Dam Marie Inger (Orcid ID: 0000-0001-7833-0606)

Sex pheromone biosynthesis in the sugarcane borer *Diatraea saccharalis:* paving the way for biotechnological production

Running title

Diatraea saccharalis sex pheromone biosynthesis pathway elucidation

Author information

Marie Inger Dam[1]*, Bao-Jian Ding[1], Glenn P. Svensson[1], Hong-Lei Wang[1], Douglas J. Melo[1,2], Jean-Marc Lassance[3], Paulo H. G. Zarbin[2], Christer Löfstedt[1]

[1] Department of Biology, Lund University, Sölvegatan 37, SE-223 62 Lund, Sweden

[2] Departamento de Química, Universidade Federal do Paraná, CP 19081, 81531-990 Curitiba-PR, Brazil

[3] Département de gestion vétérinaire des Ressources Animales (DRA), University of Liege, Bât. B36 GIGA - Neurosciences, Quartier Hôpital, avenue Hippocrate 15, 4000 Liège 1, Belgium

* Corresponding author: marie.dam@biol.lu.se

Marie Inger Dam https://orcid.org/0000-0001-7833-0606

Bao-Jian Ding https://orcid.org/0000-0003-3322-6148

Glenn Svensson https://orcid.org/0000-0001-8112-8441

Hong-Lei Wang https://orcid.org/0000-0002-1483-5135

Douglas J. Melo https://orcid.org/0000-0003-0683-6100

Jean-Marc Lassance https://orcid.org/0000-0002-3675-6956

Paulo H. G. Zarbin https://orcid.org/0000-0001-6485-7642

Christer Löfstedt https://orcid.org/0000-0002-3116-6922

Abstract

BACKGROUND: The sugarcane borer *Diatraea saccharalis* (Lepidoptera) is a key pest on sugarcane and other grasses in the Americas. Biological control as well as insecticide treatments are used for pest management, but economic losses are still significant. The use of female sex pheromones for mating disruption or mass trapping in pest management could be established for this species, provided that economical production of pheromone is available.

RESULTS: Combining *in vivo* labelling studies, differential expression analysis of transcriptome data and functional characterisation of insect genes in a yeast expression system, we reveal the biosynthetic pathway and identify the desaturase and reductase enzymes involved in the biosynthesis of the main pheromone component (9*Z*,11*E*)-hexadecadienal, and minor components hexadecanal, (9*Z*)-hexadecenal and (11*Z*)-hexadecenal. We next demonstrate heterologous production of the corresponding alcohols of the pheromone components, by expressing multiple steps of the biosynthetic pathway in yeast.

CONCLUSION: Elucidation of the genetic basis of sex pheromone biosynthesis in *D. saccharalis*, and heterologous expression in yeast, paves the way for biotechnological production of the pheromone compounds needed for pheromone-based pest management of this species.

Keywords

Lepidoptera

Desaturase

Transcriptome
Integrated pest management
Heterologous expression
Functional characterisation

1 Introduction

The sugarcane borer *Diatraea saccharalis* (Fabricius) (Lepidoptera: Crambidae) is widely distributed throughout southern USA, Central America and the tropical and subtropical zones of South America, its dispersal likely mediated by human migration and trade [1]. It is a key pest on sugarcane and maize and also uses other grasses as hosts [2]. The larvae cause damage by eating leaves and boring into stalks. In sugarcane production where 19-25 % of internodes are bored this can diminish sugar yields by 8-20 % [3-5]. The estimated annual economic loss due to pest insects on sugarcane (in addition to *D. saccharalis*, other moth genera such as *Chilo, Sesamia*, and *Scirpophaga* also cause damage to sugarcane) is more than US$ 4.5 billion in Brazil, the world´s largest sugarcane producer [4, 6, 7]. Insecticidal control of *D. saccharalis* is inefficient because the larval and pupal stages are protected inside the plant and all developmental stages are present throughout the year [8]. Integrated pest management (IPM), where the focus lies on monitoring, prevention and limited use of pesticides, optimised cultural practices using resistant cultivars, and biological control or use of sex pheromones may provide a suitable alternative solution [9]. In 2010, around 40 % of the sugarcane area in Brazil was treated by biological control, reducing pest damage by mass release of the larval parasitoid *Cotesia flavipes* (Hymenoptera: Braconidae) and the egg parasitoid *Trichogramma galloi* (Hymenoptera: Trichogrammatidae) [10, 11]. The use of sex pheromones for monitoring or mating disruption has been successfully employed for other lepidopteran pests (reviewed in Witzgall, Kirsch and Cork (2010) [12]), though not yet for *D. saccharalis*. The major female sex pheromone component of *D. saccharalis* has been identified as (9*Z*,11*E*)-hexadecadienal (Z9,E11-16:Ald), discovered by Hammond and colleagues [13, 14] and confirmed by Svatoš, Kalinová (15). Field tests with the major sex pheromone component have shown low attractiveness compared to conspecific females [16]. Three minor components identified from females, hexadecanal (16:Ald), (9*Z*)-hexadecenal (Z9-16:Ald) and (11*Z*)-hexadecenal (Z11-16:Ald), have been shown to elicit male antennal response and the more complex blends improve attraction of males in wind tunnel assays. Solid reports on field attraction to synthetic pheromone are missing, however, and suggest the need for an improved synthetic pheromone [8, 16, 17].

Studying how sex pheromones are biosynthesised by insects is critical for the development of biotechnological pheromone production for pest management [18]. Such production may have advantages over conventional chemical synthesis by being less complex, needing less specialised manufacturing infrastructure, using fewer organic solvents and expensive catalysts, and producing less toxic waste. These aspects are discussed extensively by Wang et al. 2022 [19] and the successful production of moth pheromones in yeast- and plant platforms has been demonstrated [20-22]. Moth sex pheromone biosynthesis involves genes belonging to large multigene families of fatty acyl desaturases (desaturases) and fatty acyl reductases (reductases), fatty alcohol oxidases and acyltransferases. The biosynthesis of conjugated lepidopteran sex pheromones similar to Z9,E11-16:Ald has been described in moths from several families. The main pheromone components of *Lampronia capitella* (Prodoxidae) (9*Z*,11*Z*)-tetradecadienol (Z9,Z11-14:OH), *Epiphyas postvittana* (Tortricidae) (9*E*,11*E*)-tetradecadienyl acetate (E9,E11-14:OAc), and *Spodoptera litura* (Noctuidae) (9*Z*,11*E*)-tetradecadienyl acetate (Z9,E11-14:OAc) are made by one desaturase belonging to the clade of specific lepidopteran Δ11-desaturases, making both double bonds in sequence with a chain-shortening step in between [23-25]. The major pheromone components of *Cydia pomonella* (Tortricidae) (8*E*,10*E*)-dodecadienol (E8,E10-12:OH) and *Bombyx mori* (Bombycidae) (10*E*,12*Z*)-hexadecadienol (E10,Z12-16:OH) are made by bifunctional desaturases that first introduce one double bond in the intermediate position and then turn this into the conjugated diene pheromone component [26-29]. Two different desaturases and several chain-shortening steps are involved in the biosynthesis of a pheromone component of *Dendrolimus punctatus* (Lasiocampidae), (5*Z*,7*Z*)-dodecadienol (Z5,Z7-12:OH), either using a desaturase homologous to common Δ9 acyl-CoA desaturases and a

Lepidoptera-specific Δ11-desaturase, or two Δ11-desaturases [30]. The biosynthesis of the (7*E*,9*Z*)-dodecadienyl acetate (E7,Z9-12:OAc) pheromone component of *Lobesia botrana* (Tortricidae) involves a Δ11-desaturase and chain-shortening followed by the action of an elusive desaturase that introduces a Δ7 double bond in the Z9-12:acyl pheromone precursor [31]. The biosynthetic pathways of these diunsaturated sex pheromone components may serve as hypotheses for *D. saccharalis* biosynthesis of Z9,E11-16:Ald.

In the present study, we elucidate the biosynthetic pathway towards the major female sex pheromone component of *D. saccharalis* as well as minor components hexadecanal, (9*Z*)-hexadecenal and (11*Z*)-hexadecenal using *in vivo* labelling, transcriptome analysis and functional characterisation of genes in a yeast expression system. These results can contribute to the identification of an optimized pheromone blend for *D. saccharalis* and pave the way for its biotechnological production.

2 Materials and methods

2.1 Reference- and deuterated chemicals

Reference chemicals for identification of pheromone gland (PG) compounds were purchased from Pherobank (Wijk bij Duurstede, The Netherlands), including four isomers of hexadecenal (E/Z9-16:Ald and E/Z11-16:Ald) and the four isomers of (9,11)-hexadecadienal (Z9,E11-16:Ald). Hexadecanal (16:Ald) was from our laboratory collection. The aldehydes were used to prepare corresponding acids, methyl esters and alcohols according to previously reported protocols [32-34].

Deuterium-labelled fatty acid for *in vivo* labelling were of various origin. [16,16,16-$^2H_3$]-Hexadecanoic acid (D_3-16:acid), was purchased from Larodan (Malmö, Sweden), [8,8,7,7,6,6,5,5,4,4,3,3,2,2-$^2H_{14}$]-*cis*-9-hexedecenoic acid (D_{14}-Z9-16:acid) from Cayman Chemicals (MI, USA), and [16,16,16,15,15,14,14,13,13-$^2H_9$]-*cis*-11-hexedecenoic acid (D_9-Z11-16:acid) was from our laboratory collection, its synthesis described in Löfstedt, Hansson (35). [16,16,16,15,15-$^2H_5$]-*trans*-11-Hexedecenoic acid (D_5-E11-16:acid) was prepared following Zarbin et al. 2007 (36) (Fig. S1). 1,10-Decanediol was used as starting material and subjected to monobromination with hydrobromic acid (48%) and toluene under reflux for 6 h to synthesize 10-bromodecan-1-ol. The hydroxyl group was protected with 3,4-dihydro-*2H*-pyran in dichloromethane, with *p*-toluenesulfonic acid (*p*-TSA) as catalyst at room temperature for 2 h to give 2-((10-bromodecyl)oxy)tetrahydro-*2H*-pyran. The alkyne 2-(dodec-11-yn-1-yloxy)tetrahydro-*2H*-pyran was generated coupling the bromide with lithium acetylide (ethylene diamine complexed) using dimethyl sulfoxide as solvent. The mixture was stirred for 3 h at 0 °C, then for another 12 h at room temperature. This product was converted into the anion with *n*-butyllithium in tetrahydrofuran at -78 °C and the mixture was warmed to 0 °C. [4,4,4,3,3-$^2H_5$]-1-Bromobutane (Qmx Laboratories Ltd. Dunmow, UK) was added after 30 min and the reaction was stirred overnight. The crude product was directly deprotected by *p*-TSA in methanol at room temperature for 2 h, producing [16,16,16,15,15-$^2H_5$]-hexadec-11-yn-1-ol. [16,16,16,15,15-$^2H_5$]-*trans*-11-Hexedecenol was obtained by reaction with lithium aluminium hydride in diglyme under reflux for 5 h. The oxidation of the alcohol with pyridinium dichromate in dimethylformamide for 18 h at room temperature produced D_5-E11-16:acid at a 88 % yield. All materials were purchased from Merck unless otherwise stated.

