Title

Characterization of Z chromosomal inversion and lineage-specific sweeps reveal widespread local adaptation in *Helicoverpa armigera*

Fushi Ke[a]

[a] School of Biological Sciences, The University of Hong Kong, Pok Fu Lam Road, Hong Kong, China.

Corresponding author

Fushi Ke: fjkfs@163.com

Abstract

Efficient pest management requires comprehensive knowledge of its biology, ecology, and evolution, particularly adaptation that exacerbating pest damage. *Helicoverpa armigera* (Hübner) is a notorious pest that attacks more than 200 species, adapts to diverse environments, and has developed resistance to almost all applied insecticides. However, local adaptation of this species was less investigated. Here, we analyzed genomic variation of *Helicoverpa armigera* in China and Oceania to identify adaptative loci in different lineages. Despite intensive gene flow, two differentiated Z chromosomal lineages in Central China (i.e., northern and southern lineages) were formed by chromosome inversion and local adaptation. Based on population genomic analysis, we identified genes related to circadian clock (*period*, *cycle*, *cyr2* and *daywake*), insulin secretion (*IGF2BP2*) and energy metabolism (*SLC25A25*, *CCG7766*, *D11DS*, *TSTP* and *CPQ*) were under selection in northern lineage. These genes may have contributed to its adaptation to high latitudes. In the southern lineage however, the Z chromosome inversion harboring alien alleles of two clock genes (*period* and *cycle*) and may have disrupted clock oscillator for adaptation. We further conducted selective sweep analysis in lineages from Northwestern China, Central China, and Oceania, and found several insecticide resistant genes that shared among different lineages were likely contributed by introgression. Nevertheless, many candidates, including a cadherin that involved in resistance to *Bacillus thuringiensis* protein in Oceanian populations, are lineage-specific.

Our results highlight the importance of investigating local adaptation in effective pest control under globalization of agriculture.

Key words

Cotton bollworm; local adaptation; clock genes; energy metabolism; Z chromosome inversion; lineage-specific sweep

Introduction

Agricultural insect pests cause substantial losses of crop production each year and pose a threat to global food security (Paini et al., 2016). It's difficult to implement effective pest control strategies without a thorough understanding of pest biology, ecology and evolution, particularly the adaptive evolution that exacerbates the pest damage (Roderick & Navajas, 2003; Renou, 2022). Investigating the ability and genetic basis of pests adapting to local habitats, in terms of adaptation to environments and pesticides thus, would benefit successful pest control. For most insect pests, environmental factors play a crucial role in their performance. Identifying genotypes of higher fitness and selective agents causing environmental adaptation could provide important information for their interaction with local environments, prediction of their further distribution (Pélissié et al., 2018), and implementation of efficient molecular techniques (e.g., through gene editing) for pest control. In the case of resistance to applied pesticides and toxins expressed by transgenic crops, local adaptation could imply that control strategies might need to be tuned according to which populations and regions that are targeted in to keep their efficiency (Alyokhin et al., 2015; Pélissié et al., 2018; Szendrei et al., 2012).

For insect species, forecasting seasonal changes and responding to them in an appropriate way are species' specific requirements to live in high latitudes or altitudes (Hoikkala & Poikela, 2022). Since temperature and photoperiod are generally reliable indicators of seasonality that have shaped life histories of many temperate climate organisms (Fält-Nardmann et al., 2016), efficient adaptations to thermal stress and photoperiod such as diapause ensure survival and succession of local populations during long overwintering periods (Toepfer et al., 2014). Several pathways have been identified to be underlying the genetic and physiological mechanisms of diapause regulation and evolution (Ragland et al., 2019). Associated genes involved in the circadian clock with diapause were uncovered, with examples showing strong connections between seasonal adaptations and adaptative alleles of clock genes (Ragland et al., 2019; Kozak et al., 2019; Hasebe & Shiga, 2022; Peng et al., 2023). In addition, pathways that insulin involved (Williams et al. 2006; Paaby et al., 2010) and energy metabolism related could also facilitate adaptation of insects to high latitude habitats (Hahn & Denlinger, 2011). Identification of adaptative genes/alleles in the lineages of high

latitude thus could give us a comprehensive knowledge of adaptative evolution (e.g., diapause) that could be employed in successful management of insect pests, especially those from temperate regions.

Recombination of genotypes from different genetic background would generate new genetic combinations that facilitate local adaptation of descendants (Ortiz-Barrientos et al., 2016). It could also increase interchange of maladaptative alleles that restrict adaptation of local lineages (Crispo et al., 2011; Arnold et al., 2017). Structural variations such as chromosome inversion that inhibit recombination would restrict gene reshuffling in heterozygotes and maintain adaptative or maladaptative genotypes with linkage (Kirkpatrick & Barton 2006; Roesti et al., 2022). Meanwhile, biotic and abiotic stress that mediated local adaptation would be barrier of gene flow in adaptative loci during homogenization of global genetic variation in differentiated lineages (Tigano et al., 2016). Therefore, investigating the interplay of gene flow, local adaptation and chromosome structural variation would benefit identification of adaptative genes in local populations.

The cotton bollworm, *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae), is a worldwide crop pest that attacks over 200 species (Pratissoli et al., 2015; Riaz et al., 2021) and shows its capability of adaptation to diverse habitats and hosts. It has also developed resistance to many applied insecticides with a high number of reported cases of insecticide resistance worldwide (Klai et al., 2020). Meanwhile, it could migrate over 2,000 km and mediated intensive gene flow among different populations (Nibouche et al., 1998; Jones et al., 2015; Riaz et al., 2021). Recently, this pest has invaded to South America, hybridized with its sibling species and introduced insecticide-resistant genes into local colonies (Valencia-Montoya et al., 2020). Diapause is an important strategy by which insects, such as *H. armigera*, avoid unfavorable environmental conditions (Danks, 2000). Successful range expansion, particularly at high latitudes, thus requires that species should be able to coordinate their phenology and synchronize life cycles using local temperature and/or day length as cues (Fält-Nardmann et al., 2016; Saikkonen et al., 2012). Investigating phenological adaptation of source populations to environmental cues would help in predicting sites of their invasion and implementing effective pest control. While difference in response to temperature and day length in *H. armigera* lineages have been explored based on molecular experiments (Wu et

al., 1997; Wu & Guo 1997; Wu & Guo 2007; Chen et al., 2013; Chen et al., 2016; Virachack et al., 2018), less analysis was conducted to investigate the genetic basis of its acclimation (but see Jin et al., 2023) in the presence of widespread gene flow among geographically distinct lineages.

In this study, we revisited resequencing data of *H. armigera* samples from Asia and Oceania (Zhang et al., 2022), and investigated adaptative genes that have contributed to local adaptation of different lineages. When constructing phylogeny based on Z chromosome polymorphism, we found two differentiated lineages of *H. armigera* in Central China (i.e., the southern (CC-S) and northern (CC-N) lineages) despite widespread gene flow across this region. We further verified intensive introgression between *H. a. conferta* in Oceania and *H. armigera* in Central China, while gene flow of adaptative loci into CC-N was restricted due to local adaptation, representing a model of divergent selection with gene flow (Pinho & Hey, 2010). In addition, our results support Z chromosome inversion harboring *H. a. conferta* alleles of two linked clock genes may have disrupted acclimation in CC-S, and contributed to differential gene flow of *H. a. conferta* into northern and southern *H. armigera* lineages in Central China. Finally, we identify several lineage-specific signals of selection that may be related to insecticide resistance of local populations. We further discuss our findings under current global climate change scenario and severe pest damage to agricultural industry.

113 # Materials & Methods

114 **Polymorphic sites and genome resequencing data**

115 We downloaded polymorphic SNPs from figshare (https://figshare.com/articles/dataset/population
116 _genomics_of_cotton_bollworm/17008922). This dataset includes 141 *Helicoverpa armigera*
117 individuals from China, 19 individual *H. a. conferta* from Australia, and nine *H. zea* from North
118 and South America. Samples of China collected from sites at Yellow River Region (YRR),
119 Changjiang River Region (CRR), and Northwestern Region (NR) during 2016-2018. Genomic
120 analysis shows high similarity between samples from YRR and CRR (Zhang et al., 2022), we thus
121 reassigned their geographical origin as Central China (CC). A total of 100 individuals and 41
122 individuals from CC and NR regions were employed, respectively. Among them, 47, 33 and 10
123 male individuals with ZZ genotype of sex chromosome, were found in CC, NR *H. armigera* lineages,
124 and *H. a. conferta*, respectively. See Zhang et al., (2022) for detailed information of these samples.
125

126 To further calculate population genetic parameters, such as d_{xy}, we downloaded sequencing reads
127 of 141 *H. armigera* individuals (BioProject No. PRJNA731848) and six *H. zea* individuals NCBI
128 (SRR15526515-20), and called SNPs by using Genome Analysis Toolkit 3.8 (GATK) (DePristo et
129 al. 2011). Sequencing reads of each individual were filtered by using fastq
130 (https://github.com/OpenGene/fastp) and mapped to HaSCD2 assembly (Zhang et al., 2022) using
131 BWA (Li, 2013). The standard GATK pipeline (DePristo et al. 2011) was employed to generate raw
132 sites from bam files. We further conducted filterings in GATK with default settings to exclude low-
133 quality sites, and removed loci with mean coverage depth lower than 3 or higher than 30, or with
134 missing rate higher than 0.5 using VCFtools (Danecek et al. 2011). Unless stated, this dataset with
135 high-quality non-variants was only used for π and d_{xy} calculation.

136

137 **Phylogeny tree and genetic structure**

138 Phylogeny trees were reconstructed using three methods: neighbour-joining (NJ) method, maximum
139 likelihood (ML) method based on concatenated SNPs, and species tree inferred from windowed
140 trees based on ASTRAL-III (Zhang et al., 2018). For NJ phylogeny, genetic distance matrix of each
141 individual was calculated using VCF2Dis (https://github.com/hewm2008/VCF2Dis) and used for

142 phylogeny construction in phylip-3.697 (http://evolution.genetics.washington.edu/phylip). We
143 further reconstructed the phylogeny relationship based on maximum likelihood (ML) method
144 implemented in IQ-TREE-2.1.3 (Minh et al. 2020). Bi-allelic SNPs thinned by an interval of 100bp
145 were concatenated and used in ML tree construction. We employed +ASC model in model testing
146 to include ascertainment bias correction as SNP data do not contain constant sites. Following the
147 manual, we also appended -bnni to the regular UFBoot command to reduce the risk of
148 overestimating branch support (Hoang et al. 2018). The topologies inferred by NJ and ML methods
149 based on autosomal polymorphism were similar, while topology difference was found when using
150 polymorphic data of Z chromosome. To be specific, we found two differentiated lineages (CC-S,
151 southern lineage, and CC-N, northern lineage) in Central China based on ML tree. Thus, we further
152 constructed the individual phylogeny based on ASTRAL-III using Z chromosomal SNPs. We first
153 divided Z chromosome into 100-kb windows according to coordinates and constructed window tree
154 with SNPs of each window using IQTREE (Minh et al. 2020). A summary tree was then generated
155 based on the constructed 179 window trees in ASTRAL-III (Zhang et al., 2018). Five regions in Z
156 chromosome that were divided by Fst values were further used for tree construction in IQ-TREE.
157 We further investigated population genetic structure of *H. armigera* (including 19 *H. a.*
158 *conferta* individuals) based on autosomal SNPs. We generated SNP datasets by pruning SNPs based
159 on LD value of 0.2 with an interval of 50 loci and included only SNP loci at a missing rate ≤ 20%
160 within the intergenic regions 20kb away from the genic regions. We used ADMIXTURE for genetic
161 structure analyses and performed each K from 1 to 20 with 20 replicates. The best K value was
162 identified according to cross-entropy criterion. Employing the same dataset, we conducted Principal
163 Component Analysis (PCA) based on PLINK2 (Chang et al., 2015).
164
165 **Inversion identification and genome-wide association analysis**
166 We used Asaph (Nowling et al., 2022) to detect chromosomal inversion in samples collected from
167 Central China as we found two differentiated lineages in phylogeny based on Z chromosomal
168 polymorphims. Asaph could efficiently detect inversion when samples are all from a single
169 population and SNPs are all from a single chromosome (Nowling et al., 2022). Two first PC
170 coordinates of Z chromosomal polymorphism were generated to separate samples in Central China.
171 Further, single-SNP association tests using genotypes of each SNP against the samples' first PC

