Up-regulation of cryptochrome 1 gene expression in cotton bollworm (*Helicoverpa armigera*) during migration over the Bohai Sea

Liyu Yang Equal first author, 1, Yingjie Liu Equal first author, 1, Philip Donkersley 2, Pengjun Xu Corresp. 1

1 Chinese Academy of Agricultural Sciences, Tobacco Research Institute, Qingdao, Shandong, China
2 Lancaster University, Lancaster Environment Centre, Lancaster, United Kingdom

Corresponding Author: Pengjun Xu
Email address: xupengjun@caas.cn

Cryptochromes (CRYs) are flavoproteins and play a pivotal role in circadian clocks which mediate behaviour of organisms such as feeding, mating and migrating navigation. Herein, we identified novel transcripts in *H. armigera* of six isoforms of *cry1* and seven isoforms of *cry2* by Sanger sequencing. Phylogenetic analysis showed that the transcripts of *cry1* and *cry2* align closely with other insect crys, indicating within-species divergence of *Hacry*. A dn/ds analysis revealed that the encoding sequence of the *cry1* was under purifying selection by a strong negative selection pressure whereas the *cry2* was less constraint and showed a less strong purification selection than *cry1*. In general, *Hacrys* were more abundantly transcribed in wild migrating populations than that in laboratory maintained populations, and expression of the *cry2* was lower than *cry1* in all samples tested. Moreover, when compared with the migrating parental population, offspring reared in laboratory conditions showed a significant reduction on transcription of the *cry1* but not *cry2*. These results strongly suggest that *cry1* was more related to the migration behaviour of *H. armigera* than *cry2*. 
Up-regulation of cryptochrome 1 gene expression in cotton bollworm (*Helicoverpa armigera*)
during migrating over the Bohai Sea

Liyu Yang¹, Yingjie Liu¹, Philip Donkersley², Pengjun Xu¹

¹Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, P.R. China
²Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK

Corresponding author

Pengjun Xu

Keyuanjing 4th Road, Qingdao, Shandong Province, 266101, P.R. China

E-mail: xupengjun@163.com.

The first two authors contributed equally to this paper.

Abstract

Cryptochromes (CRYs) are flavoproteins and play a pivotal role in circadian clocks which mediate behaviour of organisms such as feeding, mating and migrating navigation. Herein, we identified novel transcripts in *H. armigera* of six isoforms of *cry1* and seven isoforms of *cry2* by Sanger sequencing. Phylogenetic analysis showed that the transcripts of *cry1* and *cry2* align closely with other insect *crys*, indicating within-species divergence of *Hacry*. A *dn/ds* analysis revealed that the encoding sequence of the *cry1* was under purifying selection by a strong
negative selection pressure whereas the cry2 was less constraint and showed a less strong purification selection than cry1. In general, Hacrys were more abundantly transcribed in wild migrating populations than that in laboratory maintained populations, and expression of the cry2 was lower than cry1 in all samples tested. Moreover, when compared with the migrating parental population, offspring reared in laboratory conditions showed a significant reduction on transcription of the cry1 but not cry2. These results strongly suggest that cry1 was more related to the migration behaviour of H. armigera than cry2.

**Keywords:** Helicoverpa armigera; Migration; Cryptochromes; Transcripts; Expression analysis
Introduction

Cryptochromes (CRYs) genes were firstly characterized in *Arabidopsis thaliana* L. and then found in a wide range of organisms including animals and bacteria (Ahmad & Cashmore, 1993; Lin & Todo, 2005). According to their phylogenetic analysis, CRYs from different organisms can be divided into three subclasses: plant CRYs, animal CRYs and CRY-DASH (Selby & Sancar, 2006; Lin & Todo, 2005; Daiyasu et al., 2004). As receptors for blue and ultraviolet (UV-A) light, plant and animal CRYs mediate many light-dependent responses such as flowering in plants and circadian rhythms in animals (Emery et al. 1998; Griffin et al. 1999; Guo et al. 1998). Generally, CRYs have a high degree of sequence affinity to the members of DNA photolyase but do not possess any function on double-strand DNA repair, except for CRY-DASH, which is found in many organisms except for insects and has a single-stranded DNA-specific photolyase activity (Selby & Sancar, 2006).

In insects, two forms of CRYs, Drosophila-type CRY (CRY1, acting as a photoreceptor) and mammalian-type CRY (CRY2, acting as a transcriptional repressor), represent three kinds of circadian clocks: type 1, containing only CRY1 as in Drosophila; type 2, containing only CRY2 as in species from Hymenoptera and Coleoptera and type 3, containing both CRY1 and CRY2 as in mosquito and species in Lepidoptera (Zhu et al., 2005; Yuan et al., 2007; Zhu et al., 2008; Wang et al., 2013; Chang et al., 2019). Besides the regulation of circadian clocks, CRY proteins are also well-known for time-compensated sun compass navigation, such as in the monarch butterfly and *Drosophila* etc. (Zhu et al., 2008; Merlin et al., 2009; Yoshii & Ahmad, 2009;
Differently to diurnal butterflies and *Drosophila*, most moths are nocturnal. *Cry* genes have been reported in several moth species (Yuan et al., 2007; Chang et al., 2019). The cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae), is one of the most important long-distance migratory agricultural pests in China (Wu & Guo, 2005; Feng et al., 2009; Feng et al., 2003; Feng et al., 2004; Feng et al., 2005). Every year, the first generation of the cotton bollworm emerges in late June feeding on wheat in northern China (especially Shandong, Hebei, and Henan Provinces) and then the adults migrate to the northeast (Liaoning Province, China), where they reproduce one or two generations and the offspring migrates back to the southern China for overwintering (Wu et al., 1998; Feng et al., 2005; Wu & Guo, 2005; Wu et al., 2001). Recently, two types of CRYs were reported in *H. armigera*, named *Hacry1* and *Hacry2*, respectively (Yan et al., 2013).

