Evolution of sexual signals

Within and between species variation in a dual function sex-pheromone component in two noctuid moths

Fruitet, E.

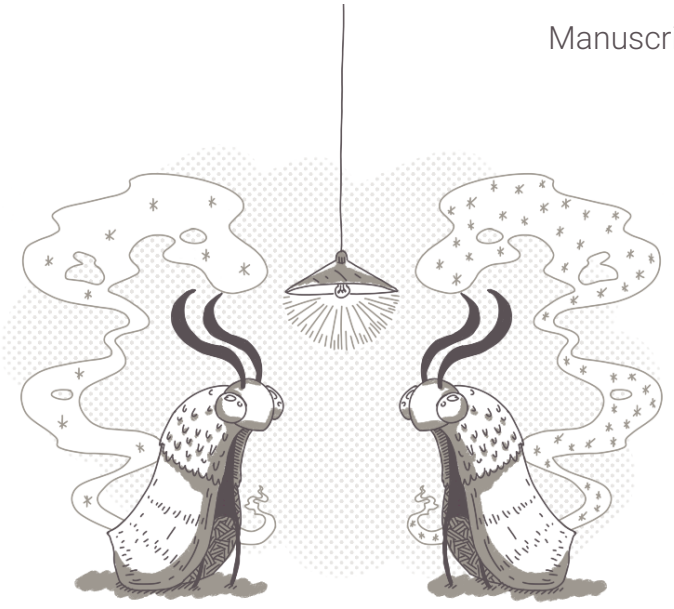
Publication date
2023

3

An esterase affects pheromone components important for reproductive isolation between two closely related moth species

Elise Fruitet, David G. Heckel, Astrid T. Groot

Manuscript in preparation



Abstract

Sexual signals are key components in mate finding and mate choice, and play an important role in speciation. Sexual signalling by sex pheromones is widely studied in moths, where females emit species-specific multi-component pheromone blends to attract males over long distances. The nature of the enzymes involved in the variation of acetate ester (acetate) levels, which occur in many moth pheromone blends, has been a long-standing gap in our understanding. Previous genetic analyses had identified two major quantitative trait loci (QTL) affecting acetate amounts in *Heliothis (Cloridea) subflexa*. Genome sequencing identified a cluster of three carboxylesterases mapping to one of the QTLs. If one or more of these esterases hydrolyzed pheromone acetates, blocking enzyme activity would be expected to increase acetate levels. Contrary to this expectation, two different knockouts of the esterase CXE21, a naturally-occurring transposable element insertion and a CRISPR/Cas9-induced frameshift, exhibited lower acetate levels. To place our findings in an evolutionary context, we then explored variation in the esterase cluster in 16 species of Lepidoptera with completely sequenced genomes. To our knowledge, this is the first report of a single gene that increases acetate levels in moth sex pheromones. These findings pave the way for further biochemical and molecular research into the role of esterases in the evolution of sex pheromone communication in moths.

Introduction

For species that reproduce sexually, it is essential to find a suitable mate. Sexual communication allows individuals to not only recognize conspecifics and avoid hybridization with individuals from different species, but also to discriminate between more and less attractive individuals. Therefore, studying the evolution of sexual signals used is crucial for the understanding of species evolution (Coyne and Orr, 2004). However, the genetic basis of variation in sexual signals is still incompletely understood. For example, sex pheromones are widely used for mate choice in many different animal species, but so far only a few genes that affect chemical signaling have been identified in *Drosophilia* (review in Jallon and Wicker-Thomas 2003; Ferveur 2005) or in moths (Lassance et al. 2010).

Moth sex pheromones are an exemplary model to investigate the evolution of sexual communication. In moths, females emit a volatile multicomponent sex pheromone blend that attracts conspecific males from a distance (Ando et al., 2004). By now, the sex pheromone blend of over 2000 moth species has been characterized (see Pherobase.com). Compounds may be shared among the blends of different species, (El-Sayed 2021), and species-specificity stems from presence or absence of different compounds as well as shifts in their ratios (Ando et al., 2004; Tillman et al., 1999).

In general, moth sex pheromone blends consist of long carbon chains of different lengths, with possibly one or more double bonds and a terminal alcohol, aldehyde, or acetate ester functional group (reviewed by (Jurenka, 2017, 2004; Tillman et al., 1999). In the 1970s and 1980s, some of the key enzymes involved in sex pheromone production were identified through detailed biochemical studies (reviewed by (Jurenka, 2004; Rafaeli and Jurenka, 2003; Tillman et al., 1999). As these enzymes are members of large protein families,

the identification of specific genes responsible for species-specific pheromone blends has been challenging.

Surprisingly, even though acetate esters (hereafter, acetates) are present in numerous species' sex pheromone blends, the genes responsible for acetate synthesis from precursor alcohols remain unknown (Ding and Löfstedt, 2015; Groot et al., 2013). Biochemical studies have shown that acetates can be synthesized from alcohols by acetyl transferases (Jurenka and Roelofs, 1989; Morse and Meighen, 1987a) or hydrolyzed back to alcohols by acetate esterases. Acetate hydrolysis activity has been found in moth sex pheromone glands since the 1980s (Morse and Meighen, 1987a, 1984; Teal and Tumlinson, 1987).

Carboxylesterases have been investigated extensively for their possible function in acetate hydrolysis of pheromone and non-pheromone esters, especially in degrading pheromones in male antennae after detection (Ferkovich et al., 1980; Ishida and Leal, 2005; Merlin et al., 2007b; Prestwich et al., 1989; Vogt, 2005). Wang et al. (2020) discovered two carboxylesterase genes expressed in the Antheraea pernyi pheromone gland, and hypothesized a potential role in the degradation of the two major sex pheromone components of this species, i.e. E6,Z11-16:OAc and E4,Z9-14:OAc. Moreover, we also recently detected two carboxylesterases involved in acetate degradation in the two closely related moth species *Heliothis subflexa* and *H. virescens* (de Fouchier et al., 2022).

The two moth species of the genus *Heliothis* were recently renamed *Chloridea*, (Pogue, 2013) but since the sex pheromone variation in these two species has been extensively studied, we retain the older name here for continuity. In summary, in *H. subflexa* acetates are present in the female sex pheromone blend and attract conspecific males, while acetates are absent in the sex pheromone blend of *H. virescens* females and repel *H. virescens* males, decreasing the chance of cross-attraction and species hybridization (Groot et al., 2007, 2006; Vickers, 2002; Vickers and Baker, 1997). Previously, two main quantitative trait loci (QTLs) involved in acetate

variation were discovered by backcrossing hybrids of *H. virescens* and *H. subflexa* (Groot et al., 2009a, 2004b; Sheck et al., 2006). After introgression one of these QTLs from *H. virescens* (on Chromosome 20) into the genomic background of *H. subflexa*, we identified two carboxylesterase genes within the introgressed region and studied their involvement in acetate degradation (de Fouchier et al., under revision). In the second QTL (on Chromosome 28), we identified three carboxylesterase genes in tandem. Here we describe their role in influencing the amounts of acetates in *H. subflexa*. Both a naturally-occuring insertion of a transposable element in one of the esterase genes, and experimental CRISPR/Cas9 mutagenesis of the same gene, affected the amounts of acetates in the pheromone gland of *H. subflexa*. A phylogenetic analysis based on 16 species of Lepidoptera provides further insight into the evolution of this carboxylesterase cluster.

