Polymorphism of DNA sequences of cryptochrome genes is not associated with the photoperiodic flowering of wild soybean along a latitudinal cline

Nobuchika Ishibashi and Hiroaki Setoguchi*

Graduate School of Human and Environmental Studies, Kyoto University,

Yoshida-nihonmatsucho, Sakyo-ku, Kyoto 606-8501, JAPAN

*Corresponding author:

Hiroaki SETOGUCHI

Tel: +81-75-753-6860/Fax: 075-735-6694/

e-mail: seto@botany.mbox.media.kyoto-u.ac.jp

The English in this document has been checked by at least two professional editors, both native speakers of English.

Abstract   Both cultivated soybean and its wild relative *Glycine soja* exhibit strong photoperiodic sensitivity at different latitudes. Recent studies have demonstrated that the blue light-absorbing cryptochrome gene, *CRY1a*, is involved in the photoperiodic flowering of soybeans. However, no sequence variation was found in the cDNA among cultivars at different latitudinal clines. In the present study, we examined whether positive selection due to polymorphisms in the cryptochrome genes of *G. soja* occurs. Partial DNA sequences, mainly exons, of cryptochrome genes *CRY1a-1d* and *CRY2a-2c* were analyzed for 18 accessions in the Japanese archipelago. The neutral evolutionary pattern of the polymorphisms for all cryptochrome genes except for *CRY1a* was

summarized using Tajima's *D* test and low nucleotide diversity was shown for all genes. Although *CRY1a* did not show neutral evolution, positive selection was recognized in the intron while not in the exon. No geographical pattern of polymorphisms was observed in cryptochrome genes. These results reject the possibility of cryptochrome genes being involved in the photoperiodic flowering of wild soybeans along a latitudinal cline.

Introduction

Soybean [*Glycine max* (L.) Merr.], one of the most important crops in the world, was domesticated in China about 3,000–5,000 years ago from its progenitor *Glycine soja* Sieb. & Zucc. (Hymowitz et al. 1970). Presently, wild soybean (*G. soja*) is widely distributed in East Asia, including China, southeastern Russia, Korea, Taiwan, and Japan. In the Japanese archipelago, it is distributed from the islands of Hokkaido to Kyushu, with the exception of the Ryukyu Islands.

Photoperiod is one of the most important environment factors in flowering in response to changing seasons (Garner and Akkard 1920). Most soybean varieties exhibit strong photoperiodic sensitivity, an adaptation to a narrow latitudinal range. Therefore, various soybean cultivars have been grown in suitable latitudinal ranges globally. This photoperiodic sensitivity to latitudinal cline originated from the nature of the wild soybean. Both *G. max* and *G. soja* exhibit strong photoperiodic sensitivities at the latitude of each habitat or cultivar [local cultivars and *G. soja* in China: Zhang et al. 2008; *G. soja* in Japan: Fukui and Kaizuma 1971; Abe (personal communication)]. These studies reported the presence of a latitudinal cline in flowering time, where northern accessions tended to flower in short days more so than those of lower latitudes. These observations have been suggested to be the result of local adaptation of photoreceptor genes within a species' latitudinal distribution.

Plants have various photoreceptors that monitor light intensity, direction, quality, and duration. Phytochromes and cryptochromes are representative members of photoreceptors that trigger germination, flowering, shade avoidance, de-etiolation, and hypocotyl elongation. They are known for absorbing red/far-red light and blue light,

respectively. Seven gene families (*PHYA–PHYE* and *CRY1* and *CRY2*) have been identified in *Arabidopsis thaliana* (Ahmad and Cashmore 1993; Guo et al. 1998; Chen et al. 2004). Recent studies showed that phytochromes or cryptochromes are involved in local adaptation along latitudes and longitudes such as *PHYB2* in *Populus tremula* (Ingvarsson et al. 2006, 2008), *PHYC* in *A. thaliana* (Balasubramanian et al. 2006), *PHYE* in *Cardamine nipponica* (Ikeda et al. 2009), and *CRY1a* in *G. max* (Zhang et al. 2008). Thus, various photoreceptors play a role in local adaptation of different species, but may function differently under certain environmental conditions (e.g., temperature: Halliday and Whitelam 2003; Heschel et al. 2007). In addition, natural variations of many photoreceptors can be suppressed by purifying selection due to their essential role in development throughout a plant's life cycle (García-Gil et al. 2003; White et al. 2004; Mathews and McBreen 2008; Ikeda et al. 2011). For example, although *PHYE* is involved in the local adaptation of *C. nipponica*, *CRY1* and *CRY2* are not associated with a narrow latitudinal range in the Japanese archipelago (Ikeda et al. 2009, 2011).

