SUPPLEMENTARY MATERIAL

Locus	Number of pairwise comparisons	Significant pairwise comparisons by Fisher's exact test	After Bonferroni correction	Significant pairwise comparisons by χ^2 test	After Bonferroni correction
Apt	990	286	83	301	202
Cycle	300	86	15	131	18
Cyp303a1	1830	1119	683	1156	633
Cyp305b1	741	97	5	126	43
Period	378	40	1	48	30
Phc	1953	325	28	397	115
SCAP	780	93	0	171	0
Tc	946	110	0	129	0
Tpi	2926	375	36	549	87

Table S1 Average r^2 and number of significant pairwise comparisons before and after Bonferroni correction

Locus	Accession		Primer sequence
Apt	AB024903.1	forward:	CTTGGTCTACTGCCGACCTCACT
		reverse:	TAGCCATCCTCATTGTTTGGACT
Cycle	BAB20632.1	forward:	GCTATGGCGAGGAAGTTAGACAA
		reverse:	TCTCTCCCGTGGGCTGAGGT
Cyp303a1	NM_143813.2	forward:	ACGGCATTTCATGGGGCGCA
		reverse:	GTGCTTGGCTAGGCCGAACGGA
		forward, nested:	GCTCGCTGGATATTTTATACCAGA
Cyp305b1	NP_001106220.1	forward:	ACACGTCTGCGTTTCTCCAA
		reverse:	GGTGTAGCCAATATACCAATCAAC
down3		forward:	TCTCCATTCGGTTTCCCTTC
		reverse:	TAATCACTGGGTTGCTTCTGG
Period	ABF21088.1	forward:	CAGGGACTCGGGTGAGATGA
		reverse:	AGCACTGGTTGGATGGTAGG
Phc	XP_002432072.1	forward:	TACGCGAAGATGTGGTACAAGG
		reverse:	ACAAGTCCATCGGCGGTCTG
SCAP	XP_974766.1	forward:	AAAGCCAAAGCGAGCATAGCA
		reverse:	TACTGAACAAGCAGCCAGACC
Tc	XP_972068.2	forward:	TGTATTCCCAAGTCCGCTGTTT
		reverse:	TTGTTGATAGCTTCGCAAGAGT
Tpi	AY736358	forward:	ATTCGTTGTTGGTGGTAACTGGA
		reverse:	TTTGCCTGCCTCCCTCTCTTCT
up3		forward:	GTTGCGAGTAAGATAGCAGCAC
		reverse:	GTTGCGTGGCAGGAAGGTAA

Table S2 Accessions of Z-linked loci and primer sequences used in this study. In cases where sequence quality was unsatisfactory, an internal (nested) primer was used for re-sequencing. The loci *down3* and *up3* refer to the regions 3 kb downstream and upstream of *Cyp303a1*, respectively, in which no known genes were identified.

	F_{ST} values for pairwise comparisons of populations		
Locus	**Dal X Orb**	**Dal X MV**	**Orb X MV**
Apt	-0.077	0.020	0.013
Cyp303a1	-0.036	-0.012	0.018
Period	-0.017	-0.005	0.025
Phc	0.059	-0.028	-0.009
Tpi	-0.044	0.003	-0.176

Table S3 F_{ST} values for five loci sampled from Dalmore, Orbost and MacIntyre Valley populations using an unbiased estimator (Hudson *et al.*, 1992)