Results

A cluster of esterases expressed in the pheromone gland

Sequencing of already published cDNA library from female pheromone glands revealed expression of three orthologous carboxylesterases in *H. virescens* (Vogel et al., 2010) and *H. subflexa* (de Fouchier et al., under revision). We named two of these (CXE5, CXE16) after apparently orthologous esterases previously described from the antenna of *Spodoptera littoralis* (N. Durand et al., 2010). The third (CXE21) was arbitrarily named to continue the numbering series, but whether an ortholog exists in *S. littoralis* is unknown. All three are represented by ESTs from a cDNA library of the female pheromone gland of *H. virescens* (Vogel et al., 2010); GenBank Accession GR967905.1 for CXE5, GT067762.1, GR969506.1,

GR959306.1, and GR969507.1 for CXE16, and GR970796.1 for CXE21). Genome sequencing showed that the three esterases appear in a tandem array on Chromosome 28, flanked by three other genes (not esterases) conserved across Lepidoptera (*H. virescens* GenBank Accession No. ON500499, Supplementary Figure S1A; *H. subflexa* GenBank Accession No. ON603978). Orthologous clusters were found in other Lepidoptera (see below).

Structure of the HsubTED retrotransposon

In exon 8 of an allele of CXE21 described above (GenBank Accession ON603978), we found a long terminal repeat (LTR) retrotransposon in *H. subflexa*. We named it HsubTED due to its similarity to the TED retrotransposon from Trichoplusia ni (Friesen and Nissen, 1990)(GenBank Accession NC_038512); it is also similar to Lydia from Lymantria dispar (Pfeifer et al., 2000) (GenBank Accession AF177773). The length of HsubTED is 8963 bp (basepairs), including identical 226 bp long terminal repeats on either side of the three open reading frames (ORFs) (Supplementary Figure S1B). The orientation of HsubTED and its three ORFs is opposite to the orientation of the esterases. There is a four-base target site duplication (tsd) GAGT flanking each LTR. The identity of the LTRs suggests a relatively recent insertion event, because paired LTRs are identical when the retrotransposon is inserted into the genome (Boeke and Stoye, 1997), and random point mutations tend to accumulate at different positions in the two LTRs over time after insertion (SanMiguel et al., 1998). Also consistent with a recent insertion is the absence of any other inactivating mutations in the coding sequence of the inserted allele of CXE21.

The presence of HsubTED in exon 8 of an allele of CXE21 produced an mRNA encoding a truncated protein of 388 amino acids (GenBank Accession ON500499), not the entire 558 residue protein encoded by the wild-type allele (GenBank Accession ON603979). Since the missing ~170 residues at the carboxy terminus contain the

glutamate serving as the third residue in the catalytic triad, the truncated protein is expected to lack any hydrolytic activity.

To determine the effect of the HsubTED insert on the female sex pheromone composition, we used the LTRs to design PCR primers to screen for the presence or absence of HsubTED in this location (Supplementary Figure S1C). Two primers (F1A and R2A) within exon 8 flanking the insertion point produced a 138 bp product from the wild-type allele. When a third primer (R2B) positioned within the left LTR was added to the mix, a 189 bp product was generated if HsubTED was present or a 170 bp product when the F1B primer within the other LTR was added. We screened individuals of our *H. subflexa* colony with these primers and set up single-pair matings to generate progeny homozygous for the insertion, homozygous for the wild-type allele, and heterozygous offspring.

Effects on the pheromone blend

Assuming that the carboxylesterase CXE21 hydrolyzes acetates in the pheromone gland, we expected that the knockout would reduce hydrolysis and increase acetate levels. However, we observed reduced acetate levels instead, which was thus opposite to our expectation. Females with both CXE21 alleles having HsubTED inserted, i.e. a double knockout CXE21 -/-, had lower acetate levels than individuals with only one or no CXE21 alleles with HsubTED inserted (Figure 1A, Games-Howell test, P =0.02 when comparing CXE21+/- and CXE21 -/- and P =5*10^{-4} when comparing CXE21+/+ and CXE21 -/-). None of the other sex pheromone compounds or the total amount of pheromone showed differences between the double knockout and the two other genotypes (Figure 2). Thus, the HsubTED insertion in CXE21 affected only the acetate levels in the sex pheromone blend of H. subflexa.

One possible explanation of this opposite-to-expected result is that new transcripts originating from the LTRs could extend into a neighboring esterase in the cluster, affecting its expression in some way. Such transcripts have been documented for the LTR of TED, inserted into a baculovirus genome, extending into adjacent viral genes (Friesen and Miller, 1987). Rather than searching for such transcripts, we knocked out CXE21 using CRISPR/Cas9 starting from the wild-type allele lacking HsubTED. Guide RNAs were designed targetting two locations in exon 8, so that the exon 8 screening primers could also be used to screen the deletions caused by the Cas9 enzyme (Supplementary Figure S1C). Knockout mutants were recovered and used to set up crosses generating the three genotypes: homozygous wild-type, homozygous CRISPR/Cas9 knockout, and heterozygous.

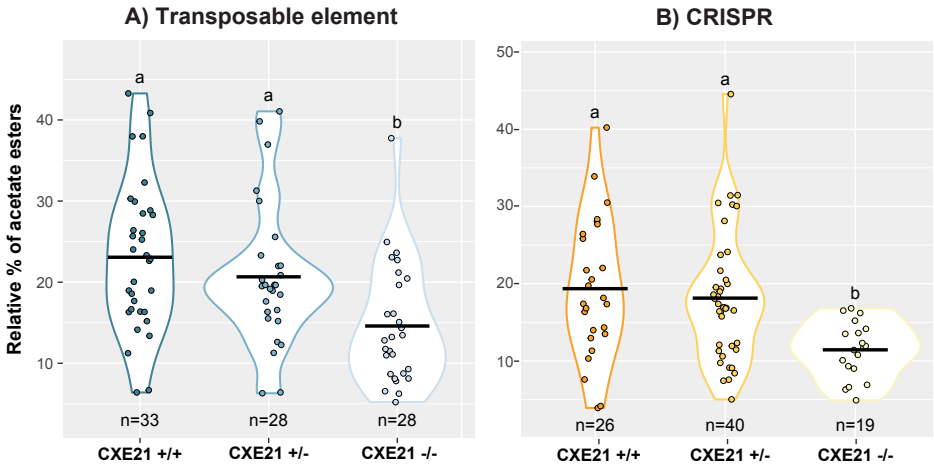


Figure 1: Effect of CXE21 knockouts on the relative percentage of acetates.

Relative percentage of the sum of the three acetates in H. subflexa females A) with the HsubTED transposable element insertion, B) from the CRISPR/Cas9 experiment. CXE21 +/+ are individuals with two functional copies of the CXE21 gene, CXE21 +/- individuals carry one functional copy, and CXE21 -/- females have no functional copies. The horizontal black bar represents the mean. Different letters above the violin plots indicate statistical differences, based on a Welch one-way ANOVA, followed by a Games-Howell post-hoc test.