Cryptochromes regulate plant development and play an important role in controlling de-etiolation, the inhibition of hypocotyl elongation, and flowering (Ahmad and Cashmore 1993; Guo et al. 1998). Soybeans have four *CRY1* genes (*CRY1a–1d*) and three *CRY2* genes (*CRY2a–2c*) as a result of allopolyploidization in the wild soybean *G. soja* (tetraploid: Gill et al. 2009). As noted above, CRY1a is significantly associated with a latitudinal cline in the photoperiodic flowering of *G. max* (Zhang et al. 2008). The photoperiod-dependent circadian rhythmic expression of the CRY1a protein correlates with photoperiodic flowering and its latitudinal distribution. However, no sequence variation was found in *CRY1a* cDNA among landraces at different latitudinal clines (Zhang et al. 2008). This result suggests that *CRY1a* gene polymorphisms could

be suppressed by purifying selection due to its critical function as a photoreceptor. However, introns of *CRY1a* or other cryptochrome genes have not been sequenced.

In this study, partial sequences (including introns) of the cryptochrome genes, *CRY1a–1d* and *CRY2a–2c*, in *G. soja* were sequenced from 18 accessions in the Japanese archipelago, which covered the entire distribution range of the species. This study aimed to elucidate whether cryptochrome gene polymorphisms in *G. soja* exhibit a geographical pattern associated with latitudinal cline.

Materials and methods

Sampling and DNA extraction

Eighteen wild soybean *G. soja* samples from 18 accessions in the Japanese archipelago were used in this study. Seeds of *G. soja* were provided by the National BioResource Project Legume Base, University of Miyazaki, Japan, covering the whole distribution range from Hokkaido (42°N) to Kyushu (31°N) (Table S1, Supporting information and Fig. 1). Seeds were soaked in 200 μL of concentrated sulfuric acid for 30 min, washed with water, soaked in water for 10 min, and sowed in the pot. Mature foliage leaves from one plant in each accession were used for DNA extraction. Fresh leaf material was pulverized with a Tissue Lyser (Qiagen, Hilden, Germany) according to the manufacturer's instructions. After removing the polysaccharides from the powder using HEPES buffer (pH 8.0; Setoguchi and Ohba 1995), DNA was extracted using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle 1990). The extracted DNA was dissolved in 100 μL of Tris-EDTA buffer and used for subsequent

amplification by polymerase chain reaction (PCR).

PCR and sequencing

Seven cryptochrome genes (*CRY1a*, *CRY1b*, *CRY1c*, *CRY1d*, *CRY2a*, *CRY2b*, and *CRY2c*) were amplified (including introns). PCR primers were designed based on sequences in the soybean genome database Phytozome (http://www.phytozome.net/soybean), and others were obtained from the literature (Matsumura et al. 2009) (Table S2). The position of each primer is shown in Fig. S1. PCR was conducted in a total reaction volume of 10 μL containing 6.75 μL of autoclaved ion-exchanged water, 0.8 μL of 2.5 mM dNTP mixture, 1.0 μL of Ex *Taq* Buffer (Takara Ex*Taq*; Takara, Kyoto, Japan), 0.25 U Ex *Taq* (Takara), 0.2 μM of each primer, and 1.0 μL of template DNA (10 ng/μL). PCR was performed with 35 cycles of 1 min at 94°C, 1 min at 56°C, and 2 min at 72°C, and 1 cycle of 5 min at 72°C. Following amplification, the PCR products were visualized on 1.5% TAE-agarose gels stained with ethidium bromide and gel-purified with glass powder using GeneCleanII (Bio 101, Vista, CA, USA). The products were sequenced using the standard methods of the BigDye™ Terminator Cycle Sequence Ready Reaction Kit (Applied Biosystems, Foster City, CA, USA) using several primers for each region and sequenced using an ABI 3130 Genetic Analyzer (Applied Biosystems). All sequences were aligned using Auto Assembler (Applied Biosystems).

All exons from samples no. 1-18 were sequenced. If single-nucleotide polymorphisms (SNPs) were identified, the neighboring introns were also sequenced. Haplotypes of each cryptochrome gene were determined by direct-sequencing of PCR products because no duplicate sequences were found in any gene. DnaSP version 5.10

(Librado and Rozas 2009) was used for DNA polymorphism analysis, haplotyping, and carrying out statistical neutrality tests, and Tajima's *D* test (Tajima 1989) was performed on each sequence. In addition, the nucleotide variation of each sequence was estimated as nucleotide diversity π (Tajima 1989).

