Evolution of the Codling Moth Pheromone Through the Member of an Ancient Desaturase Expansion

Jean-Marc Lassance*[,a,b], Bao-Jian Ding*[,a], Christer Löfstedt[a,1]

[*] These authors contributed equally to this work
[a] Department of Biology, Lund University, Sölvegatan 37, SE-223 62 Lund, Sweden
[b] Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA, 02138, USA
[1] Correspondence: christer.lofstedt@biol.lu.se

Abstract

Defining the origin of genetic novelty is central to our understanding of the evolution of novel traits. Diversification among fatty acid desaturase (FAD) genes has played a fundamental role in the introduction of structural variation in fatty acyl derivatives. Because of its central role in generating diversity in insect semiochemicals, the FAD gene family has become a model to study how gene family expansions can contribute to the evolution of lineage-specific innovations. Here we used the codling moth (*Cydia pomonella*) as a study system to decipher the proximate mechanism underlying the production of the Δ8Δ10 signature structure of Olethreutine moths. Biosynthesis of the codling moth sex pheromone, (*E*8,*E*10)-dodecadienol (codlemone), involves two consecutive desaturation steps, the first of which is unusual in that it generates an *E*9 unsaturation. The second step is also atypical: it generates a conjugated diene system from the *E*9 monoene C_{12} intermediate via 1,4-desaturation. Here we describe the characterization of the FAD gene acting in codlemone biosynthesis. We identify 27 FAD genes corresponding to the various functional classes identified in Insects and Lepidoptera. These genes are distributed across the *C. pomonella* genome in tandem arrays or isolated genes, indicating that the FAD repertoire consists of both ancient and recent duplications and expansions. Using transcriptomics, we show large divergence in expression domains: some genes appear ubiquitously expressed across tissue and developmental stages; others appear more restricted in their expression pattern. Functional assays using heterologous expression systems reveal that one gene, Cpo_CPRQ, which is prominently and exclusively expressed in the female pheromone gland, encodes an FAD that possesses both *E*9 and Δ8Δ10-desaturation activities. Phylogenetically, Cpo_CPRQ clusters within the Lepidoptera-specific Δ10/Δ11 clade of FADs, a classic reservoir of unusual desaturase activities in moths. Our integrative approach shows that the evolution of the signature pheromone structure of Olethreutine moths relied on a gene belonging to an ancient gene expansion. Members of other expanded FAD subfamilies do not appear to play a role in chemical communication. This advises for caution when postulating the consequences of lineage-specific expansions based on genomics alone.

Keywords: fatty acyl desaturase/gene family evolution/bifunctional/conjugated double bond/Tortricidae

Introduction

Establishing the origin of genetic novelty and innovation is central to our understanding of the evolution of novel traits. While genes can evolve *de novo* in the genome, the most common mechanism involves duplications of existing genes and the subsequent evolution of novel properties harbored by the encoded gene products. Expansions provide opportunities for specialization and evolution of novel biological functions within a lineage, including the breadth of expression via modifications of promoter architecture (Lespinet et al. 2002). The size of gene families is influenced by both stochastic processes and selection, and particularly large differences in genetic makeup can be indicative of lineage-specific adaptation and potentially associate with traits contributing to phenotypical differentiation between groups (Hahn et al. 2005, 2007). However, the precise functional consequences of such amplifications remain frequently unclear, even if expansions show readily discernible patterns. Therefore, deciphering the underlying genetic and molecular architecture of new phenotypic characters is necessary for a complete understanding of the role played by the accumulation of genetic variation through gene duplication.

For organisms relying on chemical communication, evolution of the ability to produce and detect a new type of molecule could allow for the expansion of the breadth of available communication channels and provide a medium with no or limited interference from other broadcasters. Since the identification of bombykol by Butenandt and co-workers in 1959 (Butenandt et al. 1959), a countless number of studies have contributed to revealing the diversity of fatty acid derivatives that play a pivotal role in the chemical communication of insects. As a consequence of homologies with well-studied metabolic pathways, our understanding of the molecular basis of pheromone biosynthesis from fatty acid intermediates has greatly advanced over the past two decades, highlighting the role of several multigene families (Löfstedt et al. 2016). Insect semiochemicals are synthesized in specialized cells in which fatty acyl intermediates are converted in a stepwise fashion by a combination of desaturation, chain-shortening and chain-elongation reactions followed by modifications of the carbonyl group, to cite a few possible steps (Tillman et al. 1999; Löfstedt et al. 2016 and references therein). Desaturation appears particularly important and contributes to producing the great diversity of structures observed in insect pheromones. This derives from the properties of the enzymes that are central to many uncanonical fatty acid synthesis pathways seen in insects. These enzymes can exhibit diverse substrate preference, introduce desaturation in either or both *cis* (*Z*) and *trans* (*E*) geometry, and give rise to variation in chain-length double bond position, number and configuration.

In insects, the group of proteins responsible for catalyzing desaturation reactions are fatty acyl-CoA desaturases (FADs). These membrane-bound acyl-lipid desaturases are biochemically and

structurally homologous to the desaturases ubiquitously found in animals, yeast, fungi and many bacteria where they play important basic biological functions in lipid metabolism and cell signaling and contribute to membrane fluidity in response to temperature fluctuation (Sperling et al. 2003). In the past decades, the integration of molecular and phylogenetic approaches has greatly advanced our understanding of the function of FAD genes in the biosynthesis of mono- and poly-unsaturated fatty acids in a range of organisms. Moreover, an ever-increasing number of acyl-CoA desaturase genes have been functionally characterized, demonstrating mechanistically their crucial role in the biosynthesis of pheromone and semiochemicals in *Drosophila* fruitflies, bees, wasps, beetles, and lepidopteran species. Variation in the number and expression of acyl-CoA desaturase genes have been shown to affect the diversity of pheromone signals between closely related species (Takahashi et al. 2001; Roelofs et al. 2002; Fang et al. 2009; Shirangi et al. 2009; Albre et al. 2012). The family is characterized by multiple episodes of expansion and contraction that occurred during the evolution of insects (Roelofs and Rooney 2003; Helmkampf et al. 2015). Consequently, the FAD gene family has become a model to study how structural and regulatory changes act in concert to produce new phenotypes.

Among the taxa available to study the molecular basis of pheromone production in an evolutionary framework leafroller moths (Lepidoptera: Tortricidae) provide a model system of choice (Roe et al. 2009). Leafrollers represent one of the largest families in the Lepidoptera with over 10,000 described species (Gilligan et al. 2018). Their larvae feed as leaf rollers, leaf webbers, leaf miners or borers in plant stems, roots, fruits or seeds. Many tortricid species are important pests and, due to their economic impact on human society, became the target of countless pheromone identification studies. In 1982, Roelofs and Brown published a comprehensive review on pheromones and evolutionary relationships among Tortricidae (Roelofs and Brown 1982). These authors related the patterns in pheromone diversity in a biosynthetic perspective to different postulated phylogenies of the Tortricidae, and specifically the two major group within Tortricidae, Tortricinae and Olethreutinae. Based on the pheromone identifications available at the time, it was suggested that species in the Tortricinae use mostly 14-carbon pheromone components (acetates, alcohols and aldehydes) whereas species in the Olethreutinae subfamily use mostly 12-carbon compounds. Building on recent advances towards a robust molecular phylogeny of Tortricidae (Regier et al. 2012; Fagua et al. 2017) and incorporating the information for the pheromones identified in 179 species and sex attractants reported for an additional 357 species we show that the proposed dichotomy is well-supported by the data currently available (Fig. 1). Furthermore, a majority of species in Tortricinae use pheromone components with double bonds in uneven positions, i.e., Δ9 isomers or Δ11 isomers. By contrast, the Olethreutinae pheromones typically contain components with double bonds in even positions (Δ8, Δ10), with doubly unsaturated Δ8Δ10:12C fatty acyl chains being one of the signature structures of the subfamily. Roelofs and Brown (1982) suggested that the use of a Lepidoptera-specific Δ11-desaturase acting on

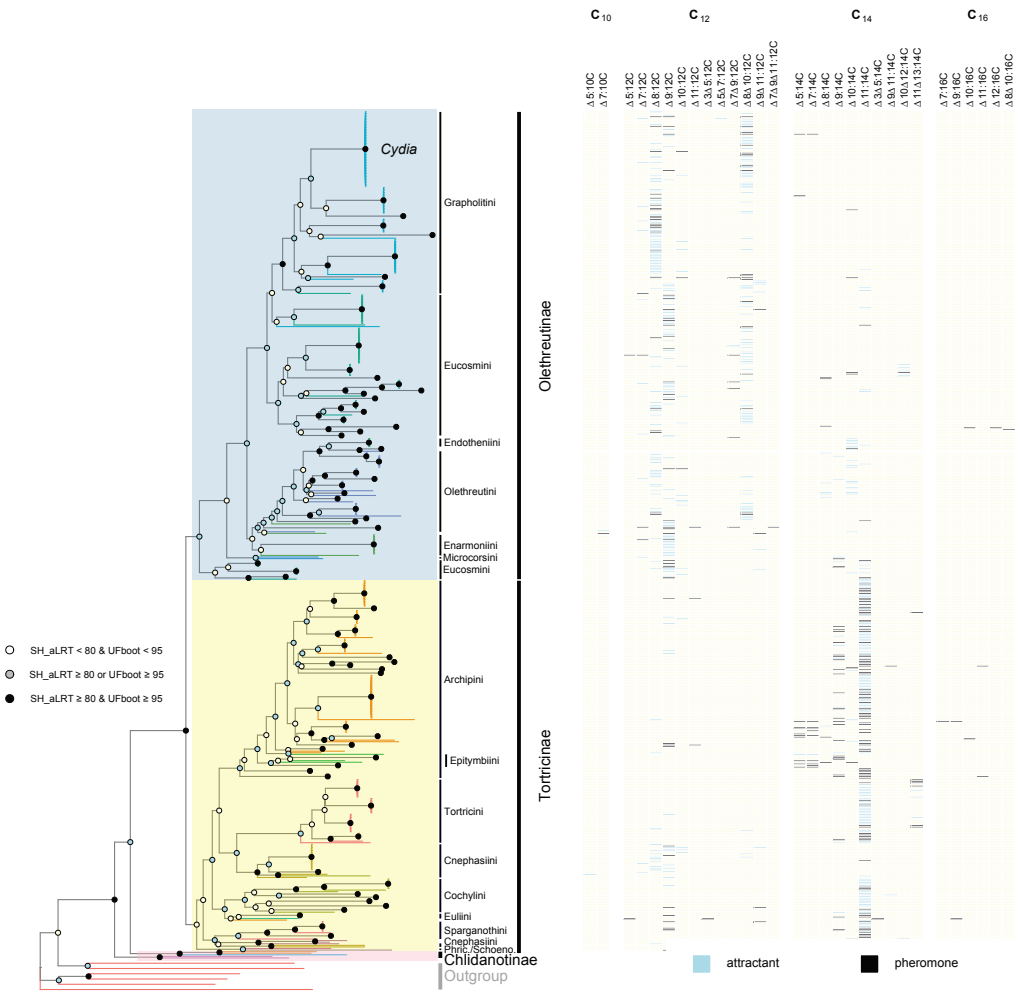


Figure 1: Phylogeny of Tortricidae and their associated female sex pheromone components. (left) The maximum likelihood tree was obtained for predicted nucleotide sequences of Tortricidae species (7591 aligned positions). The species represented comprise all tortricids for which pheromones or attractants have been reported plus some outgroups. Typically, one representative species was chosen per genus and contributed molecular evidence for all species in the genus. Outgroup species are represented by red branches whereas tortricid species from the same tribe are represented by branches of the same colour. Branch support values were calculated from 1000 replicates using the Shimodaira-Hasegawa-like approximate ratio test (SH_aLRT) and ultrafast bootstraping (UFboot). Support values for branches are indicated by colored circles, with color assigned based on thresholds of branch selection for SH-aLRT (80%) and UFBoot (95%) supports, respectively. The major subfamilies and represented tribe names are indicated (Phric: Phricanthini; Schoeno: Schoenotenini). (right) Heatmap indicating presence/absence of unsaturated fatty acid structure in bioactive molecules. Attractants correspond to compounds found to be attractive in either field or laboratory experiments; pheromone components correspond to sex attractants produced naturally by the organism and with a demonstrated biological activity on conspecific males. Double bond positions are annotated in Δ-nomenclature without referring to the geometry. Molecular and trait data retrieved from GenBank and the Pherobase, respectively.

myristic acid (C_{14}) could account biosynthetically for most of the pheromone compounds found in Tortricinae. This hypothesis was later confirmed with the functional characterization of FADs expressed in the female pheromone gland of several Tortricinae representatives. These include a desaturase that makes only the *E*11-isomer in the light brown apple moth *Epiphyas postvittana* (E11-14, E11-16 and E9E11-14)(Liu et al. 2002a)**,** as well as desaturases from the redbanded leafroller moth *Agryrotaenia velutinana* (Liu et al. 2002b) and the obliquebanded leafroller moth *Choristoneura rosaceana* (Hao et al. 2002a) which both produce a mixture of Z/E11-14:Acids. In the case of Olethreutinae*,* Δ11-desaturation followed by chain-shortening could account for the Δ9:12C compounds found in several tribes of Olethreutinae. On the other hand, the biochemical pathways leading to the pheromone components with double bonds in even positions, the Δ8 and Δ10 as well as the doubly unsaturated Δ8Δ10:12C compounds in the Olethreutinae was not obvious.