Herein, we identified several novel isoforms of *cry* gene transcripts from *H. armigera*. We also investigated mutation and expression patterns of these genes with interests in: (1) evolutionary selection pressure acting on the two types of *cry* genes, and (2) comparison of transcription patterns between migrating and non-migrant cotton bollworms to obtain evidence whether CRYs are related to the migrating behaviour in *H. armigera*.

**Materials & Methods**

**Insects and culture conditions**
Migrating adult cotton bollworms (CD) were captured in different months in 2011 by a vertical pointing trap, on Beihuang Island in Bohai sea in Shandong province (38°23.200´N, 120°54.500´E). There were no local individuals of this species (Feng et al., 2003; Feng et al., 2004; Feng et al., 2005). Trapped insects were collected every 3 h starting at 18:00 pm (Beijing time) and were directly stored in liquid N\textsubscript{2}. The population of LF (originally captured in Langfang, Hebei province, in 2005) was used as control and reared in laboratory at 25 °C with a light (L): dark (D) photoperiod of 14 h: 10 h. These control insects were 3\textsuperscript{rd} day post eclosion adults exposed to natural light cycles in July 2011. To investigate the influence of heredity on Hacrys expression, offspring of CD were reared in laboratory (population CDF3) and collected at 15:00 pm under natural light using 3\textsuperscript{rd} day post eclosion adults in August 2011.

Additional three laboratory reared populations were used for obtaining Hacrys expression profiles in different population (Population-96s, captured in Xinxiang, Henan province, in 1996; Population-1463, captured in Xiajin, Shandong province, in 2004; Population-XXF1, captured in Xinxiang, Henan province, in 2011). Except for studying the developmental stages, all insects reared in lab were the 3\textsuperscript{rd} day instar adults under natural light.

**Cloning of full length Hacry gene transcripts**

Total RNA was isolated from *H. armigera* using TRIzol reagent (Invitrogen, Carlsbad, CA, USA); cDNA was synthesized using oligo(dT) and M-MLV Reverse Transcriptase (Promega, Madison, WL, USA). To identify transcripts of *cry* genes in *H. armigera*, degenerate primers
based on insect *cry* genes [including *Hacry1* (GenBank no.: GQ896502) and *Hacry2* (GQ896503)] were used to amplify partial segments of *Hacry* gene transcripts by nested RT-PCR (Table S1), using the following PCR program: 30 s at 94 ºC, 30 s at 55 ºC, and 60 s at 72 ºC for 40 cycles. To obtain complete transcript sequences of the *Hacry* genes, 5’ and 3’ RACE (Rapid Amplification of cDNA Ends) were performed using the SMART RACE cDNA Amplification Kit (Clontech, Palo Alto, CA, USA) with gene specific primers (Table S1).

To avoid possible artifacts, specific primers amplifying the complete ORFs (open reading frames) of *Hacrys* were designed according to the RACE results (Table S1). PCR program using the cDNA templates was: 30 s at 94 ºC, 30 s at 55 ºC, and 150 s at 72 ºC for 40 cycles. PCR products were purified, cloned into the pEASY-T Cloning Vector (TransGen, Beijing, China). Positive clones of plasmids of *Hacry* were purified and sequenced by using Sanger sequencing.

**Sequence alignment and phylogenetic analyses**

Percentage identities of the nucleotide and amino acid sequences were calculated using CLUSTAL W program (Thompson et al., 1994). The longest obtained sequences of *cry1* (*Hacry1-4*) and *cry2* (*Hacry2-6*) were selected to use together with the other insect *cry1*s and *cry2*s to determine selection pressures (Table S2).

The sequences were aligned and phylogenetically analyzed using maximum likelihood (ML) methodology implemented in RAxML 7.3.2 (Stamatakis, 2006) under the GTR+G substitution model. For estimating nodal support, nonparametric bootstrap proportions (Felsenstein, 1985)
with 1000 pseudoreplicates were used. For the phylogenetic inference, we divided the dataset into three partitions according to codon position in each gene (1\textsuperscript{st}/2\textsuperscript{nd}/3\textsuperscript{rd} codon). The multiple partitions controlled for heterogeneity across dataset, such as variation in substitution rates. Three replicates were conducted for each analysis to assess whether runs failed to converge upon the optimal posterior distribution and whether likelihood values, branch lengths, tree topology or posterior probabilities differed among runs. We made two kind of trees. One contained only butterfly, moth and outgroup for selection analysis, and another one contained much more insect species for phylogenetic purpose.