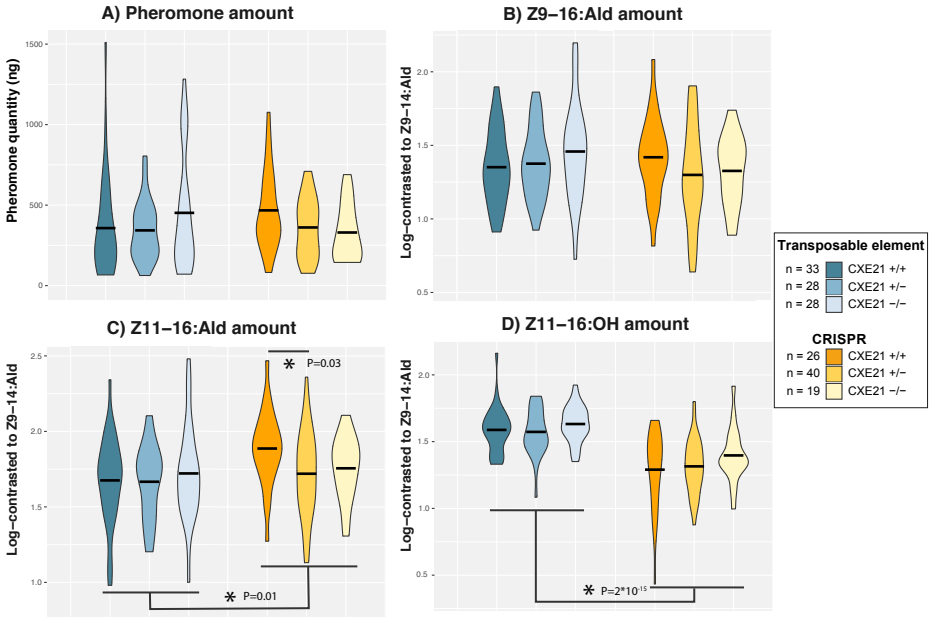


Figure 2: Effect of CXE21 knockouts on various pheromone components in H. subflexa.

the total amount of pheromone produced, B) Z9-16:Ald, C) Z11-16:Ald and D) Z11-16:OH. In blue are the data from females with or without the HsubTED transposable element insertion. In yellow are the data from females of the CRISPR/Cas9 experiment. CXE21 +/+ are individuals with two functional copies of the CXE21 gene, CXE21 +/- individuals carry one functional copy, and CXE21 -/- females have no functional copies. The horizontal black bar represents the mean. The black bar with a star indicates a statistical difference based on the results from two-way Anova, with the genotype and the type of knockout used as explanatory variables followed by a Turkey post-hoc test.

Surprisingly, we observed the same results for the new knockouts: homozygous females with two non-functional copies of CXE21 had significantly lower acetate levels compared to the two other genotypes (Figure 1B, Games-Howell test, P =2*$\mathbf{10^{-4}}$ when comparing CXE21+/- and CXE21 -/- and P =7*$\mathbf{10^{-4}}$ when comparing CXE21+/+ and CXE21 -/-). Thus, regardless of how CXE21 was inactivated in exon 8, acetate levels decreased when both copies of CXE21 were knocked out. Furthermore, in these experiments there were no significant differences between the three genotypes in the total amount of pheromone (Figure 2A), or in the two minor sex

pheromone components to which male *H. subflexa* respond, Z9-16:Ald (Figure 2B) or Z11-16:OH; (Figure 2D). However, the major sex pheromone component, Z11-16:Ald, was significantly different between heterozygous females and homozygous females with two functional copies of CXE21 (Figure 2C, Tukey test, P = 0.03).

Finally, we compared both datasets to investigate possible differences between the two knockout methods (HsubTED insert and CRISPR/Cas9). No significant differences in the total amount of pheromone or Z9-16:Ald were found. However, we did observe an experimental effect on the acetate levels, and on Z11-16:Ald and Z11-16:OH, two pheromone components that affect male behavior. No significant interaction effect was found between the knockout methods and the genotypes.

Evolution of the esterase cluster in Lepidoptera

To better understand the evolutionary history of the gene cluster represented by three carboxylesterases in *H. subflexa*, we searched for orthologous clusters in other Lepidoptera species. These could be identified with confidence because they were flanked by single-copy conserved genes (not esterases) with clear orthologs in Lepidoptera. All esterase genes in orthologous clusters that we identified from whole-genome sequencing projects possessed the catalytic triad and appeared to encode functional carboxylesterases. Furthermore, within each cluster, genes could be distinguished by the motif surrounding the catalytic serine. Both HsCXE5 and HsCXE21 have a G***C***S**A**G motif, while HsCXE16 has a G***Y***S**A**G motif.

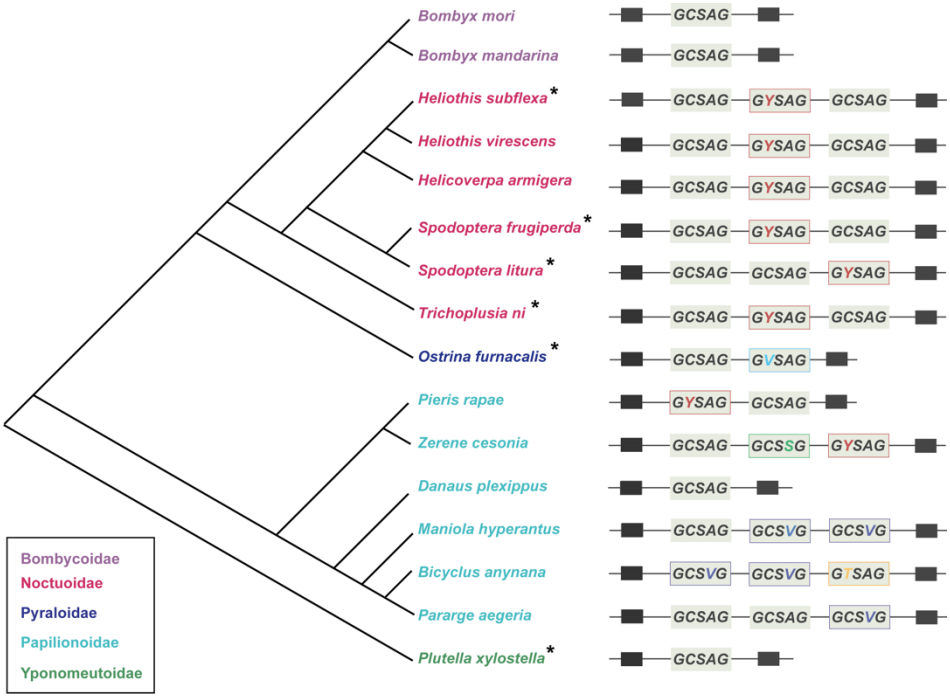


Figure 3: The chromosome 28 esterase cluster in Lepidoptera

Relatedness between the species studied is based on Kawahara and Mitter phylogenies (Kawahara et al., 2019; Mitter et al., 2017). *In light purple are superfamily Bombycoidae (Bombyx mori and B. mandarina), in pink are Noctuoidea (Heliothis subflexa, H. virescens, Helicoverpa armigera, Spodoptera frugiperda, S. litura and Trichoplusia ni), in dark blue are Pyraloidae (Ostrinia furnacalis), in light blue the superfamily Papilionoidea (Pieris rapae, Zerene cesonia, Danaus plexippus, Maniola hyperantus, Bicyclus anynana and Pararge aegeria), and in green Yponomeutoidea (Plutella xylostella). Species with a star are known to contain acetates in their sex pheromone blend. The black rectangles represent two of the flanking genes, while the grey rectangles represent the esterases. The letters written in the grey rectangles are the motifs around the catalytic serine; colored letters show changes in this motif.*

Similar three-gene clusters were found in the following Noctuoidea: *Helicoverpa armigera, Spodoptera frugiperda, S. litura* and *Trichoplusia ni* (Figure 3). By comparison, more variation was found in the Papilionoidea, where not all species had the same number of esterases (from one to three esterases were present), while species with the same number of esterases never had the exact same motif around the catalytic serine. Furthermore, both Bombycoidae species, *Bombyx mori* and *B. mandarina*, as well as the Papilionoidea *Danaus plexippus* and the one Yponomeutoidea, *Plutella xylostella*, had only one esterase flanked by the single-copy conserved genes. For all the species with only one esterase, the G***C***S<u></u>AG motif was found, suggesting that it is the ancestral form, and that esterases with the G***Y***S<u></u>AG motif diverged after gene duplication.