2.2 Insects, dissection and extraction of sex pheromone glands

Insects were obtained from the Entomology Department of the Superior School of Agriculture Luiz de Queiroz at University of Sao Paulo (Brazil) and reared in dedicated facilities at Lund University (Sweden) on an artificial diet of soy flour, sugar and wheat germ, under controlled conditions (23 °C; 70 % relative humidity; light:dark cycle of 16:8 h). Males and females were kept separately after the pupal stage. Time of dissection and number of sex pheromone glands (PGs) extracted for pheromone and total lipid/precursor analysis were based on pheromone titres reported in Batista-Pereira et al. (2002) [8]. Virgin females were used because

inhibition of pheromone biosynthesis occurs after mating in many species of moths, and this can involve reduction in expression level of critical biosynthetic genes [37]. Pheromone glands of 1- to 3-day-old virgin females were dissected 3-5 h into the scotophase and extracted in 15 µL heptane (Merck) per 5 glands. After 15 min of incubation, the solvent was recovered and transferred to a new vial for pheromone analysis by GC/MS. For total lipid extraction and precursor analysis, the same sample was extracted again with chloroform:methanol (2:1 v/v) (Merck) overnight at room temperature. The extract was transferred to a new vial and the solvent evaporated by a gentle stream of N_2. The sample was then subjected to base methanolysis for transformation of lipids into fatty acid methyl esters (FAMEs) [33]. Identification of double-bond position of monounsaturated compounds was done using dimethyl disulfide-derivatization (DMDS) [38], produced by using 100 µL DMDS (Merck) and 20 µL 5 % I_2 in diethyl ether (Merck), incubating [illegible] °C overnight, then mixing with 50 µL 5% aqueous sodium thiosulfate (Merck) to neutralize the reaction and 50 µL heptane to [illegible]ver the DMDS-adducts.

[illegible] Gas chromatography/mass spectrometry analysis of sex pheromone components and fatty acid precursors

Pheromone- and FAME PG extracts were analysed in splitless mode using an Agilent 5975 mass detector (Agilent Technologies, Palo Alto, CA, USA) coupled to an Agilent 6890 series gas chromatograph or an Agilent 5977B mass detector coupled to an Agilent [illegible]) series gas chromatograph, both fitted with an HP-INNOWax column (30 m x 0.25 mm i.d., 0.25 µm film thickness; J&W Scientific, Agilent Technologies, Santa Clara, CA, USA). The GC inlet was set to 250 °C and the oven program to 80 °C for 1 min, increase of 10 °C/min to 230 °C, held for 10 min. The temperature of the transfer line was 280 °C and the MS source 230 °C. DMDS-adducts were analysed using an Agilent 5975C mass detector coupled to an Agilent 7890A series gas chromatograph fitted with a HP-5MS column (30 m x 0.25 mm i.d., 0.25 µm film thickness, J&W Scientific, Agilent Technologies, Santa Clara, CA, USA). The GC inlet was set to 260 °C and the oven program to 80 °C for 2 min, increase of 15 °C/min to 140 °C, then 5 °C/min to [illegible] °C, held for 15 min. The temperature of the transfer line was 280 °C and the MS source 230 °C. Helium was used as carrier gas.

[illegible]. In vivo labelling

[illegible]ium-labelled fatty acids D_3-16:acid, D_9-Z11-16:acid, D_5-E11-16:acid, D_{14}-Z9-16:acid and DMSO as a control were used for *in vivo* labelling, to monitor incorporation into the pheromone components and the potential pheromone precursors in the pheromone biosynthetic pathway. A total of 16 µg labelled compound in a volume of 0.4 µL DMSO was topically applied to the extruded female pheromone gland and abdominal tip 1 h before extraction of the pheromone gland. Pheromone gland extraction and pheromone and precursor analyses by GC/MS were performed as described above.

2.5 Sequencing and transcriptome assembly and phylogenetic analyses

For transcriptome analysis, one sample of 30 female PGs and one sample of 32 male abdominal tips were dissected as described above and stored immediately at -80 °C prior to RNA extraction. RNA extraction was done using TRIzol reagent (Thermo Fisher) and RNA cleanup and concentration using the RNeasy Micro kit (QIAGEN), both steps following the manufacturers´ instructions. The RNA concentration was measured using a 2100 Bioanalyzer system (Agilent). Library preparation with Illumina TruSeq poly-A enrichment and 150 bp paired-end Illumina sequencing using a NovaSeq6000 system was done by SciLifeLab National Genomics Infrastructure (Stockholm, Sweden), for one female and one male replicate. FastQC v0.11.5 (bioinformatics.babraham.ac.uk/projects/fastqc) was used to assess quality of the reads, and low quality raw reads were filtered and adaptors removed using Trimmomatic v0.36 [39] and Prinseq v0.20.4 [40]. Assembly was done using data from both samples with the

Trinity software package v2.8.2 [41, 42] with default parameters, except --normalize_max_read_cov was set to 50, --min_kmer_cov to 2, --min_glue to 2 and --KMER_SIZE to 23. Completeness of the assembly was assessed with BUSCO v3.0.2b and the Insecta- and Endopterygota datasets [43]. TransDecoder v5.0.1 (github.com/TransDecoder/) was used to extract ORFs and predict protein coding regions. Differential expression between female and male tissues was estimated with RSEM v1.3.1 [44] and Trinity package scripts align_and_estimate_abundance.pl and abundance_estimates_to_matrix.pl.

Fatty acid desaturases were found using the *FA_desaturase* (PF00487) family of the Pfam domain database [45] together with HMMER v3.2.1 (hmmer.org), and fatty acid reductases and fatty alcohol oxidases were found by BLAST [46] homology search with other lepidopteran sequences.

For phylogenetic analyses, *D. saccharalis* desaturases and reductase amino acid sequences, together with other lepidopteran sequences available from GenBank (ncbi.nlm.nih.gov) (Table S2), were aligned using MAFFT [47] and scoring matrix BLOSUM62. Maximum-likelihood phylogenies were constructed using IQ-TREE with ModelFinder and ultrafast bootstrap [48-50] and visualised with the R package ggtree [51].

2.6 Cloning and yeast functional assay of insect genes

cDNA was synthesised from the female PG RNA sample with the Thermoscript RT-PCR kit (ThermoFisher) following the manufacturer´s protocol. To verify transcript sequences, all full-length gene ORFs were amplified and Sanger sequenced in house on an ABI PRISM Genetic Analyzer (Applied Biosystems). For yeast episomal expression, desaturase- and reductases were cloned using Gateway technology (Thermo Fisher) into pDONR221 followed by pYEX-CHT [52] or pYES-DEST52 vectors, and transformed into the yeast strain Δole1/Δelo1 (MATa elo1::HIS3 ole1::LEU2 ade2 his3 leu2 ura3) [53] or INV*Sc*1 (MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52) (ThermoFisher) using the *S.c.* EasyComp transformation kit (ThermoFisher) following the manufacturers´ protocols. Uracil prototrophs were selected and cultivated in medium containing 1.92 g/L dropout medium lacking uracil (Formedium), 6.7 g/L yeast nitrogen base (Merck), 0.08 g/L adenine (Merck), 1.5 % tergitol (Merck), 100 mM oleic acid (Merck) and 2 % glucose (Merck) or 2 % galactose (Merck) and 1 % raffinose (Merck). The medium also contained 1 mM 12:Me and 14:Me, except for expressions of desaturases *Dsac_NPTQ*, *Dsac_NPAQ* and *Dsac_KPSE*. A 24-72 h pre-cultivation was followed by 48-72 h incubation at 30 °C of 4-50 mL cultures, where the heterologous gene expression was induced with a final concentration of 0.5-1 mM $CuSO_4$ (Merck) and/or 2 % galactose. Cells were harvested and subjected to total lipid GC/MS analysis as described above, or as follows: extraction with 1 mL methanol:chloroform (2:1, v/v) and 1 mL 0.075 M acetic acid (Merck), followed by running the organic phase on a silica gel TLC plate (Silica gel 60, Merck) that was developed in heptane:DEE:HAc (85:15:1, v/v/v), the bands visualized by spraying water and target gel areas collected separately into 4 mL vials and extracted with 1 mL methanol:chloroform (2:1, v/v) in a sonication bath for 2 min. The extract was then centrifuged at 2,000 *g* for 1 min. The supernatant was transferred to a new vial and 1 mL 0.075 M acetic acid was added to partition the lipids into chloroform. The chloroform phase containing alcohols was transferred into a new vial and evaporated to dryness, followed by addition of 40 µL heptane and GC/MS analysis as described above. For yeast integrative expression, constructs were cloned by fusion PCR and Gateway technology into vector pCfB2875, modified from Maury et al. (2016), and transformed into INV*Sc*1 and cultivated as described above, in 25-100 mL and for four days. Total broth or only supernatant was extracted using 2 mL heptane/25 mL culture and analysed by GC/MS as described above.

3 Results

3.1 Analysis of sex pheromone gland composition

Extracts of PGs contained the previously identified sex pheromone components of *D. saccharalis* (16:Ald, Z9-16:Ald, Z11-16:Ald and Z9,E11-16:Ald) (no figure shown as the aldehyde compounds were identified in previous literature, also see most compounds in Fig. 2) and their corresponding fatty acid precursors (Fig. 1). Compounds were identified by their mass spectra and retention times relative to those of synthetic references, and by analysis of DMDS-adduct for the monounsaturated compounds (DMDS *m/z* 362, 189 and 173 for Δ7-16:Me, 362, 217 and 145 for Δ9-16:Me, and 362, 245 and 117 for Δ11-16:Me). Extracts also contained the three other isomers of Δ9,Δ11-16:Ald/acid, but no E9-16:Ald/acid or E11-16:Ald/acid was detected (see fatty acids in Fig. 1 b). The amount of Z9,E11-16:Ald was around 1 ng/PG, and the percent ratio between Z9,E11-16:Ald, Z11-16:Ald and 16:Ald was 66±9.9, 19±6.7 and 15±5.2, respectively. Z9-16:Ald amount was too low to be quantified. Using ion *m/z* 236, the percent ratio between Z,Δ11-16:Ald isomers Z,E, Z,Z and E,E was 87±2.6, 13±1.6 and 1±1.2, respectively, while isomer E,Z amount was too low to be quantified.

