Evolution of the Integral Membrane Desaturase Gene Family in Moths and Flies

Douglas C. Knipple,*,[1] Claire-Lise Rosenfield,* Rasmus Nielsen,[†] Kyung Man You*,[2] and Seong Eun Jeong[‡]

**Department of Entomology, New York State Agricultural Experiment Station, Cornell University, Geneva, New York 14456, [†]Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York 14850 and [‡]Department of Biological Sciences, Hannam University, Taeduk-Ku, Taejon 300-791, Korea*

Manuscript received February 18, 2002
Accepted for publication September 19, 2002

ABSTRACT

Lepidopteran insects use sex pheromones derived from fatty acids in their species-specific mate recognition system. Desaturases play a particularly prominent role in the generation of structural diversity in lepidopteran pheromone biosynthesis as a result of the diverse enzymatic properties they have evolved. These enzymes are homologous to the integral membrane desaturases, which play a primary role in cold adaptation in eukaryotic cells. In this investigation, we screened for desaturase-encoding sequences in pheromone glands of adult females of eight lepidopteran species. We found, on average, six unique desaturase-encoding sequences in moth pheromone glands, the same number as is found in the genome database of the fly, *Drosophila melanogaster*, *vs.* only one to three in other characterized eukaryotic genomes. The latter observation suggests the expansion of this gene family in insects before the divergence of lepidopteran and dipteran lineages. We present the inferred homology relationships among these sequences, analyze nonsynonymous and synonymous substitution rates for evidence of positive selection, identify sequence and structural correlates of three lineages containing characterized enzymatically distinct desaturases, and discuss the evolution of this sequence family in insects.

A striking feature of the reproductive biology of the Lepidoptera is the use of sex pheromones, which are volatile species-specific chemical signals that are synthesized in terminally differentiated glands of adult female moths and released at appropriate times to attract conspecific males for mating. The hundreds of unique chemical constituents used as sex pheromones in this large taxonomic group are derived from simple fatty acids by similar routes involving desaturation at one or more positions, variable extents of chain shortening by limited β-oxidation, and reductive functional group modification (reviewed in Bjostad *et al.* 1987; Wolf and Roelofs 1989; Tillman *et al.* 1999). It is apparent that desaturases are particularly significant in the generation of structural diversity of lepidopteran sex pheromone components, since these enzymes have evolved diverse substrate specificities, regiospecificities, and stereospecificities to produce unsaturated fatty acid (UFA) precursors with a range of chain lengths, variable positions and numbers of double bonds, and both *Z* (*cis*) and *E* (*trans*) double-bond geometries. In many species, the specificity of sex pheromone signaling is further enhanced by the use of blends of active components derived from UFA precursors produced by two or more enzymatically discrete desaturases.

The homology of lepidopteran pheromone desaturases with the integral membrane desaturases (also referred to as acyl-CoA desaturases) was initially suggested by the analysis of the biochemical properties of Δ11 desaturases isolated from pheromone glands of the cabbage looper moth, *Trichoplusia ni* (Wolf and Roelofs 1986) and the Egyptian cotton leaf worm moth, *Spodoptera littoralis* (Rodriguez *et al.* 1992). Integral membrane desaturases occur ubiquitously in eukaryotic cells, where they play a primary role in the homeostatic regulation of physical properties of lipid membranes in response to cold (Vigh *et al.* 1993; Tiku *et al.* 1996), a universal biological response termed "homeoviscous adaptation" (Hazel and Williams 1990). Integral membrane desa-

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF482898 (*Cpom* CPRQ), AF482899 (*Cpom* KPAE), AF482900 (*Cpom* KPSE), AF482901 (*Cpom* LATE), AF482902 (*Cpom* MPTQ), AF482903 (*Cpom* NPVE), AF482904 (*Cpom* SPTQ), AF482905 (*Hass* GATD), AF482906 (*Hass* KPSE), AF482907 (*Hass* KSVE), AF482908 (*Hass* LPAQ), AF482909 (*Hass* NPVE), AF482910 (*Hass* PPAE), AF482911 (*Hass* QPGE), AF482912 (*Mbra* KPSEa), AF482913 (*Mbra* KPSEb), AF482914 (*Mbra* LPAQ), AF482915 (*Mbra* NPVE), AF482916 (*Pgos* GATD), AF482917 (*Pgos* KPAQ), AF482918 (*Pgos* MPAE), AF482919 (*Pgos* NPAE), AF482920 (*Pgos* NPVE), AF482921 (*Pgos* VASQ), AF482922 (*Pint* AGTQ), AF482923 (*Pint* ASVQa), AF482924 (*Pint* ASVQb), AF482925 (*Pint* GATD), AF482926 (*Pint* IPAE), AF482927 (*Pint* KPSE), AF482928 (*Pint* NPRD), AF482929 (*Pint* NPVE), AF482930 (*Pint* VGTQ), AF482931 (*Psep* GATD), AF482932 (*Psep* KPSE), AF482933 (*Psep* LPAQ), AF482934 (*Psep* NPVE), AF482935 (*Sexi* GATD), AF482936 (*Sexi* KPSE), AF482937 (*Sexi* KSVE), AF482938 (*Sexi* LPAQ), AF482939 (*Sexi* NPVE), AF482940 (*Sexi* VPAE), AF482941 (*Slit* GATD), AF482942 (*Slit* KPSE), AF482943 (*Slit* KSVE), AF482944 (*Slit* LPSQ), and AF482945 (*Slit* NPVE).

[1]*Corresponding author:* Department of Entomology, Cornell University, New York State Agricultural Experiment Station, North St., Geneva, NY 14456. E-mail: dck2@cornell.edu

[2]*Present address:* Department of Molecular and Life Sciences, Pohang University of Science and Technology, Pohong 790-784, South Korea.

turases are structurally unrelated to soluble desaturases that occur predominantly in plants (Shanklin and Cahoon 1998).

The predicted homology of lepidopteran pheromone desaturases and integral membrane desaturases was confirmed by the cloning and functional expression of desaturase-encoding cDNAs isolated from the pheromone glands of *T. ni* (Knipple *et al.* 1998), the corn earworm moth *Helicoverpa zea* (Rosenfield *et al.* 2001), and the light brown apple moth *Epiphyas postvittana* (Liu *et al.* 2002). Each of these desaturases was shown to have enzymatic properties consistent with the known pheromone biosynthetic pathway of the major component(s) of the species from which it was isolated: in the case of *T. ni*, a Δ11 desaturase producing both *Z*11-16:1 and *Z*11-18:1 UFA precursors (Bjostad and Roelofs 1983); in *H. zea*, a Δ11 desaturase producing only the *Z*11-16:1 precursor (Klun *et al.* 1980; Pope *et al.* 1984; Teal and Tumlinson 1986; Jurenka *et al.* 1991); and, in *E. postvittana*, a Δ11 desaturase producing *E*11-16:1, *E*11-14:1, and *E*9,*E*11-14:2 precursors (Foster and Roelofs 1990). In *H. zea*, the presence of minor pheromone components derived from palmitoleic (*Z*9-16:1) and oleic (*Z*9-18:1) acid precursors suggested that a desaturase with Δ9 regiospecificity was also present in the pheromone gland (Klun *et al.* 1980; Pope *et al.* 1984; Teal and Tumlinson 1986; Jurenka *et al.* 1991). In fact, two Δ9 desaturase-encoding transcripts were subsequently found in this tissue: one encoding a desaturase with a substrate preference of 16:0 > 18:0 that is differentially expressed in the cells of the pheromone gland and another encoding a desaturase with a substrate preference of 18:0 > 16:0 that encodes the same Δ9 desaturase that is expressed in larval fat bodies (Rosenfield *et al.* 2001). An apparent ortholog of the latter desaturase having similar sequence, enzymatic properties, and tissue distribution is found in *T. ni* (Liu *et al.* 1999; Rosenfield *et al.* 2001). Two additional *E. postvittana* cDNA sequences have also been shown to encode Δ9 desaturases with enzymatic properties similar to their respective orthologs in *H. zea*, although their expression patterns remain uncharacterized (Liu *et al.* 2002). These studies demonstrate that the integral membrane desaturase gene family has evolved in the Lepidoptera to function not only in normal cellular lipid metabolism, but also in the production of chemical signals used in a sophisticated mate recognition system.

Here we describe our exploration of the sequence space of integral membrane desaturase-encoding transcripts present in the pheromone glands of eight lepidopteran species belonging to four families. We describe the inferred homology relationships of the amino acid sequences encoded by the 48 unique cDNAs isolated in this study, 16 additional lepidopteran desaturase sequences that have been described elsewhere (Knipple *et al.* 1998; Liu *et al.* 1999, 2002; Yoshiga *et al.* 2000; Rosenfield *et al.* 2001; GenBank accession nos. AAL16642, AAF44709, AAG54077, and AAF73073), and 6 desaturase sequences identified in the genome of the fly *Drosophila melanogaster* (FlyBase 1999; Dallérac *et al.* 2000). We report the results of statistical analyses of nucleotide substitution patterns. We infer some of the sequence and structural correlates of regioselectivity and substrate chain length preferences identified among functionally characterized desaturases by the application of computational filters to the protein sequence data sets. We present the results of our analyses in the contexts of the unique aspects of insect biology and, in particular, the evolution of sex and species recognition systems in insects that use compounds derived from UFA intermediates.

MATERIALS AND METHODS

Insects: The insect species from which desaturase-encoding cDNAs were isolated are shown in Figure 1 along with information about the compounds present in their respective pheromones and their inferred biosynthetic pathways. Codling moth (*Cydia pomonella*) pupae were obtained from Dr. Peter Landolt [U.S. Department of Agriculture (USDA) Agriculture Research Service (ARS), Wapato, WA] under applicable permits from the New York State Department of Agriculture and Markets (Albany, NY) and the USDA Plant Protection and Quarantine Service (Riverdale, MD). Male pupae, identified by their sexually dimorphic pigmentation pattern, were culled and the remaining females were transferred into individual 4-cm shell vials capped with a loose cork and placed in a Percival Scientific (Boone, IA) incubator at 25°, 55% relative humidity (RH), with a 16:8 light-dark cycle. Pheromone glands were dissected from adult females 2 days after emergence, 3–4 hr into the scotophase.

Pink bollworm (*Pectinophora gossypiella*) pupae were obtained from Dr. Robert T. Staten (USDA APHIS PPQP, Phoenix, AZ) under applicable permits as above. Females were selected and incubated under the conditions described above, and pheromone glands were dissected from adult females 2 days after emergence.

The following insects were maintained and used as a source of pheromone glands in Korea: the Indian meal moth (*Plodia interpunctella*), the oriental tobacco budworm (*H. assulta*), the cabbage armyworm (*Mamestra brassicae*), the common armyworm (*Pseudaletia separata*), the beet armyworm (*S. exigua*), and the common cutworm (*S. litura*). *P. interpunctella* was reared on artificial diet (Silhacek and Miller 1972) at 27° and 60% RH under a 16:8 light-dark cycle. *H. assulta* was reared on artificial diet (Jeong *et al.* 2001) at 25° and 60% RH under a 16:8 light-dark cycle. Pupae of *M. brassicae*, *P. separata*, *S. exigua*, and *S. litura*, obtained from Dr. Myung-Hee Ohh (Korea Ginseng and Tobacco Research Institute, Taejon, Korea), were kept at 25°, 60% RH under a 16:8 light-dark cycle. Female pupae were selected, and the pheromone glands were harvested as described above. RNAs were isolated and ethanol precipitated as described below and shipped to the D. C. Knipple lab.

