nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0746-5

In the format provided by the authors and unedited.

Fall webworm genomes yield insights into rapid adaptation of invasive species

Ningning Wu[1,2,6], Sufang Zhang[3,6], Xiaowei Li[1,6], Yanghui Cao[1], Xiaojing Liu[1], Qinghua Wang[3], Qun Liu[1], Huihui Liu[1,3], Xiao Hu[1], Xuguo J. Zhou[4], Anthony A. James[5], Zhen Zhang[3]*, Yongping Huang[1]* and Shuai Zhan[1]*

[1]CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. [2]University of Chinese Academy of Sciences, Shanghai, China. [3]Key Laboratory of Forest Protection, Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, State Forestry Administration, Beijing, China. [4]Department of Entomology, University of Kentucky, Lexington, KY, USA. [5]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA, USA. [6]These authors contributed equally: Ningning Wu, Sufang Zhang, Xiaowei Li. *e-mail: szhan@sibs.ac.cn; yphuang@sibs.ac.cn; zhangzhen@caf.ac.cn

Supplementary Information for:

Fall webworm genomes yield insights into rapid adaptation of invasive species

Ningning Wu[1,2#], Sufang Zhang[3#], Xiaowei Li[1#], Yanghui Cao[1], Xiaojing Liu[1], Qinghua Wang[3], Qun Liu[1], Huihui Liu[1,3], Xiao Hu[1], Xuguo J. Zhou[4], Anthony A. James[5], Zhen Zhang[3]*, Yongping Huang[1]*, Shuai Zhan[1]*

[1]CAS Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[2]University of Chinese Academy of Sciences, Shanghai 200032, China

[3]Key Laboratory of Forest Protection, Research Institute of Forest Ecology, Environment and Protection, Chinese Academy of Forestry, State Forestry Administration, Beijing 100091, China

[4]Department of Entomology, S-225 Agricultural Science Center North, University of Kentucky, Lexington, KY 40546-0091, USA

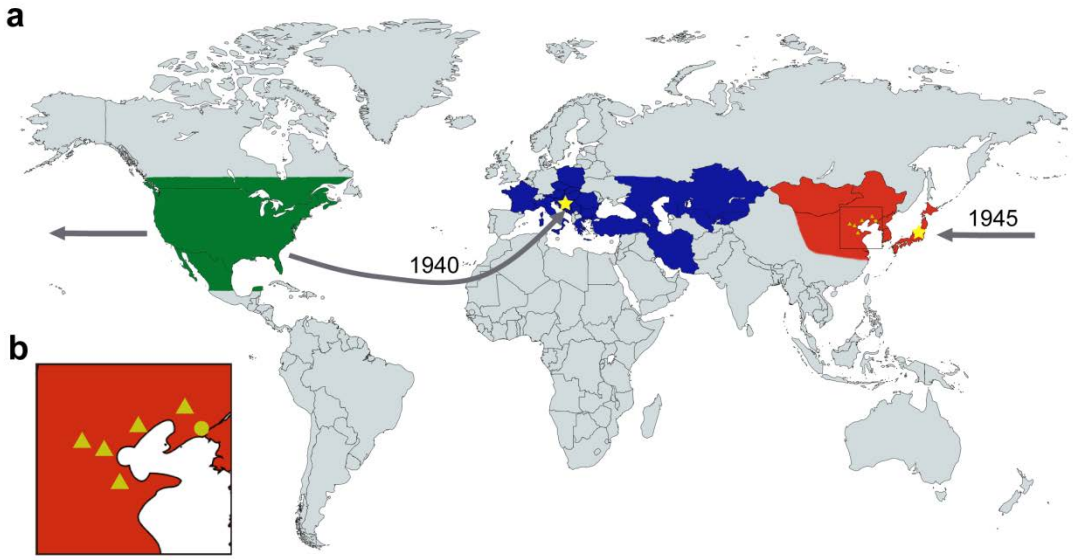
[5]Department of Molecular Biology and Biochemistry, University of California, Irvine, 3205 McGaugh Hall, Irvine, CA 92697-3900, United States

[#] These authors contributed equally to this work.

[*] To whom correspondence should be addressed: szhan@sibs.ac.cn (Sh.Z.); yphuang@sibs.ac.cn (Y.H.); zhangzhen@caf.ac.cn (Z.Z.)

Contents:

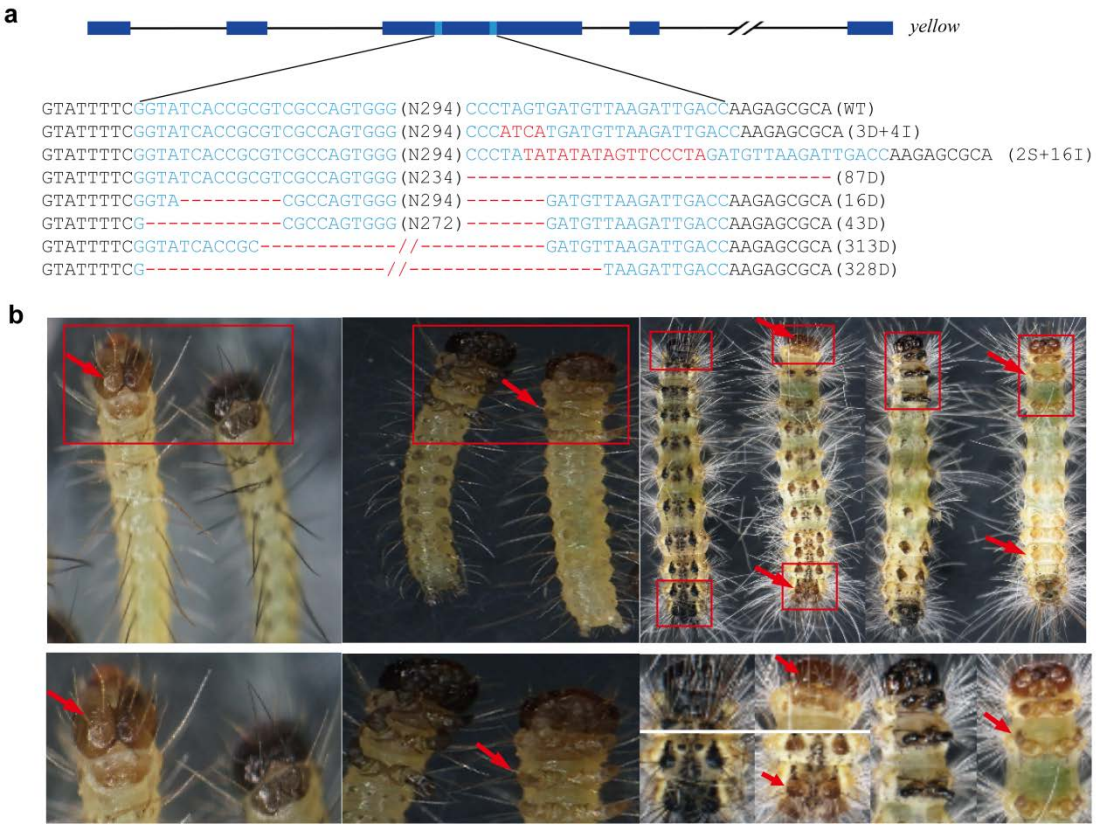
Supplementary Figure 1 ---------- 3
Supplementary Figure 2 ---------- 4
Supplementary Figure 3 ---------- 5
Supplementary Figure 4 ---------- 6
Supplementary Figure 5 ---------- 7
Supplementary Figure 6 ---------- 8
Supplementary Figure 7 ---------- 9
Supplementary Figure 8 ----------10
Supplementary Figure 9 ---------- 11
Supplementary Figure 10 ---------- 12
Supplementary Figure 11 ---------- 13
Supplementary Figure 12 ---------- 14
Supplementary Figure 13 ---------- 15
Supplementary Figure 14 ---------- 16
Supplementary Figure 15 ---------- 17
Supplementary Figure 16 ---------- 18
Supplementary Figure 17 ---------- 19
Supplementary Table 1 ---------- 20
Supplementary Table 2 ---------- 21
Supplementary Table 3 ---------- 22
Supplementary Table 4 ---------- 23
Supplementary Table 5 ---------- 24
Supplementary Table 6 ---------- 25
Supplementary Table 7 ---------- 26
Supplementary Table 8 ---------- 27
Supplementary Table 10 ---------- 28
Supplementary Table 11 ---------- 29
Supplementary Table 16 ---------- 30
Supplementary Table 18 ---------- 31
Supplementary References ---------- 32