The codling moth *Cydia pomonella* (Linnaeus) (Tortricidae: Olethreutinae: Grapholitini) is one of the most devastating pests in apple and pear orchards worldwide (Witzgall et al. 2008). With the goal of disrupting its reproduction, it has received substantial attention and been at the center of numerous studies focused on characterizing its communication system via sex pheromones. Its pheromone, (*E*8,*E*10)-dodecadien-1-ol, also known under the common name *codlemone*, was first identified using gas chromatography in combination with electroantennogram (EAG) recordings (Roelofs et al. 1971). This identification was later confirmed by fine chemical analysis (Beroza et al. 1974; McDonough and Moffitt 1974). The codling moth thus provides a relevant system to unravel the molecular pathway associated with the production of the Δ8Δ10:12C typical of Olethreutines. Previous studies support the hypothesis that the biosynthesis involves the desaturation of a Δ9 monoene intermediate. First, Arn et al. (1985) reported the presence of the unusual (*E*)-9-dodecenol (E9-12:OH) at about 10% of the doubly unsaturated alcohol in pheromone gland extracts and effluvia from *C. pomonella* females. Although this monoene does not carry any behavioral activity, its occurrence suggested that E9-12:Acyl could be an intermediate of codlemone biosynthesis in a process analogous to the biosynthesis of 10,12-dienic systems via Δ11 monoene intermediates observed in *Bombyx mori*, *Manduca sexta*, and *Spodoptera littoralis* (Yamaoka et al. 1984; Tumlinson et al. 1996; Navarro et al. 1997). The selective incorporation of deuterium-labeled E9-12 fatty acid precursors into codlemone and its direct precursor, *E*8,*E*10-dodecadienoate (E8E10-12:Acyl), supported the presence of an unusual Δ9 desaturase in *C. pomonella* and the biosynthesis of a conjugated diene system via 1,4-desaturation and the characteristic elimination of two hydrogen atoms at the allylic position of the double bond in the monoene intermediate precursor (Löfstedt and Bengtsson 1988)(Fig. 2). Similar conclusions were reached from a replicate study using *C. splendana* and *C. nigricana* females (Witzgall et al. 1996). To date, the FAD(s) central to the biosynthesis of codlemone and related fatty acyl derivates with a Δ8Δ10 system has not been characterized.

The availability of a high-quality draft genome for *C. pomonella* provides an opportunity to comprehensively annotate and analyze relevant genes (Wan et al. 2019). Here we annotated a total of 27 FAD genes and performed a phylogenetic analysis, revealing expansions of different ages in the Δ9(KPSE) clade and in the Δ10/Δ11(XXXQ/E) clade. Using a transcriptomic approach, we determine the breadth of expression of all FAD genes and identified genes with upregulated expression in the pheromone gland of female *C. pomonella*, where the biosynthesis of codlemone takes place. We tested the function of these desaturases in heterologous expression systems and identified one FAD gene conferring on the cells the dual desaturase functions playing a key role in the biosynthesis of (*E8*,*E*10)-dodecadienol in *C. pomonella* and a signature structure of many Olethreutinae moth pheromones.

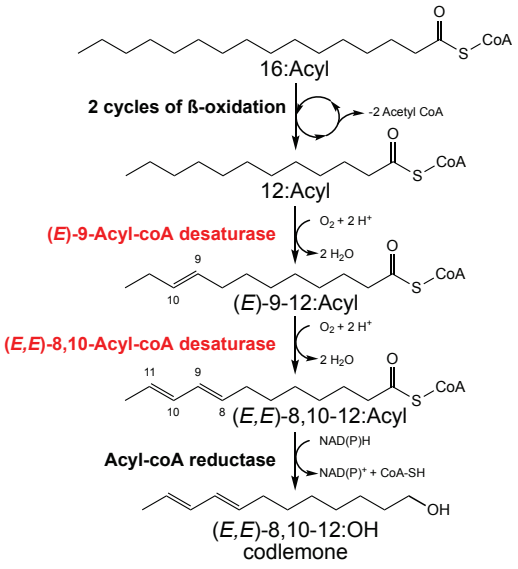


Figure 2: Pathway of codlemone biosynthesis.
Palmitic acyl (C_{16}) is first formed by the fatty acid synthesis pathway. Lauric acyl (C_{12}) is then produced following two recurring reactions of beta-oxidation. A first desaturation occurs, introducing an *E*9 double bond between carbon 9 and 10 from the carboxyl terminus. The resulting monoenoic fatty acyl is then the substrate of a second desaturation via the removal of hydrogen atoms at the carbons 8 and 11, causing the formation of an *E8E10* conjugated diene system. Finally, an unknown fatty acyl-CoA reductase converts the fatty acyl precursor into the fatty alcohol (*E*8,*E*10)-dodecadienol commonly known as codlemone. The steps characterized in this study are highlighted in red.

Results

Expansion of FADs in the Genome of C. pomonella

We identified candidates potentially involved in fatty acid synthesis by searching in the genome of *C. pomonella* for genes encoding fatty acid desaturases, which are characterized by a fatty-acid desaturase type 1 domain (PFAM domain PF00487). First, we improved the genome annotation by incorporating data from the pheromone gland transcriptome, a tissue which was not part of the panel of tissues used to generate the available annotation of *C. pomonella*. Searching our improved annotation, we identified 27 genes harboring the signature domain of FADs. While the vast majority of genes identified in our exhaustive search contain open-reading frames of 300 amino acids or longer, a small number of genes (i.e. 2-3) may represent pseudogenes or assembly errors. We adopted the nomenclature proposed by Knipple et al. (2002) in which genes are named based on the composition of 4 amino acid residues at a signature motif. Next, we looked at their genomic organization. We found that FAD genes are distributed across 12 of the 27 autosomes, with no FAD genes on either Z or W sex chromosome (chr1 and chr29, respectively) (Fig. 3). Two genes, Cpo_QPVE and Cpo_MATD(2), were found on unplaced scaffolds. As is typical for members of a multigene family, we identify several clusters corresponding to tandemly duplicated genes. Moreover, several genes are present as the single member of the family on a given autosome.

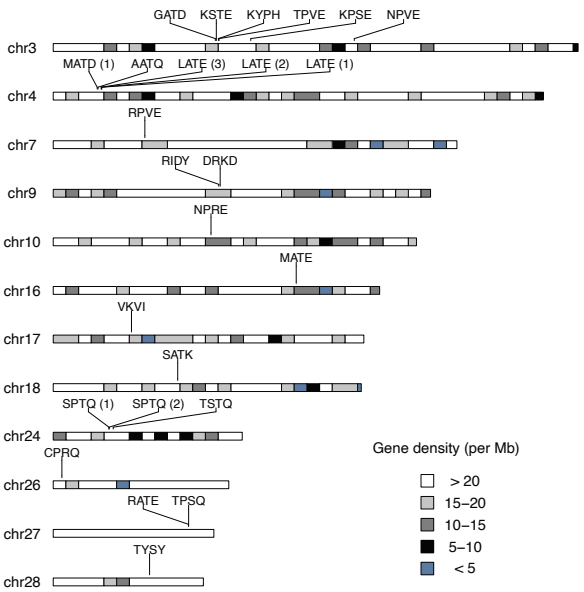


Figure 3: Genomic organization of FAD genes in *Cydia pomonella*. The position of the 27 predicted FAD genes were mapped to the genome. 25 genes could be placed on 12 autosomes; 2 genes, Cpo_MATD(2) and Cpo_QPVE, are located on unplaced scaffolds (not drawn). The distribution analysis showed that there are 5 gene clusters which contain 2 or more FAD genes. Names refer to chromosome names in the genome assembly. Chromosomes are drawn to scale and the bands represent gene density.

Phylogenetic Analyses Place Expansions in the Δ9(KPSE) and Δ10/Δ11(XXXQ/E) Clades

In order to determine which of the genes identified above could be important for codlemone biosynthesis, we assessed the functional classes represented by these genes using phylogenetic analyses. To that end, we aligned the predicted protein sequences of the *C. pomonella* genes with a panel of FAD sequences from Lepidoptera species including genes for which function have been previously characterized using *in vitro* heterologous expression. We found representatives of the eight insect acyl-CoA desaturase subfamilies *sensu* Helmkampf et al. (Helmkampf et al. 2015) (Fig.4, Fig. S1).

All subfamilies form highly supported groups, with the exception of the relationship between Desat A1 and A2, which appear weakly or strongly supported, depending on the set of genes used in our analyses. With a characteristic domain architecture (PF08557 in front of PF00487), Cpo_TYSY encodes a putative Sphingolipid Delta-4 desaturase and is homologous to *interfertile crescent* in *D. melanogaster* (Ifc). Two tandemly duplicated genes, Cpo_RIDY and Cpo_DRKD, contain a Cytochrome b5-like heme binding domain (PF00173) in front of the fatty acid desaturase domain and group with putative Front-End desaturases homologous to *Cytochrome b5-related* in *D. melanogaster* (Cyt-b5-r). These three genes bear little similarity with the other 24 FAD genes which encode First Desaturases of the subfamilies Desat A1 through E.

Desat C, D, and E are present as single-copy genes in *C. pomonella* (Cpo_QPVE, Cpo_KSTE, and Cpo_KYPH, respectively). Cpo_QPVE groups with the Δ14 desaturase identified in male and female corn borer pheromone biosynthesis (Lepidoptera: Crambidae) (Roelofs et al. 2002; Lassance and Löfstedt 2009).

Desat B, which is particularly expanded in Hymenopterans and in *Bombyx mori* (Helmkampf et al. 2015), has only two representatives in *C. pomonella*. These genes, Cpo_VKVI and Cpo_TPVE, are homologous to the *E*6 desaturase identified in *Antheraea pernyi* (Lepidoptera: Saturnidae) (Wang et al. 2010) and the *Z*5 desaturase identified in *Ctenopseustis obliquana* (Lepidoptera: Tortricidae) (Hagström et al. 2014), respectively.

Desat A2 subfamily is characterized by a single-copy gene, Cpo_GATD, which seem to be typical of Lepidoptera. The homolog identified in *Choristoneura parallela* (Lepidoptera: Tortricidae) is associated with the formation of Δ9 desaturation in saturated acyl moieties of a range of length (14C-26C) (Liu et al. 2004).

Finally, Desat A1 forms the largest group and experienced a particularly dynamic evolutionary history. With 17 genes *C. pomonella* harbors twice as many genes as *Bombyx mori* (7 in the most recent version of the genome SilkDB 3.0 (Lu et al. 2020)). While Δ9 C16<C18 (NPVE) is represented by a single-copy gene, we found two significant expansions in the group corresponding to Δ9 C16>C18 (KPSE) and Δ11, Δ10 & bifunctional (XXXQ/E) FADs. The tandemly arrayed cluster found in chr4 correspond to an extension within the A1 subfamily encoding Δ9 C16>C18 FADs. Nine genes are found in the Lepidoptera-specific A1 subfamily encoding Δ11, Δ10 & bifunctional FADs. These genes are scattered across 6 private chromosomes, none of which containing members of the other functional classes.

Based on their placement in the FAD gene tree, these genes can be sub-divided into four groups (Fig.4, Fig. S1). Group1 contains Cpo_SPTQ(1), Cpo_SPTQ(2), Cpo_SATK and Cpo_TSTQ, which are homologous to the Δ10 desaturase identified in the New Zealand tortricid *Planotortrix octo* (Lepidoptera: Tortricidae) (Hao et al. 2002). Group2 is formed with Cpo_RATE, Cpo_MATE, and Cpo_TPSQ, which are homologous to the FAD with Δ6 activity isolated from *Ctenopseutis herana* (Lepidoptera: Tortricidae) (Albre et al. 2012). Group3 is represented by Cpo_RPVE, which groups with desat2 from *Ctenopseutis obliquana* and allied species (Lepidoptera: Tortricidae) (Albre et al. 2012) and Lca_KPVQ from *Lampronia capitella* (Lepidoptera: Prodoxidae) (Liénard et al. 2008). Finally, Cpo_CPRQ falls near Dpu_LPAE from *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae) (Liénard et al. 2010).Based on our phylogenetic reconstructions, several genes are plausible candidates in the biosynthesis of codlemone. First, genes from the Δ9 C16>C18 (KPSE) subfamily have been implicated in pheromone biosynthesis and carry activities similar to those expected in *C. pomonella*. Specifically, the Dpu_KPSE from *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae) produces a range of Δ9-monounsaturated products including Z9- and E9-12:Acyl (Liénard et al. 2010). This exemplifies that enzymes in this subfamily can introduce *cis* and *trans* double bonds in lauric acid (12C). Furthermore, when supplemented with Z7- and E7-14:Acyl, Dpu_KPSE can introduce a second desaturation to produce Δ7Δ9-14:Acyl, illustrating the formation of conjugated double bonds by this type of FAD. The expansion of the KPSE clade we report in *C. pomonella* is an interesting coincidence. Second, the abundance of representatives in the Δ11, Δ10 & bifunctional (XXXQ/E) subfamily support the possibility that one or several could be involved in codlemone biosynthesis. This group contains many genes involved in the biosynthesis of conjugated double bonds, e.g. Bmo_KATQ in *Bombyx mori* (Lepidoptera: Bombycidae) (Moto et al. 2004), Mse_APTQ in *Manduca sexta* (Lepidoptera: Sphingidae) (Matoušková et al. 2007), Lca_KPVQ in *Lampronia capitella* (Lepidoptera: Prodoxidae) (Liénard et al. 2008), and Dpu_LPAE in *Dendrolimus punctatus* (Liénard et al. 2010). Cpo_RPVE groups with Lca_KPVQ, which encodes an enzyme involved in the desaturation of 16:Acyl and Z9-14:Acyl to produce the conjugated Z9Z11-14:Acyl pheromone precursor of *L. capitella* (Liénard et al. 2008). Another interesting candidate is Cpo_CPRQ, which clusters near Dpu_LPAE, an enzyme capable of catalyzing the production of E9Z11-16:Acyl and E9E11-16:Acyl (Liénard et al. 2010).