Isolation of desaturase-encoding cDNAs: Dissected pheromone glands were kept at −80° until they were extracted with TRIzol (GIBCO BRL, Gaithersburg, MD) to obtain total RNA according to the manufacturer's protocol. RNA was precipitated with ethanol and stored at −80° until further use. First-strand oligo(dT)-primed cDNA was synthesized by using 5 μg total RNA [without poly(A)$^+$ RNA isolation] and a SuperScript

Genus species common name Family	Code	Unsaturated Pheromone Components	Proposed Biosynthetic Pathways (→ known, ⇢ probable)
Cydia pomonella codling moth Tortricidae	*Cpom*	**E8E10-12OH**/Al/Ac E9-12OH E8Z10-12OH Z8E10-OH	16:● —βox→ 12:● —Δ9(E)→ E9-12:● ⇣ monoene (8,10)Ds ⇣ geometric isomers + **E8E10-12:●**
Helicoverpa assulta oriental tobacco budworm Noctuidae	*Hass*	**Z9-16Al**/Ac Z11-16Al/Ac	16:● —Δ9(Z)→ **Z9-16:●** 16:● —Δ11(Z)→ Z11-16:●
Mamestra brassica cabbage armyworm Noctuidae	*Mbra*	**Z11-16Ac**/OH E11-16Ac Z9-16Ac	16:● —Δ11(Z/E)→ **Z11-16:●** + E11-16:● 16:● —Δ9(Z)→ Z9-16:●
Pectinophora gossypiella pink bollworm Gelechiidae	*Pgos*	**Z7Z11-16Ac**/OH **Z7E11-16Ac**/OH	18:● —Δ9(Z)→ Z9-18:● ⇢βox⇢ Z7-16:● ⇣ monoene Ds(s) Δ11(Z/E) ⇣ **Z7Z11-16:● + Z7E11-16:●**
Plodia interpunctella Indian meal moth Pyralidae	*Pint*	**Z9E12-14Ac**/OH/Al	16:● —Δ11(Z)→ Z11-16:● —βox→ Z9-14:● ⇣ monoene Ds Δ12(E) ⇣ **Z9E12-14:●**
Pseudaletia separata common armyworm Noctuidae	*Psep*	**Z11-16Ac**/OH	16:● —Δ11(Z)→ **Z11-16:●**
Spodoptera exigua beet armyworm Noctuidae	*Sexi*	**Z9-14Ac**/OH **Z9E12-14Ac**/OH Z9Z12-14Ac/OH Z11-16Ac/OH	16:● —Δ11(Z)→ Z11-16:● —βox→ Z9-14:● ⇣ monoene Ds Δ12(E) ⇣ **Z9E12-14:●**
Spodoptera litura common cutworm Noctuidae	*Slit*	**Z9E11-14Ac** Z9E12-14Ac	16:● ⇢Δ11(Z)⇢ Z11-16:● ⇢βox⇢ Z9-14:● ⇣ monoene Ds: ⇙Δ12(E) ⇓Δ11(E) **Z9E12-14:●** **Z9E11-14:●**

FIGURE 1.—Lepidopteran species used in this study, their sex pheromone components, and biosynthetic pathways: *C. pomonella* (*Cpom*; ARN *et al.* 1985), *H. assulta* (*Hass*; PARK *et al.* 1996), *M. brassica* (*Mbra*; ATTYGALLE *et al.* 1987), *P. gossypiella* (*Pgos*; HUMMEL *et al.* 1973), *P. interpunctella* (*Pint*; TEAL *et al.* 1995), *P. separata* (*Psep*; KOU *et al.* 1992), *S. exigua* (*Sexi*; TUMLINSON *et al.* 1990), and *S. litura* (*Slit*; TAMAKI *et al.* 1973). The abbreviations used for chemical structures are as follows: Z or E followed by a number indicates the geometry and location of a double bond relative to the functional group; the number following the hyphen indicates fatty acyl chain length; and functional groups are indicated by OH (alcohol), Al (aldehyde), and Ac (acetate). In the pathway diagrams, chemical structures of fatty acids are represented as above, esterified to coenzyme A (symbolized by solid circles); enzymes of β-oxidation are represented by the notation βox; desaturases are indicated by symbols designating their regioselectivity and the geometry of the double bond they form [*e.g.*, Δ9(E) in *C. pomonella*] or by designations indicating the use of an unsaturated substrate and the locations of the double bond formed [*e.g.*, monoene Ds Δ12(E) in *P. interpunctella*]. Major components are shown in boldface type.

preamplification system for first-strand cDNA synthesis (GIBCO BRL no. 18089-011) according to the manufacturer's protocol.

To obtain desaturase-encoding cDNAs, each pheromone gland cDNA preparation was used in 12 separate PCR reactions, each containing 1 of 12 partially deconvoluted oligonucleotide primer pools of 5′ primers that collectively hybridize to all possible target sequences on the antisense strand of the GAHRLW(A/T/S) amino acid sequence motif (GAHR primers), and a degenerate 3′ primer that hybridizes to all possible target sequences on the sense strand of the EGFHNYH amino acid sequence motif (EGFH primers; ROSENFIELD *et al.* 2001).

PCR reactions were performed in a Perkin-Elmer (Norwalk, CT) model 480 thermal cycler. Each 50-μl reaction contained 0.2 mM dNTPs, 2 mM $MgCl_2$, 0.4 μM each of 5′ and 3′ primer pools, and variable amounts of cDNA templates. Following a 5-min preincubation at 95°, the PCR reactions were started by adding 0.4 μl Taq polymerase (Perkin-Elmer) and run for 35 cycles of 95°, 1 min; 60°, 1 min; and 72°, 2 min. PCR amplification products were analyzed by electrophoresis in agarose gels. Specific amplification products were gel purified, ligated into plasmid pCR2.1 (Invitrogen, San Diego), and cloned in *Escherichia coli*. DNA sequencing was done by a variation of the dideoxynucleotide terminator method (SANGER *et al.* 1977), using a model 377 Applied Biosystems (Foster City, CA) automated DNA sequencer and the Prism dye-labeling chemistry according to the manufacturer's protocol.

DNA and protein sequence analysis: Standard DNA and protein sequence analyses performed in the course of this investigation used BLASTn and BLASTp searches (ALTSCHUL *et al.* 1997), accessed sequence databases via the Entrez Search and Retrieval System of the National Center for Biotechnology Information and via FlyBase of the *Drosophila* Genome Consortium (FLYBASE 1999), and used other tools contained in the MacVector 7.0 software suite (Oxford Molecular, Palo Alto, CA). Multiple sequence alignments of deduced amino acid sequences were obtained by using the CLUSTALW (1.4) algorithm (THOMPSON *et al.* 1994; HIGGINS *et al.* 1996) and guide trees were generated by the neighbor-joining method (SAITOU and NEI 1987). The reliability of branching points in inferred trees was tested by generating consensus bootstrap trees (FELSENSTEIN 1985) that collapsed nodes occurring in <50% of trees from 1000 resampling iterations.

Nonsynonymous and synonymous substitution rates were estimated and likelihood-ratio tests of neutrality (NIELSEN and YANG 1998; YANG *et al.* 2000) were performed using the PAML package (YANG 2000). In this test a model (M7), which assumes that the rate ratio of nonsynonymous to synonymous substitutions ($\omega = d_N/d_S$) is β-distributed, is compared to a model (M8) that includes an extra discrete category of sites in which ω is assumed to be a free parameter. If M8 provides a significantly better fit to the data than does M7, as determined by a likelihood-ratio test, and if the estimate of ω in the extra category is larger than one, positive selection is inferred. An additional analysis was performed in which the d_N/d_S for each branch of the phylogeny was estimated (YANG and NIELSEN 1998, 2002). The analysis was conducted separately on the six major sequence groups identified in the bootstrap analysis.

Methods to identify class-specific features of encoded amino acids included a modification of a subroutine of the evolutionary trace methodology (LICHTARGE *et al.* 1996), which was