3.2 *In vivo* labelling for elucidation of sex pheromone biosynthetic pathway

For elucidation of the pathway for sex pheromone biosynthesis, deuterium-labelled potential pheromone precursors were used in *in vivo* labelling experiments (Fig. 2). For D_3-16:acid, label-incorporation was detected in all four 9,11-16:Ald/acid isomers, Z9-16:Ald/acid, Z11-16:Ald/acid and 16:acid (Fig. 2 a, upper panels, indicated by arrows pointing to the solid line). For D_{14}-Z9-16:acid, label-incorporation was detected in Z9,E11-16:Ald/acid, Z9,Z11-16:Ald/acid and Z9-16:Ald/acid (Fig. 2 b, upper panels, indicated by arrows pointing to the solid line). For D_9-Z11-16:acid and D_5-E11-16:acid, the only detected incorporation was in Z11-16:Ald/acid and E11-16:Ald, respectively (Fig. 2 c and 2 d, upper panels, indicated by arrows pointing to the solid line). No deuterium-label ions were seen in the control DMSO samples (Fig. 2 a-d, lower panels).

3.3 Transcriptome- and phylogenetic analyses of sex pheromone biosynthesis genes

Transcriptome sequencing yielded 390 M raw reads per sample, and 282-302 M reads per sample after quality filtering. The assembly using both female and male samples resulted in 267,839 total assembled transcripts with an average length of 831 bp, N50 length of 1,689 bp and an overall alignment rate of 77 %. Looking at ExN50 statistics, 25,333 transcripts are covered by 83 % of the and N50 length of this subset is 2,360 bp (Fig. S2). When compared to the 1,658 total BUSCO groups of the Insecta dataset, the completeness of the assembly was assessed to 93.9 % (single-copy 46.4 %, duplicated 47.5 %, fragmented 4.4 %, missing 1.7 %), and 87.6 % (single-copy 46.1 %, duplicated 41.5 %, fragmented 9.4 %, missing 3.0 %), when compared to the 2,442 total BUSCO groups of the Endopterygota dataset. All in all, these statistics are indicative of an assembly of good quality. The RNA-Seq data and assembly is available from ENA by study accession PRJEB59358, and the expression data from ArrayExpress by accession E-MTAB-12752.

The transcriptome assembly was screened for putative desaturase, reductase and alcohol oxidase genes. Twelve, ten and one full-length genes, respectively, were found by homology search using known lepidopteran genes (Table S1 and transcript assembly nucleotide sequences in Supplementary data). For four of these genes, the expression level was considered both female-biased (expression ratio female:male > 10, an arbitrary choice) and very high (> 100 TPM, choice based on that the top 1K of female transcripts in this transcriptome have expression > 100 TPM) (Fig. 3, genes *Dsac_NPAQ*, *Dsac_KPSE*, *Dsac_FAR3781* and *Dsac_FAO1*). Phylogenetic reconstruction of first desaturases with sequences from other lepidopteran species showed that *Dsac_KPSE* clustered within the putative Δ9-desaturase clade and *Dsac_NPAQ* within the Δ11-desaturase clade (Fig. 4a, bootstrap values in fig. S4a). The *Dsac_NPAQ* consensus sequence was assembled from two overlapping transcripts, resulting in a protein sequence with ten ambiguous amino acid residues in the overlap (Fig. S3). All full-length desaturase genes found in the transcriptome assembly, except for *Dsac_NPVE2*, could be confirmed by cDNA amplification and sequencing of the ORF. An exact

match to the transcriptome assembly sequence of *Dsac_NPAQ* could not be amplified from cDNA, instead the variant *Dsac_NPTQ* was amplified, different in seven amino acids residues compared to the ambiguous *Dsac_NPAQ* transcript and 33 amino acid residues compared to the *Dsac_NPAQ* (genome) sequence (Fig. S3). *Dsac_FAR3781* clustered within the pgFAR clade of known lepidopteran reductases (Fig. 4b, bootstrap values in fig. S4b). *Dsac_FAO1* shared 98.7 % amino acid similarity with FAO candidate from *Chilo suppressalis* (Lepidoptera: Crambidae), CsupFAO*_15570* identified in Xia, Ding (54), and 34 % similarity with the functionally characterised *ScFAO* from jojoba (*Simmondsia chinensis*) seeds [55].

3.4 Yeast functional assays of desaturases and reductase

Desaturases and reductases for which expression in the transcriptome analysis was considered both female-biased (female:male > 10) and relatively high (> 100 TPM), and all remaining full-length desaturases found in the transcriptome, were together with the desaturase *Dsac_NPAQ* from the published genome [56], functionally characterised in a yeast expression system. After expression, their products were methylated and analysed by GC/MS (Fig. 5). All mono- and diunsaturated fatty acids in the form of corresponding methyl esters (FAMEs) were identified by their mass spectra and retention times relative to those of synthetic reference compounds. The putative Δ9-desaturase *Dsac_KPSE* expressed in a Δole1/Δelo1 yeast background (knock-out of native yeast Δ9 desaturase and fatty acid elongase, respectively) and supplied with oleic acid (OA) during cultivation, produced (*Z*9)-hexadecenoic acid (Z9-16:acid) as the most abundant acid in the cells (Fig. 5a). Other monounsaturated acids produced by *Dsac_KPSE* were all isomers of Δ9-dodecenoic acid (E/Z9-12:acid), tetradecenoic acid (E/Z9-14:acid), and (*E*9)-hexadecenoic acid (E9-16:acid). No (*E*9)-octadecenoic acid (E9-18:acid) was detected (DMDS analysis). No monounsaturated acids were produced in the empty vector control cells. The putative Δ11-desaturases *Dsac_NPTQ* and *Dsac_NPAQ* (genome) were expressed separately, together with *Dsac_KPSE*, in the Δole1/Δelo1 yeast background, and both strains produced Z9-16:acid, Z11-16:acid and all four isomers of (9,11)-hexadecadienoic acid (Δ9,Δ11-16:acid), none of which were found in the empty vector control cells (Fig. 5 b). The NPTQ strain produced only trace amounts of Z11-16:acid and the NPAQ strain only trace amounts of the Δ9,Δ11-16:Me isomers. No E11-16:acid was detected. Expression of these two Δ11-desaturases without *Dsac_KPSE* resulted in production of Z11-16:acid only (chromatogram not shown). A WT yeast background with abundant native Z9-16:acid was used to simultaneously express the four genes *Dsac_KPSE*, *Dsac_NPTQ*, *Dsac_NPAQ* (genome) and *Dsac_FAR3781* (Fig. 5 c), and produced Z9,E11-16:OH at an average titre of 2.7 ± 0.4 mg/L in the broth of 25 mL-scale cultivations. The percent ratio between the four pheromone blend compound precursors Z9,E11-16:OH, Z11-16:OH, 16:OH and Z9-16:OH was 22±4.5, 45±6.2, 17±4.2 and 17±5.3, respectively. The percent ratio between the four Δ9, Δ11-16:OH isomers, Z,E, E,Z, Z,Z and E,E was 53±4.0, 11±1.4, 20±1.1 and 16±4.1, respectively. No difference was seen between the episomal and genome integrated expressions (chromatogram not shown). The remaining full-length desaturases expressed in the Δole1/Δelo1 yeast background, exhibited no desaturation activities compared to the control, except for *Dsac_NPVE1* which produced Z9-14:acid and Z9-16:acid (chromatograms not shown).

4 Discussion

We have shown that biosynthesis of the major *D. saccharalis* sex pheromone component, Z9,E11-16:Ald, involves two different desaturases acting sequentially: the Δ9 desaturase *Dsac_KPSE* acts on palmitic acid and produces Z9-16:acid, followed by a Δ11 desaturation catalyzed by *Dsac_NPTQ* to produce Z9,E11-16:acid (Fig. 6). The reductase *Dsac_FAR3781* converts this precursor into an alcohol. These results are significant for a prospective future biotechnological production of the *D. saccharalis* sex pheromone for use in pest management.

The demonstrated pathway is in agreement with the presence of the ubiquitous 16:acid and Z9-16:acid precursors in the PG while no E11-16:acid or E/Z10-16:acid were found. This suggests that saturated 16:acid is the saturated precursor and that the Z9 double bond in the diene is made first, instead of the E11 double bond or the Z9,E11 double bond being made by a bifunctional desaturase from a Δ10 double bond. *In vivo* labelling showed that 16:acid and Z9-16:acid precursors were incorporated into the pheromone component whereas E11-16:acid was not. The functional assay of the desaturases in the yeast background deficient in native Δ9 desaturation activity showed that *Dsac_KPSE* displays broad Δ9 desaturation activity with a preference for producing Z9-16:acid, the precursor of Z9,E11-16:Ald, and also the minor pheromone component Z9-16:Ald. *Dsac_NPTQ* displays Δ11 desaturation activity to produce the major pheromone component precursor, and the other three isomers, Z9,Z11-16:acid, E9,E11-16:acid and E9,Z11-16:acid. It also produces the precursor of pheromone component Z11-16:acid. This type of desaturation pathway is similar to that seen for the pine caterpillar moth *D. punctatus*, where a Δ9 (*Dpud9_KPSE*) and a Δ11 (*Dpud11_LPAE*) desaturase are involved in the biosynthesis of a diene pheromone component precursor Z5,E7-12:acid [30]. Our *in vivo* labelling experiments could establish that Z9-16:acid is the precursor of two of the dienes isomers seen in the PG, Z9,E11-16:acid and Z9,Z11-16:acid, while label incorporation was seen into the two isomers with an E9 double bond. Possibly these come from an E9-16:acid precursor produced by *Dsac_KPSE*, even though it is not seen in the PG, or they may more likely come from isomerisation of the Z9-16:acid double bond while the second double bond is introduced by *Dsac_NPAQ/NPTQ*. In the *in vivo* labelling experiment with E11-16:acid used as precursor, a labelled E11-16:Ald was produced even if label incorporation from this precursor into the major pheromone component was not seen, indicating that this precursor is also a suitable substrate for both a present pgFAR and fatty alcohol oxidase (FAO).

