

Lipases and carboxylesterases are involved in interspecific pheromone differences between two moth species

Arthur de Fouchier (✉ arthurdefouchier@gmail.com)

Sorbonne Université <https://orcid.org/0000-0003-0673-7380>

Elise Fruitet

Univerty of Amsterdam

Rik Lievers

University of Amsterdam (UvA)

Peter Kuperus

Univerty of Amsterdam

Jennifer Emerson

North Carolina State University

Fred Gould

NC State

David Heckel

Max Planck Institute for Chemical Ecology <https://orcid.org/0000-0001-8991-2150>

Astrid Groot

<https://orcid.org/0000-0001-9595-0161>

Article

Keywords:

Posted Date: January 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1156799/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on November 18th, 2023. See the published version at <https://doi.org/10.1038/s41467-023-43100-w>.

Abstract

Moth sex pheromones are a classical model for studying sexual selection. Females produce a species-specific pheromone blend that attracts males. Revealing the enzymes involved in the interspecific variation in blend composition is key for understanding the evolution of these sexual communication systems. The nature of the enzymes involved in the variation of acetate esters, which are prominent compounds in moth pheromone blends, remains unclear. We identified enzymes involved in acetate metabolism in two closely related species: *Heliothis (Chloridea) subflexa* and *H. (C.) virescens*, which differ in production of acetate esters. Through comparative transcriptomic analyses and CRISPR/Cas9 knockouts, we showed that two lipases and two esterases induce lower levels of acetate esters in female pheromones. To place our findings in an evolutionary context, we explored the molecular evolution of related lipases and esterases in Lepidoptera. Together, our results show that lipases and carboxylesterases are unexpectedly involved in tuning Lepidoptera pheromones composition.

Introduction

One of the most fascinating questions in evolutionary biology is to understand how new species arise. Sexual signals, through which individuals recognize suitable mating partners, are critical for prezygotic isolation of species. As such, sexual signals offer a seductive model to uncover evolutionary patterns that lead to speciation. Identifying genes underlying the production and detection of sexual signals opens a door to the analysis of their evolution across species.

As many other species, nocturnal moths use sex pheromones for mate recognition¹. Female moths produce a species-specific blend of a small number of pheromone components to which conspecific males are attracted². As moths are one of the most diverse groups of animals, and the sex pheromone of more than 2000 species have been identified³, moths are exemplary to understand the formation of new species through the evolution of sexual communication.

The species-specificity of moth sex pheromone blends is determined by the ratio of alcohols, aldehydes and acetate esters (hereafter also referred to as "acetates") with carbon backbones of various lengths and degrees of desaturation^{2,4}. Biochemical studies have unraveled some of the biosynthetic pathways leading to these pheromone compounds^{4,5}. Before emission, alcohols can be acetylated to acetate esters by acetyltransferases, or oxidized to aldehydes by alcohol oxidases. Acetate esterase or aldehyde reductase activities in and around the gland can convert (back) the acetate ester or aldehyde pheromone compounds into alcohols^{6,7}. The terminal steps in (de)acetylation of pheromone components determine the final species-specific pheromone composition. Surprisingly, even though the sex pheromone blend of almost all moth species contain acetate esters, the genes underlying acetate production remains elusive despite their pivotal role⁸.

To identify the gene(s) underlying acetate production, we conducted several genetic analyses with two closely related sympatric moth occurring species. These noctuid moths have recently been reclassified

from the paraphyletic genus *Heliothis* to the monophyletic genus *Chloridea*⁹ but we retain the older name here for continuity with the extensive literature on pheromones in these species. The most striking pheromone difference between these two species is the presence of (Z)-11-hexadecenyl acetate (Z11-16:OAc) and two other acetate esters (Z7-16:OAc and Z9-16:OAc) in the *H. subflexa* female pheromone, which are absent in the *H. virescens* female blend. This difference plays a critical role in the reproductive isolation of these two species, because the presence of these acetates in the pheromone blend inhibits the attraction of *H. virescens* males while increasing *H. subflexa* male attraction¹⁰⁻¹². Since these two species can also be hybridized in captivity, they are an excellent model to identify the genes underlying acetate production.

Previously, we identified two quantitative trait loci (QTL) underlying the difference in acetate production¹³⁻¹⁵. In this study, we introgressed one of these QTLs from *H. virescens* corresponding to one end of Chromosome 20 into the genomic background of *H. subflexa*. We then analyzed the transcriptomes of the sex pheromone glands of the introgressed and pure-strain females, followed by RT-qPCR. With these analyses, we found that certain lipases and esterases encoded within the introgressed QTL region were overexpressed in *H. virescens* and in the introgressed line compared to *H. subflexa*. Knock-outs of these genes through CRISPR/Cas9 correspondingly caused an increase in acetate amounts in the introgressed line, implicating these enzymes in hydrolysis of acetate esters in the pheromone gland. This reveals an unexpected critical role of specific lipases and esterases in the evolution of moth sex pheromones.

Results

Dominant effect of *H. virescens* QTL for reduced acetate level

Through continuous backcrossing one of the two major QTLs for acetate production¹³⁻¹⁵, we isolated this QTL (2.4cM at one end of Chromosome 20) into an otherwise complete *H. subflexa* genomic background. We refer to this introgressed line as DD23. Homozygous (VV) and heterozygous (VS) DD23 females showed significantly lower acetate levels than wild-type *H. subflexa* (SS) (Tukey post-hoc test, $P < 0.001$) (Figure 1). Also, the introgression of Chr20-QTL from *H. virescens* into *H. subflexa* has a dominant effect, as VS females had similarly low levels of acetates as VV (Tukey post-hoc test, $p = 0.585$).

Differential expression of lipases putatively acting on esters in the DD23 transcriptome

To determine which genes were differentially expressed in the sex pheromone glands of the different types of females, we analyzed the transcriptomes of these glands after building a reference transcriptome (see Supplementary Results). We considered differentially expressed contigs when we

found an absolute and significant log₂ fold change in expression. This first filtering revealed 89 differentially expressed contigs, none of which were annotated by Blast2Go with putative functions in pheromone biosynthesis (see Supplementary Table S1). However, among these differentially expressed contigs, five were annotated as lipases (38088_c0_seq4, 38088_c0_seq5, 38579_c0_seq3, 39873_c1_seq2 and 39873_c1_seq3) and one as putative esterase (comp39435_c1_seq17). As the GO terms predicted that the genes represented by these contigs are putatively involved in acetate degradation, we considered these contigs as candidate transcripts for explaining the phenotypic difference in acetate production between VV and SS.

In vivo expression patterns of lipases transcripts

As a first step in verifying *in silico* observations, we confirmed the sequence of candidate transcripts by Sanger sequencing (see materials). Consistent with its low expression levels in the RNAseq experiment, comp38088_c0_seq5 could not be amplified from *H. subflexa* pheromone gland cDNA. Similarly, comp38579_c0_seq3 could not be amplified from *H. virescens* pheromone gland cDNA. The sequences resulting from Sanger sequencing were named: LipX (38088_c0_seq4 ; OK556469, OK55646970), LipZ (38088_c0_seq5 ; OK556475), Est1 (comp39435_c1_seq17 ; OK556471, OK556472), Lip39873 (39873_c1_seq2 ; OK556473, OK556474) and Lip39879 (38579_c0_seq3 ; OK556476).

When we verified the differential expression levels in the RNAseq experiment with quantitative real-time PCR experiments on pheromone gland cDNA, we observed higher expression levels of LipX and Est1 in *H. virescens* and VV individuals compared to *H. subflexa* and SS females, and this difference was significant in comparing VV with *H. subflexa* and SS individuals (Games-Howell post-hoc test, $p < 0.05$, Supplementary Table S2) (Figure 2.A and 2.C). For LipZ, the expression level was significantly higher in the pheromone glands of VV and *H. virescens* females compared to SS and *H. subflexa* females (Tukey post-hoc test, $p < 0.05$, Supplementary Table S2) (Figure 2.B). For Lip39873, we observed overall very low and similar expression levels in *H. subflexa*, *H. virescens* and SS female pheromone glands (Figure 2.D) and therefore did not include this gene in CRISPR/Cas9 experiments.

Functional characterization of our candidate genes

To determine the role of the identified candidate genes in pheromone biosynthesis, we performed loss-of-function studies by generating mutant DD23 lines using CRISPR/Cas9. To avoid functional compensation between closely related enzymes, we targeted multiple genes at once in each knock-out line. Since we found two esterases in tandem in the QTL region in a draft version of the *H. subflexa* genome (data not shown), we knocked out both these esterases as well. In the esterase-only knock-out line, we found intermediate percentages of acetate levels compared to VS and SS (Games-Howell post-hoc test, $p < 0.05$, Supplementary Table S3) (Figure 3). The same pattern was observed when we log-contrasted the total amount of acetate esters to 14:Ald (to break data interdependency, see Materials and Methods) (Supplementary Figure S3.C). In the lipase-only knock-out line, we found similar acetate levels as in SS and significantly lower acetate levels in VS females (Games-Howell post-hoc test, $p < 0.05$,

Supplementary Table S3). These results were similar to when we knocked out both the lipases and esterases (Figure 3 and Supplementary Figure S3.C).