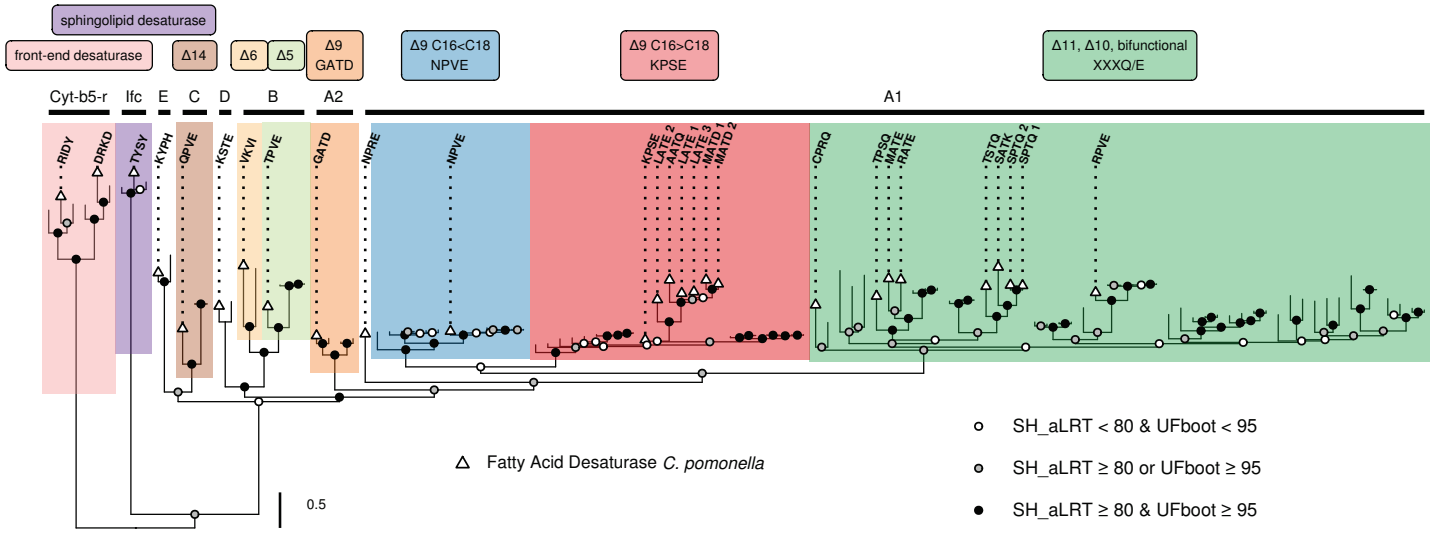


Figure 4: Phylogeny of lepidoptera FAD genes.
The maximum likelihood tree was obtained for predicted amino acid sequence of 114 FAD genes (805 aligned positions) of 28 species, with branch support values calculated from 1000 replicates using the Shimodaira-Hasegawa-like approximate ratio test (SH_aLRT) and ultrafast bootstraping (UFboot). Support values for branches are indicated by filled circles, with the color assigned based on SH-aLRT and UFBoot supports using 80% and 95% as thresholds of branch selection for SH-aLRT and UFBoot supports, respectively. The major constituent six subfamilies of First Desaturase (A1 to E) and two subfamilies of Front-End (Cyt-b5-r) and Sphingolipid Desaturases (Ifc), respectively, are indicated following the nomenclature proposed by Helmkampf et al. (2015). For First Desaturases, the different shades correspond to the indicated putative biochemical activities and consensus signature motif (if any). Triangles indicate sequences from *C. pomonella* (see also Fig. S1 for an extended version of this tree). The scale bar represents 0.5 substitutions per amino acid position.

Patterns of Gene Expression Identify Cpo_CPRQ and Cpo_SPTQ(1) as Candidates

To identify candidates for codlemone production we compared the expression levels of the twenty-seven genes found in the genome using published RNA sequencing (RNA-Seq) data from various tissues and life-stages augmented with two new pheromone gland data sets (this study). Since codlemone is found in the pheromone gland of females, which is located at the tip of the abdomen, we hypothesized that its desaturase(s) would be enriched or even exclusively expressed in this sex and in the population of cells forming this tissue.

We found strong qualitative differences in the transcription profile of the 27 examined FAD genes (Fig. 5). The genes divide into three clusters. The first two correspond to genes that appear ubiquitously expressed. These include the two putative Front-End desaturases Cpo_DRKD and Cpo_RIDY (Cyt-b5-r), the Sphingolipid Δ4 desaturase Cpo_TYSY (Ifc), and five First desaturases, namely Cpo_NPVE (A1; Δ9 C16<C18), Cpo_KPSE (A1; Δ9 C16>C18), Cpo_RPVE (A1; Δ11/Δ10), Cpo_KSTE (D), Cpo_KYPH (E). With the exception of Cpo_RPVE, these genes exhibit a high evolutionary stability characterized by a single-copy gene with rare cases of gains and losses across insects. Their broad expression profiles suggest that these genes fulfill a fundamental metabolic function, making them unlikely candidates for pheromone biosynthesis. By contrast, the other First Desaturase genes found in the third cluster are characterized by a sparse expression profile and greater tissue and/or life-stage specificity: head: Cpo_TPSQ (A1; Δ11/Δ10); ovary/female abdomen: Cpo_TPVE (B; Δ5); pheromone gland: Cpo_CPRQ (A1; Δ11/Δ10), Cpo_SPTQ(1) (A1; Δ11/Δ10); eggs: Cpo_NPRE (A1), Cpo_VKVI (B; Δ6). These expression patterns are suggestive of more specialized functions. Specifically, their expression profiles and levels indicated that Cpo_CPRQ and Cpo_SPTQ(1) (average FPKM in pheromone gland: 6082.0 and 1451.5, respectively) represent primary candidates for functional testing.

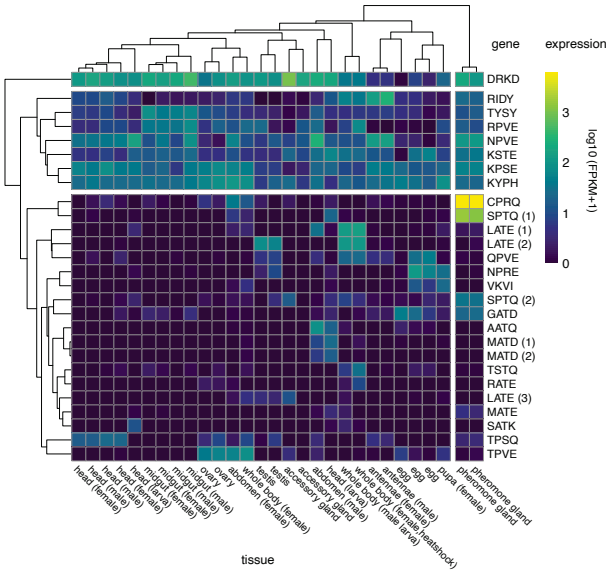


Figure 5: Expression profiling of FAD genes in different tissues of *Cydia pomonella*.
The heatmap represents the absolute expression value (log FPKM) of all FAD genes in the corresponding tissues. Genes are identified by their signature motif. Automatic hierarchical clustering of FAD genes distinguishes three clusters: cluster I and II contain genes with ubiquitous expression across tissues and developmental stages; cluster III contain genes with more divergent expression and higher tissue specificity. Automatic hierarchical clustering of tissues indicates that pheromone gland samples have a distinguishable profile, which is characterized by the over-expression of Cpo_CPRQ and Cpo_SPTQ(1). Estimates of abundance values were obtained by mapping reads against the genome. The NCBI SRA accession numbers of all RNA-Seq data sets used are given in Supplementary Table S4.

Functional Characterization Demonstrates the E9 *FAD Activity of Cpo_CPRQ and Cpo_SPTQ(1)*

Previous studies have shown that the baker's yeast *Saccharomyces cerevisiae* is a convenient system for studying *in vivo* the function of FAD genes and it has become the eukaryotic host of choice for the heterologous expression of desaturases from diverse sources including moths. Conveniently, yeast cells readily incorporate exogenous fatty acid methyl esters and convert them to the appropriate coenzyme A thioester substrates required by FADs.

First, we cloned Cpo_CPRQ and Cpo_SPTQ(1) into a copper-inducible yeast expression vector to generate heterologous expression in *S. cerevisiae*. We carried out assays with precursor supplementation to ensure availably of the medium-chain saturated fatty acids, i.e. lauric (C_{12}) and myristic acids (C_{14}), which are typically less abundant. Cpo_CPRQ conferred on the yeast the ability to produce a small amount of E9-12:Acyl and had no detectable activity outside of the C_{12} substrate (Fig. 6). Cpo_SPTQ(1) showed Δ9 desaturase activity, producing E9- and Z9-12:Acyl, Z9-14:Acyl and Z9-16:Acyl. Double bond positions in these products were confirmed by dimethyl disulfide (DMDS) derivatization (Fig. 6). In addition to retention times matching those of synthetic standards, spectra of the DMDS derivatives of the methyl esters displayed the following diagnostic ions: Z9-16:Me (M^+, m/z 362; A^+, m/z 145; B^+, m/z 217), Z9-14:Me (M^+, m/z 334; A^+, m/z 117; B^+, m/z 217), E9- and Z9-12:Me (M^+, m/z 306; A^+, m/z 89; B^+, m/z 217). No doubly unsaturated products were detected in these experiments.

In a second round of experiments we assayed 7 other First desaturases that could be amplified from pheromone gland cDNA. As predicted from their respective placement in the FAD phylogeny, Cpo_NPVE, Cpo_KPSE and Cpo_GATD encode Δ9 desaturases (Fig. S2). Cpo_NPVE is a Δ9 desaturase with preference for myristic (C_{14}) and palmitic (C_{16}) acid, showing only weak activity on lauric acid (C_{12}). Cpo_GATD and Cpo_KPSE showed Δ9 desaturase activities mainly on C_{16} but also some activity on C_{14}. The other 4 desaturases had no detectable activity in the yeast system.

Finally, we supplemented the growth medium with the monounsaturated methyl esters E9-12:Me and Z9-12:Me. The doubly unsaturated codlemone precursor, E8E10-12:Acyl, was inconsistently detected in some replicates of the assays with Cpo_CPRQ. No doubly unsaturated products were detected with the other constructs.

Functional Analysis Identify Cpo_CPRQ as the FAD Catalyzing Δ8Δ10 Desaturation

In order to evaluate the presumptive Δ8Δ10 desaturase activity of Cpo_CPRQ, we further investigated the activity of the enzyme in an insect cell line. We reasoned that the inconsistent activity of the enzyme in yeast could be due to the cellular environment of the expression system. When we expressed Cpo_CPRQ using the Sf9 cell system and supplementing the culture medium with lauric acid, we could observe the consistent production of E9-12 as well as E8E10-12, demonstrating that Cpo_CPRQ catalyzes the biosynthesis of E8E10-12:Acyl and its monounsaturated intermediate E9-12:Acyl (Fig. 7). As expected from the previous experiments in yeast, we did not detect activity on longer chain substrates. The retention time and mass spectrum of the E8E10-12:Me peak were identical to those of the synthetic standard. Double bond positions of the conjugated system were confirmed by derivatization with 4-methyl-1,2,4-triazoline-3,5-dione (MTAD)(Fig. 7). Finally, we carried out assays to further evaluate the stereospecificity of the enzyme. When provided with E9-12:Me, Cpo_CPRQ produced the doubly unsaturated product (Fig 7). In contrast, we found no evidence for activity on the corresponding diastereomer, Z9-12:Me (Fig 7).