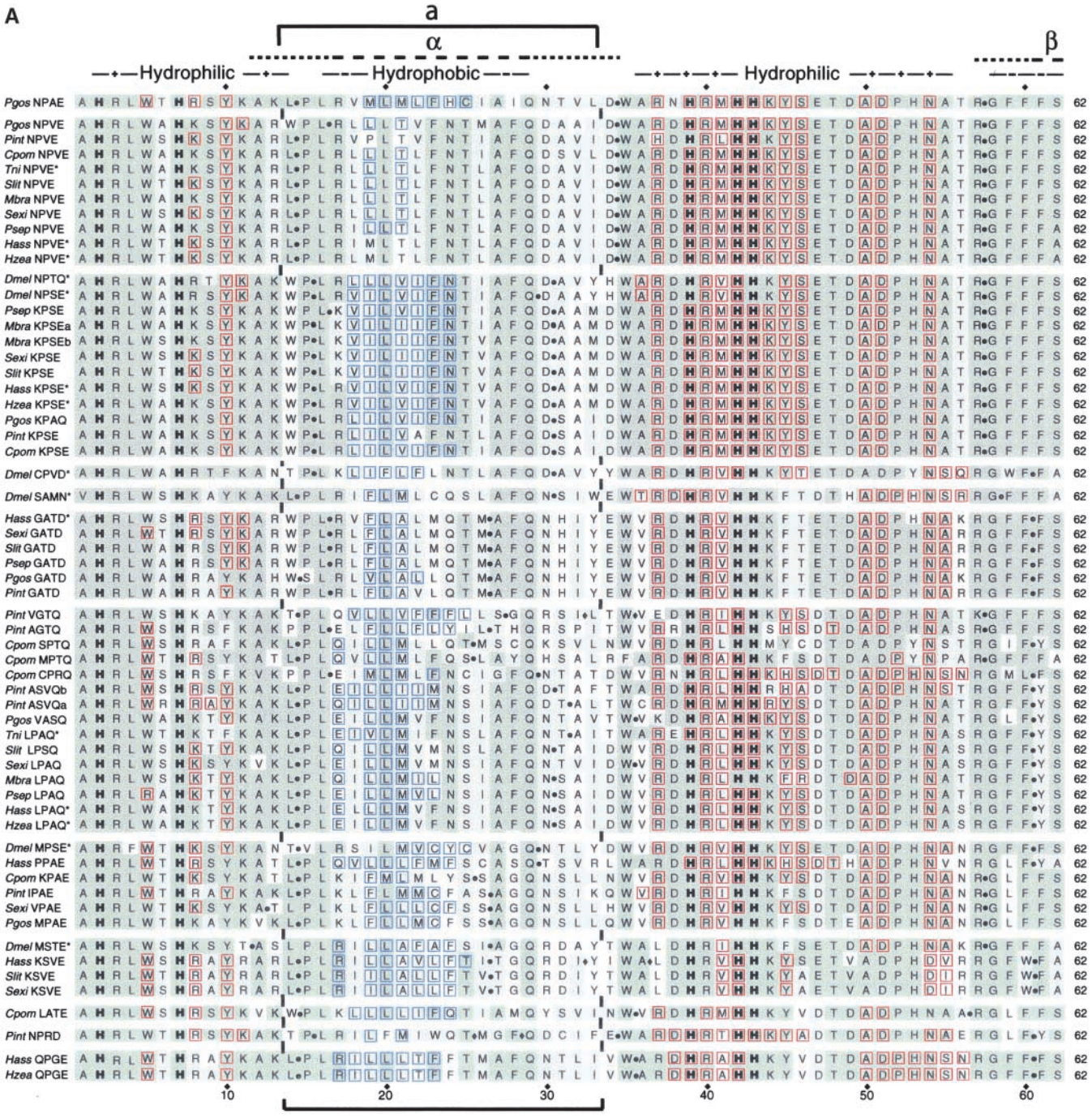


FIGURE 2.—Aligned amino acid sequences of desaturase core domains (delimited by GAHR and EGFH motifs): (A) aligned amino acid positions 1–62; (B) positions 63–124; (C) positions 125–185. Numbers in right-hand columns indicate the number of amino acids in a sequence minus gaps. Gray backgrounds indicate amino acid conservation: identities (dark gray) and conservative substitutions (light gray; see Table 2 legend). Boxed regions (identified at top as a–e) are domains analyzed by the evolutionary trace method (LICHTARGE *et al.* 1996) to search for functional subtype-specific sequence motifs (see also Figure 4). Patterns from Kyte-Doolittle hydropathy plots (KYTE and DOOLITTLE 1982) are indicated as follows: Transition points between calculated hydrophilic and hydrophobic domains are represented by solid circles (definitive transitions) or solid diamonds (transitions that reverse a short distance downstream), and amino acid positions with extreme hydrophilic or hydrophobic values are framed, respectively, in red ($\geq +1.9$) or in blue (≤ -1.9). Regions of hydrophobic or hydrophilic character are identified at the top; domains exhibiting functional class-specific patterns are designated α, β, γ, δ, and ε. Conserved histidine residues that are components of three histidine boxes that are essential for desaturase catalytic activity are shown in boldface type (see also Figures 5 and 6). Other labeled features are transmembrane domains 3 and 4 (TM3 and TM4 in B and C, respectively) and the signature motif at positions 165–168 (green residues in Figure 2C).

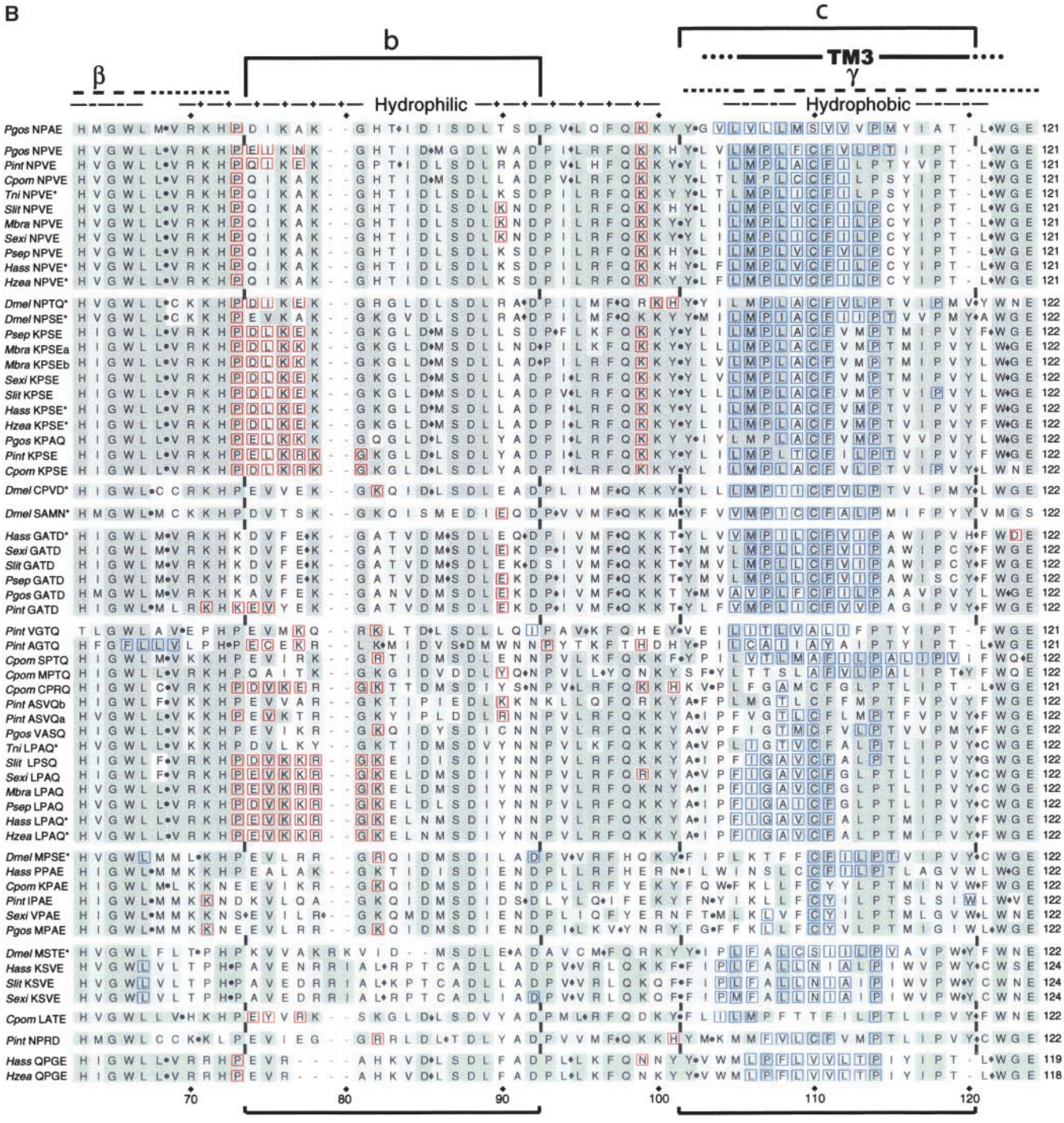


FIGURE 2.—*Continued.*

applied to sequence groups onto which desaturases with defined functional properties were mapped. The Protein Analysis Toolbox of MacVector 7.0 was used to obtain hydrophobicity and hydrophilicity values for individual amino acid residues of deduced polypeptides, from which Kyte/Doolittle plots were graphed (KYTE and DOOLITTLE 1982).

Predictions of transmembrane helices and other structural features of inferred proteins were obtained by using the following programs: TMHMM-2.0 (SONNHAMMER *et al.* 1998), HMMTOP (TUSNÁDY and SIMON 1998, 2001), TMpred (HOFMANN and STOFFEL 1993), SOSUI (HIROKAWA *et al.* 1998; MITAKU and HIROKAWA 1999; MITAKU *et al.* 1999), and MEMSAT (JONES *et al.* 1994).

RESULTS

Isolation of desaturase cDNAs: We have extended the application of a reverse transcriptase (RT)-PCR-based

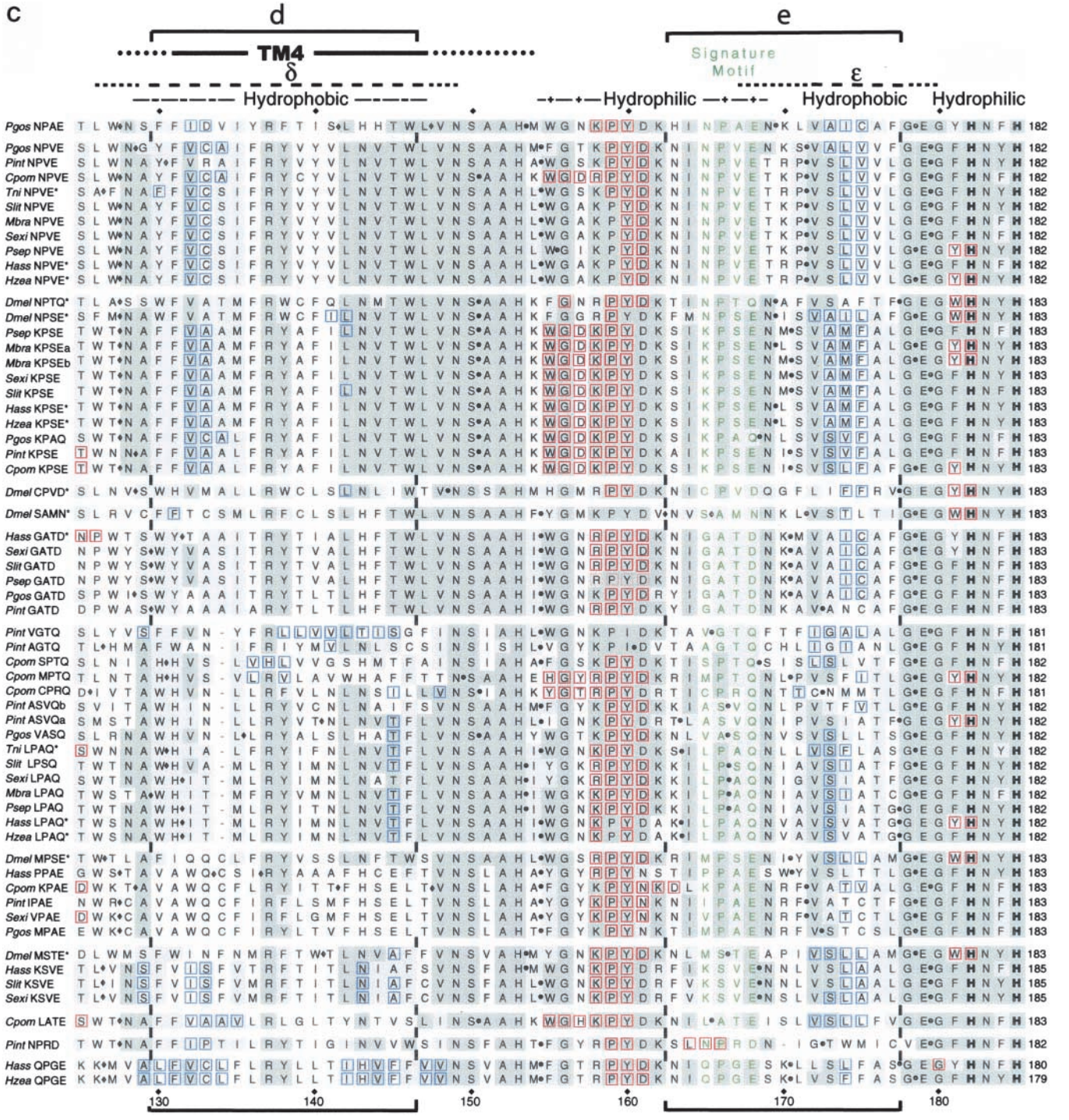
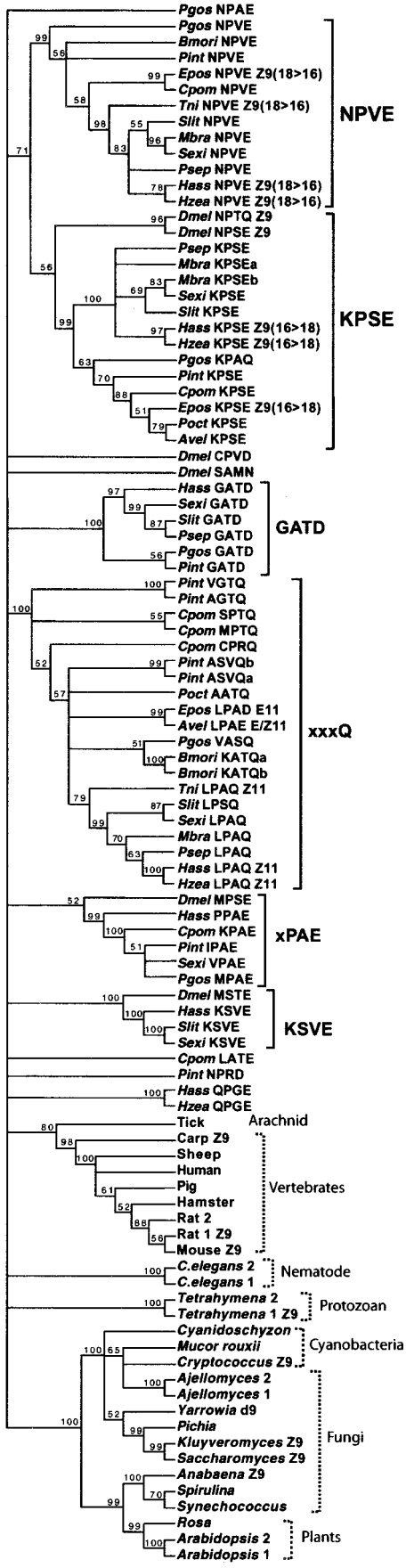