172 coordinate were performed. Without long sequencing reads to identify the accurate boundary of the
173 inversion, we set an arbitrary boundary of the inversion by picking SNPs with a $-\log_{10}(P) > 15$.
174 Within the inversion, haplotypes possessing derived alleles in over half of the differentiated SNPs,
175 we treated them as haplogroups of inversion. To further identify loci that may have contributed to
176 differentiation of two lineages in China, the GWAS analysis was extended to polymorphic SNPs in
177 autosomal SNPs of each chromosome. We set an arbitrary while widely used *P* value (5×10^{-8}) as
178 the threshold, and selected SNPs with a *P* value lower than the threshold as outliers.
179
180 **Calculation of population genetic parameters**
181 We calculated nucleotide diversity (π), genetic differentiation (F_{ST}) and genetic divergence (d_{xy}) by
182 using Simon Martin's script (https://github.com/simonhmartin/genomics_general). In addition, we
183 calculated window-based parameter f_{dM} and *D* based on Dsuite (Malinsky et al., 2021) to investigate
184 genome-wide introgression between different lineages. Only male individuals were included for
185 analysis. Note that we have regenerated a genotype matrix including non-variant sites for accurate
186 calculation of π and d_{xy}. We further calculated r^2 of SNP pairs in Z chromosome based on
187 PopLDdecay (https://github.com/BGI-shenzhen/PopLDdecay) and compared linkage
188 disequibrilum (LD) between lineages including *H. zea*, CC-N *H. armigera* and CC-S *H. armigera.*
189 To account for different sampling size, we randomly sampled eight individuals in CC-N *H.*
190 *armigera* and *H. zea* ten times and used these replicates for further comparisons. In addition, to
191 compared LD of different regions, we further divided Z chromosome into "inversion" (identified by
192 Asaph) and "the other" regions, and calculated r^2 of SNP pairs in each region.
193
194 **Selective sweep analysis**
195 To identify genetic loci underlying local adaptation in each geographically distinct lineage, selective
196 sweep analysis was performed using site frequency spectrum (SFS) in SWEEPFINDER2
197 (DeGiorgio et al., 2016). We first assigned *H. zea* as the outgroup and calculated the derived allele
198 frequency spectrum in each population. We further ran selective sweep analysis in each
199 chromosome in a step of 1000 nucleotides. Regions with a CLR value at the top 0.05% of all
200 autosomal regions in each population were identified as outliers. Further, we used a homozygosity-
201 based method and unphased multilocus genotype (MLG) to calculate G12 parameter in each

202 population (Harris et al., 2018). G12 could employ unphased genotype data and is robust similar to
203 H12 (using phased genotype, Garud et al., 2015) in identifying selective sweeps. We don't intend
204 to separate soft and hard sweeps in each population, but to identify genomic regions under positive
205 natural selection. For *H. armigera* has a similar effective population size while much higher
206 recombination rate (4.86 x 10^{-6} cM/bp, Zhang et al., 2022) compared with *Drosophila melanogaster*
207 (5 x 10^{-7}cM/bp, Garud et al., 2015), we used a smaller window of 100 SNPs in a step of 20 SNPs.
208 Windows with G12 values over 99.95% of autosomal windows in each population were identified
209 as outliers.
210
211 **Genetic data illustration**
212 For the phylogeny tree, we visualized the tree by using Itol (https://itol.embl.de/). Several R
213 packages, including rworldmap (South, 2011), reshape (Wickham, 2007) and rworldxtra (South,
214 2012) were used for illustrating inversion and ancestral haplogroups on the map. Pearson correlation
215 was conducted in R and plotted using ggpubr (Kassambara & Kassambara, 2020) and ggplot2
216 (Wickham, 2011). Package CMplot (Yin, 2020) was used for generation of Manhattan plot. In
217 addition, RectChr (https://github.com/BGI-shenzhen/RectChr) were employed for generating
218 genotype matrix plot.
219
220 # Results
221 **Two differentiated lineages identified by Z chromosomal variation**
222 The phylogeny reconstructed using Z chromosomal variation identified two differentiated lineages
223 in Central China, of which all samples in the southern lineage (CC-S) were from sites near
224 Changjiang River Region (lower latitudes, Figure 1A). This lineage (CC-S) clustered with samples
225 from Oceania (*H. a. conferta*) and differentiated with other *H. armigera* samples in China (Figure
226 1A). This topology was further supported by summary tree generated by ASTRAL-III based on 179
227 window (100kb) trees using Z chromosomal polymorphisms (Supplementary Figure 1). The genetic
228 pattern however, was differentiated with the neighbour-joint (NJ) phylogeny (Zhang et al., 2022)
229 and genetic structure (Figure 1B; Supplementary Figure 2) based on autosomal SNPs. To further
230 characterize regional variation across Z chromosome, we calculated genetic differentiation between

231 CC-S and CC-N (CC-N vs. AUS for comparison), and divided Z chromosome into five regions. Of
232 these five regions, two major topologies were identified, with topology one consistent with the
233 autosomal phylogeny (Zhang et al., 2022) and topology two clustering CC-S *H. armigera* lineage
234 with *H. a. conferta*. Tree topology in region 4 was consistent with topology two. Topologies in
235 region 1, region 2 and region 5 that show similarity with topology two have individuals from CC-
236 N or NR clustering with *H. a. conferta* individuals. Only region 3 has a tree topology consistent
237 with topology one based on autosomal SNPs (Supplementary Figure 3). In addition, we found high
238 correlation between $Fst_{(CC\text{-}N\ vs.\ CC\text{-}S)}$ and $Fst_{(CC\text{-}N\ vs.\ AUS)}$ (Figure 1C), which support widespread
239 introgression of *H. a. confera* into *H. armigera*. This was further supported by window-based
240 introgression analysis (Figure 1D; Supplementary Figure 4).
241
242 **Z chromosomal inversion in *H. armigera***
243 The first principal component (PC) of the Z chromosomal variation divided samples in Central
244 China into two distinguished clusters (Supplementary Figure 5), which is consistent with two
245 differentiated lineages in Central China (Figure 1A). PC1 further identified three individuals (Aq4,
246 Jz42 and DF44) with mixed genetic background. Similar to samples from CC-S, these three
247 individuals were all from the Changjiang River region. Based on the identified PC, we further
248 conducted single-SNP association tests using genotypes each SNP against the samples' first PC
249 coordinate in Asaph. This analysis identified one distinguished peak at Z chromosome, with the
250 boundary ranges from 6.7 - 12.4 Mb, by setting a threshold of -log10(P) > 15 (Figure 2A). We
251 further isolated SNPs with $-\log_{10}(P) > 15$ in the inversion region and constructed genotype matrix
252 of these differentiated SNPs in male individuals of Central China (Figure 2B, see Supplementary
253 Figure 6 for genotype matrix of all individuals). This inversion region in CC-S lineage was with
254 higher linkage disequilibrilum (LD) compared with other regions of CC-S as well as CC-n lineage
255 and AUS (*H. a. conferta*) (Figure 2C). The genotype matrix identified nine individuals with derived
256 alleles and were with Z chromosomal inversion. The geographical distribution of this inversion
257 showed all individuals possessing this inversion are restricted to Changjiang River region (lower
258 latitudes, Figure 2D).
259 Genetic variation of CC-N was higher than CC-S, the regions with lower π ratio ($\pi_{CC\text{-}N}/\pi_{CC\text{-}S}$) and
260 higher $-\log_{10}(P)$ thus would be indicatives of selection signals on CC-N lineage (i.e., reduced genetic

261 variation). We treated regions with π ratio lower than 99% of Z chromosome regions (a total of 37
262 20kb-windows) as candidates and further considered the $-\log_{10}(P)$ more than 7.3 ($P = 5 \times 10^{-8}$).
263 Interestingly, we identified several genes that maybe potentially important in adaptation of CC-N
264 and differentiation of CC-N and CC-S lineages. Of which, *SLC25A25* and *IGF2BP2* were located
265 outside the inversion, while period circadian protein (*per*) and Aryl hydrocarbon receptor nuclear
266 translocator−like protein 1 (*ARNTL*, or *cycle*) were within the inversion (Figure 2A). Genotype
267 matrix of *per* and *cycle* show that the male individuals in CC-S have similar derived genotype with
268 those in *H. a. conferta* males (Figure 3A). In addition, three male individuals including Aq4, DF44,
269 and Jz22 that are collected from Changjiang River region have the same derived genotypes at both
270 *per* and *cycle*. Several additional individuals from both CC-N (Ay13, DF28, DF46, Xj7, Kf2, Kf21,
271 Kf6) and NR (Hami12, Hami7, Hami9, and Kel26) have only derived genotype in *cycle* (Figure
272 3A). We also found the genetic differentiation (Fst) and divergence (dxy) of these two genes
273 elevated compared with surrounding regions, and nucleotide diversity of two genes were lower in
274 CC-N lineage compared with those in CC-S lineage (Figure 3B). When conducted protein modeling
275 of two divergent haplotypes dominated in each group, we found difference not only in amino acid
276 but also in protein secondary structure (Figure 3C). In CC-S lineage, these two clock genes, although
277 separated by a distance of around 4 Mb, show highly linkage within and between genes (Figure 3D).
278

279 **Genes contributed to differentiation of southern and northern lineages**

280 In addition to genes in Z chromosome that differentiated between southern and northern *H. armigera*
281 lineages. Genome-wide association analysis of the autosomal SNP genotype against first principal
282 component (of Z chromosome variation) further isolated several regions with significant -log10(*P*)
283 value (> 7.3) and elevated genetic differentiation (Fst). These regions spread across the genome and
284 distributed at more than five chromosomes (Figure 4A). We further checked the annotated genes in
285 these regions, and identified acyl-CoA Delta (11) desaturase (*D11DS*) and bifunctional trehalose-6-
286 phosphate synthase/phosphatase (*TPSP*) in Chr12, phosphorylase b kinase regulatory subunit
287 (*CG7766*), carboxypeptidase Q (*CPQ*) of Chr 13, and cryptochrome-2 (*Cry2*) from Chr15, and
288 circadian clock-controlled protein *daywake* (*annon-3B1.2*) in Chr23. Interestingly, these genes all
289 have elevated Fst and dxy, which supported a divergent selection with gene flow in formation of
290 these divergent peaks. In addition, decreased π ratio ($\pi_{CC\text{-}N}/\pi_{CC\text{-}S}$) of these genes compared with