Locus	n	Number of sites (bp)[a]	Number of indel events		π(i)	
Apt **(55)**	Dalmore (8)	779–840	8	23	0.004	0.005
	Orbost (5)		4		0.003	
	M.Valley (41)		20		0.005	
Cycle **(21)**	M.Valley (20)	919		11		0.002
Cyp303a1 **(83)**[b]	Dalmore (14)	774–780	13	20	0.005	0.006
	Orbost (10)		7		0.005	
	M.Valley (56)		17		0.006	
Cyp305b1 **(22)**	M.Valley (21)	696		8		
Period **(36)**	Dalmore (13)	486–502	27	46	0.015	0.014
	Orbost (5)		12		0.011	
	M.Valley (17)		32		0.013	
Phc **(36)**	Dalmore (12)	723–754	28	38	0.011	0.010
	Orbost (7)		20		0.010	
	M.Valley (16)		25		0.009	
SCAP **(12)**	M.Valley (11)	974		20		
Tc **(16)**	M.Valley (15)	914		16		0.004
Tpi **(33)**	Dalmore (11)	551–577	18	21	0.009	0.007
	Orbost (2)		3		0.005	
	M.Valley (19)		14		0.005	

Table S4 Indel diversity for nine loci surveyed in this study. Where estimates are presented in two columns under a single heading, the left column represents estimates for an individual population while the right column represents estimates after pooling sequences of all three populations. Figures in brackets after the locus name represent the total number of sequences surveyed including the reference strain.

* *$p<0.05$*

[a] *the number of sites is presented as a range due to the differing subsets of indel polymorphisms present in different populations. Since gapped sites are included in this analysis, the upper boundary represents the number of sites considered when alleles from all three populations are pooled.*

[b] *includes two sequences from a laboratory-maintained colony*

Site	*Del200*	*Ins200*	*H. assulta*	*H. punctigera*	*H. zea*
1	ATAA	-	x	x	GTAA
2	G(1A)	C	C	x	C
3	C	T	T	x	T
4	T	C	C	x	T
5	G	A/T	T	x	T
6	T(1G)	A	A/C	x	A
7	C	G	G	x	G
8	A	G	G	x	G
9	G	A	G	x	G
10	C(1G)	A	C	x	C
11	G	T	T	x	T
12	G	C	G/A	x	C
13	T	A	T	x	T
14	G	T	G	x	T
15	A	G	G	x	G
16	T	C	T/-	x	C
17	G	T	T/-	x	T
18	C	G	C	x	C
19	C	A	C	x	C
20	A	G	G	x	-
21	A	G	G	x	-
22	A	G	A	x	G
23	A	C	C	x	C
24	T	A/C	T	T	T
25	A	C	C	C	C
26	A	T	T	T	T
27	T	C	C	C	C
28	C	A	A	A	C
29	T	A	A	A	G
30	AACG	-	AATG	AATN	-
31	-	A	GA	-	A
32	A	C	A	A	G

33	A	C	A	A	G
34	A	G	A	A	T
35	G	A	A	A	A
36	G	A	AG	-	G
Number of sites with state identical or closer to the					
Del200 haplogroup			10	14	5
Ins200 haplogroup			20	19	6
Ambiguous unknown:			6	3	25

Table S5 State of the outgroup consensus sequences at sites where fixed differences occur between the *H. armigera* insertion and deletion haplogroups. '-' represents a gap in the alignment flanked by matching nucleotides whereas 'x' represents a gap that forms part of a deletion extending beyond the boundaries of the sites where the fixed differences occur. All sites marked with an 'x' are deemed ambiguous. Sites with a '-' are considered unambiguous if they match one of the two states in the *H. armigera* alleles e.g. site 20 in *H. zea* is considered ambiguous but site 30 is not. Sites which require more than one change from either allele are also deemed ambiguous e.g. site 30 in *H. assulta* is considered to be closer to the insertion state but the same site in *H. punctigera* is deemed ambiguous because a single event is insufficient to produce a match to either state. Similarly, site 31 in *H. assulta* is ambiguous because only one event (a mono- or di-nucleotide insertion) is sufficient to produce a match to either state. Where polymorphisms are present in the outgroup sequences, they are deemed unambiguous if at least one of the states is identical to that of *H. armigera* and all possible states are observed to be exclusive of the alternate *H. armigera* allele. Sites 2, 5, 6, 10 and 24 are, strictly speaking, not fixed differences, but have been included because no overlapping states have been observed between the two *H. armigera* alleles.