FIGURE 2.—*Continued.*

homology probing screen that we previously used to isolate multiple desaturase-encoding cDNA fragments from pheromone gland RNA of *H. zea* (ROSENFIELD *et al.* 2001). In this screen, we isolated and sequenced cDNA clones containing ~550-bp PCR amplification products encoding open reading frames (ORFs) delimited by the conserved GAHR and EGFH sequence motifs corresponding to desaturase core domains from the following eight species of lepidopteran insects: *C. pomonella*, *H. assulta*, *M. brassicae*, *P. gossypiella*, *P. interpunctella*, *P. separata*, *S. exigua*, and *S. litura*. Among the 462 cDNA clones sequenced, we identified a total of 48 unique cDNAs with between 4 and 9 from each species. The number of clones recovered for each unique se-



quence was highly variable, ranging from as few as 1 (in 11 instances) to as many as 56.

Homology relationships of insect desaturases: We used the CLUSTALW program (Thompson *et al.* 1994; Higgins *et al.* 1996) to align amino acid sequences encoded by the 48 unique desaturase cDNAs isolated in this study along with the following previously described amino acid sequences of desaturase core domains: two of *T. ni* (Knipple *et al.* 1998; Liu *et al.* 1999), four of *H. zea* (Rosenfield *et al.* 2001), two of *D. melanogaster* (Dallérac *et al.* 2000), and four additional sequences of *D. melanogaster* identified in a BLAST search of the Drosophila genome database (Figure 2). A bootstrap tree (Felsenstein 1985) shows the statistically supported homology relationships of the above desaturase sequences in addition to 3 previously described sequences from the silk moth, *Bombyx mori* (family Bombycidae; Yoshiga *et al.* 2000); 3 from *E. postvittana* (family Tortricidae; Liu *et al.* 2002); 2 from the redbanded leafroller moth, *Argyrotaenia velutinana* (family Tortricidae; GenBank accession nos. AAF73073 and AAL16642); 2 from *Planotortrix octo* (family Tortricidae; GenBank accession nos. AAF73073, AAG54077); and 28 others from distantly related phylogenetic groups (Figure 3).

The nomenclature that we propose in this article for insect desaturases incorporates an abbreviated biological species name [*e.g.*, *Spodoptera litura* (*Slit*)] and a sequence-specifying name (signature motif) based on a group of four amino acids at positions 165–168 (Figure 2C). In a few instances, two closely related sequences with identical signature motifs were isolated from the same species, in which case they were differentiated with suffixes, *e.g.*, *Mbra* KPSEa and *Mbra* KPSEb. This scheme also provides the basis for naming the sequence lineages defined in Figure 3, specifically, by deriving a consensus name on the basis of the most prevalent signature motif within a supported grouping of lepidopteran sequences. Thus, the sequence lineages were named NPVE, KPSE, GATD, xxxQ, KSVE, and xPAE. Four orthologs of *D. melanogaster* mapped onto three of these lineages: *Dmel* desat1 and *Dmel* desat2 (Dallérac *et al.* 2000; here called *Dmel* NPSE and *Dmel* NPTQ, respectively) onto the KPSE lineage, *Dmel* MSTE onto the KSVE lineage, and *Dmel* MSPE onto the xPAE lineage. The five lepidop-

Figure 3.—Bootstrapped sequence tree displaying homology relationships of the 60 aligned sequences in Figure 2, plus 3 previously described sequences from the silk moth, *B. mori* (family Bombycidae; Yoshiga *et al.* 2000); 3 from the light brown apple moth, *E. postvittana* (family Tortricidae; Liu *et al.* 2002); 2 from the redbanded leafroller moth, *A. velutinana* (family Tortricidae; GenBank accession nos. AAF73073 and AAL16642); 2 from *P. octo* (family Tortricidae; GenBank accession nos. AAF73073 and AAG54077); and 28 others from distantly related phylogenetic groups.

TABLE 1

Parameter estimates based on model M8 of Yang *et al.* (2000)

Lineage	No. of sequences	Positive selection	d_N/d_S	Parameter estimates[a]				
				p	q	κ	ω	p_0
KPSE	15	No	0.028	0.30	9.56	1.60	N/A	0.0
NPVE	12	No	0.033	0.21	5.67	1.71	N/A	0.0
GATD	6	No	0.034	0.37	10.15	1.79	0.36	0.005
xxxQ	20	No	0.072	0.67	8.55	1.62	N/A	0.0
xPAE	6	No	0.049	0.57	10.55	1.36	N/A	0.0
KSVE	4	No	0.039	0.54	99.00	2.22	0.09	0.38

[a] p and q are the parameters of the β-distribution. κ is the estimate of the d_N/d_S ratio for the category of sites, of frequency p_0, for which the d_N/d_S ratio is free to vary. ω is the transition/transversion rate ratio and d_N/d_S is the average rate ratio on nonsynonymous to synonymous mutations. The presence of positive selection has been tested using a likelihood-ratio test as described in Yang *et al.* (2000).

teran sequences *Pgos* NPAE, *Cpom* LATE, *Pint* NPRD, *Hass* QPGE, and HzPGDs3 (Rosenfield *et al.* 2001; the latter here called *Hzea* QPGE) and the two fly sequences *Dmel* CPVD and *Dmel* SAMN are sufficiently divergent that their assignment to one of the six named lineages is not supported.

Analysis of d_N/d_S nucleotide substitution ratios: None of the likelihood-ratio tests for positive selection within the analyzed lineages revealed any positive results (Table 1). The average d_N/d_S ratio was very low in all six lineages, varying between 0.028 and 0.072, suggesting that strong purifying selection is acting and that all lineages remain functional. Other parameter estimates were also consistent among groups, with the transition/transversion rate ratio varying between 1.36 and 2.22 and the estimates of the distributions of d_N/d_S ratios among sites being very similar. Estimates of the d_N/d_S ratio obtained for all branches of the phylogeny revealed none with d_N/d_S values >1, with the exception of the branch leading to sequences *Pint* AGTQ and *Pint* VGTQ of the xxxQ lineage. Using the codon-based likelihood method to detect adaptation at individual sites (Yang and Nielsen 2002), positive selection was inferred for amino acid positions 71, 117, 131, 160, 164, 166, and 173 (data not shown), of which the latter four closely flank or map onto the signature motif, which is hypervariable in the xxxQ lineage.

Amino acid sequence conservation within lineages: We calculated sequence identities and similarities (identities plus conservative substitutions) for the six named lineages, both with and without *D. melanogaster* orthologs where present (Table 2). Considering only lepidopteran sequences, four lineages exhibit substantial conservation, specifically KPSE, NPVE, GATD, and KSVE. The xxxQ lineage is the most degenerate, with only 32% similarities, and the xPAE lineage is intermediate between these extremes.

Within the xxxQ lineage, the subgrouping of sequences from species comprising the more ancestral families, specifically *C. pomonella* (this study), *E. postvittana* (Liu *et al.* 2002), *A. velutinana* (GenBank accession no. AAF44709), *P. octo* (GenBank accession no. AAF73073) of the family Tortricidae, *P. interpunctella* of the family Pyralidae, *P. gossypiella* of the family Gelechiidae (this study), and *B. mori* of the family Bombicidae (Yoshiga *et al.* 2000), was highly divergent compared to the subgrouping of sequences from species comprising the modern family Noctuidae, specifically *H. zea* (Rosenfield *et al.* 2001), *H. assulta* (S. E. Jeong, K. M. You, C.-L. Rosenfield and D. C. Knipple, unpublished data), *M. brassica*, *P. separata*, *S. exigua*, and *S. litura* (this study; Table 2). In contrast, both modern and ancestral subgroupings of the KSPE lineage have high conservation.

Functional class-specific amino acid sequence motifs: We used a modification of the evolutionary trace (ET) method to search for positions in the aligned desaturase amino acid sequences that are variable between functional classes but conserved within them (Lichtarge *et al.* 1996). In principle, positions identified by the ET method reflect the effect of selection acting to preserve structural determinants of unique functional properties, which, in the present case, are the discrete enzymatic properties of the integral membrane desaturase family of proteins. For the purpose of this analysis, we defined the following functional classes of lepidopteran sequences: the NPVEs (containing sequences encoding Δ9 desaturases with an 18-carbon substrate chain-length preference), KPSEs (containing sequences encoding Δ9 desaturases with a 16-carbon substrate chain-length preference), the combined NPVE and KPSE subgroups (containing sequences encoding Δ9 desaturases), and the modern subgroup of the xxxQ lineage (containing sequences encoding Δ11 desaturases that use saturated fatty acid substrates).

In the first step of this procedure, we aligned the deduced amino acid sequences encoded by lepidopteran desaturase cDNAs for which the above functional properties are established (Knipple *et al.* 1998; Liu *et al.* 1999; Rosenfield *et al.* 2001; S. E. Jeong, K. M. You, C.-L. Rosenfield and D. C. Knipple, unpublished data).

TABLE 2

Amino acid sequence conservation within lineages and subgroups of insect desaturases

Lineage or subgroup[a]	n[b]	AAs[c]	% identities[d]	% similarities[e]
KPSE	13	183	79	92
KPSE (ancestral branch)	6	183	85	95
KPSE (modern branch)	7	183	92	98
KPSE (+*Dmel*)	15	183	70	87
NPVE	12	182	61	80
GATD	6	183	76	89
xxxQ	20	181/182	23	32
xxxQ (ancestral branch)	13	181/182	23	33
xxxQ (modern branch)	7	182	64	85
xPAE	5	183	50	66
xPAE (+*Dmel*)	6	183	38	54
KSVE	3	185	89	97
KSVE (+*Dmel*)	4	185	58	74

Comparisons reflect aligned amino acid sequences of desaturase core domains (Figure 2).