291 nearby regions supported local adaptation of CC-N population may have contributed to this
292 divergence.
293
294 **Widespread selective sweep signals in different *H. armigera* lineages**
295 Several selective sweeps were identified based on strict threshold in different populations (Figure
296 5). Compared with SWEEPFINDER2, haplotype-based parameter G12 generally identified more
297 regions with sweep signals. Among them, several linked P450 genes identified including
298 CYP337B3V1, CYP6B2, and CYP6a21 were shared across three populations (Figure 5). As we
299 found many regions contributed to differentiation of southern and northern *H. armigera* lineages of
300 Central China, and intensive while diverse application of insecticide for controlling different *H.*
301 *armigera* populations (Wu et al., 2005; Li et al., 2010; Zhang et al., 2012; Yang et al., 2013; Lu et
302 al., 2021; Downes et al., 2017), different *H. armigera* lineages should have adapted to local climates
303 (Chen et al., 2013) and regionally specific insecticides. Indeed, we found many lineage-specific
304 signals of selective sweeps in populations in Central China, Northwestern China, and Oceania (*H.*
305 *a. conferta*) based on G12 parameter (Figure 5). In Central China, we found several GSTs on Chr7,
306 carboxylic ester hydrolase (CCE001c) and Juvenile hormone esterase (JHER) of Chr 15, and
307 carboxyl/choline esterase (CCE014a) on Chr30 were under selection. In addition, trehalose-6-
308 phosphate synthase/phosphatase (TPSP) located at Chr12, and Cytochrome P450 12a4 (CYP12A4)
309 and CPQ of Chr13 that were identified as outliers based on G12. In Northwestern China, we found
310 several genes with signals of sweeps except CYP337B3 (Figure 5). In Oceania, cadherin (CadN,
311 Chr 4), choline O-acetyltransferase (ChAT, Chr6), cadherin-89D (Cad89D, Chr6), GST(Chr7),
312 carboxylic ester hydrolase (CCE001c, Chr15), and juvenile hormone esterase (JHER, Chr15)
313 showed selective signals. SWEEPFINDER2 identified less but unique selective signals among
314 different populations. For example, in Central China lineage, PTP10D on Chr13 showed higher
315 CLR values compared with those of other regions. Similarly, two genes (D11DS, and Scd1) on
316 Chr12 have higher CLR values in NC lineage.
317
318

319 # Discussion

320 Understanding the genetic basis of local adaptation in different insect pest lineages is important in
321 effective pest control and management. To investigate genetic differentiation and local adaptation
322 of two *H. armigera* lineages in Central China identified by Z chromosome polymorphism, we
323 combined genetic differentiation and gene flow analysis to disentangle effects of introgression, local
324 adaptation and chromosome structure on genome-wide genetic variation. Compared with southern
325 lineage in Central China, we identified several genes in northern lineage of Central China with
326 sweep signals (e.g., low nucleotide diversity and high differentiation), restricting introgression from
327 the other subspecies *H. a. conferta*, and may have contributed to adaptation of this lineage to high
328 latitudes. On the contrary, a Z chromosome inversion in southern lineage, keeping divergent alleles
329 of two circadian clock genes introgressed from *H. a. conferta*, have contributed to differentiation of
330 these two lineages in Central China, and may have led to disorder of clock oscillator for acclimation.
331 Further, we conducted selective sweep analysis based on two parameters and identified several
332 genes that may have contributed to local adaptation in different *H. armigera* lineages (Northwestern
333 China, Central China, and Oceania). Overall, our results reveal widespread local adaptation of
334 different *H. armigera* lineages that should be considered for predicting *H. armigera* range expansion
335 and executing effective pest management.

336

337 **Regional climate contributed to differentiation of southern and northern *H. armigera* lineages**
338 **in Central China**

339 We conducted analysis on *H. armigera* resequencing data and identify two differentiated lineages
340 in Central China that are differentiated at Z chromosome. Based on population genetic
341 differentiation analysis and gene flow, we found a pattern of divergent selection with gene flow in
342 Central China that was contributed by local adaptation of northern lineage from the high latitudes.
343 A total of four clock genes (Z chromosome: *per* and *cycle*; autosome chromosome: *cry2* and
344 *daywake*) were under selection in the northern lineage. While in the southern lineage, a Z
345 chromosome inversion locked alleles of two core clock genes (*per* and *cycle*) from *H. a. conferta*
346 that further boost divergence of these two lineages. Introduction divergent alleles of these two clock
347 genes into local population may have disrupted sensing system to temperature and photoperiod. In

348 addition, apart from southern lineage, linked introgressed alleles of *per* and *cycle* were only
349 identified in three individuals (Aq4, DF44, and Jz22) from Changjiang River Region without an Z
350 chromosome inversion. These results suggest a combination of certain genotypes in clock genes
351 may be needed to maintain normal function of seasonal sensing system.
352
353 Studies shows acclimation to high latitudes through diapause was contributed by multiple loci with
354 minor effects and support a polygenic model (Ragland et al., 2019). In our study, in addition to four
355 clock genes, we further isolated several genes under selection in northern lineage and may have
356 contributed to local adaptation of this lineage to temperature and photoperiod of high latitudes. In
357 Z chromosome, we found *SLC25A25*, a Ca^{2+}-sensitive ATP carrier in the inner mitochondrial
358 membrane, fluctuates in a circadian manner in mouse skeletal muscle and might be involve in
359 muscle thermogenesis (Nakao et al., 2017). The other gene on Z chromosome is *IGF2BP2*, which
360 is from insulin-like growth factor2 mRNA binding protein family in human and may associated
361 with impaired insulin secretion (Dai et al., 2020). Phosphorylation of this gene is critical for
362 coordinating cellular function and nutrient metabolism (Dai et al., 2020). In the autosomal GWAS
363 analysis by correlation analysis with each SNP and PCA1 identified by Z chromosome
364 polymorphism, *CCG7766*, a Phosphorylase b kinase regulatory subunit that involved in
365 carbohydrate metabolic process and glycogen metabolic process was identified as an outlier and
366 show local adaptation in northern lineage. The other gene that appears on this gene panel was Acyl-
367 CoA Δ^{11}Z-desaturases (D11DS) that occurred ubiquitously in animal and fungal kingdoms and
368 plays essential roles in fatty acid metabolism and response to temperature fluctuations by regulation
369 of cell membrane fluidity (Knipple et al., 1998). In addition, bifunctional trehalose-6-phosphate
370 synthase/phosphatase (TSTP) that shows selective sweep signal in northern lineage is important in
371 energy production and other biological processes of invertebrates (Tang et al., 2018). Additionally,
372 we found an adaptative CPQ gene that encodes carboxypeptidase Q, which is involved in proteolytic
373 functions and resistant to deficit water status (Gheyas et al., 2021; Mohamadipoor Saadatabadi et
374 al., 2021), such as diapause in long overwintering periods.
375
376 Overall, our studies identified candidate genes in northern lineage of Central China that involved in
377 circadian clock and could sense seasonal variation of temperature and photoperiod, and in energy

378 metabolism that may have contributed to local adaptation during stressful and long overwintering
379 periods. However, although verified by bioinformatic analysis, additional evidence from
380 transcriptome data and molecular experiment are needed to depict a more robust and comprehensive
381 understanding of adaptation to high latitude of this pest. We also propose a more intensive sampling
382 in southern China as well as southeast Asia to investigate the distribution of Z chromosome
383 inversion and identify genetically distinct lineages that may have adapted to different temperature
384 and photoperiod (Chen et al., 2013; Virachack et al., 2018).
385
386 **Adaptation to insecticide leads to widespread local adaptation in different *H. armigera***
387 **lineages**
388 Up to now, *H. armigera* shows the highest number of reported cases of insecticide resistance
389 worldwide and have developed resistance against most of applied chemicals and several *Bacillus*
390 *thuringiensis* (Bt) toxins (Valencia-Montoya et al., 2020). Resistance to insecticides such as
391 carbamates, organophospates, pyrethroids families, or to Bt has been reported all over the world for
392 *H. armigera*, while others including spinosyns, indoxacarb, emamectin, and benzoate were reported
393 regionally (https://irac-online.org/). The differentiation in resistance in regional populations indicate
394 widespread local adaptation to applied insecticides of different *H. armigera* lineages.
395
396 *CYP337B3* shows selection signals in all populations, which is consistent with widespread
397 distribution of this P450 gene in field that was resistant to fenvalerate (Joußen et al., 2012; Xu et
398 al., 2014; Han et al., 2015; Rasool et al., 2015). The resistant genotype of *CYP337B3* was also
399 introduced along with populations of *H. armigera* into Brazil (Durigan et al., 2017) and introgressed
400 to its sibling species (i.e., *H. zea*) in South America. In addition, we found selection signals in
401 several lineage-specific genes that indicate local adaptation by development of insecticide resistance.
402 For example, in Central China, the glutathione S-transferase genes that were also identified in
403 previous study (Zhang et al., 2022) took part in insecticide and plant secondary metabolites (Francis
404 et al., 2005; Gawande et al., 2011; Rane et al., 2019). This gene, not identified in Northwestern
405 China (e.g., Xinjiang) but in Oceania, may indicate introgression between *H. armigera* and *H. a.*
406 *conferta*. In addition, one esterase (*CCE001*) that have involved in fenvalerate resistance was also
407 identified in these two regions (Central China and Oceania), which support widespread introgression

408 of insecticide-resistant genes.

409

410 In Central China, one lineage-specific gene encodes carboxyl esterase (*CCE014a*) that could
411 mediate metabolic resistance to organophosphates and carbamates (Venkatesan et al., 2022) and
412 Cytochrome P450 (*CYP12A4*) that could facilitate adaptation to lufenuron (Bogwitz et al., 2005)
413 were identified, while no lineage-specific insecticide resistant genes were isolated in Northwestern
414 China. This may be due to shorter insecticide applied history in Northwestern China compared with
415 that in Central China, differentiation of cropping pattern or geographical and climatic environment
416 (Li et al., 2010; Lu, 2021). On the contrast, we found many population-specific genes (over the
417 threshold) that may have contributed to insecticide resistance in Oceania (i.e., *H. a. conferta* lineage).
418 For example, two cadherins (*CadN* and *Cad89D*) could have contributed to *Bt* resistance (Fabrick
419 & Wu 2015; Du et al., 2019) of transgenic crops in this region were outliers. In addition, one gene
420 that encodes choline acetyltransferase (*ChAT*) could be underlying resistance of this lineage to
421 insecticides that targeted on neurons (Liu et al., 2022). We noted that *Cad89D* and *ChAT* were with
422 weak peaks in Central China (although not being outliers base on threshold), while *CadN* sweep
423 was strong in Oceania and not identified in Central China. The time to start Bt transgeneic crop
424 cultivation in China (since 1997, Huang et al., 2002) and Australia (since 1996, Fitt, 2002) were
425 similar, and the difference in identification of lineage-specific Bt genes could be due to sampling
426 bias or due to the difference in development of *Bt* resistance, which needs further investigation.

427

428 **Management of agricultural insect pest required more comprehensive understandings of local**
429 **adaptation and introgression**

430 For most insect pests, environmental factors are important for successful reproduction and
431 succession of local populations (Fält-Nardmann et al., 2016). Among the genetic agents underlying
432 adaptation of lineages to high latitudes, circadian clock genes could influence insect behaviors such
433 as locomotion, courtship, and photoperiodism should receive more attention (Khyati et al., 2017).
434 In addition, due to more hostile environments of high latitudes, successful range expansion of
435 agricultural pests requires that species are able to use local temperature and/or day length as cues to
436 coordinate their phenology (Saikkonen et al., 2012), pre-adaptation of insects to similar
437 environments thus would facilitate rapid colonization to new habitats (Li et al., 2015). For their

importance in viability of insects, techniques such as molecular editing and sterile insect technique (SIT) that involved clock genes (and other genes related to phenological adaptation) should benefit integrated pest management to minimize pesticide use and prevent rapid resistance to insecticides (Khyati et al., 2017).