The pgFAR *Dsac_FAR3781* was shown to have a broad specificity and the ability to reduce all the *D. saccharalis* pheromone precursor components into their respective alcohols in the yeast assay. This pgFAR and the two described desaturases being part of the pheromone biosynthesis pathway is corroborated by their high expression levels and female PG biased expressions. Even if mRNA expression level does not always have an obvious biological significance and is not an undisputable indicator of protein expression level [57], transcript expression level and PG bias has in sex pheromone biosynthesis studies so far been a good indicator for involvement in the pathway [58-60]. A less expressed or not tissue biased gene candidate might be involved [61], but lower expression could also just represent a suboptimal sampling time or less discriminate tissue dissection. A putative FAO responsible for the last step in *D. saccharalis* pheromone biosynthesis, namely conversion of the alcohol precursors into their aldehyde pheromone counterparts, was identified in the transcriptome based on similarity to other putative moth FAOs [54], its high expression level and female PG bias. An assay for functional characterisation of a FAO or other aldehyde oxidases by heterologous expression in yeast and monitoring of aldehyde production has not yet been established, and may be difficult due to the toxicity of aldehydes in such a system. An alternative route for functional characterisation of such a gene is deletion of the gene in the insect.

An unambiguous *Dsac_NPAQ* transcript could not be inferred from the *de novo* transcriptome assembly, and neither could a transcript identical to the one ultimately amplified from cDNA, *Dsac_NPTQ*. The completeness of the assembly was assessed to be high at 93.9 % when comparing to the BUSCO dataset of conserved Insecta genes. Ambiguities in transcriptome assemblies can be related to allelic variation or gene duplication (paralogs), which could explain the difficulty observed in resolving this sequence that possibly correspond to two highly similar transcripts [62]. Lepidopteran desaturases are part of a large gene family of both ancient and more recent duplicated and evolved genes [23, 28]. The ambiguous sequence might be two very similar genes merged because of stringency in assembly parameters, creating fragmented or chimeric sequences. Short-read data will often not be able to resolve paralog sequences, as they will be part of the same assembly graph. Sometimes fragmented sequences can be completed by increasing the *k-mer* size in the assembly parameters, but highly similar transcripts resulting in chimeras (or lowly expressed paralogs, with low coverage) might possibly only be resolved by lowering the *k-mer* size, which in turn can result in a highly fragmented assembly [63]. If the consensus sequence for *Dsac_NPAQ* is chosen as the transcript assembly most similar to *Dsac_NPAQ*

from the available *D. saccharalis* genome assembly [56], it is 17 amino acid residues different from the *Dsac_NPTQ* sequence amplified from cDNA. It might not be possible to resolve two such highly similar sequences with similar expression levels, by short-read transcriptome sequencing and *de novo* assembly. Two such sequences would be more easily resolved if one of them had lower expression, and then a certain coverage cut-off level could be used to unambiguously assemble only the sequence with high expression level [63]. The *D. saccharalis* genome assembly presented in Borges dos Santos, Paulo Gomes Viana (56) is 87 % complete assessed by the BUSCO Insecta dataset, and does resolve two full-length genes, *Dsac_NPAQ* (genome) and *Dsac_NPTQ* (genome), differing in 35 amino acid residues and also in their intron sequences. These are highly similar to *Dsac_NPAQ* from this study´s transcriptome assembly and *Dsac_NPTQ* amplified from the transcriptome cDNA, respectively (Fig. S3). These could represent different alleles [64], or a recent duplication event creating two paralogs that are still both functional and exhibit high expression levels, where one might evolve to become redundant in future pheromone biosynthesis of *D. saccharalis*. It could also be that both paralogs contribute to the specificity of the *D. saccharalis* pheromone blend. Both versions of this desaturase expressed in yeast in this study (*Dsac_NPTQ* (cDNA) and *Dsac_NPAQ* (genome)) were functionally capable of producing the pheromone component precursors Z11-16:acid and Z9,E11-16:acid (and its isomers), although at different ratios. These sequences differ at 33 amino acid residues, possibly revealing positions in the primary sequence that are functionally redundant, but also responsible for the ratio difference. 30 out of the 33 are considered conservative substitutions, and none are within domains characterised as important for the catalytic function, such as the conserved histidine-residues, the CoA-binding site or the coordinating carboxylates [65]. 14 substitutions are within either trans-membrane domains or alpha-helices as characterised in the crystal structure of the mammalian stearoyl-CoA desaturase in Bai, McCoy (65), and 19 are in domains not structurally or functionally defined.

The alcohol precursor titres presented in this study are modest and cannot be considered immediately relevant for industrial biotechnological production and use in IPM applications. Expression of *Dsac_KPSE*, 2x *Dsac_NPTQ/NPAQ* and *Dsac_FAR3781* produced a titre of only 2.7 ± 0.4 mg/L Z9,E11-16:OH precursor in the broth of small-scale cultivations. Before a large-scale yeast production of *D. saccharalis* pheromone precursor(s) can be initiated using the identified biosynthetic pathway elements, it will be crucial to optimise the production with respect to yeast host, platform strain and fermentation conditions, possibly even enzyme engineering for more active and specific desaturases. For production of pheromones in the oleaginous yeast *Yarrowia lipolytica*, optimisation has involved manipulating the flux in fatty acid biosynthesis pathways, the copy number of pheromone biosynthesis genes and eliminating production of unwanted side-products [22, 66]. A biotechnological production could consist of heterologous synthesis of all the pheromone component precursors at a biologically relevant ratio, for trapping and monitoring purposes, but this would be difficult. The ratios of pheromone component precursors produced in yeast in this study are much different from what is seen in the *D. saccharalis* females, differences that are expected considering the two very different biological systems and their possibly incomparable gene expression levels. It would be most efficient to develop biotechnological production of the major and also most complex structure of the blend components, for possible use in mating disruption. The minor components might be produced more easily by already published biological production platforms, especially 16:acid and Z9-16:acid as they are native products in relevant heterologous platforms [22, 67]. A platform for production of Z9,E11-16:OH alone would facilitate down-stream processing, where the diene could be purified from saturated- and monounsaturated products by urea-complexation [68] if necessary. It would be difficult to isolate Z9,E11-16:acid from its isomers, and there are many examples of geometric isomers or pheromone analogues that disrupt pheromone communication by acting as behavioural antagonists [69-72].

Despite the observed deviations of the yeast extract from an optimal *D. saccharalis* pheromone blend, it might still be useful for IPM purposes, as seen for a crude extracts of the codling moth *Cydia pomonella* pheromone produced by engineered oilseed plants *Camelina sativa* [73]. In this case, traps with crude pheromone-producing plant extracts still caught males, although not at the levels of purified extracts or synthetic pheromone. Field tests with sticky traps baited with live *D. saccharalis* females were already in the 1960s shown to reduce the damage on sugarcane in Louisiana [74]. However, it has also been shown in the field that traps baited with

Z9,E11-16:Ald alone is much less attractive to males compared to live females [15], and the newly identified four-component blend [17] has shown similar inferior attractiveness (P. H. G. Zarbin, unpublished data). Research on optimisation of component ratio and dosage in trap lures may improve attractiveness in the field. The low trap attractiveness could also be explained by a pheromone blend that is still incomplete, or enhancing or synergistic effects of compounds present in the gland extract [75, 76] that do not necessarily elicit detectable antennal responses in GC-EAD analyses for this species. Pheromone blend components could remain elusive because this species show low electrophysiological responses to minor pheromone components, as observed by Batista-Pereira, Santangelo (8) to Z11-16:Ald, and also by Kalinova, Kindl (16), where > 100 ng was needed for a significant response as opposed to < 1 ng of Z9,E11-16:Ald. Biosynthesis studies such as the present could aid in elucidation of putative missing blend components, in the form of specific precursors observed in the pheromone gland and/or genes showing female-biased expression and a specialised function or producing minor by-products related to the major pheromone components. Further studies will have to be carried out to reveal any unknown components in the pheromone blend, compounds that are possibly produced in very low quantities in the pheromone gland but may be crucial in eliciting male attraction to a synthetic blend comparable to that of conspecific females.

Mating disruption has not been assessed for *D. saccharalis*, and might still work when relying only on the major pheromone component, as seen for other lepidopteran pests, for example the European grapevine moth (*Lobesia botrana*) [77]. Mating disruption based on biotechnologically-produced sex pheromones, in a green and cheap production, can be the way forward in controlling *D. saccharalis* in the vast Brazilian sugarcane fields. Half of the cultivated land area in Brazil is covered by sugarcane crops [7, 10], and it is a crop important for both sugar- and ethanol industries, smallholder farmers and as a renewable biomass [4, 78, 79]. Because *D. saccharalis* damage has a significant economic impact and is difficult to control using conventional insecticides, it is pertinent and pressing to develop IPM technologies to control this pest.

Acknowledgements

The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Thanks to Daniel Powell for advice on trancriptome assembly and Etka Yapar and Jadranka Rota for advice on phylogenetic reconstruction and visualisation.

Conflict of interest declaration

The authors declare the following competing interests: CL is co-founder of SemioPlant AB. MID, HLW, BJD and CL are inventors on patent applications WO2016/207339, WO2018/109167 and WO2018/109163, and BJD and CL on WO/2023/275374, WO/2021/123128 and EP3167070, which describe production of insect pheromones in yeast or plant.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 760798 (OLEFINE), a CAPES/STINT grant for Swedish-Brazilian collaboration (STINT BR2014-5866 and CAPES 756/2016), Formas (2015-01336, 2021-00933) and the Jörgen Lindström Scholarship Fund. Open access funding provided by Lund University.

Availability of data and material

The RNA-Seq data and assembly is available from ENA by study accession PRJEB59358, and the expression data from ArrayExpress by accession E-MTAB-12752.

Author Contribution Statement

CL, MID, and HLW conceived and designed the study. CL, GPS, and PHGZ obtained funding. MID, BJD, HLW, GPS, JML and DJM conducted the experiments and analysed data. MID drafted the manuscript with inputs from all authors. All authors read and approved the final version of the manuscript.