[a] Groupings are as shown in the bootstrap tree in Figure 3. The designation (+*Dmel*) indicates that the orthologous *D. melanogaster* sequences were included in the comparison.

[b] Number of sequences in the grouping.

[c] Number of amino acids in each sequence category excluding alignment gaps.

[d] Percentage of aligned positions with identical amino acids.

[e] Percentage of aligned sites with identical amino acids or conservative substitutions as follows: L, V, I, M (large apolar); G, A, S, T, P (small apolar); F, Y, W, (H) (cyclic); D, E, N, Q (acidic); K, R, (H) (basic).

We then identified sequence domains between the GAHR and EGFH sequence motifs that are at least 15 amino acids in length and that have ≥50% of positions that are variable between functional classes but conserved within at least one functional class. These variable domains were mapped onto the aligned sequences shown in Figure 2 (bracketed regions labeled a, b, c, d, and e). Domains c and d contain significant portions of the TM3 and TM4 transmembrane helices, respectively, and domain e contains the signature motif (Figure 2C). The derived consensus sequences of these domains for the four functional classes defined above are shown in Figure 4. In contrast to the stringent criterion of invariance (amino acid identity) used in the original evolutionary trace method (Lichtarge *et al.* 1996), we allowed conservative substitutions (Bordo and Argos 1991; Table 2, Figure 4) to increase the sensitivity of the test. ET analysis of domains c and e found significant numbers of functional class-specifying positions in comparisons of Δ9 desaturase functional classes, *i.e.*, KPSE and NPVE desaturases with different substrate preferences (Figure 4; shaded positions and upper scores in CSPS column). ET analysis of domains a, c, and d revealed significant numbers of functional class-specifying positions in comparisons of the Δ9 and Δ11 desaturase functional classes (Figure 4; boxed positions and lower scores in CSPS column). Besides amino acid substitutions, gaps in the aligned sequences in both domains c and d were conspicuous class-specifying characters of the NPVE and xxxQ functional classes, respectively.

Hydropathy analysis of sequence lineages: We performed hydropathy analysis (Kyte and Doolittle 1982) on the sequences shown in Figure 2 and applied the results of two filters onto the linear sequence data to search for lineage- and functional class-specific patterns of hydrophobic and hydrophilic features. The first filter identified the transition points between hydrophilic (positive) and hydrophobic (negative) regions (Figure 2; solid circles represent definitive transition points, and solid diamonds represent transition points that are reversed within a short distance), whereas the second filter identified amino acid positions with hydrophilic values ≥1.9 and hydrophobic values ≤−1.9 (Figure 2; red boxes indicate positions exceeding the hydrophilic threshold, and blue boxes, positions exceeding the hydrophobic threshold).

The filter that maps the positions and lengths of regions delimited by hydrophobic-to-hydrophilic transitions identified five interesting hydrophobic areas displaying functional class- or lineage-specific patterns (designated α, β, γ, δ, and ε in Figure 2). Hydrophobic domain α mapped onto ET domain a (Figure 2A); β, a short distance upstream of ET domain b (Figure 2, A and B); γ, onto ET domain c (TM3; Figure 2B); δ, onto ET domain d (TM4; Figure 2C); and ε, onto ET domain e downstream from the signature motif (Figure 2C). The locations of the transition points of these regions were fairly uniform within any given lineage, with the exception of the xxxQ lineage, where they were quite heterogeneous.

The filter that marks amino acid positions exceeding the hydrophilic/hydrophobic thresholds is extremely sensitive to amino acid substitutions and, consequently, most of the patterns formed within classes lack sharp

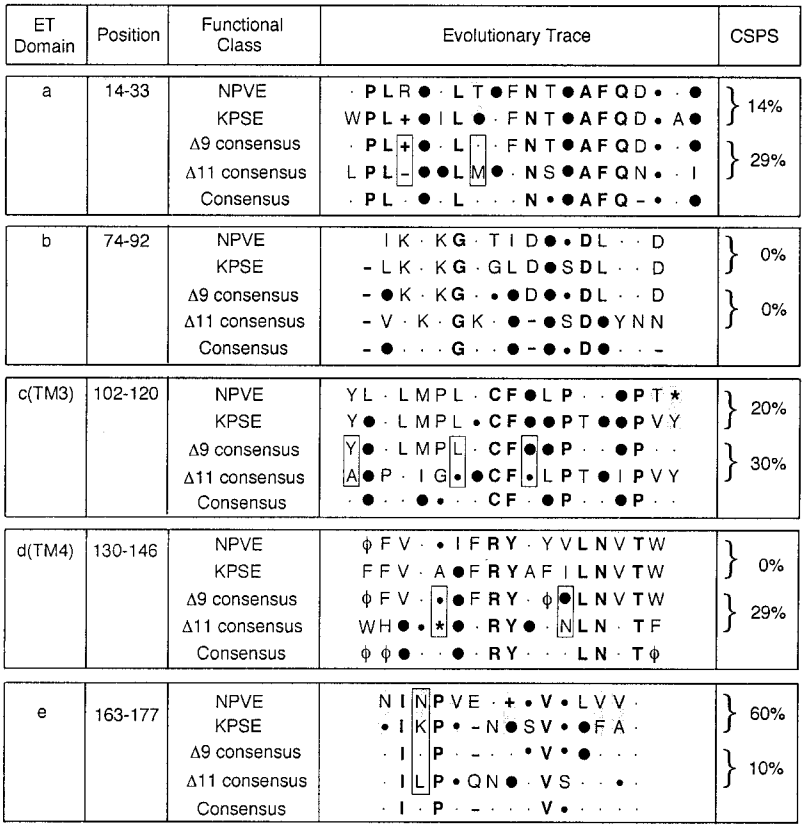


FIGURE 4.—Evolutionary trace (ET) analysis (LICHTARGE *et al.* 1996) of five nonconserved domains of lepidopteran desaturases, showing consensus sequences of the functional classes NPVE, KPSE, desaturases with Δ9 regiospecificity (NPVEs + KPSEs), and desaturases with Δ11 regiospecificity (modern xxxQs). ET domains and position coordinates are as in Figure 1. Positions that are conserved in all members of a class are indicated by the single-letter code of that amino acid. Conservatively substituted amino acids are indicated as follows: ●, large apolar (L, V, I, and M); •, small apolar (G, A, S, T, and P); ϕ, cyclic (F, Y, W, and H); −, acidic (D, E, N, and Q); +, basic (K, R, and H). Gaps (*) are considered as valid characters. Nonconserved positions are indicated with a small dot. Functional class-specific differences between NPVE and KPSE Δ9 desaturases are shaded; those between desaturases with Δ9 regiospecificity and Δ11 regiospecificity are boxed. The class-specific position score (CSPS) is the percentage of class-specifying positions among nonconserved positions.

definition; *i.e.*, they are "fuzzy." Nevertheless, a number of differences are clearly discernible between discrete lineages or functional classes. In particular, two regions were found between the GAHR and EGFH sequence motifs, having significant numbers of positions with extreme hydrophobic character, *i.e.*, hydrophobic domains α (Figure 2A) and γ (TM3; Figure 2B). In addition, three regions with significant numbers of positions with extreme hydrophilic character, *i.e.*, between hydrophobic domains α and β (inclusive of the second histidine box; Figure 2A), between β and γ (Figure 2B), and in the interval between δ (TM4) and the signature motif (Figure 2C), were identified.

Predictions of transmembrane domains: We used five public domain transmembrane prediction programs (see MATERIALS AND METHODS) to analyze each of the partial sequences of this study in addition to all available full-length lepidopteran desaturase sequences. We found that the predicted locations of transmembrane helices TM1, TM2, and TM3 were conserved for all of the available full-length sequences and that the location of the TM3 helix was conserved across all sequences of the present data set. In contrast, the predicted location of the TM4 helix varied substantially between the major subgroups of the two largest lineages and the three smaller lineages, but was conserved within these groupings (summarized in Figure 5). An unanticipated result was the identification by some programs of hydrophobic domain α as a transmembrane domain in many of the sequences. We also examined the positions of tryptophan residues, which frequently occur in integral membrane proteins at the boundaries of hydrophilic domains and transmembrane helices. We found that tryptophans were highly conserved in most positions, except in the short hydrophilic interval between TM3 and TM4 where polymorphisms cleanly defined each of the functional classes and lineages (Figure 5).

DISCUSSION

Expansion of the desaturase gene family in a common ancestor of moths and flies: The number of desaturase-encoding sequences that we found in lepidopteran pheromone glands is significantly larger than that found in the characterized genomes of other eukaryotic organisms (typically, one to three genes per haploid genome), with the exception of the fly *D. melanogaster*, which has six (FLYBASE 1999). In any given species, the number of unique cDNAs we isolated exceeded the minimum number of transcripts required to encode enzymatically distinct desaturases capable of producing the known or inferred UFA precursors of sex pheromone components as well as the normal cellular UFAs. The antiquity of the duplication events that led to the establishment of the six desaturase lineages defined in this investigation is supported by two observations. First, when desaturase

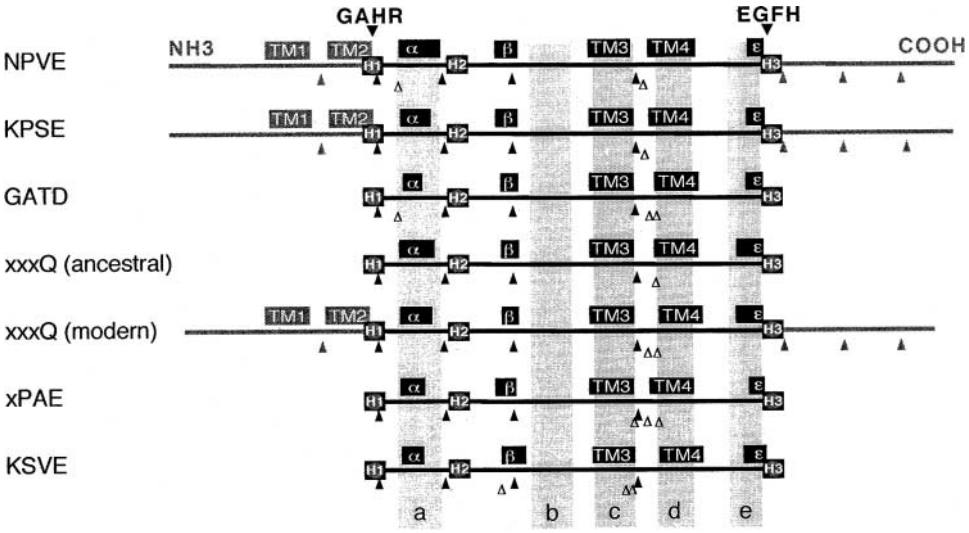


FIGURE 5.—Diagram showing the positions and sizes of transmembrane domains calculated by five structure prediction programs (TMHMM, HMMTOP, TMpred, SOSUI, and MEMSAT) in the GAHR to EGFH regions of the desaturase lineages described in this study. The amino-terminal ends (including transmembrane domains TM1 and TM2) and carboxy-terminal ends shown (shaded) for NPVE, KPSE, and modern xxxQ sequences are derived from previously characterized full-length amino acid sequences. The predicted transmembrane domains are located relative to the following elements corresponding to the amino acid positions in Figure 2: conserved histidine boxes H1–H3; ET domains a–e; and hydrophobic domains α, β, γ (TM3), δ (TM4), and ε, determined from hydropathy plots. The locations of conserved tryptophan residues are identified with triangles: large solid triangles are conserved across all sequences; small open triangles are lineage specific (except for *Pint* VGTQ and *Pint* AGTQ of the xxxQ lineage, which have the conserved pattern of the KSVEs).

sequences from other taxa are used in constructing a bootstrap tree (FELSENSTEIN 1985) only four additional strongly supported sequence lineages result (Figure 3): one composed of sequences from fungi, cyanobacteria, and higher plants; another composed of vertebrate sequences and an arachnid sequence; and two composed of two sequences each from the worm *Caenorhabditis elegans* and the ciliated protozoan *Tetrahymena thermophila* (Figure 3). Second, the orthology of four *D. melanogaster* sequences with three of the lineages defined by lepidopteran sequences is consistent with the existence of these gene lineages in an arthropod ancestor prior to the divergence of dipteran and lepidopteran orders from a common ancestor ~280 million years ago (BRIGGS *et al.* 1993). The occurrence of pheromones derived from UFA precursors in flies and other insect orders (TILLMAN *et al.* 1999) indicates that the recruitment of desaturase function for mate recognition must have occurred relatively early in the evolution of the class Insecta. We suggest that the large repertoire of desaturase genes in the Lepidotera revealed by this study reflects the preservation of specific sequence lineages of an expanded gene family as a result of mutational and selective forces acting according to the subfunctionalization model of duplicate gene evolution (FORCE *et al.* 1999; LYNCH and FORCE 2000).

