terrestris* LG (Fig. 6 and Supplementary Table 5).

570 The level of heterologous expression of bumblebee FARs was assayed by Western blot
571 analysis of the whole-cell extracts (obtained via sonication) using anti-6×His-tag antibody-HRP
572 conjugate (Sigma-Aldrich) and SuperSignal West Femto Maximum Sensitivity Substrate kit
573 (Thermo Fisher Scientific).

574

575 **Lipid extraction and transesterification**

576 Lipids were extracted from bumblebee tissue samples under vigorous shaking using a 1:1 mixture
577 of CH_2Cl_2/MeOH, followed by addition of an equal amount of hexane and sonication. The
578 extracts were kept at −20 °C prior to GC analysis.

579 Base-catalysed transesterification was performed as described previously[89] with
580 modifications: the sample was shaken vigorously with 1.2 mL 2:1 CH_2Cl_2/MeOH and glass beads

(0.5 mm) for 1 h. After brief centrifugation to remove particulate debris, 1 mL supernatant was evaporated under nitrogen, and the residue was dissolved using 0.2 mL 0.5 M KOH in methanol. The mixture was shaken for 0.5 h and then neutralized by adding 0.2 mL Na_2HPO_4 and KH_2PO_4 (0.25 M each) and 35 µL 4 M HCl. The obtained FAMEs were extracted with 600 µL hexanes and analysed by gas chromatography.

For quantification purposes, either 1-bromodecane (10:Br) or 1-bromoeicosane (20:Br) were added to the extracts as internal standards.

Gas chromatography and fatty alcohol ratio determination

Standards of *Z*9,*Z*12,*Z*15-18:3OH and *Z*15-20:1OH were prepared from their corresponding acids/FAMEs by reduction with $LiAlH_4$. The *Z*9-18:1Me standard was prepared by reacting oleoyl chloride with methanol. Other FAME and fatty alcohol standards were obtained from commercial suppliers. The FA-derived compounds in extracts were identified based on the comparison of their retention times with the standards and comparison of measured MS spectra with those from spectral libraries. Double bond positions were assigned after derivatization with dimethyl disulfide[90].

The fatty alcohol ratio was calculated as the molar percentage of fatty alcohol relative to the total fatty alcohol and fatty acyls (e.g. free FAs, fatty acyl-CoAs, and triacylglycerols) containing the same fatty chain structure, i.e., the same chain length and double bond position. The fatty alcohol ratio represents the hypothetical degree of conversion of total fatty acyls (all the quantified fatty acyls counted as fatty acyl-CoAs) to the respective fatty alcohol and reflects the apparent FAR specificity in the investigated bumblebee tissue or yeast cell.

GC-FID

A flame-ionization detector was used for quantitative assessment of the FA-derived compounds.

606 The separations were performed on a Zebron ZB-5ms column (30 m × 250 µm I. D. × 0.25 µm
607 film thickness, Phenomenex) using a 6890 gas chromatograph (Agilent Technologies) with
608 following parameters: helium carrier gas, 250 °C injector temperature, and 1 mL.min^{-1}column
609 flow. The following oven temperature program was used: 100 °C (held for 1 min), ramp to
610 285 °C at a rate of 4 °C.min^{-1} and a second ramp to 320 °C at a rate of 20 °C.min^{-1} with a final
611 hold for 5 min at 320 °C. The analytes were detected in FID at 300 °C using a makeup flow of
612 25 mL.min^{-1} (nitrogen), hydrogen flow of 40 mL.min^{-1}, air flow of 400 mL.min^{-1} and acquisition
613 rate of 5 Hz. The collected data were processed in Clarity (DataApex).

614 **GC×GC-MS**

615 This approach was used to identify analytes by comparing their retention characteristics and mass
616 spectra with those of synthetic standards. The following conditions were employed using a
617 6890N gas chromatograph (Agilent Technologies) coupled to a Pegasus IV D time-of-flight
618 (TOF) mass selective detector (LECO Corp.): helium carrier gas, 250 °C injector temperature,
619 1 mL.min^{-1}column flow, modulation time of 4 s (hot pulse time 0.8 s, cool time 1.2 s), modulator
620 temperature offset of +20 °C (relative to secondary oven) and secondary oven temperature offset
621 of +10 °C (relative to primary oven). Zebron ZB-5ms (30 m × 250 µm I. D. × 0.25 µm film
622 thickness, Phenomenex) was used as a non-polar primary column and BPX-50 (1.5 m × 100 µm
623 I. D. × 0.10 µm film thickness, SGE) was used as a more polar secondary column. The primary
624 oven temperature program was as follows: 100 °C (1 min), then a single ramp to 320 °C at a rate
625 of 4 °C.min^{-1} with a final hold for 5 min at 320 °C.

626 The mass selective detector was operated in electron ionization mode (electron voltage
627 −70 V) with a transfer line temperature of 260 °C, ion source temperature of 220 °C, 100 Hz
628 acquisition rate, mass scan range of 30–600 u and 1800 V detector voltage. ChromaTOF software
629 (LECO Corp.) was used to collect and analyse the data.

Organic synthesis

Methyl *Z*15-eicosenoate (*Z*15-20:1Me, **4**) was synthesized by a new and efficient four-step procedure, starting from inexpensive and easily available cyclopentadecanone. The C1–C15 part of the molecule was obtained by Baeyer-Villiger oxidation of cyclopentadecanone, followed by subsequent methanolysis of the resulting lactone **1** and Swern oxidation of the terminal alcohol group of **2**; the C16–C20 fragment was then connected to the aldehyde **3** by Wittig olefination.

All reactions were conducted in flame- or oven-dried glassware under an atmosphere of dry nitrogen. THF, CH_2Cl_2 and MeOH were dried following standard methods under a nitrogen or argon atmosphere. Petroleum ether (PE, 40–65 °C boiling range) was used for chromatographic separations. TLC plates (silica gel with fluorescent indicator 254 nm, Fluka or Macherey-Nagel) were used for reaction monitoring. Flash column chromatographic separations were performed on silica gel 60 (230–400 mesh, Merck or Acros).

IR spectra were taken on an ALPHA spectrometer (Bruker) as neat samples using an ATR device. ^{1}H and ^{13}C NMR spectra were recorded in $CDCl_3$ on an AV III 400 HD spectrometer (Bruker) equipped with a cryo-probe or an AV III 400 spectrometer (Bruker) equipped with an inverse broad-band probe at 400 MHz for ^{1}H and 100 MHz for ^{13}C. ^{1}H NMR chemical shifts were provided in ppm using TMS as external standard; ^{13}C NMR chemical shifts were referenced against the residual solvent peak. The connectivity was determined by ^{1}H-^{1}H COSY experiments. GC-MS (EI) measurements were performed on an Agilent 5975B MSD coupled to a 6890N gas chromatograph (Agilent Technologies). High-resolution MS (HRMS) spectra were measured on a Q-Tof micro spectrometer (resolution 100000 (ESI), Waters) or GCT Premier orthogonal acceleration TOF mass spectrometer (EI and CI, Waters).

1-Oxacyclohexadecan-2-one (**1**)

Cyclopentadecanone (500 mg, 2.23 mmol) was dissolved in dry CH_2Cl_2 (6 mL) and *meta*-chloroperbenzoic acid (*m*CPBA*)* (687 mg, 2.79 mmol, 70%) was added at 0 °C. The reaction mixture was stirred at room temperature (r.t.), occasionally concentrated under a flow of nitrogen, and the solid residue was re-dissolved in dry CH_2Cl_2. After stirring for four days, the conversion was still not complete; additional *m*CPBA (164 mg, 667 µmol, 70%) was added at 0 °C and stirring was continued at r.t. for 48 h. The mixture was diluted with CH_2Cl_2 (20 mL), and the organic layer was washed with saturated $NaHCO_3$ solution (5×5 mL) and brine (5 mL). The organic layer was dried over Na_2SO_4, filtered and evaporated. The crude product was purified by column chromatography (50 mL silica gel, PE/CH_2Cl_2 1:1) providing product **1** (426 mg, 80%) as a colorless waxy solid.

1: Melting point (m.p.) <30 °C. R_f (PE/Et_2O 95:5) = 0.5. IR (film): ν = 2925, 2855, 1733, 1459, 1385, 1349, 1234, 1165, 1108, 1070, 1013, 963, 801, 720 cm^{-1}. HRMS (+EI TOF) *m*/*z*: ($C_{15}H_{28}O_2$) calc.: 240.2089, found: 240.2090. ^1H NMR (400 MHz, $CDCl_3$) δ = 4.13 (t, J = 5.7 Hz, 2H, H16), 2.33 (t, J = 7.0 Hz, 2H, H3), 1.72–1.56 (m, 4H, H4, H15), 1.48–1.37 (m, 2H, H14), 1.36–1.23 (m, 18H, H5–H13). ^{13}C NMR (100 MHz, $CDCl_3$) δ = 174.2, 64.1, 34.6, 28.5, 27.9, 27.28, 27.26, 27.1, 26.8, 26.5, 26.2, 26.1, 26.0, 25.3, 25.1.

Methyl 15-hydroxypentadecanoate (2)

MeOK (74 µL, 208 µmol, 2.81M in MeOH) was added dropwise at 0° C to a mixture of lactone **1** (50 mg, 208 µmol), dry THF (0.5 mL) and MeOH (1 mL). The mixture was stirred at r.t. for 48 h, by which point the reaction was complete as indicated by TLC. The solution was quenched with a few drops of water and diluted with Et_2O (5 mL). After stirring for 30 min, the layers were separated and the aqueous layer was extracted with Et_2O (3×3 mL). The combined organic layers were washed with brine and water, dried over Na_2SO_4, filtered and evaporated to obtain nearly pure product. Purification by column chromatography (5 mL silica gel, PE/EtOAc 9:1) provided

product **2** (55 mg, 97%) as a colorless solid.

2: m.p. 47–48 °C. R_f (PE/Et_2O 95:5) = 0.2. IR (film): ν = 3285, 2917, 2849, 1740, 1473, 1463, 1435, 1412, 1382, 1313, 1286, 1264, 1240, 1217, 1196, 1175, 1117, 1071, 1061, 1049, 1025, 1013, 992, 973, 926, 884, 731, 720, 701 cm^{-1}. HRMS (+ESI) *m*/*z*: ($C_{16}H_{32}O_3Na$) calc.: 295.2244, found: 295.2245. ^1H NMR (400 MHz, $CDCl_3$) δ = 3.64 (s, 3H, OCH_3), 3.61 (t, *J* = 6.7 Hz, 2H, H15), 2.28 (t, *J* = 7.5 Hz, 2H, H2), 1.72 (s, 1H, OH), 1.59 (quint, *J* = 7.1 Hz, 2H, H3), 1.54 (quint, *J* = 7.1 Hz, 2H, H14), 1.37–1.14 (m, 20H, H4–H13). ^{13}C NMR (101 MHz, $CDCl_3$) δ = 174.5, 63.1, 51.6, 34.2, 32.9, 29.71 (3C), 29.68 (2C), 29.5 (2C), 29.4, 29.3, 25.9, 25.1.

Methyl 15-oxopentadecanoate (**3**)

Dry DMSO (110 µL, 1.54 mmol) was added at −78 °C dropwise to a mixture of oxalyl chloride (90 µL, 1.03 mmol) and CH_2Cl_2 (2 mL) in a 25 mL flask, and the reaction mixture was stirred for 15 min. The hydroxy ester **2** (140 mg, 0.51 mmol) in dry CH_2Cl_2 (2 mL) was added dropwise via a cannula; the white, turbid reaction mixture was stirred for 40 min, and dry triethylamine (432 µL, 3.08 mmol) was added dropwise. The mixture was stirred at –78 °C for 1 h and warmed to 0 °C over 30 min at which point the reaction was complete according to TLC. The reaction mixture was diluted with CH_2Cl_2 (10 mL), quenched with saturated NH_4Cl solution (5 mL) and water (5 mL), and warmed to r.t. The layers were separated and the aqueous layer was extracted with CH_2Cl_2 (3×10 mL). The combined organic layers were washed with brine, dried over $MgSO_4$, filtered and evaporated. The crude product was purified by flash chromatography (10 mL silica gel, PE/EtOAc 95:5) giving aldehyde **3** (109 mg, 78%) as a colorless waxy solid.

3: m.p. <37 °C. R_f (PE/EtOAc 9:1) = 0.4. IR (film): ν = 2923, 2852, 2752, 1738, 1465, 1436, 1362, 1315, 1243, 1197, 1172, 1120, 1017, 985, 958, 883, 811, 719 cm^{-1}. HRMS (+CI TOF) *m*/*z*: ($C_{16}H_{31}O_3$) calc.: 271.2273, found: 271.2277. ^1H NMR (400 MHz, $CDCl_3$) δ = 9.76 (t, *J* = 1.9 Hz, 1H, H15), 3.66 (s, 3H, OCH_3), 2.41 (td, *J* = 7.4, 1.9 Hz, 2H, H14), 2.29 (t, *J* = 7.5 Hz,

2H, H2), 1.67–1.56 (m, 4H, H3,H13), 1.35–1.20 (m, 18H, H4-H12). ^{13}C NMR (100 MHz, $CDCl_3$) δ = 203.1, 174.5, 51.6, 44.1, 34.3, 29.72, 29.70 (2C), 29.58, 29.56, 29.5, 29.4, 29.31, 29.29, 25.1, 22.2.

Methyl *Z*15-eicosenoate (**4**, *Z*15-20:1Me)

NaHMDS (614 µL, 0.614 mmol, 1.0 M in THF) was added dropwise at −55 °C over 10 min to a suspension of high vacuum-dried (pentyl)triphenylphosphonium bromide (282 mg, 0.68 mmol)[91] in dry THF (3 mL) in a flame dried round-bottomed Schlenk flask. The bright orange reaction mixture was stirred while warming to −40 °C for 50 min, and a solution of aldehyde **3** (92 mg, 0.34 mmol) in dry THF (1.5 mL) was added dropwise via cannula at −45 °C. Stirring was continued for 1 h, and the reaction mixture was warmed to r.t. over 90 min. The reaction mixture was diluted with PE (25 mL); filtered through a short silica gel plug, which was washed with PE; and evaporated. The crude product was purified by flash chromatography (silica gel, gradient PE/EtOAc 100:0 to 95:5) to give methyl ester **4** (88 mg, 79%) as a colorless oil.

4: R_f (PE/Et_2O 95:5) = 0.6. IR (film): ν = 3005, 2922, 2853, 1743, 1699, 1684, 1653, 1541, 1521, 1507, 1489, 1436, 1362, 1196, 1169, 1106, 1017, 880, 722 cm^{-1}. GC-MS (EI) t_R [60 °C (4 min) → 10 °C/min to 320 °C (10 min)] 21 min; *m*/*z* (%): 324 (4) [M^+], 292 (26), 250 (10), 208 (9), 152 (7), 123 (12), 111 (22), 97 (48), 87 (40), 83 (52), 74 (56), 69 (70), 59 (14), 55 (100), 41 (43), 28 (26). HRMS (+EI TOF) *m*/*z*: ($C_{21}H_{40}O_2$) calc.: 324.3028, found: 324.3026. ^{1}H NMR (401 MHz, $CDCl_3$) δ = 5.39–5.30 (m, 2H, H15,H16), 3.66 (s, 3H, OCH_3), 2.30 (t, J = 7.5 Hz, 2H, H2), 2.07–1.96 (m, 4H, H14,H17), 1.61 (quint, J = 7.5 Hz, 2H, H3), 1.37–1.14 (m, 24H, H4-H13,H18,H19), 0.89 (t, J = 7.2 Hz, 3H, H20). ^{13}C NMR (100 MHz, $CDCl_3$) δ = 174.5, 130.1, 130.0, 51.6, 34.3, 32.1, 29.9, 29.81, 29.79, 29.74, 29.70, 29.68, 29.6, 29.5, 29.4, 29.3, 27.4, 27.1, 25.1, 22.5, 14.2.