Organism	π (10^2)	Approximate distance (bp) at which E(r^2)≈0.2	References
Lepidoptera			
Bombyx mori	1	>1600[a]	Xia *et al.* (2009); Guo *et al.* (2011)
Bombyx mandarina	1–2	20–200[a]	Xia *et al.* (2009); Guo *et al.* (2011)
Heliconius erato	2	<500	Counterman *et al.* (2010)
Heliconius melpomene	1	>500	Baxter *et al.* (2010)
Helicoverpa armigera	3	10–200[b]	this study
Melitaea cinxia	n.a	3000	Ahola *et al.* (2014)
Other insects			
Acyrthosiphon pisum	0.6	1000	Brisson *et al.* (2009)
Anopheles arabiensis	0.2–0.3	<200	Marsden *et al.* (2014)
Anopheles gambiae	0.8–25	<200	Wilding *et al.* (2009); Harris *et al.* (2010)
Drosophila melanogaster	0.6–2	30[c]–640[d]	Langley *et al.* (2012); Pool *et al.* (2012)

Table S6 Estimates of nucleotide diversity, π and linkage disequilibrium, E(r^2) in different species. Unless otherwise indicated, LD was imputed from diploid sequences.

[a] *haploid by cloning amplicons prior to sequencing*

[b] *haploid by sequencing sex chromosomes in hemizygous individuals*

[c] *haploid embryos*

[d] *haploid by using inbred lines*

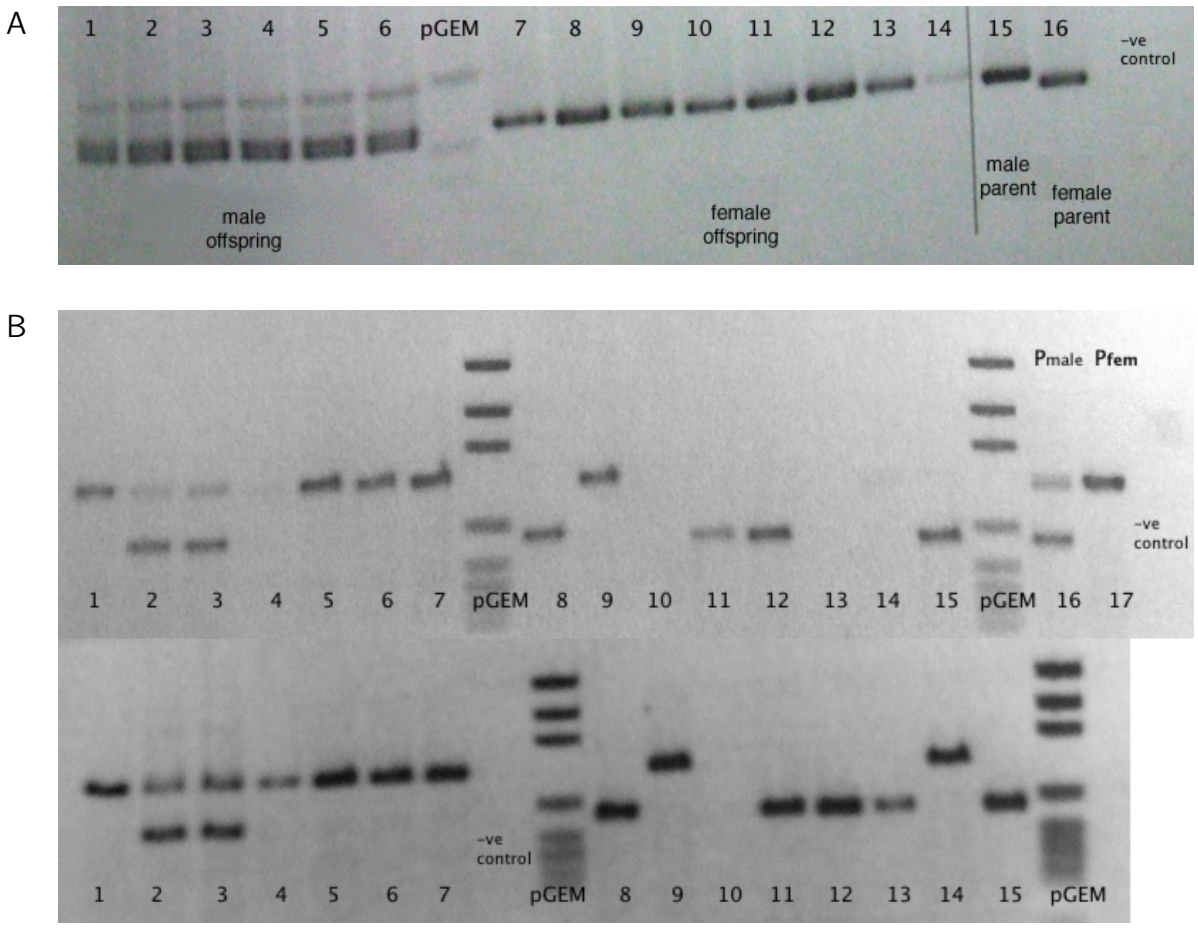


Figure S1 Pedigree analysis of (A) *Period*. All male offspring (lanes 1–6) present double bands inherited from both parents while all female offspring (7–14) inherit a single band from the male parent (15). The female parent has a slightly shorter product (16) and her allele is only seen in male offspring. Pedigree analysis of (B) *Phc*. The upper gel shows male offspring (lanes 1–7), female offspring (8–15), male parent (16) and female parent (17). The male parent is heterozygous. The lower gel stems from a technical replicate (different PCR) to account for lack of product in samples 4, 10, 13 and 14. Male offspring exhibit heterozygosity (2,3) while female offspring (excluding sample 10) exhibit hemizygosity.