We did not find significant differences between the five genotypes in the total amount of pheromone produced, except between VS and the esterase knock-out line, which is most probably due to one VS outlier (Supplementary Figure S3.A). In comparing differences in the three other components that affect male response (Z9-16:Ald, Z11-16:Ald, and Z11-16:OH), we found that Z9-16:Ald levels were different only between VV and SS (Supplementary Figure S3.D). However, the lipase knock-out line as well as the four-gene knock-out line had similar levels of Z11-16:Ald and Z11-16:OH as SS females, but significantly lower levels than VS females, while, the esterase knock-out line had similar levels as VS females and significantly higher levels than SS females (Supplementary Figure S3.E and S3.F). These results thus show that these lipases and to a lesser extent the esterases are involved in moth sex pheromone acetate production.

Phylogeny of insect acidic lipases

To investigate the evolution of the acidic lipase gene family, to which LipX and LipZ belong, we built a phylogeny of these genes, by exploring the transcriptomes and genomes of 16 insects (including nine Lepidoptera; Figure 4). With this analysis, we discovered that the insect acidic lipases are split across multiple order-specific supported clades (bootstrap ≥ 70). This is consistent with what has been observed in earlier work on the phylogeny of insect acidic lipases¹⁶. In this phylogeny, we observed ten Lepidoptera-specific supported clades annotated from A-J (Figure 4). Four of these clades include sequences from the DD23 *H. subflexa* pheromone gland transcriptome.

Phylogeny of LipX and LipZ orthologues in Lepidoptera and analyses of selection pressures

LipX and LipZ both belong to the C clade, which can be divided into three clades encompassing sequences from multiple Lepidoptera super-families (clades C.1 to C.3). These three clades show recent duplication events, as they contain genus or species-specific sub-clades (Figure 5.A). To determine selective pressures in the Lepidoptera lipase sequences belonging to the C clade, we built a separate phylogenetic tree using only Lepidoptera sequences (Figure 5.A). We detected positive selection at the basal node of the C.1 and C.2 subclades ($\omega > 1$, likelihood ratio tests < 0.05 , see Supplementary table S4, Supplementary Figure S1). We also observed multiple other instances of positive selection in the tree, most of them are linked with supported clades encompassing genus, or species-specific duplications.

The Noctuidea-specific clade, which includes HsubLipX, HvirLipX and HvirLipZ, encompasses two genomic sequences from *Helicoverpa armigera* (XP_021190894.1 and XP_021190895.1)¹⁸, one from *Spodoptera litura* (XP_022818761.1) and two from *Trichoplusia ni* (XP_026728530.1 and XP_026728531.1) (Figure 5.B). The LipX/Z clade also includes sequences from pheromone gland transcriptomes: one from *Agrotis segetum* (AIN34700.1), one from *Helicoverpa assulta* (ATJ44570.1) and one from *Spodoptera exigua* (ARD71223.1). Within this Noctuidae clade, we observed significant positive

selection in the clade that encompass all Heliothinae LipX and LipZ homologues (Figure 5.B). We also detected positive selection at the node of the HvirLipZ clade and its *Helicoverpa armigera* homologue.

Phylogeny of Est1 and Est2 orthologues in Lepidoptera and analyses of selection pressures

To study selective pressures in Est1 and Est2 orthologues, we built a phylogenetic tree using esterase sequences from multiple Lepidoptera and used Tsubota and Shiotsuki's carboxylesterase sequences as outliers (see Tsubota and Shiotsuki, 2010a, 2010b) (Figure 6.A). We observed multiple events of positive selection ($\omega > 1$, likelihood ratio tests $p < 0.05$, see Supplementary Table S5, Supplementary Figure S2) across the tree. Interestingly, we found a Noctuidae-specific clade composed of Est1 and Est2 homologues that are clustered in two separated sub-clades. The Est1 clade encompasses genomic sequences from *T. ni* (XP_026728424.1) and *S. litura* (XP_022818360.1), and transcriptomic sequences from *H. armigera* (ATJ44487.1) and *H. assulta* (ATJ44554.1) pheromone glands, as well as from of *S. littoralis* (ACV60233) and *Sesamia inferens* (All21982) male antennae. The Est2 clade includes genomic sequences from *T. ni* (XP_026728423.1), *S. litura* (XP_022818359) and *H. armigera* (XP_021196515.1), as well as transcripts from *Sesamia nonagrioides* (ABH01082) and *S. littoralis* (ACV60239) male antennae. We found evidence of positive selection at the base of the Est1 and Est2 clades and along the evolution of the Est1 *H. armigera* and *H. assulta* homologues.

Discussion

Enzymes degrading acetates in *Heliothis* pheromone gland

By conducting RNAseq on female sex pheromone glands from two closely related species with acetate esters (*H. subflexa*) or without them (*H. virescens*) as well as acetate QTL introgression lines (VV and SS), we identified genes involved in moth sex pheromone acetate production. Surprisingly, we found that *Heliothis* females with low acetate levels over-express transcripts of two lipases and an esterase, indicating that acetate esters can be degraded in the species without acetates. This result was unexpected, as the general assumption is that species producing acetate esters possess an acetyltransferase that non-acetate-producing species lack; so, there should be no need to degrade an acetate ester that is never made. Using a gene inactivation approach, we demonstrated that lipases and esterases indeed reduced the amount of acetate esters in the pheromone of *Heliothis* females. These observations could be explained by a model where acidic lipase and esterase enzymes hydrolyze the acetate pheromone compounds that are produced in the gland of both species.

Acetates are produced but degraded in the non-acetate producing species

Even though our results and model seem counter-intuitive, as not producing a compound seems more parsimonious than degrading it, the presence of an acetate ester hydrolysis activity in the pheromone

gland of *H. virescens* and *Choristoneura fumiferana* (Tortricidae) had been suggested before²¹⁻²³. We can think of three possible evolutionary scenarios that would have led to the selection of acetate hydrolysis enzymatic activities in pheromone glands in *Heliothis* spp. or other moths: i) A trace acetates removing scenario, where acetate hydrolysis might counterbalance the spontaneous acetylation of pheromone compounds thus ensuring that acetates are absent in the pheromone, ii) The anti-basal acetate production scenario where acetate enzymes leading to production of acetate pheromones might be involved in other physiological enzymatic pathways and cannot be under-expressed without undesirable side-effects thus leading to significant production of acetate pheromone compounds and making degradation of those compounds mandatory, iii) The fine tuning of pheromone composition scenario where acetate hydrolysis would serve to fine-tune the species-specific relative amounts of acetate-esters, alcohols and aldehydes in the pheromone blend. Below, we discuss the implications of gene families we highlighted in the evolution of moth sex pheromone biosynthesis.

Lipases and pheromone biosynthesis

So far, lipases have been rarely studied in relation to moth pheromone metabolism. The function of lipases has been mainly considered in digestion, following work in vertebrates. For example, the human gastric lipase is the most well-known acidic lipase and predominantly hydrolyzes triglycerides and cholesterol esters²⁴. The expression of gastric-type and pancreatic-type lipases is dependent on diet in the lightbrown apple moth²⁵. Lipases in general are known to act on water-insoluble lipids²⁶ which is consistent with the involvement of lipases in pheromone hydrolysis. Furthermore, there is a body of evidence supporting that lipases are involved in moth pheromone biosynthesis: 1) Lipases have been shown to be involved in the biosynthesis of long carbon chain pheromones in other insects^{27,28}. 2) In moths, the hormone PBAN activates lipase activity in the pheromone gland in some Apoditryia (Lepidoptera) species^{29,30}. 3) Lipase enzymes have been used in hydrolysis reactions of Lymantriinae moth pheromone precursors^{31,32}. 4) Du *et al.* highlighted the expression of some lipases, including three acidic lipases, in the female pheromone glands of *Bombyx mori*³³. Using RNAi, they found that two of the acidic lipases increase the proportion of bombykol in the female pheromone³³. Thus, even though our results are unexpected, there are precedents for the involvement of acidic lipases in Lepidoptera sex pheromone biosynthesis.

Duplication of LipX and LipZ and acetate production in Noctuidae

Our phylogenetic study revealed two orthologues of LipX and LipZ in some Heliothinae moths, but not in the closely related noctuid *Spodoptera* species. Genomic data support a scenario where the two lipase genes result from a duplication event that happened in a common ancestor of *H. virescens* and *H. armigera*. This duplication may have allowed a faster evolution of LipZ compared to LipX and positive selection specifically in LipZ.