Statistical analysis

All lipid quantifications in yeast and LGs and FBs of bumblebees and transcript quantifications in bumblebee tissues were performed using three biological replicates (in addition, technical duplicates were used for RT-qPCR). The results are reported as mean value ± S.D. Significant differences were determined by one-way analysis of variance (ANOVA) followed by post-hoc Tukey's honestly significant difference (HSD) test or by a two-tailed *t*-test as indicated in the Results section.

Acknowledgements

This work was supported by the Czech Science Foundation (grant no. 15-06569S) and by the Ministry of Education of the Czech Republic (project n. LO1302). We thank Dr. Stefan Jarau for stimulating discussions. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures" and by the Okinawa Institute of Science and Technology.

References

1. Koonin, E. V. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* **39,** 309–338 (2005).
2. Lespinet, O. The Role of Lineage-Specific Gene Family Expansion in the Evolution of Eukaryotes. *Genome Res.* **12,** 1048–1059 (2002).
3. Demuth, J. P., Bie, T. De, Stajich, J. E., Cristianini, N. & Hahn, M. W. The Evolution of Mammalian Gene Families. *PLoS One* **1,** e85 (2006).
4. Rispe, C. *et al.* Large Gene Family Expansion and Variable Selective Pressures for Cathepsin B in Aphids. *Mol. Biol. Evol.* **25,** 5–17 (2007).
5. Cortesi, F. *et al.* Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes. *Proc. Natl. Acad. Sci.* **112,** 1493–1498 (2015).

6. Niimura, Y. & Nei, M. Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J. Hum. Genet.* **51,** 505–517 (2006).

7. Ohno, S. *Evolution by Gene Duplication*. (1970).

8. Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **18,** 292–298 (2003).

9. Lynch, M. The Evolutionary Fate and Consequences of Duplicate Genes. *Science (80-.).* **290,** 1151–1155 (2000).

10. Innan, H. & Kondrashov, F. The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11,** 97–108 (2010).

11. Eirín-López, J. M., Rebordinos, L., Rooney, A. P. & Rozas, J. The Birth-and-Death Evolution of Multigene Families Revisited. **7,** 170–196 (2012).

12. Wang, M. *et al.* Five Fatty Acyl-Coenzyme A Reductases Are Involved in the Biosynthesis of Primary Alcohols in Aegilops tauschii Leaves. *Front. Plant Sci.* **8,** (2017).

13. Cheng, J. B. & Russell, D. W. Mammalian wax biosynthesis. I. Identification of two fatty acyl-Coenzyme A reductases with different substrate specificities and tissue distributions. *J. Biol. Chem.* **279,** 37789–97 (2004).

14. Jaspers, M. H. J. *et al.* The fatty acyl-CoA reductase Waterproof mediates airway clearance in Drosophila. *Dev. Biol.* **385,** 23–31 (2014).

15. Metz, J. G., Pollard, M. R., Anderson, L., Hayes, T. R. & Lassner, M. W. Purification of a jojoba embryo fatty acyl-coenzyme A reductase and expression of its cDNA in high erucic acid rapeseed. *Plant Physiol.* **122,** 635–44 (2000).

16. Teerawanichpan, P. & Qiu, X. Fatty acyl-CoA reductase and wax synthase from Euglena gracilis in the biosynthesis of medium-chain wax esters. *Lipids* **45,** 263–73 (2010).

17. Teerawanichpan, P. & Qiu, X. Molecular and functional analysis of three fatty acyl-CoA reductases with distinct substrate specificities in copepod *Calanus finmarchicus*. *Mar. Biotechnol. (NY).* **14,** 227–36 (2012).

18. Nagan, N. & Zoeller, R. A. Plasmalogens: biosynthesis and functions. *Prog. Lipid Res.* **40,** 199–229 (2001).

19. Lassance, J.-M. *et al.* Functional consequences of sequence variation in the pheromone biosynthetic gene pgFAR for *Ostrinia* moths. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 3967–72 (2013).

20. Lassance, J.-M. *et al.* Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* **466,** 486–9 (2010).

21. Smadja, C. & Butlin, R. K. On the scent of speciation: the chemosensory system and its role in premating

isolation. *Heredity (Edinb).* **102,** 77–97 (2009).

22. Liénard, M. A., Hagström, A. K., Lassance, J.-M. & Löfstedt, C. Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 10955–60 (2010).

23. Liénard, M. A., Wang, H.-L., Lassance, J.-M. & Löfstedt, C. Sex pheromone biosynthetic pathways are conserved between moths and the butterfly Bicyclus anynana. *Nat. Commun.* **5,** (2014).

24. Antony, B. *et al.* Pheromone-gland-specific fatty-acyl reductase in the adzuki bean borer, *Ostrinia scapulalis* (Lepidoptera: Crambidae). *Insect Biochem. Mol. Biol.* **39,** 90–95 (2009).

25. Tupec, M., Buček, A., Valterová, I. & Pichová, I. Biotechnological potential of insect fatty acid-modifying enzymes. *Zeitschrift für Naturforsch. C* **72,** 387–403 (2017).

26. Teerawanichpan, P., Robertson, A. J. & Qiu, X. A fatty acyl-CoA reductase highly expressed in the head of honey bee (*Apis mellifera*) involves biosynthesis of a wide range of aliphatic fatty alcohols. *Insect Biochem. Mol. Biol.* **40,** 641–9 (2010).

27. Hu, Y.-H., Chen, X.-M., Yang, P. & Ding, W.-F. Characterization and functional assay of a fatty acyl-CoA reductase gene in the scale insect, Ericerus pela Chavannes (Hemiptera: Coccoidae). *Arch. Insect Biochem. Physiol.* **97,** e21445 (2018).

28. Ayasse, M. & Jarau, S. Chemical ecology of bumble bees. *Annu. Rev. Entomol.* **59,** 299–319 (2014).

29. Buček, A. *et al.* Exploring complex pheromone biosynthetic processes in the bumblebee male labial gland by RNA sequencing. *Insect Mol. Biol.* **25,** 295–314 (2016).

30. Goulson, D. *Bumblebees - behaviour, ecology and conservation*. (Oxford University Press, Oxford, 2010).

31. Bertsch, A., Schweer, H., Titze, A. & Tanaka, H. Male labial gland secretions and mitochondrial DNA markers support species status of Bombus cryptarum and B. magnus (Hymenoptera, Apidae). *Insectes Soc.* **52,** 45–54 (2005).

32. Rasmont, P. *et al.* Cephalic secretions of the bumblebee subgenus Sibiricobombus Vogt suggest Bombus niveatus Kriechbaumer and Bombus vorticosus Gerstaecker are conspecific (Hymenoptera, Apidae, Bombus). *Apidologie* **36,** 571–584 (2005).

33. Coppée, A., Terzo, M., Valterová, I. & Rasmont, P. Intraspecific Variation of the Cephalic Labial Gland Secretions in Bombus terrestris (L.) (Hymenoptera: Apidae). *Chem. Biodivers.* **5,** 2654–2661 (2008).

34. Bergström, G., Kullberg, B. J. & Ställberg-Stenhagen, S. Studies on natural odoriferous compounds: VII.

Recognition of two forms of *Bombus lucorum* L. (Hymenoptera, Apidae) by analysis of the volatile marking secretion from individual males. *Zoon* **1,** 31–42 (1973).

35. Kullenberg, B., Bergstrom, G., Bringer, B., Carlberg, B. & Cederberg, B. Observations on scent marking by *Bombus* Latr. and *Psithyrus* Lep. males (Hym. Apidae) and localization of site of production of the secretion. *Zoon* **1,** 23–30 (1973).

36. Kullenberg, B., Bergström, G. & Ställberg-Stenhagen, S. Volatile components of the cephalic marking secretion of male bumble bees. *Acta Chem. Scand.* **24,** 1481–5 (1970).

37. Urbanová, K., Valterová, I., Hovorka, O. & Kindl, J. Chemotaxonomical characterisation of males *Bombus lucorum* (Hymenptera: Apidae) collected in Czech republic. *Eur. J. Entomol.* **127,** 111–115 (2001).

38. Šobotník, J. *et al.* Age-dependent changes in structure and function of the male labial gland in *Bombus terrestris*. *J. Insect Physiol.* **54,** 204–14 (2008).

39. Luxová, A., Valterová, I., Stránský, K., Hovorka, O. & Svatoš, A. Biosynthetic studies on marking pheromones of bumblebee males. *Chemoecology* **87,** 81–87 (2003).

40. Žáček, P. *et al.* Comparison of age-dependent quantitative changes in the male labial gland secretion of *Bombus terrestris* and *Bombus lucorum*. *J. Chem. Ecol.* **35,** 698–705 (2009).

41. Tillman, J. a, Seybold, S. J., Jurenka, R. a & Blomquist, G. J. Insect pheromones--an overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* **29,** 481–514 (1999).

42. Buček, A. *et al.* The role of desaturases in the biosynthesis of marking pheromones in bumblebee males. *Insect Biochem. Mol. Biol.* **43,** 724–731 (2013).

43. Žáček, P. *et al.* Biosynthetic studies of the male marking pheromone in bumblebees by using labelled fatty acids and two-dimensional gas chromatography with mass detection. *Chempluschem* **80,** 839–850 (2015).

44. Moto, K. *et al.* Pheromone gland-specific fatty-acyl reductase of the silkmoth, *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 9156–61 (2003).

45. Stolle, E. *et al.* A second generation genetic map of the bumblebee *Bombus terrestris* (Linnaeus, 1758) reveals slow genome and chromosome evolution in the Apidae. *BMC Genomics* **12,** 48 (2011).

46. Angov, E. Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnol. J.* **6,** 650–659 (2011).

47. Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I. & Koonin, E. V. Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Res.* **11,** 555–65 (2001).

48. Radivojac, P. *et al.* A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10,** 221–227 (2013).

49. Elsik, C. G. *et al.* Hymenoptera Genome Database: integrating genome annotations in HymenopteraMine. *Nucleic Acids Res.* **44,** D793–D800 (2016).

50. Johnson, B. R. *et al.* Phylogenomics Resolves Evolutionary Relationships among Ants, Bees, and Wasps. *Curr. Biol.* **23,** 2058–2062 (2013).

51. Mao, M., Gibson, T. & Dowton, M. Higher-level phylogeny of the Hymenoptera inferred from mitochondrial genomes. *Mol. Phylogenet. Evol.* **84,** 34–43 (2015).

52. Ramírez, S. R. *et al.* A molecular phylogeny of the stingless bee genus Melipona (Hymenoptera: Apidae). *Mol. Phylogenet. Evol.* **56,** 519–525 (2010).

53. Pils, B. & Schultz, J. Inactive Enzyme-homologues Find New Function in Regulatory Processes. *J. Mol. Biol.* **340,** 399–404 (2004).

54. Hines, H. M. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera : Apidae : Bombus). *Syst. Biol.* **57,** 58–75 (2008).

55. Jarau, S., Hrncir, M., Zucchi, R. & Barth, F. G. A stingless bee uses labial gland secretions for scent trail communication (Trigona recursa Smith 1863). *J. Comp. Physiol. A Sensory, Neural, Behav. Physiol.* **190,** 233–239 (2004).

56. Jarau, S. *et al.* Hexyl Decanoate, the First Trail Pheromone Compound Identified in a Stingless Bee, Trigona recursa. *J. Chem. Ecol.* **32,** 1555–1564 (2006).

57. Jarau, S. *et al.* The Trail Pheromone of a Stingless Bee, Trigona corvina (Hymenoptera, Apidae, Meliponini), Varies between Populations. *Chem. Senses* **35,** 593–601 (2010).

58. Tumlinson, J. H. *et al.* Identification of a pheromone blend attractive to *Manduca sexta* (L.) males in a wind tunnel. *Arch. Insect Biochem. Physiol.* **10,** 255–271 (1989).

59. Janoušek, V., Laukaitis, C. M., Yanchukov, A. & Karn, R. C. The Role of Retrotransposons in Gene Family Expansions in the Human and Mouse Genomes. *Genome Biol. Evol.* **8,** 2632–2650 (2016).

60. Deininger, P. L., Moran, J. V, Batzer, M. A. & Kazazian, H. H. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* **13,** 651–658 (2003).

61. Reams, A. B. & Roth, J. R. Mechanisms of Gene Duplication and Amplification. *Cold Spring Harb. Perspect. Biol.* **7,** a016592 (2015).

62. Sadd, B. M. *et al.* The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* **16,** 76 (2015).

63. Elsik, C. G. *et al.* Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* **15,** 86 (2014).

64. Weinstock, G. M. *et al.* Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443,** 931–949 (2006).

65. Amsalem, E., Kiefer, J., Schulz, S. & Hefetz, A. The effect of caste and reproductive state on the chemistry of the cephalic labial glands secretion of bombus terrestris. *J. Chem. Ecol.* **40,** 900–12 (2014).

66. Oba, Y., Ojika, M. & Inouye, S. Characterization of CG6178 gene product with high sequence similarity to firefly luciferase in *Drosophila melanogaster*. *Gene* **329,** 137–145 (2004).

67. Matsumoto, S. *et al.* Characterization of acyl-CoA-binding protein (ACBP) in the pheromone gland of the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* **31,** 603–9 (2001).

68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

69. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

70. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–8 (2008).

71. Woodard, S. H. *et al.* Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7472–7477 (2011).

72. Bonasio, R. *et al.* Genomic Comparison of the Ants Camponotus floridanus and Harpegnathos saltator. *Science (80-.).* **329,** 1068–1071 (2010).

73. Nygaard, S. *et al.* The genome of the leaf-cutting ant Acromyrmex echinatior suggests key adaptations to advanced social life and fungus farming. *Genome Res.* **21,** 1339–1348 (2011).

74. Werren, J. H. *et al.* Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species. *Science (80-.).* **327,** 343–348 (2010).

75. Patalano, S. *et al.* Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proc. Natl. Acad. Sci.* **112,** 13970–13975 (2015).

76. Peters, R. S. *et al.* Evolutionary History of the Hymenoptera. *Curr. Biol.* **27,** 1013–1018 (2017).

77. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250,** 194–199 (1974).

78. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222–D226 (2015).

79. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30,** 1188–1195 (2013).

80. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14,** 587–589 (2017).

81. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. <scp>ggtree</scp> : an <scp>r</scp> package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* (2016). doi:10.1111/2041-210X.12628

82. R Core Team. R: A language and environment for statistical computing. (2016).

83. Darling, A. C. E. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Res.* **14,** 1394–1403 (2004).

84. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

85. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13,** 134 (2012).