References

1. Francischini FJB, Cordeiro EMG, de Campos JB, Alves-Pereira A, Viana JPG, Wu X, et al. *Diatraea saccharalis* history of colonization in the Americas. The case for human-mediated dispersal. PLoS One. 2019;14(7):e0220031.
2. Roe RM, Hammond Jr. AM, Reagan TE, Hensley SD. A bibliography of the sugarcane borer, *Diatraea saccharalis* (Fabricius), 1887-1980. U.S. Department of Agriculture: Agricultural Reviews and Manuals; 1981.
3. Grimi DA, Parody B, Ramos ML, Machado M, Ocampo F, Willse A, et al. Field-evolved resistance to Bt maize in sugarcane borer (*Diatraea saccharalis*) in Argentina. Pest Manage Sci. 2018;74(4):905–13.
4. Goebel F-R, Sallam N. New pest threats for sugarcane in the new bioeconomy and how to manage them. Current Opinion in Environmental Sustainability. 2011;3(1):81-9.
5. de S. Rossato Jr JA, Costa GH, Madaleno LL, Mutton MJ, Higley LG, Fernandes OA. Characterization and impact of the sugarcane borer on sugarcane yield and quality. Agron J. 2013;105(3):643-8.
6. Oliveira C, Auad A, Mendes S, Frizzas M. Crop losses and the economic impact of insect pests on Brazilian agriculture. Crop Protect. 2014;56:50-4.
7. FAOSTAT. Sugarcane Crop statistics 2019: Food and Agriculture Organization of the United Nations (FAO); 2021 [updated 2021/01/27/. Available from: http://www.fao.org/faostat/en/#data/QC.
8. Batista-Pereira LG, Santangelo EM, Stein K, Unelius CR, Eiras AE, Correa AG. Electrophysiological studies and identification of possible sex pheromone components of Brazilian populations of the sugarcane borer, *Diatraea saccharalis*. Zeitschrift Für Naturforschung. 2002;57(7-8):753-8.
9. US-EPA. Introduction to Integrated Pest Management: United States Environmental Protection Agency; 2017 [updated 2017//08/18/. Available from: https://www.epa.gov/ipm/introduction-integrated-pest-management.
10. Parra JRP. Biological control in Brazil: an overview. Scientia Agricola. 2014;71(5):420-9.
11. Parra JRP, Botelho PSM, Pinto AdS. Biological control of pests as a key component for sustainable sugarcane production. LAB Cortez (eds), Sugarcane Bioethanol: Rand for Productivity and Sustainability, SP, Brazil. 2010:441-50.
12. Witzgall P, Kirsch P, Cork A. Sex pheromones and their impact on pest management. J Chem Ecol. 2010;36(1):80-100.
13. Carney RL, Liu ST, inventors; Zoecon Corporation, Palo Alto, California, assignee. Insect pheromone1982.
14. Hammond A, Fisher N. Abstract: The sex pheromone of the sugar cane borer, *Diatraea saccharalis* (F.), and its potential as a management tool. Sugar y Azucar. 1982;77:32.
15. Svatoš A, Kalinová B, Kindl J, Kuldová J, Hovorka O, Do Nascimento RR, et al. Chemical Characterization and Synthesis of the Major Component of the Sex Pheromone of the Sugarcane Borer *Diatraea saccharalis*. Collection of Czechoslovak Chemical Communications. 2001;66(11):1682–90.
16. Kalinova B, Kindl J, Hovorka O, Hoskovec M, Svatos A. (11*Z*)-hexadec-11-enal enhances the attractiveness of *Diatraea saccharalis* main pheromone component in wind tunnel experiments. J Appl Entomol. 2005;129(2):70-4.

17. da Silva MR, Cortés AMP, Svensson GP, Löfstedt C, Lima ER, Zarbin PHG. Identification of Two Additional Behaviorally Active Gland Constituents of Female *Diatraea saccharalis* (Fabricius) (Lepidoptera: Crambidae). Journal of the Brazilian Chemical Society. 2021;32:225-30.

18. Löfstedt C, Xia Y-H. 3 - Biological production of insect pheromones in cell and plant factories. In: Blomquist GJ, Vogt RG, editors. Insect Pheromone Biochemistry and Molecular Biology (Second Edition). London: Academic Press; 2021. p. 89-121.

19. Wang H-L, Ding B-J, Dai J-Q, Nazarenus TJ, Borges R, Mafra-Neto A, et al. Insect pest management with sex pheromone precursors from engineered oilseed plants. Nature Sustainability. 2022.

20. Ding B-J, Hofvander P, Wang H-L, Durrett TP, Stymne S, Löfstedt C. A plant factory for moth pheromone production. Nature communications. 2014;5(1):1-7.

21. Hagström ÅK, Wang H-L, Liénard MA, Lassance J-M, Johansson T, Löfstedt C. 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. 2013;12(1):125.

22. Holkenbrink C, Ding B-J, Wang H-L, Dam MI, Petkevicius K, Kildegaard KR, et al. Production of moth sex pheromones for pest control by yeast fermentation. Metab Eng. 2020;62:312-21.

23. Liénard MA, Strandh M, Hedenstrom E, Johansson T, Löfstedt C. Key biosynthetic gene subfamily recruited for pheromone production prior to the extensive radiation of Lepidoptera. BMC Evol Biol. 2008;8:270.

24. Liu W, Jiao H, Murray NC, O'Connor M, Roelofs WL. Gene characterized for membrane desaturase that produces (*E*)-11 isomers of mono- and diunsaturated fatty acids. PNAS. 2002;99(2):620-4.

25. Xia Y-H, Zhang Y-N, Ding B-J, Wang H-L, Löfstedt C. Multi-Functional Desaturases in Two *Spodoptera* Moths with Δ11 and Δ12 Desaturation Activities. J Chem Ecol. 2019;45(4):378-87.

26. Löfstedt C, Bengtsson M. Sex pheromone biosynthesis of (*E,E*)-8,10-dodecadienol in codling moth *Cydia pomonella* involves E9 desaturation. J Chem Ecol. 1988;14(3):903-15.

27. Moto K, Suzuki MG, Hull JJ, Kurata R, Takahashi S, Yamamoto M, et al. Involvement of a bifunctional fatty-acyl desaturase in the biosynthesis of the silkmoth, *Bombyx mori*, sex pheromone. PNAS. 2004;101(23):8631-6.

28. Lassance J-M, Ding B-J, Löfstedt C. Evolution of the codling moth pheromone via an ancient gene duplication. BMC Biol. 2021;19(1):83.

29. Yamaoka R, Taniguchi Y, Hayashiya K. Bombykol biosynthesis from deuterium-labeled (*Z*)-11-hexadecenoic acid. Experientia. 1984;40(1):80-1.

30. Liénard MA, Lassance JM, Wang H-L, Zhao CH, Piskur J, Johansson T, et al. Elucidation of the sex-pheromone biosynthesis producing 5,7-dodecadienes in *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae) reveals Delta 11- and Delta 9-desaturases with unusual catalytic properties. Insect Biochem Mol Biol. 2010;40(6):440-52.

31. Ding B-J, Xia Y, Wang H-L, Andersson F, Hedenström E, Gross J, et al. Biosynthesis of the Sex Pheromone Component (*E,Z*)-7,9-Dodecadienyl Acetate in the European Grapevine Moth, *Lobesia botrana*, Involving Δ11 Desaturation and an Elusive Δ7 Desaturase. J Chem Ecol. 2021;47:248-264.

32. Corey E, Schmidt G. Useful procedures for the oxidation of alcohols involving pyridinium dichromate in aprotic media. Tetrahedron Lett. 1979;20(5):399-402.

33. Bjostad LB, Roelofs WL. Sex pheromone biosynthetic precursors in *Bombyx mori*. Insect Biochemistry. 1984;14(3):275-8.

34. Corso TN, Lewis BA, Brenna JT. Reduction of fatty acid methyl esters to fatty alcohols to improve volatility for isotopic analysis without extraneous carbon. Analytical Chemistry. 1998;70(18):3752-6.

35. Löfstedt C, Hansson BS, Tòth M, Szöcs G, Buda V, Bengtsson M, et al. Pheromone differences between sibling taxa *Diachrysia chrysitis* (Linnaeus, 1758) and *D. tutti* (Kostrowicki, 1961) (Lepidoptera: Noctuidae). J Chem Ecol. 1994;20(1):91-109.

36. Zarbin PH, Lorini LM, Ambrogi BG, Vidal DM, Lima ER. Sex pheromone of *Lonomia obliqua*: daily rhythm of production, identification, and synthesis. J Chem Ecol. 2007;33(3):555-65.

37. Jurenka R. Regulation of pheromone biosynthesis in moths. Current Opinion in Insect Science. 2017;24:29-35.

38. Buser HR, Arn H, Guerin P, Rauscher S. Determination of double bond position in mono-unsaturated acetates by mass spectrometry of dimethyl disulfide adducts. Analytical Chemistry. 1983;55(6):818-22.

39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

40. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27(6):863-4.

41. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-U130.

42. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols. 2013;8(8):1494-512.

43. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-2.

44. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323.

45. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2020;49(D1):D412-D9.

46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

47. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059-66.

48. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2014;32(1):268-74.

49. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587-9.

50. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2017;35(2):518-22.

51. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8(1):28-36.

52. Patel O, Satchell J, Baell J, Fernley R, Coloe P, Macreadie I. Inhibition studies of sulfonamide-containing folate analogs in yeast. Microb Drug Resist. 2003;9(2):139-46.

53. Schneiter R, Tatzer V, Gogg G, Leitner E, Kohlwein SD. Elo1p-dependent carboxy-terminal elongation of C14:1Delta(9) to C16:1Delta(11) fatty acids in *Saccharomyces cerevisiae*. J Bacteriol. 2000;182(13):3655-60.

54. Xia Y-H, Ding B-J, Dong S-L, Wang H-L, Hofvander P, Löfstedt C. Release of moth pheromone compounds from Nicotiana benthamiana upon transient expression of heterologous biosynthetic genes. BMC Biol. 2022;20(1):80.

55. Rajangam AS, Gidda SK, Craddock C, Mullen RT, Dyer JM, Eastmond PJ. Molecular Characterization of the Fatty Alcohol Oxidation Pathway for Wax-Ester Mobilization in Germinated Jojoba Seeds Plant Physiol. 2012;161(1):72-80.

56. Borges dos Santos L, Paulo Gomes Viana J, José Biasotto Francischini F, Victoria Fogliata S, L. Joyce A, de Souza A P, et al. A first draft genome of the Sugarcane borer, Diatraea saccharalis. [version 1; peer review: 1 approved with reservations]. F1000Research. 2020;9:1269.

57. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Scientific Reports. 2015;5(1):10775.

58. Strandh M, Johansson T, Ahrén D, Löfstedt C. Transcriptional analysis of the pheromone gland of the turnip moth, *Agrotis segetum* (Noctuidae), reveals candidate genes involved in pheromone production. Insect Mol Biol. 2008;17(1):73-85.