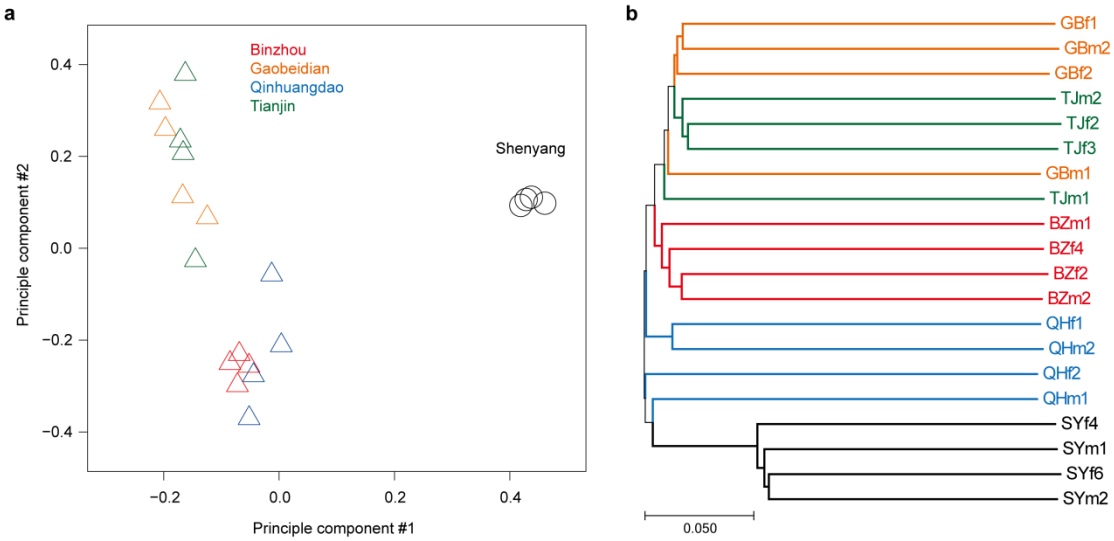
Supplementary Figure 1 Invasive paths and distribution of the fall webworm
a, The fall webworm, is native to North America (in green). In 1940s, fall webworms were first found in East Europe and East Asia, respectively (indicated by yellow stars). Since then, the invasive fall webworms have spread widely during the last years and colonized most range of northern hemisphere (in blue and red, respectively). The invasive populations in this study were sampled in China, which were supposed to be introduced to China in Dandong in 1979 (the yellow circle). The background world map was created with mapchart.net (https://mapchart.net/). Invasive paths and distribution were based on previous studies[1,2]. **b**, Sequencing data in this study was sampled from the populations indicated by yellow triangles. See details in **Supplementary Table 9**.

Supplementary Figure 2 Fall webworms cause damage to host plants in China

As shown, the larvae of fall webworms caused serious damage to the colonized plants by creating massive webs. The photographs were taken by Qinghua Wang in Tangshan, China in 2015. Images of representative developmental stages in the life cycle of fall webworms are shown inside.

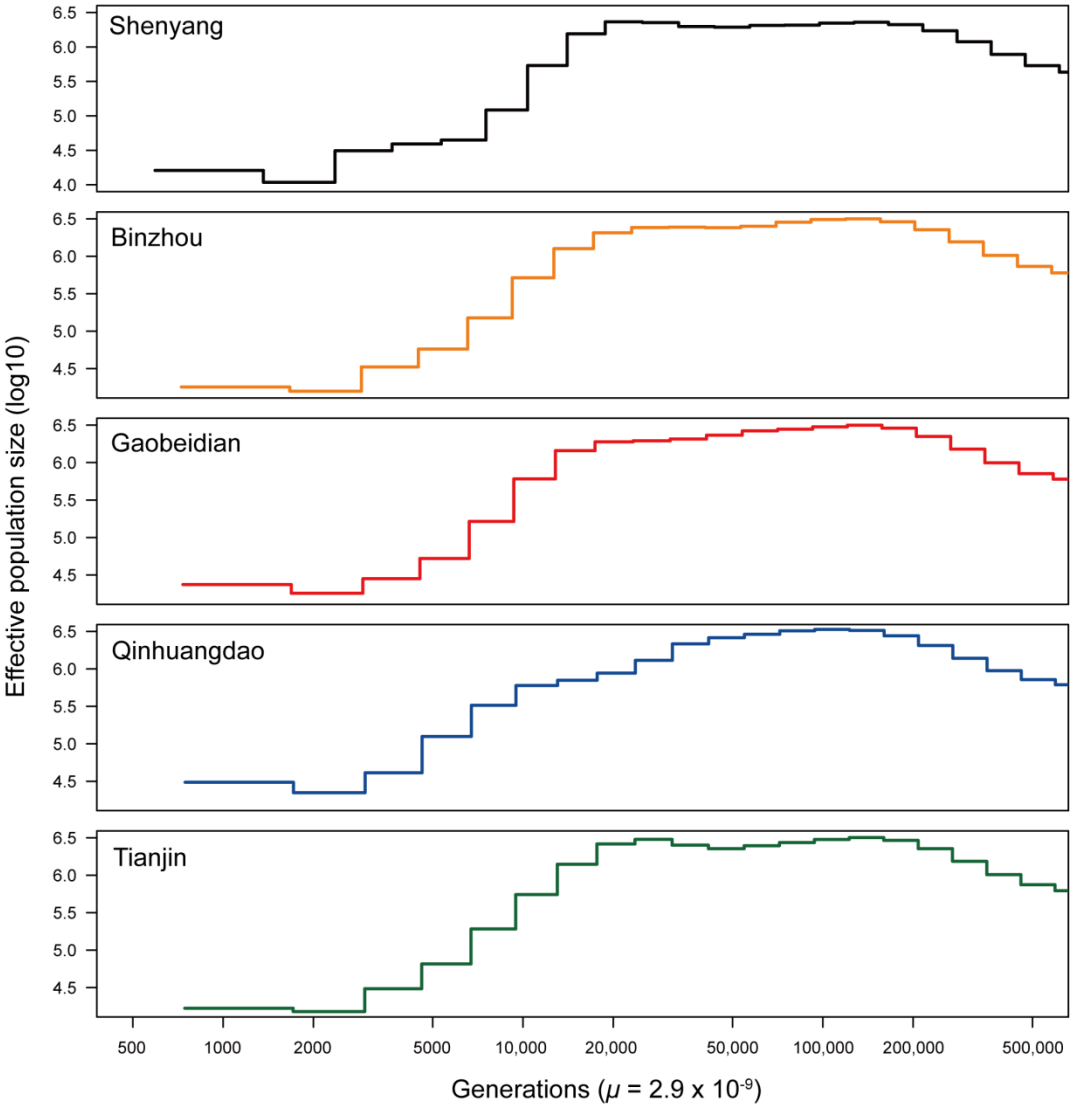


Supplementary Figure 3 CRISPR/Cas9- mediated knockout of *yellow* in *Hyphantria cunea*
a, Gene structure of Hc-*yellow* and the target sites used in designing of dsRNAs. Examples of genotyping result, including both insertions and deletions, indicate a high level of mutation in G0 mutants. **b**, Visible phenotypic changes were found in many G0 mutations (indicated by red arrows and enlarged in red boxes).

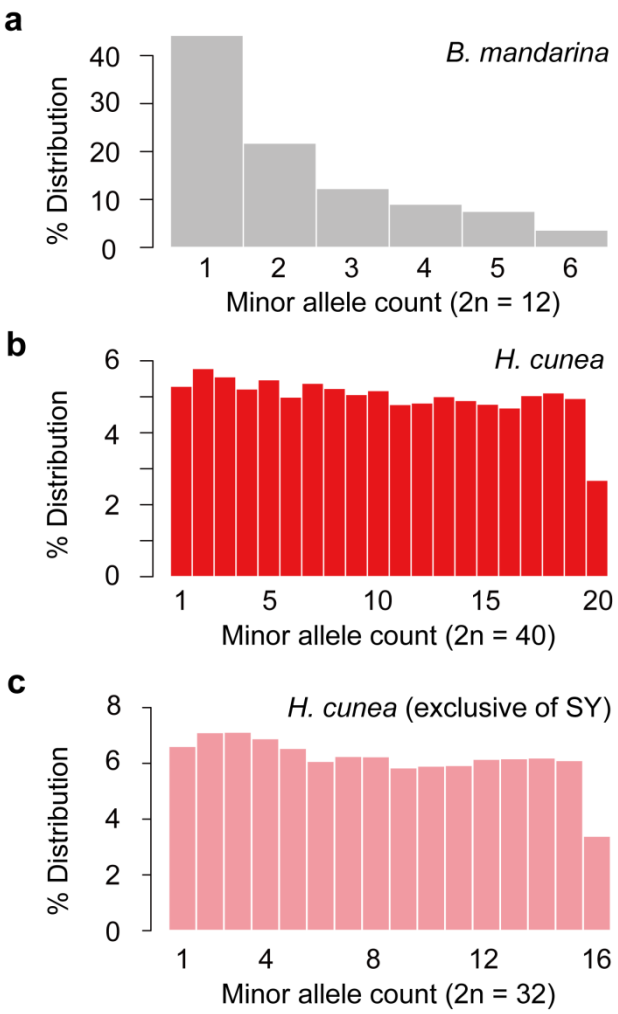


Supplementary Figure 4 Population structure and phylogenetic relationship of sequenced invasive fall webworms in China

a, Principle component analysis (PCA) of genome-wide synonymous SNPs. The first two components are presented for all 20 sequenced individuals. Detailed sampling information is presented in **Supplementary Fig. 1** and **Supplementary Table 9**. **b**, Neighbor-joining phylogenetic tree based on genome-wide SNPs. Note that Shenyang is close to the site of first introduction of fall webworms in China. Documented expansion ranges showed fall webworms were subsequently dispersed both southwards and westwards. Qinhuangdao is the necessary from Shenyang to other dispersed regions.

