Supporting Information

Lassance et al. 10.1073/pnas.1208706110

SI Materials and Methods

Phylogenetic Analysis. We amplified the five loci by PCR using pheromone gland cDNAs as template. The amplified regions were as follows: *elongation factor-1α* (*EF-1α*), *malate dehydrogenase* (*MDH*), *isocitrate dehydrogenase* (*IDH*), and *ribosomal protein S5* (*RpS5*) from the nuclear genome and *cytochrome oxidase subunit I* (*COI*) from the mitochondrial genome. Primers were taken from Wahlberg and Wheat (1), and the PCR amplifications were performed using the following touchdown profile: 2 min at 94 °C, 15 cycles of 30 s at 94 °C, 30 s of primer annealing at an initial temperature of 55 °C dropped 0.3 °C per cycle, and 1 min at 72 °C, followed by 25 cycles of 30 s at 94 °C, 30 s at 55 °C, and 1 min at 72 °C. PCR products were treated with a mixture containing exonuclease I (EXO; Fermentas) and shrimp alkaline phosphatase (SAP; Fermentas) and sequenced in both directions using the T7 and T3 universal primers with the BigDye terminator kit v1.1 (Applied Biosystems), followed by analysis on an ABI PRISM 3130xl genetic analyzer (Applied Biosystems). Data were analyzed using the BioEdit sequence analysis software version 7 (2). Any heterozygous position was coded following the International Union of Pure and Applied Chemistry (IUPAC) ambiguity code. The combined nucleotide sequence dataset comprised 4,804 bp that were partitioned by gene. Phylogenetic reconstruction was inferred using MrBayes v3.2 (3). The substitution model (GTR+I +Γ) was selected following model testing using MultiPhyl (4). The Bayesian analysis was performed for 1 million generations, with parameter values estimated separately for each gene. Every 100th generation was sampled, and the first 1,000 were discarded as burn-in. We selected *Ostrinia latipennis* as an outgroup.

For the pheromone biosynthetic fatty-acyl reductase gene *pgFAR* gene phylogeny reconstruction, the sequence data used corresponded to the coding region of the gene only. We excluded the UTRs because no reliable alignment of those could be obtained. In particular, the 3′ UTRs are extensively diverged from one species to the other. The phylogenetic tree was inferred with MrBayes implementing the GTR+Γ model with the nucleotide sequences and the JTT+I+Γ model with the amino acid sequences. The tree topologies were the same irrespective of the dataset used. Again, *O. latipennis* was enforced as an outgroup.

Functional assays. Functional assays were carried out in a yeast expression system following the procedures described previously (5, 6). Briefly, the open reading frames of the various *pgFAR* genes were cloned into the expression vector pYES2.1/V5-His TOPO (Invitrogen) and the resulting recombinant vectors were used to transform yeast, Saccharomyces cerevisae (INVSc1 strain, Invitrogen). When testing the conversion of individual precursors, aliquots of yeast cultures were suspended in induction medium containing 0.5 mM of the FAME precursor diluted in ethanol. After 24 h of incubation at 30°C, alcohol products were extracted in n-hexane spiked with 150 ng of (Z)-11-tridecenol (Z11-13:OH) used as internal standard. Samples were analyzed by gas chromatography-mass spectrometry under the conditions described earlier (5). For each substrate, the alcohol production was calculated using the internal standard method. For each pgFAR, the relative abundance of each fatty alcohol was derived using the formula:

$$\text{ratio of alcohol } y \text{ in vitro} = \frac{\text{amount of } y \text{ (in ng)}}{\sum \text{amount of all alcohols (in ng)}}.$$

The predicted pheromone ratio was calculated using the following formula: predicted ratio of y in pheromone blend = (ratio of precursor of alcohol y in the gland) × (ratio of alcohol y in vitro) (results were scaled so that the sum of predicted ratios would equal 1).