Addressing the genetic basis of pesticide resistance is the best way to find a sustainable crop plant protection strategy (Barzman et al., 2015). Our analysis of adaptative genes that are related to different insecticides not only identify target genes underlying molecular basis of insecticide resistance, but also revealed that gene flow may have contributed to widespread and rapid development of insecticide resistance of *H. armigera*. This information would help putting forward efficient pest management strategy according to the resistant list of local/regional pest populations. Further analysis to investigate the frequency and distribution of insecticide resistant alleles (introgressed or locally adaptative) across Southeast Asia and Oceania, as well as other regions, are needed for a more successful regional integrated pest management strategy.

Conclusion

We combined association analysis and population genetic parameters to isolate adaptative agents underlying local adaptation of different *H. armigera* lineages. Genes related acclimation and insecticide resistance were identified despite widespread gene flow. In addition, Z chromosome inversion could disrupt local adaptation by locking maladaptive alleles of introgressed genes. Our results show widespread local adaptation of different *H. armigera* populations, which should further be implemented in integrated pest management and accounted for efficient regional management of insecticide resistance.

References

Alyokhin A, Mota-Sanchez D, Baker M, et al. The Red Queen in a potato field: integrated pest management versus chemical dependency in Colorado potato beetle control. Pest Manag. Sci., 2015, 71(3): 343-356.

Arnold M L, Kunte K. Adaptive genetic exchange: a tangled history of admixture and evolutionary innovation. Trends Ecol. Evol., 2017, 32(8): 601-611.

Minh BQ, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol., 37:1530-1534.

Barzman, M., Bàrberi, P., Birch, A.N.E. et al. Eight principles of integrated pest management. Agron. Sustain. Dev. 2015, 35: 1199–1215.

Bogwitz MR, Chung H, Magoc L, Rigby S, Wong W, O'Keefe M, McKenzie JA, Batterham P, Daborn PJ. Cyp12a4 confers lufenuron resistance in a natural population of Drosophila melanogaster. Proc Natl Acad Sci U S A. 2005, 102(36):12807-12.

Chang C C, Chow C C, Tellier L C A M, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 2015, 4(1): s13742-015-0047-8.

Crispo E, Moore J S, Lee-Yaw J A, et al. Broken barriers: Human-induced changes to gene flow and introgression in animals: An examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. BioEssays, 2011, 33(7): 508-518.

Dai, N. The diverse functions of IMP2/IGF2BP2 in metabolism. Trends Endocrinol. Metab., 2020, 31(9): 670-679.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. Bioinformatics, 27(15): 2156-2158.

Danks, H. V. (2000) Measuring and reporting life-cycle duration in insects and arachnids. Eur. J. Entomol. 97: 285–303.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M. and McKenna, A., 2011. A framework for variation discovery

491 and genotyping using next-generation DNA sequencing data. Nat. Genet., 43(5): 491-498.
492 Downes S, Kriticos D, Parry H, et al. A perspective on management of *Helicoverpa armigera*: transgenic
493 Bt cotton, IPM, and landscapes. Pest Manag. Sci., 2017, 73(3): 485-492.
494 Du, L., Chen, G., Han, L. et al. Cadherin CsCad plays differential functional roles in Cry1Ab and Cry1C
495 intoxication in *Chilo suppressalis*. Sci. Rep., 2019, 9: 8507.
496 Durigan MR, Corrêa AS, Pereira RM, Leite NA, Amado D, de Sousa DR, Omoto C. High frequency of
497 CYP337B3 gene associated with control failures of *Helicoverpa armigera* with pyrethroid
498 insecticides in Brazil. Pestic Biochem Physiol. 2017, 143:73-80.
499 Fabrick, J.A., Wu, Y. and Wu, Y., 2015. Roles of insect midgut cadherin in Bt intoxication and
500 resistance. *Bt resistance: characterization and strategies for GM crops producing Bacillus*
501 *thuringiensis toxins*, 2015 pp.69-86.
502 Fält-Nardmann, J., Klemola, T., Roth, M., Ruohomäki, K., & Saikkonen, K. (2016). Northern geometrid
503 forest pests (Lepidoptera: Geometridae) hatch at lower temperatures than their southern
504 conspecifics: Implications of climate change. Eur. J. Entomol., 2016: 113.
505 Fitt G. Deployment and impact of transgenic Bt cotton in Australia. Econ. Environ. Impacts
506 Agrobiotechnol., 2003:141-161.
507 Francis F, Vanhaelen N, Haubruge E. Glutathione S-transferases in the adaptation to plant secondary
508 metabolites in the *Myzus persicae* aphid. Arch. Insect Biochem. Physiol., 2005, 58(3):166-74.
509 Garud N R, Messer P W, Buzbas E O, et al. Recent selective sweeps in North American Drosophila
510 melanogaster show signatures of soft sweeps. PLoS Genet., 2015, 11(2): e1005004.
511 Gawande ND, Subashini S, Murugan M, Subbarayalu M. Molecular screening of insecticides with sigma
512 glutathione S-transferases (GST) in cotton aphid Aphis gossypii using docking. Bioinformation.
513 2014: 10(11):679-83.
514 Gheyas A A, Vallejo-Trujillo A, Kebede A, et al. Integrated environmental and genomic analysis reveals
515 the drivers of local adaptation in African indigenous chickens. Mol. Biol. Evol., 2021, 38(10): 4268-
516 4285.
517 Han Y, Yu W, Zhang W, Yang Y, Walsh T, Oakeshott JG, Wu Y. Variation in P450-mediated fenvalerate
518 resistance levels is not correlated with *CYP337B3* genotype in Chinese populations of *Helicoverpa*
519 *armigera*. 2015, 121:129-35.
520 Hahn, D.A. and Denlinger, D.L., 2011. Energetics of insect diapause. Annu. Rev. Entomol., *56*, pp.103-

521 121.
522 Harris, AM, Garud NR, DeGiorgio M. Detection and classification of hard and soft sweeps from
523 unphased genotypes by multilocus genotype identity. Genetics, 2018: 1429-1452.
524 Hasebe M, Shiga S. Clock gene-dependent glutamate dynamics in the bean bug brain regulate
525 photoperiodic reproduction. PLoS Biol., 2022, 20(9): e3001734.
526 Hoang D.T., O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh. UFBoot2: Improving the ultrafast
527 bootstrap approximation. Mol. Biol. Evol., 2018, 35:518–522.
528 Hoikkala, A. and Poikela, N., 2022. Adaptation and ecological speciation in seasonally varying
529 environments at high latitudes: *Drosophila virilis* group. Fly, 16(1): 85-104.
530 Huang JK, RF Hu, CH Fan, CE Pray, S Rozelle, Bt cotton benefits, costs, and impacts in
531 China. AgBioForum 2002, 5: 153–166.
532 Jin, MH, Henry L. North, Yan Peng, Hangwei Liu, Ruiqing Pan, Yan Zhou, Weigang Zheng et al.
533 Adaptive evolution to the natural and anthropogenic environment in a global invasive crop pest, the
534 cotton bollworm. The Innovation (2023), 4(4): 100454.
535 Jones, C. M., Papanicolaou, A., Mironidis, G. K., Vontas, J., Yang, Y., Lim, K. S., Oakeshott, J. G.,
536 Bass, C., & Chapman, J. W. Genome wide transcriptional signatures of migratory flight activity in
537 a globally invasive insect pest. Mol. Ecol., 2015, 24: 4901– 4911.
538 Joußen N, Agnolet S, Lorenz S, et al. Resistance of Australian *Helicoverpa armigera* to fenvalerate is
539 due to the chimeric P450 enzyme CYP337B3. Proc. Natl. Acad. Sci. U.S.A., 2012, 109(38): 15206-
540 15211.
541 Kassambara, A. and Kassambara, M.A. Package 'ggpubr'. R package version 0.1, 2020, 6(0).
542 Khyati, Malik I, Seth R K. Insect clocks: implication in an effective pest management. Biol. Rhythm
543 Res., 2017, 48(5): 777-788.
544 Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics, 2006,
545 173(1): 419-434.
546 Klai, K., Chenais, B., Zidi, M., Djebbi, S., Caruso, A., Denis, F., Confais, J., Badawi, M., Casse, N. and
547 Mezghani Khemakhem, M., 2020. Screening of *Helicoverpa armigera* mobilome revealed
548 transposable element insertions in insecticide resistance genes. Insects, 2020, 11(12): 879.
549 Knipple D C, Rosenfield C L, Miller S J, et al. Cloning and functional expression of a cDNA encoding
550 a pheromone gland-specific acyl-CoA Δ11-desaturase of the cabbage looper moth, *Trichoplusia ni*.

Proc. Natl. Acad. Sci. U.S.A., 1998, 95(26): 15287-15292.

Kozak G M, Wadsworth C B, Kahne S C, et al. Genomic basis of circannual rhythm in the European corn borer moth. Curr. Biol., 2019, 29(20): 3501-3509. e5.

Li S, Cadotte M W, Meiners S J, et al. The effects of phylogenetic relatedness on invasion success and impact: deconstructing Darwin's naturalisation conundrum. Ecol. Lett., 2015, 18(12): 1285-1292.

Li, G., Feng, H., Gao, Y., Wyckhuys, K. A., & Wu, K. Frequency of Bt resistance alleles in *Helicoverpa armigera* in the Xinjiang cotton-planting region of China. Environ. Entomol., 2010, 39(5), 1698-1704.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Liu J, Gao S, Wei L, et al. Choline acetyltransferase and vesicular acetylcholine transporter are required for metamorphosis, reproduction, and insecticide susceptibility in *Tribolium castaneum*. Gene, 2022, 842: 146794.

Lu YH. Ever-evolving advances in the researches of cotton insect pest management in China. Acta Phytophylacica Sin., 2021, 48(5): 937-939.

Malinsky, M, Matschiner, M, Svardal, H. Dsuite - Fast D-statistics and related admixture evidence from VCF files. Mol Ecol Resour. 2021; 21: 584– 595.DeGiorgio M, Christian D. Huber, Melissa J. Hubisz, Ines Hellmann, Rasmus Nielsen, Sweep Finder 2: increased sensitivity, robustness and flexibility. Bioinformatics, 2016, 32(12): 1895-1897.

Mohamadipoor Saadatabadi L, Mohammadabadi M, Amiri Ghanatsaman Z, et al. Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. BMC Vet. Res., 2021, 17(1): 1-9.

Nakao R, Shimba S, Oishi K. Ketogenic diet induces expression of the muscle circadian gene Slc25a25 via neural pathway that might be involved in muscle thermogenesis. Sci. Rep., 2017, 7(1): 1-11.

Nibouche, S., Buès, R., Toubon, J.- F., & Poitout, S. Allozyme polymorphism in the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae): Comparison of African and European populations. Heredity, 1998, 80: 438– 445.

Ortiz-Barrientos D, Engelstädter J, Rieseberg L H. Recombination rate evolution and the origin of species.Trends Ecol. Evol., 2016, 31(3): 226-236.

Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS: Identification of a candidate adaptive polymorphism

for Drosophila life history by parallel independent clines on two continents. Mol. Ecol., 2010, 19:760-774.

Paini D R, Sheppard A W, Cook D C, et al. Global threat to agriculture from invasive species. Proc. Natl. Acad. Sci. U.S.A., 2016, 113(27): 7575-7579.

Pélissié B, Crossley M S, Cohen Z P, et al. Rapid evolution in insect pests: the importance of space and time in population genomics studies. Curr. Opin. Insect. Sci., 2018, 26: 8-16.

Peng, Yan, Minghui Jin, Zhimin Li, Hongran Li, Lei Zhang, Songmiao Yu, Zhuting Zhang et al. Population-genomics provide insights into the evolution and adaptation of the Asia corn borer. Mol. Biol. Evol., 2023, 40(5): msad112.