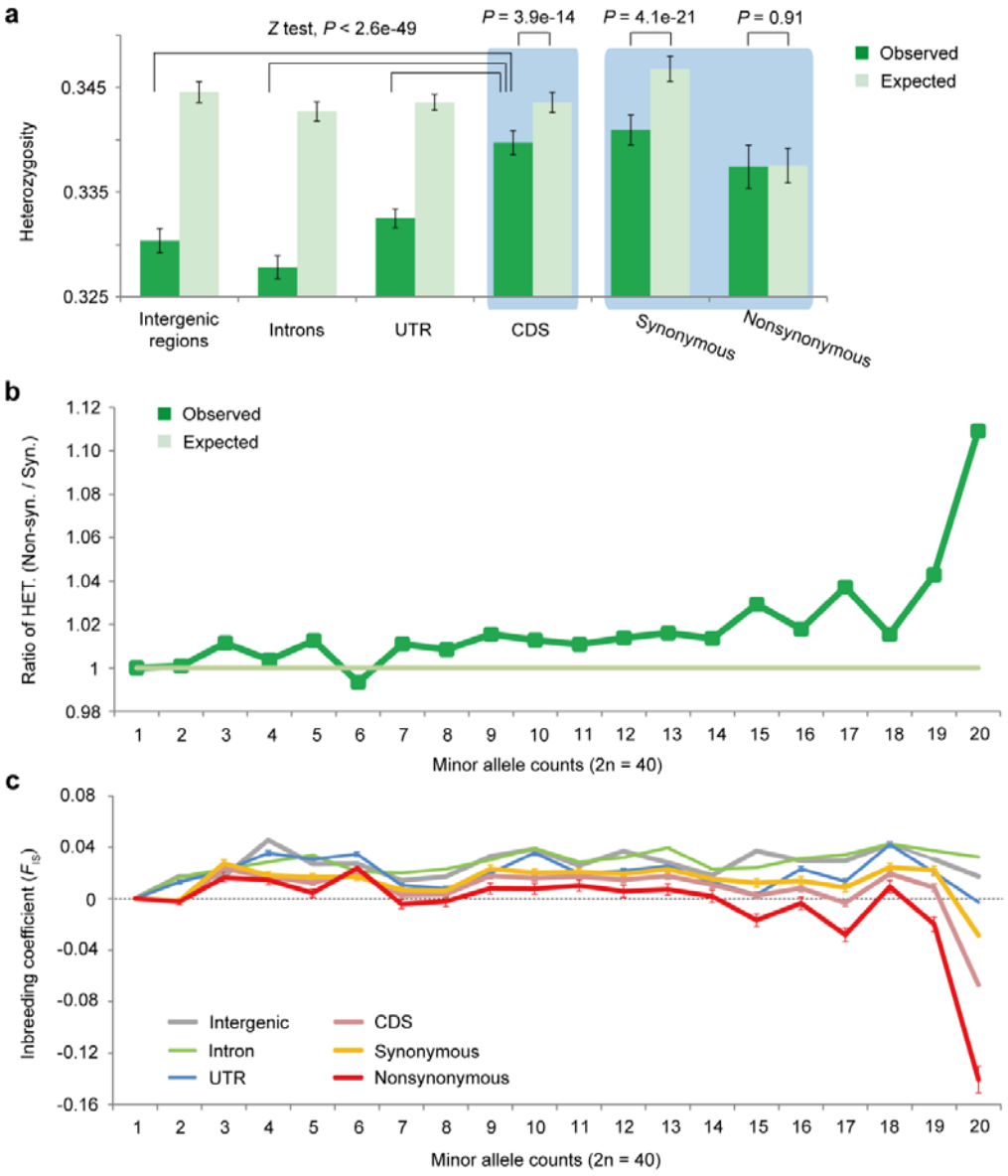


Supplementary Figure 5 Inferred population history of the invasive fall webworms in China
The demographic analysis was reconstructed by the multiple sequentially Markovian coalescent (MSMC)[3]. Both generations and effective population sizes were converged from the scaled coalescence rate by an estimated mutation rate of Lepidoptera (*Heliconius*; 2.9×10^{-9})[4]. See the result of an alternative approach, SMC++, in **Fig. 2c**.



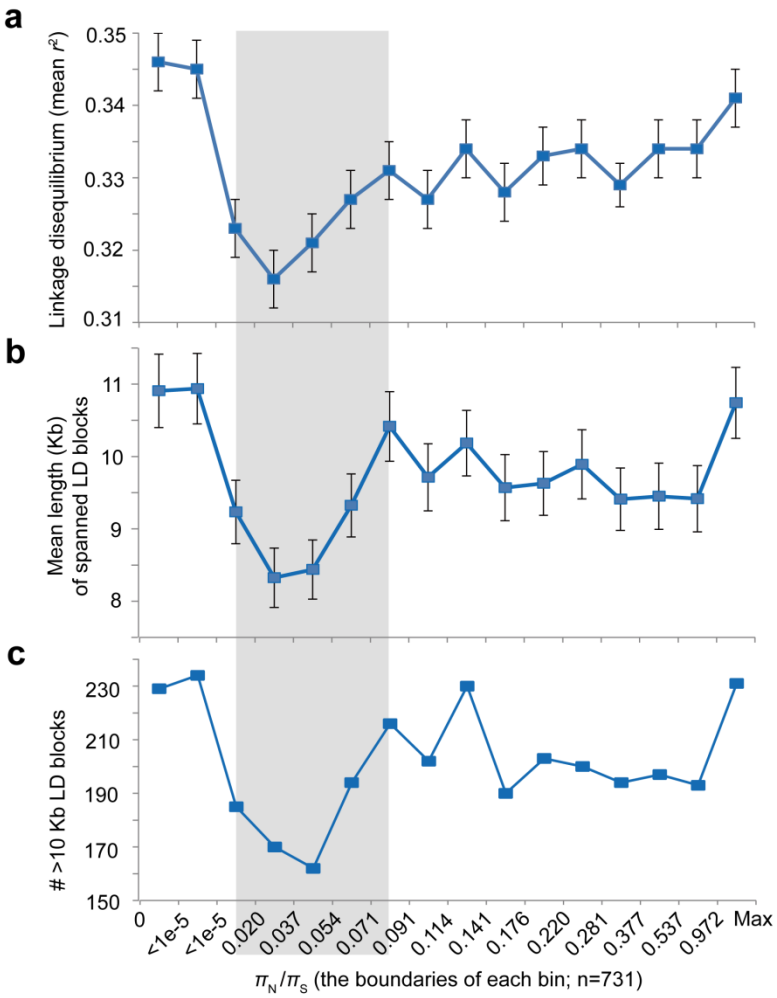
Supplementary Figure 6 Distribution comparison of entire folded site frequency spectrums between *Hyphantria cunea* and *Bombyx mandarina*

a, MAF distribution of six sequenced *B. mandarina* (the wild silkworm, a Chinese local species). **b**, MAF of 20 sequenced *H. cunea* (the invasive population of fall webworms in China). **c**, MAF of 16 sequenced *H. cunea*, in which the population of Shenyang, a partially isolated population from the other four subpopulations, was excluded from the analysis. The pattern without Shenyang samples (as shown in **c**) is marginally different with that based on all samples (as shown in **b**), suggesting that the difference between *H. cunea* and *B. mandarina* (**a** versus **b**) is unlikely due to the inclusion of this partially isolated population (see **Supplementary Fig. 4**).



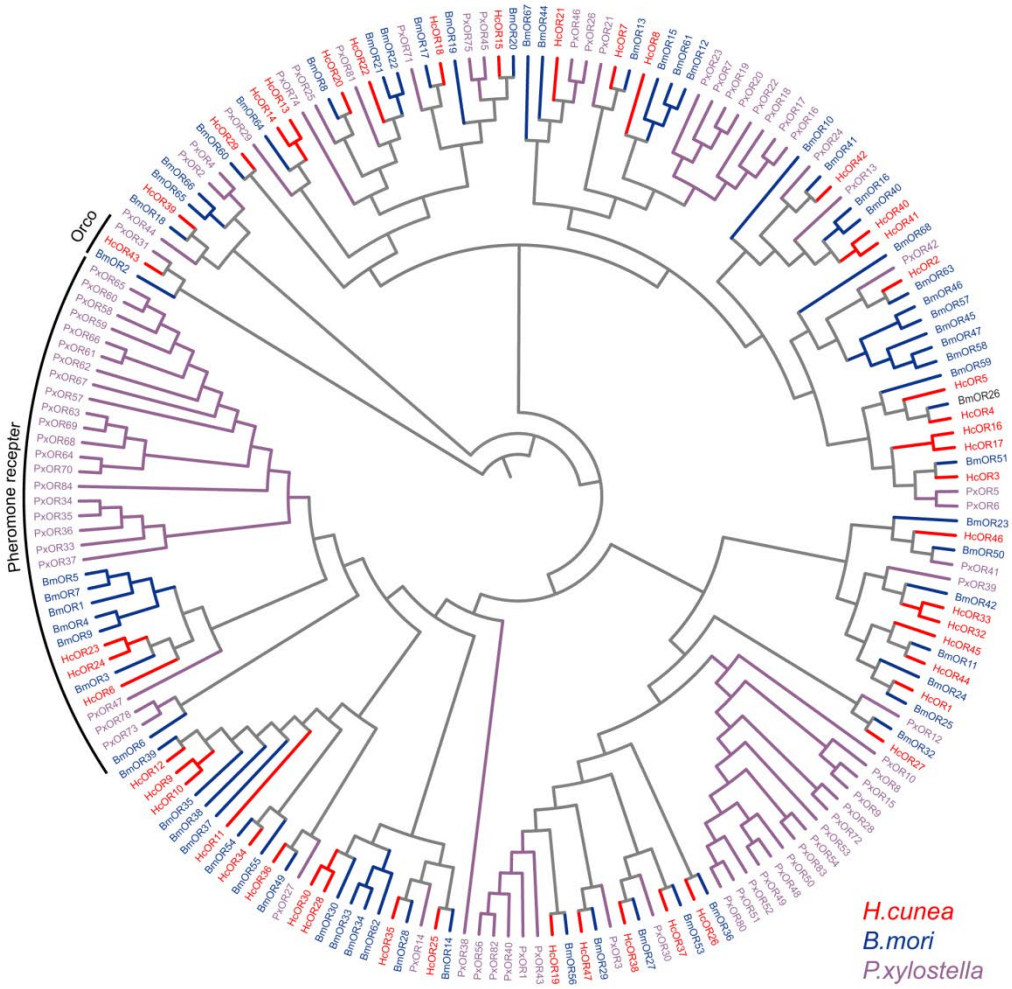
Supplementary Figure 7 Excess heterozygotes in nonsynonymous SNPs in invasive fall webworm genomes