When testing the alcohol production of yeast supplemented with precursor blends in the ratio as found in female pheromone glands, the total concentration of precursors was 0.5 mM and the culture medium was 4 ml, whereas all other parameters were kept unchanged. After concentration under a gentle stream of pure nitrogen, yeast extracts were analyzed on a Hewlett Packard HP 6890 gas chromatograph (Agilent Technologies) coupled to a mass-selective detector HP 5973 operated in electron impact mode (70 eV). The GC was equipped with a HP-INNOWax capillary column (30 m × 0.25 mm × 0.15 μm; Agilent Technologies). Helium was used as carrier gas (velocity 30 cm s^{-1}) and the injection port was maintained at 230°C. One μl of sample was injected in spitless mode using a HP 7683 automatic liquid sampler (Agilent Technologies). In order to obtain good separation of the different isomers, the oven temperature program was set as follows: 2 min at 100°C, then the temperature was raised to 150°C at a rate of 5°C min^{-1} then to 200°C at a rate of 2°C min^{-1}. A post-run of 10 min at 230°C was used to clean the column between runs.

1. Wahlberg N, Wheat CW (2008) Genomic outposts serve the phylogenomic pioneers: Designing novel nuclear markers for genomic DNA extractions of lepidoptera. *Syst Biol* 57(2):231–242.
2. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98.
3. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
4. Keane TM, Naughton TJ, McInerney JO (2007) MultiPhyl: A high-throughput phylogenomics webserver using distributed computing. *Nucleic Acids Res* 35(Web Server issue):W33-7.
5. Lassance J-M, et al. (2010) Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* 466(7305):486–489.
6. Liénard MA, Hagström ÅK, Lassance J-M, Löfstedt C (2010) Evolution of multicomponent pheromone signals in small ermine moths involves a single fatty-acyl reductase gene. *Proc Natl Acad Sci USA* 107(24):10955–10960.