In *A. segetum*, *S. exigua* and *H. assulta*, only the homologue of LipX is expressed, which matches the expression profile observed in *H. subflexa*. Interestingly, all these species display acetates in their sex pheromones³. However, in *S. Litura* no homologue of LipX is present in the pheromone gland transcriptome, while the female sex pheromone blend contains a significant amount of acetates^{34,35}. Together, these findings suggest that expression of LipX alone is not sufficient to hydrolyze the acetates in female sex pheromone glands. Intriguingly, LipZ is found in two Heliothinae species without acetates in their pheromone blend, *H. virescens* and *H. armigera*, while their close relatives with acetates (*H. subflexa* and *H. assulta*, respectively) do not appear to express a LipZ homologue in their pheromone gland^{36,37}. Thus, LipZ expression correlates with the absence of acetates in Noctuidae pheromone glands and *vice versa*. This supports that the function of LipZ evolved following the evolutionary “trace acetates removing” scenario or “anti-basal acetate production” scenario. Having information on the expression of HarmLipZ in the female pheromone gland could further support this correlation, unfortunately, Li *et al.* didn't report the expression of lipases in their work³⁷.

LipX homologues and pheromone biosynthesis in other Lepidoptera

While LipX doesn't appear sufficient to hydrolyze acetates in the pheromone gland of noctuid species, two acidic lipases of the C.1 clade have been found expressed in Lepidoptera pheromone gland transcriptomes: Px001710 and GBXH01009737.1, which respectively belong to *Plutella xylostella* (Plutellidae)³⁸ and *Cadra cautella* (Pyralidae)³⁹. As both these species have acetate esters in their pheromone blend, these expression patterns indicate a potential role of LipX Lepidoptera orthologues in pheromone metabolism. Moreover, in our phylogeny of the C clade of acidic lipases, we observed duplications in multiple super-families of Lepidoptera. These observations suggest an adaptive role of these acidic lipases in fine-tuning moth sex pheromone ratios (fine tuning of pheromone composition scenario) or to counterbalance unavoidable production of acetate pheromone compounds (anti-basal acetate production evolutionary scenario).

Acidic lipases and pheromone biosynthesis in Lepidoptera

The numerous supported clades of acidic lipases encompassing lipases from one order at a time underlines a diversification of the acidic lipase family during insects' evolution. The high number of Lepidoptera acidic lipases clades in the phylogeny supports that the diversity of these enzymes was already high in the basal species of Lepidoptera, suggesting a diversity of functions early in their evolution, which makes it plausible that of them might be involved in pheromone biosynthesis.

Esterases and pheromone biosynthesis in Lepidoptera

In contrast to lipases, carboxylesterases have been extensively studied for their potential role in degradation of acetate esters in the pheromone gland and elsewhere, mainly in male antennae⁴⁰⁻⁴⁵. Only recently, carboxylesterases have been found to possibly also play a role in pheromone biosynthesis⁴⁶.

Thus, carboxylesterases may not only be involved in the degradation of pheromone components in the context of olfactory reception, but also in the production of Lepidoptera sex pheromone blends.

Our phylogenetic analyses revealed the presence of orthologues of Est1 and Est2 in the Noctuidae clade, except for *H. assulta*, *S. inferens* and *S. nonagrioides* for which we found only one sequence, namely either Est1 or Est2. As the sequences from these three species were obtained from transcriptomic data, it is possible that only one copy is expressed, or one duplicated gene may have been lost. Based on these findings, the most parsimonious explanation is that a duplication event happened at the basal node of the Noctuidae clade, leading to positive selective pressure in both Est1 and Est2 clades. Since we could not find expression patterns of these esterases in female pheromone glands that are consistent with acetate esters in the female sex pheromone³⁷, the role of each esterases in moth sex pheromone biosynthesis remains elusive. As with LipX homologues in Lepidoptera, this favors that the function of Est1 and Est2 evolved following the “fine tuning of pheromone composition” evolutionary scenarios and “anti-basal acetate production” evolutionary scenarios.

More enzymes involved in acetate synthesis in *Heliothis virescens*

We and previous workers²³ have never found any trace of acetate esters in *H. virescens* female pheromone glands. The hypothetical acetyltransferase has eluded all previous attempts at discovery⁸, including our own. Our previous finding of two QTLs underlying acetate production⁴⁷ indicates that additional enzymes are involved in acetate production, perhaps in the second QTL. Unfortunately, even after 10 years of effort, we have never been able to introgress this second QTL region from *H. virescens* into *H. subflexa* and we are currently exploring different approaches to investigate the presence of potential hydrolyzing genes in this QTL and possible interaction effects.

Conclusion

Our study underlines the effectiveness of introgressing a major QTL into the genomic background of a parental species to conduct transcriptome and loss-of-function analyses, to identify new genes underlying a phenotype. Through these analyses, we revealed that lipases and carboxylesterases are involved in moth sex pheromone metabolism, specifically in the degradation of acetate esters. As the role of these enzyme families in moth sex pheromone biosynthesis was previously unexpected, this paves the way for new biochemical and molecular analyses to gain insights in the specific functions of these enzymes in moth sex pheromone biosynthesis. Our study also highlights that pheromone degradation needs to be considered as an important mechanism in the evolution of moth sex pheromones.

Methods

Insects rearing

Moths were lab-reared as described in detail in⁴⁸. The so-called DD23 strain was generated by hybridizing *H. virescens* and *H. subflexa*, followed by continuous backcrossing to *H. subflexa*, as described in detail in^{13–15}. This resulted in the introgression of Chr20-QTL from *H. virescens* into a complete *H. subflexa* genomic background, which led to significantly lower acetate levels in the pheromone blend¹³. In this study, we used animals homozygous (VV), heterozygous (VS) and wild-type (SS) for this QTL. As this QTL is dominant, i.e. one V-copy is enough to reduce the acetate levels¹³ (Figure 1), we used VS females in the loss-of-function analyses. Distinction between the three genotypes was made through PCR amplification of marker sequence where an EcoRI restriction site is present in V-allele and absent in the S-allele.

Pheromone extraction

For all pheromone analyses in this study, we used 2 to 4 days old virgin female moths. Females, whose glands were used in the RNAseq and the real-time quantitative PCR, were injected with pheromone biosynthesis activating neuropeptide (PBAN) 1–2 hours before gland extraction to stimulate pheromone production, following^{11,15,49}. To minimize possible side-effects of PBAN injections in our subsequent functional analysis experiments, pheromone extractions were performed only on calling females, 3-4 hours into scotophase. Pheromone gland extractions were conducted as described in detail in⁴⁸, and summarized here. Pheromone gland were cut with microdissection scissors, soaked for 20–30 min in 50 μ L of hexane containing 200 ng of pentadecane as an internal standard, and stored at -20C until analysis. For analysis, samples were reduced to 1–2 μ L under a gentle stream of N₂ and injected into a splitless inlet of a HP6890 gas chromatograph coupled with a high-resolution polar capillary column [DB-WAXetr (extended temperature range); 30 m 90.25 mm 90.5 Mm] and a flame ionization detector (FID). The area under the pheromone peak was estimated using integration software implemented in Agilent ChemStation (version B.04.03). The absolute amounts (in ng) of the eleven identified sex pheromone of *H. subflexa* pheromone blend (14:Ald, Z7-14:Ald, 16:Ald, Z7-16:Ald, Z9-16:Ald, Z11-16:Ald, Z7-16:OAc, Z9-16:OAc, Z11-16:OAc, 16:OH and Z11-16:OH; see⁵⁰) were calculated relative to our 200 ng pentadecane internal standard and divided by the total amount to get the relative amount. To break data interdependency and to correct for individual variation in the amount of pheromone produced that play a role in male response (see⁵¹), the total amount of the three acetate-esters, Z9-16:Ald, Z11-16:Ald and Z11-16:OH were divided by the amount of 14:Ald. We chose 14:Ald as divisor, because of its not affecting male response behavior⁵⁰ and was present in comparable amounts in all extracts, except SS and the esterase knock-out line (Supplementary figure S1.b and Supplementary table S3). Log transformation was performed when this corrected residuals non-normality. Samples with less than a total pheromone amount of 40 ng were excluded, because the ratios of the minor compounds could not be accurately quantified in the chromatograms in these samples.

DD23 strain pheromone gland transcriptome assembly

For our transcriptome analysis, we extracted RNA from the pheromone extracted pheromone glands of 10 females, five of VV genotype and five of SS, using Ambion™ TRIzol™ Reagent (Thermo Fisher Scientific,

Waltham, Massachusetts, USA) and Direct-zol™ RNA MiniPrep Kit (Zymo Research, Irvine, California, USA). The RNA was sequenced using Illumina HiSeq2500 by BaseClear (Leiden, The Netherlands). Raw reads of used to assemble the pheromone gland transcriptome were deposited on NCBI Sequence Read Archive under the bioproject number PRJNA493752. Resulting reads were trimmed of adapters and for a Phred quality score of 28 or more using Trimmomatic⁵². We then removed ribosomal reads using Ribopicker⁵³. We also removed reads without a mate pair. The remaining reads were normalized and *de novo* assembled using Trinity⁵⁴. We then aligned the Trinity output with the R1 reads used for assembly with Bowtie2^{55,56} and removed all contigs matching less than five reads. The resulting contigs will be further referred to as the DD23 pheromone gland unfiltered transcriptome. To improve the quality of the transcriptome, we sequentially used three tools as described in⁵⁷ and summarized here. We first filtered out the contigs without an ORF of 100 amino or more using Transdecoder (<http://transdecoder.github.io/>). We then used RSEM⁵⁸ to remove all contigs with an TPM expression value lesser than one. Finally, we clustered and removed redundant contigs using CD-HIT EST (90% similarity threshold and word size of eight)^{59,60}. To assess this reference transcriptome quality, we used BUSCO⁶¹ and Transrate⁶².