Pinho, C. and Hey, J. Divergence with gene flow: models and data. Annu. Rev. Ecol. Evol. S., 2010, 41: 215-230.

Pratissoli, D., Lima, V. L., Pirovani, V. D., & Lima, W. L. Occurrence of Helicoverpa armigera (Lepidoptera: Noctuidae) on tomato in the Espírito Santo state. Hortic. Bras., 2015, 33: 101– 105.

Ragland G J, Armbruster P A, Meuti M E. Evolutionary and functional genetics of insect diapause: a call for greater integration. Curr. Opin. Insect. Sci., 2019, 36: 74-81.

Rane, R.V., Ghodke, A.B., Hoffmann, A.A., Edwards, O.R., Walsh, T.K. and Oakeshott, J.G., 2019. Detoxifying enzyme complements and host use phenotypes in 160 insect species. Curr. Opin. Insect. Sci., 2019 31: 131-138.

Rasool A, Joußen N, Lorenz S, Ellinger R, Schneider B, Khan SA, Ashfaq M, Heckel DG. An independent occurrence of the chimeric P450 enzyme CYP337B3 of *Helicoverpa armigera* confers cypermethrin resistance in Pakistan. Insect Biochem. Mol. Biol. 2014, 53:54-65.

Renou M. Is the evolution of insect odorscapes under anthropic pressures a risk for herbivorous insect invasions?. Curr. Opin. Insect. Sci., 2022: 100926.

Riaz S, Johnson J B, Ahmad M, et al. A review on biological interactions and management of the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). J. Appl. Entomol., 2021, 145(6): 467-498.

Nowling RJ, F Fallas-Moya, et al. Fast, low-memory detection and localization of large, polymorphic inversions from SNPs. PeerJ, 2022: e12831.

Roderick GK and Navajas M.Genes in new environments: genetics and evolution in biological control. Nat. Rev. Genet., 2003, 4: 889-899.

Roesti M, Gilbert KJ, Samuk K. Chromosomal inversions can limit adaptation to new environments. Mol. Ecol., 2022, 31(17):4435-4439.

Saikkonen K, Taulavuori K, Hyvönen T, et al. Climate change-driven species' range shifts filtered by photoperiodism. Nat. Clim. Chang., 2012, 2(4): 239-242.

South, A. rworldmap: A new R package for mapping global data. R J., 2011 3(1).

South, A. rworldxtra: Country boundaries at high resolution. R package version 1.01. 2012.

Szendrei Z, Grafius E, Byrne A, et al. Resistance to neonicotinoid insecticides in field populations of the Colorado potato beetle (Coleoptera: Chrysomelidae). Pest Manag. Sci., 2012, 68(6): 941-946.

Tang B, Wang S, Wang S G, et al. Invertebrate trehalose-6-phosphate synthase gene: genetic architecture, biochemistry, physiological function, and potential applications. Front. Physiol., 2018, 9: 30.

Tigano, Anna, and Vicki L. Friesen. Genomics of local adaptation with gene flow. Mol. Ecol., 2016, 25(10): 2144-2164.

Toepfer, S., Li, H., Pak, S.G., Son, K.M., Ryang, Y.S., Kang, S.I., Han, R. and Holmes, K., 2014. Soil insect pests of cold temperate zones of East Asia, including DPR Korea: A review. J. Pest Sci., 87: 567-595.

Valencia-Montoya W A, Elfekih S, North H L, et al. Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. Mol. Biol. Evol., 2020, 37(9): 2568-2583.Venkatesan, T., B. R. Chethan, and M. Mani. Insecticide resistance and its management in the insect pests of horticultural crops. Trends Hortic. Entomol., 2022: 455-490.

Virachack, Hatsada, et al. Temperature effects on the development of Vietnamese cotton bollworm Helicoverpa armigera (Hübner). J. Fac. Agr., Kyushu Univ., 2018, 63 (2): 303-309.

Wickham, H. Reshaping data with the reshape package. J. Stat. Softw., 2007, 21:1-20.

Wickham, H., 2011. ggplot2. Wiley Interdiscip. Rev. Comput. Stat., 3(2): 180-185.

Williams KD, Busto M, Suster ML, So AKC, Ben-Shahar Y, Leevers SJ, Sokolowski MB: Natural variation in Drosophila melanogaster diapause due to the insulin-regulated PI3- kinase. Proc. Natl. Acad. Sci. U. S. A., 2006, 103:15911-15915.

Wu K, Mu W, Liang G, et al. Regional reversion of insecticide resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae) is associated with the use of Bt cotton in northern China. Pest Manag. Sci., 2005, 61(5): 491-498.

Wu KM, Guo YY. Genotype differentiation and regional migratory regularity of *Helicoverpa armigera* in China. Plant Prot., 2007, 33(5):5-11.

Wu KM, Guo YY. Investigations on the migration and diapause in *Helicoverpa armigera* -Diapause responses of different geographical populations of cotton bollworm *Helicoverpa armigera* Hübner. Zhongguo Nong Ye Ke Xue, 1997, 30 (3): 1-6.

Wu Kongming, Guo Yuyuan, Wei Jianfu, Sun Fuzai, On the cold hardiness of cotton bollworm Helicoverpa armigera Hübner. Sheng Tai Xue Bao, 1997, 17(3): 298-302.

Xu L, Li D, Qin J, Zhao W, Qiu L. Over-expression of multiple cytochrome P450 genes in fenvalerate-resistant field strains of *Helicoverpa armigera* from north of China. Pestic. Biochem. Physiol., 2016, 132:53-8.

Yang Y, Li Y, Wu Y. Current status of insecticide resistance in *Helicoverpa armigera* after 15 years of Bt cotton planting in China. J. Econ. Entomol., 2013, 106(1): 375-381.

Yin, L. CMplot: circle manhattan plot. R package version, 2020, 3(2).

Chen YS and Tu XY, Geographic variation in developmental duration of the cotton bollworm , *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae). J. Environ. Entomol., 2016 ,38(4):771 − 778.

Chen, Y.S., Chen, C., He, H.M., Xia, Q.W. and Xue, F.S. Geographic variation in diapause induction and termination of the cotton bollworm, *Helicoverpa armigera* Hübner (Lepidoptera: Noctuidae). J. Insect Physiol., 2013, *59*(9):855-862.

Zhang C, Rabiee M, Sayyari E, et al. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinform., 2018, 19(6): 15-30.

Zhang H, Tian W, Zhao J, et al. Diverse genetic basis of field-evolved resistance to Bt cotton in cotton bollworm from China. Proc. Natl. Acad. Sci. U.S.A., 2012, 109(26): 10275-10280.

Zhang J, Zhang F, Tay W T, et al. Population genomics provides insights into lineage divergence and local adaptation within the cotton bollworm. Mol. Ecol. Resour., 2022, 22(5): 1875-1891.

Figure Legends

Figure 1. Z chromosome divergence of *H. armigera* despite widespread gene flow. (A) Maximum likelihood phylogeny based on concatenated SNPs from Z chromosome. Central China southern lineage (CC-S, with arrows) and Central China northern lineage (CC-N) were labeled with blue and cyan, respectively. Similar topology was reconstructed based on Astral (Supplementary Figure 1) (B) Principal component analysis based on autosomal SNPs. Note the CC-S and CC-N samples were clustered together. (C) Genetic differentiation of *H. armigera* males among different lineages. Two major topologies were illustrated, with topology one (①) consistent with the species tree, while topology two (②) gathering *H. a. conferta* (AUS) and CC-S individuals. "②※" indicates a window tree similar to topology two, but with CC-S lineage including more individuals from CC-N lineage. See supplementary Figure 2 for more information. The lower panel shows significant correlation between genetic differentiations of CC-S vs. CC-N and AUS vs. CC-N in Z chromosome. (D) Widespread gene flow between *H. a. conferta* from Oceania and *H. armigera* populations in Central China.

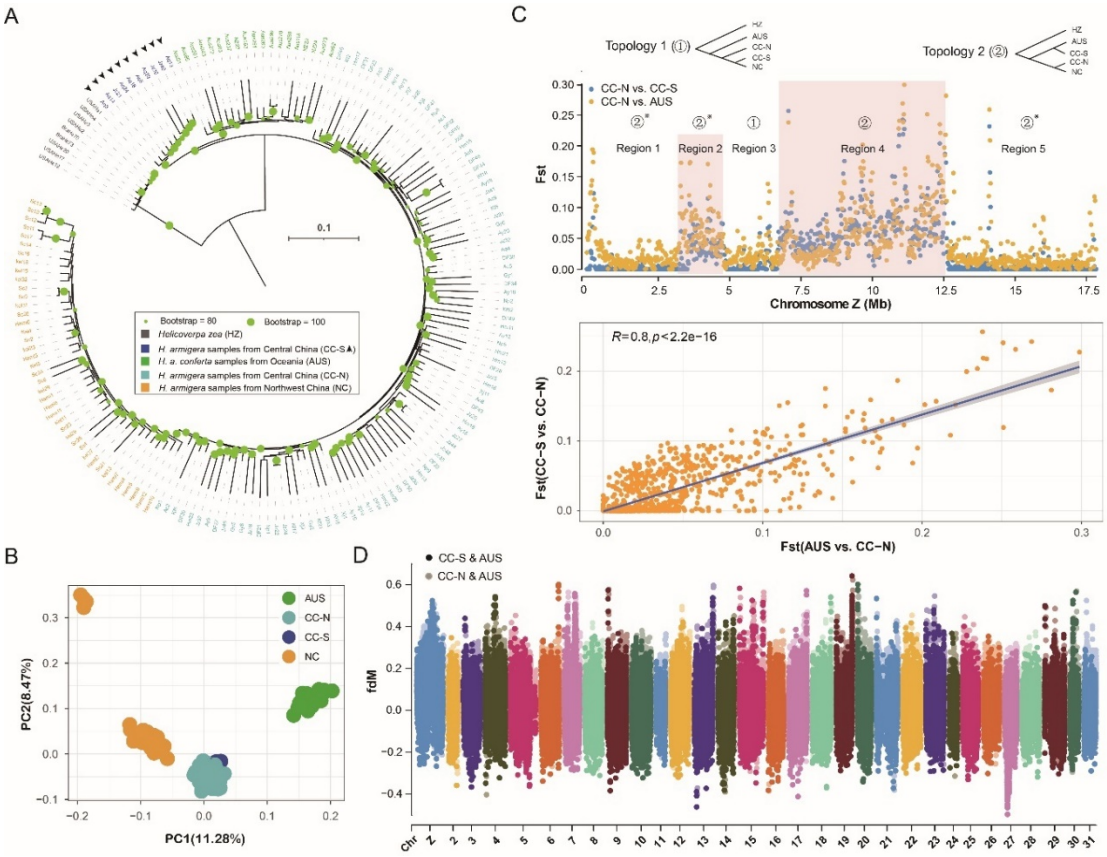
Figure 2. Detection and characterization of putative Z chromosome inversion in *Helicoverpa armigera*. (A) Manhattan plot of GWAS analysis between Z chromosome polymorphism and principal component 1 (PC1). The break point of the inversion was set at the site with $-\log_{10}(p)$ value more than 25. The central panel is π ratio ($\pi_{CC\text{-}N}/\pi_{CC\text{-}S}$) calculated in 5kb window, with labeled genes identified in 1% lowest π ratio regions. The lower panel is *fdM* distribution of CC-N vs. AUS and CC-S vs. AUS calculated with 100-SNP window and 25-SNP step. (B) Genotype matrix of highly differentiated SNPs ($-\log_{10}(p) > 15$) in the putative chromosome inversion. A total of 3,237 SNPs were used for analysis. (C) Linkage disequilibrium analysis in different population and genomic regions; Inversion indicates the putative inversion region, and other regions are the rest of Z chromosome excluding the inversion. AUS are male individuals sampled from *H. a. conferta*. (D) Geographical distribution of two haplogroups (ancestral and inversion) in Central China. Only CC-S and CC-N individuals included for analysis as inversion identified in Central China (see Supplementary Figure 4 for the genotype matrix of all individuals).