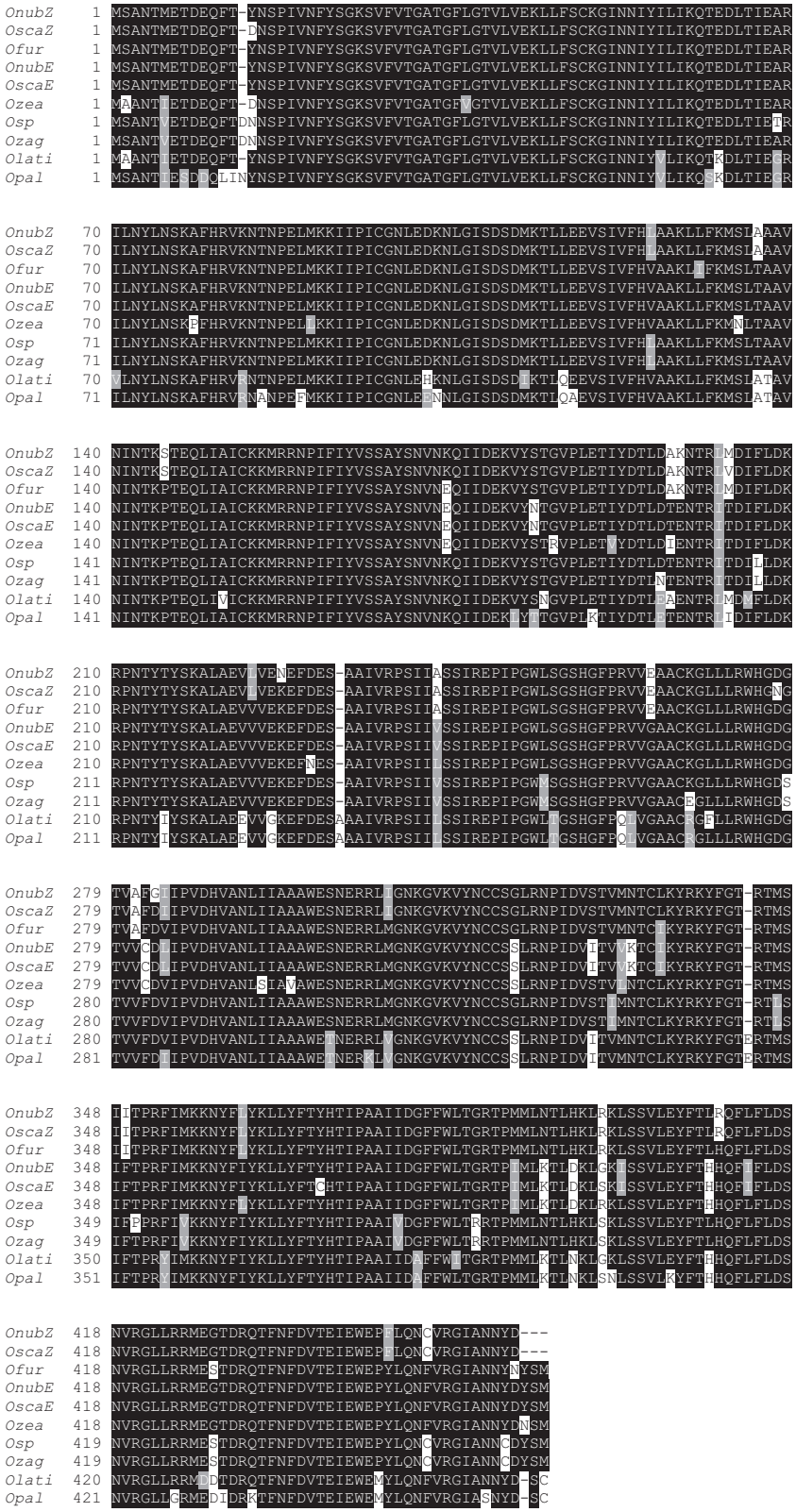

Fig. S1. Alignment of the *Ostrinia* pheromone biosynthetic reductase *pgFAR*. Identical amino acid residues and conservative substitutions are shaded in black and gray, respectively.

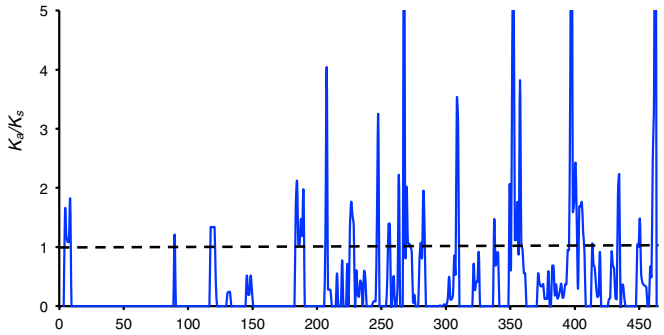

Fig. S2. Ratio of K_a/K_s calculated throughout the coding region of *pgFAR* estimated over a sliding window (window size: five codons; step size: one codon) using Li's synonymous/nonsynonymous method implemented in SWAAP 1.0.2 (1). Peaks above the dashed line indicate an excess of nonsynonymous substitutions over the neutral expectations ($K_a/K_s > 1$). Values are presented with an arbitrary cutoff of 5.