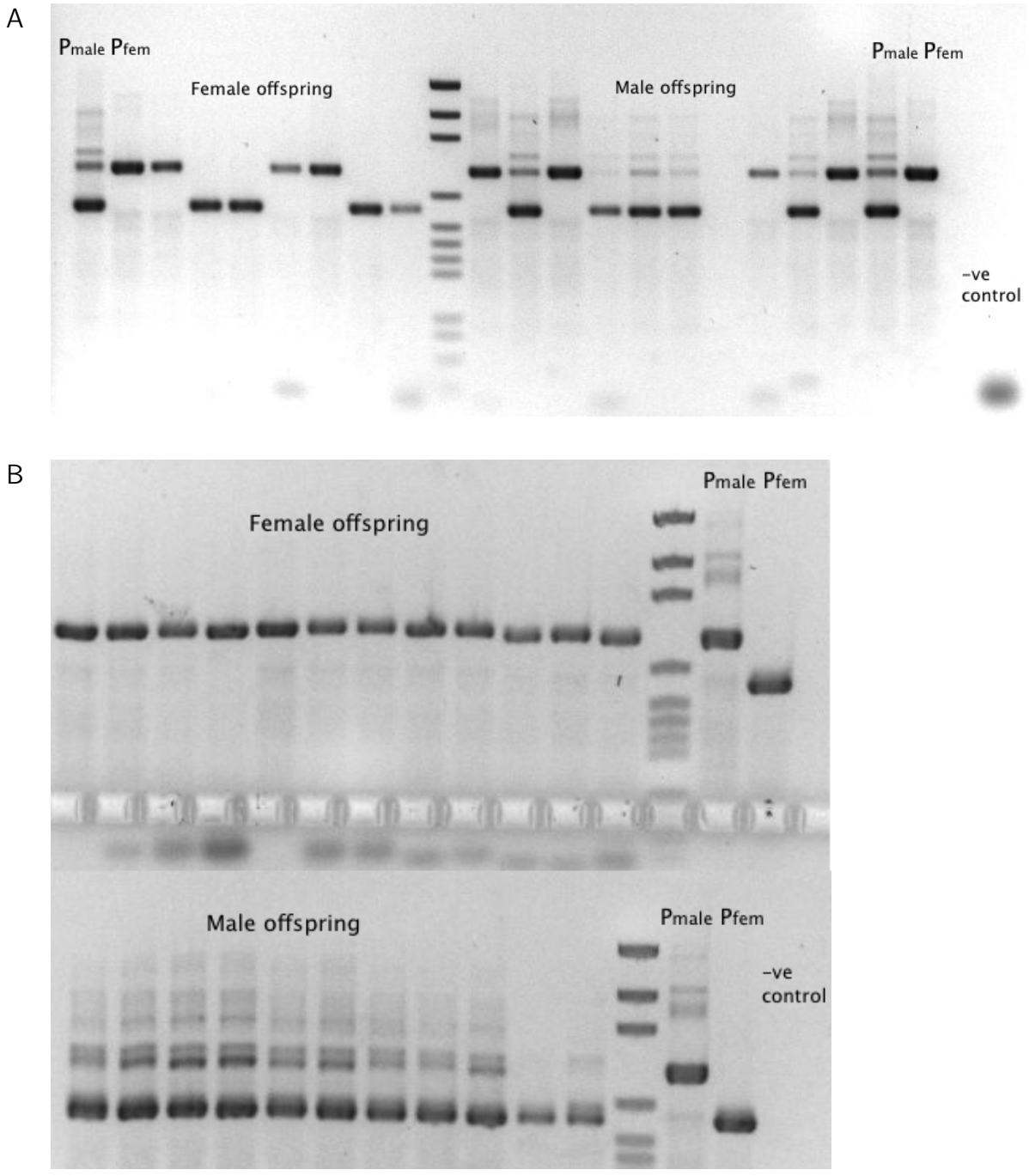


Figure S2 Pedigree analysis on the *Cyp303a1* insertion and deletion variants. The male parent of family 1 (A) is heterozygous while that of family 2 (B) is homozygous for the insertion. In both cases, female offspring exhibit hemizygosity whereas male offspring may be homozygous or heterozygous.