86. Horňáková, D., Matoušková, P., Kindl, J., Valterová, I. & Pichová, I. Selection of reference genes for real-time polymerase chain reaction analysis in tissues from *Bombus terrestris* and *Bombus lucorum* of different ages. *Anal. Biochem.* **397,** 118–20 (2010).

87. Holz, C., Hesse, O., Bolotina, N., Stahl, U. & Lang, C. A micro-scale process for high-throughput expression of cDNAs in the yeast Saccharomyces cerevisiae. *Protein Expr. Purif.* **25,** 372–8 (2002).

88. Baker Brachmann, C. *et al.* Designer deletion strains derived fromSaccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14,** 115–132 (1998).

89. Matoušková, P. *et al.* A delta9 desaturase from *Bombus lucorum* males: investigation of the biosynthetic pathway of marking pheromones. *Chembiochem* **9,** 2534–41 (2008).

90. Carlson, D. A., Roan, C. S., Yost, R. A. & Hector, J. Dimethyl Disulfide Derivatives of Long-Chain Alkenes,

Alkadienes, and Alkatrienes for Gas-Chromatography Mass-Spectrometry. *Anal. Chem.* **61,** 1564–1571 (1989).

91. Prasad, V. P. *et al.* Synthesis of fluorinated analogues of sphingosine-1-phosphate antagonists as potential radiotracers for molecular imaging using positron emission tomography. *Bioorg. Med. Chem.* **22,** 5168–5181 (2014).

Figures, tables

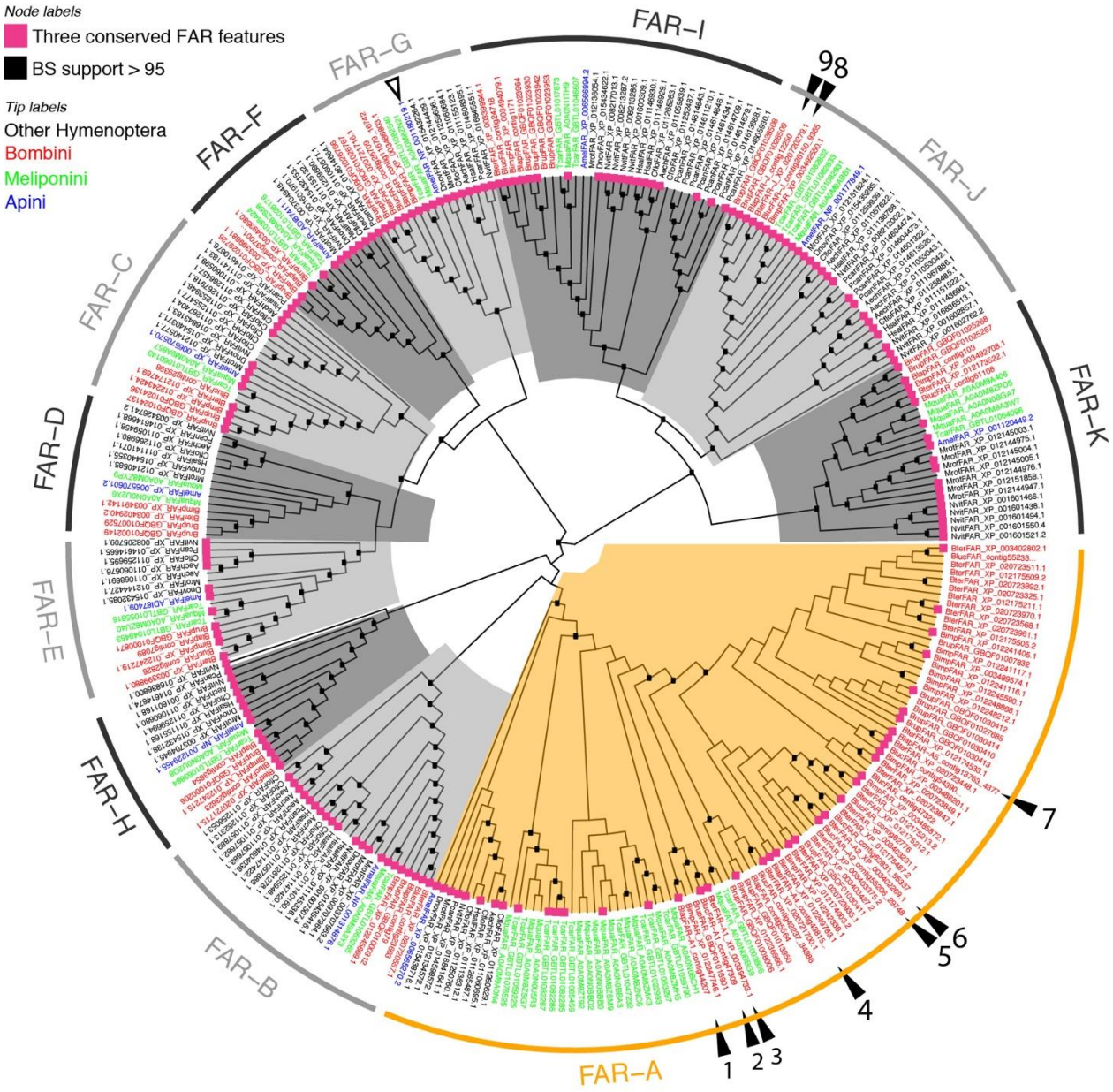
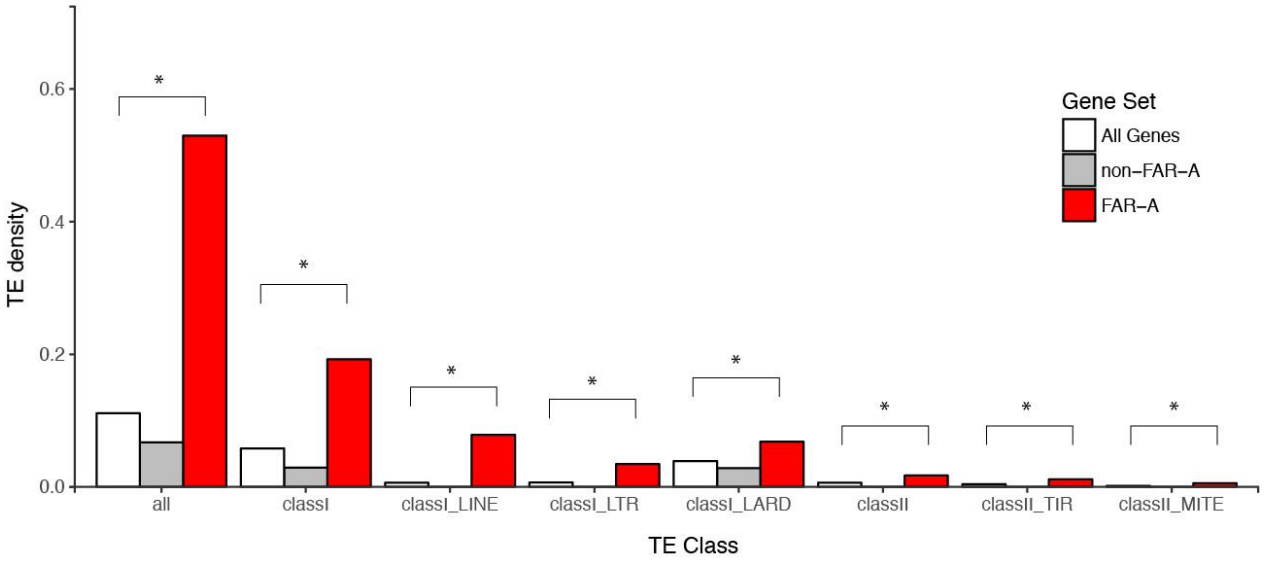


Figure 1. Hymenopteran FAR gene tree. Tree tips are coloured according to taxonomy: red, bumblebee FARs (*B. terrestris*, *B. lucorum*, *B. lapidarius*, *B. impatiens*, *B. rupestris*); green, stingless bee FARs (*Tetragonula carbonaria*, *Melipona quadrifasciata*); blue, *A. mellifera* FARs; and black, FARs from other hymenopteran species. The FAR-A orthology group is highlighted orange; other orthology groups in shades of grey. Functionally characterized bumblebee FARs

943 from this study are indicated by filled triangles and numbered.1: *Blap*FAR-A1, 2: *Bluc*FAR-A1,
944 3: *Bter*FAR-A1, 4: *Blap*FAR-A4, 5: *Bluc*FAR-A2, 6: *Bter*FAR-A2, 7: *Blap*FAR-A5, 8: *Bter*FAR-
945 J, and 9: *Blap*FAR-J. The functionally characterized *A. mellifera* FAR is indicated by an empty
946 triangle. Internal nodes highlighted with black boxes indicate bootstrap support >95%. Violet
947 squares at the tree tips indicate FARs for which CDD search yielded all three FAR conserved
948 features—active site, putative $NAD(P)^+$ binding site and substrate binding site (see Table S1 for
949 complete CDD search results).
950
951

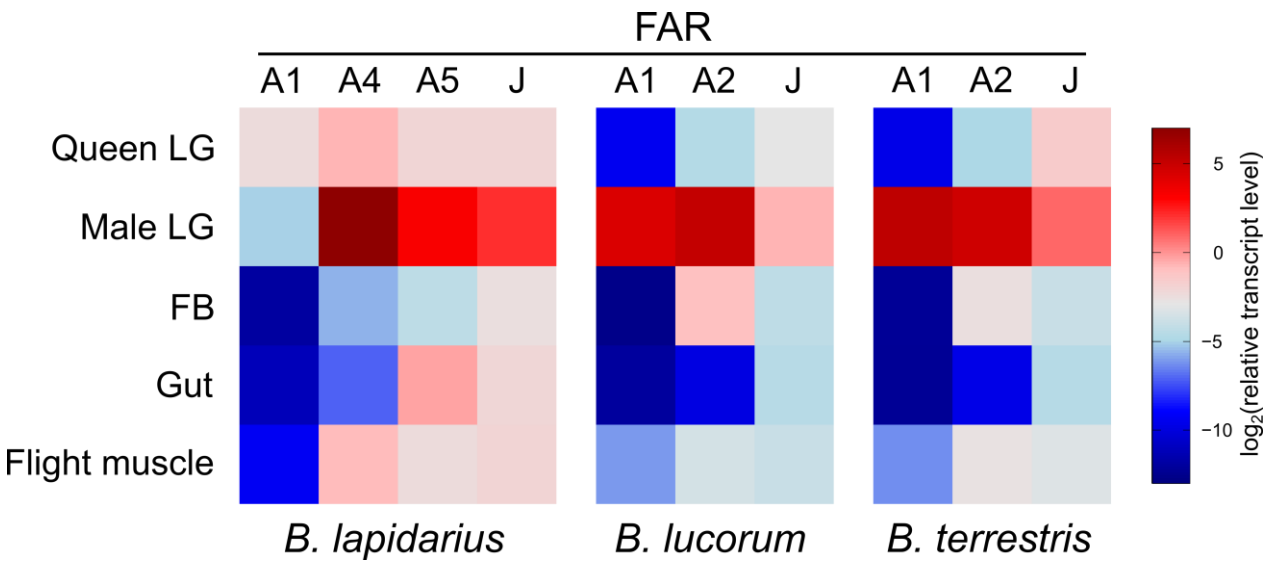
952



953

954 **Figure 2.** Average TE densities in 10 kb windows around groups of *B. terrestris* genes. "All
955 Genes" represents randomly selected sets of *B. terrestris* genes, and non-FAR-A includes FAR
956 genes belonging to the non-FAR-A orthology groups. Densities were analyzed for all TEs (all)
957 and separately for Class I, Class II and the most abundant TE families within each class (LINEs,
958 LTRs, LARDs, TIRs, MITEs). Significant differences ($p < 0.05$, two-tailed *t*-test) are marked
959 with asterisks.

960

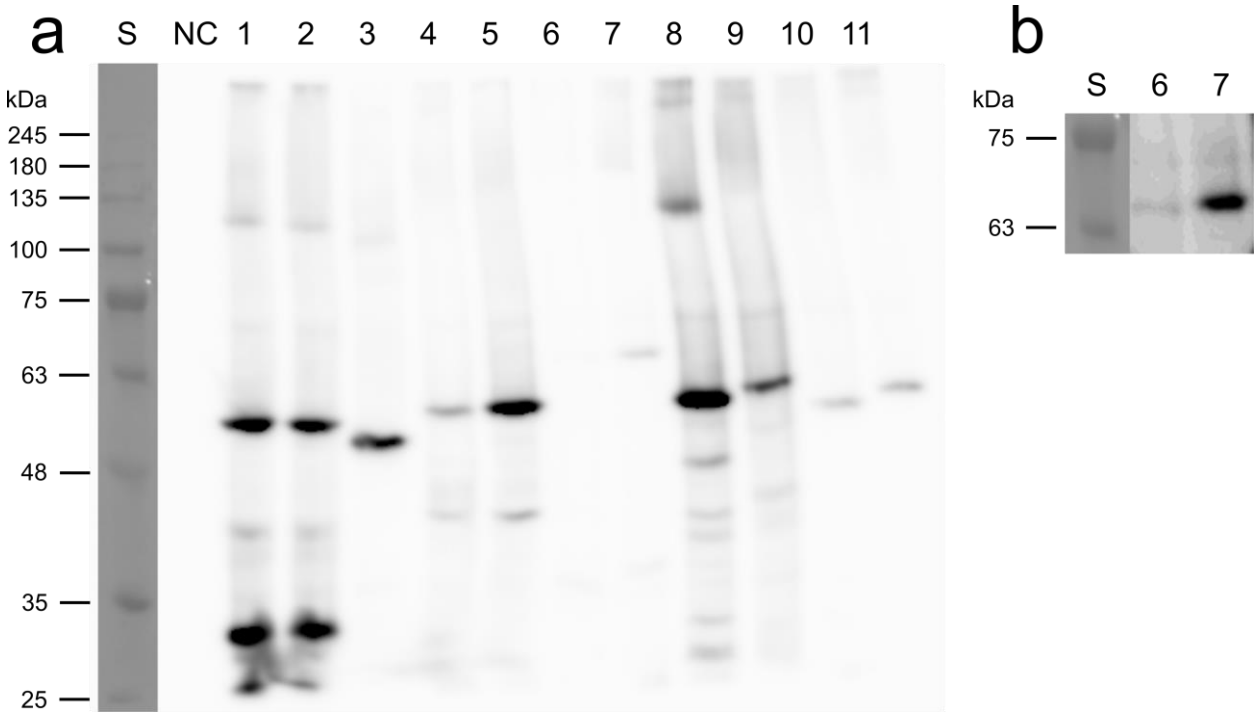


961

962 **Figure 3.** Relative transcript levels of FAR candidates across bumblebee tissues. The transcript
963 levels were assayed by RT-qPCR in male tissues (LG, FB, gut, flight muscle) and in LGs
964 of virgin queens.