1. Li W-H (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J Mol Evol* 36(1):96–99.

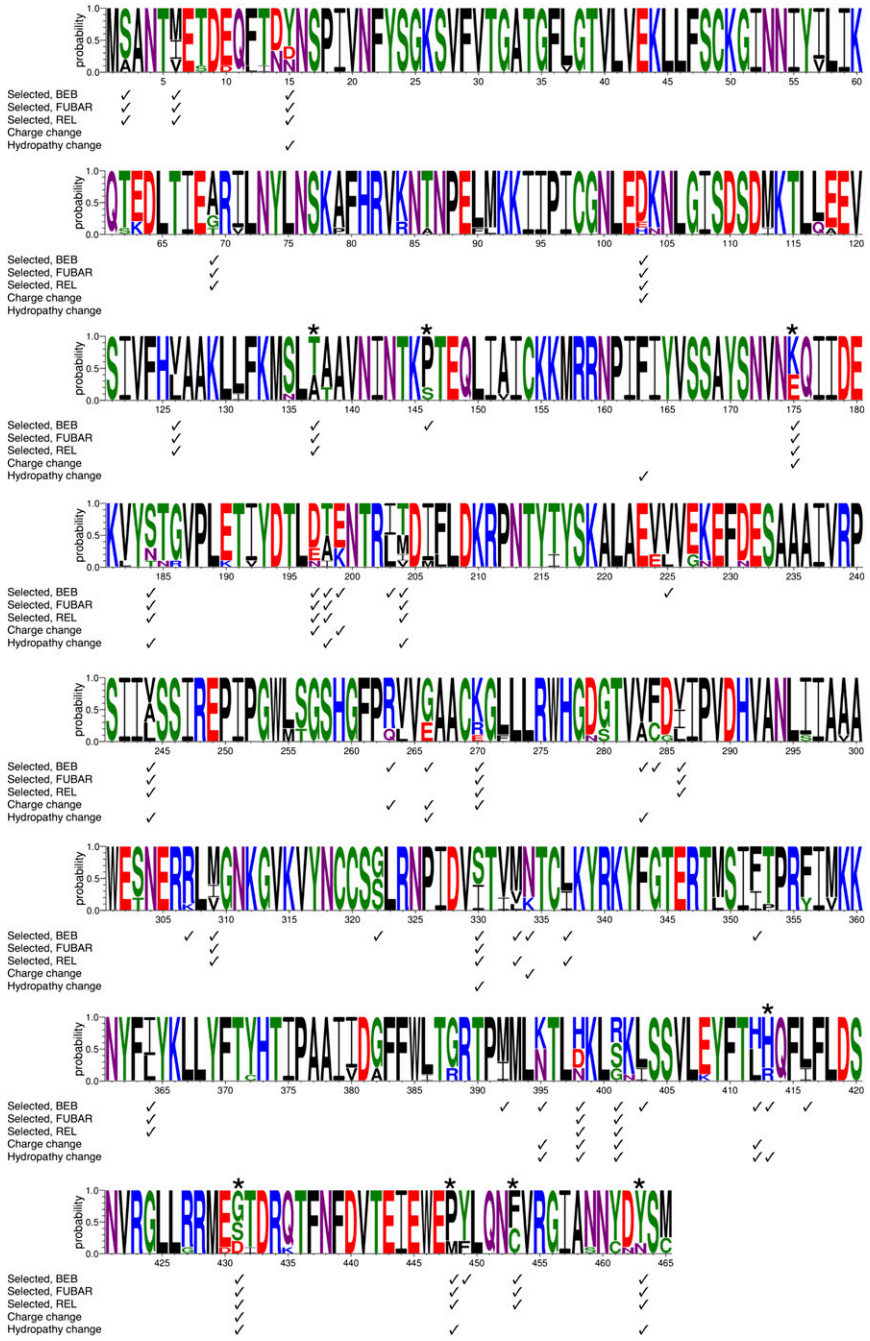

Fig. S3. Sequence logo of the aligned *Ostrinia pgFAR* orthologs. The logo displays the frequencies of amino acids at each position, with the height of a letter reflecting the raw residue frequencies. Amino acids are colored according to chemical properties (green: polar residues (G, S, T, Y, C); purple: neutral residues (Q, N); blue: basic residues (K, R, H); red: acidic residues (D, E); black: hydrophobic residues (A, V, L, I, P, W, F, M)). Positively selected sites detected using the Bayes empirical Bayes (BEB) method, the random effects likelihood method (REL), and/or the fast unbiased approximate Bayesian analysis (FUBAR) are indicated. Transitions in charge and/or hydrophobicity are indicated for the positively selected sites. Sites subjected to site-directed mutagenesis in this study are marked with asterisks. The logo was created using WebLogo 3.1 (1).

1. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14(6):1188–1190.