59. Antony B, Soffan A, Jakše J, Alfaifi S, Sutanto KD, Aldosari SA, et al. Genes involved in sex pheromone biosynthesis of *Ephestia cautella*, an important food storage pest, are determined by transcriptome sequencing. BMC Genomics. 2015;16(1):532.

60. Zhang S, Liu X, Zhu B, Yin X, Du M, Song Q, et al. Identification of differentially expressed genes in the pheromone glands of mated and virgin *Bombyx mori* by digital gene expression profiling. Plos one. 2014;9(10):e111003.

61. Liénard MA, Wang H-L, Lassance J-M, Löfstedt C. Sex pheromone biosynthetic pathways are conserved between moths and the butterfly *Bicyclus anynana*. Nature Communications. 2014;5(1):3957.

62. Razo-Mendivil FG, Martínez O, Hayano-Kanashiro C. Compacta: a fast contig clustering tool for de novo assembled transcriptomes. BMC Genomics. 2020;21(1):148.

63. Gruenheit N, Deusch O, Esser C, Becker M, Voelckel C, Lockhart P. Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. BMC Genomics. 2012;13(1):92.

64. Lassance JM, Groot AT, Liénard MA, Antony B, Borgwardt C, Andersson F, et al. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature. 2010;466(7305):486-9.

65. Bai Y, McCoy JG, Levin EJ, Sobrado P, Rajashankar KR, Fox BG, et al. X-ray structure of a mammalian stearoyl-CoA desaturase. Nature. 2015;524(7564):252-6.

66. Petkevicius K, Koutsoumpeli E, Betsi PC, Ding B-J, Kildegaard KR, Jensen H, et al. Biotechnological production of the European corn borer sex pheromone in the yeast *Yarrowia lipolytica*. Biotechnology Journal. 2021;16(6):2100004.

67. Xia Y-H, Ding B-J, Wang H-L, Hofvander P, Jarl-Sunesson C, Löfstedt C. Production of moth sex pheromone precursors in *Nicotiana spp*.: a worthwhile new approach to pest control. J Pest Sci. 2020;93(4):1333-46.

68. Hayes DG, Bengtsson YC, Van Alstine JM, Setterwall F. Urea complexation for the rapid, ecologically responsible fractionation of fatty acids from seed oil. J Am Oil Chem Soc. 1998;75(10):1403-9.

69. Wang F-M, Shen Z-J, Schal C, Zhu JJ, Zhou G-X, Wang Y-L, et al. Pheromone antagonism in *Plutella xylostella* (Linnaeus) by sex pheromones of two sympatric noctuid moths. Pest Manage Sci. 2022;78(1):379-87.

70. Juárez ML, Ruiz MJ, Fernández PC, Goane L, Villagrán ME, Arce OEA, et al. Communication interference in sympatrically occurring moth species. Entomol Exp Appl. 2016;158(1):25-33.

71. Eizaguirre M, Albajes R, López C, Sans A, Gemeno C. Inhibition of pheromone response in Sesamia nonagrioides by the pheromone of the sympatric corn borer, *Ostrinia nubilalis*. Pest Manage Sci. 2007;63(6):608-14.

72. Witzgall P, Bäckman A-C, Svensson M, Koch U, Rama F, El-Sayed A, et al. Behavioral observations of codling moth, *Cydia pomonella*, in orchards permeated with synthetic pheromone. BioControl. 1999;44(2):211-37.

73. Xia Y-H, Wang H-L, Ding B-J, Svensson GP, Jarl-Sunesson C, Cahoon EB, et al. Green Chemistry Production of Codlemone, the Sex Pheromone of the Codling Moth (*Cydia pomonella*), by Metabolic Engineering of the Oilseed Crop Camelina (Camelina sativa). J Chem Ecol. 2021;47(12):950-67.

74. Hammond AM, Hensley SD. The sugarcane borer sex attractant. Entomophaga. 1971;16(2):159-64.

75. Meier LR, Zou Y, Millar JG, Mongold-Diers JA, Hanks LM. Synergism between Enantiomers Creates Species-Specific Pheromone Blends and Minimizes Cross-Attraction for Two Species of Cerambycid Beetles. J Chem Ecol. 2016;42(11):1181-92.

76. Hall DR, Farman D, Domínguez JC, Pajares JA. Female Sex Pheromone of the Cone Moth, *Dioryctria mendacella*: Investigation of Synergism between Type I and Type II Pheromone Components. J Chem Ecol. 2017;43(5):433-42.

77. Ioriatti C, Anfora G, Tasin M, De Cristofaro A, Witzgall P, Lucchi A. Chemical Ecology and Management of *Lobesia botrana* (Lepidoptera: Tortricidae). J Econ Entomol. 2011;104(4):1125-37.

78. Dias HB, Sentelhas PC. Sugarcane yield gap analysis in Brazil – A multi-model approach for determining magnitudes and causes. Sci Total Environ. 2018;637-638:1127-36.

79. da Cunha Borges Filho R, Sturza VS, Bernardi D, da Cunha US, Pinto AS, dos Anjos e. Silva SD, et al. Population Dynamics of Pests and Natural Enemies on Sugar Cane Grown in a Subtropical Region of Brazil. Fla Entomol. 2019;102(3):526–30.

80. Knipple DC, Rosenfield C-L, Nielsen R, You KM, Jeong SE. Evolution of the integral membrane desaturase gene family in moths and flies. Genetics. 2002;162(4):1737-52.

Figure legends

Fig. 1 a Total ion chromatogram of GC/MS analysis of total fatty acid composition from female *Diatraea saccharalis* pheromone gland. 12:Me, methyl dodecanoate; 14:Me, methyl tetradecanoate; 16:Me, methyl hexadecanoate; Z9-16:Me and Z11-16:Me, methyl- (9*Z*)- and (11*Z*)-hexadecenoate; Z9,E11-16:Me, (9*Z*,11*E*)-hexadecadienoate; 18:Me, methyl octadecenoate; Z9-18:Me, methyl (9*Z*)-octadecenoate; 18:2-Me, methyl octadecadienoate; 18:3-Me, methyl octadecatrienoate. No methyl- (9*E*)- or (11*E*)-hexadecenoate were detected. **b** Selected ions from the same analysis, showing 16:2-Me (hexadecadienoate) and 16:1-Me (hexadecenoate) compounds by their diagnostic ions *m/z* 266 (solid line) and 236 (dashed line), respectively.

Fig. 2 Figures of chromatograms of GC/MS selected ion monitoring (SIM) analyses of aldehydes from female *Diatraea saccharalis* pheromone gland extracts, showing incorporation from deuterated fatty acid precursors. Lower panels show control labelling with only DMSO. Left panels show mono-unsaturated compounds, right panels show di-unsaturated compounds. Unlabelled (native) compounds shown by dashed lines and corresponding deuterium-labelled compounds by solid lines and indicated by arrows. **a** Incorporation from [16,16,16-$^2H_3$] hexadecanoic acid (D_3-16:acid) into (9*Z*,11*E*)-hexadecadienal (Z9,E11-16:Ald) and its isomers, (9*Z*)-hexadecenal (Z9-16:Ald) and (11*Z*)-hexadecenal (Z11-16:Ald) (top) compared with the control treatment of DMSO alone (bottom). **b** Incorporation from [16,16,16,15,15,14,14,13,13-$^2H_9$]-*cis*-11-hexedecenoic acid (D_9-Z11-16:acid) into Z11-16:Ald, and no observed incorporation into Z9,E11-16:Ald and its isomers (top) compared with the control treatment of DMSO alone (bottom). **c** Incorporation from [8,8,7,7,6,6,5,5,4,4,3,3,2,2-$^2H_{14}$]-*cis*-9-hexedecenoic acid (D_{14}-Z9-16:acid) into Z9,E11-16:Ald, (9*Z*,11*Z*)-hexadecadienal (Z9,Z11-16:Ald) and Z9-16:Ald (top) compared with the control treatment of DMSO alone (bottom). **d** Incorporation from [16,16,16,15,15-$^2H_5$]-*trans*-11-hexedecenoic acid (D_5-E11-16:acid) into E11-16:Ald, and no observed incorporation into Z9,E11-16:Ald and its isomers (top) compared with the control treatment of DMSO alone (bottom) (the peak seen in the control at the same RT as E11-16:Ald, is not that compound). There is no native E11-16:Ald peak.

Fig. 3 Female and male tissue transcript expression levels of *Diatraea saccharalis* putative desaturase, reductase and fatty alcohol oxidase genes involved in sex pheromone biosynthesis, identified from the transcriptome assembly. **a** Expression level of full-length desaturase genes, where the naming convention follows Knipple, Rosenfield (80). *Dsac_NPAQ*-end and *Dsac_NPAQ*-start refer to two different transcripts in the assembly that was verified as one gene by amplification from cDNA and sequencing, but with a few nt differences to the transcript sequence. **b** Expression level of full-length reductase (FAR) and fatty alcohol oxidase (FAO) genes, named after their respective transcript ID´s.

Fig. 4 Phylogeny of *Diatraea saccharalis* full-length first desaturase and reductase sequences identified in transcriptome analysis, in relation to select sequences from other lepidopteran species. Desaturase (**a**) and reductases (**b**). The maximum-likelihood tree

was constructed using amino acid sequences aligned with MAFFT and scoring matrix BLOSUM62, and IQ-TREE. *D. saccharalis* sequences are indicated with a triangle, with signature motif (desaturases). Published sequences are named by species prefixes (see table S2) followed by biochemical activity or signature motif (desaturases), and accession number. *D. saccharalis* sequences are marked with a triangle. Only nodes where ultrafast bootstrap support values are < 90 % are coloured (see all bootstrap values in fig. S4). Scale bar represents 0.5 substitutions per amino acid position.