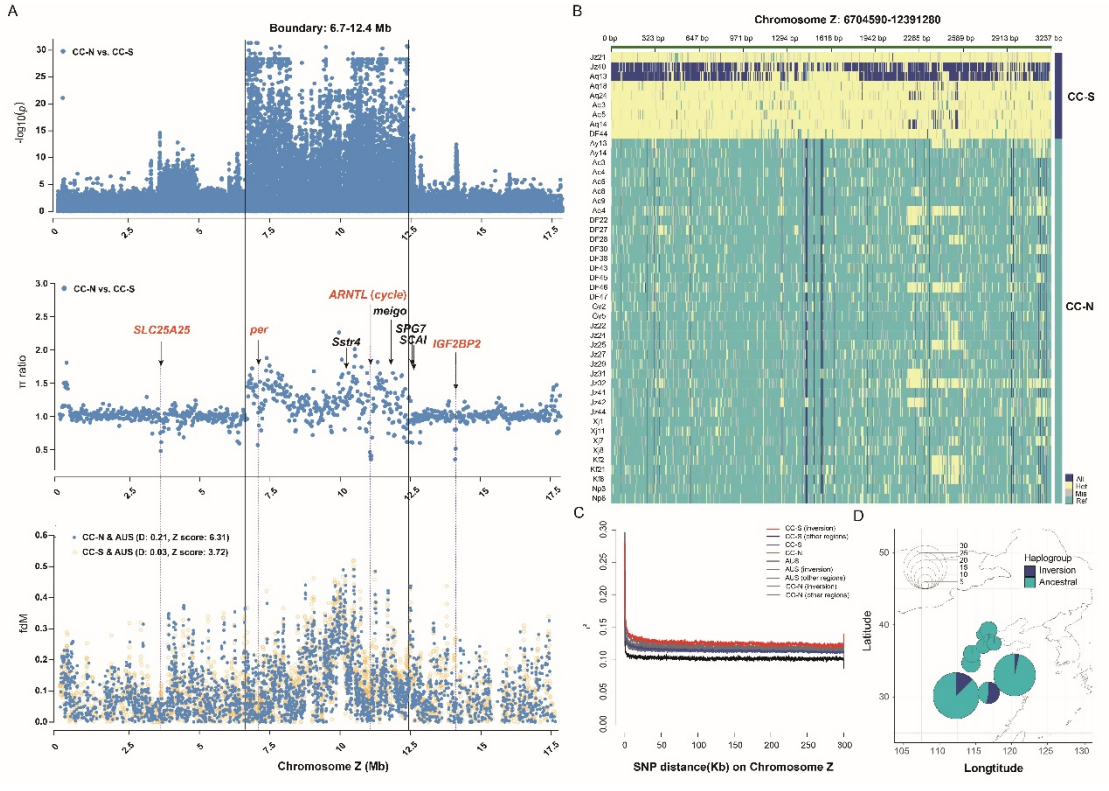
Figure 3. Characterization of *per* and *ARNTL* (*cycle*) in the inversion region. (A) Genotype matrix of polymorphism at CDS regions of *per* and *cycle* genes. (B) Population genetic statistics of *per* and *cycle*. The red dashed line in each panel was the mean value of each statistic calculated by Z chromosomal genetic polymorphism. (C) Protein structures of two differentiated haplotypes in *per* and *cycle*. Cyan structure was constructed from the CC-N dominant haplotype, and additional blue structure that shows the difference was constructed with the CC-N dominant one. (D) Patterns of LD blocks between and within genes (*per* and *cycle*). Red and purple rectangles denote CDS and UTR, respectively. Only male individuals were used in Z chromosome-related analysis.

Figure 4. Autosomal genome-wide association tests show differentiation between two *Helicoverpa armigera* lineages. (A) Manhattan plot of genome-wide polymorphisms. Peak regions with more than one highly differentiated SNPs (p value lower than 5.0×10^{-8}) were used for gene identification. The lower panel is genetic differentiation (Fst) between CC-S and CC-N lineages. (B)

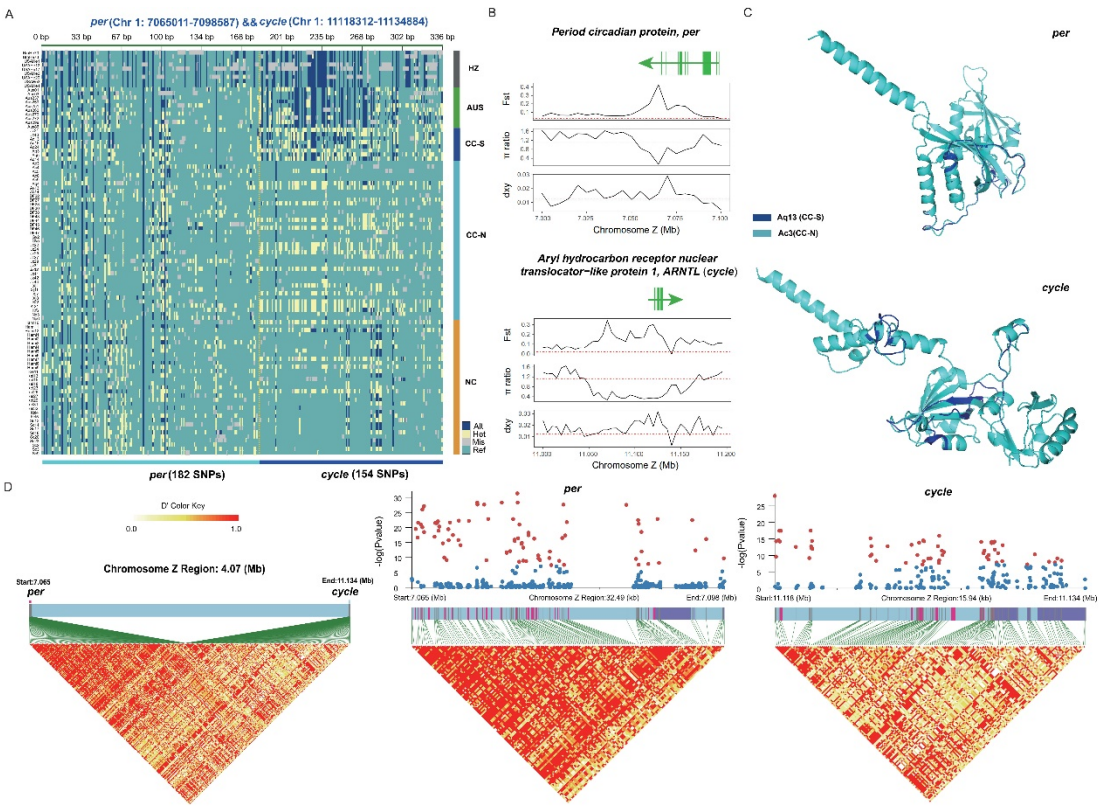
Characterization of six highly differentiated regions. π ratio indicates $\pi_{CC\text{-}N}/\pi_{CC\text{-}S}$, and d_{xy} is nucleotide divergence between CC-S and CC-N.

Figure 5. Widespread local adaptation in *Helicoverpa armigera* lineages. Selective sweep signals in different populations from Central China (CC), Northwest China (NC), and Australia (*H. a. conferta*, AUS) based on G12 (A) and CLR (B) statistics. Horizontal dash lines denote threshold values of G12/CLR over 99.95% autosomal windows in each lineage.


A
0.1
Bootstrap = 80
Bootstrap = 100
Helicoverpa zea (HZ)
H. armigera samples from Central China (CC-S)
H. a. conferta samples from Oceania (AUS)
H. armigera samples from Central China (CC-N)
H. armigera samples from Northwest China (NC)
B
AUS
CC-N
CC-S
NC
PC2(8.47%)
PC1(11.28%)
C
Topology 1 (①)
Topology 2 (②)
CC-N vs. CC-S
CC-N vs. AUS
Region 1
Region 2
Region 3
Region 4
Region 5
Fst
Chromosome Z (Mb)
R=0.8, p<2.2e-16
Fst(CC-S vs. CC-N)
Fst(AUS vs. CC-N)
D
CC-S & AUS
CC-N & AUS
fdM
Chr

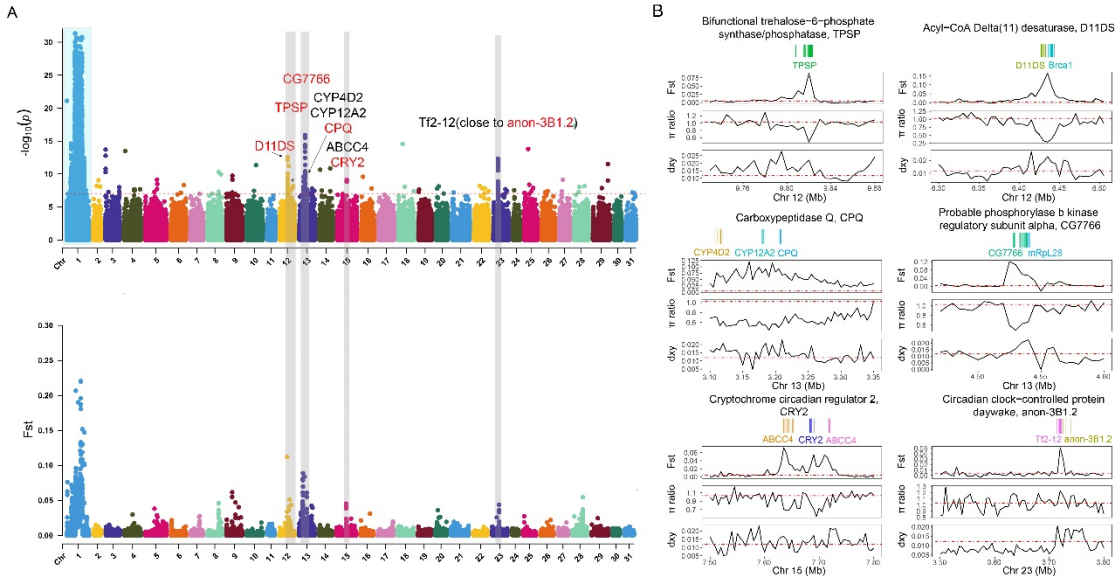

A
Boundary: 6.7-12.4 Mb
CC-N vs. CC-S
-log10(p)
CC-N vs. CC-S
π ratio
SLC25A25
per
Sstr4
ARNTL (cycle)
meigo
SPG7
SCAI
IGF2BP2
CC-N & AUS (D: 0.21, Z score: 6.31)
CC-S & AUS (D: 0.03, Z score: 3.72)
fdM
Chromosome Z (Mb)
B
Chromosome Z: 6704590-12391280
CC-S
CC-N
C
SNP distance(Kb) on Chromosome Z
D
Haplogroup
Inversion
Ancestral
Latitude
Longtitude