Discussion

A counterintuitive effect of a putative esterase

Our results clearly show that the absence of two enzymatically active copies of CXE21 significantly changed the acetate levels in the pheromone gland of *H. subflexa* females, while none of the other pheromone components were affected by the CXE21 double knockout. Importantly, these results were obtained using two different knockout methods, one a naturally-occuring retrotransposon and the other an experimental disruption of the gene. The two comparisons were conducted at different times with about one year in between, therefore differences in overall pheromone amounts is likely due to temporal variation which we have also observed in other studies of this laboratory colony over several years. However, the qualitative result is the same, and

opposite to typical expectations of knocking out an active carboxylesterase.

Because esterases typically hydrolyze acetates, yielding alcohols and acetic acid, we expected acetates to accumulate when CXE21 was not functional, and thus to find higher acetate levels when this gene was knocked out. However, our results were opposite to this expectation, as the acetate levels were *lower* when the HsubTED transposable element was inserted and when the gene was knocked out using CRISPR. This begs the question of how acetate levels can be lower when CXE21 cannot hydrolyze them. One possibility could be that CXE21 possesses some acetyltransferase activity. Another possibility could be that CXE21 has negative interactions with other enzymes (other esterases or acetyl-transferase). We discuss these two possible explanations below (Figure 4).

H1: Possible acetyltransferase activity

The first explanation of our finding is that CXE21 behaves like an acetyltransferase. In this case, it would esterify alcohols into acetates instead of hydrolyzing acetates into alcohols. It is important to note that both reactions are reversible, but that hydrolysis is favored in an aqueous environment (Bornscheuer and Kazlauskas, 2005). Since sex pheromone biosynthesis occurs in sex pheromone glands, and reactions may occur in the gland membrane (Teal and Tumlinson 1987), a suitable non-aqueous environment may be available for the exclusion of water that would otherwise drive hydrolysis. However, CXE21 would need to be expressed in a heterologous system to confirm possible acetyltransferase activity. A previous attempt with 34 different candidates from the moth Agrotis segetum failed to yield any detectable acetyltransferase activity (Ding and Löfstedt, 2015).

In a recent study with Dalotia rove beetles, silencing of the esterase TG-aEst using systemic RNA interference caused the loss

of three fatty acid esters of ethanol or isopropanol from the defensive gland, suggestive of acyltransferase activity (Brückner et al., 2021). Moreover, diverse esterases from the bacterial hormone-sensitive lipase family have been described to have promiscuous acyltransferase activity (Müller et al., 2020).

Interestingly, the acetyl-transferase activity of some bacterial carboxylesterases was due to a single nucleotide substitution (Müller et al., 2021): A mutation in one of the three amino acids forming the motif adjacent to the catalytic triad switched the activity of an esterase without promiscuous acetyltransferase activity to an enzyme with acetyltransferase activity. Conversely, the acetyltransferase activity was lost after mutation of one of these three-amino acids. Thus, a possible acetyltransferase activity of an esterase like CXE21 is not completely without precedent.

Even if CXE21 possesses some acetyltransferase activity, it cannot be responsible for all of the enzymatic activity observed, for example in pheromone gland extracts of *Choristoneura fumiferana* (Morse and Meighen, 1984) or *Trichoplusia ni* (Jurenka and Roelofs, 1989). Both studies provided evidence of an acetyltransferase activity that utilizes acetyl-CoA and fatty alcohols of different chain lengths. The enzyme or enzymes responsible for this activity have not yet been identified, despite strenuous efforts (Ding and Löfstedt, 2015).

H2: Negative interactions with other enzymes

Another possible explanation of our opposite-to-expected result is that mutations in CXE21 have negative interactions with its two neighbors (CXE5 and CXE16) or other esterases hydrolyzing acetates in the pheromone gland. The inactivation of a gene can sometimes lead to an increase in expression of its paralogs. This can happen 1) when the transcription factors of the deleted gene repress the expression of its paralog, or 2) when the transcriptional regulation network is modified due to the expression of the paralogs

being regulated by products of the same biosynthetic pathway (Diss et al., 2014). If other esterases are upregulated in the absence of CXE21 expression, more acetates in total could be hydrolyzed resulting in lower acetate levels (Figure 4).

Yet another possibility is a negative interaction with an acetyltransferase. It is conceivable that the 388-residue protein produced from the gene with the HsubTED insertion, or the 415-residue protein produced from the delta28 mutation (GenBank Accession ON641837) or the 414-residue protein produced from the delta31 mutation (GenBank Accession ON641838), could bind to some of the alcohol precursors and prevent their access by the active acetyltransfererase(s), reducing the synthesis of the acetates.

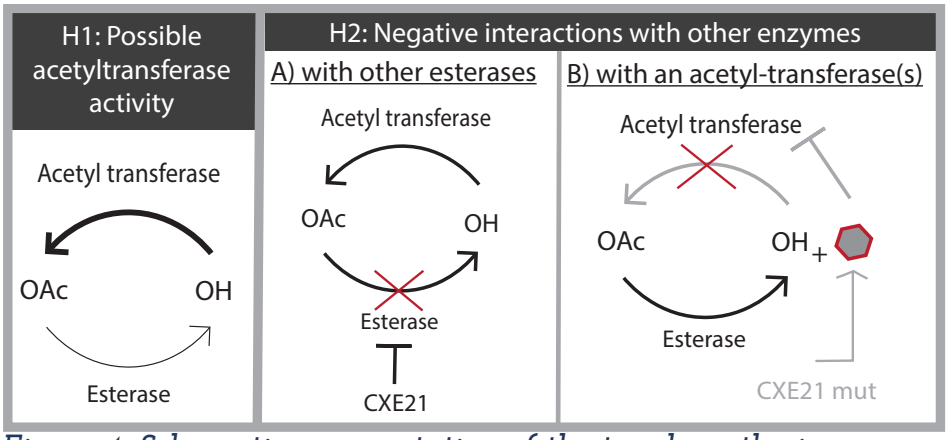


Figure 4: Schematic representation of the two-hypothesis developed in the discussion

To explain the opposite-to-expected result that a knock-out of CXE21 leads to lower acetate levels in the sex pheromone on H. subflexa, two-hypothesis are proposed in the discussion and schematized here.

Conserved esterase cluster in Lepidoptera

By investigating homologous esterase clusters in Lepidoptera species with fully sequenced genomes, we found that species from three different superfamilies had only one esterase, with a G**C**S**A**G motif surrounding the catalytic serine, which is probably the ancestral enzyme. All species with two or more esterases had the ancestral motif as well as at least one derived motif. All investigated Noctoidea had three esterases in tandem, two with a G**C**S**A**G motif and one with a G**Y**S**A**G motif, pointing to two duplication events in this superfamily.