965

966



967

968 **Figure 4.** Western blot analysis of FAR protein expression in yeast cell lysates. (**a**) Photograph of
969 the membrane after detection with anti-6×His-tag antibody. (**b**) Detail of the FAR-J full-length
970 protein region with increased contrast. *Lanes*: S, protein standard (VI, AppliChem); NC, negative
971 control (yeast carrying empty vector); 1–11, yeast strains carrying plasmids with *Bluc*FAR-A1
972 (1), *Bter*FAR-A1 (2), *Blap*FAR-A1 (3), *Bluc*FAR-A2 (4), *Bter*FAR-A2 (5), *Bter*FAR-J (6),
973 *Blap*FAR-J (7), *Blap*FAR-A4 (8), *Blap*FAR-A5 (9), *Bluc*FAR-A1-opt (10) and *Bluc*FAR-A2-opt
974 (11).

975

976

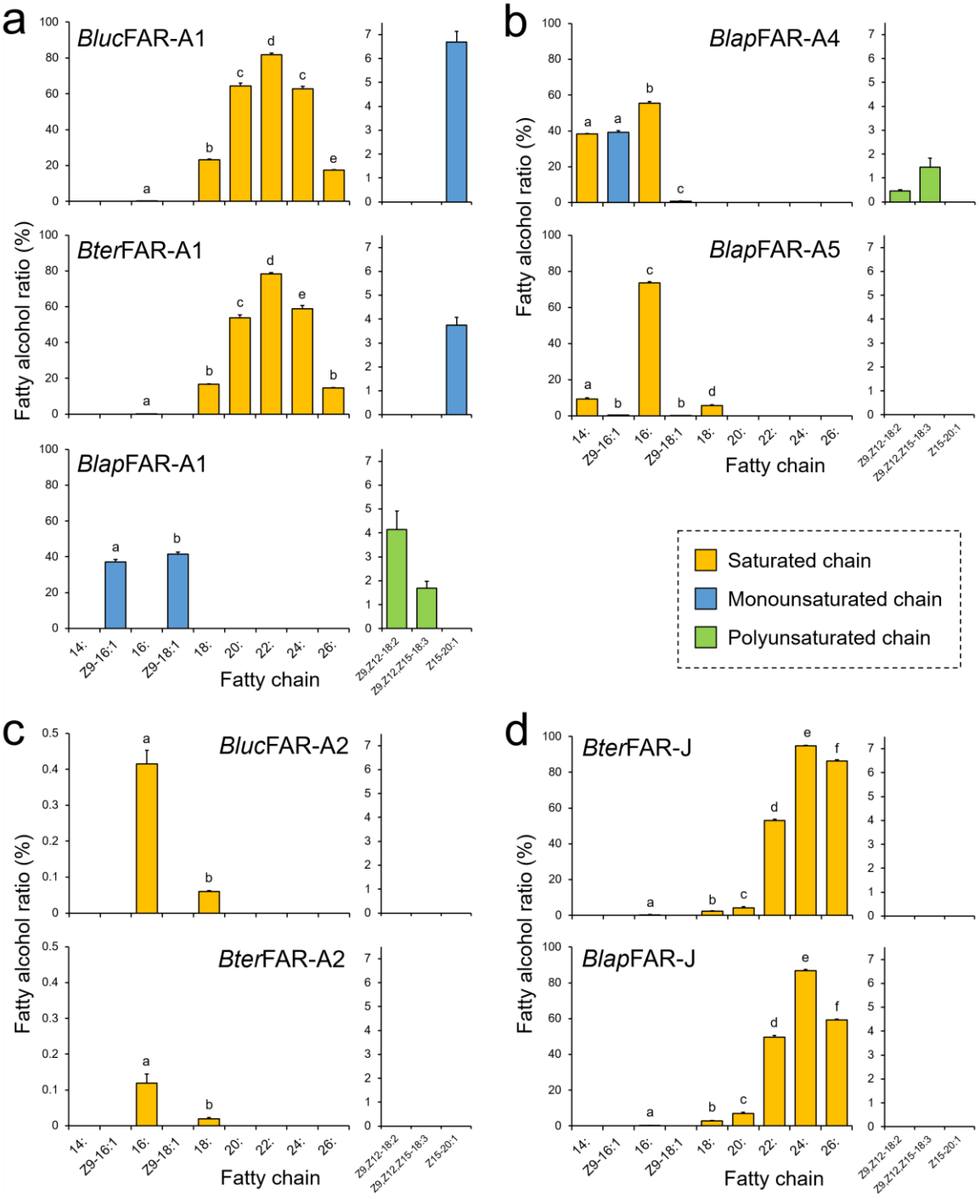
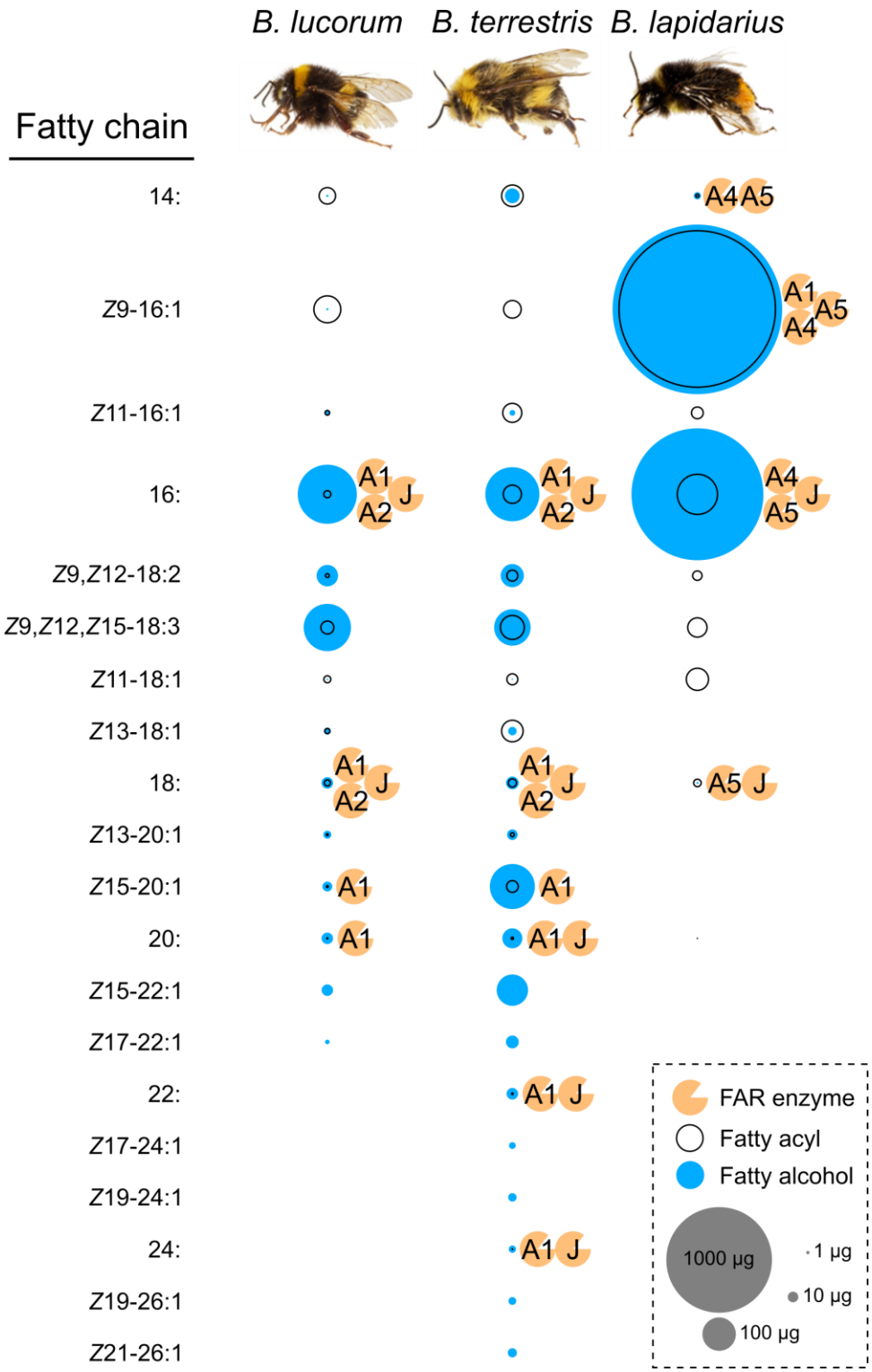


Figure 5. Fatty alcohol ratios in yeast strains expressing bumblebee FARs assayed on yeast native lipids and after supplementation of yeast with either *Z*9,*Z*12-18:2, *Z*9,*Z*12,*Z*15-18:3 or *Z*15-20:1 acyls. Significant differences ($p < 0.01$, one-way ANOVA followed by *post-hoc* Tukey's HSD test) are marked with different letters. See Methods for description of fatty alcohol ratio

982 calculation.

983

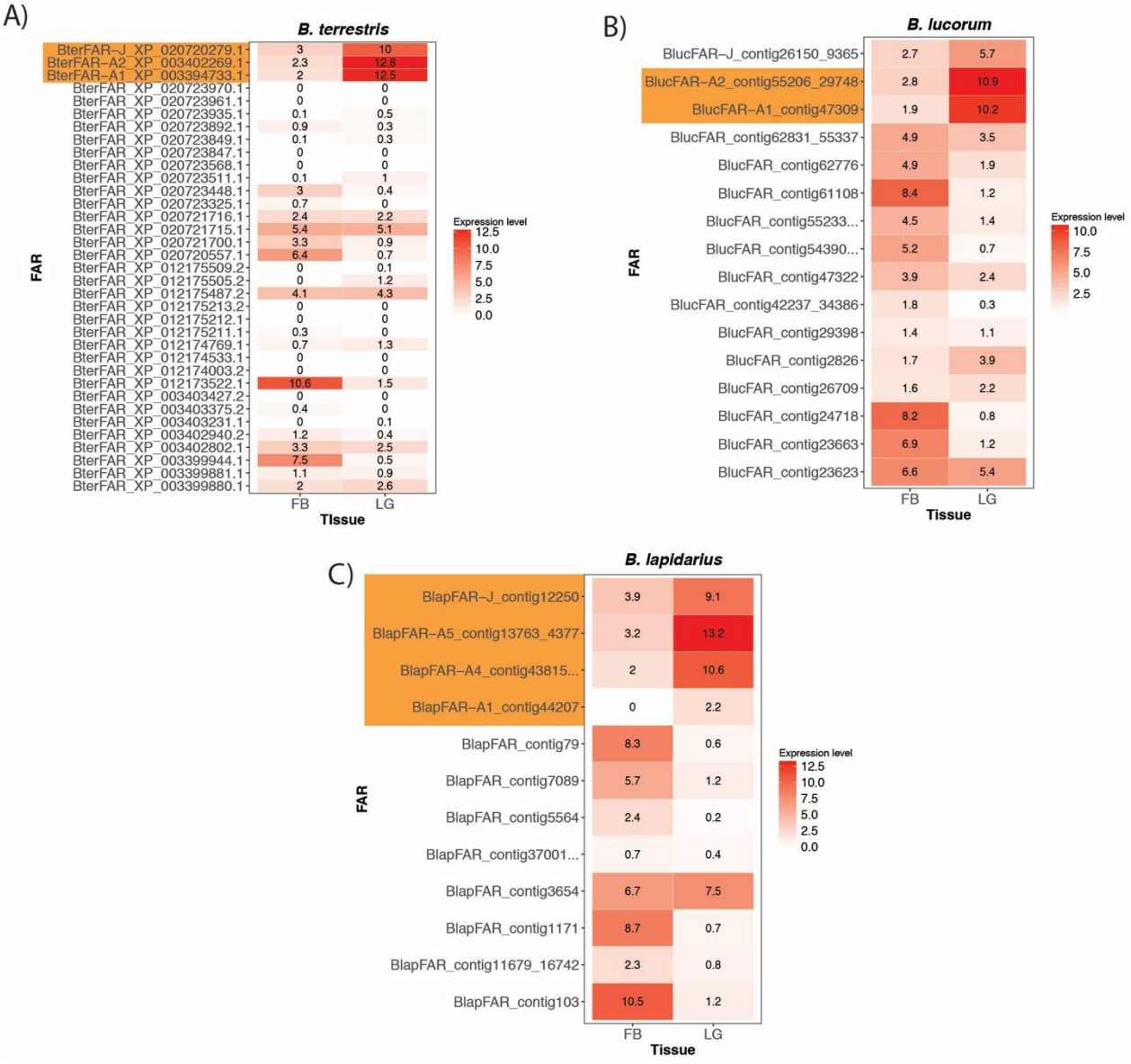


984

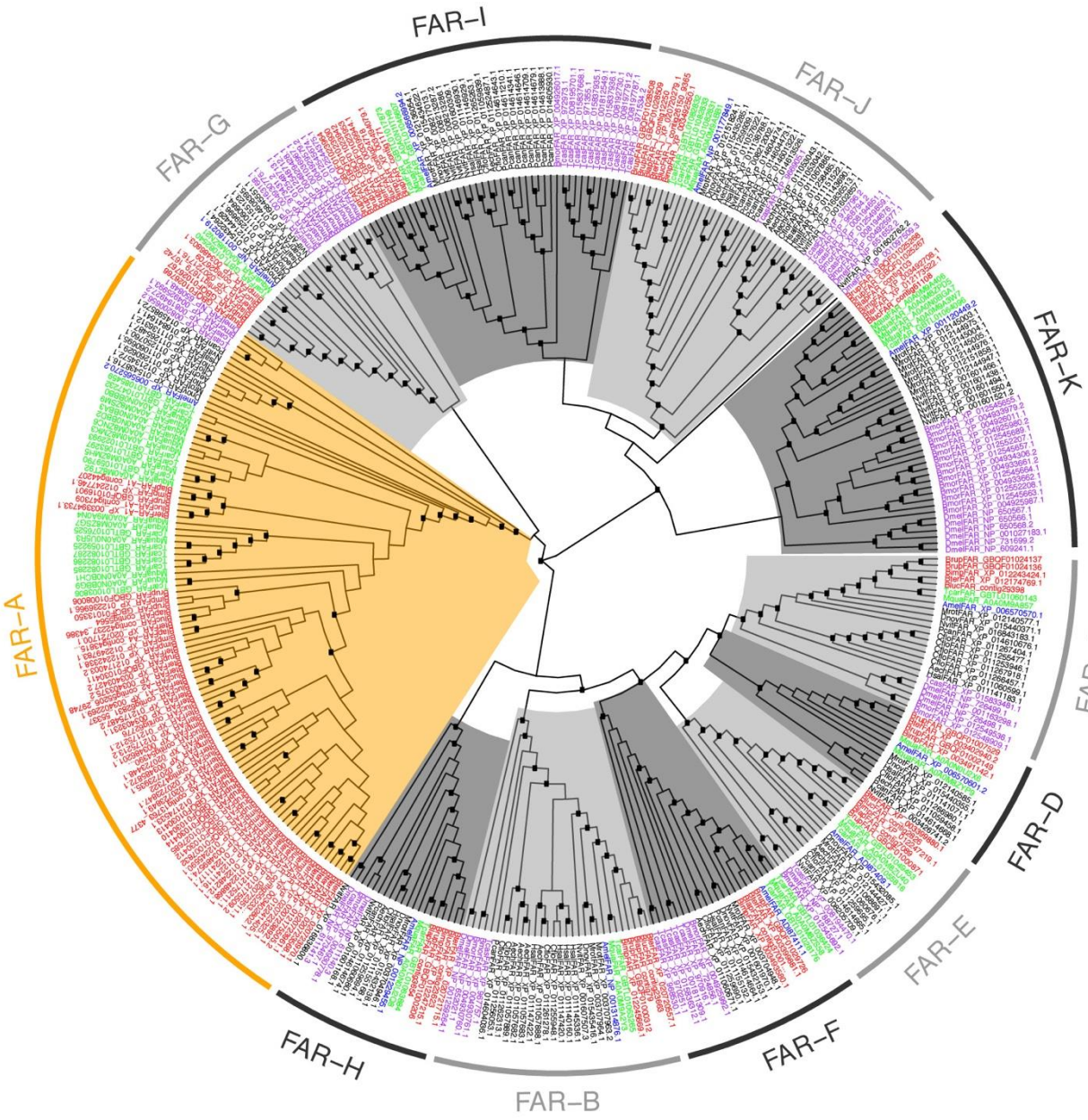
985

986 **Figure 6.** Fatty alcohol and fatty acyl composition in male LGs of *B. lucorum*, *B. terrestris* and
987 *B. lapidarius* together with the proposed participation of FARs. The fatty acyls were determined
988 as corresponding methyl esters. The size of fatty acyl and fatty alcohol circle, respectively,
989 represents the average quantity in single male LG (Supplementary Table 5).
990

Supplementary Materials



Supplementary Figure 1. Expression of FARs in male labial gland and male fat body of *Bombus terrestris* (**A**), *B. lucorum* (**B**) and *B. lapidarius* (**C**). The expression values shown are log2-transformed normalized counts of reads (RPKM values) mapping to the FAR coding regions. FARs functionally characterized in this study are highlighted in orange.