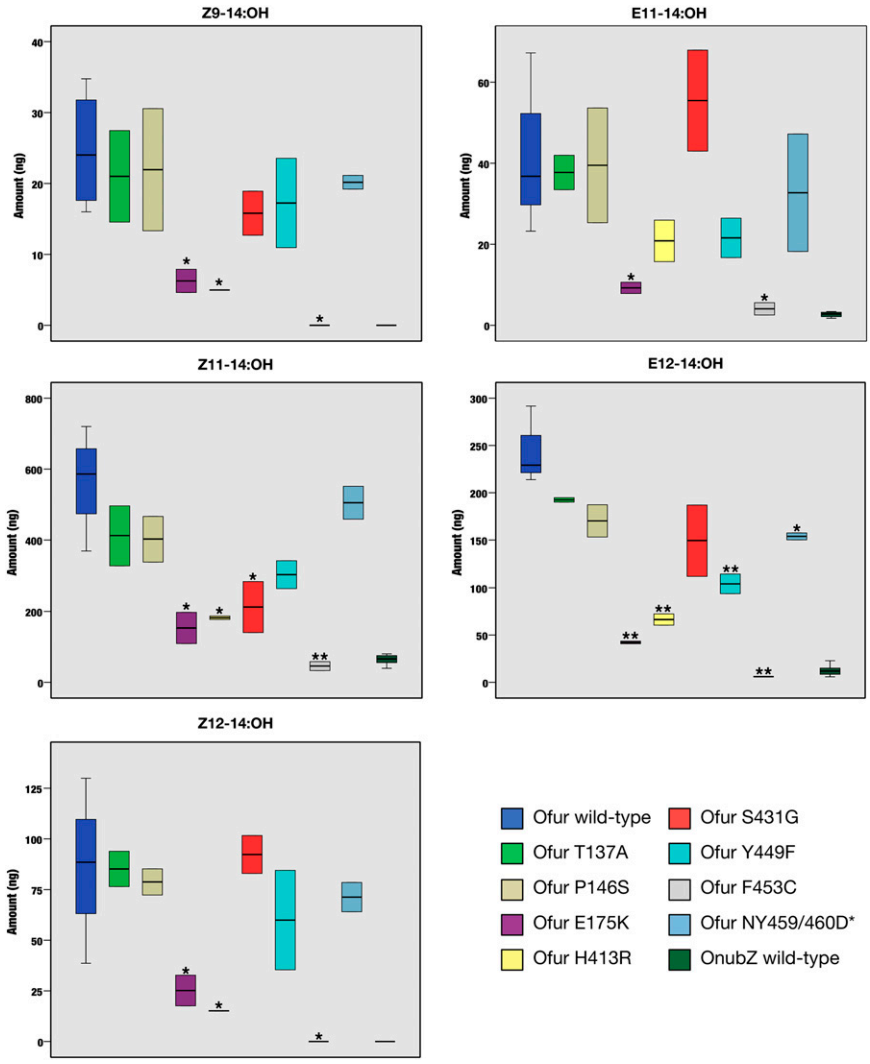

Fig. S4. Fatty alcohols produced by pgFAR WT and mutants expressed in *Saccharomyces cerevisae* supplemented with fatty-acyl precursors. Student *t* tests were used to compare the mean production of mutants with the mean production of the parental *Ofurnacalis* WT pgFAR ($n = 2–4$). The mean production of *OnubilalisZ* WT pgFAR is given for illustrative purposes. One or two asterisks denote significant difference at the 0.05 and 0.01 levels, respectively.