Branch and Branch-site test of selection

Selection can promote the evolution of sensory systems and make organisms to adapt to local ecological conditions. To investigate the potential signal of positive selection acting on the cry genes between nocturnal moths and diurnal butterflies, Maximum Likelihood approach (Nielsen & Yang, 1998) was employed to test for differences in selection pressure, using the CODEML program of PAML version 4.5 (Yang, 2007). Branch models and branch-site models were employed to detect positive selection on the lineages. These tests of selection were phylogeny-based tests with requirements of unrooted input-trees. The RAxML trees mentioned above were converted to unrooted trees. For branch models, five hypotheses were evaluated: 1) one dn/ds ratio for all branches; 2) dn/ds ratio = 1 for all branches; 3) moth lineage and butterfly
lineage have different $dn/ds$ ratio ($\omega_2$ and $\omega_1$); 4) neutral evolution for moth ($\omega_2 = 1$); and 5) the free-ratio model with free $dn/ds$ ratio for each branch. For branch-site modes, the moth lineage was defined as foreground, and the rest lineages were defined as background branch, which were specified in the tree file by using branch labels. Likelihood ratio test (LRT) was used to investigate if the alternative model, indicating positive selection, was superior to the null model.

To investigate non-synonymous mutations within $Hacry1$ and $Hacry2$, we calculated the average $dn/ds$ ratio under one ratio model in PAML. In this analysis, all the detected isoforms of the $cry1$ and $cry2$ transcripts were used.

**Real-time PCR analysis**

With $\beta$-actin (GenBank accession no.: X97614) as the reference gene, Real-time PCR (qPCR) reactions were carried out with the first strand cDNAs using the TaqMan method in 20 $\mu$l reaction agent comprised of 1 $\mu$l of template cDNA, 2*Premix Ex Taq™ (Takara, Japan), 0.2 $\mu$M each primer and 0.4 $\mu$M probe (Table S1) on a 7500 Fast Real-time PCR System (Applied Biosystems). Thermal cycling conditions were: 45 cycles of 95 °C for 15 s, 60 °C for 34 s. cDNA sample of each group was replicated 3 times. At least three groups of individuals were tested for each data-point. Fold differences of $Hacry1$ or $Hacry2$ transcripts were calculated according to the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen, 2001).

**Statistical analysis**
Statistical analyses were conducted using the STATA package (version 9.0) and GraphPad InStat 3 (GraphPad Software, Inc., San Diego, CA, USA). The Student’s t-test and/or ANOVA were used to determine the significant differences.

Results

Transcripts of cry genes in *H. armigera*

By RT-PCR and RACE method, the complete sequences of *cry1* and *cry2* gene transcripts with variable 5’ and 3’ end sequences were obtained from *H. armigera* (Fig. 1). BLAST searches performed at NCBI revealed that these *Hacry* transcripts were highly similar to *cry1* or *cry2* from the other insects, respectively. The encoding nucleotide sequences of the *cry1* of *H. armigera* displayed homology of 91-99.8% with each other whereas the *cry2* encoding sequences displayed much more polymorphisms including large deletions (Table S3 and S4, Fig. 1 and Fig. S1). All deduced amino acid sequences of the *cry1* contained fragments of the C-terminal extension (CCE) and the conserved photolyase homology region (PHR) which included a flavin adenine dinucleotide (FAD) domain and an additional pterin derivative domain (methenyltetrahydrofolate, MTHF) (Fig. 1 and Fig. S1). However, *Hacry2-1, Hacry2-2, Hacry2-3* and *Hacry2-4* lacked a complete PHR encoding region (Fig. 1 and Fig. S1).
Phylogenetic analysis and selection pressure

Tree based analysis indicated that transcripts of cry1 and cry2 cloned in our study clustered with the one from species in Lepidoptera (Fig. 2). The free-ratio model provided a better fit for the two cry genes than the other four models. No significant differences in $dn/ds$ were detected between the clades of moths and butterflies (Table 1). The branch-specific $\omega$ values of moths were 0.0287 and 0.0219 for cry1 and cry2, respectively. The corresponding $\omega$ values of butterflies were 0.0409 and 0.0130, respectively (Fig. 2). Interestingly, for cry1, the $dn/ds$ ratio of moths was lower than that of butterflies, suggesting increased constraint on nonsynonymous substitutions. In contrast, for cry2, the $dn/ds$ ratio of moths was much higher than that of butterflies. This suggested that the moth cry2 genes might have accumulated higher proportion of nonsynonymous mutations than the butterfly cry2 in the evolutionary history, suggesting a less strong purification selection than cry1. However, no positive selection could be detected by PAML.

Then we calculated the $dn/ds$ for all the transcript isoforms of the cry1 and the cry2 (as shown in Table S3 and Table S4). Surprisingly, highly contrasting results were obtained. The $dn/ds$ of Hacry1 transcripts was 0.0082, which is in line with the expectation based on results of the cry1. Four replacement changes and 127 synonymous changes were detected. However, for the cry2, $dn/ds$ increased dramatically to 0.684. Fourteen replacement changes and eight synonymous changes were found. It indicated that nonsynonymous mutations were accumulated with much higher rate in these cry2 isoforms than in the cry1 genes.
Expressions Hacrys in different tissues, adult instars, populations and months

According to our preliminary experiment and experience investigating gene expression pattern in *H. armigera* (Mao et al., 2007; Sui et al., 2009; Zhang et al., 2011), \( \beta \)-actin was selected as reference gene in this study. Individuals of females and males from CD were used to detect tissue distribution of Hacrys mRNA transcripts including antenna, head, thorax, abdomen and leg, by quantitative RT-PCR analyses. Expressions were detected in all examined tissues but no statistically significant difference was detected among them (ANOVA: for Hacry1, in day time, \( F(4,19) = 0.92, P = 0.4731 \); in night time: \( F(4,18) = 0.57, P = 0.6862 \); for Hacyr2, in day time, \( F(4,19) = 0.61, P = 0.6628 \), in night \( F(4,18) = 0.57, P = 0.6905 \)) (Fig. 3A, B).