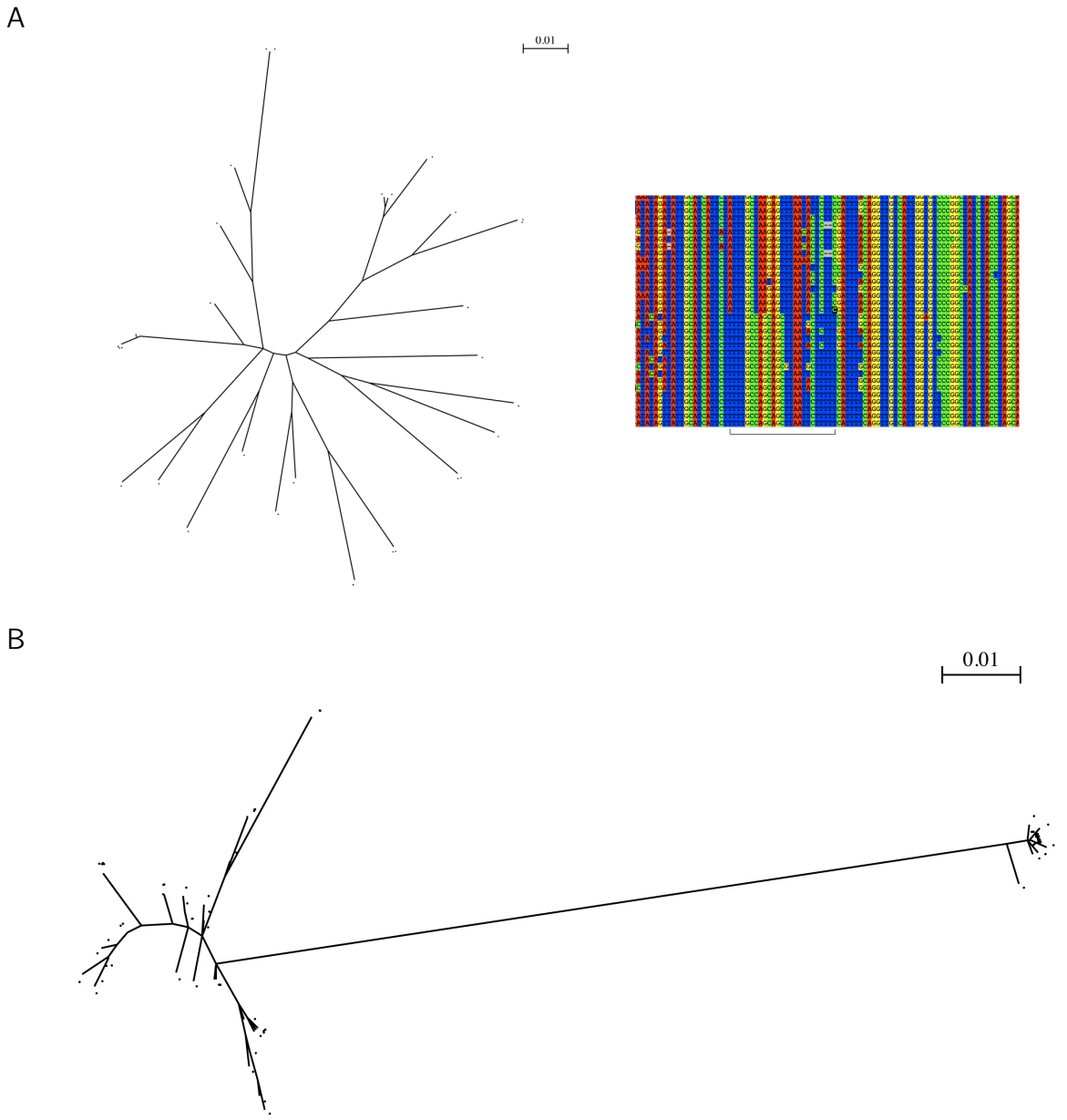