After duplication, genes may accumulate mutations, resulting in pseudogenization, subfunctionalization or neofunctionalization. The latter usually results in function close to the ancestral gene (Conant and Wolfe, 2008; Zhang, 2003). Interestingly, in the yeast *Saccharomyces cerevisiae,* a duplicated alcohol dehydrogenase was found to convert ethanol to acetaldehyde, while the reverse reaction was performed by the ancestral enzyme (Thomson et al., 2005). Heterologous expression of all esterases in the cluster would be required to determine which might be prone to acetyltransferase activity.

Although all Noctoidea examined have three esterases in the Chromosome 28 cluster, *H. virescens* and *Helicoverpa armigera* females do not produce any acetates. A QTL mapping to Chromosome 20 affected acetate variation between *H. subflexa* and *H. virescens* (Groot et al., 2009a; Sheck et al., 2006) and carboxylesterases and lipases mapping to Chromosome 20 region vary in expression level among the two species (de Fouchier et al., under revision). Thus, other mechanisms, like expression or repression of the esterase cluster in interaction with trans-acting effects on carboxylesterases on another chromosome, are possibly involved in generating species-specific pheromone blends.

Conclusion

To conclude, we found a decrease in levels of acetates in the pheromone blend of *H. subflexa* females when the gene encoding the carboxylesterase CXE21 is mutated by natural or experimental methods. An increase in acetates would have been expected if the wild-type CXE21 hydrolyzes them, as do similar esterases expressed in the moth antenna. The results do not have a simple explanation; nevertheless, CXE21 is the first gene identified in a lepidopteran species that affects the amounts of a very common constituent of moth pheromones, namely acetates. Metabolism of acetates in the female pheromone gland is likely more complicated than biosynthesis by a single acetyltransferase enzyme and degradation by a single carboxylesterase. These results thus pave the way to new biochemical and molecular studies on the role of esterases in moth sex pheromone biosynthesis and evolution.

Materials and Methods

Insect rearing

The laboratory population of *H. subflexa*, originally collected in 2006 in the field in North Carolina, has been reared at the University of Amsterdam since 2011. As lepidoptera are very sensitive to inbreeding, occasional exchange happened among *H. subflexa* populations at North Carolina State University, the University of Amsterdam, and the Max Planck Institute for Chemical Ecology in Jena. Moreover, single pair matings were used to maintain genetic diversity.

Eggs collected from single pair matings were kept in Petri dishes (Ø 85 mm) with artificial wheat germ/soy flour-based diet (BioServ Inc., Newark, DE, USA) at room temperature. To avoid cannibalism, L3 larvae were isolated in individual 37-mL cups filled with the same artificial diet. From that stage on until their death, individuals were kept in climate chambers with 14h:10h light-dark cycle, at 25°C and 60% relative humidity. After emerging, adults were sexed and supplied with cotton soaked in sugar water (10% sucrose). Pairs of males and females, from distinct families, were placed in 375 mL paper cups that contained cotton soaked in sugar water. The cup was then covered with gauze and to stimulate oviposition, a freshly cut gooseberry fruit was placed on top of it. Once the eggs turned black or began to hatch, the gauze, the eggs, and larvae on it, were transferred to Petri dishes as described above.

Phenotyping

To assess changes in pheromone blend composition, pheromone gland extractions from individual females were conducted. We used the same protocol as described in detail in (Lievers and Groot, 2016), which is summarized here. Pheromone gland extractions were performed on 3-4 days old virgin females and in the 4th hour of scotophase, as this is the peak of pheromone production in *H. subflexa* (Heath et al., 1991). Female abdomens were held with forceps to extrude their pheromone gland, which was cut with microdissection scissors and soaked in a glass vial with 50 µL hexane containing 200 ng pentadecane as an internal standard. After 30 minutes, the gland was removed from the vial and the extracts were stored at -20°C until analysis.

For pheromone analysis, the volume of each extract was first concentrated to approximately 2 µL under a gentle stream of nitrogen. To avoid evaporation 1 µL of octane was added to each sample before being injected into a splitless inlet of a 7890A GC (Agilent Technologies, Santa Clara, CA, USA). Integration of the area

under the pheromone peaks of the 11 pheromone components present in *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, Z9-16:OH and Z11-16:OH) was performed using Agilent ChemStation software (version B.04.03). As separation of Z7-16:Ald and Z9-16:Ald peaks was difficult, they were integrated as one peak (referred to Z9-16:Ald).

For every pheromone sample, the absolute amount (in ng) of each pheromone compound was calculated relative to a 200 ng pentadecane internal standard. As the ratios of the components in too low titers cannot be reliably measured in the chromatogram, samples containing < 50 ng were excluded. To correct for individual variation in the amount of pheromone produced which is highly variable (Groot et al., 2005; Heath et al., 1991); we calculated the relative percentage of the sum of the three acetates (Z7-16:OAc, Z9-16:OAc, Z11-16:OAc) by dividing their absolute amounts by the total amount of pheromone produced as done in Groot et al. (2013). The amount of the sum of the three acetates was also analyzed, which gave similar results (see Supplementary Figure S3B).

cDNA libraries

To reveal the three esterases cluster we used cDNA previously collected from *H. virescens* (Vogel) and *H. subflexa* (de Fouchier et al., under revision). as summarized below. For *H. virescens*, RNA from the pheromone glands of 32 females was extracted using Ambion™ TRIzol™ Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and RNeasy MinEluteClean up Kit (Qiagen). The construction of the cDNA library was generated using a combination of the SMART cDNA library construction kit (Clontech) and the Trimmer Direct cDNA normalization kit (Evrogen). For *H. subflexa*, RNA from the pheromone glands of 10 females was extracted using Ambion™ TRIzol™ Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Direct-zol™ RNA MiniPrep Kit

(Zymo Research, Irvine, California, USA) and sequenced using Illumina HiSeq2500 by BaseClear (Leiden, The Netherlands). To assemble the pheromone gland transcriptome, multiple filtration steps were performed as described in detail in de Fouchier et al. (2022). The quality of the transcriptome was than assess using BUSCO66 and Transrate67.

CXE21 loss-of function analysis

To explore the role of CXE21 in pheromone biosynthesis, we performed the following loss-of-function studies.

a. Transposable element knockout:

To detect the presence of the transposable element insert HsubTED, we designed PCR primers flanking the insertion site. To differentiate the three genotypes, three primers were used simultaneously, HsubTED-F1A (5'-ACCGAATAGCTCCGATCTCA-3'), HsubTED-F1B (5'-AGGGCTAAATCAGGGTTGCT-3'), and HsubTED-R2A (5'-CATACAGCTTGCCAAGTCCA-3'). Alleles without the insert in CXE21 produced a 138-bp amplicon, alleles carrying the HsubTED insert within CXE21 produced a 170-bp amplicon. Alternatively, when HsubTED-R2B (5'-CTCAGCTAGCCAGCCAGTTC-3') was used instead of HsubTED-F1B, alleles carrying the HsubTED insert within CXE21 produced a 189-bp amplicon (Supplementary Figure S2).

b. CRISPR/Cas9 knockout protocol:

To verify that the HsubTED insert effectively knocks out CXE21 and is responsible for the phenotypic change in the pheromone blend (Supplementary Figure S1), we conducted CRISPR/Cas9 knockout experiment as follows. First, for this experiment only *H. subflexa* eggs that were not carrying HsubTED transposable element insert were used. The IDT system for CRISPR/Cas9 was used (Integrated DNA Technologies) and gene-specific guide RNAs were designed (HsCXE21-T7: 5'-TAACGAGTCCACTGACAAGC-3', HsCXE21-T6: 5'-

AGAAAGGCGTACCTTGGAGA-3', Fig 3). Gene-specific crRNAs were annealed with tracrRNAs and incubated with Cas9 enzyme to form ribonucleoproteins. Before injection into the eggs, the two scRNAs were dissolved in 1 nmol of tracrRNA to get a final concentration of 50 µM, and later 200 pmol of this scRNA+tracrRNA mix was combined with 100 pmol of IDT Alt-R Cas9. This solution was then loaded in handmade glass needles and using a Femtojet, 30min-1h old eggs were micro-injected. The gauze with the injected embryos was housed in Petri-dishes at lab temperature (20°C) and examined daily. Newly hatched neonates were segregated in individual cups with wheat germ/soy flour-based diet (BioServ Inc., Newark, DE, USA).