Fig. 5 Functional characterisation of *Diatraea saccharalis* candidate desaturases and reductase in a yeast expression system. Yeast cultivation was followed by GC/MS analyses of fatty acid methyl esters (FAMEs) or alcohols from the cells or cultivation broth. **a** Total ion chromatogram (TIC) of desaturase *Dsac_KPSE* in the Δole1/Δelo1 yeast background, supplied with oleic acid (OA), producing (9*Z*)-hexadecenoic acid (Z9-16:acid) and other Δ9-monounsaturated acids (top), compared with the empty vector control (bottom). **b** TIC of desaturases *Dsac_KPSE-NPAQ* (top panel) and *Dsac_KPSE-NPTQ* (middle panel) in the Δole1/Δelo1 yeast background, supplied with oleic acid (OA), producing Z9-16:acid, Z11-16:acid and (9*Z*,11*E*)-hexadecenoic acid (Z9,E11-16:acid) and its three isomers, compared with the empty vector control (bottom panel). The NPTQ strain produced only trace amounts of Z11-16:acid and the NPAQ strain only trace amounts of the four Δ9,Δ11-16:acid isomers. **c** TIC of co-expressed desaturases *Dsac_KPSE*, *Dsac_NPTQ*, *Dsac_NPAQ* (genome) and reductase *Dsac_FAR3781*, producing (9*Z*)-hexadecenol (Z9-16:OH), (11*Z*)-hexadecenol (Z11-16:OH) and (9*Z*,11*E*)-hexadecenol (Z9,E11-16:OH) and its isomers (top), compared with the empty vector control (bottom). 12:acid, dodecanoic acid; E/Z9-12:acid, dodecenoic acid; 14:acid, tetradecanoic acid; E/Z9-14:acid, tetradecenoic acid; 18:acid, octadecenoic acid; Z9-18:acid, (9*Z*)-octadecenoic acid.

Fig. 6 Pathway for biosynthesis of *Diatraea saccharalis* pheromone component fatty alcohol precursors, as shown in this study. Three genes are involved in biosynthesis of the four precursors; *Dsac_KPSE* is a Δ9 desaturase, *Dsac_NPTQ* is a Δ11 desaturase and *Dsac_FAR3781* is a reductase.

Fig. S1 Synthesis of [16,16,16,15,15-$^2H_5$]-*trans*-11-hexedecenoic acid (D_5-E11-16:acid). 1,10-decanediol (**1**) was used as starting material and subjected to monobromination with hydrobromic acid (HBr) (48%) and toluene under reflux for 6 h to synthesize 10-bromodecan-1-ol (**2**) at 91 % yield. The hydroxyl group was protected with 3,4-dihydro-*2H*-pyran (DHP) in dichloromethane (CH_2Cl_2), with *p*-toluenesulfonic acid (*p*-TSA) as catalyst at room temperature for 2 h to give 2-((10-bromodecyl)oxy)tetrahydro-*2H*-pyran (**3**) with 96 % yield. The alkyne 2-(dodec-11-yn-1-yloxy)tetrahydro-*2H*-pyran (**4**) was generated coupling (**3)** with lithium acetylide (ethylene diamine complexed) using dimethyl sulfoxide (DMSO) as solvent. The mixture was stirred for 3 h at 0 °C, then for another 12 h at room temperature, and the product was obtained at 83 % yield. In the next step (**4)** was converted into the anion with *n*-butyllithium (*n*-BuLi) in tetrahydrofuran (THF) at -78 °C and the mixture was warmed to 0 °C. [4,4,4,3,3-$^2H_5$]-1-bromobutane was added after 30 min and the reaction was stirred overnight. The crude product was directly deprotected by *p*-TSA in methanol (MeOH) at room temperature for 2 h, producing [16,16,16,15,15-$^2H_5$]-hexadec-11-yn-1-ol (**5**) in two steps with 92 % yield. [16,16,16,15,15-$^2H_5$]-*trans*-11-hexedecenol (**6**) was obtained at 72 % yield by reacting (**5)** with lithium aluminum hydride ($LiAlH_4$) in diglyme under reflux for 5 h. The oxidation of (**6)** with pyridinium dichromate (PDC) in dimethylformamide (DMF) for 18 h at room temperature yielded D_5-E11-16:acid (**7**) at 88 % yield.

Fig. S2 ExN50 statistics for the transcriptome assembly. N50 in bp by read expression percentile (solid line, left y-axis) and number of transcripts by read expression percentile (dashed line, right y-axis). Vertical dotted line goes through the maximum of the ExN50 curve, showing that 25,333 transcripts are covered by 83 % of reads and N50 length of this assembly subset is 2360 bp.

Fig. S3 Nucleotide alignment (**a**) and protein translation alignment (**b**) of assembled transcripts *Dsac_NPAQ-start, Dsac_NPAQ-end,* amplified cDNA sequence *Dsac_NPTQ Dsac_NPAQ* (genome) and *Dsac_NPTQ* (genome). Highlights show disagreements. The three first from this study, the two latter from Borges dos Santos, Paulo Gomes Viana (56). Alignment done using MAFFT.

Fig. S4 Phylogeny of *Diatraea saccharalis* full-length first desaturase and reductase sequences identified in transcriptome analysis, in relation to select sequences from other lepidopteran species. Desaturase (**a**) and reductases (**b**). The maximum-likelihood tree was constructed using amino acid sequences aligned with MAFFT and scoring matrix BLOSUM62, and IQ-TREE. *D. saccharalis* sequences are indicated with a triangle, with signature motif (desaturases). Published sequences are named by species prefixes (see table S2) followed by biochemical activity or signature motif (desaturases), and accession number. *D. saccharalis* sequences are marked with a triangle. Numbers at nodes represent branch support as ultrafast bootstrap and scale bar represents 0.5 substitutions per amino acid position.