721
722

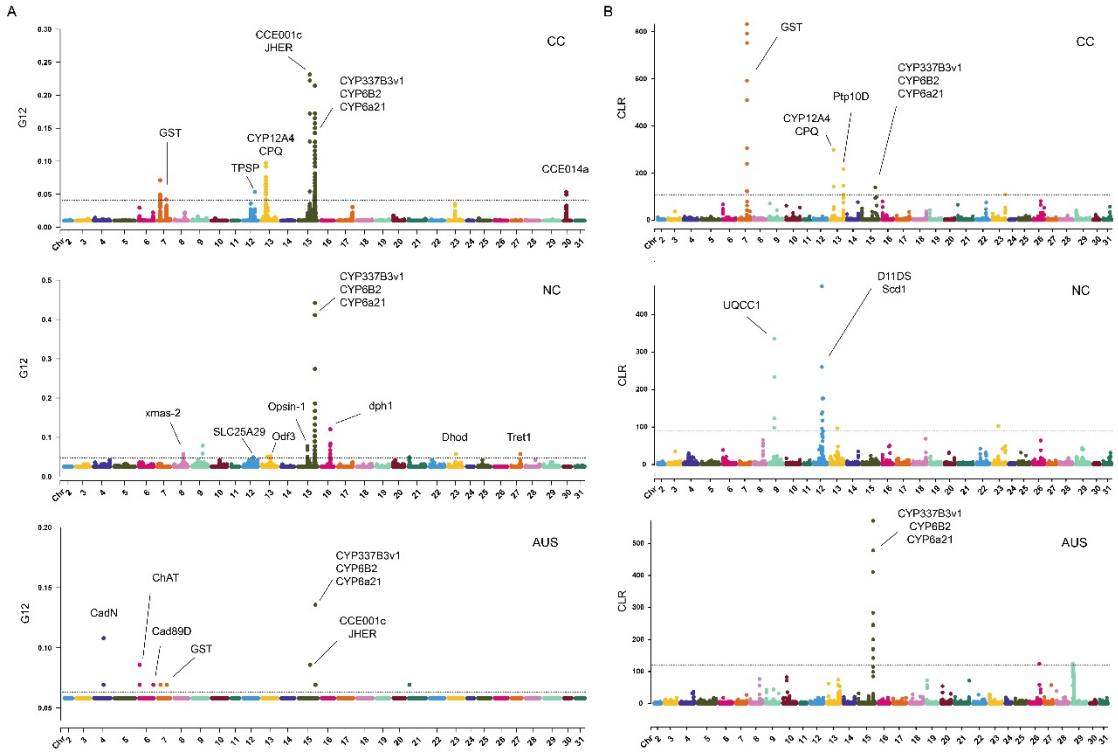

A
per (Chr 1: 7065011-7098587) && cycle (Chr 1: 11118312-11134884)
HZ
AUS
CC-S
CC-N
NC
Alt
Hot
Ms
Ref
per (182 SNPs)
cycle (154 SNPs)
B
Period circadian protein, per
Fst
π ratio
dxy
Chromosome Z (Mb)
Aryl hydrocarbon receptor nuclear translocator−like protein 1, ARNTL (cycle)
C
per
Aq13 (CC-S)
Ac3(CC-N)
cycle
D
D' Color Key
Chromosome Z Region: 4.07 (Mb)
per
cycle
-log(Pvalue)

723

724

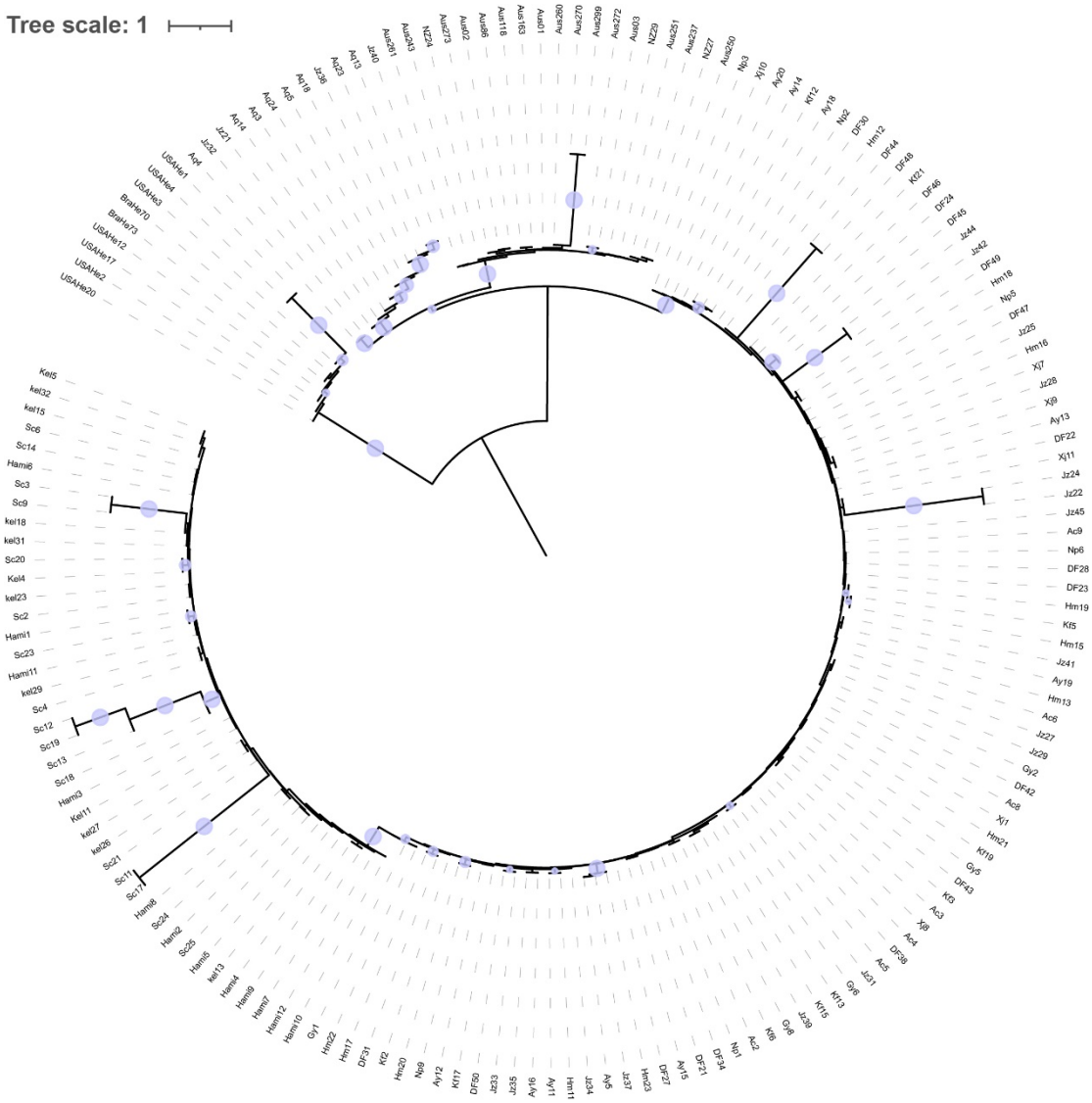

A
-log10(p)
CG7766
TPSP
CYP4D2
CYP12A2
CPQ
D11DS
ABCC4
CRY2
Tf2-12(close to anon-3B1.2)
Fst
Chr
B
Bifunctional trehalose−6−phosphate synthase/phosphatase, TPSP
TPSP
Acyl−CoA Delta(11) desaturase, D11DS
D11DS Brca1
Carboxypeptidase Q, CPQ
CYP4D2 CYP12A2 CPQ
Probable phosphorylase b kinase regulatory subunit alpha, CG7766
CG7766 mRpL28
Cryptochrome circadian regulator 2, CRY2
ABCC4 CRY2 ABCC4
Circadian clock−controlled protein daywake, anon-3B1.2
Tf2-12 anon-3B1.2
Fst
π ratio
dxy
Chr 12 (Mb)
Chr 13 (Mb)
Chr 15 (Mb)
Chr 23 (Mb)

725
726

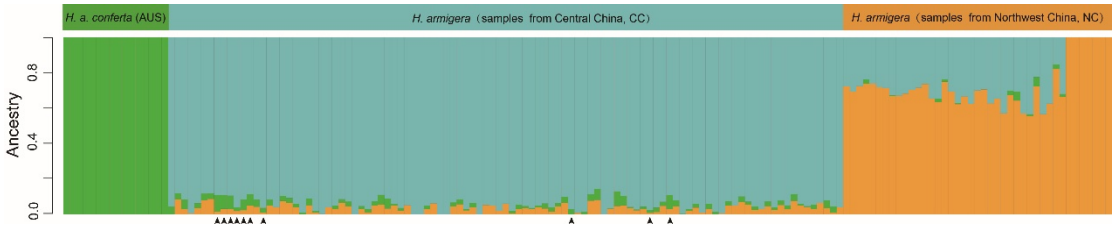

A
B
CC
NC
AUS
G12
CLR
CCE001c
JHER
CYP337B3v1
CYP6B2
CYP6a21
GST
CYP12A4
CPQ
TPSP
CCE014a
Ptp10D
xmas-2
SLC25A29
Odf3
Opsin-1
dph1
Dhod
Tret1
UQCC1
D11DS
Scd1
ChAT
CadN
Cad89D

727
728

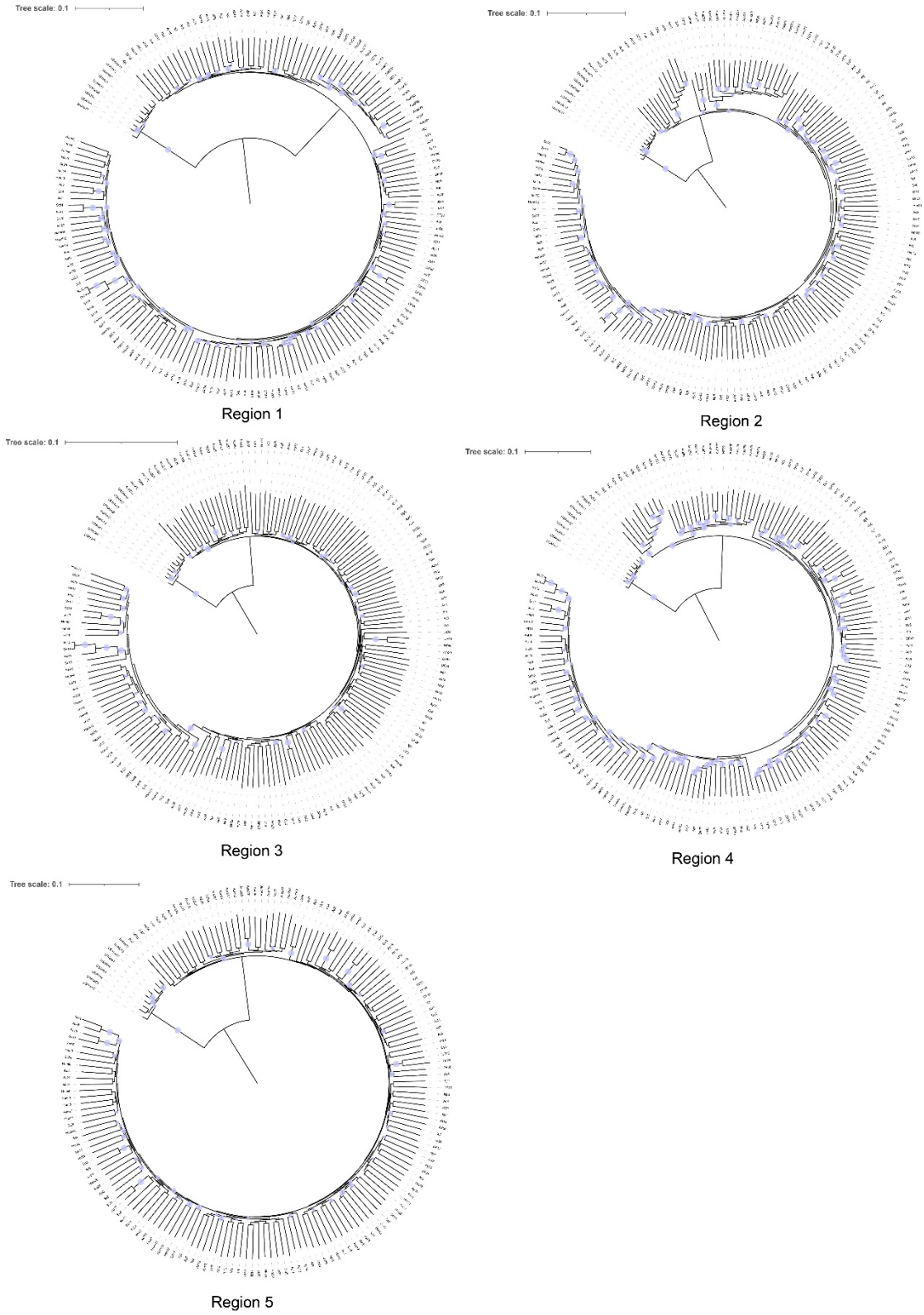
Supplementary Figure 1. Phylogenetic tree of Z chromosome based on ASTRAL-III.