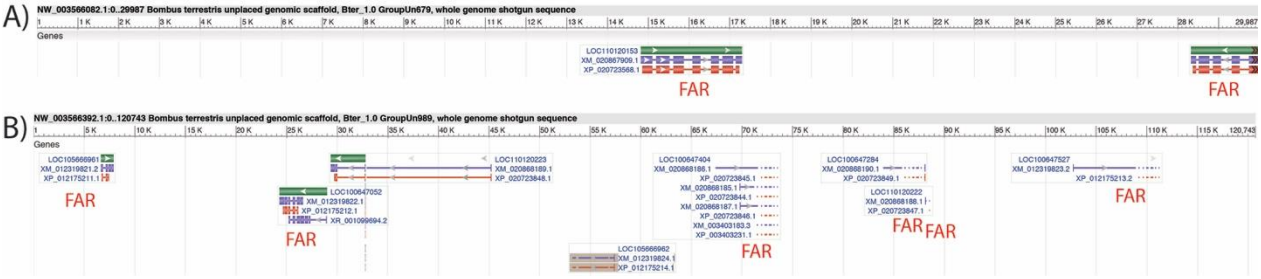


999

1000 **Supplementary Figure 2.** FAR gene tree including non-hymenopteran (*Drosophila*
1001 *melanogaster*, *Bombyx mori*, *Tribolium castaneum*) and hymenopteran FARs. Tree tips are
1002 coloured according to taxonomy: red, bumblebee FARs (*B. terrestris*, *B. lucorum*, *B. lapidarius*,
1003 *B. impatiens*, *B. rupestris*); green, stingless bee FARs (*Tetragonula carbonaria*, *Melipona*
1004 *quadrifasciata*); blue, *A. mellifera* FARs; black, FARs from other hymenopteran species; purple,
1005 non-hymenopteran species. The FAR-A orthology group is highlighted in orange; other
1006 orthologous groups are in shades of grey. Internal nodes highlighted with black boxes indicate

1007 bootstrap support >95 %.

1008



1009

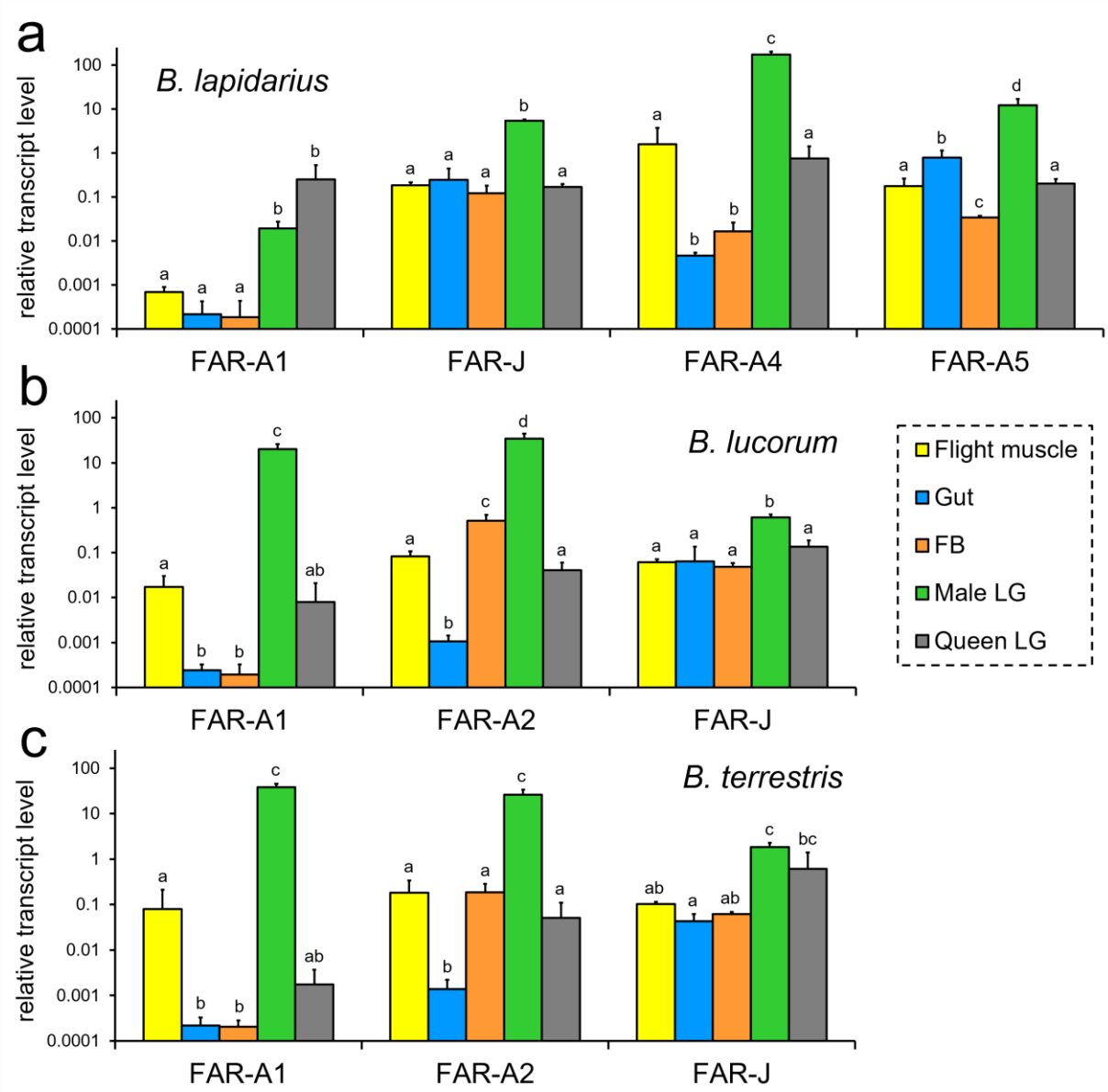
1010 **Supplementary Figure 3.** Clusters of *B. terrestris* FAR-A genes on *B. terrestris* genomic

1011 scaffolds Un679 (**A**) and Un989 (**B**). The horizontal axis shows genomic coordinates of genes

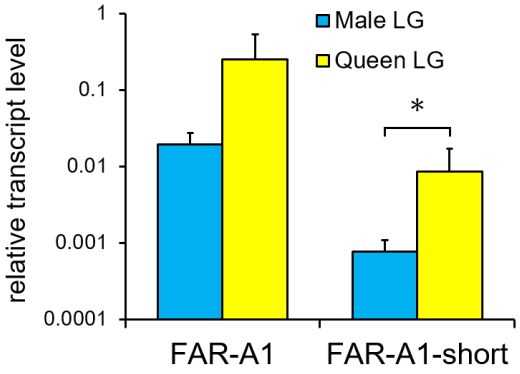
1012 within the scaffold. FAR genes are labelled.

1013

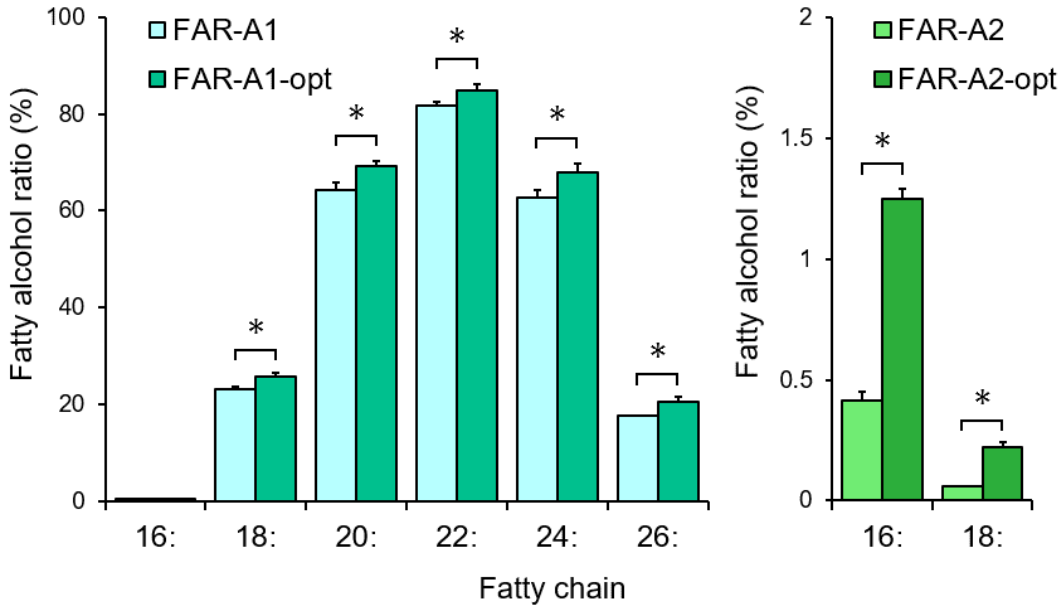
1014



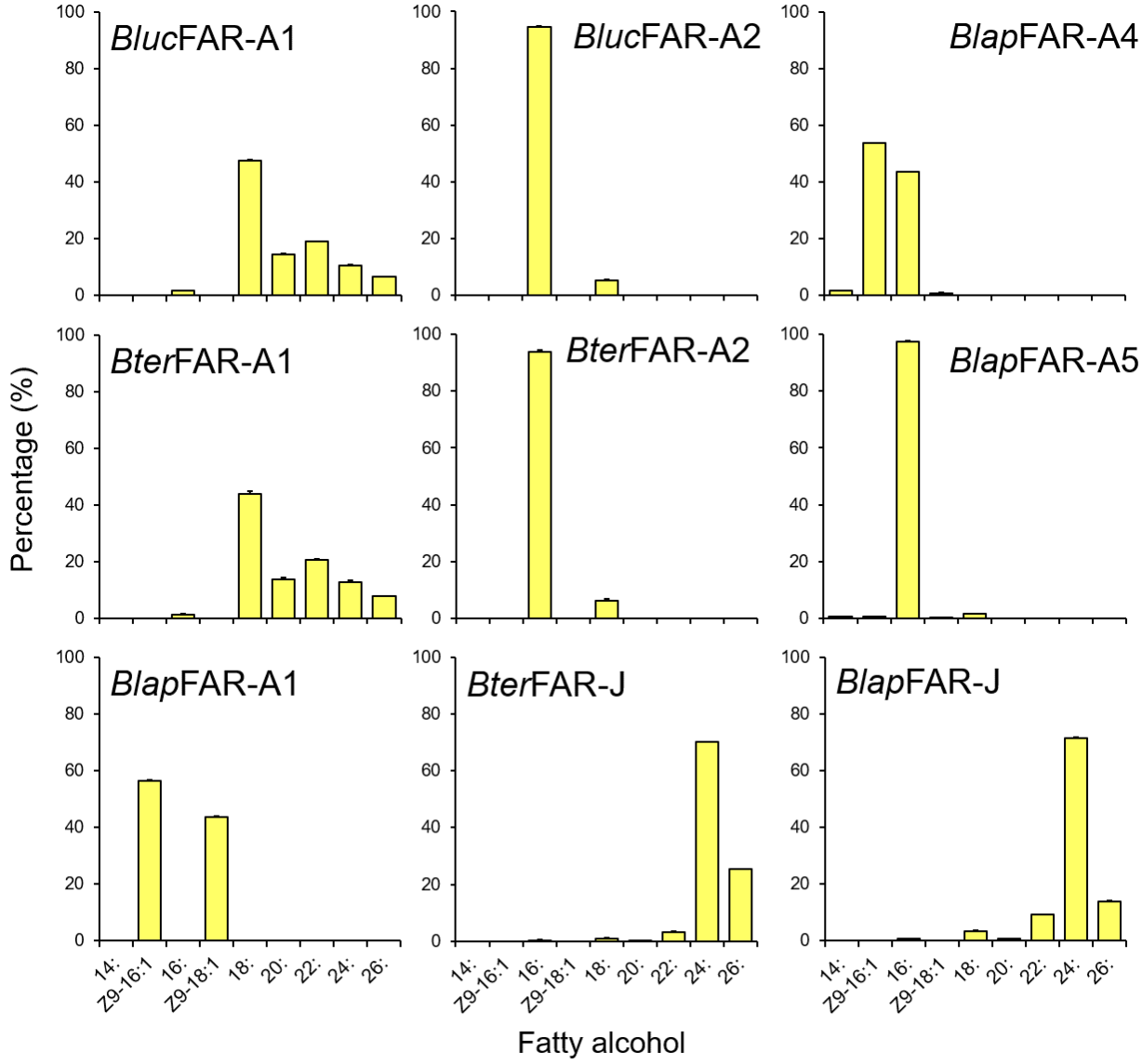
Supplementary Figure 4. Relative transcript levels of FARs in the tissues of *B. lapidarius* (**a**), *B. lucorum* (**b**) and *B. terrestris* (**c**). Significant differences ($p < 0.05$, one-way ANOVA followed by *post-hoc* Tukey's HSD test) are marked with different letters.