Differential expression analysis and differentially expressed transcripts annotation

Differences in expression levels were explored by aligning the reads used for the assembly onto the reference transcriptome, using RSEM⁵⁸, after which we conducted differential expression analysis, using Deseq2⁶² and edgeR⁶³ tools. We selected contigs differentially expressed between SS and VV genotypes by filtering contigs with: a ± 2 log₂ fold change in expression value (Deseq2) and an associated adjusted *P*-value (from DEseq2 and edgeR) of 0.05 or lower. We annotated these differentially expressed transcripts by blasting them against the NR NCBI database, using blastx, and attributed GO terms using Blast2GO.

Sequencing of candidate transcripts and quantitative real-time PCR

To confirm our *in silico* differential expression results, we first had to confirm the sequences of the candidate transcripts, which we did by extracting RNA from *H. virescens* and *H. subflexa* female pheromone glands, as described above. RNA was retro-transcribed using Verso cDNA Synthesis Kit (Thermo Fisher Scientific), the candidate gene transcripts were amplified using DreamTaq™ DNA Polymerase (Thermo Fisher Scientific) and primers were designed to amplify an approximately 1000 nucleotides region on cDNA from *H. virescens* and *H. subflexa* pheromone glands (see Supplementary Table S6). Amplicons were sequenced using Sanger sequencing by Marcogen (Amsterdam, the Netherlands). Sequences resulting from genomic DNA Sanger sequencing are available on Genbank, accession numbers OK556469 to OK556476. From these sequences, we designed primers for real-time PCR to get approximately 100 bp amplicons from the alleles of both species for each target gene.

To assess the *in vivo* expression of the candidate genes, we extracted RNA as described above from 5-6 pheromone-extracted pheromone glands of *H. subflexa*, *H. virescens*, SS and VV females. We used 20 µl reaction of 5x HOT FIREPol® EvaGreen® qPCR Mix Plus (ROX) (Solis BioDyne, Tartu, Estonia), 2 ng of cDNA and 1 µM of primer couples for the target genes and used RPS18 as a reference gene. For primer used, see Supplementary Table S6. All biological replicates were also technically replicated. qPCR reactions and measurements were made using an Applied Biosystem 7500 Real-Time PCR System (Thermo Fisher Scientific).

Knock-out of the candidate genes using CRISPR/Cas9

To functionally characterize the candidate genes discovered in the previous steps, CRISPR/Cas9 knock-out experiments were performed as follows. The IDT system for CRISPR/Cas9 was used (Integrated DNA Technologies). Gene-specific crRNAs are annealed with tracrRNAs and incubated with Cas9 enzyme to form ribonucleotide particles for injection into eggs. Gene-specific guide RNAs were designed for the two esterases (VVest1-T1 and VVest1-T2 for est1, VVest2-T1 and VVest2-T2) and two lipases (VVLipX-T1 and VVLipX-T2 for LipX, VVLipZ-T1 and VVLipZ-T2 for LipZ, all sequences are in Supplementary Table S6). First, VV eggs were injected with Est1-Est2 guide RNA to knock-out the two esterases. In a second step, VV and VV-Est1-Est2-ko eggs were injected with LipX-LipZ guide RNAs to obtain two lipase knock-outs and four gene knock-outs, respectively. All eggs were 30 min-1 hour old and micro-injected using a Femtojet with a solution previously loaded in home-made glass needles. The solution was prepared by first dissolving each four scRNAs (Supplementary Table S6) in 1 nmol of tracrRNA to get a final concentration on 50 µM, after which 200 pmol of this scRNA+tracrRNA mix was combined with 100 pmol of IDT Alt-R Cas9.

The gauze with the injected embryos was kept in Petri-dishes at lab temperature (20°C) and checked daily. After three days, newly hatched neonates were separated in individual cups with wheat germ/soy flour-based diet (BioServ Inc., Newark, DE, USA). To identify mutations in the genes, we performed DNA extractions on newly emerged adults by soaking one foreleg in 30 µl of 10% Chelex and 2.5 µl of Proteinase K during three hours at 56 degrees C, then 8 min at 98 degrees C and by finally freezing the mixture before usage (Adapted protocol from Biorad). Distinction between the genotypes was made through PCR amplification with gene-specific primers: VVest1-scr-Fd and VVest1-scr-R for est1, VVest2-scr-F and VVest2-scr-R for est2, VVLipX-scr-Fd and VVLipX-scr-R for LipX, and VVLipZ-scr-F and VVLipZ-scr-R for LipZ; all sequences in Supplementary Table S6. Amplification followed the Thermo Fisher Phire Hot start II 3-step protocol with 33-35 cycles of 30s (10s for each step) and an annealing temperature of 60 degrees. The reactions were performed in 10 µl with 2 µl of DNA, 2 µl PCR Buffer, 2 µl of 1 mM dNTP's, 2 µM for each primer and 0.1 µl of Phire hot start II polymerase. PCR products were analyzed by electrophoresis on 2% agarose gels in TAE buffer (40 mM Tris-acetate, 2 mM EDTA), and the resulting bands were visualized with midori green. Est1-Est2 mutants were cloned and sequenced at the Max Planck Institute for Chemical Ecology and LipX-LipZ mutants were cloned and then sequenced by MacroGen EZ-seq (Amsterdam, Netherlands) (Supplementary Table S6). Individuals with mutations were subsequently outcrossed with *H. subflexa* and their progeny was reared as above. Every generation, this

procedure was repeated to genotype all adults, while pheromone gland extractions were performed on a subset of females.

Data analysis

Data and statistical analysis were performed using R Studio version 1.0.136 with R version 3.3.2. Bartlett's test (function *bartlett.test*, package *stats*) was used to test for homogeneity of variance of the difference between the Ct observed for the reference and the target genes in the qPCR data, pheromone data of DD23 line and of the five genotypes in the loss-of-function analysis. Homogeneity of the means of the same values were tested using a Welch one-way ANOVA (function *oneway.test*, package *stats*). A Tukey or a Games-Howell post-hoc test (custom R script, <http://aoki2.si.gunma-u.ac.jp/R/src/tukey.R>) was used for the groups with homogenous or non-homogenous variance, respectively. Quantitative PCR results were expressed in target copy number relative to 1000 copies of RPS18, similarly as in⁴⁷.

Phylogeny of insect acidic lipase

For the phylogenetical analysis of the acidic lipase family across insect genomes, we used the official gene sets from multiple insect genomes, as listed in Supplementary Table S7. We also used the protein sequences encoded by the contigs of the DD23 reference pheromone glands transcriptome, as predicted by Transdecoder (<http://transdecoder.github.io/>). We blasted (blastp, e-value cutoff 0.01) amino acids sequence of the CDS of LipX and LipZ against the insects OGS and the DD23 pheromone glands proteins. Blast hits were blasted against the NCBI NR database (e-value cutoff, 0.01). All sequences from the insect OGS and DD23 hitting a lipase or an hydrolase sequence were blasted against the $\alpha\beta$ -hydrolase 6 family from the ESTHER database⁶⁴ to confirm that they belong to the acidic lipase family. To obtain a suitable outgroup, we blasted the peptide sequences of LipX and LipZ against the $\alpha\beta$ -hydrolase 6 protein sequences, except acidic lipases, from the ESTHER database. Sequences from the top seven hit from this blast were added to the sequences obtained through the previous procedure. The sequences were aligned (alignment query is available as Supplementary File S1) using MAFFT⁶⁵ (<http://mafft.cbrc.jp/alignment/server/>). Positions in the alignment that contained more than 50% of gaps were removed using the Gap Strip/Squeeze (v2.1.0, <https://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html>). With ProtTest v3.4.2⁶⁶ we identified the substitution model LG (Gamma shape parameter: 1.377 and proportion of invariable sites: 0.005.) as the best model to perform a phylogeny on this alignment using PhyML (NNI tree improvement, bootstrap x1000)⁶⁷.