Initially, the expression levels of both Hacry1 and Hacry2 were not significantly different in female or male adults. However, the expression level of Hacry1 mRNA was significantly higher in male than female adults during day 3, day 5, and day 7 (students t test, day3, \( t = 2.1696, p = 0.0365 \); day5, \( t = 1.9515, p = 0.0494 \); day 7, \( t = 2.3900, p = 0.0270 \)). The expression level of Hacry2 mRNA was significantly higher in males on day 1 (students t test, \( t = 2.1190, p = 0.0392 \)) and day5 (students t test, \( t = 3.6869, p = 0.0071 \)) than female adults (Fig. 3C, D), which might be due to the differences of rhythm and biological function between females and males.

From CD female and male adults, the Hacry1 and Hacry2 mRNA level were significant higher than other populations including LF, 96s, 1463 and XX (ANOVA: for Hacry1, \( F(5,30) = 6.63, P = 0.0003 \); for Hacry2, \( F(5,26) = 2.69, P = 0.0436 \)) (Fig 3E, F). The Hacry1 genes in CD were...
significantly differently expressed among individuals from different months (F(4,43) = 3.337, P
= 0.0182), however, there were no significant differences in the expression levels of \(\textit{Hacry}2\) genes (F(4,41) = 0.4738, P = 0.7545) (Fig 3G, H).

To investigate the effect of heredity, we detected the mRNA levels of \(\textit{Hacrys}\) using CD third generation reared indoor (CDF3). The level of \(\textit{Hacry}1\) mRNA was significantly down-regulated in CDF3 (Bonferroni multi comparison, \(p = 0.009\) between CD and CDF3) and similar to other populations detected in this study (Fig. 3E, F). However, the levels of \(\textit{Hacry}2\) mRNA were stable between CD and CDF3 (Bonferroni multi comparison, \(p = 0.686\) between CD and CDF3), which were significantly higher than other populations except for LF (Fig. 3E, F).

Diel changes in the expression of \(\textit{cry}1\) and \(\textit{cry}2\) from \(\textit{H. armigera}\) in migrating and reared populations

We selected CD and an artificial reared population (LF, 3\(^{rd}\) day moths post eclosion) to detect the diurnal fluctuation of the expression levels of \(\textit{Hacry}1\) and \(\textit{Hacry}2\). qPCR analysis revealed a significant fluctuation of the two \(\textit{cry}\)s in both CD and LF. The two populations had a similar rhythm of \(\textit{Hacry}1\) expression level from 0:00 to 12:00. However they show different profiles, with peaks and troughs at 12:00 and 9:00 for outdoor population (CD), 0:00 and 18:00 for indoor population (LF), respectively (Fig. 4A, B, C, D).

The diurnal-rhythm of \(\textit{Hacry}2\) mRNA was different between CD and LF, which was low during the day in CD while was high in LF at 0:00 and 18:00 (Fig. 4A, B, C, D). The amount of \(\textit{Hacry}1\) mRNA expression in CD was significantly higher than in LF in all of the detected
samples except for at 6:00 and 15:00 (Fig. 4E, F). The amount of Hacry2 mRNA did not show significant temporal variance in LF, but was significantly higher in CD at 0:00, 12:00 and 18:00 than in LF (Fig. 4E, F).

Discussion

Recently, Yan et al. (2013) reported two types of cry genes from H. armigera, each with one transcript respectively. Here, we identified six novel transcripts of cry1 and seven novel transcripts of cry2. The transcripts of HaCRY1s contained a conserved PHR motif and the alternative splicing region located in the CCE domain which was also found in cry1s of M. brassicae and Spodotera exigua, suggesting that the cry1 genes might be evolving at a special period in Lepidoptera (Merlin et al., 2006; Berndt et al., 2007).

Similarly, the seven detected isoforms of the Hacry2 are diverse in sequence length and nucleotides, which may be from alternative splicing of the same gene or different genes derived from duplication events. The CRY2s of insects act as a transcriptional repressor in a light-independent manner (Zhu et al., 2005; Yuan et al., 2007; Ikeno et al., 2011). The CCE domain is key to CRYs maintenance and can be active in the absence of a PHR domain (Fankhauser & Ulm, 2011; Mazzotta et al., 2013). Therefore, we suggest that HaCRY2-1, HaCRY2-2, HaCRY2-3 and HaCRY2-4 might be function like other HaCRY2s, despite lacking a complete PHR domain.