CRISPR/Cas9 knockout rearing and screening for mutations

To identify possible mutations in CXE21, adults were genotyped as explained below. Cloning and sequencing of the mutants was performed at the Max Planck Institute for Chemical Ecology and highlighted a deletion of either 28 or 31bp. After two generations, a subset of females from the different genotype were used for pheromone extraction. To avoid inbreeding, half of CXE21+/- females were crossed with CXE21+/+ *H. subflexa* male from the main rearing and the other-half were paired with either CXE21+/+ or CXE21+/- males to obtain double mutants. This procedure was repeated every generation until a decent samples size was reached.

Genotyping

For both knockout methods, distinction between the genotypes was made through PCR amplification following 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 of reverse and forward primers 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. DNA extractions were conducted 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, then 8 min at 98 degrees and lastly freezing the mixture before use (Adapted protocol from Biorad).

Data analysis

To estimate differences in pheromone blend composition, statistical analyses were done using R Studio version 1.0.136 with R version R 3.6.1 (R Core Team, 2019). In a first step, as two different methods were used to knock out CXE21 (HsubTED insert and CRISPR method), we analyzed both datasets separately. Homogeneity of the means were tested using a one-way ANOVA. When the assumption of homogeneity of variances was violated, a Welch test (function "oneway.test", package "stats") was performed. If this assumption was satisfied but the residuals were not normally distributed, a Kruskal-Wallis test (function "kruskal.test", package "dplyr") was used. To test for homogeneity of variances, we used Bartlett's test (function "bartlett.test", package "stats"). To evaluate if the residuals were normally distributed, we used a Shapiro test (function "shapiro.test", package "stats"). If Ho was rejected, to compare which specific genotypes differed, a Tukey test was performed when both ANOVA assumptions (homogeneity of the variance and normal distribution of the residuals) were respected, while a Games-Howell post-hoc test was performed when these assumptions were violated.

In a second step, datasets of both experiments were combined to be analyzed together, using a two-way ANOVA, with the genotype and the type of knockout used as explanatory variables. As no interaction effect was found between the two variables, this was removed from the model. In case the residuals were not normally

distributed, log or box-cox transformation was used (function "bcPower", package "car"), after which Tukey's post-hoc tests were performed (function "glht", package "multcomp").

First, we investigated if the three genotypes (CXE21+/+, CXE21+/-, CXE21-/-) showed differences in the relative percentage of the sum of the three acetates, as calculated in Groot et al. (2010). Then, we compared the total amount of pheromone produced.

Secondly, to normalize and break data interdependency (see Groot et al. 2010), the sum of the three acetates and other sex pheromone component that have been shown to be important for male attraction (i.e: Z9-16:Ald, Z11-16:Ald and Z11-16:OH) were log-contrasted by the amount of Z9-14:Ald. We chose Z9-14:Ald as the divisor, because this compound does not alter male response behavior (Heath et al., 1990) and was present in equivalent quantities in the different genotypes (Supplementary Figure S3A).

CXE cluster in Lepidoptera

To investigate the evolutionary history of the three esterase clusters found in *H. subflexa* and *H. virescens*, we searched for the presence of similar clusters in other Lepidoptera species. Only species with full genome assembly were used, which amounted to a total of 16 genomes. The flanking gene sequences Hs-NRK-pep and Hv-HDphosphohydrolase-Pep (sequences in Supplementary Table 1) were blasted against the Lepidoptera clade. For each species, the top hit of both flanked sequences was then observed in the gene viewer tool of NHBI. The number of esterases and the motif around the catalytic serine was noted.

References

Ando, T., Inomata, S., Yamamoto, M., 2004. Lepidopteran Sex Pheromones, Current Chemistry.

Boeke, J., Stoye, J., 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: Coffin, J., Hughes, S., Varmus, H. (Eds.), Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Bornscheuer, U.T., Kazlauskas, R.J., 2005. Catalytic Promiscuity in Biocatalysis: Using Old Enzymes to Form New Bonds and Follow New Pathways.

Brückner, A., Badroos, J.M.;, Learsch, R.W.;, Yousefelahiyeh, M., Kitchen, S.A., Parker, J., 2021. Evolutionary assembly of cooperating cell types in an animal chemical defense system. Cell 184, 6138-6156.e28.

Conant, G.C., Wolfe, K.H., 2008. Turning a hobby into a job: How duplicated genes find new functions. Nat Rev Genet 9, 938–950.

Coyne, J.A., Orr, A.H., 2004. Speciation. Sinauer Associates, Sunderland, MA.

Ding, B.J., Löfstedt, C., 2015. Analysis of the agrotis segetum pheromone gland transcriptome in the light of Sex pheromone biosynthesis. BMC Genomics 16.

Diss, G., Ascencio, D., Deluna, A., Landry, C.R., 2014. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. J Exp Zool B Mol Dev Evol 322, 488–499.

Durand, N., Carot-Sans, G., Chertemps, T., Montagné, N., Jacquin-Joly, E., Debernard, S., Maïbèche-Coisne, M., 2010. A diversity of putative carboxylesterases are expressed in the antennae of the noctuid moth Spodoptera littoralis. Insect Mol Biol 19, 87–97.

El-Sayed A. M., 2021. The Pherobase: Database of Pheromones and Semiochemicals [WWW Document].

Ferkovich, S.M., van Essen, F., Taylor, T.R., 1980. Hydrolysis of sex pheromone by Antennal Esterases of the Cabbage looper, Trichoplusia ni. Chem Senses 5, 33–46.

Ferveur, J.F., 2005. Cuticular hydrocarbons: Their evolution and roles in Drosophila pheromonal communication. Behav Genet.

Friesen, P.D., Miller, L.K., 1987. Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the HindIII K genome

fragment of the baculovirus Autographa californica nuclear polyhedrosis virus. J Virol 61, 2264–2272.

Friesen, P.D., Nissen, M.S., 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol Cell Biol 10, 3067–77.

de Fouchier, A., Fruitet E., Lievers R., Kuperus P., Emerson J., Gould F., Heckel D. G., Groot A. T., 2023. Lipases and carboxylesterases are involved in interspecific pheromone differences between two moth species [manuscript under revision]

Groot, A.T., Estock, M.L., Horovitz, J.L., Hamilton, J., Santangelo, R.G., Schal, C., Gould, F., 2009. 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.

Groot, A.T., Fan, Y., Brownie, C., Jurenka, R.A., Gould, F., Schal, C., 2005. Effect of PBAN on pheromone production by mated Heliothis virescens and Heliothis subflexa females. J Chem Ecol 31, 15–28.

Groot, A.T., Horovitz, J.L., Hamilton, J., Santangelo, R.G., Schal, C., Gould, F., 2006. Experimental evidence for interspecific directional selection on moth pheromone communication. Proc Natl Acad Sci U S A 103, 5858–5863.