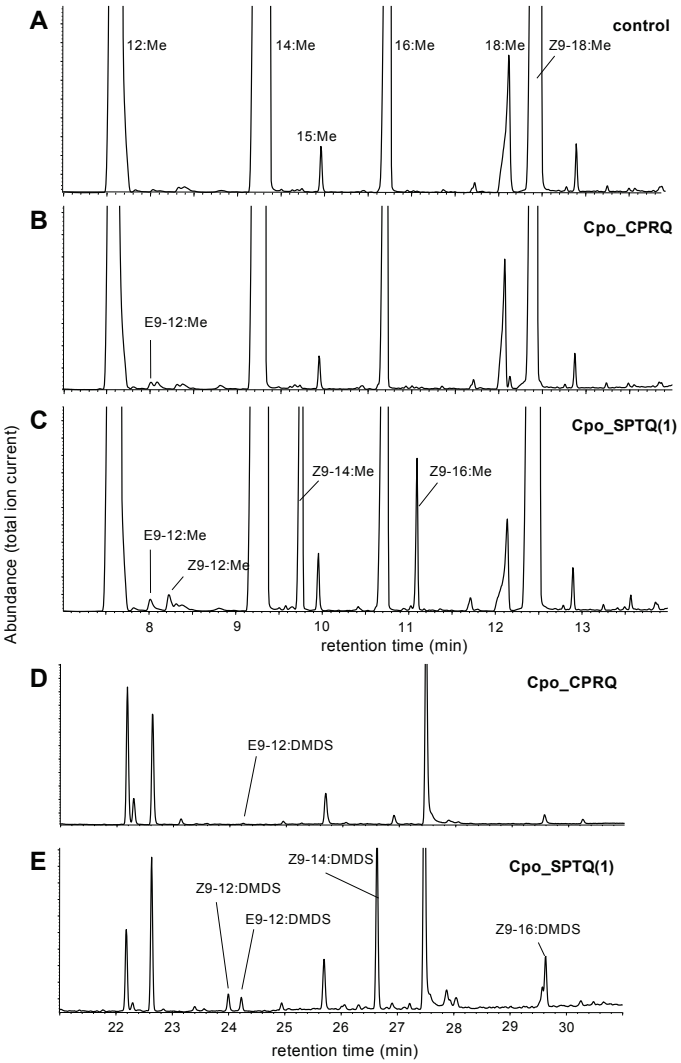


Figure 6: Functional characterization of desaturase activity of candidate genes in yeast.
Total ion chromatograms of fatty acid methyl ester (FAME) products of Cu2+-induced ole1 elo1 S. cerevisiae yeast supplemented with saturated acyl precursors and transformed with (A) empty expression vector (control), (B) pYEX-CHT-Cpo_CPRQ and (C) pYEX-CHT-Cpo_SPTQ(1). Confirmation of identity of enzyme products was obtained by comparison of retention time, mass spectra of synthetic standards. Confirmation of double bond positions by mass spectra of DMDS adducts (D, E).

Altogether, these results show that Cpo_CPRQ exhibit both the ability to catalyze the *trans* desaturation of lauric acid to form E9-12:Acyl and the 1,4-desaturation of the latter to form the E8E10-12 conjugated pheromone precursor of codlemone.

Discussion

In this study, we show the key role of a desaturase with a dual function in the biosynthesis of codlemone. We confirm the results suggested by the labelling experiments reported by Löfstedt and Bengtsson (1988) and demonstrate that Cpo_CPRQ can conduct both the initial desaturation of 12:Acyl into E9-12:Acyl, and then transform E9-12:Acyl into E8E10-12:Acyl, the immediate fatty acyl precursor of E8E10-12:OH, the sex pheromone of *C. pomonella* (Fig. 2). Phylogenetically, Cpo_CPRQ clusters in the Lepidoptera-specific XXXQ/E clade (Fig 4). Several

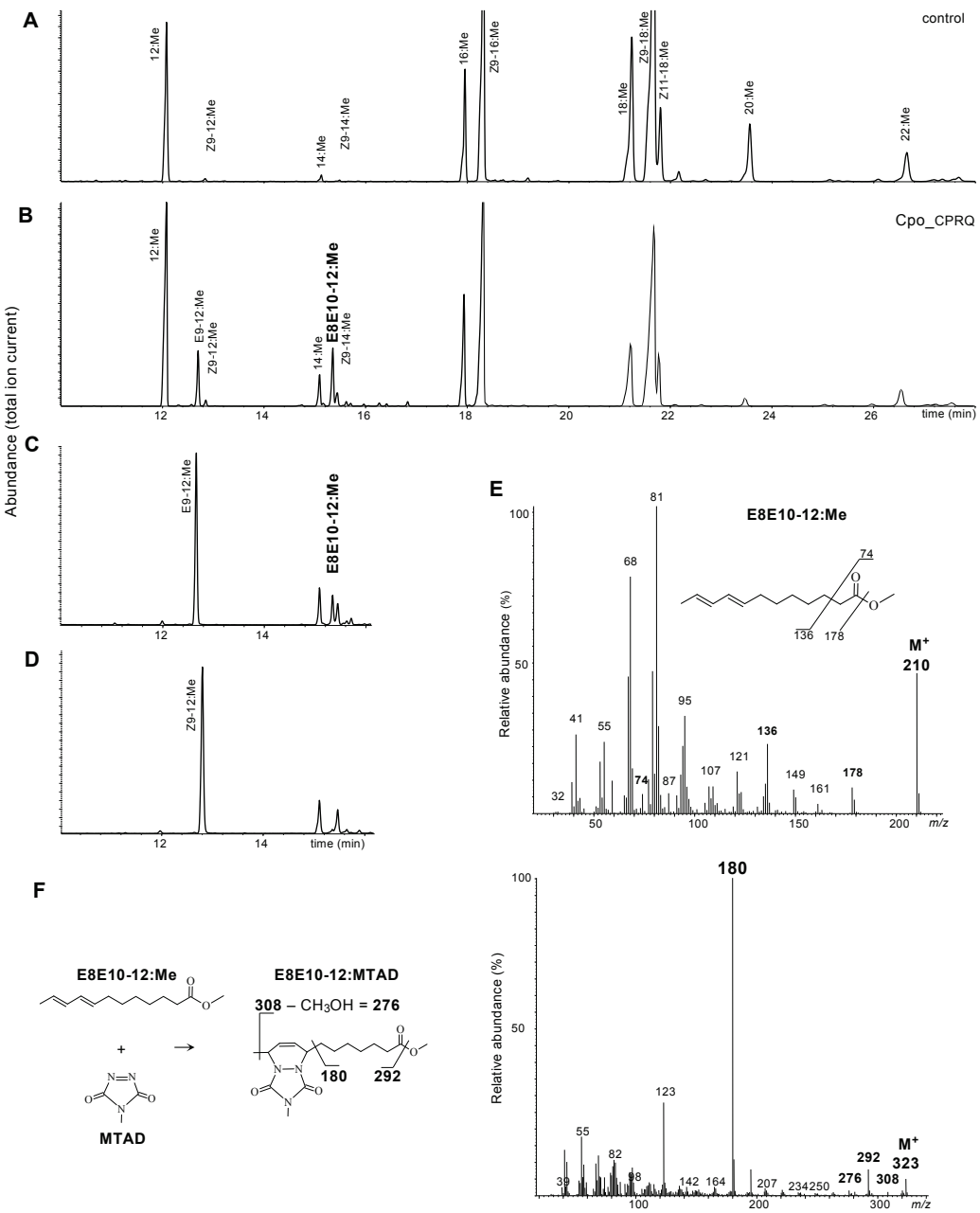


Figure 7: Functional characterization of desaturase activity of Cpo_CPRQ in insect cells.
Total ion chromatograms of methyl esters (FAME) samples from Sf9 cells supplemented with lauric methyl ester (C12) infected with (A) empty virus (control) or (B) recombinant baculovirus expressing Cpo_CPRQ. Cpo_CPRQ produces large amount of E9-12:Me and E8E10-12:Me. Sf9 insect cells infected with bacmid expressing Cpo_CPRQ in medium in the presence of the monoenic intermediate (C) (*E*)-9-dodecenoic methyl ester (E9-12:Me) or (D) (*Z*)-9-dodecenoic methyl ester (Z9-12:Me). The retention time and mass spectrum of the E8E10-12:Me peaks observed after addition of 12:Me and E9-12:Me were identical with those of the synthetic standard. The relatively long retention time (compound eluting later than 14:Me) is in agreement with what is expected from a diene with conjugated double bonds. (E) Spectrum of the E8E10-12:Me peak in panel B. The relatively abundant molecular ion m/z 210 is in agreement with the expectation for a diene with conjugated double bonds. (F) Analyses of MTAD-derivatized samples displayed diagnostic ions at m/z 323 (M+), m/z 308 and m/z 180 (base peak), confirming the identification of the conjugated double bonds system.

FADs found in that clade have been demonstrated to be involved in the pheromone biosynthesis of many moth species, including the species of Tortricinae using Δ11 (*Epiphyas postvittana*, *Choristoneura rosaceana, Agryrotaenia velutinana*) and Δ10 desaturations (*Planotortrix octo*) (Liu et al 2002a; Hao et al. 2002a; Liu et al 2002b; Hao et al. 2002b). Our phylogenetic reconstruction indicates that Cpo_CPRQ does not appear orthologous to any of those genes (Fig. S1), suggesting the recruitment of a different ancestral duplicate early in the evolution of Olethreutinae. In addition to Cpo_CPRQ, Cpo_SPTQ(1) clustered in the same clade and had a similar activity profile on lauric acid, although it produces more Z9- than E9-12:Acyl. With an expression at about 25% the level of Cpo_CPRQ, Cpo_SPTQ(1) is the second most highly expressed FAD gene in the pheromone gland, a tissue in which this FAD appear exclusively expressed. Although Cpo_CPRQ appears sufficient to produce E8E10-12:Acyl, we cannot exclude the possibility that Cpo_SPTQ(1) contributes to the production of the monoene intermediate and is involved in the pheromone biosynthesis of codlemone or E9-12:OH, a pheromone component eliciting antennal response but with no apparent behavioral effect (Witzgall et al. 2001). Interestingly, even though they are phylogenetically clustered within the Δ11/Δ10-desaturase subfamily, Cpo_CPRQ and Cpo_SPTQ(1) exhibit primarily Δ9-desaturase activity, which likely represent the ancestral state. Whether this represents a conservation or a reversal to the ancestral function awaits further investigations. This finding provides exciting opportunities to study the structure-function relationships in FADs, in particular the structural determinant of regioselectivity.

Expansion of gene families is usually regarded as playing a key role in contributing to phenotypic diversity. The acyl-CoA desaturase

gene family is characterized by a highly-dynamic evolutionary history in insects (Roelofs and Rooney 2003; Helmkampf et al. 2015). Bursts of gene duplication have provided many opportunities for evolution to explore protein space and generate proteins with unusual and novel functions. In moths, many examples highlight the role played by the diversification in FAD functions in allowing the expansion of the multi-dimensional chemical space available for communication via sex pheromones. Careful re-annotation of the *C. pomonella* genome and phylogenetic analyses of identified desaturase genes allowed us to identify 27 genes representative of all desaturase subfamilies previously identified in insects. Interestingly, all functional classes previously identified in moths (i.e. Δ5, Δ6, Δ9, Δ10, Δ11, Δ14) have homologs in *C. pomonella*. This indicates that FAD genes tend to be preserved in the genomes for long periods of time. The availability of several chromosome-level assemblies in Lepidoptera has revealed that the genome architecture is generally conserved, even among distant species, with few to no structural rearrangements, i.e. inversions and translocations, being observed (but see (Hill et al. 2019)). This is attested by the high level of synteny existing between *C. pomonella* and *Spodoptera litura* (Lepidoptera: Noctuidae), two representatives of lineages which shared their last common ancestor ca. 120 MYA (Kawahara et al. 2019; Wan et al. 2019). This stability over long evolutionary time means that we can draw tentative conclusions from the genomic organization of gene families. Specifically, the genomic location of FAD genes in *C. pomonella* suggests that the diversification of this gene family is ancient. First Desaturase genes from all groups *sensu* Helmkampf et al. (2015) have representatives in species that have shared their last common ancestor over 300 MYA. Moreover, we found several expansions with differing patterns within Δ9 C16>C18 (KPSE) and the Lepidoptera-specific gene subfamily Δ11, Δ10 & bifunctional (XXXQ/E). Genes clustered in tandem arrays represent evolutionary recent duplications, as exemplified by the expansion in Δ9 C16>C18 (KPSE). These duplicates are likely the product of imperfect homologous recombination. By contrast, with 9 representatives spread across 6 chromosomes, the Δ11/Δ10 subfamily comprises both recent duplicates (e.g., Cpo_SPTQ(1) and Cpo_SPTQ(2)) as well as unlinked FAD genes (such as Cpo_CPRQ). The latter genes are likely to be the products of more ancient events of duplications and chromosomal rearrangements. The location of the majority of genes within this clade suggests that the diversification within this subfamily essential for the pheromone biosynthesis occurred early in the evolution of the Lepidoptera. As genome data from more species become available, it will be possible to follow the evolutionary trajectories of this gene family which has been a key contributor to pheromone evolution.

Expansion of gene families is often followed by a relaxation of selection, which translates into an acceleration of the mutation rate leading to the accumulation of non-synonymous nucleotide substitutions (Bielawski and Yang 2003; Taylor and Raes 2004). Looking at the impact of selection on desaturase genes in insects, Helmkapmf and co-authors (2015) found little evidence for positive selection but rather purifying selection acting on all desaturase genes in insects, even in expanded subfamilies. This could be indicative that large constraints along the coding sequences to not change sites affecting the proper enzymatic function outweigh the potential for diversifying selection. Modification of the regulation of duplicated genes is also expected, with some paralogs specializing to a limited set of tissues or developmental stages (Taylor and Raes 2004). Data from a panel representing a relatively large variety of tissue types and developmental stages reveal a large expression divergence between FAD genes. Interestingly, we see that genes present in recent expansion (i.e. Δ9 C16>C18 (KPSE)), of which members form tandem arrays in the genome, have low to no expression in the large panel of samples we analyzed. The same is also true for *Bombyx mori*. The evidence at hand suggest that these may have limited physiological relevance. This advocates for caution when interpreting the importance and functional consequences of expansions. Some expansions can contribute to lineage-specific adaptation and an increased demand for variability in associated traits. However, it is almost certain that many expansions will not reflect a response to changes and be responsible for the evolution of novel traits but rather have their evolutionary origin in the stochastic nature of the gene duplication process. Functional testing is paramount to advance our mechanistic understanding of the proximate consequences of variation in the size of multigene families.