Transcription analyses of cry1 and cry2 from insects including D. melanogasta, A. mellifera,
Riptortus pedestris, M. brassicae, Danaus plexippus revealed that each insect species contained only one copy of cry1 and/or cry2 without alternative splicing (Emery et al., 1998; Zhu et al., 2005; Merlin et al., 2006; Rubin et al., 2006; Ikeno et al., 2008). The cry genes of H. armigera were the most polymorphic among all of the known insect cry genes, which might be related to the lifestyle and environmental adaptability of H. armigera. dn/ds analysis did not detect any positive selective evolution between the cry genes of nocturnal moths and diurnal butterflies. However, the dn/ds analysis of H. armigera transcripts suggested that the encoding sequence of the Hacry1 was negatively selected possibly due to functional constraint, but the cry2 showed less strong selection than cry1. For the branch-site model, the neutral model could not be rejected. These results suggested that the Hacry1 might play a more important role than the Hacry2 in H. armigera.

mRNA levels of Hacrys were neither tissue-specific nor developmental-stage-specific (Yan et al., 2013). Based on ovary development, most of the individuals from CD population were 3-4 day instars after eclosion. Therefore, we used whole individuals of 3 day adult instar in laboratory populations to compare with wild CD individuals. To our surprise, the mRNA levels of the cry1 were significantly higher in CD than LF while the Hacry2 was not as strong as the Hacry1.

To confirm that the levels of Hacrys transcripts were correlated to the migrating behavior, we reared offspring of CD for three generations (CDF3) in laboratory and investigated the mRNA levels of Hacrys using the 3rd day adults after eclosion. The expression level of Hacry1 was significantly down-regulated in CDF3 population, suggesting that HaCRY1 might play a
role in the migrating behaviour of *H. armigera*. Detected expression of *Hacry2*, however, was not significantly changed in CDF3, suggesting that the *cry2* might not have specific function in the migrating behaviour of *H. armigera*. Recently, CRY proteins were reported as magnetoreceptors and using light signal for their function in navigation (Baik et al., 2017; Marley et al., 2014; Bazalova et al., 2016; Rodgers & Hore, 2009; Gegear et al., 2010). Our results warrant further investigation on the functions of CRY proteins in the migrating behaviour of moths, in areas of both circadian biology and magnetochemical biology. This potential role in migration may have important implications for our understanding of migratory behaviour of key crop pest species.

Conclusions

We found the *Hacry1* and *Hacry2* of *H. armigera* occurred alternative splicing which generated six and seven transcripts, respectively. Selection pressure analysis suggested that the *Hacry1* was under purifying selection by a strong negative selection pressure whilst the *Hacry2* was less constrained and showed a less strong purification selection than *cry1*. At mRNA levels by qPCR, *Hacrys* were more abundantly transcribed in wild migrating populations than in laboratory maintained populations, and expression of the *Hacry2* was lower than *Hacry1* in all samples tested. Interestingly, when compared with the migrating parental population, offspring reared in laboratory conditions showed a significant reduction on transcription of the *Hacry1* but not *Hacry2*. These results suggest that *Hacry1* is more related to the migration of *H. armigera*.
than Hacry2.

Table

Table 1. Selective patterns for Cryptochrome genes.

Figure Legends

Figure 1. Transcripts of Hacry1 (A) and Hacry2 (B) in H. armigera. “■” stands for the open reading frame (ORF) region. “□” stands for untranslated region (UTR). “-” stands for gap. The conserved MTHF and FAD binding regions were shown on the top. The names and GenBank accession numbers of transcripts were shown on the right.

Figure 2. Phylogenetic reconstruction of cry genes in species from Lepidoptera based on maximum likelihood. (A) tree based analysis of cry1 genes using the ones of Drosophila melanogaster and Acyrthosiphon pisum as outgroup. (B) tree based analysis of cry2 genes using the one of Apid mellifera as outgroup. Values on the nodes are the nonparametric bootstrap proportions (MLBPs). Moths denoted in blue and butterflies in red. Branch-specific \( \omega \) values are shown on nodes of the common ancestors. The names and GenBank accession numbers of crys were shown in Table S2.

Figure 3. Relative expression level of Hacry1 and Hacry2 in different tissues (n=3-5), day instar stage (n=4), population (n=4-8) and months (n=8-11) of adults. (A) Relative expression level of Hacry1 in different tissues of adults. (B) Relative expression level of Hacry2 in different tissues of adults. (C) Relative expression level of Hacry1 in different day instar stage of adults. (D)
Relative expression level of *Hacry2* in different day instar stage of adults. (E) Relative expression level of *Hacry1* in different population of adults. (F) Relative expression level of *Hacry2* in different population of adults. (G) Relative expression level of *Hacry1* in different months of adults. (H) Relative expression level of *Hacry2* in different months of adults. Mean ± SE. The “*” and “**” denote statistical significance of the expression levels in (C) and (D) (p<0.05 and p<0.01, respectively). The different letters were used to show significant difference in (E) and (F) (p<0.05, by Student’s t-test).

**Figure 4.** The expression level of *Hacrys* in CD and LF population. (A) Diel changes of relative expression level of *Hacry1* in CD population. (B) Diel changes of relative expression level of *Hacry2* in CD population. (C) Diel changes of relative expression level of *Hacry1* in LF population with moths at three-day after eclosion. The different letter indicates significant differences between groups (p<0.05, by ANOVA). (D) Diel changes of relative expression level of *Hacry2* in LF population with moths at three-day after eclosion. The different letter indicates significant differences between groups (p<0.05, by ANOVA). (E) Relative expression level of *Hacry1* between CD and LF individuals. Mean ± SE. The “*” and “**” denote statistical significance of the expression levels between CD and LF individuals at the same time (p< 0.05 and p< 0.01, respectively, n= 3-4). (F) Relative expression level of *Hacry2* between CD and LF individuals. Mean ± SE. The “*” and “**” denote statistical significance of the expression levels between CD and LF individuals at the same time (p< 0.05 and p< 0.01, respectively, n= 3-4).