Groot, A.T., Santangelo, R.G., Ricci, E., Brownie, C., Gould, F., Schal, C., 2007. Differential attraction of Heliothis subflexa males to synthetic pheromone lures in Eastern US and Western Mexico. J Chem Ecol 33, 353–368.

Groot, A.T., Staudacher, H., Barthel, A., Inglis, O., Schöfl, G., Santangelo, R.G., Gebauer-Jung, S., Vogel, H., Emerson, J., Schal, C., Heckel, D.G., Gould, F., 2013. One quantitative trait locus for intra- and interspecific variation in a sex pheromone. Mol Ecol 22, 1065–1080.

Groot, A.T., Ward, C., Wang, J., Pokrzywa, A., O'Brien, J., Bennett, J., Kelly, J., Santangelo, R.G., Schal, C., Gould, F., 2004. Introgressing pheromone QTL between species: Towards an evolutionary understanding of differentiation in sexual communication. J Chem Ecol 30, 2495–2514.

Heath, R.R., Mclaughlin, J.R., Proshold, F., Teal, P.E.A., 1991. Periodicity of Female Sex Pheromone Titer and Release in Heliothis subflexa and H. virescens (Lepidoptera: Noctuidae). Ann Entomol Soc Am 84, 182–189.

Ishida, Y., Leal, W.S., 2005. Rapid inactivation of a moth pheromone. Proc Natl Acad Sci U S A 102, 14075–14079.

Jallon, J.-M., Wicker-Thomas, C., 2003. Genetic studies on pheromone production in Drosophila. In: The Biosynthesis and Detection of Pheromones and Plant Volatiles. pp. 253–281.

Jurenka, R., 2004. Insect Pheromone Biosynthesis.

Jurenka, R., 2017. Regulation of pheromone biosynthesis in moths. Curr Opin Insect Sci 24, 29–35.

Jurenka, R.A., Roelofs, W.L., 1989. Regulationofpheromonebiosynthesisinmoths. Current Oin 19, 639–644.

Kawahara, A.Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E.F.A., Donath, A., Gimnich, F., Frandsen, P.B., Zwick, A., dos Reis, M., Barber, J.R., Peters, R.S., Liu, S., Zhou, X., Mayer, C., Podsiadlowski, L., Storer, C., Yack, J.E., Misof, B., Breinholt, J.W., 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. Proc Natl Acad Sci U S A 116, 22657–22663.

Lassance, J.M., Groot, A.T., Liénard, M.A., Antony, B., Borgwardt, C., Andersson, F., Hedenström, E., Heckel, D.G., Löfstedt, C., 2010. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. Nature 466, 486–489.

Lievers, R., Groot, A.T., 2016. Disposable polydimethylsiloxane (PDMS)-coated fused silica optical fibers for sampling pheromones of moths. PLoS One 11, 1–15.

Merlin, C., Rosell, G., Carot-Sans, G., François, M.C., Bozzolan, F., Pelletier, J., Jacquin-Joly, E., Guerrero, A., Maïbèche-Coisne, M., 2007. Antennal esterase cDNAs from two pest moths, Spodoptera littoralis and Sesamia nonagrioides, potentially involved in odourant degradation. Insect Mol Biol 16, 73–81.

Mitter, C., Davis, D.R., Cummings, M.P., 2017. Phylogeny and Evolution of Lepidoptera. Annu Rev Entomol 62, 265–283.

Morse, D., Meighen, E., 1984. Detection of pheromone biosynthetic and degradative enzymes in vitro. Journal of Biological Chemistry 259, 475–480.

Morse, D., Meighen, E., 1987. Biosynthesis of the acetate ester precursor of the spruce budworm sex pheromone by an acetyl CoA: Fatty alcohol acetyltransferase. Insect Biochem 17, 53–59.

Müller, H., Becker, A., Palm, G.J., Berndt, L., Badenhorst, C.P.S., Godehard, S.P., Reisky, L., Lammers, M., Bornscheuer, U.T., 2020.

Sequence-Based Prediction of Promiscuous Acyltransferase Activity in Hydrolases. Angewandte Chemie 132, 11704–11709.
Müller, H., Godehard, S.P., Palm, G.J., Berndt, L., Badenhorst, C.P.S., Becker, A.K., Lammers, M., Bornscheuer, U.T., 2021. Discovery and Design of Family VIII Carboxylesterases as Highly Efficient Acyltransferases. Angewandte Chemie - International Edition 60, 2013–2017.
Pfeifer, T.A., Ring, M., Grigliatti, T.A., 2000. Identification and analysis of Lydia, a LTR retrotransposon from Lymantria dispar. Insect Mol Biol 9, 349–356.
Pogue, M.G., 2013. Revised status of Chloridea Duncan and (Westwood), 1841, for the Heliothis virescens species group (Lepidoptera: Noctuidae: Heliothinae) based on morphology and three genes. Syst Entomol 38, 523–542.
Prestwich, G.D., Graham, S.M.G., Handley, M., Latli, B., Streinz, L., Tasayco, M.L.J., 1989. Enzymatic processing of pheromones and pheromone analogs. Experientia 45, 263–270.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. [WWW Document]. R Foundation for Statistical Computing, Vienna, Austria.
Rafaeli, A., Jurenka, R.A., 2003. PBAN regulation of pheromone biosynthesis in female moths. Insect Pheromone Biochemistry and Molecular Biology 107–136.
SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., 1998. The paleontology of intergene retrotransposons of maize. Nat Genet 20, 43–45.
Sheck, A.L., Groot, A.T., Ward, C.M., Gemeno, C., Wang, J., Brownie, C., Schal, C., Gould, F., 2006. Genetics of sex pheromone blend differences between Heliothis virescens and Heliothis subflexa: A chromosome mapping approach. J Evol Biol 19, 600–617.
Teal, P.E.A., Tumlinson, J.H., 1987. The role of alcohols in pheromone biosynthesis by two noctuid moths that use acetate pheromone components. Arch Insect Biochem Physiol 4, 261–269.
Thomson, J.M., Gaucher, E.A., Burgan, M.F., de Kee, D.W., Li, T., Aris, J.P., Benner, S.A., 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. Nat Genet 37, 630–635.
Tillman, J. a, Seybold, S.J., Jurenka, R. a, Blomquist, G.J., 1999. Insect pheromones—an overview of biosynthesis and endocrine regulation. Insect Biochem Mol Biol 29, 481–514.

Vickers, N.J., 2002. Defining a synthetic pheromone blend attraectiv to male Heliothis subflexa under wind tunnel conditions. J Chem Ecol 28, 1255–1267.

Vickers, N.J., Baker, T.C., 1997. 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 A 180, 523–536.