Previously characterized FADs that catalyze the formation of conjugated fatty acids among moth sex pheromones precursors fall into two groups: (1) double bonds can be inserted via sequential 1,2-desaturation to produce a diene, either by the action of two different desaturases operating consecutively as postulated in *Dendrolimus punctatus* (Lasiocampidae) (Liénard et al., 2010), or by the same desaturase operating twice with an intermediate chain-shortening step as suggested for *Epiphyas postvittana* (Tortricidae) (Liu et al. 2002a), *L. capitella* (Liénard et al. 2008), and *Spodoptera littoralis* (Noctuidae) (Xia et al. 2019); (2) bifunctional desaturases can introduce the first double bond and then rearrange the monoene to produce a diene with conjugated double bonds via 1,4-desaturation, as seen in *Bombyx mori* (Bombycoidae) (Moto et al. 2004), *S. littoralis* (Noctuidae) (Serra et al. 2006) and *Manduca sexta* (Sphingidae) (Matoušková et al. 2007). The *C. pomonella* desaturase falls into the latter category: Cpo_CPRQ introduces first an *E*9 double bond in the 12:Acyl saturated substrate and then transforms the *E*9 unsaturation into the *E*8*E*10 conjugated double bonds. The independent evolution of similar features in enzymes used by distantly related species offers the opportunity to further study mechanistically the determinants of the switch between 1,2 and 1,4-dehydrogenation.

Recent molecular phylogenetic studies (Regier et al. 2012; Fagua et al. 2017) provide strong support for the monophyly of Tortricinae and Olethreutinae, the two most speciose subfamilies within Tortricidae, and the paraphyly of Chlidanotinae. Our tribal-level tree for Tortricidae combined with data on moth pheromones and attractants corroborate the pattern of pheromone components in Tortricinae and Olethreutinae reviewed and discussed by Roelofs and Brown (1982). Tortricinae species use mostly 14-carbon pheromone components with unsaturation introduced via Δ11 desaturation. Molecular and biochemical studies have confirmed the paramount role of Δ11 FADs in the pheromone biosynthesis of not only tortricines, but the vast majority of ditrysian Lepidoptera (see for instance (Jurenka 2003; Roelofs and Rooney 2003; Liénard et al. 2008, 2014; Liénard and Löfstedt 2010)). No pheromones are reported for the Chlidanotinae. However, the sex attractants reported for three species include Z9-14:OAc and Z11-16:OAc, which supports the idea that Δ11 desaturation in combination with chain-shortening may be typical of this subfamily. By contrast, the Olethreutinae subfamily use mainly 12-carbon compounds. The Δ9-12 chain structures used by many species could be accounted for the action of a Δ11 enzyme producing Δ11-14:Acyl as an intermediate following by chain-shortening to Δ9-12:Acyl. For example, the host races of the larch budworm *Zeiraphera diniana*(Tortricidae: Olethreutinae: Eucosmini) developing on larch (*Larix decidua*) or Cembra pine (*Pinus cembra*) used respectively E11-14:OAc and E9-12:OAc as pheromone (Guerin et al. 1984), and it is parsimonious to hypothesize that polymorphism in the activity of the chain-shortening enzyme is responsible for this difference between host races that show little to moderate genomic differentiation (Emelianov et al. 2004). Our data suggest however that an alternative pathway may exist. Indeed, two of the enzymes characterized in the present study (i.e. Cpo_SPTQ(1) and Cpo_CPRQ) can catalyze the production of Δ9-12:Acyl directly from lauric acid (C_{12}), bypassing the

need for prior Δ11 desaturation and chain-shortening. In addition to Δ9-12 chain structures, 12-carbon components with Δ8 and Δ8Δ10 systems are frequently used by olethreutines. Specifically, these occur in the three major tribes: Grapholitini, Eucosmini, and Olethreutini (Fig. 1). The identification of the bifunctional desaturase catalyzing the biosynthesis of Δ8Δ10-12 fatty acid derivatives sheds new light on the pathway associated with the production of pheromone of a large number of tortricids. Monounsaturated compounds with double bonds in Δ8 position frequently co-occur with conjugated Δ8Δ10, including in trace amount in *C. pomonella* (Witzgall et al. 2001). However, since we did not detect any Δ8 activity towards the lauric acid from Cpo_CPRQ or Cpo_SPTQ(1), we anticipate that a different FAD plays a role in the pheromone biosynthesis of Δ8 monoenes. This is supported by the pheromone bouquets identified for several species of olethreutines. In the podborer *Matsumuraeses falcana* (Tortricidae: Olethreutinae: Grapholitini), the pheromone consists of a mixture of E8-12:OAc and E8E10-12:OAc, plus minor amounts of (*E*7,*Z*9)- and (*E*7,*E*9)-dodecadienyl isomers, and the monoenes E10-12:OAc and E10-14:OAc (Wakamura 1985). The major pheromone component in *Hedya nubiferana* (Tortricidae: Olethreutinae: Olethreutini) is E8E10-12:OAc but in addition the pheromone contains relatively large amounts of Z8- and E8-12:OAc (Frerot et al. 1979). The same holds for *Epiblema foenella* (Tortricidae: Olethreutinae: Eucosmini) in which the major pheromone component was identified as Z8Z10-12:OAc: in addition to the other three geometric isomers of the diene, the females produce also E8- and Z8-12:OAc but no E9-12:OAc (Witzgall et al. 1996). Biosynthesis of these bouquets suggests the involvement of a FAD with Δ8 or Δ10 activity. A Δ10 FAD is the key enzyme involved in the production of monounsaturated compounds with double bonds in Δ8 position. Specifically, in *Planotrotrix sp.* and *Ctenopseustis sp.* (Tortricidae: Tortricinae: Archipini), Z8-14:OAc is bionsynthesized via Δ10 desaturation of palmitic acid (C_{16}) followed by chain shortening, whereas the production of Z10-14:OAc involves the same FAD acting on myristic acid (C_{14}) (Foster and Roelofs 1987; Hao et al. 2002; Albre et al. 2012). Our phylogenetic analysis of FAD genes shows that the *C. pomonella* genome contains several genes clustering with the tortricine Δ10 FADs, indicating that this biochemical activity may in theory still be carried out in olethreutines. Additional work will be necessary to elucidate the routes towards the biosynthesis of monounsaturated pheromone components with double bond at even positions.

Biological production of moth pheromones has been suggested as an environmentally-friendly alternative to conventional chemical synthesis. The concept is proven and large-scale production is on the horizon (Nešněrová et al. 2004; Hagström et al. 2013; Ding et al. 2014; Holkenbrink et al. 2020). Identification and characterization of the key enzymes underlying production of pheromone is a crucial step to identify the exogenous building blocks central to the engineering of plant and cell factories. In 2010, the estimated annual production of synthetic codlemone was 25 metric tons for the treatment by mating disruption of 200,000 ha of orchards worldwide (Witzgall et al. 2010). The identification and characterization of the desaturase involved in pheromone biosynthesis in *C. pomonella* is of particular interest as it may open up for biological production of codlemone in plant and cell factories by metabolic engineering (Löfstedt and Xia 2020) and represent an important step towards the biological production of pheromone for the integrated pest management of several economically important olethreutines.

Materials and Methods

Semiochemical Data Set

To compile a dataset of the semiochemicals used as sex attractants by tortricid moths we manually retrieved data from the Pherobase, a freely accessible database of pheromones and semiochemicals which compiles information from the literature and peer-reviewed publications (El-Sayed 2020). Altogether, we gathered data for 536 taxa belonging to the family Tortricidae. Specifically, we were interested in collecting information about the chemical structure of the bioactive fatty acyl derivatives, i.e. double bond position, number and length of the aliphatic chain. Following the database convention, we classified bioactive compounds as *pheromone* or *attractant*. The term pheromone denotes a component of the female sex attractant, identified from females, and with demonstrated biological activity; attractant characterizes those bioactive compounds found by male responses alone, such as by systematic field screening or by significant non-target capture in traps baited with pheromone lures. Although these compounds are not necessarily used in the natural communication system, they have in most cases a high probability of being pheromone components. Details are in Supplementary Table S1.

Molecular Data Set

To infer the evolutionary relationship among the species represented in our trait dataset we used sequence data from 137 terminal taxa in our phylogenetic analysis, including 131 species of Tortricidae as ingroup. The outgroup contained six taxa, each a species representative of Cossidae, Galacticidae, Heliocosmidae, Lacturidae, Limacodidae and Sesiidae, respectively. Our data set is based largely on Regier et al (2012) and Fagua et al. (2017), augmented with a few additional Genbank accessions. Following the aforementioned authors, we used regions of five nuclear genes and one mitochondrial gene. The nuclear genes were: *carbamoyl-phosphate synthetase II* (CAD), *dopa decarboxylase* (DDC), *enolase* (Eno), *period* (PER), *wingless* (WG); the mitochondrial gene was the fragment of *cytochrome oxidase I* (COI) used for barcoding. Details are in Supplementary Table S2. For each genus, we used one representative species as the type species.Our main goal was to place the major transition in the pheromone communication system of Tortricidae and thus, to infer the phylogenetic relationships with good resolution at the subfamily and tribe level. We retrieved a total of 340 Genbank accessions (see Supplementary Table S2 for accession numbers).

Phylogenetic Tree

We concatenated the molecular data corresponding to one taxon and obtained an alignment of the entire molecular data set using MAFFT (Katoh et al. 2002) as implemented in Geneious (Biomatters). The final alignment comprised 7591 sites. To define the best scheme of partitions using the MAFFT alignment, we used PartitionFinder v2.1.1 (Lanfear et al. 2017). We tested models of nucleotide substitution with partitions representing codon position and genes using *greedy* search (Lanfear et al. 2012), *AICC* model selection, and setting the phylogeny program to *phyml* (Guindon et al. 2010)(default value for all other settings). A 16-partitions scheme was selected and used to perform the parametric phylogenetic analysis in IQ-TREE 1.6.11 (Nguyen et al. 2015; Chernomor et al. 2016). The maximum likelihood analysis was performed using default settings and by calculating Shimodaira-Hasegawa-like approximate ratio test (SH-aLRT) support and ultrafast bootstrap support (UFBoot; (Hoang et al. 2018)) after 1000 replicates each. We used the R package *ggtree* (Yu et al. 2017) for visualizing and annotating the phylogenetic tree with associated semiochemical data.

Insects, Tissue Collection, and RNA-Seq

Pupae of *C. pomonella* were purchased from Entomos AG (Switzerland) and were sexed upon arrival to our laboratory. Adult males and females emerged in separate jars in a climate chamber at 23±1ºC under a 16h:8h light: dark cycle. On the third day after emergence, pheromone glands of female moths were dissected, immersed in TRIzol Reagent (Thermo Scientific), and stored at -80ºC until RNA extraction. Two pools of 30 glands were collected 3 hours before and after the onset of the photophase, respectively. Total RNA was extracted according to the manufacturer's instructions. RNA concentration and purity were initially assessed on a NanoDrop2000 (Thermo Fisher Scientific). Library preparation and paired-end Illumina sequencing (2x100bp) were performed by BGI (Hong-Kong, PRC). We obtained ~66 million QC-passing reads per sample.

Genome Annotation Improvements

To ensure that all FAD genes were correctly annotated in the C. pomonella genome we used our RNA-Seq data of pheromone gland to improve the annotation (cpom.ogs.v1.chr.gff3) which we obtained from InsectBase (http://www.insect-genome.com/cydia/). We performed low-quality base trimming and adaptor removal using cutadapt version 1.13 (Martin 2011) and aligned the trimmed read pairs against the genome using HISAT2

version 2.2.0 (Kim et al. 2019). The existing annotation was used to create a list of known splice sites using a python script distributed with HISAT2. We used StringTie version 2.1.3b (Pertea et al. 2015) with the *conservative* transcript assembly setting to improve the annotation and reconstruct a non-redundant set of transcripts observed in any of the RNA-Seq samples. We applied Trinotate version 3.2.1 (Bryant et al. 2017) to generate a functional annotation of the transcriptome data. We identified candidate FAD genes by searching for genes harboring the PFAM domain corresponding to a fatty-acid desaturase type 1 domain (PF00487) in their predicted protein sequences. We used the R package *KaryoploteR* (Gel and Serra 2017) to visualize the location of the FAD genes.

Desaturase Phylogenetic Analysis

To investigate the evolutionary relationship of the *C. pomonella* FADs we carried out a phylogenetic analysis with other moth FAD proteins. Our sampling includes all the functional classes previously identified in moths (i.e. Δ5, Δ6, Δ9, Δ10, Δ11, Δ14) and representatives of Tortricidae as well as other moth families (see Supplementary Table S3 for list of taxa). Protein sequences for these enzymes were downloaded from GenBank. Protein sequences for *C. pomonella* were predicted from the genome using gffread (Pertea and Pertea 2020) with our improved annotation.