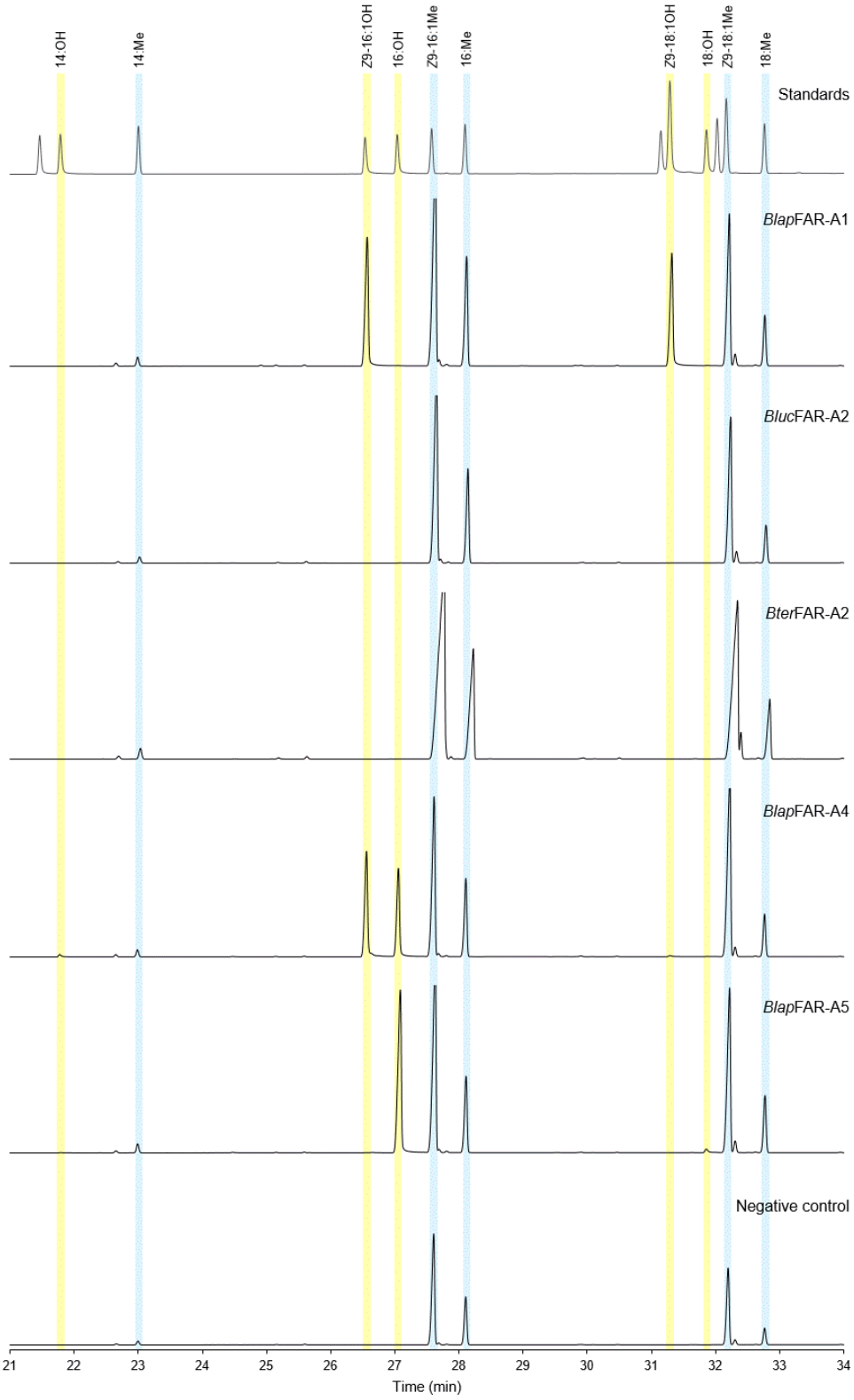
Supplementary Figure 5. Relative transcript levels of FAR1 and FAR1-short in male and queen LGs of *B. lapidarius*. Significant differences ($p < 0.05$, two-tailed *t*-test) are indicated with an asterisk.



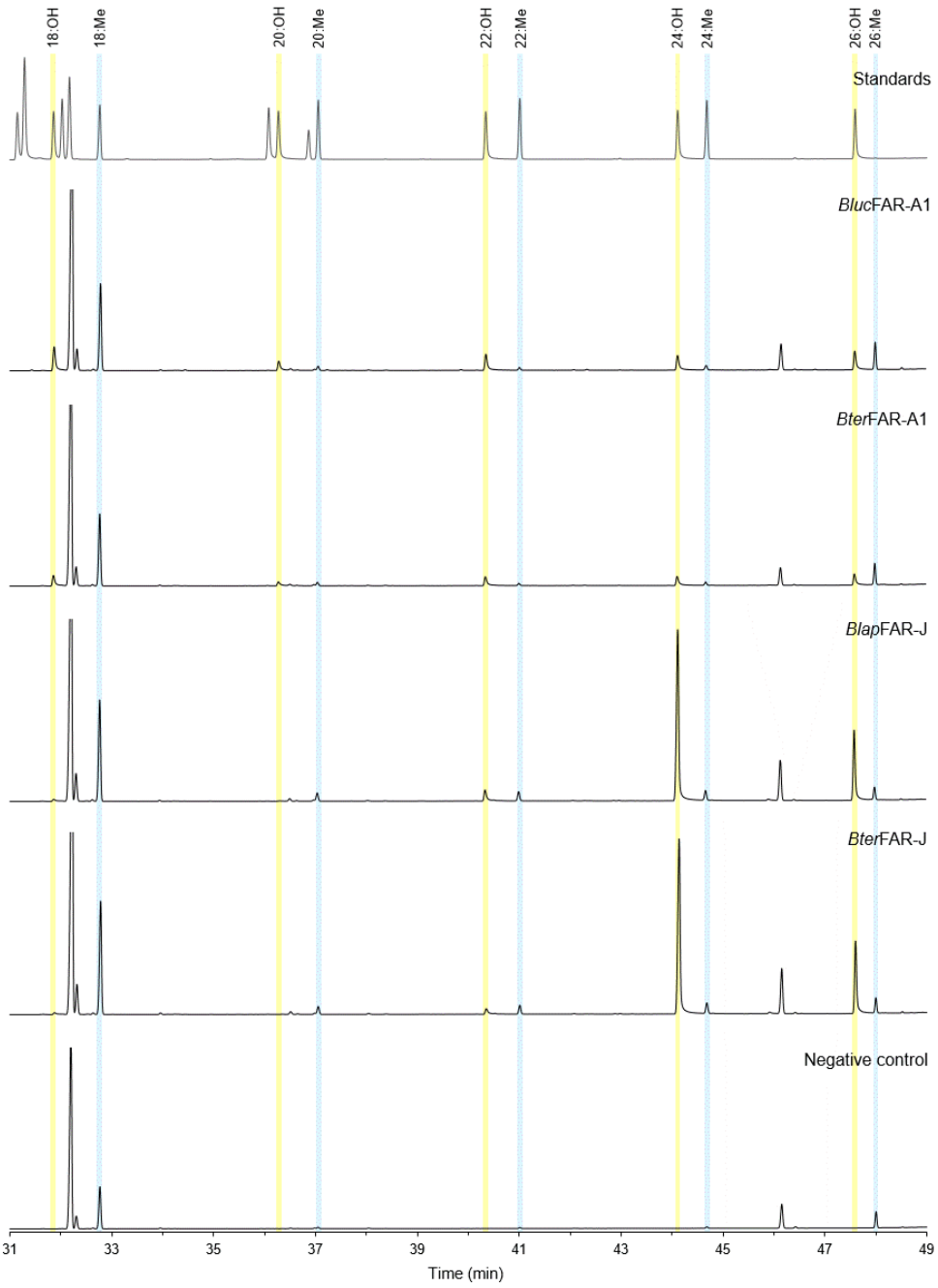
Supplementary Figure 6. Fatty alcohol production by FARs from *B. lucorum* expressed from yeast codon-optimized and wild-type nucleotide sequences. Significant differences ($p < 0.05$, two-tailed *t*-test) are marked with asterisks. See Methods for a description of fatty alcohol ratio calculation.



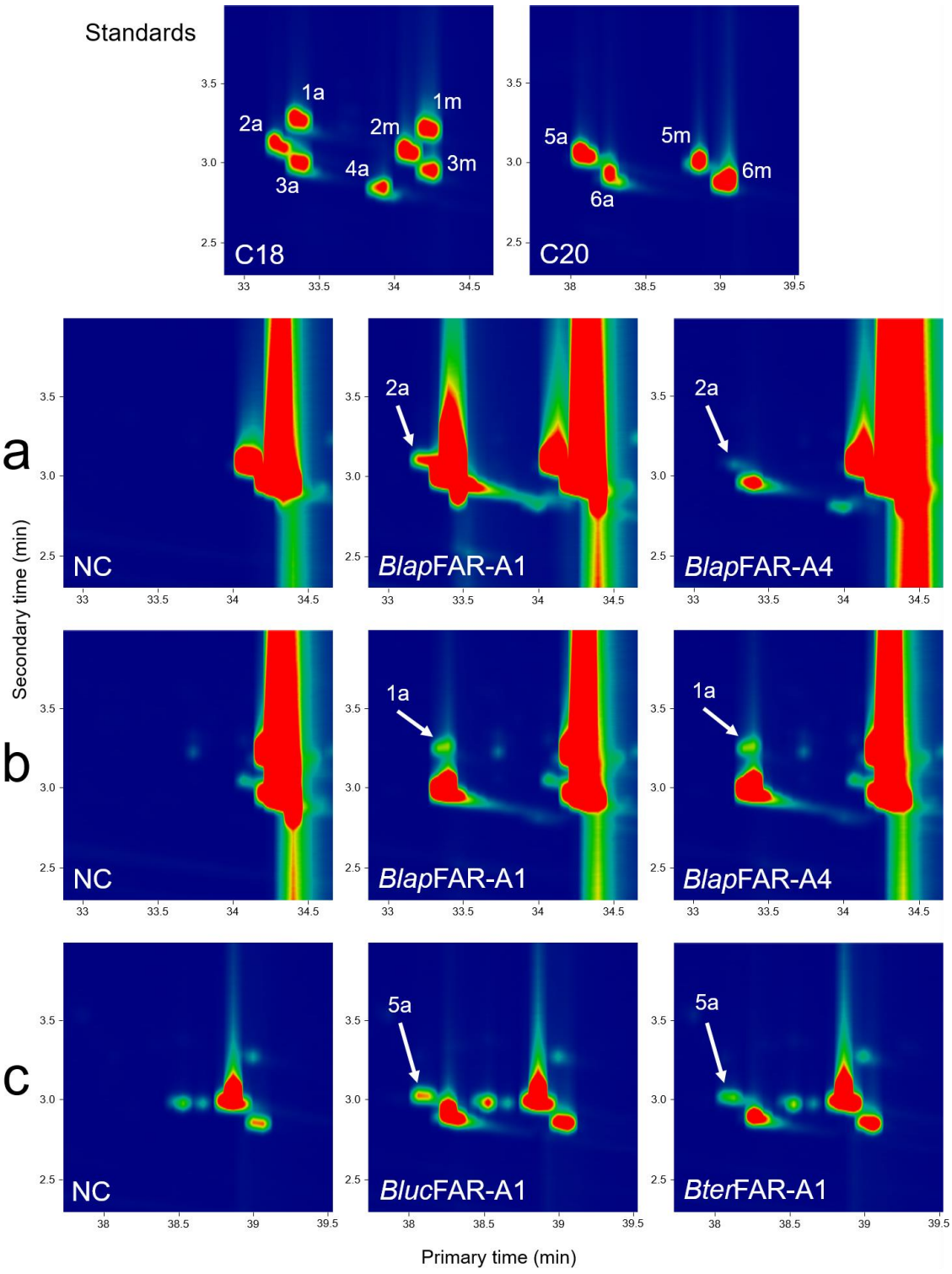
Supplementary Figure 7. Fatty alcohol production in yeast strains expressing bumblebee FARs calculated as ratios of a particular alcohol to the total produced fatty alcohols.



Supplementary Figure 8. GC-FID chromatograms of extracts from control yeast strain and strains expressing FARs specific for long chain fatty acyls (C14–C18).



Supplementary Figure 9. GC-FID chromatograms of extracts from control yeast strain and strains expressing FARs specific for very long chain fatty acyls (C20–C26).



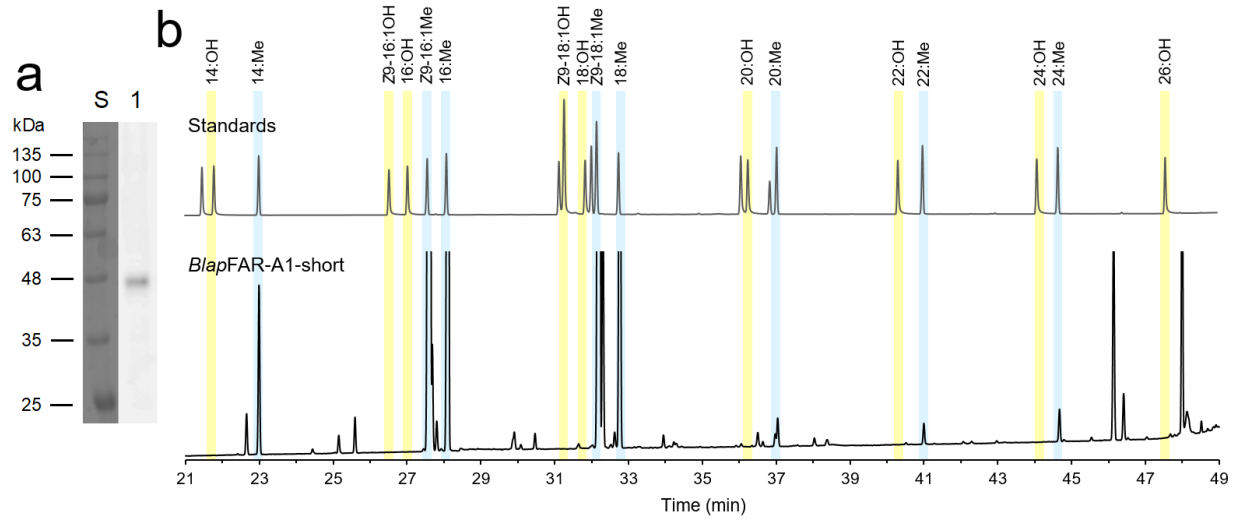
1044

1045 **Supplementary Figure 10.** GC×GC-MS chromatograms of extracts from yeast strains

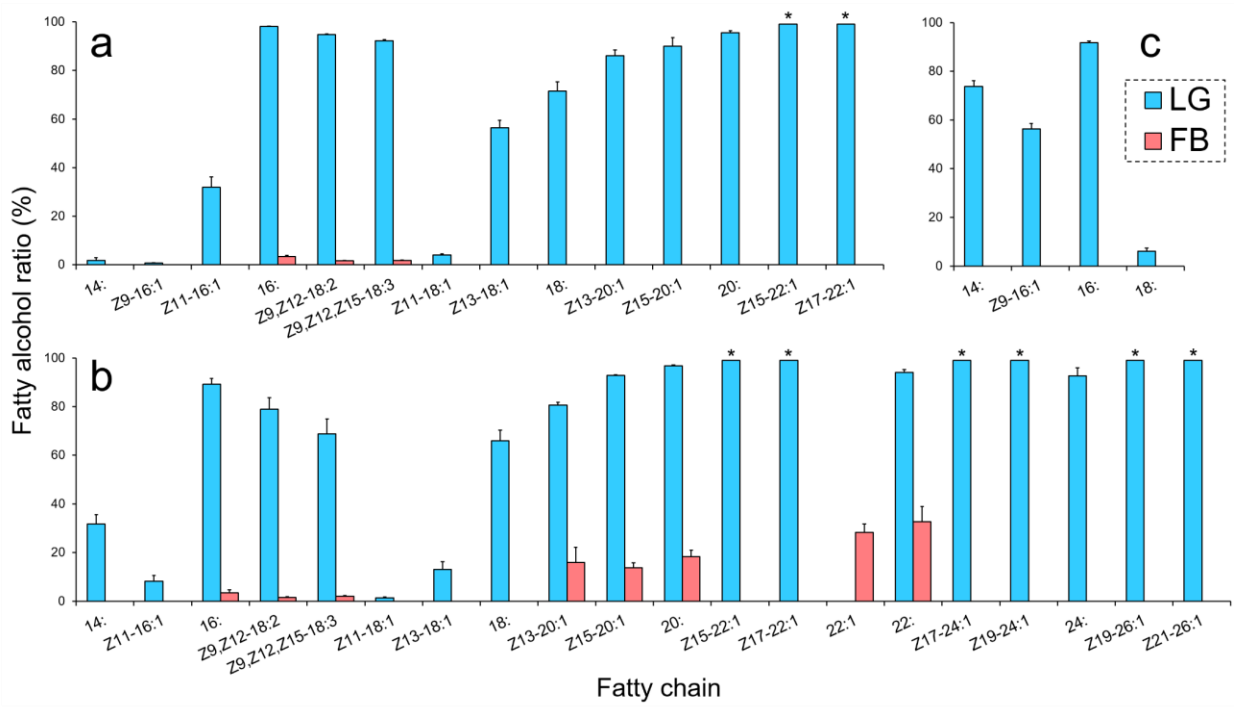
1046 supplemented with *Z*9,*Z*12-18:2 (**a**), *Z*9,*Z*12,*Z*15-18:3 (**b**) and *Z*15-20:1 (**c**) acyls. NC, negative

1047 control (yeast carrying empty vector); 1a, *Z*9,*Z*12,*Z*15-18:3OH; 1m, *Z*9,*Z*12,*Z*15-18:3Me; 2a,

*Z*9,*Z*12-18:2OH; 2m, *Z*9,*Z*12-18:2Me; 3a, *Z*9-18:1OH; 3m, *Z*9-18:1Me; 4a, 18:OH; 5a, *Z*15-20:1OH; 5m, *Z*15-20:1Me; 6a, 20:OH; 6m, 20:Me. Selected masses: (**a**) 55+67, (**b**) 55+79, (**c**) 55+74, standards 55+67+74+79.