a, Observed frequency of heterozygotes (H_o) is significantly elevated within protein-coding regions (CDS) in comparison to any other type of genomic regions. The highest *P* value of three sets of Z-test is presented. CDS was further classified as synonymous and nonsynonymous sites (in blue shadows). In comparison to synonymous sites, H_o in nonsynonymous sites approaches the expected value of heterozygosity (H_e) without significant difference (P=0.91). Error bars indicate ± S.E.M * 3. **b**, Heterozygosity ratios between nonsynonymous and synonymous sites across the entire MAF spectrum. Expected values of heterozygosity (H_e) are equal between nonsynonymous and synonymous sites across each minor allele count bin, due to genetic diversity being directly associated with allelic frequency. In contrast, the ratios of observed heterozygotes between nonsynonymous and synonymous sites are substantially elevated along the increased MAF. **c**, Inbreeding coefficients (F_{IS}) of different types of SNPs across each MAF bin. The F_{IS} pattern of nonsynonymous sites is overall negative than any other type of SNPs, especially when MAF is increased. Error bars indicate ± S.E.M..



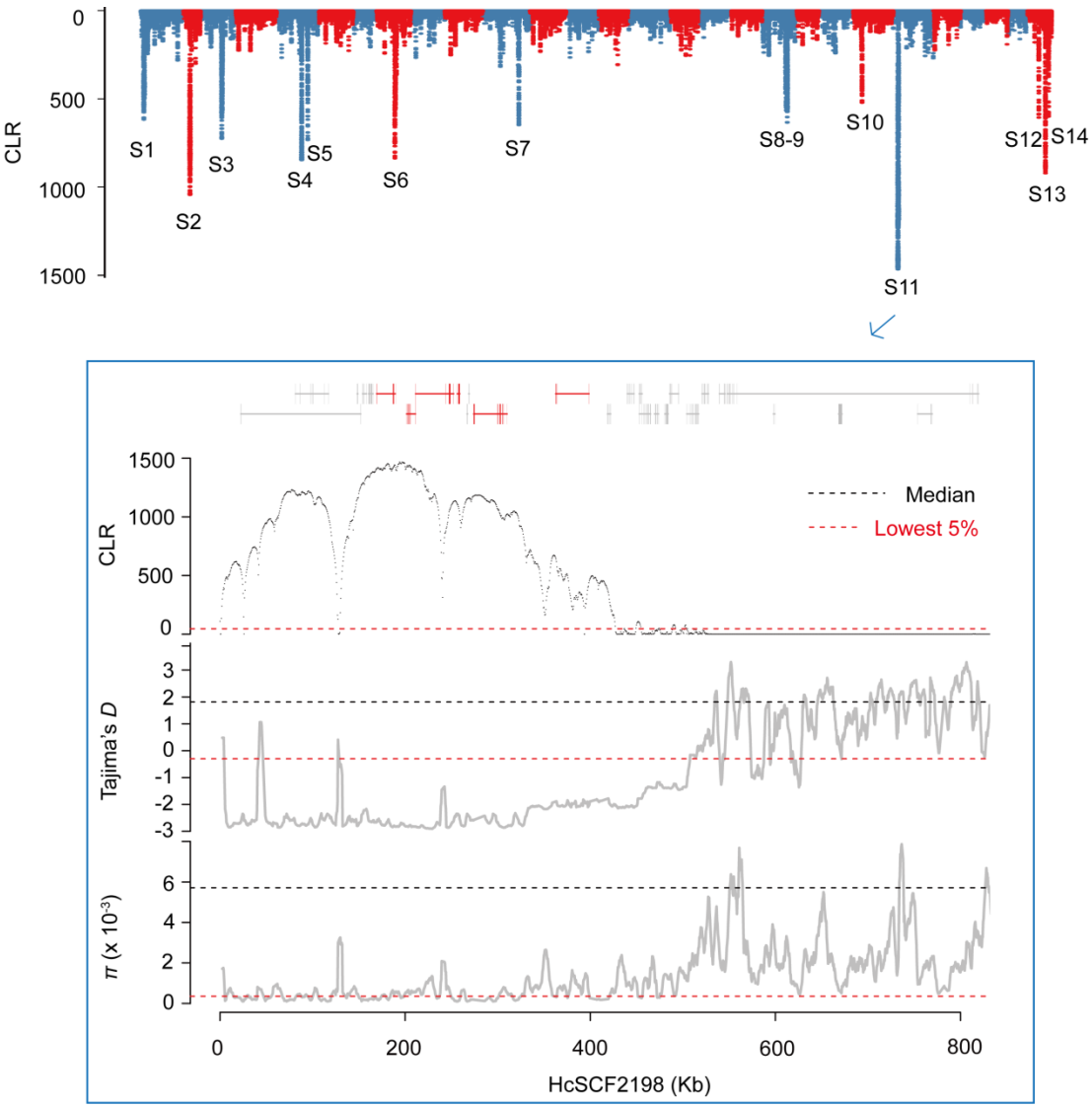
Supplementary Figure 8 Linkage disequilibrium-based signatures across genes of different ratios of π_N to π_S

Since the distribution of π_N / π_S across all genes was not even, we sorted all genes based on values of π_N / π_S and equally placed them into 16 continuous bins of increased π_N / π_S; each bin included 731 genes, which approached the gene number of $\pi_N > \pi_S$. Statistics were performed for each bin independently. **a**, Mean r^2 of all genes located in the corresponding π_N / π_S bin; **b**, mean length of spanned LD blocks; **c**, numbers of genes overlapped with a long haplotype block (>10Kb). We note that π_N is generally much lower than π_S for most genes, due to the fact that a much higher fraction of amino-acid-changing mutations is deleterious in neutral populations. In a population experiencing contraction, an increase in π_N relative to π_S can arise as a consequence of relaxed selection, matching with the observed valleys of above LD-based signatures (in grey shadows). Along the values of π_N / π_S continue rising, haplotype signatures were found recovered, suggesting that potential selection favoring diversity may occur on the corresponding genes.



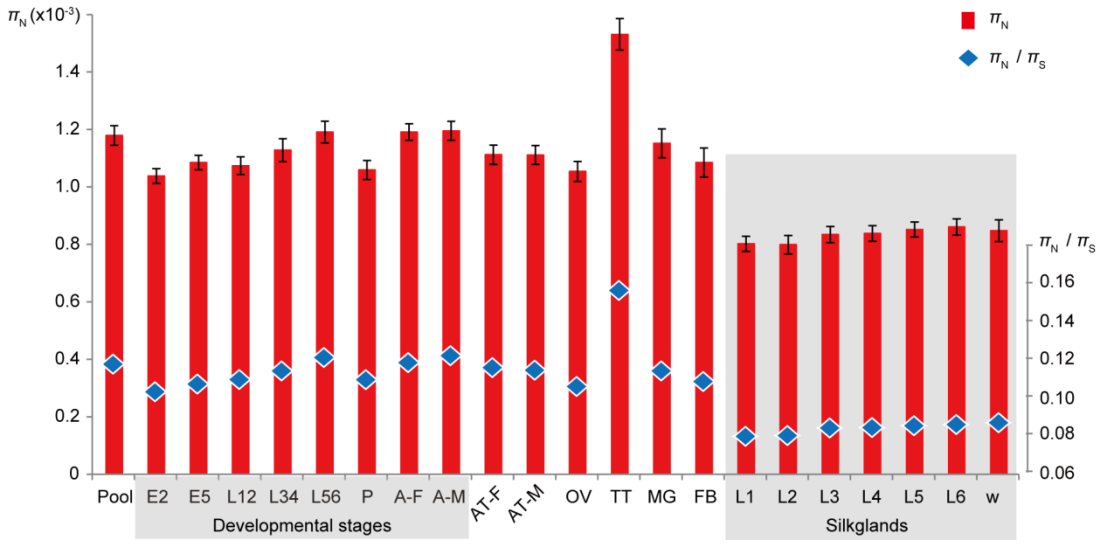
Supplementary Figure 9 Maximum-likelihood phylogenetic tree of olfactory receptor genes among three lepidopteran species

Unlike gustatory receptor genes (**Fig. 4**), we did not find evident expansion for most clades of olfactory receptor genes in the fall webworm.