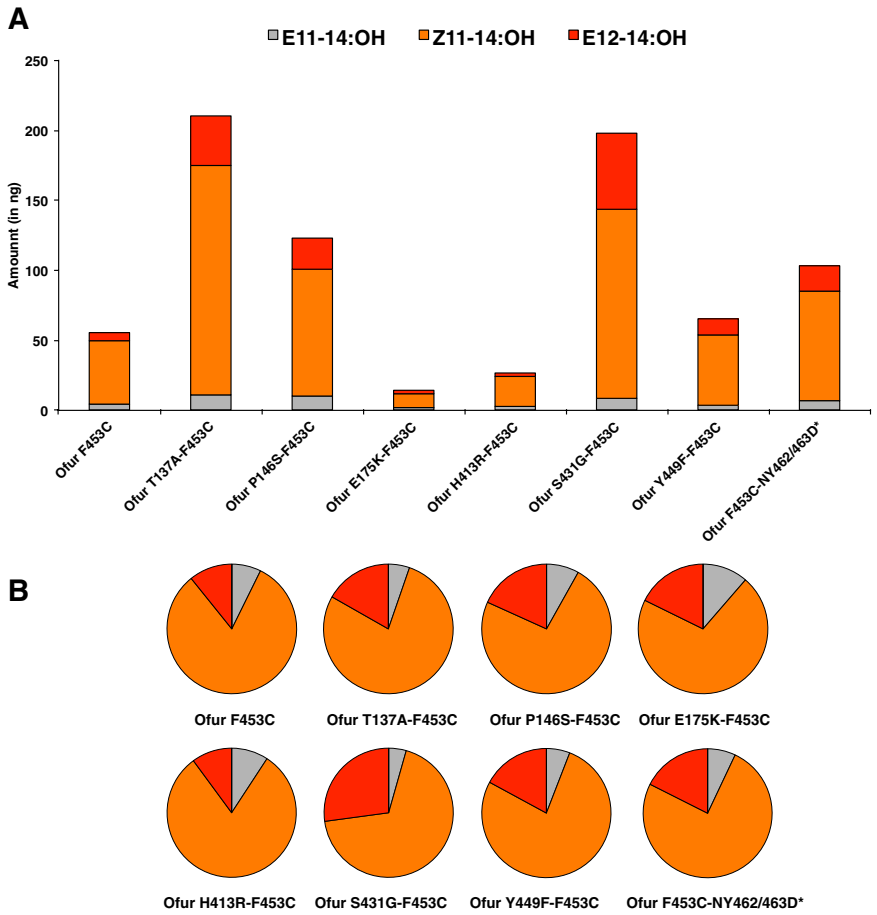

Fig. S5. Fatty alcohols produced by the *Ofurnacalis*-pgFAR^{F453C} mutant and double mutants. (*A*) Each bar in the stacked bar graph represents absolute amounts of fatty alcohol produced by yeast expressing each of the mutant when supplemented with fatty-acyl precursors (n = 2). (*B*) Relative proportions of fatty alcohol illustrating the substrate-preference of mutant enzymes.

Table S1. Pairwise distance matrix of uncorrected p-distance between the coding region of the *pgFAR* orthologs

Species	*OnubZ*	*OscaZ*	*OscaE*	*OnubE*	*Ozag*	*Osp*	*Ozea*	*Ofur*	*Opal*
Nucleotide									
OnubZ									
OscaZ	0.004								
OscaE	0.038	0.038							
OnubE	0.038	0.038	0.004						
Ozag	0.033	0.033	0.032	0.032					
Osp	0.030	0.030	0.031	0.031	0.004				
Ozea	0.040	0.040	0.019	0.020	0.034	0.032			
Ofur	0.018	0.018	0.033	0.033	0.030	0.028	0.036		
Opal	0.084	0.084	0.079	0.079	0.084	0.082	0.080	0.083	
Olati	0.071	0.071	0.069	0.069	0.071	0.070	0.066	0.070	0.053
Amino acid									
OnubZ									
OscaZ	0.009								
OscaE	0.079	0.077							
OnubE	0.077	0.074	0.004						
Ozag	0.070	0.068	0.066	0.066					
Osp	0.068	0.066	0.066	0.066	0.007				
Ozea	0.077	0.074	0.046	0.044	0.072	0.072			
Ofur	0.039	0.039	0.063	0.061	0.066	0.063	0.063		
Opal	0.138	0.136	0.123	0.123	0.136	0.138	0.127	0.138	
Olati	0.112	0.112	0.103	0.103	0.112	0.114	0.096	0.107	0.068