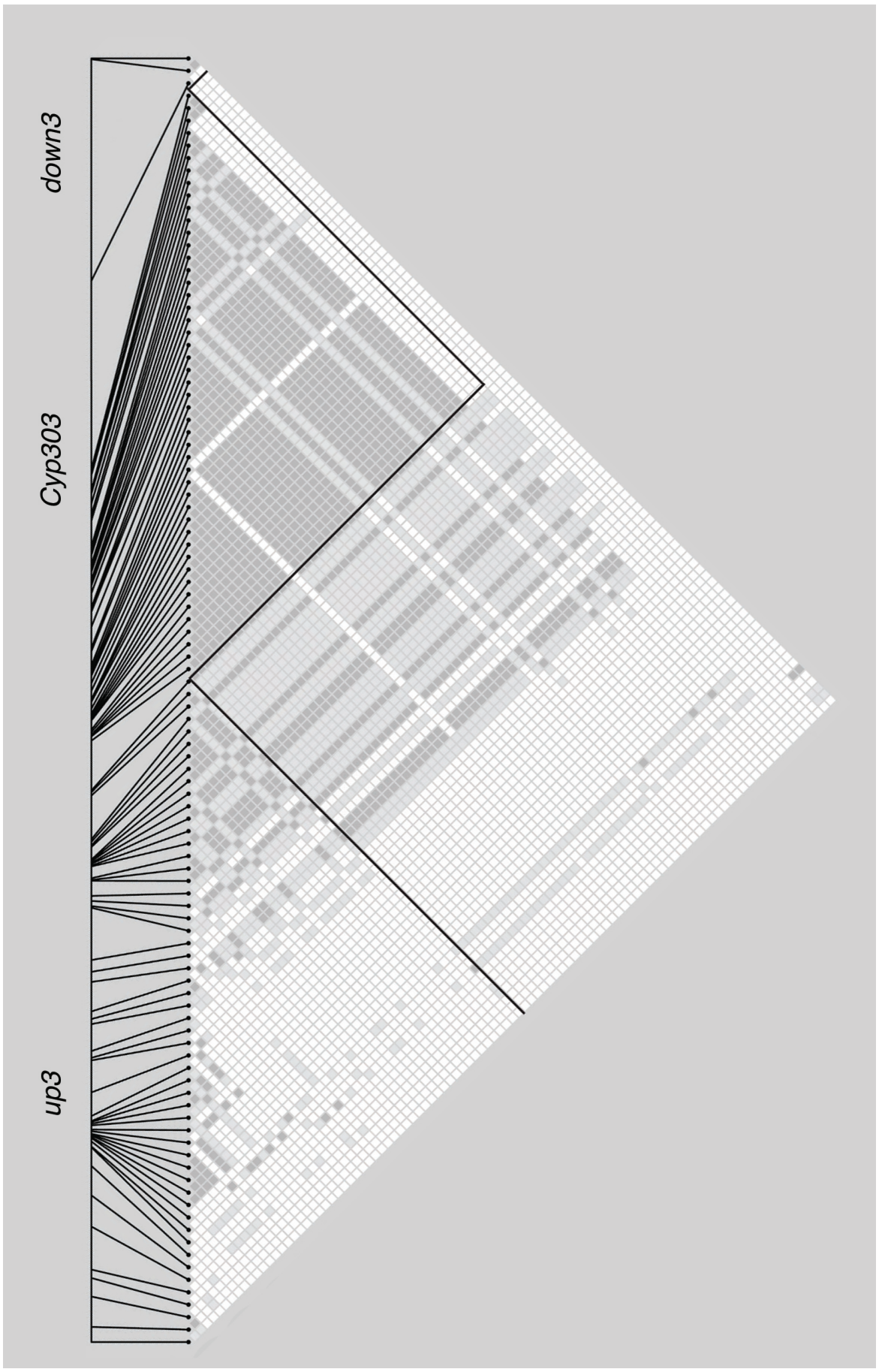