Detection of positive selection

For the selection pressure analysis, we built two phylogenetic trees from Lepidoptera sequences homologous to either LipX and LipZ or Est1 and Est2. Sequences were obtained by blasting genes of interest against the Lepidoptera sequences in the Genbank NR database (blastp, e-value cutoff: 10^{-3}). The sequences hit by the previous blast were blasted against the sequences from the previous phylogenetic tree (acidic lipase tree). Sequences with best hits belonged to the C clade and the protein

sequence of LipX from *H. virescens* were selected as query for building a tree (alignment query is available as Supplementary File S1). Similar approach was used for the Est1 and Est2, except that the sequences were blasted against the sequences from *Bombyx mori* Carboxyl/cholinesterases phylogeny made by Tsubota and Shiotsuki²⁰ (alignment query is available as Supplementary File S2). Trees were built using PhyML with smart model selection.

CDS nucleotide sequences of the proteins used for building the tree were obtained from NCBI NR database, genome databases or the sequencing data of this study. These sequences were used to generate a nucleotide alignment based on the alignment used for the phylogenetic tree of lipases or esterases, using PAL2NAL⁶⁸. This nucleotide alignment and the phylogenetic tree were used as input for selective pressure analysis. We used codeML to infer episodic change in natural selection acting on protein sequence⁶⁹. We used branch site models with default parameters and following the procedure suggested by the user manual of codeML, for each tested clade, we calculated a posterior probability using the procedure described in⁷⁰.

Declarations

Competing interests

The authors declare no competing interests.

Author contributions

A.F., E.F., F. G., A.T.G. and D.G.H. conceptualized the study. A.F., E.F., R.L., P.K. and J.E. performed the experiments. A.F., E.F. analyzed the data. A.F., E.F., A.T.G. and D.G.H. wrote and edited the paper. All authors approved the submitted manuscript.

References

1. Johansson, B. G. & Jones, T. M. The role of chemical communication in mate choice. *Biol. Rev.* **82**, 265–289 (2007).
2. Ando, T., Inomata, S. & Yamamoto, M. Lepidopteran sex pheromones. *Top. Curr. Chem.* **239**, 51–96 (2004).
3. El-Sayed, A. M. The Pherobase: Database of Pheromones and Semiochemicals. <http://www.pherobase.com> (2021).
4. Tillman, J. A., Seybold, S. J., Jurenka, R. A. & Blomquist, G. J. Insect pheromones - An overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* **29**, 481–514 (1999).
5. Jurenka, R. A. Insect pheromone biosynthesis. *Top. Curr. Chem.* **239**, 97–132 (2004).

6. Morse, D. & Meighen, E. A. Differences in oxidase and esterase activities involved in pheromone biosynthesis in two species of *Choristoneura*. *J. Chem. Ecol.* **16**, 1485–1493 (1990).
7. Groot, A. T., Dekker, T. & Heckel, D. G. The Genetic Basis of Pheromone Evolution in Moths. *Annu. Rev. Entomol.* **61**, annurev-ento-010715-023638 (2016).
8. Ding, B. & Löfstedt, C. Analysis of the *Agrotis segetum* pheromone gland transcriptome in the light of sex pheromone biosynthesis. *BMC Genomics* **16**, 711 (2015).
9. Pogue, M. G. Revised status of *Chloridea* Duncan and (Westwood), 1841, for the *Heliiothis virescens* species group (Lepidoptera: Noctuidae: Heliiothinae) based on morphology and three genes. *Syst. Entomol.* **38**, 523–542 (2013).
10. Vickers, N. J. & Baker, T. C. Chemical communication in heliothine moths. VII. Correlation between diminished responses to point-source plumes and single filaments similarly tainted with a behavioral antagonist. *J. Comp. Physiol. - Sens. Neural Behav. Physiol.* **180**, 523–536 (1997).
11. Groot, A. T. *et al.* Experimental evidence for interspecific directional selection on moth pheromone communication. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5858–5863 (2006).
12. Groot, A. T. *et al.* Differential attraction of *Heliiothis subflexa* males to synthetic pheromone lures in Eastern US and Western Mexico. *J. Chem. Ecol.* **33**, 353–368 (2007).
13. Groot, A. T. *et al.* Introgressing pheromone QTL between species: Towards an evolutionary understanding of differentiation in sexual communication. *J. Chem. Ecol.* **30**, 2495–2514 (2004).
14. Sheck, A. L. *et al.* Genetics of sex pheromone blend differences between *Heliiothis virescens* and *Heliiothis subflexa*: A chromosome mapping approach. *J. Evol. Biol.* **19**, 600–617 (2006).
15. Groot, A. T. *et al.* QTL analysis of sex pheromone blend differences between two closely related moths: Insights into divergence in biosynthetic pathways. *Insect Biochem. Mol. Biol.* **39**, 568–577 (2009).
16. Horne, I., Haritos, V. S. & Oakeshott, J. G. Comparative and functional genomics of lipases in holometabolous insects. *Insect Biochem. Mol. Biol.* **39**, 547–567 (2009).
18. Pearce, S. L. *et al.* Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol.* **15**, 63 (2017).
19. Tsubota, T. & Shiotsuki, T. Genomic and phylogenetic analysis of insect carboxyl/cholinesterase genes. *J. Pestic. Sci.* **35**, 310–314 (2010).
20. Tsubota, T. & Shiotsuki, T. Genomic analysis of carboxyl/cholinesterase genes in the silkworm *Bombyx mori*. *BMC Genomics* **11**, 377 (2010).

21. Morse, D. & Meighen, E. Detection of pheromone biosynthetic and degradative enzymes in vitro. *J. Biol. Chem.* **259**, 475–480 (1984).
22. Morse, D. & Meighen, E. Pheromone biosynthesis and role of functional groups in pheromone specificity. *J. Chem. Ecol.* **12**, 335–351 (1986).
23. Teal, P. E. a. & Tumlinson, J. H. The role of alcohols in pheromone biosynthesis by two noctuid moths that use acetate pheromone components. *Arch. Insect Biochem. Physiol.* **4**, 261–269 (1987).
24. Gargouri, Y., Moreau, H. & Verger, R. Gastric lipases: biochemical and physiological studies. *Biochim. Biophys. Acta BBALipids Lipid Metab.* **1006**, 255–271 (1989).
25. Christeller, J. T., Poulton, J., Markwick, N. M. & Simpson, R. M. The effect of diet on the expression of lipase genes in the midgut of the lightbrown apple moth (*Epiphyas postvittana* Walker; Tortricidae). *Insect Mol. Biol.* **19**, 9–25 (2010).
26. Ali, Y. Ben, Verger, R. & Abousalham, A. Lipases or esterases: Does it really matter? toward a new bio-physico-chemical classification. in *Methods in Molecular Biology* vol. 861 31–51 (Humana Press, 2012).
27. Castillo, C. *et al.* Biosynthesis of ethyl oleate, a primer pheromone, in the honey bee (*Apis mellifera* L.). *Insect Biochem. Mol. Biol.* **42**, 404–416 (2012).
28. Brabcová, J. *et al.* Characterization of neutral lipase BT-1 isolated from the labial gland of *Bombus terrestris* males. *PLoS ONE* **8**, e80066 (2013).
29. Foster, S. P. Fatty acyl pheromone analogue-containing lipids and their roles in sex pheromone biosynthesis in the lightbrown apple moth, *Epiphyas postvittana* (walker). *J. Insect Physiol.* **47**, 433–443 (2001).
30. Zhang, S. D. *et al.* Molecular identification of a pancreatic lipase-like gene involved in sex pheromone biosynthesis of *Bombyx mori*. *Insect Sci.* **21**, 459–468 (2014).
31. Bianchi, D., Cabri, W., Cesti, P., Francalanci, F. & Rama, F. Enzymatic resolution of 2,3-epoxyalcohols, intermediates in the synthesis of the gypsy moth sex pheromone. *Tetrahedron Lett.* **29**, 2455–2458 (1988).
32. Muto, S. & Mori, K. Synthesis of all Four Stereoisomers of Leucomalure, Components of the Female Sex Pheromone of the Satin Moth, *Leucoma salicis*. *Eur. J. Org. Chem.* **2003**, 1300–1307 (2003).
33. Du, M. *et al.* Identification of lipases involved in PBAN stimulated Pheromone production in *Bombyx mori* using the DGE and RNAi approaches. *PLoS ONE* **7**, e31045 (2012).