**Supplemental Information**
Supplemental material 1

Table S1. Listing of primers used in this study.

Supplemental material 2

Table S2. Listing of insect species used to determine selection pressures based on cry1s and cry2s sequences.

Supplemental material 3

Table S3. Nucleotide identities among ORFs of the Hacry1 from H. armigera. The length of aligned fragments and the numbers of variant nucleotides were shown in bracket.

Supplemental material 4

Table S4. Nucleotide identities among ORFs of the Hacry2 from H. armigera. The length of aligned fragments and the numbers of variant nucleotides were shown in bracket.

Supplemental material 5

Fig S1. Amino acid sequence alignment of the HACRY1s (A) and HACRY2s (B) in H. armigera. Region shaded in red correspond to the PHR region. Gray shown the CCE region. “—” stands for MTHF and FAD binding region, in which “=” stands for the shared region by MTHF and FAD. “.” stands for identical sites with the first row sequence. “-” stand for gap. dCRY: Drosophila melanogaster (accession no.: AAK92938), MbCRY1: Mamestra brassicae (accession no.: AAY23345), SeCRY1: Spodoptera exigua (accession no.: ADY17887), AgCRY2: Anopheles gambiae (accession no.: XP_313179).

Supplemental material 6
Dataset S1. The original data used to plot Figure 3 and Figure 4.

Additional Information and Declarations

Author Contributions

Pengjun Xu, Liyu Yang and Yingjie Liu designed research;
Pengjun Xu, Liyu Yang, Yingjie Liu, He Yuan, Bin Lu and Yonghao Dong performed research;
Pengjun Xu analyzed data;
Pengjun Xu, Philip Donkersley, Liyu Yang and Yingjie Liu wrote the paper.

Data Availability

The following information was supplied regarding data availability:
The raw data are provided in a Supplemental File.

Acknowledgements

We would like to thank Dr. Xiaowei Fu for collecting samples on Beihuang Island.

References

Ahmad M, Cashmore AR. 1993. HY4 gene of A. thaliana encodes a protein with characteristics of a blue-light photoreceptor. Nature 366(6451):162.
Baik LS, Fogle KJ, Roberts L, Galschiodt AM, Chevez JA, Recinos Y, Nguy V, Holmes TC. 2017. CRYPTOCHROME mediates behavioral executive choice in response to UV light.
Cryptochrome 2 mediates directional magnetoreception in cockroaches. *Proceedings of the National Academy of Sciences* 113(6):1660-1665.

Berndt A, Kottke T, Breitkreuz H, Dvorsky R, Hennig S, Alexander M, Wolf E. 2007. A novel photoreaction mechanism for the circadian blue light photoreceptor *Drosophila* cryptochrome. *Journal of Biological Chemistry* 282(17):13011-13021.

Chang H, Guo J, Fu X, Wang M, Hou Y, Wu K. 2019. Molecular Characterization and Expression Profiles of Cryptochrome Genes in a Long-Distance Migrant, *Agrotis segetum* (Lepidoptera: Noctuidae). *Journal of Insect Science* 19(1):8.

Daiyasu H, Ishikawa T, Kuma K, Iwa Si, Todo T, Toh H. 2004. Identification of cryptochrome DASH from vertebrates. *Genes Cells* 9(5):479-495.

Emery P, So WV, Kaneko M, Hall JC, Rosbash M. 1998. CRY, a Drosophila clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* 95(5):669-679.

Fankhauser C, Ulm R. 2011. Light-regulated interactions with SPA proteins underlie cryptochrome-mediated gene expression. *Gene&Development* 25(10):1004-1009.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39(4):783-791.

Feng H, Wu K, Cheng D, Guo Y. 2003. Radar observations of the autumn migration of the beet
armyworm *Spodoptera exigua* (Lepidoptera: Noctuidae) and other moths in northern China.

*Bulletin of Entomological Research* 93(2):115-124.

Feng H, Wu K, Cheng D, Guo Y. 2004. Northward migration of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and other moths in early summer observed with radar in northern China. *Journal of Economic Entomology* 97(6):1874-1883.

Feng H, Wu K, Cheng D, Guo Y. 2005. Return migration of *Helicoverpa armigera* (Lepidoptera: Noctuidae) during autumn in northern China. *Bulletin of Entomological Research* 95(4):361-370.

Feng H, Wu X, Wu B, Wu K. 2009. Seasonal migration of *Helicoverpa armigera* (Lepidoptera: Noctuidae) over the Bohai Sea. *Journal of Economic Entomology* 102(1):95-104.

Gegear RJ, Foley LE, Casselman A, Reppert SM. 2010. Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism. *Nature* 463(2010):804-807.

Griffin EA, Staknis D, Weitz CJ. 1999. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* 286(5440):768-771.

Guo H, Yang H, Mockler TC, Lin C. 1998. Regulation of flowering time by Arabidopsis photoreceptors. *Science* 279(5355):1360-1363.

Ikeno T, Numata H, Goto SG. 2011. Photoperiodic response requires mammalian-type cryptochrome in the bean bug *Riptortus pedestris*. *Biochemical and biophysical research communications* 410(3):394-397.