We aligned amino acid sequences using MAFFT version 7.427 (Katoh et al. 2002; Katoh and Standley 2013). We used the L-INS-i procedure and the BLOSUM30 matrix as scoring matrix. The final multiple sequence alignment contained 114 sequences with 805 amino-acid sites. The phylogenetic tree was constructed in IQ-TREE 1.6.11 (Nguyen et al. 2015). Automatic model search was performed using ModelFInder (Kalyaanamoorthy et al. 2017) with the search restricted to models including the WAG, LG and JTT substitution models. The maximum likelihood analysis was performed using default settings and by calculating Shimodaira-Hasegawa-like approximate ratio test (SH-aLRT) support and ultrafast bootstrap support (UFBoot; (Hoang et al. 2018)) after 1000 replicates each. We used the R package *ggtree* (Yu et al. 2017) for visualizing and annotating the phylogenetic tree, with midpoint rooting performed using the midpoint.root function implemented in the *phytools* package (Revell 2012).

Transcript Abundance Estimation

In addition to our pheromone gland data sets, we retrieved 27 already published Illumina paired-end RNA-Seq data sets for *C. pomonella* available on the Sequence Read Archive (SRA) (see Supplementary Table S4 for accession numbers). These data correspond to various tissues or life-stages. Following QC, low-quality base trimming and adaptor removal with cutadapt, read pairs were aligned against the genome using HISAT2. The obtained alignment BAM files were used to estimate transcript abundance using StringTie together with our improved annotation. The abundance tables from StringTie were imported into R using the *tximport* package (Soneson et al. 2016), which was used to compute gene-level abundance estimates reported as FPKM. We used the R package *pheatmap* (Kolde 2019) to visualize the expression level of FADs genes.

Cloning of Desaturases

First-strand cDNAs were synthesized from one microgram of total RNA using the SuperScript IV cDNA synthesis kit (Thermo Scientific). The cDNA products were diluted to 50 µL. PCR amplification (50 µL) was performed using 2.5 µL of cDNA as template with *attB*-flanked gene-specific primers (Supplementary Table S5) in a Veriti Thermo Cycler (Thermo Fisher Scientific) using Phusion polymerase master mix (Thermo Fisher Scientific). Cycling parameters were as follows: an initial denaturing step at 94ºC for 5 min, 35 cycles at 96ºC for 15 s, 55ºC for 30 s, 72ºC for 1 min followed by a final extension step at 72ºC for 10 min. Specific PCR products were cloned into the pDONR221 vector (Thermo Fisher Scientific) in presence of BP clonase (Invitrogen). Plasmid DNAs were isolated according to standard protocols and recombinant plasmids were subjected to sequencing using universal M13 primers and the BigDye terminator cycle sequencing kit v1.1 (Thermo Fisher Scientific). Sequencing products were EDTA/ethanol-precipitated, dissolved in formamide and loaded for analysis on an in-house capillary 3130xl Genetic analyzer (Applied Biosystems).

Pheromone Compounds and Corresponding Fatty Acid Precursors

All synthetic pheromone compounds used as references for identification came from our laboratory collection unless otherwise indicated. Pheromone compounds are referred to as short forms, e.g. (*E*8,*E*10)-8,10-dodecadienol is E8E10-12:OH, the corresponding acetate would be E8E10-12:OAc, corresponding acyl moiety (dodecadienoate) is E8E10-12:Acyl (although it may occur as CoA-derivative) and the methyl ester thereof is referred to as E8E10-12:Me. The dodecanoic methyl ester (methyl laurate, 12:Me) and tetradecanoic methyl ester (methyl myristate, 14:Me) were purchased from Larodan Fine Chemicals (Malmö, Sweden). The (*E*)-9-dodecenol (E9-12:OH) was purchased from Pherobank (Wageningen, The Netherlands) and then converted into the corresponding methyl ester (E9-12:Me). E8E10-12:OAc was purchased from Bedoukian (Danbury, CT, USA) and converted to the corresponding alcohol by hydrolysis using a 0.5 M solution of KOH in methanol. Fatty alcohols were oxidized to the corresponding acid with pyridinium dichromate in dimethylformamide as described by Bjostad and Roelofs (1984). The methyl esters were prepared as described under fatty acid analysis (see below).

Heterologous Expression of Desaturases in Yeast

Plasmids containing the genes of interest were selected as entry clones and subcloned into the copper-inducible pYEX-CHT vector (Patel et al. 2003). Sequences of recombinant constructs were verified by Sanger sequencing. The pYEX-CHT recombinant expression vectors harboring *C. pomonella* FADs were introduced into the double deficient *elo1*Δ/*ole1*Δ strain (*MATa elo1::HIS3 ole1::LEU2 ade2 his3 leu2 ura3*) of the yeast *Saccharomyces cerevisiae*, which is defective in both desaturase and elongase functions (Schneiter et al. 2000) using the *S.c.* EasyComp Transformation kit (Thermo Fisher Scientific). For selection of uracil and leucine prototrophs, the transformed yeast was allowed to grow on SC plate containing 0.7 % YNB (w/o AA; with Ammonium sulfate) and a complete drop-out medium lacking uracil and leucine (Formedium, UK), 2% glucose, 1% tergitol (type Nonidet NP-40, Sigma), 0.01% adenine (Sigma) and containing 0.5 mM oleic acid (Sigma) as extra fatty acid source. After 4 days at 30ºC, individual colonies were selected and used to inoculate 10 mL selective medium, which was grown at 30ºC at 300 rpm in a shaking incubator for 48 h. Yeast cultures were diluted to an OD_{600} of 0.4 in 10 mL of fresh selective medium containing 2 mM $CuSO_4$ for induction, with or without addition of a biosynthetic precursor. Yeast cells contain sufficient quantities of naturally-occurring palmitic acid, hence supplementation with 16:Me was typically not necessary. Lauric acid and myristic acid were added to the medium as methyl esters (12:Me and 14:Me). In subsequent assays the monounsaturated methyl esters E9-12:Me and Z9-12:Me were also supplemented. Each FAME precursor was prepared at a concentration of 100 mM in 96% ethanol and added to reach a final concentration of 0.5 mM in the culture medium. Yeasts were cultured in 30ºC with Cu^{2+}-induction. After 48 h yeast cells were harvested by centrifugation at 3000 rpm and the supernatant was discarded. The pellets were stored at –80°C until fatty acid analysis. Yeast expression experiments were conducted with three independent replicates.

Heterologous Expression of Desaturases in Insect Cells

The expression construct for Cpo_CPRQ in the baculovirus expression vector system (BEVS) donor vector pDEST8_CPRQ was made by LR reaction. Recombinant bacmids were made according to instructions for the Bac-to-Bac® Baculovirus expression system given by the manufacturer (Invitrogen) using DH10EMBacY (Geneva Biotech). Baculovirus generation was done using *Spodoptera frugiperda* Sf9 cells (Thermo Fisher Scientific), Ex-Cell 420 serum-free medium (Sigma) and baculoFECTIN II (OET). The virus was then amplified once to generate a P2 virus stock using Sf9 cells and Ex-Cell 420 medium. Viral titer in the P2 stock was determined using the BaculoQUANT all-in-one qPCR kit (OET) and found to be: 3 x 10^8 pfu/mL for Cpo_CPRQ.

Insect cells lines Sf9 were diluted to 2 x 10^6 cells/mL. Heterologous expression was performed in 20-ml cultures in Ex-Cell 420 medium and the cells infected at an MOI of 1. The cultures were incubated in 125 mL Erlenmeyer flasks (100 rpm, 27 °C), with fatty-acid methyl-ester substrates supplemented at a final concentration of 0.25 mM after one day. After three days, 7.5-mL samples were taken from the culture and centrifuged for 15 min at 4500*g* at 4° C. The pellets were stored at -80°C until fatty acid analysis. Sf9 expression experiments were conducted in three replicates.

Fatty Acid Analysis

Total lipids from cell pellets were extracted using 3.75 mL of methanol/chloroform (2:1 v/v) in clear glass tubes. One mL of 0.15M acetic acid and 1.25 mL of water were added to wash the chloroform phase. Tubes were vortexed vigorously and centrifuged at 2000 rpm for 2 min. The bottom chloroform phase (~ 1 m), which contains the total lipids, was transferred to a new glass tube. Fatty acid methyl esters (FAMEs) were produced from this total lipid extract. The solvent was evaporated to dryness under gentle nitrogen flow. One mL of sulfuric acid 2 % in methanol was added to the tube, vortexed vigorously, and incubated at 90ºC for 1

hour. After incubation, 1 mL of water was added, mixed well, and 1 mL of hexane was used to extract the FAMEs.

The methyl ester samples were subjected to GC-MS analysis on a Hewlett Packard 6890 GC (Agilent) coupled to a mass selective detector HP 5973 (Agilent). The GC was equipped with an HP-INNOWax column (30 m × 0.25 mm × 0.25 µm; Agilent), and helium was used as carrier gas (average velocity: 33 cm/s). The MS was operated in electron impact mode (70eV), and the injector was configured in splitless mode at 220ºC. The oven temperature was set to 80ºC for 1 min, then increased at a rate of 10ºC/min up to 210ºC, followed by a hold at 210ºC for 15 min, and then increased at a rate of 10ºC/min up to 230ºC followed by a hold at 230ºC for 20 min. The methyl esters were identified based on mass spectra and retention times in comparison with those of our collection of synthetic standards.

Double-bond positions of monounsaturated compounds were confirmed by dimethyl disulfide (DMDS) derivatization (Buser et al. 1983), followed by GC-MS analysis. FAMEs (50 µL) were transferred to a new tube, 50 µL DMDS were added and the mixture was incubated at 40ºC overnight in the presence of 5 µL of iodine (5% in diethyl ether) as catalyst. Hexane (200 µL) was added to the sample and the reaction was neutralized by addition of 50-100µL sodium thiosulfate ($Na_2S_2O_3$; 5% in water). The organic phase was recovered and concentrated under a gentle nitrogen stream to a suitable volume (Liénard et al., 2008). DMDS adducts were analyzed on an Agilent 6890 GC system equipped with a HP-5MS capillary column (30 m × 0.25 mm × 0.25 µm; Agilent) coupled with an HP 5973 mass selective detector. The oven temperature was set at 80ºC for 1 min, raised to 140ºC at a rate of 20ºC/min, then to 250ºC at a rate of 4ºC/min and held for 20 min (Wang et al., 2010).

Double-bond positions of di-unsaturated compounds with conjugated double-bonds were confirmed by derivatization with 4-methyl-1,2,4-triazoline-3,5-dione (MTAD) (Marques et al. 2004). MTAD adducts were prepared using 15 µL of the methyl ester extracts transferred into a glass vial. The solvent was evaporated under gentle nitrogen flow and 10 µL of dichloromethane (CH_2Cl_2) was added to dissolve the FAMEs. The resulting solution was treated with 10 µL of MTAD (Sigma; 2 mg/mL in CH_2Cl_2). The reaction was ran for 10 min at room temperature and 2 µL was subjected to GC-MS analysis on an Agilent 6890 GC system equipped with a HP-5MS capillary column (30 m × 0.25 mm × 0.25 µm; Agilent) coupled with an HP 5973 mass selective detector. The injector temperature was set to 300 °C, the oven was set to 100 °C and then increased by 15 °C/min up to 300 °C followed by a hold at 300 °C for 20 min.

End Matter

Author Contributions and Notes

J.-M.L, B.-J.D. and C.L. designed research; B.-J.D. collected tissue and processed samples for RNA-Seq; J.-M.L. performed analyses related to phylogenetics and bioinformatics; B.-J.D. performed functional assays and chemical analyses; J.-M.L. prepared the figures; J.-M.L. wrote the paper with input from B.-J.D. and C.L..
The authors declare no conflict of interest.
This article contains supporting information online.

Acknowledgments

The computations in this paper were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University. This project has received funding from the European Union's Horizon 2020 research and innovation programme (grant agreement nº 760798)(CL), the Swedish Foundation for Strategic Research (grant nº RBP 14–0037, Oil Crops for the Future) (CL), the Swedish Research Council (CL), Formas (CL) and the Royal Physiographic Society in Lund (JML).