Supplementary Figure 11. (**a**) Western blot analysis of cell lysate of yeast transformed with *Blap*FAR-A1-short. *Lanes*: S, protein standard (VI, AppliChem); 1, yeast carrying plasmid with the sequence of *Blap*FAR-A1-short. (**b**) GC-FID chromatogram of an extract from the yeast strain expressing *Blap*FAR-A1-short.



Supplementary Figure 12. Fatty alcohol ratios in LGs and FBs of 3-day-old *B. lucorum* (**a**), *B. terrestris* (**b**) and *B. lapidarius* (**c**) males. Ratios >99% are marked with asterisks. In the case of 22:1 in FB of *B. terrestris*, the assignment of individual alcohol isomers was not possible due to low amount. See Methods for a description of fatty alcohol ratio calculation.

Supplementary Table 1. Predicted protein sequence lengths and conserved domains detected in predicted FAR coding regions via Conserved Domain Database search. The presence of a domain or conserved feature is marked with "+".

Supplementary Table 2. Genomic linkage group or scaffold and orthology group of *B. terrestris* and *A. mellifera* FARs. *B. terrestris* FARs functionally characterized in this study and *A. mellifera* FAR previously functionally characterized are highlighted in green; FAR-A gene orthologs are

1080 highlighted in orange; genomic scaffolds not placed into linkage groups are grey.

Bombus terrestris		
NCBI accession number	**orthologous group**	**linkage group/scaffold**
XP_020721700.1	FAR-A	LG B12
XP_020723892.1	FAR-A	GroupUn1136
XP_020723935.1	FAR-A	GroupUn1199
XP_003403375.2	FAR-A	GroupUn1674
XP_003402269.1	FAR-A	GroupUn17
XP_012175487.2	FAR-A	GroupUn3510
XP_012175505.2	FAR-A	GroupUn4066
XP_020723961.1	FAR-A	GroupUn4067
XP_012175509.2	FAR-A	GroupUn4179
XP_020723448.1	FAR-A	GroupUn442
XP_020723970.1	FAR-A	GroupUn4704
XP_003403427.2	FAR-A	GroupUn4997
XP_020723325.1	FAR-A	GroupUn51
XP_012174003.2	FAR-A	GroupUn56
XP_020723511.1	FAR-A	GroupUn610
XP_003402802.1	FAR-A	GroupUn648
XP_012174533.1	FAR-A	GroupUn679
XP_020723568.1	FAR-A	GroupUn679
XP_003403231.1	FAR-A	GroupUn989
XP_012175211.1	FAR-A	GroupUn989
XP_012175212.1	FAR-A	GroupUn989
XP_012175213.2	FAR-A	GroupUn989
XP_020723847.1	FAR-A	GroupUn989
XP_020723849.1	FAR-A	GroupUn989
XP_003394733.1	FAR-A	LG B04
XP_020720557.1	FAR-B	LG B01
XP_012174769.1	FAR-C	GroupUn732
XP_003402940.2	FAR-D	GroupUn732
XP_003399880.1	FAR-E	LG B12
XP_003399881.1	FAR-F	LG B12
XP_020721716.1	FAR-G	LG B12
XP_020721715.1	FAR-H	LG B12
XP_003399944.1	FAR-I	LG B12
XP_020720279.1	FAR-J	LG B01
XP_012173522.1	FAR-K	LG B17

Apis mellifera		
NCBI accession number	**orthologous group**	**linkage group/scaffold**
XP_006565270.2	FAR-A	LG12
NP_001314876.1	FAR-B	LG1
XP_006570570.1	FAR-C	LG6
XP_006570601.2	FAR-D	LG6
ADI87409.1	FAR-E	LG12
ADI87411.1	FAR-F	LG12
NP_001180219.1	FAR-G	LG12
NP_001229455.1	FAR-H	LG12
XP_006566994.2	FAR-I	LG12
NP_001177849.1	FAR-J	Un
XP_001120449.2	FAR-K	LG5

1081

Supplementary Table 3. Primers used in this study. Restriction sites are underlined.

Name	Sequence 5'→3'	Origin
BlapFAR-A5_RACE_5end-R	CGCACTATCCCAACTGGCAG	
BlapFAR-A5_RACE_3end-F	GCATCAAAGCAAAACACCAGTCAG	
BterBlucFAR-A1_SphI-F	AAAAAAGCATGCATGAATACGGAATTTACAG	
BterBlucFAR-A1_NotI-R	AAAAAAGCGGCCGCTTATTAGTACATAACA	
BlucFAR-A1-opt_BamHI-F	AAAGGATCCAACACTGAGTTCACTGAAAAG	
BlucFAR-A1-opt_ EcoRI-R	AAAGAATTCTCAATACATAACTCTCAAAATAATG	
BterFAR-A2_BamHI-F	AAAAAAGGATCCATGAATACGATCAATAGAG	
BlucFAR-A2_BamHI-F	AAAAGGATCCATGGATACGATCAATAGA	
BterBlucFAR-A2_NotI-R	AAAAGCGGCCGCTTATTAGTATACAAAATAT	
BlucFAR-A2-opt_BamHI-F	AAAGGATCCGACACCATTAAC	
BlucFAR-A2-opt_EcoRI-R	AAAGAATTCTCAATAGACGAAATAC	
BterBlucFAR-J-TOPO-F	ATGGTGGAAGTTCTGGTG	
BterBlucFAR-J-TOPO-R	TTACGCAAATTGAGGTATACAATTTC	
BterFAR-J-InFusion-F	TCACCATCACGGATCCGTGGAAGTTCTGGTGGAAC	
BterFAR-J-InFusion-R	GGAAGTTAATGAATTCTTACGCAAATTGAGGTATACA	
BlapFAR-A1_BamHI-F	CGTGGATCCAATACAAAACTTAATGAAAACGAGATAAATG	
BlapFAR-A1_NotI-R	TTTAGCGGCCGCTTACGGCGCCGCATTTATAGATTTAG	
BlapFAR-A1-syn_BamHI-F	AAGGATCCAATACAAAACTTAATG	
BlapFAR-A1-syn_NotI-R	AATCAAGCGGCCGCTTACGGC	
BlapFAR-J_SphI-F	ACAATGCATGCGTGGAAGTTCTGGTGGAA	
BlapFAR-J_NotI-R	ACAAGAATGCGGCCGCTTACGCAAATTGAGGTATACAA	
BlapFAR-A4_BamHI-F	CTTGGATCCGATACAATCAATAAAGAAAG	This study
BlapFAR-A4_EcoRI-R	CATGAATTCTCAATAAACGATACAGTATAC	
BlapFAR-A5_BamHI-F	GACAAGGATCCGATACAACCGATAAA	
BlapFAR-A5_NotI-R	TAACTGCGGCCGCTCAGTACACAAAATA	
BterBlucFAR-A1_qPCR-F	ACACGGCAATGGTCCTTCAA	
BterBlucFAR-A1_qPCR-R	AAAAATAAATTTCTTTATTCCTATGACGCA	
BterBlucFAR-A2_qPCR-F	AGCAGAGCAAATTGTAGCAAGC	
BterBlucFAR-A2_qPCR-R	ACTACATCCACTCTTCCATCTCTCC	
BterBlucBlapFAR-J_qPCR-F	TGCGAGGATGGATCGACAAC	
BterBlucBlapFAR-J_qPCR-R	TCTTTGGTAGTTGCAGATTTTCGAC	
BlapFAR-A1_qPCR-F	AAGAAAAGGTCTACACCACGAATC	
BlapFAR-A1_qPCR-R	TCGCAACAGTCAATCCTTTG	
BlapFAR1-A1-short_qPCR-F	AATCTATCAGACATCGAAGAATTGATTA	
BlapFAR1-A1-short_qPCR-R	CAGGGAACGGTTCTTTTAGC	
BlapFAR-A4_qPCR-F	AGTGACAGCATATTTCGCTCTG	
BlapFAR-A4_qPCR-R	AAACGATACAGTATACAATCAGTAGTG	
BlapFAR-A5_qPCR-F	AGTTGAAAAATGCCGTGTTGAAG	
BlapFAR-A5_qPCR-R	GTGCGTGGTGAAATATGTTACCGA	
BlapEEF1A_qPCR-F	AGAATGGACAAACCCGCGAG	
BlapEEF1A_qPCR-R	CACAAATGCCACCGCAACAG	
BlapPLA2_qPCR-F	GGTCACACCGAAACCAAATT	
BlapPLA2_qPCR-R	TCGCAACATTTCGTCATTTC	
BterBlucEEF1A_qPCR-F	AGAATGGACAAACCCGTGAG	
BterBlucEEF1A_qPCR-R	CACAAATGCTACCGCAACAG	Ref [86]
BterBlucPLA2_qPCR-F	GGTCACACCGAAACCAGATT	
BterBlucPLA2_qPCR-R	TCGCAACACTTCGTCATTTC	

Supplementary Table 4. Protein sequence identities of bumblebee male marking pheromone (MMP)-biosynthetic FAR candidates. The colors indicate % identity (red: highest, blue: lowest).

	1	2	3	4	5	6	7	8	9	10
1: BterFAR-J_XP_020720279.1	100									
2: BlucFAR-J_contig26150_9365	99.65	100								
3: BlapFAR-J_contig12250	97.52	97.23	100							
4: BlapFAR-A5_contig13763_4377	32.8	32.8	32.8	100						
5: BlapFAR-A4_contig43815...	34	34	34	78.77	100					
6: BterFAR-A2_XP_003402269.1	32.41	32.41	32.41	77.78	79.37	100				
7: BlucFAR-A2_contig55206_29748	32.8	32.8	32.8	75.79	77.98	94.84	100			
8: BlapFAR-A1_contig44207	33.9	33.9	34.11	57.75	56.69	54.56	54.99	100		
9: BterFAR-A1_XP_003394733.1	33.6	33.6	33.6	56.65	57.26	53.83	52.82	60.9	100	
10: BlucFAR-A1_contig47309	33.4	33.4	33.4	56.85	57.26	53.83	52.82	61.11	99.4	100

Supplementary Table 5. Comprehensive GC analysis of fatty alcohols and transesterifiable fatty acyls in LGs of 3-day-old *B. terrestris*, *B. lucorum* and *B. lapidarius* males. The acyls were determined as corresponding methyl esters. *nd*, not detected.

Tissue	Labial gland					
Species	*B. lapidarius*		*B. lucorum*		*B. terrestris*	
Fatty chain	Amount (µg per tissue)					
	Acyl	**Alcohol**	**Acyl**	**Alcohol**	**Acyl**	**Alcohol**
*Z*7-12:	*nd*	*nd*	0.84 ± 0.07	*nd*	*nd*	*nd*
*Z*9-12:	*nd*	*nd*	*nd*	*nd*	1.58 ± 0.08	*nd*
12:	0.61 ± 0.04	*nd*	122 ± 4	*nd*	406 ± 24	*nd*
*Z*5-14:1	*nd*	*nd*	2.39 ± 0.14	*nd*	*nd*	*nd*
*Z*7-14:1	0.19 ± 0.02	*nd*	*nd*	*nd*	*nd*	*nd*
*Z*9-14:1	2.13 ± 0.57	*nd*	1419 ± 192	*nd*	103 ± 18	*nd*
14:	1.87 ± 0.21	4.70 ± 1.00	28.7 ± 2.9	0.45 ± 0.28	47.5 ± 4.7	19.5 ± 2.3
*Z*7-16:1	*nd*	*nd*	14.2 ± 2.1	*nd*	5.59 ± 0.30	*nd*
*Z*9-16:1	2236 ± 447	2577 ± 416	72.2 ± 5.7	0.44 ± 0.09	32.6 ± 0.6	*nd*
*Z*11-16:1	15.8 ± 2.0	*nd*	3.40 ± 0.90	1.47 ± 0.56	37.7 ± 4.0	2.94 ± 0.63
16:	157 ± 28	1560 ± 216	6.53 ± 0.47	313 ± 52	34.8 ± 1.9	263 ± 46
*Z*9,*Z*12-18:2	10.6 ± 2.5	*nd*	2.50 ± 0.53	41.3 ± 8.2	13.9 ± 1.2	48.1 ± 10.0
*Z*9,*Z*12,*Z*15-18:3	38.7 ± 6.8	*nd*	18.7 ± 4.6	201 ± 42	57.7 ± 0.5	119 ± 31
*Z*9-18:1	119 ± 29	*nd*	57.0 ± 9.2	*nd*	129 ± 7	*nd*
*Z*11-18:1	50.4 ± 10.7	*nd*	6.44 ± 0.87	0.25 ± 0.07	14.0 ± 1.3	0.18 ± 0.05
*Z*13-18:1	*nd*	*nd*	3.85 ± 0.80	4.48 ± 0.45	47.5 ± 3.3	6.50 ± 1.56
18:	7.27 ± 1.51	0.41 ± 0.02	5.16 ± 0.29	11.9 ± 2.4	8.35 ± 0.98	14.7 ± 1.4
*Z*11-20:1	*nd*	*nd*	0.70 ± 0.03	*nd*	2.81 ± 0.51	*nd*
*Z*13-20:1	*nd*	*nd*	1.02 ± 0.20	5.79 ± 0.76	2.65 ± 0.40	10.1 ± 1.4
*Z*15-20:1	*nd*	*nd*	1.01 ± 0.23	8.77 ± 2.37	15.5 ± 1.3	182 ± 23
20:	0.21 ± 0.04	*nd*	0.60 ± 0.05	11.9 ± 1.4	1.31 ± 0.08	35.9 ± 6.9
*Z*15-22:1	*nd*	*nd*	*nd*	11.8 ± 2.8	*nd*	88.1 ± 9.5
*Z*17-22:1	*nd*	*nd*	*nd*	1.88 ± 0.24	*nd*	14.9 ± 1.4
22:	*nd*	*nd*	0.21 ± 0.09	*nd*	0.81 ± 0.03	12.2 ± 2.6
*Z*17-24:1	*nd*	*nd*	*nd*	*nd*	*nd*	4.08 ± 0.37
*Z*19-24:1	*nd*	*nd*	*nd*	*nd*	*nd*	6.46 ± 1.42
24:	*nd*	*nd*	*nd*	*nd*	0.38 ± 0.10	4.79 ± 1.14
*Z*19-26:1	*nd*	*nd*	*nd*	*nd*	*nd*	5.44 ± 1.17
*Z*21-26:1	*nd*	*nd*	*nd*	*nd*	*nd*	7.51 ± 2.29
26:	*nd*	*nd*	*nd*	*nd*	0.90 ± 0.30	*nd*
28:	*nd*	*nd*	*nd*	*nd*	0.12 ± 0.03	*nd*

Supplementary Table 6. Comprehensive GC analysis of fatty alcohols and transesterifiable fatty acyls in FBs of 3-day-old *B. terrestris*, *B. lucorum* and *B. lapidarius* males. The acyls were determined as corresponding methyl esters. *nd*, not detected.