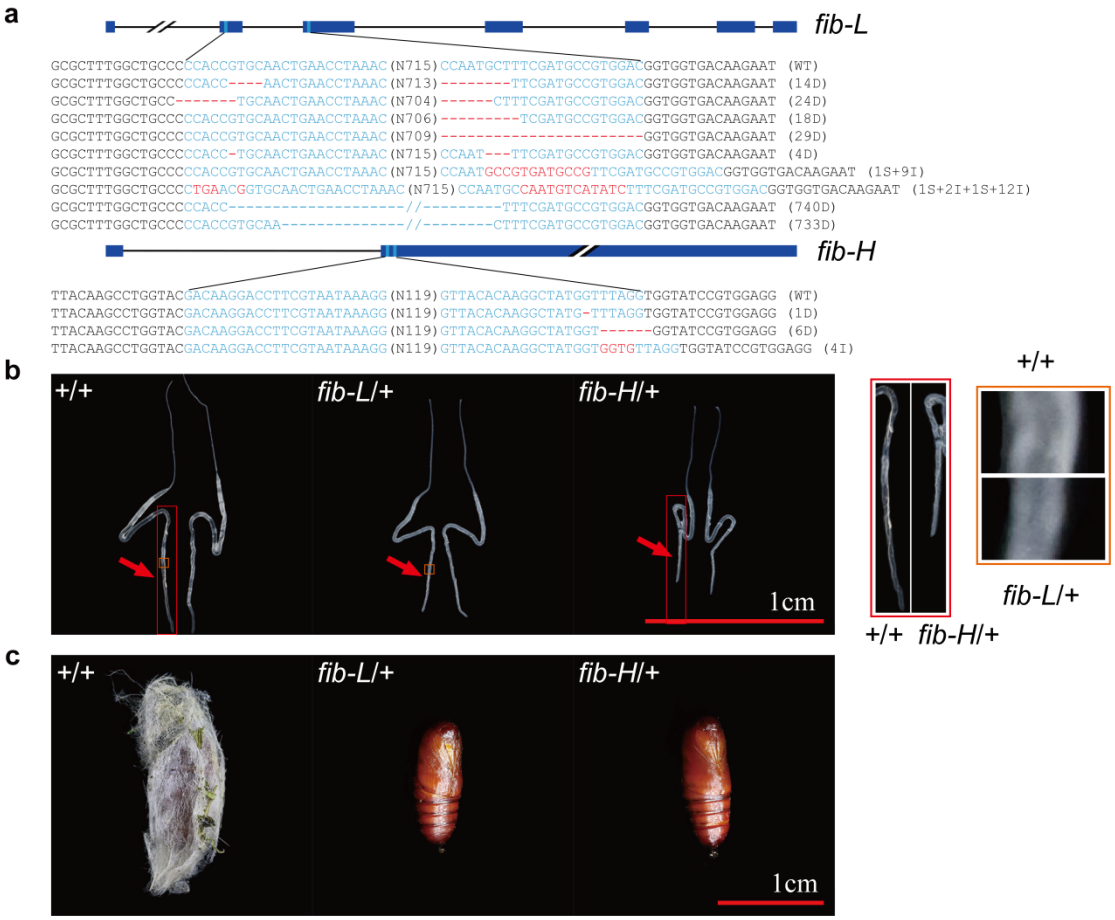
Supplementary Figure 10 Selective sweeps of highest CLR scores

Across the whole genome, 14 genomic windows presented the highest CLR signatures (>500). See **Supplementary Table 12** for details of all 14 sweeps. The sweep S11 which shows the highest CLR is presented with spanned gene models and additional polymorphism signatures (Tajima's *D* and π) as below. Organic cation transporter genes and genes containing the major facilitator superfamily domain (IPR020846) are shown in red.



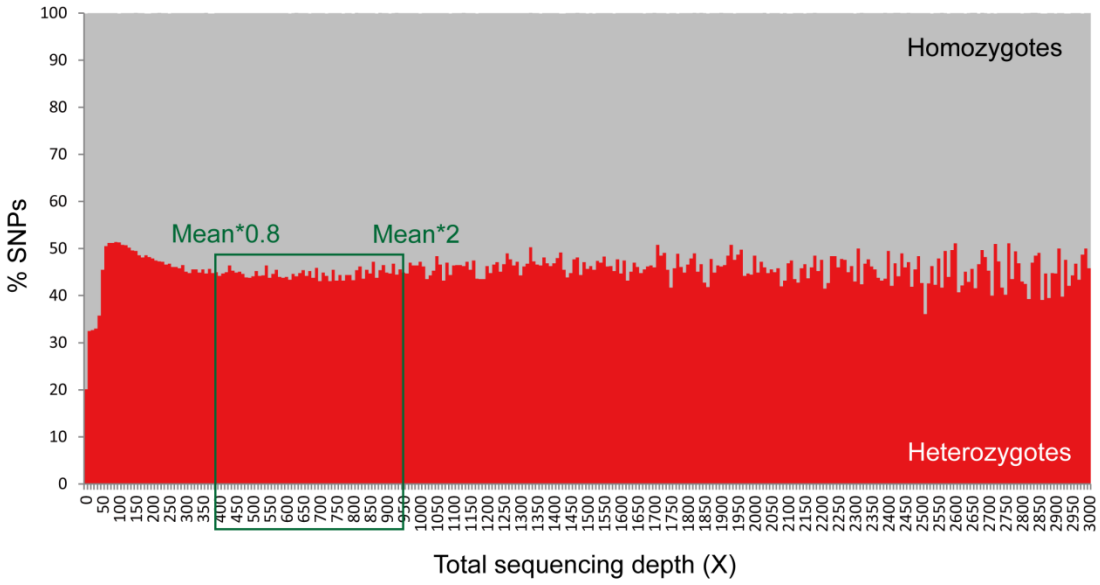
Supplementary Figure 11 Distribution of π_N and π_N/π_S within genes of reliable expression across different developmental stages and tissues

Samples of silkglands are indicated in the whole grey shadow, which presented an overall lower level of π_N/π_S (indicated by blue diamonds) than all other tissues. See the detailed distribution of π_N/π_S in **Supplementary Table 13**.



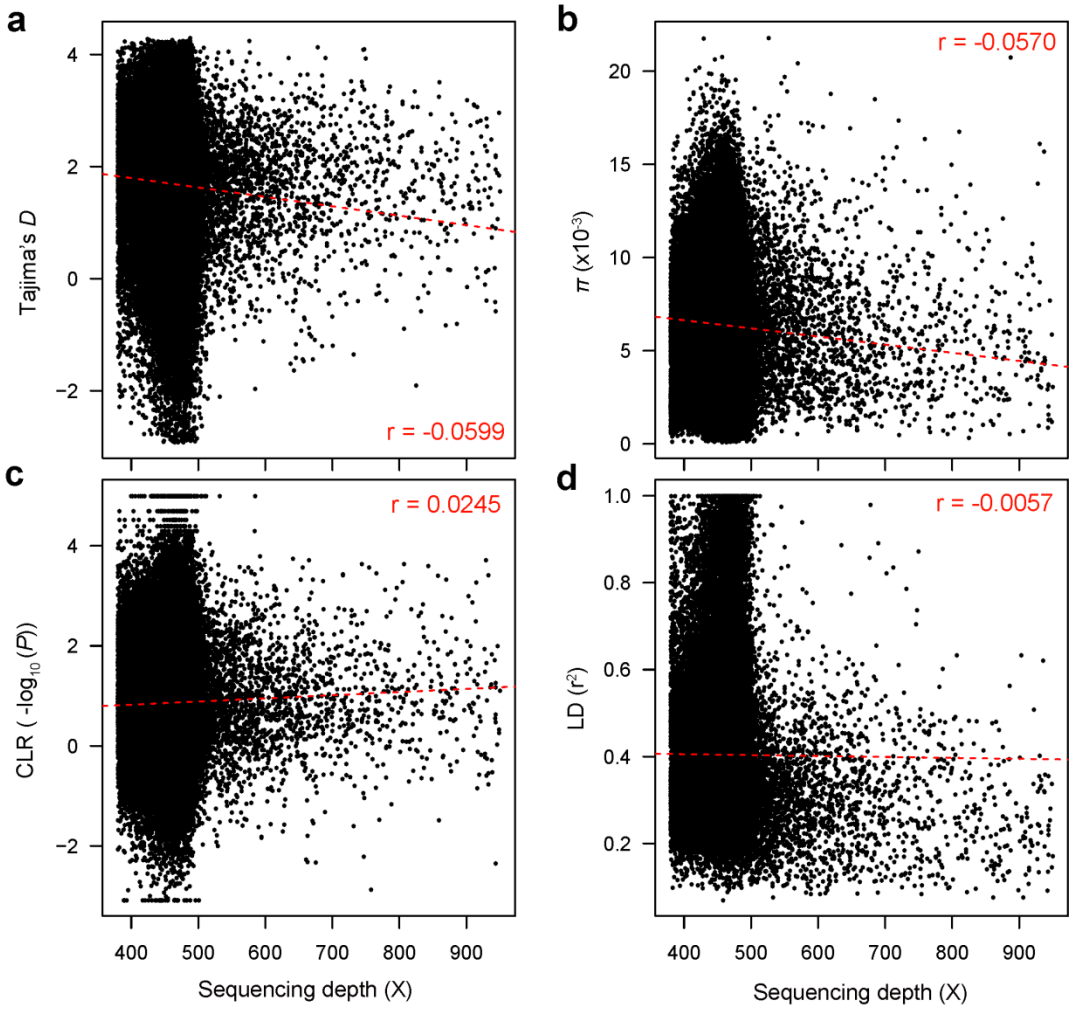
Supplementary Figure 12 Mutagenesis of fibroin related genes in *Hyphantria cunea*