Table S2. Sites identified as having evolved under positive selection

		Method				
			FUBAR		REL	
Codon	Amino acid	BEB [posterior probability ($\omega > 1$)]*	Posterior probability	Empirical Bayes factor	Posterior probability	Bayes factor
2	S/A	0.838	0.831	7.004	0.756	22.187
6	M/T/V	0.840	0.796	5.574	0.904	67.572
15	Y/D/N	0.947	0.904	13.459	0.930	95.260
69	A/G/T	0.834	0.771	4.796	0.700	16.703
103	D/E/H	0.639	0.705	3.402	0.690	15.960
126	V/L	0.928	0.779	5.006	0.725	18.888
137	T/A	0.926	0.764	4.612	0.702	16.839
146	P/S	0.852	—	—	—	—
175	K/E	0.975	0.811	6.123	0.801	28.833
184	S/N/T	0.936	0.714	3.550	0.679	15.113
197	D/E/N	0.598	0.701	3.343	0.678	15.042
198	T/A/I	0.997	0.917	15.688	0.931	96.597
199	E/K	0.893	—	—	—	—
203	I/L	0.957	—	—	—	—
204	T/M/I/V	0.997	0.936	20.971	0.962	180.966
225	V/L	0.687	—	—	—	—
244	V/A/L	0.999	0.858	8.573	0.885	54.879
263	R/Q	0.739	—	—	—	—
266	G/E	0.938	—	—	—	—
270	K/R/E	0.725	0.756	4.421	0.705	17.096
283	V/A	0.951	—	—	—	—
284	F/C	0.952	—	—	—	—
286	V/I/L	0.986	0.912	14.812	0.932	98.537
307	R/K	0.511	—	—	—	—
309	M/I/V	0.869	0.712	3.520	0.660	13.921
322	G/S	0.989	—	—	—	—
330	S/I	0.944	0.738	4.011	0.705	17.131
333	M/V/L	0.978	—	—	0.671	14.615
334	N/K	0.606	—	—	—	—
337	L/I	0.955	—	—	0.706	17.205
352	F/I	0.925	—	—	—	—
364	I/L	0.991	0.787	5.258	0.738	20.146
392	M/I	0.769	—	—	—	—
395	K/N	0.982	—	—	—	—
398	H/D/N	0.990	0.731	3.864	0.702	16.837
401	R/S/G	0.999	0.940	22.251	0.971	236.251
403	L/I	0.819	—	—	—	—
412	H/L	0.992	—	—	—	—
413	H/R	0.798	—	—	—	—
416	L/I	0.819	—	—	—	—
431	G/S/D	0.950	0.862	8.861	0.921	83.162
448	P/M	0.790	0.761	4.523	0.690	15.951
449	Y/F	0.778	—	—	—	—
453	F/C	0.936	0.735	3.951	0.697	16.484
463	Y/N/*	0.805	0.797	5.586	0.754	21.990

*Positively selected sites identified using site model M8 in PAML and the Bayes empirical Bayes procedure.

Table S3. Likelihood ratio tests for positive selection using PAML site-specific models for *Ostrinia pgFAR* genes

Model	Parameter estimates	Parameters	lnL	Likelihood ratio test
M1a: nearly neutral	$p_0 = 0.64648$ $p_1 = 0.35352$ $\omega_0 = 0.02327$ $\omega_1 = 1.00000$	2	−3,604.518	
M2a: selection	$p_0 = 0.87127$ $p_1 = 0.00000$ $p_2 = 0.12873$ $\omega_0 = 0.18631$ $\omega_1 = 1.00000$ $\omega_2 = 3.59987$	4	−3,581.530	M1a vs. M2a 45.98*
M7: neutral, beta	$P = 0.00531$ $q = 0.00702$	2	−3,604.833	
M8: selection, beta + $\omega > 1$	$p_0 = 0.87170$ $p_1 = 0.12830$ $P = 22.86913$ $q = 99.00000$ $\omega = 3.60710$	4	−3,581.564	M7 vs. M8 46.54*
M8a: neutral, beta + $\omega = 1$	$p_0 = 0.64528$ $p_1 = 0.35472$ $P = 2.31757$ $q = 99.00000$ $\omega = 1.00000$	4	−3,604.537	M8a vs. M8 45.95*

The site-specific models implemented in Codeml (PAML 4.4 package) were used to test for positive selection acting on the *pgFAR* open-reading frames. The log likelihoods (lnL) of each model were compared in likelihood ratio tests. *P* values were determined from χ^2 distributions.
*Significant at the 1% level.