Ikeno T, Numata H, Goto SG. Molecular characterization of the circadian clock genes in the bean bug, *Riptortus pedestris*, and their expression patterns under long- and short-day conditions. *Gene* 419(1-2):56-61.
Lin C, Todo T. 2005. The cryptochromes, *Genome Biology* 6(5):220.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2 (-Delta Delta C(T)) method. *Methods* 25(4):402-408.

Mao Y, Cai W, Wang J, Hong G, Tao X, Wang L, Huang Y, Chen X. 2007. Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. *Nature biotechnology* 25(11):1307-1313.

Marley R, Giachello CN, Scrutton NS, Baines RA, Jones AR. 2014. Cryptochrome-dependent magnetic field effect on seizure response in *Drosophila* larvae. *Scientific Reports.* 4:5799.

Mazzotta G, Rossi A, Leonardi E, Mason M, Bertolucci C, Caccin L, Spolaore B, Martin AJM, Schlichting M, Grebler R, Helfrich-Förster C, Mammi S, Costa R, Tosatto SCE. Fly cryptochrome and the visual system. *Proceedings of the National Academy of Sciences.* 110(15):6163-6168.

Merlin C, François M, Queguiner I, Maïbèche-Coisné M, Jacquin-Joly E. 2006. Evidence for a putative antennal clock in *Mamestra brassicae*: molecular cloning and characterization of two clock genes-period and cryptochrome-in antennae. *Insect Molecular Biology* 15(2):137-145.

Merlin C, François M, Queguiner I, Maïbèche-Coisné M, Jacquin-Joly E. 2006. Evidence for a putative antennal clock in *Mamestra brassicae*: molecular cloning and characterization of two clock genes-period and cryptochrome-in antennae. *Insect Molecular Biology* 15(2):137-145.

Merlin C, Gegear RJ, Reppert SM. 2009. Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies. *Science* 325(5948):1700-1704.

Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites
and applications to the HIV-1 envelope gene. *Genetics* 148(3):929-936.

Rodgers CT, Hore PJ. 2009. Chemical magnetoreception in birds: the radical pair mechanism. *Proceedings of the National Academy of Sciences*. 106(2):353-360.

Rubin EB, Shemesh Y, Cohen M, Elgavish S, Robertson HM, Bloch G. 2006. Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. *Genome Research* 16(11):1352-1365.

Selby CP, Sancar A. 2006. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proceedings of the National Academy of Sciences* 103(47):17696-17700.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690.

Sui Y, Liu X, Chai L, Wang J, Zhao X. 2009. Characterization and influences of classical insect hormones on the expression profiles of a molting carboxypeptidase A from the cotton bollworm (*Helicoverpa armigera*). *Insect Molecular Biology* 18(3):353-363.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22):4673-4680.

Wang B, Xiao J, Bian S, Gu H, Huang D. 2013. Adaptive evolution of vertebate-type cryptochrome in the ancestors of Hymenoptera. *Biology Letters* 9(1):20120958.

Wu K, Cheng D, Xu G, Zhai B, Guo Y. 2001. Radar observation of autumn migration of insects in
northern China. Acta Ecologica Sinica 21(11):1833-1838.

Wu K, Guo Y. 2005. The evolution of cotton pest management practices in China. Annual Review. Entomology 50:31-52.

Wu K, Xu G, Guo Y. 1998. Observations on migratory activity of cotton bollworm moths across the Bohai Gulf in China. Acta Phytophylacica Sinica 25(4):337-340.

Yan S, Ni H, Li H, Zhang J, Liu X, Zhang Q. 2013. Molecular cloning, characterization, and mRNA expression of two Cryptochrome genes in Helicoverpa armigera (Lepidoptera: Noctuidae). Journal of economic entomology 106(1):450-462.

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13(5):555-556.

Yoshii T, Ahmad M, Helfrich-Forster C. 2009. Cryptochrome mediates light-dependent magnetosensitivity of Drosophila's circadian clock. Plos Biology 7(4):e1000086.

Yuan Q, Metterville D, Briscoe AD, Reppert SM. 2007. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Molecular biology and evolution 24(4): 948-955.

Zhang T, Gu S, Wu K, Zhang Y, Guo Y. 2011. Construction and analysis of cDNA libraries from the antennae of male and female cotton bollworms Helicoverpa armigera (Hübner) and expression analysis of putative odorant-binding protein genes. Biochemical and Biophysical Research Communications 407(2):393-399.

Zhu H, Sauman I, Yuan Q, Casselman A, Emery-Le M, Emery P, Reppert SM. 2008. Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may
underlie sun compass navigation. *Plos Biology* 6:e4.

Zhu H, Yuan Q, Briscoe AD, Froy O, Casselman A, Reppert SM. 2005. The two CRYs of the butterfly. *Current Biology* 15(23):R953-R954.
Figure 1

Transcripts of *Hacry1* (A) and *Hacry2* (B) in *H. armigera*.

“■” stands for the open reading frame (ORF) region. “□” stands for untranslated region (UTR). “-” stands for gap. The conserved MTHF and FAD binding regions were shown on the top. The names and GenBank accession numbers of transcripts were shown on the right.
Figure 2

Phylogenetic reconstruction of *cry* genes in species from Lepidoptera based on maximum likelihood.