a, Gene structure of *fib-L* (upper) and *fib-H* (lower) in *H. cunea* and associated mosaic genotypes. Like mutagenesis of *yellow*, the established CRISPR/Cas9 system in *H. cunea* resulted in a relatively high mutation rate in mutagenesis of *fib-L* and *fib-H*, including deletion, insertion, and substitution. **b-c**, Example of phenotypes in mosaic mutants of *fib-L*/+ and *fib-H*/+, respectively. Unlike visible phenotypic changes in *yellow* mutants, the mosaic mutants of fibroin related genes caused instable and subtle phenotypic changes. In the developmental stage in larvae, we found that the silk glands of a majority of mosaic mutants turn softer and more cracked in dissection. We also found the silk glands of some mutants develop in a relatively lower thickness (the middle panel in **b**) or length (the right panel in **b**) than wildtypes. In pupation, a higher proportion (~30-40%) of both *fib-L* and *fib-H* mosaic mutants failed to spin a cocoon, causing naked pupae (as shown in **c**). The abnormal proportion of pupation in wildtypes was ~10%. Although we failed to obtain homozygous mutants of fibroin related, we note that the phenotypes of such mutants in the silkworm have been reported as spinning of cocoons of much thinner shells[5].



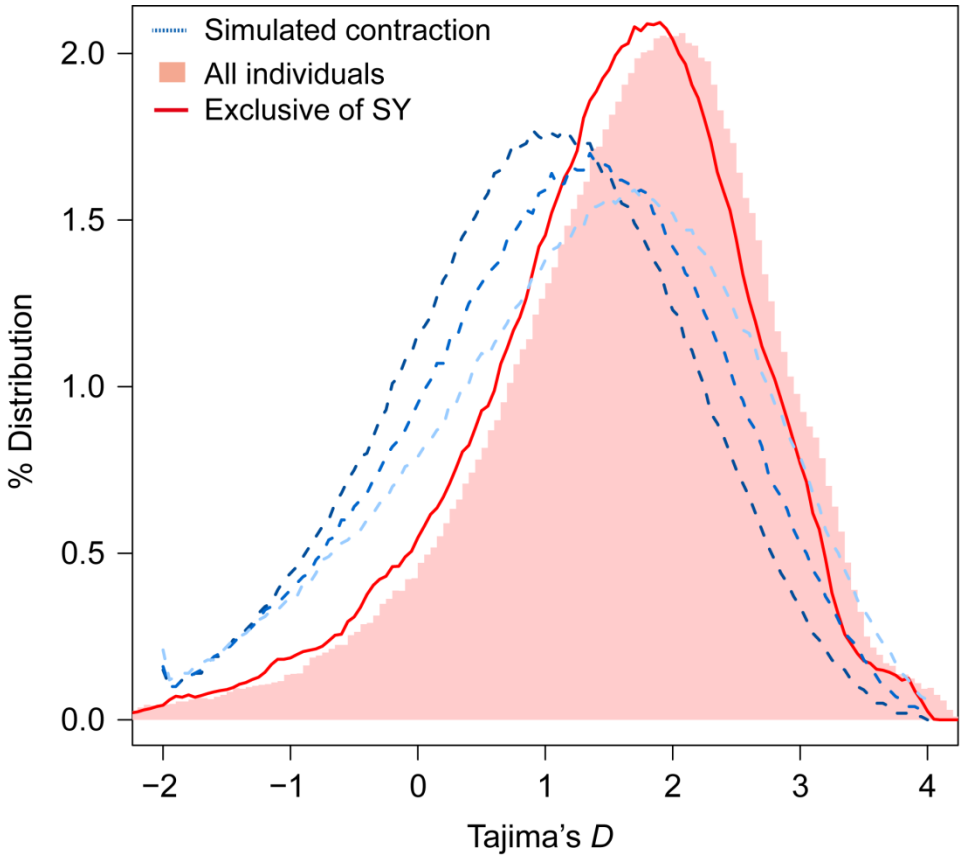
Supplementary Figure 13 Identified proportions of heterozygotes and homozygotes across different sequencing coverage intervals

Calculation of heterozygosity is probably affected by sequencing coverage. Low sequencing coverage may cause false negatives of heterozygous genotypes, which call for additional sequencing reads to present multiple alleles. On the other hand, copy number variations (duplications), reflecting by extremely high coverage, may cause false positives of heterozygotes by merging different loci. The green box outlines the regions that we used to estimate average level of heterozygosity for each type of polymorphism.



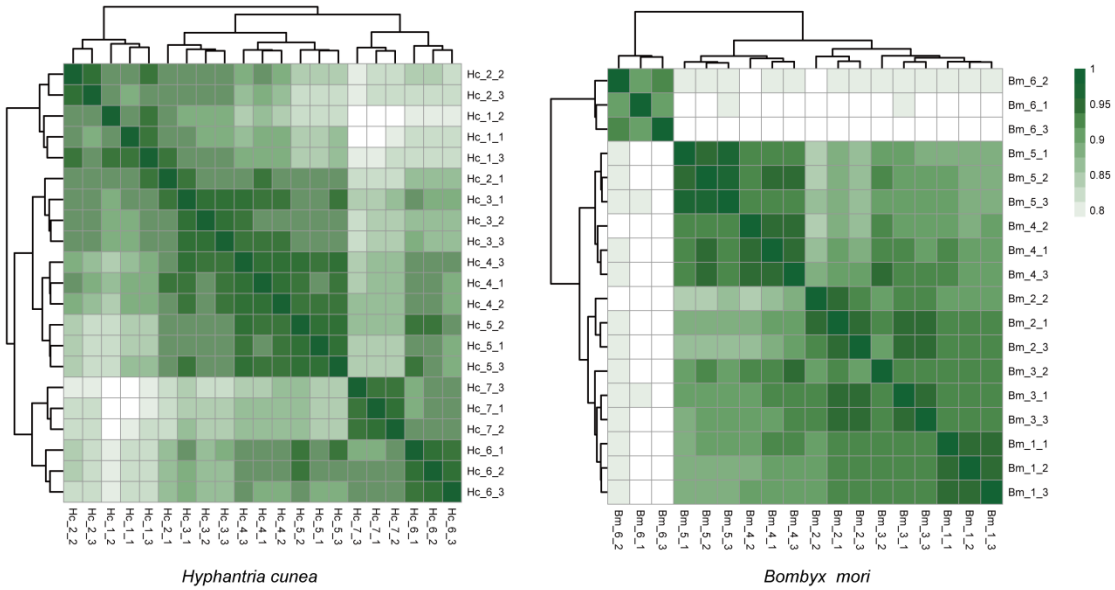
Supplementary Figure 14 Relationships between main signatures of population genetics and sequencing coverage

Abnormal sequencing coverage may cause bias to some kinds of population genetic analyses. We performed most population genetics analyses by limiting SNPs of normally average sequencing coverage (see **Supplementary Fig. 13**). For each 5-Kb genomic window, the population genetic signature was plotted against the mean sequencing depth. The presented correlated patterns (**a**, Tajima's *D*; **b**, π; **c**, CLR of SweepFinder; **d**, linkage disequilibrium) suggest that excess sequencing coverage did not cause systematic bias in this study.

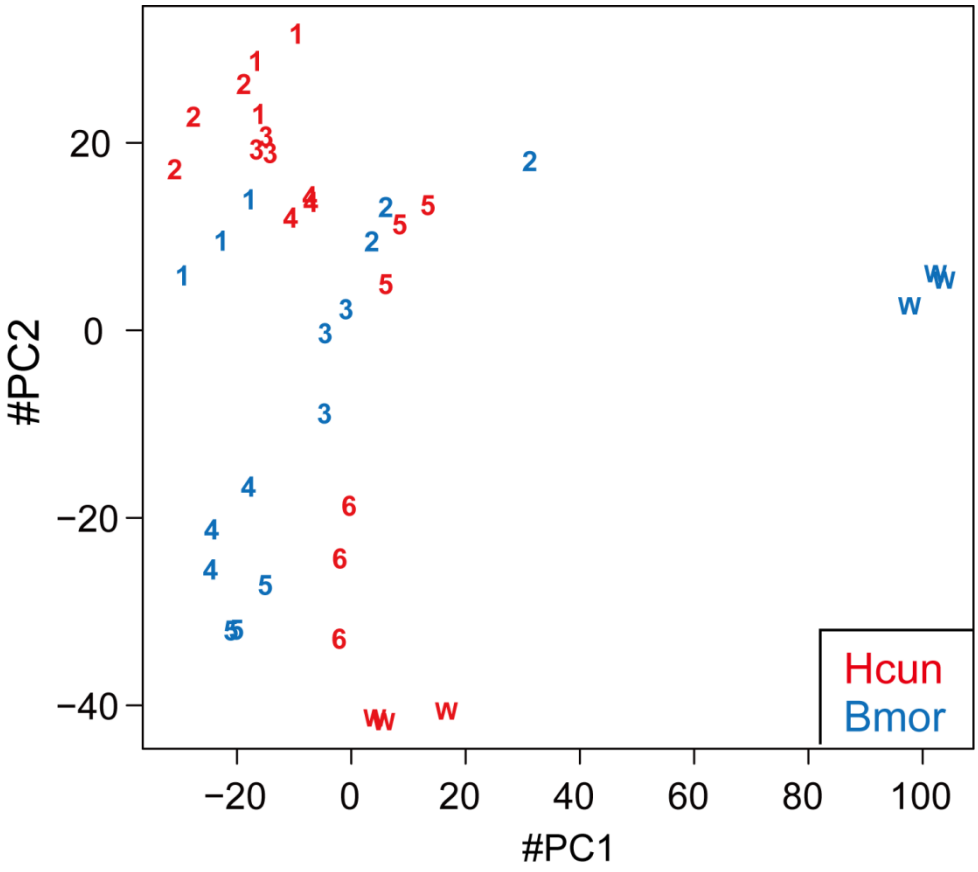


Supplementary Figure 15 Distribution of Tajima's *D* between the analyses using all samples and excluding Shenyang samples