34. Saveer, A. M. *et al.* Mate recognition and reproductive isolation in the sibling species *Spodoptera littoralis* and *Spodoptera litura*. *Front. Ecol. Evol.* **2**, (2014).
35. Zhang, Y. N. *et al.* Identification and expression profiles of sex pheromone biosynthesis and transport related genes in *Spodoptera litura*. *PLoS ONE* **10**, e0140019 (2015).
36. Li, Z. Q. *et al.* Transcriptome comparison of the sex pheromone glands from two sibling *Helicoverpa* species with opposite sex pheromone components. *Sci. Rep.* **5**, 9324 (2015).
37. Li, R. T. *et al.* Expressional divergences of two desaturase genes determine the opposite ratios of two sex pheromone components in *Helicoverpa armigera* and *Helicoverpa assulta*. *Insect Biochem. Mol. Biol.* **90**, 90–100 (2017).
38. Chen, D. S., Dai, J. Q. & Han, S. C. Identification of the pheromone biosynthesis genes from the sex pheromone gland transcriptome of the diamondback moth, *Plutella xylostella*. *Sci. Rep.* **7**, 16255 (2017).
39. Antony, B. *et al.* Genes involved in sex pheromone biosynthesis of *Ephestia cautella*, an important food storage pest, are determined by transcriptome sequencing. *BMC Genomics* **16**, 532 (2015).
40. Ferkovich, S. M., van Essen, F. & Taylor, T. R. Hydrolysis of sex pheromone by antennal esterases of the cabbage looper, *Trichoplusia ni*. *Chem. Senses* **5**, 33–46 (1980).
41. Prestwich, G. D. *et al.* Enzymatic processing of pheromones and pheromone analogs. *Experientia* **45**, 263–270 (1989).
42. Ishida, Y. & Leal, W. S. Rapid inactivation of a moth pheromone. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14075–14079 (2005).
43. Vogt, R. G. Molecular basis of pheromone detection in insects. in *Comprehensive Molecular Insect Science* vol. 3 753–804 (2005).
44. Merlin, C. *et al.* Antennal esterase cDNAs from two pest moths, *Spodoptera littoralis* and *Sesamia nonagrioides*, potentially involved in odourant degradation. *Insect Mol. Biol.* **16**, 73–81 (2007).
45. Durand, N., Chertemps, T. & Maibèche, M. Antennal carboxylesterases in a moth, structural and functional diversity. *Commun. Integr. Biol.* **5**, 284–6 (2012).
46. Wang, Q.-H. *et al.* Identification of genes involved in sex pheromone biosynthesis and metabolic pathway in the Chinese oak silkworm, *Antheraea pernyi*. *Int. J. Biol. Macromol.* **163**, 1487–1497 (2020).
47. Groot, A. T. *et al.* One quantitative trait locus for intra- and interspecific variation in a sex pheromone. *Mol. Ecol.* **22**, 1065–1080 (2013).
48. Lievers, R. & Groot, A. T. Disposable Polydimethylsiloxane (PDMS)-Coated Fused Silica Optical Fibers for Sampling Pheromones of Moths. *PloS One* **11**, e0161138 (2016).

49. Groot, A. T. *et al.* Effect of PBAN on pheromone production by mated *Heliothis virescens* and *Heliothis subflexa* females. *J. Chem. Ecol.* **31**, 15–28 (2005).
50. Heath, R. R., Mitchell, E. R. & Cibrian Tovar, J. Effect of release rate and ratio of (Z)-11-hexadecen-1-ol from synthetic pheromone blends on trap capture of *Heliothis subflexa* (Lepidoptera: Noctuidae). *J. Chem. Ecol.* **16**, 1259–1268 (1990).
51. Groot, A. T. *et al.* Phenotypic plasticity in sexual communication signal of a noctuid moth. *J. Evol. Biol.* **23**, 2731–2738 (2010).
52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
53. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28**, 433–435 (2012).
54. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52 (2011).
55. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
56. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).
57. Kerkvliet, J., Fouchier, A., Wijk, M. & Groot, A. T. The Bellerophon pipeline, improving de novo transcriptomes and removing chimeras. *Ecol. Evol.* 495754 (2019) doi:10.1002/ece3.5571.
58. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
59. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
60. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
61. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
62. Smith-Unna, R. D. *et al.* TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 021626 (2016).

63. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
64. Hotelier, T. *et al.* ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins. *Nucleic Acids Res.* **32**, D145-7 (2004).
65. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
66. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
67. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
68. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 609–612 (2006).
69. Xu, B. & Yang, Z. PamlX: A graphical user interface for PAML. *Mol. Biol. Evol.* **30**, 2723–2724 (2013).
70. Bielawski, J. P., Baker, J. L. & Mingrone, J. Inference of Episodic Changes in Natural Selection Acting on Protein Coding Sequences via CODEML. *Curr. Protoc. Bioinforma.* 6.15.1-6.15.32 (2016) doi:10.1002/cpbi.2.

Figures

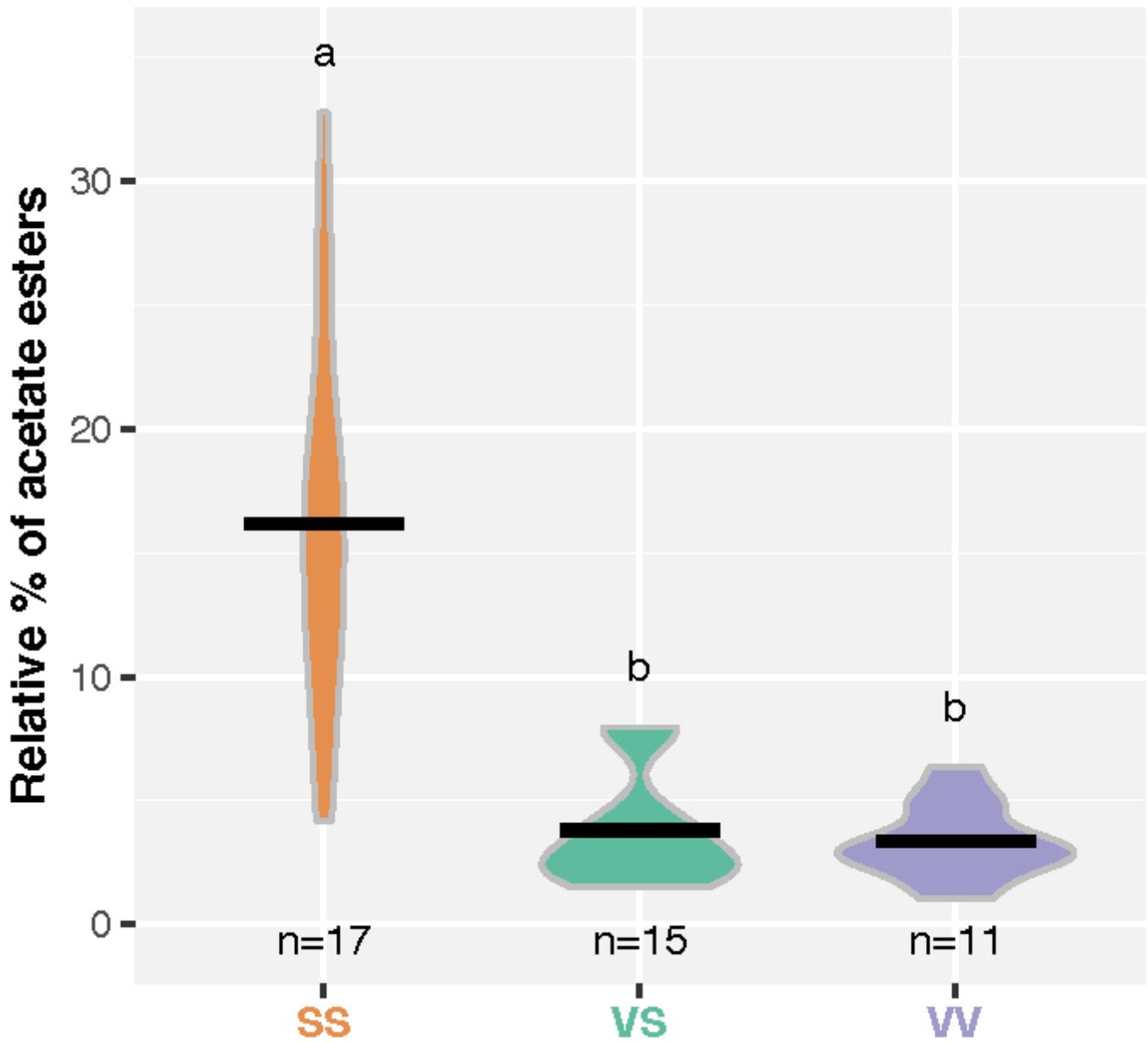


Figure 1

Acetate levels in introgressed lines.

VV individuals (purple) carry two copies of *H. virescens* QTL-Chr20, VS individuals (green) carry only one copy and SS individuals (orange) don't have *H. virescens* introgression. Letters represent group of statistical similarity based on Welch one-way ANOVA followed by Games-Howell post-hoc test (p-value= 0.000013 for SS-VS, 0.00000908 for SS-VV and 0.834 for VS-VV).