Tissue	Fat body					
Species	*B. lapidarius*		*B. lucorum*		*B. terrestris*	
Fatty chain	Amount (µg per tissue)					
	Acyl	**Alcohol**	**Acyl**	**Alcohol**	**Acyl**	**Alcohol**
12:	76.3 ± 17.0	*nd*	36.3 ± 10.6	*nd*	120 ± 34	*nd*
*Z*5-14:1	*nd*	*nd*	2.60 ± 0.58	*nd*	0.84 ± 0.21	*nd*
*Z*7-14:1	5.02 ± 0.92	*nd*	1.24 ± 0.36	*nd*	0.60 ± 0.22	*nd*
*Z*9-14:1	20.4 ± 5.2	*nd*	58.1 ± 16.1	*nd*	106 ± 21	*nd*
14:	128 ± 30	*nd*	158 ± 45	*nd*	456 ± 105	*nd*
*Z*7-16:1	*nd*	*nd*	25.6 ± 7.3	*nd*	27.4 ± 5.7	*nd*
*Z*9-16:1	817 ± 60	*nd*	67.5 ± 16.3	*nd*	134 ± 40	*nd*
*Z*11-16:1	20.2 ± 2.7	*nd*	11.9 ± 6.5	*nd*	136 ± 22	*nd*
16:	662 ± 112	*nd*	430 ± 108	13.82 ± 4.39	1333 ± 194	42.7 ± 11.8
*Z*9,*Z*12-18:2	74.0 ± 7.5	*nd*	76.3 ± 18.2	1.12 ± 0.30	303 ± 37	4.35 ± 1.33
*Z*9,*Z*12,*Z*15-18:3	245 ± 49	*nd*	597 ± 132	9.81 ± 2.32	1678 ± 435	32.2 ± 10.7
*Z*9-18:1	788 ± 111	*nd*	1217 ± 269	*nd*	2951 ± 487	*nd*
*Z*11-18:1	134 ± 8	*nd*	40.4 ± 19.3	*nd*	232 ± 30	*nd*
*Z*13-18:1	*nd*	*nd*	31.7 ± 15.7	*nd*	69.5 ± 15.5	*nd*
18:	60.8 ± 8.8	*nd*	119 ± 19	*nd*	224 ± 34	*nd*
*Z*11-20:1	*nd*	*nd*	1.33 ± 0.39	*nd*	5.29 ± 0.68	*nd*
*Z*13-20:1	*nd*	*nd*	0.66 ± 0.34	*nd*	4.40 ± 1.59	0.79 ± 0.47
*Z*15-20:1	*nd*	*nd*	0.18 ± 0.12	*nd*	4.01 ± 1.56	0.56 ± 0.11
20:	5.02 ± 0.21	*nd*	3.37 ± 1.06	*nd*	9.97 ± 2.24	2.10 ± 0.79
*Z*13-22:1	*nd*	*nd*	*nd*	*nd*	0.54 ± 0.19	*nd*
*Z*15-22:1	*nd*	*nd*	*nd*	*nd*	1.50 ± 0.49	*nd*
*Z*17-22:1	*nd*	*nd*	*nd*	*nd*	0.15 ± 0.07	*nd*
22:1 (sum)	*nd*	*nd*	*nd*	*nd*		0.81 ± 0.35
22:	*nd*	*nd*	0.17 ± 0.08	*nd*	1.64 ± 0.56	0.78 ± 0.46
24:	*nd*	*nd*	*nd*	*nd*	0.65 ± 0.35	*nd*
26:	*nd*	*nd*	*nd*	*nd*	0.92 ± 0.50	*nd*
28:	*nd*	*nd*	*nd*	*nd*	1.07 ± 0.64	*nd*

Supplementary Data 1: Synthetic nucleotide sequences. Restriction sites are underlined.

*Bluc*FAR-A1-opt (codon-optimized for yeast)

GGATCCAACACTGAGTTCACTGAAAAGTCTAACAAGGTCAACTCCATCGAAGGTTTCTACGCTGG
TACAGGTATCTTTATCACAGGTGCCTCAGGTTTTGTCGGTAAAGGTTTGTTGGAAAAGTTGATCA
GAGTTTGTCCTAGAATCGTTGTATTATTCATCTTGGTTAGACCAAAGAAACATCAAACAATGGAA
CAAAGATACAAGGAAATCATGGATGACCCTATCTTTGATGACATCAAAGCTAAGAATCCATCCGC
ATTGAAAAAGGTCCATCCTGTTGAAGGTGACATTTCCTTACCAAAGTTGGGTTTGAGTCAAGAAG
ATAGAAACATGTTGATAGAAAACGTCAACATCTTGTTTCACGTTGCTGCATCTTTGAACTTCAAG

GAACCATTGAACGCCGCTGTAAATACTAACGTCAAGGGTACATTTTCTATAATCGAATTGTGTAA
CGAATTGAAGCATGTTATATCAGCTGTACACGTCTCTACAGCATATTCAAATGCCAACTTGCCTG
AAATAGAAGAAAAGGTTTACTCCACTATCTTACAACCATCTTCAGTAATTGAAACATGCGACAGT
TTGGATAAGGAATTGATTAAGTTGTTGGAAGAAAGAATTTTGAAAATACATCCTAACACGTACAC
CTTCACGAAGAATTTGGCAGAACAAATCTTGTCCAGTTCTTCAACTAACTTCCCAATAGCAATCG
TTAGACCTTCTATCATTTCCGCCAGTTTAAAAGAACCATGTCCTGGTTGGTTGGGTAATATTACA
GCCCACATAGCTTTGGGTTTGTTTATTTCAAGAGGTTTCGCCAAGATCACCTTAGCTAACCCTGA
CACTATCACAGATACCGTACCATTAGACTATGTCGTTGATACAATTTTGTGTGCAGCCTGGCATG
TTACCTTGCACAGAGATATGAACGTTAAGGTATACAACTGCACCAATAACGCCAGACCAATTAAT
TACGGTGAATTGAAGGACACTTTTGTCAAGTACGCTATTCAAATACCTATGGATGGTTTAGTTTG
GTATCCATGTTGCGCAATGGTTTCTAACAGATACGTATACTCAATCTTGACCTTATTCTTGCATA
CTTTGCCTGCTTTTATTATGGATATTTTCTTAAGATTGCAAGGTTCAAAGCCAAGAATGATGAAG
ATCTCTAAGTACTACGATACAATGTCAATCGTTACCAACTACTTCTCCACTAGACAATGGAGTTT
CAAAAAGGATAACGTTATTAATATGATGAAAGAAGTCAAGACTTTGGAAGATTCTGACATCGTTA
GATTAGATTTGCAAGATATGGACTGGGATAAGTACATCGCTATATGCGTTATCGGTATCAAAAAG
TTTATTTTCAAAGAAGACCCAAAGTCCTTAGATGCTGCATTGAGAAGATTGAGTATCTTTTACTG
GATTCATCAAATGACTAAAGCCTTCGCTATTATTATCTTATTGACCATTATTTTGAGAGTTATGT
ATTGAGAATTC

*Bluc*FAR-A2-opt (codon-optimized for yeast)

GGATCCGACACCATTAACAGAGAAAAGAACGAAAACGCCATTAACAAGGGTTTGAACAAGTTGAA
TACATTAGAAGAATTTTACGTCGGTAGTGGTATTTTGTTAACTGGTGCAACAGGTTTTGTTGGTA
AAGCTGTTTTGGAAAAGTTGATCAGAATGTGTCCAAGAATTGCTGCAATTTTCTTGTTGTTTAGA
CCAAAGACTGATGAAACAATCGAACAAAGATTCAAGAAATTGATTGATGATCCAATCTATGATGA
TATCAAGGCAAAGCATCCATCAACTTTGTCTAGAGTTTATCCAATGAGAGGTGACTTGTCATTGC
CAGATTTGGGTTTGTCTAGAGAAGATAGAAATTTGTTGTTGGAAAAGGTTAACATCGTTTTCCAT
GCTGCAGCTACTGTTATGTTTAATGAACCATTGCAAGTTACAATTAATGTTAACACTAAAGGTAC
AGCTAGAGTTATTGATTTGTGGAACGAATTGAAGCATCCAATCTCATTCGTTCATGTTTCAACTG
CATTTTCTAACGCTAACATCCATGAAATCGGTGAAAGAGTTTACACTACATCATTGAAACCATCT
GAAGTTATTGATATTTGTAATAAGTTTGATAAAACATCTATTAATCAAATCGAAAAGAAAATTTT
GAAAACTTATCCAAATATCTATACTTTTTCTAAAAATTTGGCAGAACAAATCGTTGCTTCTAACT
GTAAGGATATGCCAGTTGCTATTGTTAGACCATCAGTTATTGGTCCATCTATGGAAGAACCATGT
CCAGGTTGGATTCAAAACATCTCTGCAATCACTGGTATCATGGTTTTGATCGGTAGAGGTTGTGC
TACAGGTATTAGAGGTAGAAGAGATGGTAGAGTTGATGTTGTTCCATTGGATTACGTTGTTGATA
TGATCATCTGTACTGCATGGCATGTTACATTGCATCCAAAGCATGAAGTTAAGGTTTACAACTGT
ACATCTTCAGCTAACCCAATTAGATGGGGTCAAATGCAACAATTGGTTTTGAAGCATTCAAGAGA
AACTCCATTAAACGATACATTGTGGTATCCAAGATGTCCAATCATCGCAAATAAGTACATTTTCA
ATGTTTTATGTGTTATCCCATACGTTTTGCCAGCTTTTATTATCGATATTTTCTTGAGATTGAGA
GGTTCTAAGCCAATCATGATGAAGTTGTTGAAGTTCGGTTACAAGTTGTCTACTTCAGTTTCTCA
TTTCACTATGAACGAATGGACATTCCAAAGAGATAACTGTTCAGATTTGGCATCTAAGGTTAAGA
TGTTGCATGATTCAGATATGGTTAAATTAGATTTGAGAGATATGAAGTGGGAAAAGTACATCGTT
ATATATTTGATGGGTATTAGAAAGTTTATTTTGAAACAAGAATTTCAACCAACAGCTAGACAAAG
ATTGTCTAGATTGTACTGGATTCATCAAATCACTAAGATTTCAGGTATCATCAGTTTGTTATGGA
TTATTTTGTATTTCGTCTATTGAGAATTC

*Blap*FAR-A1

GGATCCAATACAAAACTTAATGAAAACGAGATAAATGAAAAATTACGTAATGTGAATTCCATTGG

GGGATTCTACGCCGGAACTGGAATTCTTATTACTGGTGGGACAGGTTTCGTGGGCAAAGGACTCC
TCGAAAAACTGATACGCACGTGTTCACACATCGCTGCTATTTTTATATTGATCCGTCCGAAACGT
AACCAAACGATAGAACAACGATTTAAGAAGATAATAGATGATCCGATTTTCGATGGTGTCAGAGC
ACAAAACCCAGCAATTTTCTATAAAATTCATCTCGTGGAGGGCGACGTGACTCTACCAGATTTAG
GTCTTTTGCAAAAAGACAGAGATATGTTGATAGAGAATGTAAACATAGTGTTCCACATTGCGGCC
ACTATAAATTTCCATCAACCATTGGATATGATTGTCAATGTAAATGTGAAAGGTACCGCTAATAT
TATCAAACTGTGCAAGGAACTCAAGCATGTAATTAGCGTTGTCTATGTGAGCACAGCTTACAGTA
ATCCGAATCTATCAGACATCGAAGAAAAGGTCTACACCACGAATCTAGATCCCTCTCTCGTGATG
GATATATGCGACCGACAAGACAAAGAATTGATTAATCTGATCGAAGAAAGAATTTTAAAAACGCA
TCCGAACACATACACGTTCACCAAGAATCTTGCAGAGCAGACAATATCCAACAATAGCAAAGGAT
TGACTGTTGCGATAGTGCGACCAAGTATAATTTCTTCCTCGCTAAAAGAACCGTTCCCTGGTTGG
TTGGTATCTTTTGCTGGACAATCAGGTATCTTCAAGAATATCGGCAATGGTATGGCAAAAGTACT
ATTGGGTAGGGGAGATGTAATATCAGATATAGTGCCTGTTGATTATGTAGTCGACGCGATAATGT
GTGCCGCGTGGCACGTCACGCTACAAATTGATAATAATGTCAAAGTTTACAACTGTACGAGCAGC
GCACGTCCCATCAAATTGGGTGAAATCGTAAATATCTTCGTGGAATGTAGCAGAGAAATACCGAT
GAAAAATACGTTGTGGTATCCGAGTTGTACGATAGTAGCAAACAGATTTGTTTACAATGTACTGA
ATATACTTCTAAATGTTTTACCTGCGTTTGCCGTGGATATCTTTTTAAGGCTTCGAGGTGGTAAA
CCAATGGCAATGAATATGAACAAATATTACAATAAATTGGTCGTAGCGACAAGCTACTTCAACTC
GAATGAATGGTCCTTCAAAAGAGATAACATTGCCGATATGATAAACAAGGTGAATACCTTGGAAG
ATGGAAATATTGTTAAACTGGACTTGCAGGATATGGTTTGGAGGAAATATATAGCAAATTACTTG
GCGGGAATTAAGAAATTTATTCTGAAAGAAGACCCTAAATCTATAAATGCGGCGCCGTAAGCGGC
CGC