Fig1.jpg

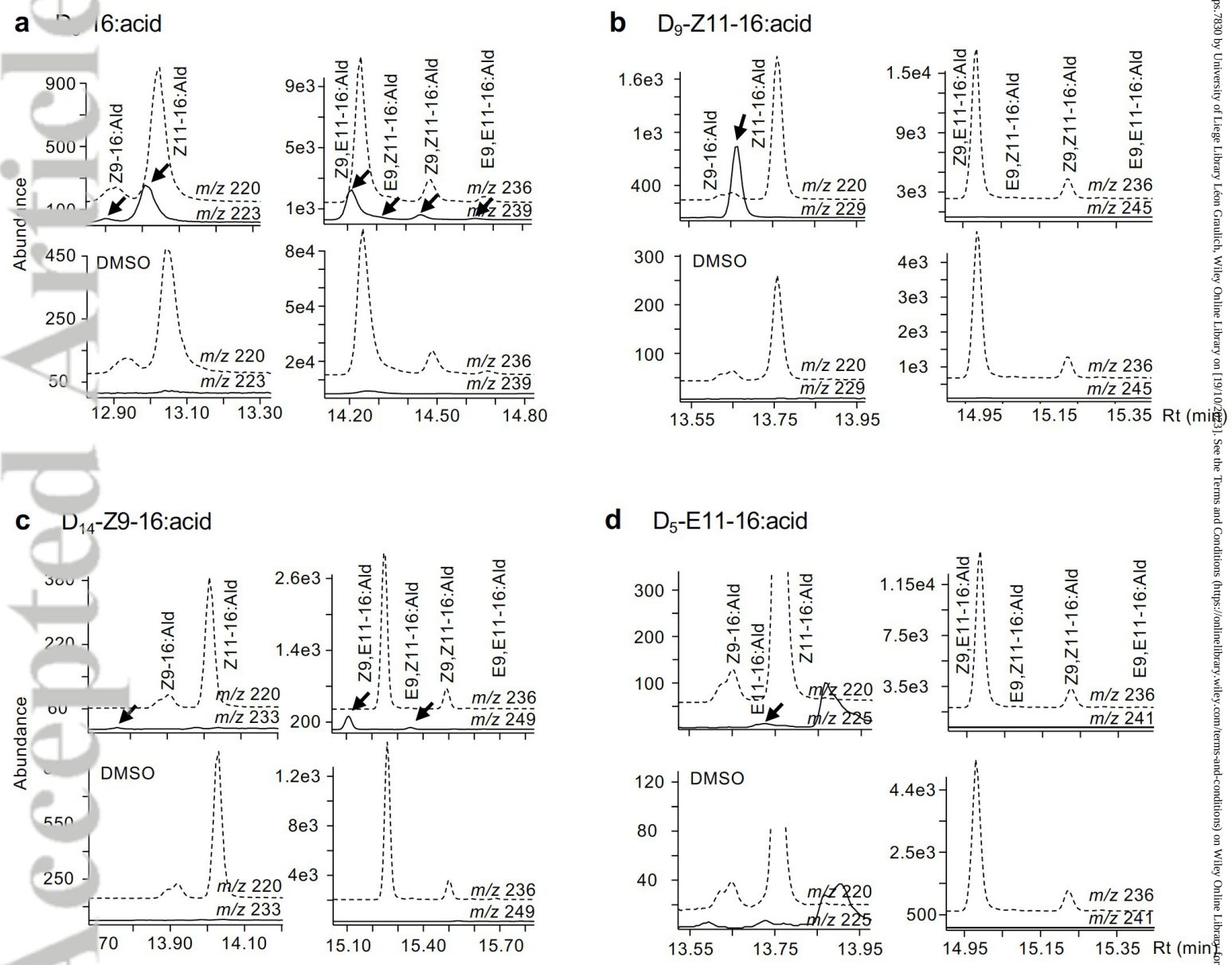

Fig2.jpg

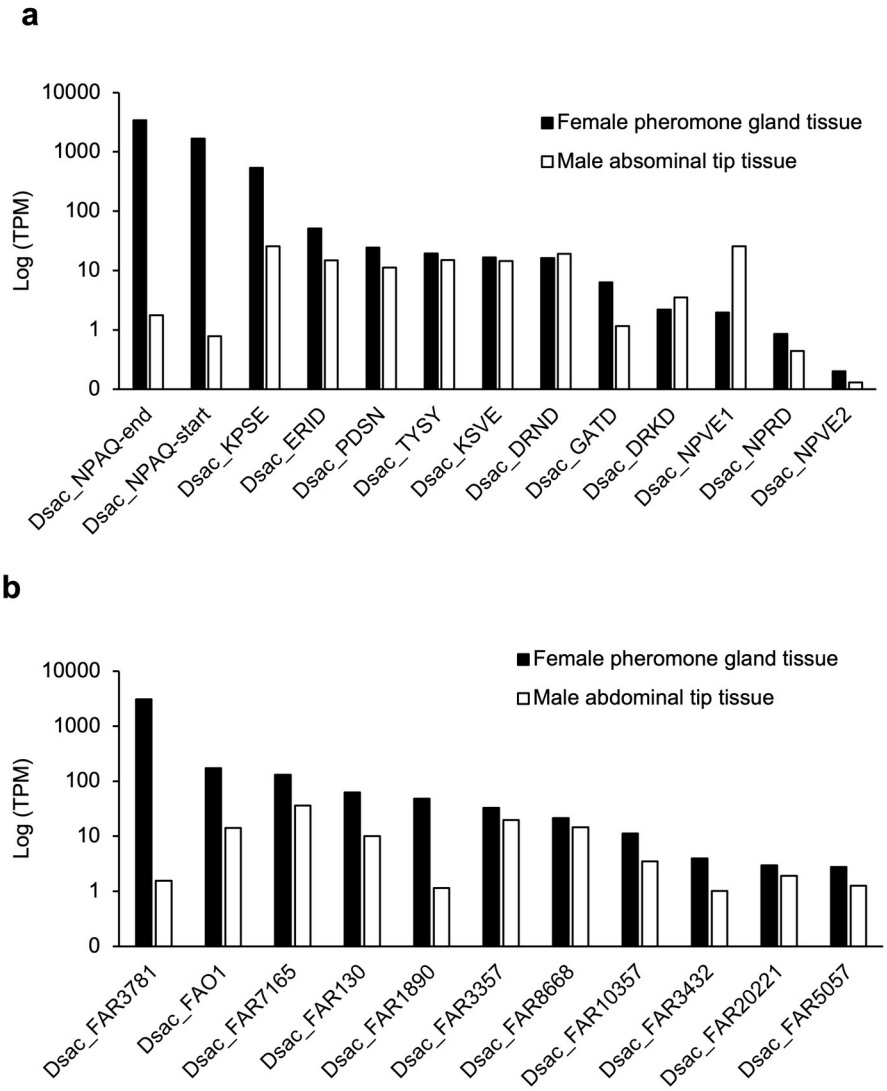

Fig3.jpg

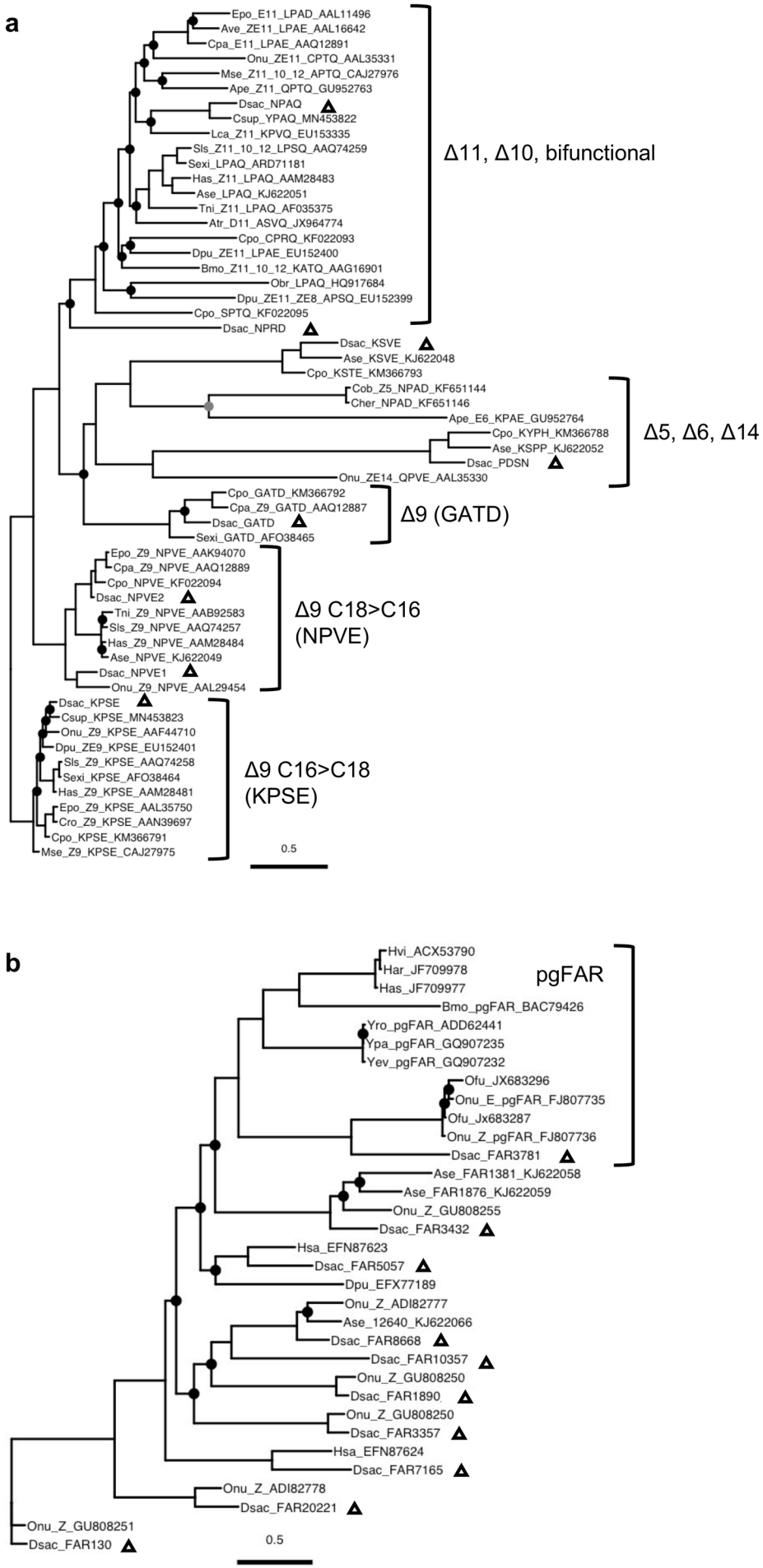

Fig4.jpg

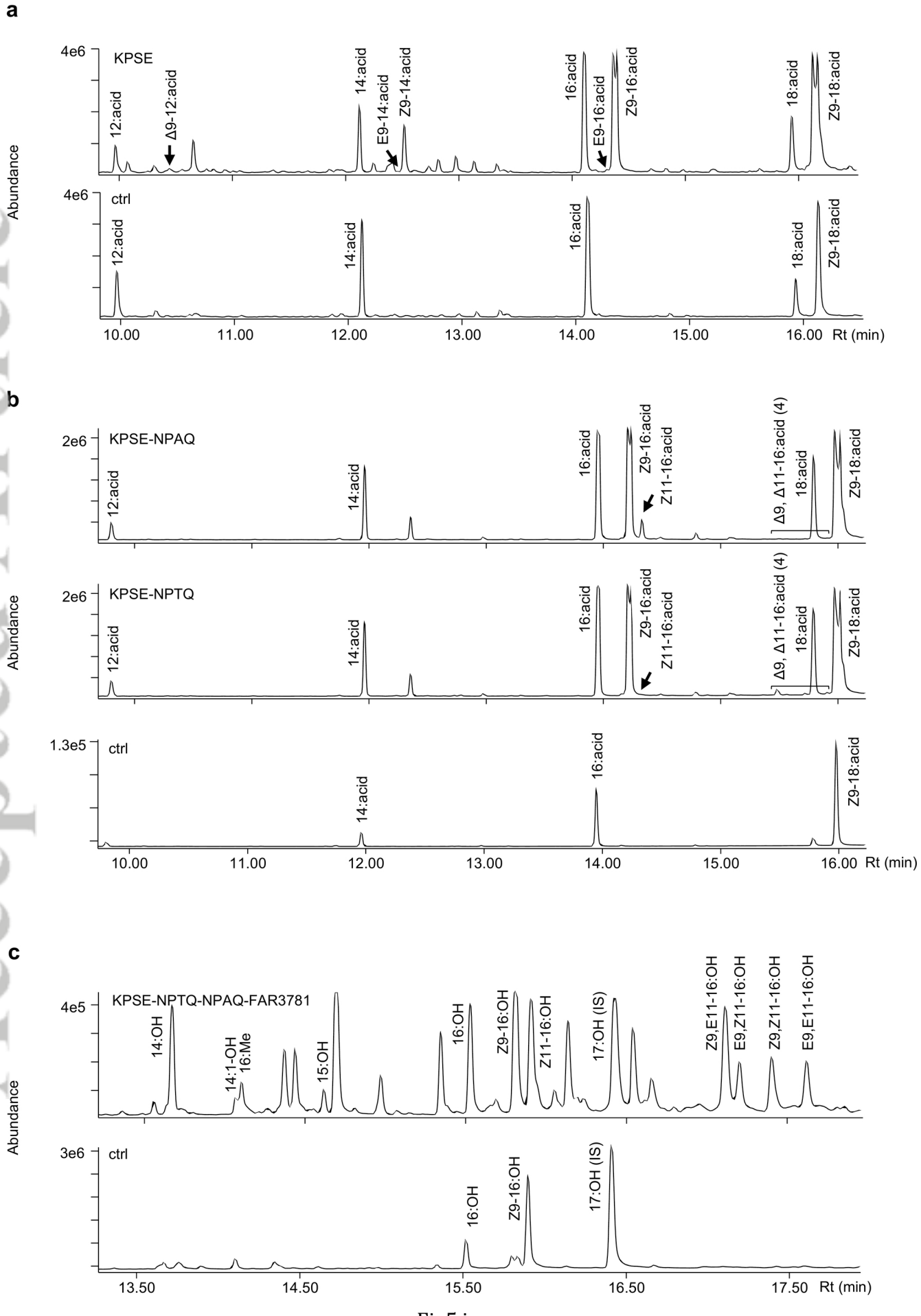

Fig5.jpg

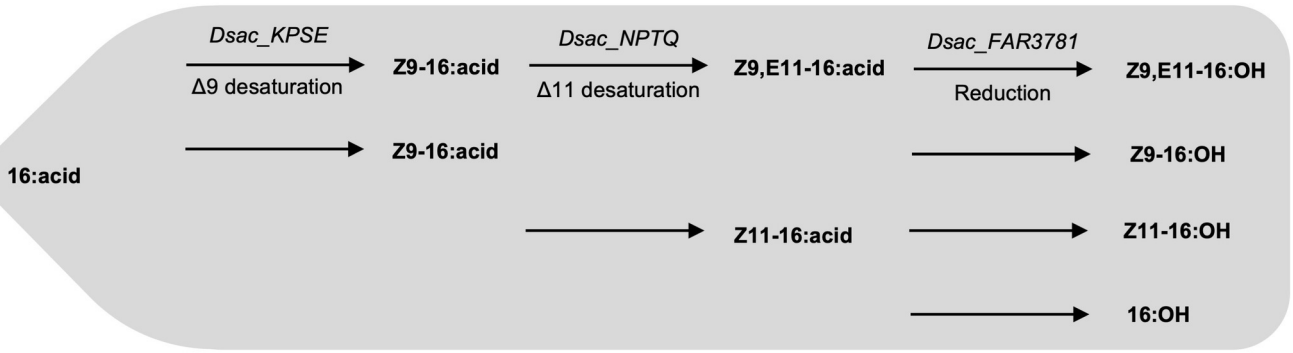

Fig6.jpg