References

Albre J, Liénard MA, Sirey TM, et al (2012) Sex Pheromone Evolution Is Associated with Differential Regulation of the Same Desaturase Gene in Two Genera of Leafroller Moths. PLOS Genetics 8:e1002489. https://doi.org/10.1371/journal.pgen.1002489

Arn H, Guerin PM, Buser HR, et al (1985) Sex pheromone blend of the codling moth, *Cydia pomonella*: Evidence for a behavioral role of dodecan-1-ol. Experientia 41:1482–1484. https://doi.org/10.1007/BF01950048

Beroza M, Bierl BA, Moffitt HR (1974) Sex Pheromones: (E,E)-8,10-Dodecadien-1-ol in the Codling Moth. Science 183:89–90. https://doi.org/10.1126/science.183.4120.89

Bielawski JP, Yang Z (2003) Maximum likelihood methods for detecting adaptive evolution after gene duplication. In: Meyer A, Van de Peer Y (eds) Genome Evolution: Gene and Genome Duplications and the Origin of Novel Gene Functions. Springer Netherlands, Dordrecht, pp 201–212

Bjostad LB, Roelofs WL (1984) Sex pheromone biosynthetic precursors in *Bombyx mori*. Insect Biochemistry 14:275–278. https://doi.org/10.1016/0020-1790(84)90060-X

Bryant DM, Johnson K, DiTommaso T, et al (2017) A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. Cell Reports 18:762–776. https://doi.org/10.1016/j.celrep.2016.12.063

Buser HR, Arn H, Guerin P, Rauscher S (1983) Determination of double bond position in mono-unsaturated acetates by mass spectrometry of dimethyl disulfide adducts. Anal Chem 55:818–822. https://doi.org/10.1021/ac00257a003

Butenandt A, Beckmann R, Stamm D, Hecker E (1959) Über den Sexuallockstoff des Seidenspinners *Bombyx mori*—Reindarstellung und Konstitution. Z Naturforsch 14:283–284

Chernomor O, von Haeseler A, Minh BQ (2016) Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst Biol 65:997–1008. https://doi.org/10.1093/sysbio/syw037

Ding B-J, Hofvander P, Wang H-L, et al (2014) A plant factory for moth pheromone production. Nature Communications 5:3353. https://doi.org/10.1038/ncomms4353

El-Sayed AM (2020) The Pherobase: Database of Pheromones and Semiochemicals. http://www.pherobase.com

Emelianov I, Marec F, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proceedings of the Royal Society of London Series B: Biological Sciences 271:97–105. https://doi.org/10.1098/rspb.2003.2574

Fagua G, Condamine FL, Horak M, et al (2017) Diversification shifts in leafroller moths linked to continental colonization and the rise of angiosperms. Cladistics 33:449–466. https://doi.org/10.1111/cla.12185

Fang S, Ting C-T, Lee C-R, et al (2009) Molecular Evolution and Functional Diversification of Fatty Acid Desaturases after Recurrent Gene Duplication in *Drosophila*. Mol Biol Evol 26:1447–1456. https://doi.org/10.1093/molbev/msp057

Foster SP, Roelofs WL (1987) Sex pheromone differences in populations of the brownheaded leafroller, *Ctenopseustis obliquana*. J Chem Ecol 13:623–629. https://doi.org/10.1007/BF01880104

Frerot B, Priesner E, Gallois M (1979) A Sex Attractant for the Green Budworm Moth, *Hedya nubiferana*. Zeitschrift für Naturforschung C 34:1248–1252. https://doi.org/10.1515/znc-1979-1229

Gel B, Serra E (2017) karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. Bioinformatics 33:3088–3090. https://doi.org/10.1093/bioinformatics/btx346

Gilligan TM, Baixeras J, Brown JW (2018) T@RTS: Online World Catalogue of the Tortricidae (Ver. 4.0). http://www.tortricid.net/catalogue.asp

Guerin PM, Baltensweiler W, Arn H, Buser H-R (1984) Host race pheromone polymorphism in the larch budmoth. Experientia 40:892–894. https://doi.org/10.1007/BF01952013

Guindon S, Dufayard J-F, Lefort V, et al (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol 59:307–321. https://doi.org/10.1093/sysbio/syq010

Hagström ÅK, Albre J, Tooman LK, et al (2014) A Novel Fatty Acyl Desaturase from the Pheromone Glands of *Ctenopseustis obliquana* and *C. herana* with Specific Z5-Desaturase Activity on Myristic Acid. J Chem Ecol 40:63–70. https://doi.org/10.1007/s10886-013-0373-1

Hagström ÅK, Wang H-L, Liénard MA, et al (2013) A moth pheromone brewery: production of (Z)-11-hexadecenol by heterologous co-expression of two biosynthetic genes from a noctuid moth in a yeast cell factory. Microbial Cell Factories 12:125. https://doi.org/10.1186/1475-2859-12-125

Hahn MW, Bie TD, Stajich JE, et al (2005) Estimating the tempo and mode of gene family evolution from comparative genomic data. Genome Res 15:1153–1160. https://doi.org/10.1101/gr.3567505

Hahn MW, Han MV, Han S-G (2007) Gene Family Evolution across 12 *Drosophila* Genomes. PLOS Genetics 3:e197. https://doi.org/10.1371/journal.pgen.0030197

Hao G, Liu W, O'Connor M, Roelofs WL (2002a) Acyl-CoA Z9- and Z10-desaturase genes from a New Zealand leafroller moth species, *Planotortrix octo*. Insect Biochemistry and Molecular Biology 32:961–966. https://doi.org/10.1016/S0965-1748(01)00176-X

Hao G, O'Connor M, Liu W, Roelofs WL (2002b) Characterization of Z/E11- and Z9-desaturases from the obliquebanded leafroller moth, *Choristoneura rosaceana.* Journal of Insect Science 2:26. https://doi.org/10.1093/jis/2.1.26

Helmkampf M, Cash E, Gadau J (2015) Evolution of the Insect Desaturase Gene Family with an Emphasis on Social Hymenoptera. Mol Biol Evol 32:456–471. https://doi.org/10.1093/molbev/msu315

Hill J, Rastas P, Hornett EA, et al (2019) Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Science Advances 5:eaau3648. https://doi.org/10.1126/sciadv.aau3648

Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281

Holkenbrink C, Ding B-J, Wang H-L, et al (2020) Production of moth sex pheromones for pest control by yeast fermentation. bioRxiv 2020.07.15.205047. https://doi.org/10.1101/2020.07.15.205047

Jurenka RA (2003) Biochemistry of female moth sex pheromones. In: Insect Pheromone Biochemistry and Molecular Biology. Academic Press, pp 53–80

Kalyaanamoorthy S, Minh BQ, Wong TKF, et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kawahara AY, Plotkin D, Espeland M, et al (2019) Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. PNAS 116:22657–22663. https://doi.org/10.1073/pnas.1907847116

Kim D, Paggi JM, Park C, et al (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology 37:907–915. https://doi.org/10.1038/s41587-019-0201-4

Knipple DC, Rosenfield C-L, Nielsen R, et al (2002) Evolution of the Integral Membrane Desaturase Gene Family in Moths and Flies. Genetics 162:1737–1752

Kolde R (2019) pheatmap: Pretty Heatmaps. Version 1.0.12URL https://CRAN.R-project.org/package=pheatmap

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Mol Biol Evol 29:1695–1701. https://doi.org/10.1093/molbev/mss020

Lanfear R, Frandsen PB, Wright AM, et al (2017) PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34:772–773. https://doi.org/10.1093/molbev/msw260

Lassance J-M, Löfstedt C (2009) Concerted evolution of male and female display traits in the European corn borer, *Ostrinia nubilalis*. BMC Biology 7:10. https://doi.org/10.1186/1741-7007-7-10

Lespinet O, Wolf YI, Koonin EV, Aravind L (2002) The Role of Lineage-Specific Gene Family Expansion in the Evolution of Eukaryotes. Genome Res 12:1048–1059. https://doi.org/10.1101/gr.174302

Liénard MA, Lassance J-M, Wang H-L, et al (2010) Elucidation of the sex-pheromone biosynthesis producing 5,7-dodecadienes in *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae) reveals Δ11- and Δ9-desaturases with unusual catalytic properties. Insect Biochemistry and Molecular Biology 40:440–452. https://doi.org/10.1016/j.ibmb.2010.04.003

Liénard MA, Löfstedt C (2010) Functional flexibility as a prelude to signal diversity? Role of a fatty acyl reductase in moth pheromone evolution. Communicative & Integrative Biology 3:586–588. https://doi.org/10.4161/cib.3.6.13177

Liénard MA, Strandh M, Hedenström E, et al (2008) Key biosynthetic gene subfamily recruited for pheromone production prior to the extensive radiation of Lepidoptera. BMC Evolutionary Biology 8:270. https://doi.org/10.1186/1471-2148-8-270

Liénard MA, Wang H-L, Lassance J-M, Löfstedt C (2014) Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. Nature Communications 5:3957. https://doi.org/10.1038/ncomms4957

Liu W, Jiao H, Murray NC, et al (2002a) Gene characterized for membrane desaturase that produces (*E*)-11 isomers of mono- and diunsaturated fatty acids. PNAS 99:620–624. https://doi.org/10.1073/pnas.221601498

Liu W, Jiao H, O'Connor M, Roelofs WL (2002b) Moth desaturase characterized that produces both Z and E isomers of Δ11-tetradecenoic acids. Insect Biochemistry and Molecular Biology 32:1489–1495. https://doi.org/10.1016/S0965-1748(02)00069-3

Liu W, Rooney AP, Xue B, Roelofs WL (2004) Desaturases from the spotted fireworm moth (*Choristoneura parallela*) shed light on the evolutionary origins of novel moth sex pheromone desaturases. Gene 342:303–311. https://doi.org/10.1016/j.gene.2004.08.017

Löfstedt C, Bengtsson M (1988) Sex pheromone biosynthesis of (E,E)-8,10-dodecadienol in codling moth Cydia pomonella involves E9 desaturation. J Chem Ecol 14:903–915. https://doi.org/10.1007/BF01018782

Löfstedt C, Wahlberg N, Millar JM (2016) Evolutionary patterns of pheromone diversity in Lepidoptera. In: Pheromone communication in moths: evolution, behavior and application. University of California Press, Berkeley, pp 43–82

Löfstedt C, Xia Y-H (2020) Biological production of insect pheromones in cell and plant factories. In: Blomquist GJ, Vogt RG (eds) Insect Pheromone Biochemistry and Molecular Biology (Second Edition). Academic Press, London, pp 89–121

Lu F, Wei Z, Luo Y, et al (2020) SilkDB 3.0: visualizing and exploring multiple levels of data for silkworm. Nucleic Acids Res 48:D749–D755. https://doi.org/10.1093/nar/gkz919

Marques FA, Millar JG, McElfresh JS (2004) Efficient method to locate double bond positions in conjugated trienes. Journal of Chromatography A 1048:59–65. https://doi.org/10.1016/j.chroma.2004.07.017

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12. https://doi.org/10.14806/ej.17.1.200

Matoušková P, Pichová I, Svatoš A (2007) Functional characterization of a desaturase from the tobacco hornworm moth (*Manduca sexta*) with bifunctional Z11- and 10,12-desaturase activity. Insect Biochemistry

and Molecular Biology 37:601–610. https://doi.org/10.1016/j.ibmb.2007.03.004

McDonough LM, Moffitt HR (1974) Sex Pheromone of the Codling Moth. Science 183:978–978. https://doi.org/10.1126/science.183.4128.978

Moto K, Suzuki MG, Hull JJ, et al (2004) Involvement of a bifunctional fatty-acyl desaturase in the biosynthesis of the silkmoth, *Bombyx mori*, sex pheromone. PNAS 101:8631–8636. https://doi.org/10.1073/pnas.0402056101

Navarro I, Mas E, Fabriàs G, Camps F (1997) Identification and biosynthesis of (*E*,*E*)-10,12-tetradecadienyl acetate in *Spodoptera littoralis* female sex pheromone gland. Bioorganic & Medicinal Chemistry 5:1267–1274. https://doi.org/10.1016/S0968-0896(97)00072-2

Nešněrová P, Šebek P, Macek T, Svatoš A (2004) First semi-synthetic preparation of sex pheromones. Green Chemistry 6:305–307. https://doi.org/10.1039/B406814A

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300

Patel O, Fernley R, Macreadie I (2003) *Saccharomyces cerevisiae* expression vectors with thrombin-cleavable N- and C-terminal 6x(His) tags. Biotechnology Letters 25:331–334. https://doi.org/10.1023/A:1022384828795

Pertea G, Pertea M (2020) GFF Utilities: GffRead and GffCompare. F1000Res 9:304. https://doi.org/10.12688/f1000research.23297.1

Pertea M, Pertea GM, Antonescu CM, et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology 33:290–295. https://doi.org/10.1038/nbt.3122

Regier JC, Brown JW, Mitter C, et al (2012) A Molecular Phylogeny for the Leaf-Roller Moths (Lepidoptera: Tortricidae) and Its Implications for Classification and Life History Evolution. PLoS ONE 7:e35574. https://doi.org/10.1371/journal.pone.0035574

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Roe AD, Weller SJ, Baixeras J, et al (2009) Evolutionary framework for Lepidoptera model systems. In: Genetics and Molecular Biology of Lepidoptera. CRC Press, Boca Raton, pp 1–24

Roelofs W, Comeau A, Hill A, Milicevic G (1971) Sex Attractant of the Codling Moth: Characterization with Electroantennogram Technique. Science 174:297–299. https://doi.org/10.1126/science.174.4006.297

Roelofs WL, Brown RL (1982) Pheromones and Evolutionary Relationships of Tortricidae. Annual Review of Ecology and Systematics 13:395–422. https://doi.org/10.1146/annurev.es.13.110182.002143

Roelofs WL, Liu W, Hao G, et al (2002) Evolution of moth sex pheromones via ancestral genes. Proc Natl Acad Sci USA 99:13621. https://doi.org/10.1073/pnas.152445399

Roelofs WL, Rooney AP (2003) Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. PNAS 100:9179–9184. https://doi.org/10.1073/pnas.1233767100a

Schneiter R, Tatzer V, Gogg G, et al (2000) Elo1p-Dependent Carboxy-Terminal Elongation of C14:1Δ9 to C16:1Δ11 Fatty Acids in *Saccharomyces cerevisiae*. Journal of Bacteriology 182:3655–3660. https://doi.org/10.1128/JB.182.13.3655-3660.2000

Serra M, Piña B, Bujons J, et al (2006) Biosynthesis of 10,12-dienoic fatty acids by a bifunctional Δ11desaturase in *Spodoptera littoralis*. Insect Biochemistry and Molecular Biology 36:634–641. https://doi.org/10.1016/j.ibmb.2006.05.005