Figure S3 Unrooted maximum-likelihood trees for (A) *Tpi* and (B) *Cyp303a1* after exclusion of the 25-bp stretch and 200-bp indel, respectively. The long internal branch is still apparent in *Cyp303a1*. The fixed differences between the insertion and deletion alleles cannot be accounted for by a consecutive stretch of nucleotides that would constitute a single event. In contrast, the alignment in (A) illustrates that excluding a single 25-bp stretch in *Tpi* (underlined with black bar) largely eliminates the long internal branch.


down3
Cyp303
up3

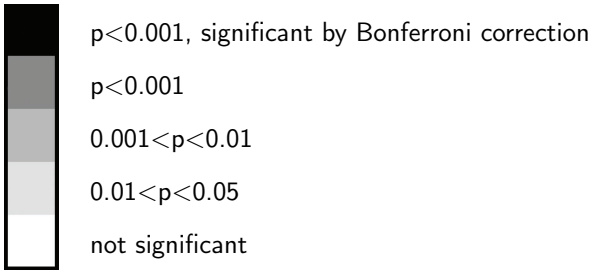


Figure S4 ***(preceding page)*** LD plotted as significance of the r^2 value for *Cyp303a1* and its flanking regions. The sequence under consideration was generated by concatenating the *up3*, *Cyp303a1* and *down3* loci (n=11). The range of each locus is bounded by black lines. Approximately 3 kb separate *Cyp303a1* from its flanking sequences. The downstream locus contained very few segregating sites, hence the small area. LD is observable between the 3' end of the upstream locus and *Cyp303a1*.