(A) tree based analysis of *cry1* genes using the ones of *Drosophila melanogaster* and *Acyrthosiphon pisum* as outgroup. (B) tree based analysis of *cry2* genes using the one of *Apid mellifera* as outgroup. Values on the nodes are the nonparametric bootstrap proportions (MLBPs). Moths denoted in blue and butterflies in red. Branch-specific $\omega$ values are shown on nodes of the common ancestors. The names and GenBank accession numbers of *crys* were shown in Table S2.
Figure 3

Relative expression level of *Hacry1* and *Hacry2* in different tissues (n=3-5), day instar stage (n=4), population (n=4-8) and months (n=8-11) of adults.

(A) Relative expression level of *Hacry1* in different tissues of adults. (B) Relative expression level of *Hacry2* in different tissues of adults. (C) Relative expression level of *Hacry1* in different day instar stage of adults. (D) Relative expression level of *Hacry2* in different day instar stage of adults. (E) Relative expression level of *Hacry1* in different population of adults. (F) Relative expression level of *Hacry2* in different population of adults. (G) Relative expression level of *Hacry1* in different months of adults. (H) Relative expression level of *Hacry2* in different months of adults. Mean ± SE. The “*” and “**” denote statistical significance of the expression levels in (C) and (D) (*p* < 0.05 and *p* < 0.01, respectively). The different letters were used to show significant difference in (E) and (F) (*p* < 0.05, by Student’s *t*-test).
Figure 4

The expression level of *Hacrys* in CD and LF population.

(A) Diel changes of relative expression level of *Hacry1* in CD population. (B) Diel changes of relative expression level of *Hacry2* in CD population. (C) Diel changes of relative expression level of *Hacry1* in LF population with moths at three-day after eclosion. The different letter indicates significant differences between groups ($p<0.05$, by ANOVA). (D) Diel changes of relative expression level of *Hacry2* in LF population with moths at three-day after eclosion. The different letter indicates significant differences between groups ($p<0.05$, by ANOVA). (E) Relative expression level of *Hacry1* between CD and LF individuals. Mean ± SE. The "*" and "**" denote statistical significance of the expression levels between CD and LF individuals at the same time ($p<0.05$ and $p<0.01$, respectively, n= 3-4). (F) Relative expression level of *Hacry2* between CD and LF individuals. Mean ± SE. The "*" and "**" denote statistical significance of the expression levels between CD and LF individuals at the same time ($p<0.05$ and $p<0.01$, respectively, n= 3-4).
Table 1 (on next page)

Selective patterns for Cryptochrome genes.
### Table 1 Selective patterns for Cryptochrome genes.

| Genes | Model                  | np | \( \text{Ln L} \) | Estimates of \( \omega \) | Models compared | LRT \( ^\text{c} \) | \( P \)-Values |
|-------|------------------------|----|------------------|--------------------------|-----------------|-----------------|---------------|
| Cry1  | Branch model           |    |                  |                          |                 |                 |               |
|       | A: one ratio           | 19 | -9129.37         | \( \omega=0.0278 \)     |                 |                 |               |
|       | B: one ratio \( \omega=1 \) | 18 | -10508.58        | \( \omega=1 \)          | B vs. A        | 2758.42         | \(<0.001\)    |
|       | C: the moth branch has \( \omega_2 \) | 21 | -9127.01         | \( \omega_2=0.0287, \omega_1=0.0409 \) | A vs. C        | 4.72            | 0.09          |
|       | the butterfly branch has \( \omega_1 \) |    |                  |                          |                 |                 |               |
|       | D: the moth branch has \( \omega_2 \) | 20 | -10114.21        | \( \omega_2=1, \omega_1=0.0270 \) | D vs. C        | 1974.4          | \(<0.001\)    |
|       | E: each branch has its own \( \omega \) | 35 | -9111.95         | Variable \( \omega \) by branch | A vs. E        | 34.84           | 0.004         |
|       | Branch-site models     |    |                  |                          |                 |                 |               |
|       | G: the moth branch     | 22 | -9061.95         |                          |                 |                 |               |
|       | H: the moth branch has \( \omega=1 \) | 21 | -9061.95         |                          | H vs. G        | 0               | 1             |
| Cry2  | Branch model           |    |                  |                          |                 |                 |               |
|       | I: one ratio           | 13 | -5538.87         | \( \omega=0.0186 \)     |                 |                 |               |
|       | J: one ratio \( \omega=1 \) | 12 | -6629.16         | \( \omega=1 \)          | J vs. I        | 2180.58         | \(<0.001\)    |
|       | K: the moth branch has \( \omega_2 \) | 15 | -5534.15         | \( \omega_2=0.0219, \omega_1=0.0130 \) | I vs. K        | 9.44            | 0.002         |
|       | the butterfly branch has \( \omega_1 \) |    |                  |                          |                 |                 |               |
|       | L: the moth branch has \( \omega_2 \) | 14 | -6121.45         | \( \omega_2=1, \omega_1=0.0064 \) | L vs. K        | 1174.6          | \(<0.001\)    |
|       | M: each branch has its own \( \omega \) | 23 | -5525.41         | Variable \( \omega \) by branch | I vs. M        | 26.92           | 0.003         |
|       | Branch-site models     |    |                  |                          |                 |                 |               |
|       | N: the moth branch     | 16 | -5513.53         |                          |                 |                 |               |
|       | O: the moth branch has \( \omega=1 \) | 15 | -5513.53         |                          | N vs. O        | 0               | 1             |