Functionality of insect desaturase sequences: To date, all of the functionally characterized insect desaturases map onto either the highly conserved NPVE and KSVE lineages (LIU *et al.* 1999; ROSENFIELD *et al.* 2001; LIU *et al.* 2002; S. E. JEONG, K. M. YOU, C.-L. ROSENFIELD and D. C. KNIPPLE, unpublished data) or the divergent xxxQ lineage (KNIPPLE *et al.* 1998; ROSENFIELD *et al.* 2001; LIU *et al.* 2002; S. E. JEONG, K. M. YOU, C.-L. ROSENFIELD and D. C. KNIPPLE, unpublished data; GenBank accession no. AF416738; Figure 3). Whereas all three of the above lineages contain at least one cDNA from each of the eight species of this study, the other three lineages were incomplete. Specifically, we did not find members of the GATD lineage in two species (the noctuid *M. brassicae* and the tortricid *C. pomonella*); we did not find KSVE sequences in five species (*C. pomonella*, *M. brassicae*, *P. gossypiella*, *P. interpunctella*, and *P. separata*); and we did not find xPAE sequences in three (*M. brassicae*, *P. separata*, and *S. litura*). Plausible explanations for not obtaining these desaturase-encoding cDNAs include divergence in the sequences targeted by the oligonucleotide primers used in the PCR screen, negligible transcript accumulation in the pheromone glands, and gene loss. Although no function has been ascribed to any sequence that maps onto the latter three lineages, the codon-based likelihood analyses performed with PAML (YANG 2000) indicate that strong purifying selection is acting in all analyzed sequences (Table 1) and that these three lineages are also likely to be protein coding and functional. Of the seven sequences that did not map onto any of the sequence lineages defined here, no data suggest functionality, with the exception of *Dmel* SAMN, which has a recessive lethal phenotype associated with a transposable element insertion in the coding region of the *Dmel SAMN* gene (FLYBASE 1999). This sequence is most similar to the lepidopteran GATD sequences, but its orthology to that lineage is not statistically supported.

Divergent patterns of expression and biological roles of NPVE and KPSE Δ9 desaturases in the Lepidoptera: NPVE transcripts are found in both sexes in diverse tissues throughout development in the three species in which their spatial and temporal patterns of occurrence

have been investigated (Liu *et al.* 1999; Yoshiga *et al.* 2000; Rosenfield *et al.* 2001). These broad patterns of expression and the UFA products formed by the functionally characterized NPVEs (*i.e.*, *Z*9-18:1 > *Z*9-16:1; Liu *et al.* 1999; Rosenfield *et al.* 2001; Liu *et al.* 2002; S. E. Jeong, K. M. You, C.-L. Rosenfield and D. C. Knipple, unpublished data) are consistent with a primary biological role in cold adaptation, which we suggest may be shared by all the members of this highly conserved lineage. The fact that there is no *D. melanogaster* ortholog of the NPVE lineage indicates that this lineage either was lost from the genome of a dipteran ancestor of *D. melanogaster* or was established by a gene duplication event in the Lepidoptera after the divergence of flies and moths from a common ancestor. In either case, it is apparent that *D. melanogaster* must use a desaturase encoded by another sequence lineage for cold adaptation.

The differential expression of KPSE sequences in the pheromone glands of *H. assulta* (S. E. Jeong, K. M. You, C.-L. Rosenfield and D. C. Knipple, unpublished data) and *H. zea* (Rosenfield *et al.* 2001) implicates the KPSE lineage in sex pheromone biosynthesis in these species. The pheromone of *H. assulta* consists of a major component derived from a *Z*9-16:1 precursor and an essential minor component derived from a *Z*11-16:1 precursor (Cork *et al.* 1992; Park *et al.* 1996), whereas *H. zea* has the opposite major and minor components (Klun *et al.* 1980; Pope *et al.* 1984; Teal and Tumlinson 1986; Jurenka *et al.* 1991). Functional expression studies of these and other members of the KPSE lineage show that the substrate preference of these desaturases (*i.e.*, 16:0 > 18:0) is consistent with the formation of the precursor of the major pheromone component of *H. assulta* and the minor component of *H. zea*. Furthermore, the relative abundance of KPSE transcripts in pheromone gland RNA *vs.* the desaturase-encoding transcripts that produce the *Z*11-16:1 precursor (*i.e.*, *Hass* LPAQ and *Hzea* LPAQ) is positively correlated with the major and minor products in both species (*i.e.*, the KPSE transcript is the most abundant in *H. assulta* and the second most abundant in *H. zea*).

While the evidence for a primary biological role of KPSE desaturases in sex pheromone biosynthesis in heliothine species is compelling, the deduced pathways of the pheromone components of many other species, including some of this study (Figure 1), either do not require the Δ9 UFA precursors produced by these desaturases, *e.g.*, in *P. separata* (Kou *et al.* 1992), or are produced by Δ11 desaturation followed by chain shortening, *e.g.*, in *S. exigua* (Jurenka 1997). Nevertheless, we isolated KPSE cDNAs from pheromone glands of all of the species used in this study, which suggests either post-transcriptional downregulation in the pheromone glands of some species, as has been postulated for the Δ9 desaturase-encoding transcripts in the pheromone gland of *H. zea* (Rosenfield *et al.* 2001), or very low transcript levels. Thus, the primary biological role of the KPSE lineage in the Lepidoptera is ambiguous at this time. It is conceivable that the recruitment of KPSEs for the production of UFA sex pheromone precursors occurred relatively recently in an ancestor of contemporary heliothine moths and that the primary biological role of KPSE desaturases in most moths is more similar to the ancestral function of the NPVE desaturases.

Two *D. melanogaster* orthologs of the lepidopteran KPSE lineage encode Δ9 desaturases that function as pheromone biosynthetic enzymes: In contrast to the situation in moths where females release volatile unsaturated derivatives of fatty acids from sex pheromone glands, flies use unsaturated cuticular hydrocarbons as contact pheromones that specify sex and species recognition (Jallon 1984; Coyne and Oyama 1995). Characterized pheromone biosynthetic pathways of Diptera use fatty acid synthesis, desaturation, chain elongation, and reductive decarboxylation (Blomquist *et al.* 1987; Pennanec'h *et al.* 1991) and thus share the early steps with lepidopteran pheromone biosynthetic pathways (reviewed in Tillman *et al.* 1999). Two *D. melanogaster* desaturases implicated in pheromone biosynthesis, desat 1 and desat 2 (Dallérac *et al.* 2000), are of considerable interest in the context of the divergent evolution of the mate recognition systems of flies and moths.

The *desat 1* and *desat 2* genes map <4 kb apart in the same cytogenetic interval containing the genetic locus of a female-specific cuticular hydrocarbon polymorphism (Wicker-Thomas *et al.* 1997; Coyne *et al.* 1999; Dallérac *et al.* 2000). The desat 1 desaturase catalyzes the formation of Δ9 UFA precursors and has a substrate preference like that of the *Hzea* KPSE enzyme (*i.e.*, 16:0 > 18:0). However, the *D. melanogaster* desat 1 is expressed in both males and females, where its major UFA product, palmitoleic acid (*Z*9-16:1), is a precursor in the syntheses of sex-specific cuticular hydrocarbon contact pheromone components (Pennanec'h *et al.* 1997). The desat 2 desaturase has specificity for the Δ9 desaturation of myristic acid (14:0; Dallérac *et al.* 2000). In contrast to *desat 1*, the *desat 2* gene is expressed only in females and its absence of expression in some geographically derived strains due to a defect in its promotor region provides the genetic basis for the cuticular diene hydrocarbon polymorphism of the major component of the female contact pheromone (Coyne *et al.* 1999; Dallérac *et al.* 2000; Takahashi *et al.* 2001). It is, thus, apparent that members of the KPSE lineage are used by both flies and moths to synthesize UFA precursors of their respective sex pheromones, suggesting the functional commitment of this lineage to this role prior to the divergence of flies and moths from a common ancestor. In addition, the expression of *desat 1* in both sexes, as well as in sexually immature developmental stages, suggests that the encoded desaturase could possibly have an additional functional role in flies in cold adaptation. Whether this is borne out by further analysis

or not, it is clear that flies and moths both require UFAs for cold adaptation and mate recognition and have evolved to use different members of the desaturase gene family in different ways to meet these needs.

The xxxQ lineage encodes pheromone biosynthetic desaturases with diverse enzymatic properties: The fact that there is no *D. melanogaster* ortholog of the xxxQ lineage indicates that it, like the NPVE lineage encoding Δ9 desaturases, either was established as a lepidopteran innovation or existed in a common ancestor of flies and moths and was subsequently lost (or substantially diverged) in the course of dipteran evolution leading to *D. melanogaster*. Evidence for a primary biological role of the xxxQ lineage in sex pheromone biosynthesis is provided by the differential and abundant expression of xxxQ transcripts in pheromone glands of three noctuid species and the simple monoene Δ11 UFA products formed by their encoded desaturases, which are pheromone precursors in these species (Knipple *et al.* 1998; Rosenfield *et al.* 2001; S. E. Jeong, K. M. You, C.-L. Rosenfield and D. C. Knipple, unpublished data). These findings and the relatively high sequence conservation within the modern branch of the xxxQ lineage are consistent with the conclusion drawn from early biochemical studies of lepidopteran pheromones that simple monoene Δ11 UFAs are the most prevalent precursors used in the formation of major sex pheromone components in the modern Lepidoptera (Roelofs and Bjostad 1984; Roelofs and Wolf 1988), as exemplified by the pathways for the sex pheromone components of the noctuids *H. assulta*, *M. brassica*, and *P. separata* (Figure 1).