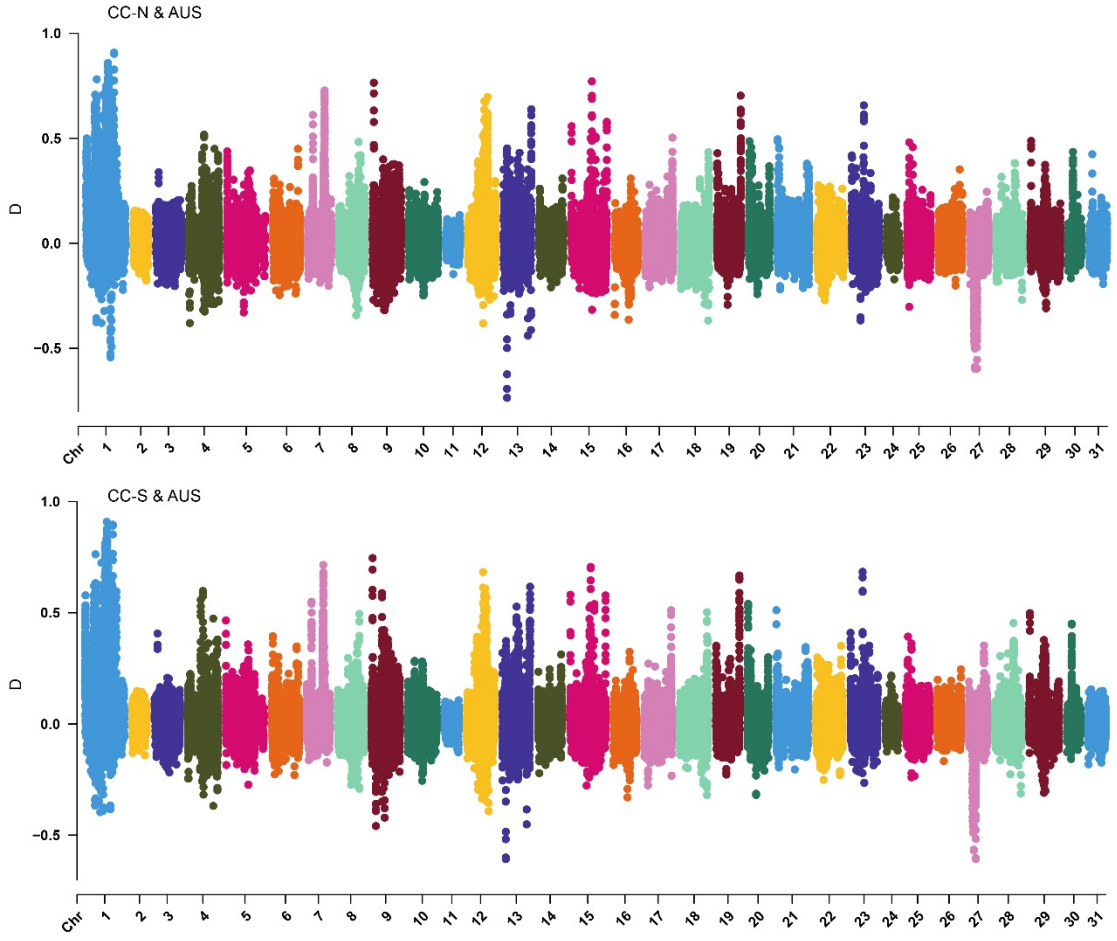
Supplementary Figure 2. Genetic structure of all *H. armigera* individuals based on autosomal polymorphisms.



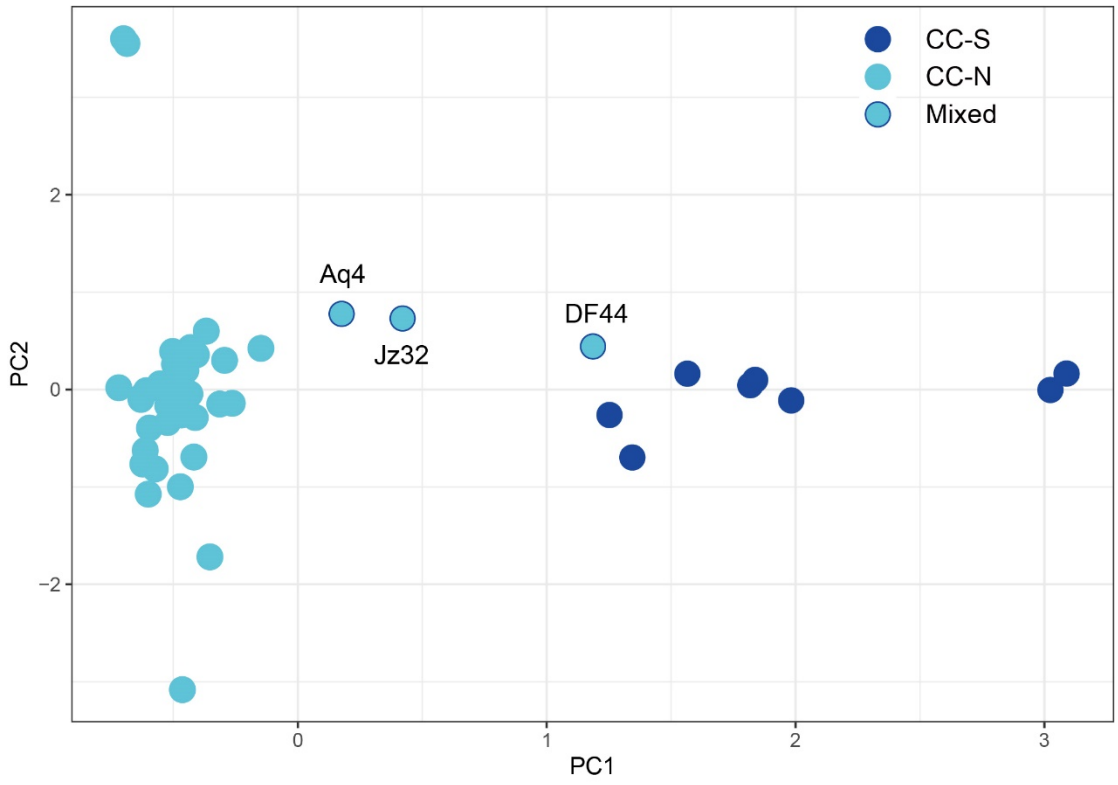
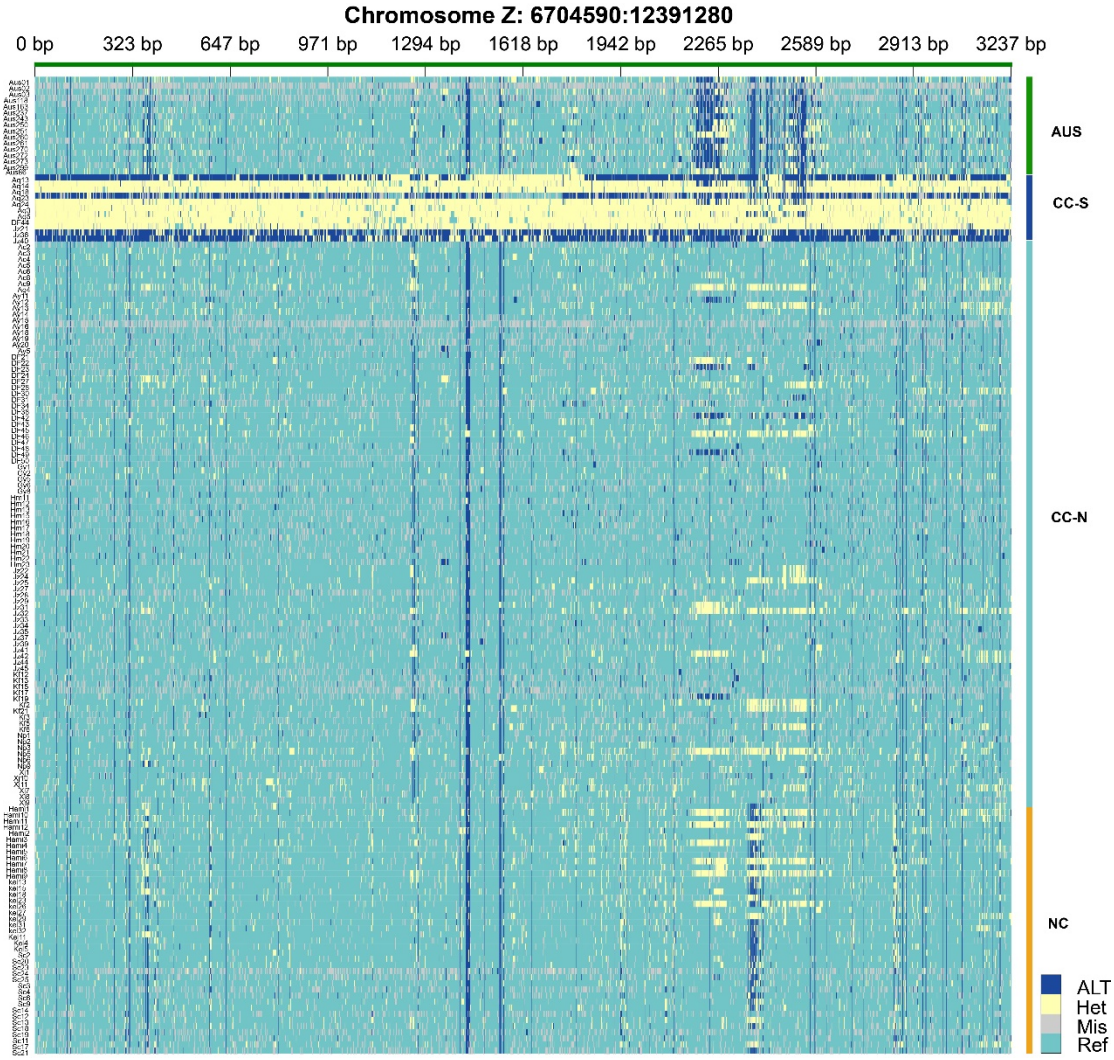
Supplementary Figure 3. Tree topologies of five regions in Z chromosome.



Supplementary Figure 4. Genome-wide calculation of D value between two *H. armigera* lineages (CC-N and CC-S) in Central China with Oceanian *H. a. conferta* lineage.



Supplementary Figure 5. Principal component analysis of *H. armigera* individuals sampled from Central China based on Z chromosomal polymorphisms.



Supplementary Figure 6. Genotype matrix of all sampled individuals in the putative inversion region of Z chromosome (6.7 - 12.4 Mb).