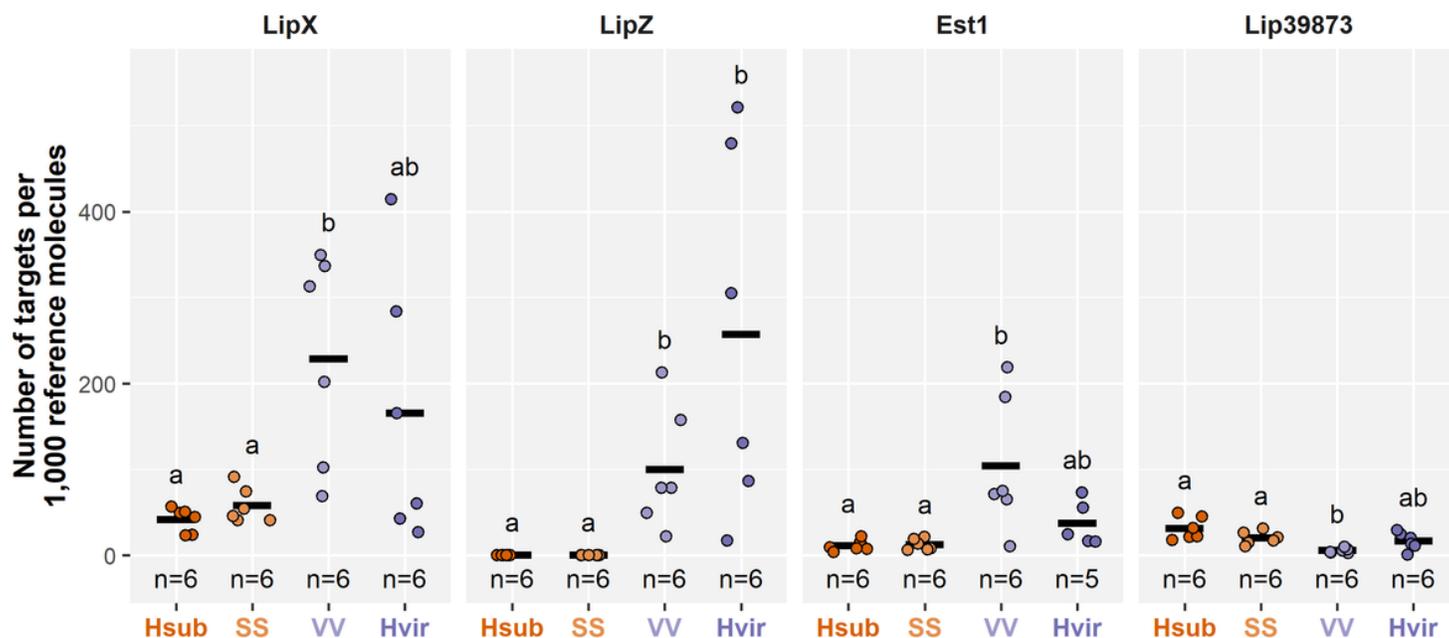


Figure 2

Expression level of candidate genes in *H. subflexa*, *H. virescens* and DD23 female pheromone glands

qPCR of the transcripts of LipX, LipZ, Est1 and Lip39873 in target transcripts per 1000 reference molecules in pheromone glands cDNA from *H. subflexa* (orange), *H. virescens* (purple) and DD23 SS (light orange) and VV (light purple) females. Letters represent group of statistical similarity based on Welch one-way ANOVA followed by Tukey or Games-Howell post-hoc test for LipZ and Est1 or LipX and Lip39873 Respectively (exact p-values in Supplementary Table S2).

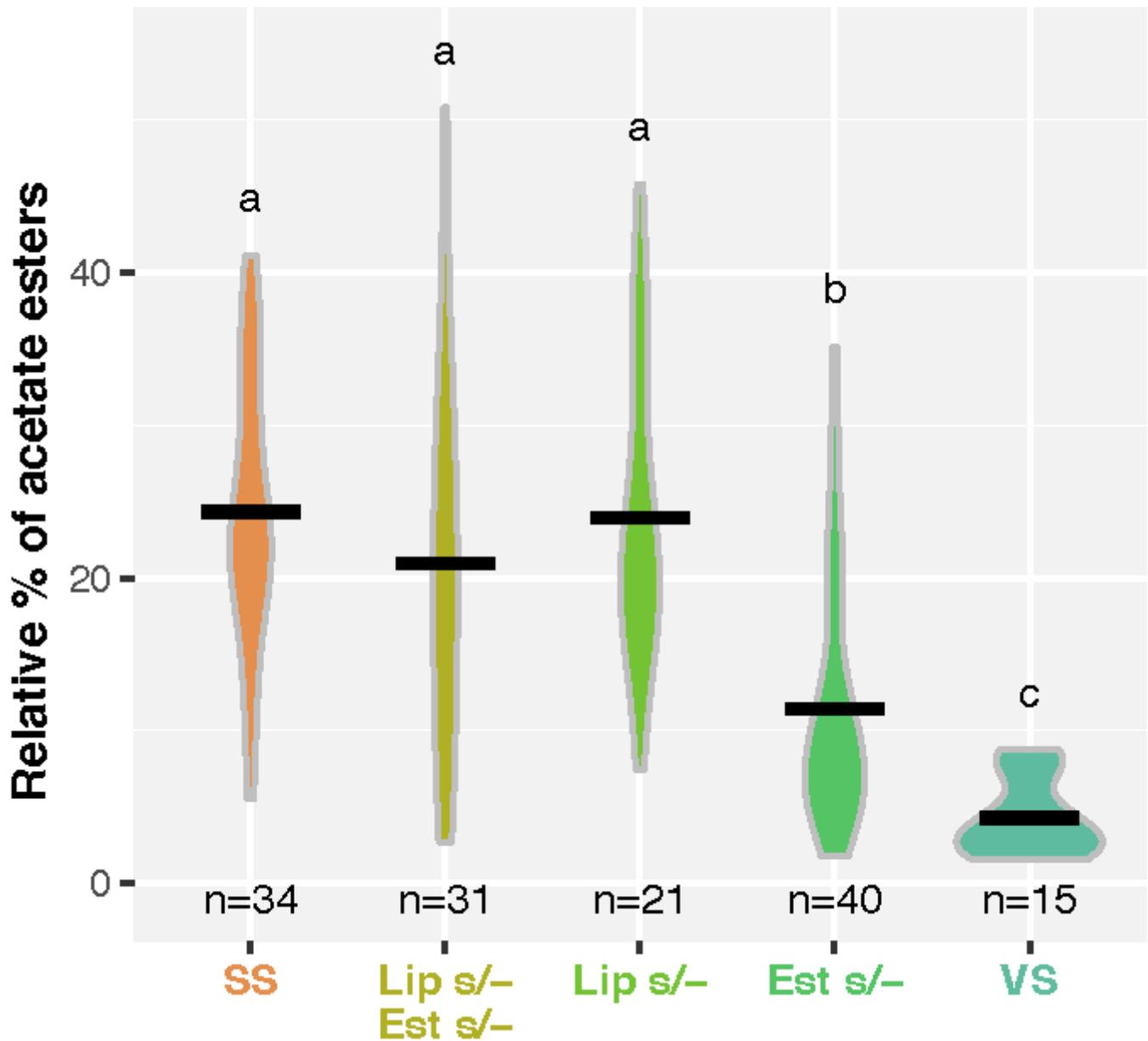


Figure 3

Effect of the different knock-outs on the acetate levels

Relative percentage of the sum of the three acetates, in *H. subflexa* females. SS individuals are wild type, VS individuals carry one copy of *H. virescens* Chr20-QTL. Est s/- stands for VS individuals for which Est1 and Est2 have been knocked-out, Lip s/- are also VS individuals but this time LipX and LipZ are not functional. Finally, Lip s/- Est s/- stands for VS females that have the four aforementioned genes knocked-out. Letters represent group of statistical similarity based on Welch one-way ANOVA followed by Games-Howell post-hoc test (exact p-values in Supplementary Table S3)