The genus Spodoptera is unusual in the context of the family Noctuidae insofar as only 2 of its 16 species with characterized sex pheromones have major components deriving from simple monoene Δ11 UFAs, whereas the other 14 use diene UFA precursors, which are more characteristic of the ancestral lepidopteran families (Arn *et al.* 1992). Two species in our study from the genus Spodoptera, *i.e.*, *S. litura* and *S. exigua*, are in the latter category (Tamaki *et al.* 1973; Tumlinson *et al.* 1990). The routes to their diene UFA precursors require a *Z*9-14:1 intermediate, which is subjected to a second desaturation reaction initiating at either the eleventh or the twelfth carbon atoms (Roelofs and Bjostad 1984; Jurenka 1997; Figure 1). We suggest that these pathways reflect a reversion to the ancestral state at an early point in the establishment of this genus, perhaps by recruitment of a recently inactivated gene encoding an xxxQ monoene desaturase with a Δ12 regioselectivity.

The low conservation among sequences of the more ancestral subgroup of the xxxQ lineage (Table 2) is consistent with the diverse enzymatic properties identified among the small number of its functionally characterized members (Figure 3). It is interesting that the d_N/d_S ratio of the xxxQ lineage, although low, was the highest of the six lineages (Table 1), suggesting that there might have been a relaxation of constraints, or an increase in the amount of positive selection, in this group. Furthermore, the only signature of positive selection detected by the codon-based likelihood analysis was found in a branch of this lineage leading to the sequences *Pint* VGTQ and *Pint* AGTQ. Of the seven amino acid sites identified as being under positive selection, six occur in regions of the desaturase protein that are also implicated in the adaptation to new function by the computational methods that we used to analyze the protein sequence data sets, discussed in more detail below. We suggest that the lack of evidence for positive selection in other subgroups of this lineage could be caused by the very high levels of sequence divergence, which can reduce the power in the statistical analysis. We anticipate that other unique enzymatic properties predicted from the analysis of pheromone biosynthetic pathways (*e.g.*, as depicted in Figure 1) will be revealed by the functional characterization of additional members of the xxxQ lineage from the more ancestral species, which this study suggests have a greater number of duplicated genes (*e.g.*, four sequences in *P. interpunctella* and three in *C. pomonella*).

Sequence and structural correlates of biochemically distinct classes of lepidopteran desaturases: We identified several correlations of structure and biochemical function in comparisons of the sequence groups containing functionally characterized lepidopteran desaturases, specifically, in the NPVEs encoding Δ9 desaturases with substrate chain length preferences 18:0 > 16:0, in the KPSEs encoding Δ9 desaturases with substrate chain length preferences 16:0 > 18:0, and in the modern xxxQs encoding Δ11 desaturases that use saturated fatty acid substrates. Sequences or structural features implicated in desaturase regioselectivity were identified by three separate methods of analysis: by the evolutionary trace method (Lichtarge *et al.* 1996), *i.e.*, sequences in domains a, c, and d; by hydropathy analysis (Kyte and Doolittle 1982), *i.e.*, size and location of hydrophobic domains β, δ, and ε relative to conserved elements of the aligned sequences; and by using transmembrane prediction algorithms (see materials and methods), *i.e.*, the location of the predicted transmembrane domain TM4 relative to conserved elements of the aligned sequences (summarized in Figure 5). An alignment gap in TM4 in all of the sequences of the xxxQ lineage (Figure 2) and functional class-specifying patterns of tryptophan polymorphisms between TM3 and TM4 (Figures 2 and 5) provide additional evidence for the involvement of mutational changes in and around the base of TM4 in the evolution of novel regioselectivities in lepidopteran pheromone desaturases. Furthermore, this region is implicated in functional adaptation by the codon-based likelihood test (Yang and Nielsen 2002), which identified two sites that may be under positive selection in the transmembrane helices of TM3 and

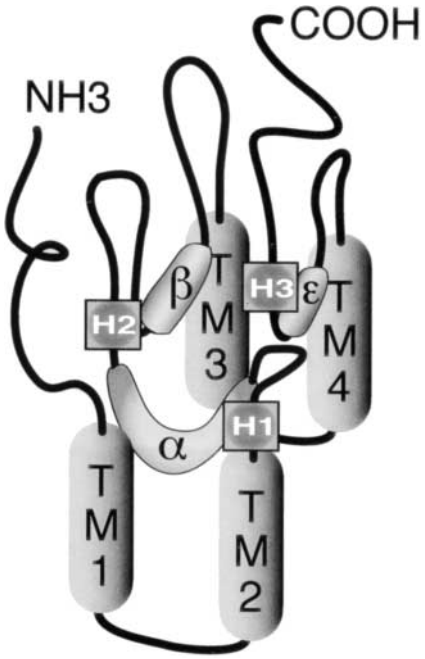


FIGURE 6.—Proposed model of integral membrane fatty acid desaturases showing a possible arrangement of transmembrane domains TM1–TM4; hydrophobic domains α, β, and ε; and hydrophilic domains (thick lines). In this proposed model, desaturase regioselectivity is determined by variable class-specific transmembrane domains and hydrophobic elements that shift the position of the catalytic site, consisting of spatially coordinated iron-binding histidine boxes (H1–H3), relative to the hydrophobic substrate-binding pocket on the cytosolic face of the enzyme.

TM4 immediately flanking the short hydrophilic interval separating them, *i.e.*, positions 117 and 131, in a subgroup of the xxxQ lineage. Four other positive sites identified in the same test, *i.e.*, positions 160, 164, 166, and 173, are clustered in the interval containing the signature motif between TM4 and H3. Taken together, these results lead us to postulate that the discontinuous hydrophobic domains of integral membrane desaturases play a crucial role in the positioning of the two histidine boxes, H2 and H3 (Figure 6), which coordinate iron atoms at the catalytic site of the enzyme, and that these elements are shifted toward the hydrocarbon end of the fatty acid substrate in the modern xxxQs, resulting in desaturation initiating at the Δ11 carbon atom rather than at the Δ9 position.

Evolutionary trace and hydropathy analyses identified three correlations between structure and substrate chain length preferences. The first of these includes sequences defined by the overlapping ET domain c and the hydrophobic domain γ, which correspond to TM3. Associated with these differences are the occurrence of an alignment gap in the NPVE class at position 120, functional class-specifying tryptophan polymorphisms between TM3 and TM4, and associated changes in the relative positions of the latter transmembrane domains relative to conserved elements of the aligned sequences. A second correlation between structure and substrate chain length preference occurs in ET domain e, which includes the signature motif. This domain has the highest class-specifying positional score of any category (Figure 4). A third correlation between structure and substrate chain length preferences is a significantly larger α hydrophobic domain in the NPVEs, which results in a shortened hydrophilic region between α and H2. On the basis of these results and the model of integral membrane desaturase structure represented in Figure 6, we hypothesize that several discontinuous sequence domains contribute to the formation of a hydrophobic substrate binding pocket and that, in the course of the evolution of pheromone desaturases in the Lepidoptera, selection for shorter chain length compounds with higher volatility resulted in the preservation of mutational changes in one or more of these domains that reduced the effective size of the substrate binding pocket.

The above evolutionary inferences are informed by only a small number of functional expression studies performed to date, but nevertheless are largely consistent with the body of knowledge of lepidopteran pheromone biosynthetic pathways that has accumulated over the last three decades (reviewed in TILLMAN *et al.* 1999). Clearly, additional experiments to establish biochemical identities of encoded proteins and to quantify tissue-specific transcript levels will be necessary to test some of the hypotheses that we present here. Furthermore, the application of computational approaches to the growing sequence data set of lepidopteran desaturase-encoding cDNAs should provide additional insights into the evolution of this enzyme family as well as useful predictions for the design of experiments to identify specific correlates of structure and biochemical function.

We express our gratitude to Peter Landolt [U.S. Department of Agriculture (USDA) Agricultural Research Service, Wapato, WA], Robert T. Staten (USDA APHIS PPQP, Phoenix, AZ), and Myung-Hee Ohh (Korea Ginseng and Tobacco Research Institute, Taejon, Korea) for providing insects used in this study. We also thank Kwai Weng Wong and Patricia Marsella-Herrick of the D.C.K. lab for technical assistance, Wendell Roelofs for communicating data on the *E. postvittana* desaturases prior to their publication, and Jarek Meller for helpful discussions and advice on structural analysis of integral membrane proteins. This study was supported by USDA grants 97-35302-4345 and 2001-35302-09926 and Environmental Protection Agency/National Science Foundation (NSF) grant BES-9728367 to D.C.K. and NSF grant DEB-0089487 and HFSP grant RGY0055/2001-M to R.N.

LITERATURE CITED

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHENG *et al.*, 1997 Gapped Blast and Psi-Blast: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

ARN, H., P. M. GUERIN, H. R. BUSER, S. RAUSCHER and E. MANI, 1985 Sex pheromone blend of the codling moth, *Cydia pomo-*

nella: evidence for a behaviour role of dodecan-1-ol. Experientia **41:** 1482–1484.

Arn, H., M. Tóth and E. Priessner, 1992 *List of Sex Pheromones of Lepidoptera and Related Attractants*, Ed. 2. International Organization for Biological Control, Montfavet, France.

Attygalle, A. B., M. Herrig, O. Vostrowsky and H. J. Bestmann, 1987 Technique for injecting intact glands for analysis of sex pheromones of Lepidoptera by capillary gas chromatography. J. Chem. Ecol. **13:** 1299–1311.

Bjostad, L. B., and W. L. Roelofs, 1983 Sex pheromone biosynthesis in *Trichoplusia ni*: key steps involve Δ11 desaturation and chain shortening. Science **220:** 1387–1389.

Bjostad, L. B., W. A. Wolf and W. L. Roelofs, 1987 Pheromone biosynthesis in lepidopterans: desaturation and chain shortening, pp. 77–120 in *Pheromone Biochemistry*, edited by J. D. Prestwich and G. J. Blomquist. Academic Press, London.

Blomquist, G. J., J. W. Dillwith and T. S. Adams, 1987 Biosynthesis and endocrine regulation of sex pheromone production in Diptera, pp. 217–250 in *Pheromone Biochemistry*, edited by J. D. Prestwich and G. J. Blomquist. Academic Press, London.

Bordo, D., and P. Argos, 1991 Suggestions for "safe" residue substitutions in site-directed mutagenesis. J. Mol. Biol. **217:** 721–729.

Briggs, D. E. G., M. J. Weedon and M. A. White, 1993 Arthropoda (Crustacea excluding Ostracoda), pp. 321–342 in *The Fossil Record 2*, edited by M. J. Benton. Chapman & Hall, London.

Cork, A., K. S. Boo, E. Dunkelblum, D. R. Hall, K. Jee-Rajunga *et al.*, 1992 Female sex pheromone of oriental tobacco budworm, *Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae): identification and field testing. J. Chem. Ecol. **18:** 403.

Coyne, J. A., and R. Oyama, 1995 Localisation of pheromonal sexual dimorphism in *Drosophila melanogaster* and its effect on sexual isolation. Proc. Natl. Acad. Sci. USA **92:** 9505–9509.

Coyne, J. A., C. Wicker-Thomas and J.-M. Jallon, 1999 A gene responsible for a cuticular hydrocarbon polymorphism in *Drosophila melanogaster*. Genet. Res. **73:** 189–203.

Dallérac, R., C. Labeur, J.-M. Jallon, D. C. Knipple, W. L. Roelofs *et al.*, 2000 A Δ9-desaturase gene with a different substrate specificity is responsible for the cuticular diene hydrocarbon polymorphism in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **97:** 9449–9454.