Tajima's *D* is sensitive to effects of demographic processes. Both population contraction and subdivision may lead a positive Tajima's *D*. Since the population of Shenyang has been shown partially divergent with other populations, we tested the distribution difference between analyses with and without Shenyang samples. As a result, analysis excluding Shenyang samples was shown marginally less positive than the analysis based on all samples and substantially more positive than all three data sets we simulated for populations under recent contraction (identical to those contracted sets in **Fig. 2e**, suggesting that partial population subdivision contributes minor effect on the distribution pattern we discussed in this study.



Supplementary Figure 16 Correlation heatmaps across biological replicates of silkglands RNAseq samples



Supplementary Figure 17 Principle component analysis of independent silk gland samples based on their overall expression profiles

The first two eigenvectors were plotted. Identical numbers indicate three biological replicates of each larval instar. See the analysis based mean values of each stage in **Fig. 5e**.

Supplementary Table 1 Statistics of genomic sequencing data for *de novo* assembling

Sequencing library	Sequencing platform (Illumina)	Sequencing data (Gb)	Coverage (X)
Paired-end 500 bp	MiSeq 300 bp * 2	27.7	53.9
Paired-end 500 bp	HiSeq 150 bp * 2	27.6	53.8
Paired-end 800 bp	HiSeq 150 bp * 2	20.3	39.6
Mate pair 3 Kb	HiSeq 150 bp * 2	35.7	69.5
Mate pair 8 Kb	HiSeq 150 bp * 2	26.7	51.9
Mate pair 13 Kb	HiSeq 150 bp * 2	10.1	19.6
	Total	148.2	288.5

Supplementary Table 2 Assembly features of *H. cunea* and other lepidopteran moths

	H. cunea	*B. mori*[a]	*M. sexta*	*P. xylostella*	*O. brumata*
Assembled size (Mb)	513.8	480.8	419.4	394.1	638.2
# Scaffolds	4,402	43,622	20,871	1,819	25,801
N50 size (Kb)	1,128[b]	3,998	664	737	66
N90 size (Kb)	274	61	46	152	14
Max length (Mb)	6.73	16.17	3.25	3.49	0.50
CEGMA complete (%)	84.3	79.0	89.5	80.7	65.0
CEGMA partial (%)	98.8	97.2	98.8	93.6	94.0
BUSCO complete (%)[c]	98.1	95.7	96.7	87.7	94.7
BUSCO partial (%)	98.8	97.6	98.7	90.5	97.7

[a] The gene sets that were available by the end of 2016 were used for comparison. *B. mori*, *Bombyx mori*[6]; *M. sexta*, *Manduca sexta*[7]; *P. xylostella*, *Plutella xylostella*[8]; *O. brumata*, *Operophtera brumata*[9].

[b] Corrected N50 is 966 Kb for *H. cunea*, which refers to the N50 of the assembly after breaking the original assembly at breakpoints called by REAPR[10]. Other species were not processed with this process since REAPR requires raw sequencing reads of long-inserts which are not available for us.

[c] The insecta_odb9 data set was used to evaluate the quality.

Supplementary Table 3 Genomic features of *H. cunea* and other lepidopteran moths

	H. cunea	*B. mori* [a]	*M. sexta*	*P. xylostella*	*O. brumata*
% Repeat	35.6	43.6	25.8	34.0	53.5
% G+C	35.5	37.7	35.3	38.3	38.6
% Coding	4.51	3.72	5.04	6.35	2.85
# Protein-coding genes	15,799	14,623	15,451	18,073	16,912
Mean CDS length (bp)	1,467	1,224	1,367	1,385	1,074

[a] The gene sets that were available by the end of 2016 were used for comparison. *B. mori*, *Bombyx mori*[6]; *M. sexta*, *Manduca sexta*[7]; *P. xylostella*, *Plutella xylostella*[8]; *O. brumata*, *Operophtera brumata*[9].

Supplementary Table 4 Statistics of independent gene sets used for generating consensus set

Gene set		#Genes	Average CDS length	Average Gene length	#Mean Exon per gene	Average Exon length
	GLEAN	15,799	1,460	17,662	7	212
Ab initio	AUGUSTUS	14,368	1,596	14,012	7	219
	SNAP	23,929	900	15,060	7	131
	GeneMark	39,079	737	7,705	4	168
	Genscan	16,591	1,352	19,674	6	219
	GlimmerHMM	45,793	722	10,087	5	144
RNA-seq	Cufflinks	17,142	1,975	11,174	5	399
	Trinity	29,614	896	4,758	3	288
Homolog [a]	*B. mori*	17,417	978	5,058	4	230
	D. plexippus	16,282	1,157	7,969	5	220
	P. xylostella	16,937	983	6,595	4	228
	D. melanogaster	8,832	1,117	7,902	5	218
	Uniprot	15,172	1,209	8,438	5	227

[a] The gene sets that were available by June, 2016 were used for comparison. *B. mori*, *Bombyx mori*[6]; *D. plexippus*, *Danaus plexippus*[11]; *P. xylostella*, *Plutella xylostella*[8].

Supplementary Table 5 Statistics of transcriptome sequencing data in this study

	#Samples	Total data (Gb)	Mean (Gb)	Max (Gb)	Min (Gb)
Pooled	1	9.0			
Developmental stages	8	61.5	7.7	9.8	6.5
Representative tissues	6	42.5	7.1	9.1	6.0
Silkglands (*H. cunea*)	21	175.0	8.3	11.0	6.1
Silkglands (*B. mori*)	18	161.9	9.0	11.6	6.6
Total	54	449.9			

Supplementary Table 6 Annotation features of the *Hyphantria cunea* gene set

	# Annotated genes	% percentage of all genes
Species-specific genes	1,648	10.4
w/ UniRef terms	13,858	87.7
w/ RefSeq terms	13,932	88.2
w/ InterPro domains	10,728	67.9
w/ GO terms	7,955	50.4
w/ PFAM domains	10,192	64.5
w/ KO terms	8,815	55.8
w/ mapped RNAseq reads	15,687	99.3

Supplementary Table 7 Quality control of lepidopteran gene set

	H. cunea	*B. mori* [a]	*M.sexta*	*P. xylostella*	*O. brumata*
CEGMA complete (%)[b]	85.8	75.3	85.6	73.4	56.1
CEGMA partial (%)	98.5	97.4	98.5	92.6	97.6
BUSCO complete (%)[c]	95.0	89.9	93.7	74.9	75.9
BUSCO partial (%)	98.7	96.7	97.0	80.5	90.9

[a] The gene sets that were available by the end of 2016 were used for comparison. *B. mori*, *Bombyx mori*[6]; *M. sexta*, *Manduca sexta*[7]; *P. xylostella*, *Plutella xylostella*[8]; *O. brumata*, *Operophtera brumata*[9].

[b] Query proteins were recovered by genblastA with coverage of at least 85%.

[c] The insecta_odb9 data set was used to evaluate the quality.