SUPPLEMENTARY REFERENCES

Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Valimaki N, Paulin L, Kvist J, Wahlberg N, Tanskanen J, Hornett EA, Ferguson LC, Luo S, Cao Z, de Jong MA, Duplouy A, Smolander OP, Vogel H, McCoy RC, Qian K, Chong WS, Zhang Q, Ahmad F, Haukka JK, Joshi A, Salojarvi J, Wheat CW, Grosse-Wilde E, Hughes D, Katainen R, Pitkanen E, Ylinen J, Waterhouse RM, Turunen M, Vaharautio A, Ojanen SP, Schulman AH, Taipale M, Lawson D, Ukkonen E, Makinen V, Goldsmith MR, Holm L, Auvinen P, Frilander MJ and Hanski I (2014). The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* **5**

Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Dawson A, Beltran M, Perez-Espona S, Chamberlain N, Ferguson L, Clark R, Davidson C, Glithero R, Mallet J, McMillan WO, Kronforst M, Joron M, ffrench Constant RH and Jiggins CD (2010). Genomic hotspots for adaptation: The population genetics of Mullerian mimicry in the *Heliconius melpomene* clade. *PLoS Genet* **6**: e1000794

Brisson J, Nuzhdin S and Stern D (2009). Similar patterns of linkage disequilibrium and nucleotide diversity in native and introduced populations of the pea aphid, *Acyrthosiphon pisum*. *BMC Genetics* **10**: 22

Counterman BA, Araujo-Perez F, Hines HM, Baxter SW, Morrison CM, Lindstrom DP, Papa R, Ferguson L, Joron M, ffrench Constant RH, Smith CP, Nielsen DM, Chen R, Jiggins CD, Reed RD, Halder G, Mallet J and McMillan WO (2010). Genomic hotspots for adaptation: The population genetics of Mullerian mimicry in *Heliconius erato* **6**

Guo Y, Shen YH, Sun W, Kishino H, Xiang ZH and Zhang Z (2011). Nucleotide diversity and selection signature in the domesticated silkworm, *Bombyx mori*, and wild silkworm, *Bombyx mandarina*. *Journal of Insect Science* **11**: 155

Harris C, Rousset F, Morlais I, Fontenille D and Cohuet A (2010). Low linkage disequilibrium in wild *Anopheles gambiae s.l.* populations. *BMC Genetics* **11**: 81

Hudson RR, Slatkin M and Maddison WP (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589

Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, Fang S, Nista PM, Holloway AK, Kern AD, Dewey CN, Song YS, Hahn MW and Begun DJ (2012). Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* **192**: 533–598

Marsden CD, Lee Y, Kreppel K, Weakley A, Cornel A, Ferguson HM, Eskin E and Lanzaro GC (2014). Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector Anopheles arabiensis. *G3 (Bethesda)* **4**: 121–131

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, Duchen P, Emerson JJ, Saelao P, Begun DJ and Langley CH (2012). Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet* **8**: e1003080

Wilding C, Weetman D, Steen K and Donnelly M (2009). High, clustered, nucleotide diversity in the genome of *Anopheles gambiae* revealed through pooled-template sequencing: implications for high-throughput genotyping protocols. *BMC Genomics* **10**: 320

Xia Q, Guo Y, Zhang Z, Li D, Xuan Z, Li Z, Dai F, Li Y, Cheng D, Li R, Cheng T, Jiang T, Becquet C, Xu X, Liu C, Zha X, Fan W, Lin Y, Shen Y, Jiang L, Jensen J, Hellmann I, Tang S, Zhao P, Xu H, Yu C, Zhang G, Li J, Cao J, Liu S, He N, Zhou Y, Liu H, Zhao J, Ye C, Du Z, Pan G, Zhao A, Shao H, Zeng W, Wu P, Li C, Pan M, Li J, Yin X, Li D, Wang J, Zheng H, Wang W, Zhang X, Li S, Yang H, Lu C, Nielsen R, Zhou Z, Wang J, Xiang Z and Wang J (2009). Complete resequencing of 40 genomes reveals domestication events and genes in silkworm *(Bombyx)*. *Science* **326**: 433–436