Shirangi TR, Dufour HD, Williams TM, Carroll SB (2009) Rapid Evolution of Sex Pheromone-Producing Enzyme Expression in *Drosophila*. PLOS Biology 7:e1000168. https://doi.org/10.1371/journal.pbio.1000168

Soneson C, Love MI, Robinson MD (2016) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4:1521. https://doi.org/10.12688/f1000research.7563.2

Sperling P, Ternes P, Zank TK, Heinz E (2003) The evolution of desaturases. Prostaglandins, Leukotrienes and Essential Fatty Acids 68:73–95. https://doi.org/10.1016/S0952-3278(02)00258-2

Takahashi A, Tsaur S-C, Coyne JA, Wu C-I (2001) The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. PNAS 98:3920–3925. https://doi.org/10.1073/pnas.061465098

Taylor JS, Raes J (2004) Duplication and Divergence: The Evolution of New Genes and Old Ideas. Annu Rev Genet 38:615–643. https://doi.org/10.1146/annurev.genet.38.072902.092831

Tillman JA, Seybold SJ, Jurenka RA, Blomquist GJ (1999) Insect pheromones—an overview of biosynthesis and endocrine regulation. Insect Biochemistry and Molecular Biology 29:481–514. https://doi.org/10.1016/S0965-1748(99)00016-8

Tumlinson JH, Teal PEA, Fang N (1996) The integral role of triacyl glycerols in the biosynthesis of the aldehydic sex pheromones of *Manduca sexta* (L.). Bioorganic & Medicinal Chemistry 4:451–460. https://doi.org/10.1016/0968-0896(96)00025-9

Wakamura S (1985) Identification of Sex-Pheromone Components of the Podborer, Matsumuraeses falcana (Walshingham) (Lepidoptera : Tortricidae). Applied Entomology and Zoology 20:189–198. https://doi.org/10.1303/aez.20.189

Wan F, Yin C, Tang R, et al (2019) A chromosome-level genome assembly of *Cydia pomonella* provides insights into chemical ecology and insecticide resistance. Nat Commun 10:4237. https://doi.org/10.1038/s41467-019-12175-9

Wang H-L, Liénard MA, Zhao C-H, et al (2010) Neofunctionalization in an ancestral insect desaturase lineage led to rare Δ6 pheromone signals in the Chinese tussah silkworm. Insect Biochemistry and Molecular Biology 40:742–751. https://doi.org/10.1016/j.ibmb.2010.07.009

Witzgall P, Bengtsson M, Rauscher S, et al (2001) Identification of further sex pheromone synergists in the codling moth, *Cydia pomonella*. Entomologia Experimentalis et Applicata 101:131–141. https://doi.org/10.1046/j.1570-7458.2001.00898.x

Witzgall P, Chambon J-P, Bengtsson M, et al (1996) Sex pheromones and attractants in the Eucosmini and Grapholitini (Lepidoptera, Tortricidae). Chemoecology 7:13–23. https://doi.org/10.1007/BF01240633

Witzgall P, Kirsch P, Cork A (2010) Sex Pheromones and Their Impact on Pest Management. J Chem Ecol 36:80–100. https://doi.org/10.1007/s10886-009-9737-y

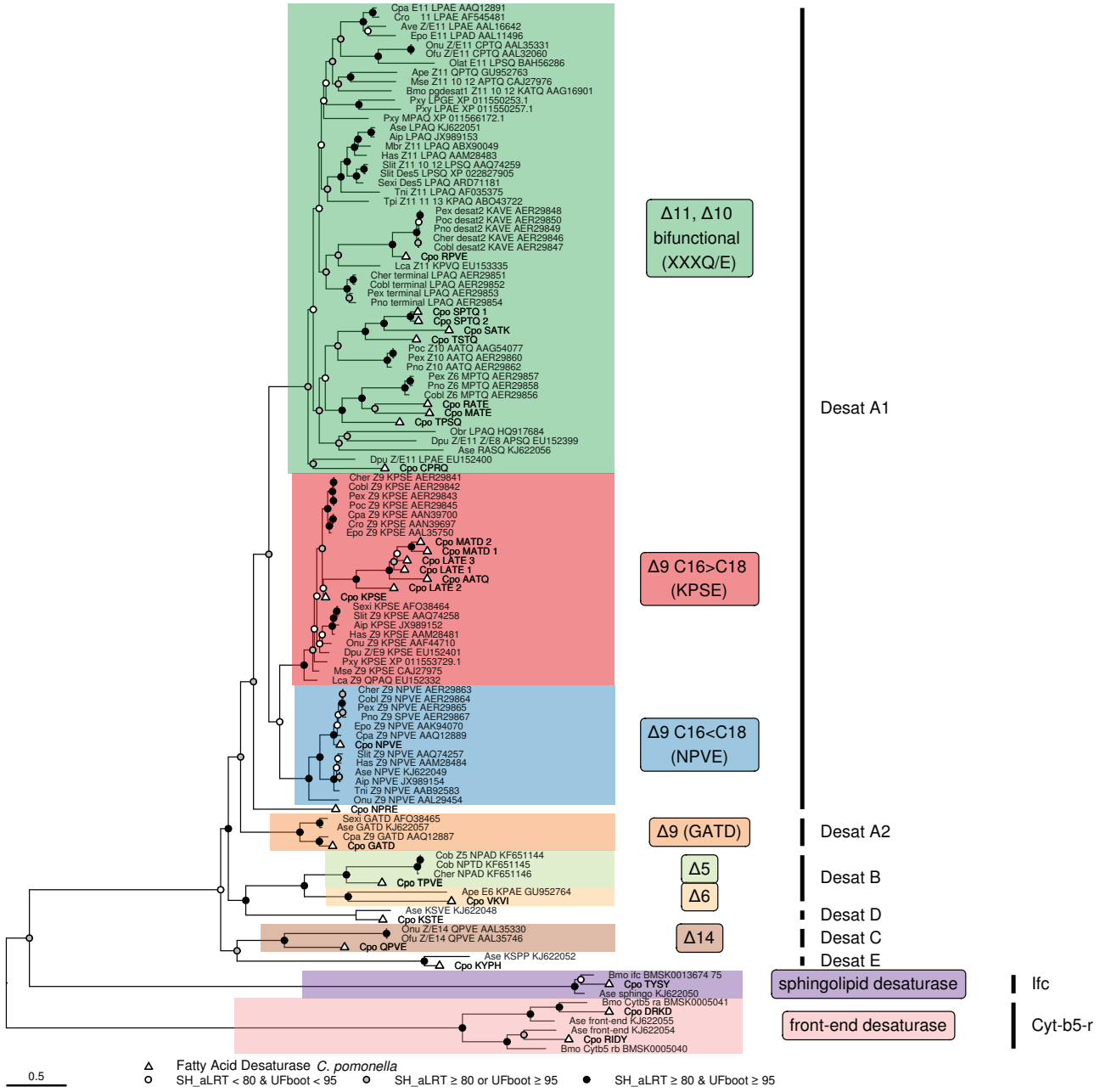
Witzgall P, Stelinski L, Gut L, Thomson D (2008) Codling Moth Management and Chemical Ecology. Annual Review of Entomology 53:503–522. https://doi.org/10.1146/annurev.ento.53.103106.093323

Xia Y-H, Zhang Y-N, Ding B-J, et al (2019) Multi-Functional Desaturases in Two *Spodoptera* Moths with ∆11 and ∆12 Desaturation Activities. J Chem Ecol 45:378–387. https://doi.org/10.1007/s10886-019-01067-3

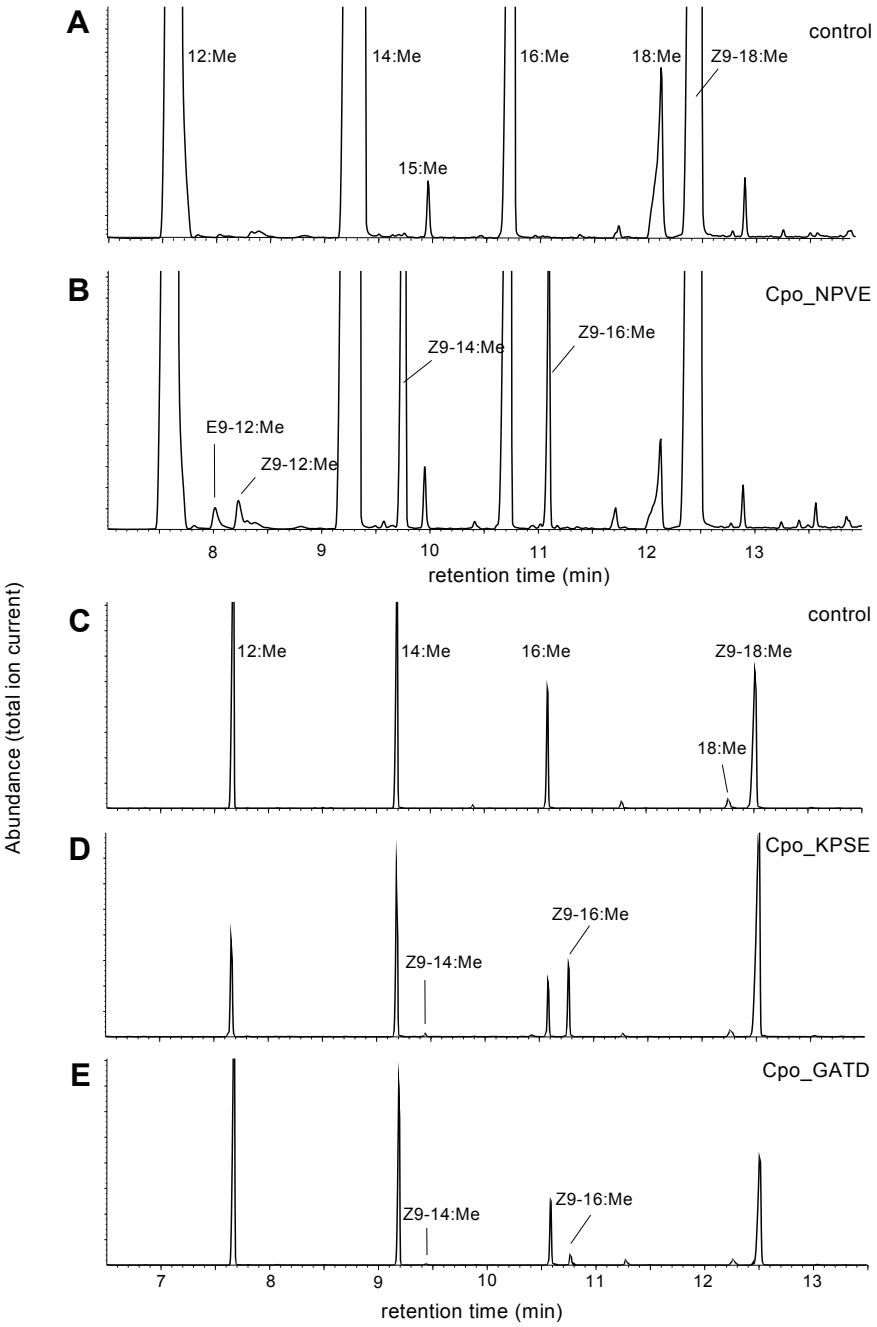
Yamaoka R, Taniguchi Y, Hayashiya K (1984) Bombykol biosynthesis from deuterium-labeled (Z)-11-hexadecenoic acid. Experientia 40:80–81. https://doi.org/10.1007/BF01959112

Yu G, Smith DK, Zhu H, et al (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28–36. https://doi.org/10.1111/2041-210X.12628

Appendix: Supplementary figures



Supplementary Figure 1: Phylogeny of lepidoptera FAD genes.
Extended version of the maximum likelihood tree displayed in Fig 4. The tree was obtained for predicted amino acid sequence of 114 FAD genes (805 aligned positions) of 28 species, with branch support values calculated from 1000 replicates using the Shimodaira-Hasegawa-like approximate ratio test (SH_aLRT) and ultrafast bootstraping (UFboot). Support values for branches are indicated by colored circles, with color assigned based SH-aLRT and UFBoot supports with 80% and 95% as thresholds of branch selection for SH-aLRT and UFBoot supports, respectively. The major constituent six subfamilies of First Desaturase (A1 to E) and two subfamilies of Front-End (Cyt-b5-r) and Sphingolipid Desaturases (Ifc), respectively, are indicated following the nomenclature proposed by Helmkampf et al. (2015). For First Desaturases, the different shades correspond to the indicated putative biochemical activities and consensus signature motif (if any). Triangles indicate sequences from *C. pomonella*. The scale bar represents 0.5 substitutions per amino acid position. Species are indicated by three- or four-letter prefixes (see Supplementary Table S3 for details). Biochemical activities (or signature motif) are indicated after the abbreviated species name, followed by the accession number.



Supplementary Figure 2: Functional characterization of desaturase activity of First-desaturases.
Total ion chromatograms of fatty acid methyl ester (FAME) products of Cu^{2+}-induced *ole1 elo1 S. cerevisae* yeast supplemented with saturated acyl precursors and transformed with (A & C) empty expression vector (control), (B) pYEX-CHT-Cpo_NPVE, (D) pYEX-CHT-Cpo_KPSE, and (E) pYEX-CHT-Cpo_GATD.