Supplementary Table 8 Selective IPR domains with specific expansion in the fall webworm

Rank[a]	IPR_ID	C[b]	Annotation	**H**[c]	B	M	O	P
4	IPR006326	S	UDP-glycosyltransferase, MGT	**3**	1	0	0	1
7	IPR010585	D	DNA repair protein XRCC4	**3**	1	0	0	0
8	IPR014751	D	DNA repair protein XRCC4, C-term.	**3**	1	1	0	0
10	IPR018325	D	Rad4/PNGase transglutaminase-like	**3**	1	1	1	1
11	IPR029158	S	Stimulator of interferon genes protein	**3**	1	1	0	0
13	IPR007808	R	Transcription elongation factor 1	**3**	0	1	0	0
20	IPR005874	R	Eukaryotic translation initiation factor SUI1	**2**	1	1	1	1
21	IPR006600	R	HTH CenpB-type DNA-binding domain	**18**	8	9	9	1
30	IPR017943	S	Bactericidal permeability-increasing protein, alpha/beta domain	**2**	1	1	1	1
31	IPR018326	D	Rad4 beta-hairpin domain 1	**2**	1	1	0	1
32	IPR018327	D	Rad4 beta-hairpin domain 2	**2**	1	1	0	1
33	IPR018328	D	Rad4 beta-hairpin domain 3	**2**	1	1	1	1
45	IPR001916	S	Glycoside hydrolase, family 22	**14**	3	6	3	3
52	IPR001557	S	L-lactate/malate dehydrogenase	**8**	5	4	3	3
53	IPR001236	S	Lactate/malate dehydrogenase, N-terminal	**11**	7	7	7	6
55	IPR000197	R	Zinc finger, TAZ-type	**3**	1	1	2	1
57	IPR002515	R	Zinc finger, C2HC-type	**3**	2	2	1	2
61	IPR009089	D	DNA double-strand break repair and VJ recombination XRCC4, N-terminal	**3**	1	2	0	0
65	IPR009318		Gustatory receptor*	**10**	2	5	5	4
67	IPR022383	S	Lactate/malate dehydrogenase, C-terminal	**11**	8	6	5	5
71	IPR000111	S	Glycoside hydrolase, clan GH-D	**4**	3	3	2	3
79	IPR006619	S	Peptidoglycan recognition protein family domain, metazoa/bacteria	**17**	8	13	4	9
95	IPR015955	S	Lactate dehydrogenase/glycoside hydrolase, family 4, C-terminal	**13**	8	8	9	10
99	IPR007577	S	Glycosyltransferase, DXD sugar-binding motif	**7**	1	0	6	0
116	IPR008979	S	Galactose-binding domain-like	**35**	30	33	27	33

[a] Out of 143 IPR domains of specific expansions in the Fall webworm, 19 IPR domains are related to regulation (R), DNA repair (D), sugar metabolism (S), and gustatory reception(*; IPR009318), which we discussed in the main text. Actual rank in the entire IPR list of webworm-specific expansion is listed.

[b] Potential functional categories: R, regulation; D, DNA repair; S, sugar metabolism.

[c] H, *Hyphantria cunea*; B, *Bombyx mori*[6]; M, *Manduca sexta*[7]; O, *Operophtera brumata*[9]; P, *Plutella xylostella*[8].

Supplementary Table 10 Enriched pathways in genes with higher frequency of nonsynonymous polymorphisms

KEGG ID	Pathway	#Genes in the list	#Genes in the genome	*P* value	FDR corrected *P* value
ko00524	Butirosin and neomycin biosynthesis [a]	7	17	1.26E-08	1.53E-06
ko04930	Type II diabetes mellitus [b]	7	39	7.08E-06	0.00043
ko00051	Fructose and mannose metabolism	8	56	9.04E-06	0.00036
ko00010	Glycolysis / Gluconeogenesis	9	84	2.69E-05	0.00081
ko05230	Central carbon metabolism in cancer [b]	8	65	2.79E-05	0.00067
ko04066	HIF-1 signaling pathway	9	94	6.63E-05	0.00134
ko00052	Galactose metabolism	8	82	2.58E-03	0.00260
ko04973	Carbohydrate digestion and absorption	7	65	3.15E-03	0.00322
ko01200	Carbon metabolism	11	162	3.15E-03	0.00318
ko00520	Amino sugar and nucl. sugar metabolism	8	94	4.67E-03	0.00470
ko00500	Starch and sucrose metabolism	7	85	1.21E-03	0.01217

[a] Most genes involved in this pathway were found with best BLASTP hit of a piggyBac element derived from Hemiptera. Given the potential contamination of these genes, we removed this pathway for further analysis.

[b] Pathways related with human disease were removed from consideration; we, however, note that these pathways are also related to carbohydrate metabolism.

Supplementary Table 11 Manual annotations of chemoreception and detoxification related genes

	Chemoreception					Detoxification				
	OR	GR	IR	OBP	CSP	P450	UGT	CCE	GST	ABC
H. cunea	47	147	26	36	22	72	15	76	19	49
B. mori[a]	64	65	18	44	20	80	45	69	23	49
M. sexta	63	99	31	47	21	97	38	86	17	47
P. xylostella	81	47	35	33	32	74	18	55	11	97
D. melanogaster	60	60	66	52	4	85	34	34	38	56

[a] The gene count of these families were independently determined by a consensus approach that was used in analysis of *H. cunea* (see **Methods**). The genomes that were available by the end of 2016 were used for comparison. *B. mori*, *Bombyx mori*[6]; *M. sexta*, *Manduca sexta*[7]; *P. xylostella*, *Plutella xylostella*[8]. Note that, the genome of *Operophtera brumata*[9] was also available by then, but it was not included in the comparison due to the incomplete status of the released assembly.

Supplementary Table 16 Datasets used in orthology analysis of this study

Species	UTR downloaded
Drosophila melanogaster[12]	flybase.org
Bombyx mori[6]	silkworm.genomics.org.cn
Manduca sexta[7]	i5k.nal.usda.gov/data/Arthropoda/mansex-(Manduca_sexta)
Plutella xylostella[8]	iae.fafu.edu.cn/DBM
Operophtera brumata[9]	ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001266575.1_ASM126657v1
Danaus plexippus[11]	monarchbase.umassmed.edu/
Heliconius melpomene[13]	butterflygenome.org/
Papilio xuthus[14]	papilio.bio.titech.ac.jp/papilio.html
Melitaea cinxia[15]	www.helsinki.fi/science/metapop/research/mcgenome.html
Lerema accius[16]	prodata.swmed.edu/LepDB/

Supplementary Table 18 Sequencing data of other Lepidoptera species native to China

Species	Sampling site	Sex	Data (Gb)	Used in this study
A. ipsilon	Beijing (39.92°N 116.43°E)	M	21.9	K-mer
E. grisecens	Hangzhou (30.37°N 120.17°E)	F	24.9	K-mer[a]
O. furnacalis	Handan (36.60°N 114.48°E)	M	21.2	K-mer
B. mandarina (downloaded from NCBI#SRP119041)[17]				
W_SAX2	Xi'an (34.16°N 108.54°E)	M	26.1	K-mer
Cwild	Shanghai (31.22°N 121.47°E)	n/a	30.2	SMC++/MAF/Tajima
C1wild	Lanzhou (36.05°N 103.67°E)	n/a	29.2	LD/SMC++/MAF/Tajima
C2wild	Kunming (25.06°N 102.68°E)	n/a	29.3	LD/SMC++/MAF/Tajima
C4wild	Xi'an (34.16°N 108.54°E)	n/a	18.1	LD/SMC++/MAF/Tajima
C5wild	Hangzhou (30.37°N 120.17°E)	n/a	25.3	LD/SMC++/MAF/Tajima
C6wild	Chengdu (30.65°N 104.05°E)	n/a	32.8	LD/SMC++/MAF/Tajima

[a] Male wild-caught samples of *E. grisescens* were unavailable for us when the project was in process, a female individual under ~1 year laboratory crossing was used. We note that the genomic heterozygosity in *E. grisescens* is partially overestimated due to the existence of unpaired sex chromosomes and also partially underestimated due to a short period of laboratory crossing.

Reference

1. Sullivan, G.T., Karaca, I., Ozman-Sullivan, S.K. & Kara, K. Tachinid (Diptera: Tachinidae) parasitoids of overwintered hyphantria cunea (Drury) (Lepidoptera: Arctiidae) pupae in hazelnut plantations in Samsun province, Turkey. *Journal of the Entomological Research Society* **14**, 21-30 (2012).
2. Gomi, T., Muraji, M. & Takeda, M. Mitochondrial DNA analysis of the introduced fall webworm, showing its shift in life cycle in Japan. *Entomological Science* **7**, 183-188 (2004).
3. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-25 (2014).
4. Keightley, P.D. et al. Estimation of the spontaneous mutation rate in Heliconius melpomene. *Mol Biol Evol* **32**, 239-43 (2015).
5. Inoue, S. *et al.* A fibroin secretion-deficient silkworm mutant, Nd-sD, provides an efficient system for producing recombinant proteins. *Insect Biochem Mol Biol* **35**, 51-9 (2005).
6. International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm Bombyx mori. *Insect Biochem Mol Biol* **38**, 1036-45 (2008).
7. Kanost, M.R. *et al.* Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, Manduca sexta. *Insect Biochem Mol Biol* **76**, 118-147 (2016).
8. You, M. *et al.* A heterozygous moth genome provides insights into herbivory and detoxification. *Nat Genet* **45**, 220-5 (2013).
9. Derks, M.F. *et al.* The Genome of Winter Moth (Operophtera brumata) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. *Genome Biol Evol* **7**, 2321-32 (2015).
10. Hunt, M. *et al.* REAPR: a universal tool for genome assembly evaluation. *Genome Biol* **14**, R47 (2013).
11. Zhan, S., Merlin, C., Boore, J.L. & Reppert, S.M. The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**, 1171-85 (2011).
12. FlyBase, C. FlyBase: a Drosophila database. *Nucleic Acids Res* **26**, 85-8 (1998).
13. The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94-8 (2012).
14. Nishikawa, H. *et al.* A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nat Genet* **47**, 405-9 (2015).
15. Ahola, V. *et al.* The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat Commun* **5**, 4737 (2014).
16. Cong, Q., Borek, D., Otwinowski, Z. & Grishin, N.V. Skipper genome sheds light on unique phenotypic traits and phylogeny. *BMC Genomics* **16**, 639 (2015).
17. Xiang, H. *et al.* The evolutionary road from wild moth to domestic silkworm. *Nat Ecol Evol* (2018).