Felsenstein, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

FlyBase, 1999 The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Res. **27:** 85–88.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-L. Yan *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:** 1531–1545.

Foster, S. P., and W. L. Roelofs, 1990 Biosynthesis of a monoene and a conjugated diene sex pheromone component of the lightbrown apple moth by Δ11 desaturation. Experientia **46:** 269–273.

Hazel, J. R., and E. E. Williams, 1990 The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. Prog. Lipid Res. **29:** 167–227.

Higgins, D. G., J. D. Thompson and T. J. Gibson, 1996 Using CLUSTAL for multiple sequence alignments, pp. 383–401 in *Methods in Enzymology*, edited by R. F. Doolittle. Academic Press, San Diego.

Hirokawa, T. S., S. Boon-Chieng and S. Mitaku, 1998 SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics **14:** 378–379.

Hofmann, K., and W. Stoffel, 1993 TMbase—a database of membrane spanning protein segments. Biol. Chem. Hoppe-Seyler **374:** 166.

Hummel, H. E., L. K. Gaston, H. H. Shorey, R. S. Kaae, K. J. Byrne *et al.*, 1973 Clarification of the chemical status of the pink bollworm pheromone. Science **181:** 873–875.

Jallon, J. M., 1984 A few chemical words exchanged by *Drosophila* during courtship and mating. Behav. Genet. **14:** 441–478.

Jeong, S. E., Y. Lee, J. H. Hwang and D. C. Knipple, 2001 Effects of the sap of the common oleander *Nerium indicum* (Apocyanaceae) on male fertility and spermatogenesis in the oriental tobacco budworm *Helicoverpa assulta* (Lepidoptera, Noctuidae). J. Exp. Biol. **204:** 3935–3942.

Jones, D. T., W. R. Taylor and J. M. Thornton, 1994 A model recognition approach to the prediction of all-helical membrane protein structure and topology. Biochemistry **33:** 3038–3049.

Jurenka, R. A., 1997 Biosynthetic pathway for producing the sex pheromone component (*Z*,*E*)-9,12-tetradecadienyl acetate in moths involves a Δ12 desaturase. Cell. Mol. Life Sci. **53:** 501–505.

Jurenka, R. A., E. Jacquin and W. L. Roelofs, 1991 Control of the pheromone biosynthetic pathway in *Helicoverpa zea* by the pheromone biosynthetic activating neuropeptide. Arch. Insect Biochem. Physiol. **17:** 81–91.

Klun, J. A., J. R. Plimmer, B. A. Bierl-Leonhardt, A. N. Sparks, M. Primiani *et al.*, 1980 Sex pheromone chemistry of female corn earworm moth, *Heliothis zea*. J. Chem. Ecol. **6:** 165–175.

Knipple, D. C., C.-L. Rosenfield, S. J. Miller, W. Liu, J. Tang *et al.*, 1998 Cloning and functional expression of a cDNA encoding a pheromone gland-specific acyl-CoA Δ11-desaturase of the cabbage looper moth, *Trichoplusia ni*. Proc. Natl. Acad. Sci. USA **95:** 15287–15292.

Kou, R., Y.-S. Chow and H.-Y. Ho, 1992 Chemical composition of sex pheromone gland extract in female oriental armyworm *Pseudaletia separata* Walker (Lepidoptera: Noctuidae) in Taiwan. Bull. Inst. Zool. Acad. Sin. **31:** 246–250.

Kyte, J., and R. F. Doolittle, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 105–132.

Lichtarge, O., H. R. Bourne and F. E. Cohen, 1996 An evolutionary trace method defines binding surfaces common to protein families. J. Mol. Biol. **257:** 342–358.

Liu, W., P. W. K. Ma, P. Marsella-Herrick, C.-L. Rosenfield, D. C. Knipple *et al.*, 1999 Cloning and functional expression of a cDNA encoding a metabolic acyl-CoA Δ9-desaturase of the cabbage looper moth, *Trichoplusia ni*. Insect Biochem. Mol. Biol. **29:** 435–443.

Liu, W., H. Jiao, N. C. Murray, M. O'Conner and W. L. Roelofs, 2002 Gene characterized for membrane desaturase that produces (*E*)-11 isomers of mono- and diunsaturated fatty acids. Proc. Natl. Acad. Sci. USA **99:** 620–624.

Lynch, M., and A. G. Force, 2000 The origin of interspecific genomic incompatibility via gene duplication. Am. Nat. **156:** 590–605.

Mitaku, S., and T. Hirokawa, 1999 Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. Protein Eng. **12:** 953–957.

Mitaku, S., M. Ono, T. Hirokawa, S. Boon-Chieng and M. Sonoyama, 1999 Proportion of membrane proteins in proteomes of 15 single-cell organisms analyzed by the SOSUI prediction system. Biophys. Chem. **82:** 165–171.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.

Park, K. C., A. Cork and K. S. Boo, 1996 Intrapopulational changes in sex pheromone composition during scotophase in oriental tobacco budworm, *Helicoverpa assulta* (Guenée) (Lepidoptera: Noctuidae). J. Chem. Ecol. **22:** 1201–1210.

Pennanec'h, J., F. Feveur, D. B. Pho and J.-M. Jallon, 1991 Insect fatty acid related pheromones. A review of their biosynthesis, hormonal regulation and genetic control. Ann. Soc. Entomol. Fr. **27:** 245–263.

Pennanec'h, M., L. Bricard, G. Kunesh and J. M. Jallon, 1997 Incorporation of fatty acids into cuticular hydrocarbons of male and female *Drosophila melanogaster*. J. Insect Physiol. **43:** 1111–1116.

Pope, M. M., L. K. Gaston and T. C. Baker, 1984 Composition, quantification, and periodicity of sex pheromone volatiles from individual Heliothis zea females. J. Insect Physiol. **30:** 943–945.

Rodriguez, F., D. L. Hallahan, J. A. Pickett and F. Camps, 1992 Characterization of the Δ11-palmitoyl-CoA desaturase from *Spodoptera littoralis* (Lepidoptera:Noctuidae). Insect Biochem. Mol. Biol. **22:** 143–148.

Roelofs, W. L., and L. Bjostad, 1984 Biosynthesis of Lepidopteran pheromones. Bioorg. Chem. **12:** 279–298.

Roelofs, W. L., and W. A. Wolf, 1988 Pheromone biosynthesis in Lepidoptera. J. Chem. Ecol. **14:** 2019–2031.

Rosenfield, C.-L., K. M. You, P. Marsella-Herrick, W. L. Roelofs and D. C. Knipple, 2001 Structural and functional conservation and divergence among acyl-CoA desaturase-encoding genes of two Noctuid species, the corn earworm, *Helicoverpa zea*, and the

cabbage looper, *Trichoplusia ni*. Insect Biochem. Mol. Biol. **31:** 949–964.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sanger, F., S. Nicklen and A. R. Coulsen, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

Shanklin, J., and E. B. Cahoon, 1998 Desaturation and related modifications of fatty acids. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49:** 611–641.

Silhacek, D. L., and G. L. Miller, 1972 Growth and development of the Indian meal moth, *Plodia interpunctella* (Lepidoptera, Phycitidae) under laboratory mass-rearing conditions. Ann. Ent. Soc. Am. **65:** 1084–1087.

Sonnhammer, E. L. L., G. von Heijne and A. Krogh, 1998 A hidden Markov model for predicting transmembrane helices in protein sequences, pp. 175–182 in *Proceedings of Sixth International Conference on Intelligent Systems for Molecular Biology*, edited by J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff *et al.* AAAI Press, Menlo Park, CA.

Takahashi, A., S.-C. Tsaur, J. A. Coyne and C.-I Wu, 2001 The nucleotide changes governing cuticular hydrocarbon variation and their evolution in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **98:** 3920–3925.

Tamaki, Y., H. Noguchi and T. Yushima, 1973 Sex pheromone of *Spodoptera litura* (F.) (Lepidoptera:Noctuidae): isolation, identification, and synthesis. Appl. Entomol. Zool. **8:** 200–203.

Teal, P. E. A., and J. H. Tumlinson, 1986 Terminal steps in pheromone biosynthesis by *Heliothis virescens* and *Heliothis zea*. J. Chem. Ecol. **12:** 353–366.

Teal, P. E. A., R. R. Heath, B. D. Dueben, J. A. Coffelt and K. W. Vick, 1995 Production and release of (*Z,E*)-9,12-tetradecadienal by sex pheromone glands of females of *Plodia interpunctella* (Lepidoptera:Pyralidae). J. Chem. Ecol. **21:** 787–799.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Tiku, P. E., A. Y. Gracey, A. I. Macartney, R. J. Beyton and A. R. Crossins, 1996 Cold-induced expression of Δ9-desaturase in carp by transcriptional and posttranslational mechanisms. Science **271:** 815–818.

Tillman, J. A., S. J. Seybold, R. A. Jurenka and G. J. Blomquist, 1999 Insect pheromones—an overview of biosynthesis and endocrine regulation. Insect Biochem. Mol. Biol. **29:** 481–514.

Tumlinson, J. H., E. R. Mitchell and H. S. Yu, 1990 Analysis and field evaluation of volatile blend emitted by calling virgin females of beet armyworm moth, *Spodoptera exigua* (Huebner). J. Chem. Ecol. **16:** 3411–3423.

Tusnády, G. E., and I. Simon, 1998 Principles governing amino acid composition of integral membrane proteins: applications to topology prediction. J. Mol. Biol. **283:** 489–506.

Tusnády, G. E., and I. Simon, 2001 The HMMTOP transmembrane topology prediction server. Bioinformatics **17:** 849–850.

Vigh, L., D. A. Los, I. Hovath and N. Murata, 1993 The primary signal in the biological perception of temperature: Pd-catalyzed hydrogenation of membrane lipids stimulated the expression of the *des*A gene in *Synechocystis* PCC6803. Proc. Natl. Acad. Sci. USA **90:** 9090–9094.

Wicker-Thomas, C., C. Henriet and R. Dallérac, 1997 Partial characterization of a fatty acid desaturase gene in *Drosophila melanogaster*. Insect Biochem. Mol. Biol. **27:** 963–972.

Wolf, W. A., and W. L. Roelofs, 1986 Properties of the Δ11-desaturase enzyme used in cabbage looper moth sex pheromone biosynthesis. Arch. Insect Biochem. Physiol. **3:** 45–52.

Wolf, W. A., and W. L. Roelofs, 1989 Enzymes involved in the biosynthesis of sex pheromones in moths, pp. 323–331 in *Biocatalysis in Agricultural Biotechnology*, edited by J. R. Whitaker and P. E. Sonnet. American Chemical Society, Washington, DC.

Yang, Z., 2000 *Phylogenetic Analysis by Maximum Likelihood (PAML)*, Version 3.0. University College, London.

Yang, Z., and R. Nielsen, 1998 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46:** 409–418.

Yang, Z., and R. Nielsen, 2002 Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. **19:** 908–917.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Yoshiga, T., K. Okano, K. Mita, T. Shimada and S. Matsomoto, 2000 cDNA cloning of acyl-CoA desaturase homologs in the silkworm, *Bombyx mori*. Gene **246:** 339–345.

Communicating editor: S. W. Schaeffer