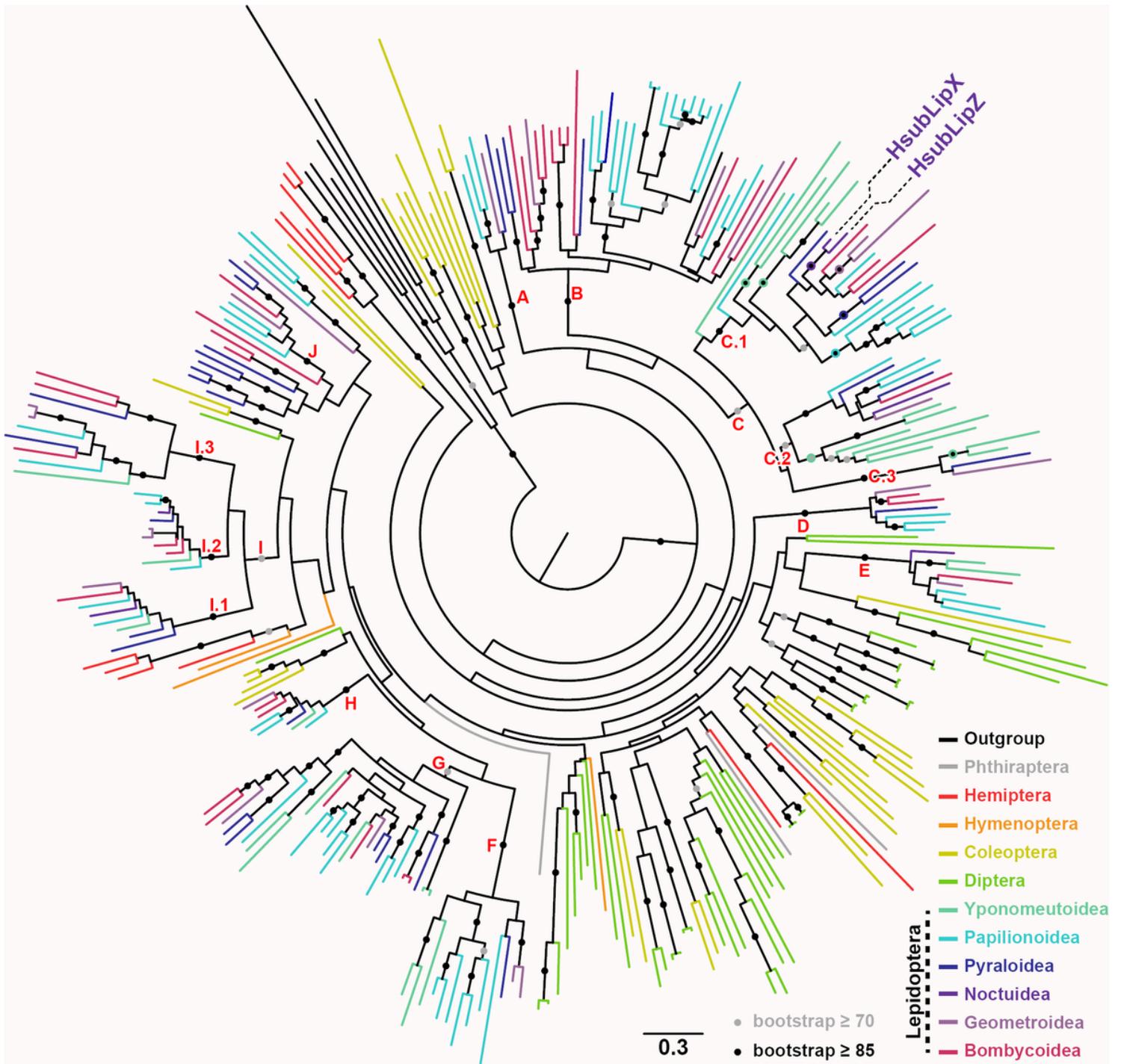


Figure 4

Maximum-likelihood phylogeny of insect acidic lipases.

Phylogenetic tree built from amino-acid sequences of insect acidic lipases obtained from the genomes of Phthiraptera (*Pediculus humanus*, branches colored in grey) ; Hemiptera (*Acyrtosiphon pisum*, red) ; Hymenoptera (*Apis mellifera*, orange) ; Coleoptera (*Tribolium castaneum* and *Dendroctonus ponderosae*, yellow) ; Diptera (*Aedes aegypti* and *Drosophila melanogaster*, light green) and multiple clades of Lepidoptera: Yponomeutoidea (*Plutella xylostella*, green) ; Papilionoidea (*Danaus plexippus*, *Heliconius melpomene* and *Papilio xuthus*, cyan) ; Pyraloidea (*Amyelois transitella* and *Chilo suppressalis*, dark blue)

; Geometroidea (*Operophtera brumata*, violet) ; Bombycoidea (*Bombyx mori* and *Manduca sexta*, pink) as well as sequences from the transcriptome of *Heliothis subflexa* (purple). The sixth sequences most similar to HvirLipX not listed as acidic lipase in the Esther database⁶⁴ have been used as outgroup (black). Bootstrap support of the clades is depicted by dots on the branches: gray for bootstrap ≥ 70 and black for bootstrap ≥ 85 . The letters (A to J) represent the Lepidoptera specific clades and their subclades containing sequences from multiple Lepidoptera family. The color stroke around dots in the C clade mark clades encompassing multiple sequences from only one species. The scale bar represents 0.3 expected amino-acid substitutions per site.

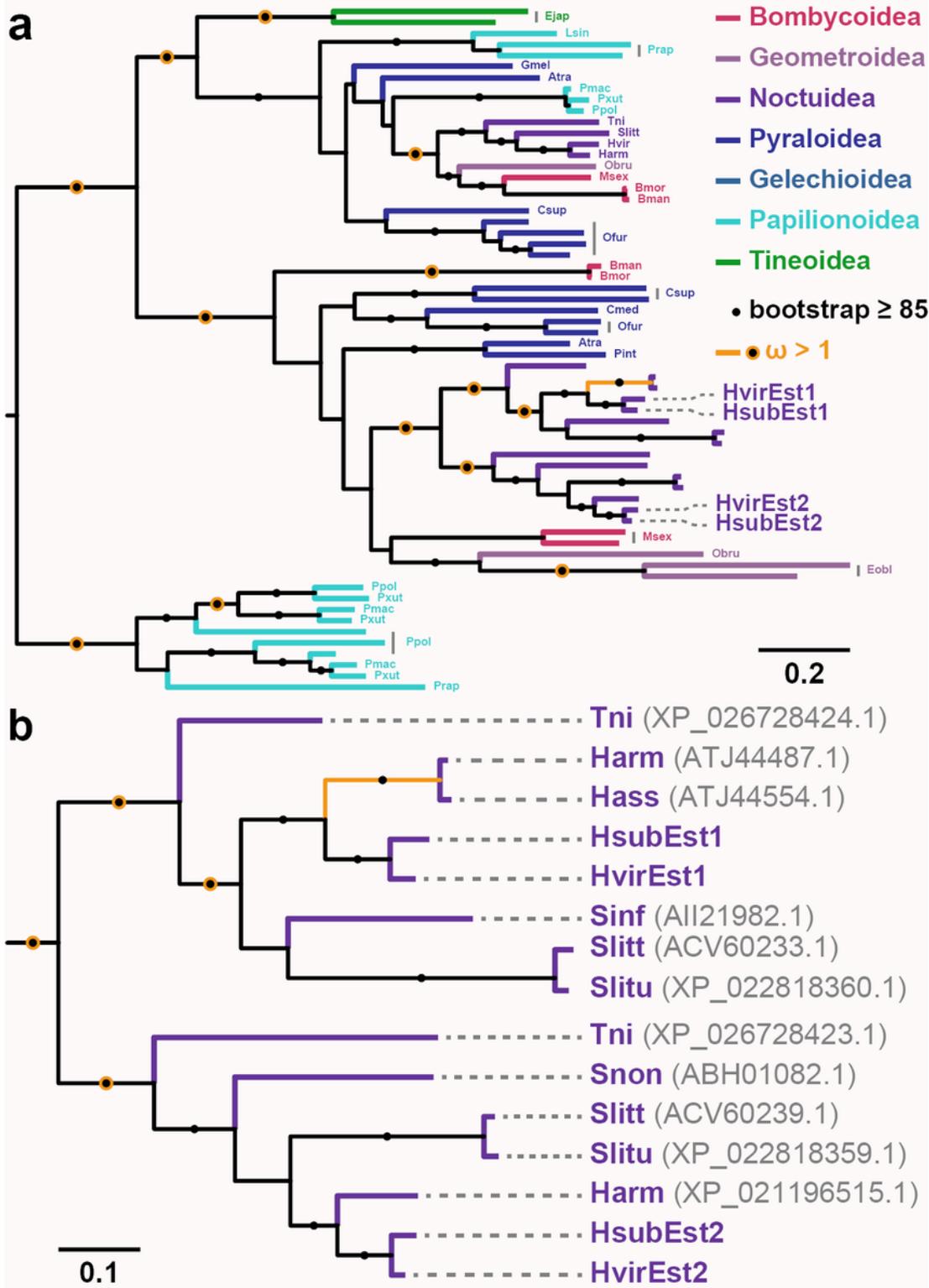


Figure 5

Maximum-likelihood phylogeny of the C clade of Lepidoptera acidic lipases and selective pressure analysis.

a) Phylogenetic tree built from amino acid sequences of Lepidoptera acidic lipases belonging to the clade C from Gelechioidea (branches colored in blue) and other Lepidoptera super-families as in Figure 4. Dots

Maximum-likelihood phylogeny of Est1 and Est2 orthologues and selective pressure analysis.

a) Phylogenetic tree built from amino acid sequences of Lepidoptera esterases from Psychidae (branches colored in green) and from other super-families as in figures 4 and 5 with similar color coding. Dots represent bootstrap support ≥ 85 . Orange dots represent detection of positive at the highlighted node using codeML branch-site model ($\omega > 1$, likelihood ratio test < 0.05). The scale bar represents 0.2 expected amino-acid substitutions per site. **b)** Noctuidae specific clade of A.2 clade from phylogenetic tree of A with similar annotation apart from scale bar representing 0.1 expected amino-acid substitutions per site.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [DD23acetatesuppmatNatcomm.docx](#)
- [DD23acetatedataNatcomm.xlsx](#)
- [rs.pdf](#)