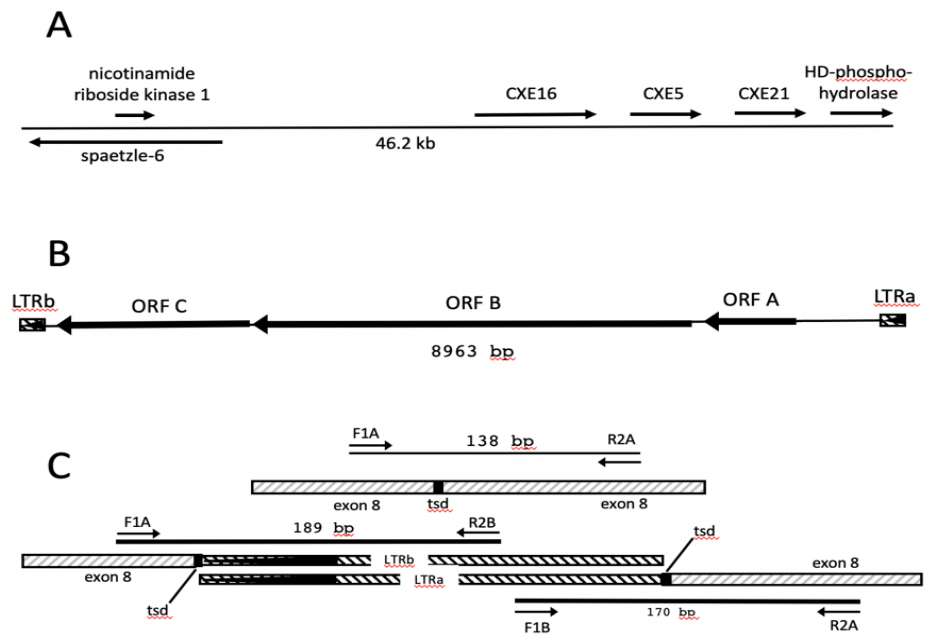
Vogel, H., Heidel, A.J., Heckel, D.G., Groot, A.T., 2010. Transcriptome analysis of the sex pheromone gland of the noctuid moth Heliothis virescens. BMC Genomics 11, 16–18.

Vogt, R.G., 2005. Comprehensive Insect Physiology, Biochemistry, Pharmacology and Molecular Biology, Elsevier.

Wang, Q.H., Gong, Q., Fang, S.M., Liu, Y.Q., Zhang, Z., Yu, Q.Y., 2020. 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.

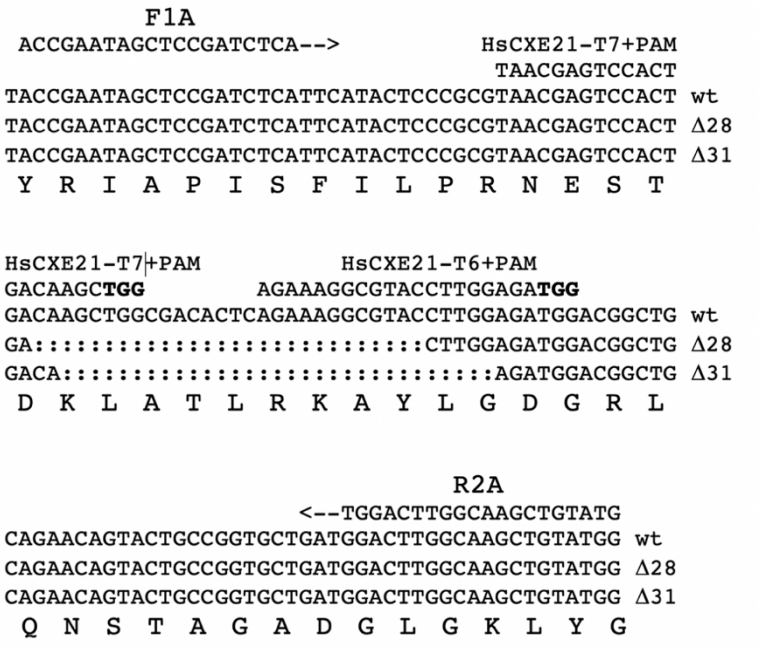
Zhang, J., 2003. Evolution by gene duplication: An update. Trends Ecol Evol 18, 292–298.

Appendix



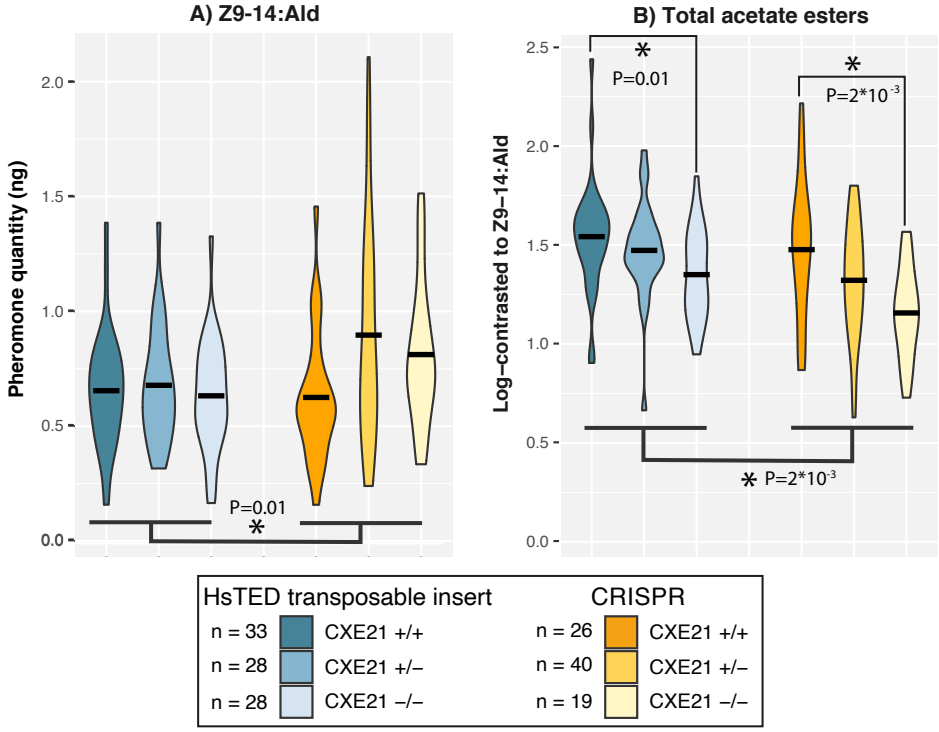
Supplementary Figure S1: Occurrence and detection of HsubTED in the Chromosome 28 esterase cluster.

A) Schematic depiction of the esterase cluster in H. virescens (GenBank Accession ON500499). The gene nicotinamide riboside kinase 1 occurs within an intron of spaetzle-6. H. subflexa has the same orientation of flanking genes and esterases (GenBank Accession ON603978). B) Structure of LTR retrotransposon HsubTED. The three ORFs (open reading frames) are shown, as well as the two identical LTRs (long terminal repeats), GenBank Accession ON603978). C) PCR strategy for detecting the presence of HsubTED in exon 8 of CXE21. tsd = 4 bp target site duplication in exon 8



Supplementary Figure S2: The two CRISPR/Cas9-induced deletions in exon 8 of H. subflexa CXE21.

Screening primers HsubTED-F1A and HsubTED-R2A are the same as in Supplementary Figure S1. Also shown are the two targets T7 and T6.



Supplementary Figure S3: Effect of CXE21 knockouts on the amount of acetates and Z9-14:Ald

A) Z9-14:Ald amount, B) Sum of the three acetate amounts log contrasted by Z9-14:Ald. In blue are the data collected using HsubTED transposable element insertion, and in yellow are the data from the CRISPR experiment. CXE21 +/+ are individuals with two functional copies of CXE21 gene, CXE21 +/- individuals carry one functional copy, and CXE21 -/- have no functional copies. The horizontal black bar represents the mean. The black bar with a star indicates a statistical difference based on the results from one-way Anova followed by a Turkey or Games-Howell post-hoc test to compare the genotype of both experiment separately as well as a two-way Anova, with the genotype and the type of knockout used as explanatory variables.