Results

Sequences and haplotypes of *CRY* genes in *Glycine soja*

The primers amplified the following PCR products: 3,400 bp (*cry1a*), 3,300 bp (*cry1b*), 3,200 bp (*cry1c*), 3,300 bp (*cry1d*), 3,200 bp (*cry2a*), 2,000 bp (*cry2b*), and 2,800 bp (*cry2c*). The genetic structure, sequenced regions, and position of detected SNPs are shown in Fig. 2. Most of the genes harbored many SNPs (e.g., *CRY1a*: 14, *CRY1b*: 6, *CRY1c*: 21, *CRY1d*: 11, *CRY2a*: 12, *CRY2b*: 2, and *CRY2c*: 5). All genes except *CRY1b* and *CRY2*c had fewer SNPs in the exons (*CRY1a*: 5, *CRY1b*: 3, *CRY1c*: 5, *CRY1d*: 3, *CRY2a*: 4, *CRY2b*: 1, *CRY2c*: 5) than in the introns (*CRY1a*: 9, *CRY1b*: 3, *CRY1c*: 16, *CRY1d*: 8, *CRY2a*: 8, *CRY2b*: 1, *CRY2c*: 0). In particular, *CRY1a*, *CRY1b*, *CRY1c*, and *CRY1d* contained many SNPs within introns, whereas *CRY2a*, *CRY2b*, and *CRY2c* had few SNPs. More synonymous changes (*CRY1a*: 3, *CRY1b*: 3, *CRY1c*: 3, *CRY1d*: 3, *CRY2a*: 4, *CRY2b*: 0, *CRY2c*: 3) were detected than nonsynonymous changes (*CRY1a*: 2, *CRY1b*: 0, *CRY1c*: 2, *CRY1d*: 0, *CRY2a*: 0, *CRY2b*: 1, *CRY2c*: 2) in each exon, with the exception of *CRY2b*. Haplotype relationships (parsimony networks) and geographical distributions are shown in Fig. 3. All genes harbored plural haplotypes, ranging from 2 (*CRY2b*) to 12 (*CRY1c*). Haplotypes 1 and 13 of *CRY1a* were distinguished by 15 steps

of nucleotide substitutions, while only one step was assumed between the two *CRY2b* haplotypes. The geographical distribution of haplotypes from each photoreceptor gene showed no particular structure (Fig. 3). In addition, we attempted to analyze the phylogenetic relationships of the 18 wild soybean accessions based on the DNA sequences of exons, or exons and introns of the seven cryptochrome genes. However, no geographic structure was found (data not shown).

Neutrality and nucleotide diversity

Tajima's *D* test showed that almost all cryptochrome genes evolved in a neutral manner in each neutral test, D_{total} (entire sequence), D_a (nonsynonymous sites), D_s (synonymous sites), and D_{sil} (synonymous sites and intron) (Table 1). However, only D_{tolal} and D_{sil} of *CRY1a* 1 were not neutral ($D_{total} = 3.07$, $P < 0.01$; $D_{sil} = 3.07$, $P < 0.01$).

The nucleotide diversities of most *CRY1* and *CRY2* exons of *G. soja* were lower (*CRY1a* 1: $\pi_{cod} = 0.00142$, *CRY1a* 2: $\pi_{cod} = 0.00088$, *CRY1b* 1: $\pi_{cod} = 0.00024$, *CRY1b* 2: $\pi_{cod} = 0.00050$, *CRY1c* 1: $\pi_{cod} = 0.00069$, *CRY1c* 2: $\pi_{cod} = 0$, *CRY1c* 3: $\pi_{cod} = 0.00015$, *CRY1c* 4: $\pi_{cod} = 0.00087$, *CRY1d* 1: $\pi_{cod} = 0$, *CRY1d* 2: $\pi_{cod} = 0.00091$, *CRY1d* 3: $\pi_{cod} = 0.00037$, *CRY2a* 1: $\pi_{cod} = 0$, *CRY2a* 2: $\pi_{cod} = 0.00033$, *CRY2a* 3: $\pi_{cod} = 0.00110$, *CRY2b* 1: $\pi_{cod} = 0.00043$, *CRY2c* 1: $\pi_{cod} = 0.000111$) (Table 1) compared to 116 functional genes in *G. soja* (average $\pi_{cod} = 0.00105$) from a wide distribution range across China, Korea, Taiwan, Russia, and Japan (23–50°N and 106–140°E) (Hyten et al. 2006). Although the *CRY1a* 1, *CRY2a* 3, and *CRY2c* 1 coding regions showed a higher nucleotide diversity than those of the 102 functional genes of *G. soja* (Hyten et al. 2006), π_a (nonsynonymous sites) of each region showed a lower nucleotide diversity (*CRY1a*1:

$\pi_a = 0$, *CRY2a* 3: $\pi_a = 0$, *CRY2c* 1: $\pi_a = 0.00072$) than those reported in a previous study ($\pi_a = 0.00096$; Table 1) (Hyten et al. 2006).

Thus, the coding region of each cryptochrome gene harbors conventional nucleotide diversity, and its evolutionary pattern is neutral.

Discussion

Photoreceptors play an important role in plant development throughout the life cycle (from seed germination to flowering). The presence of polymorphisms in photoreceptor genes can cause significant changes in phenotype and/or plant phenology. Although Zhang et al. (2008) demonstrated an association between GmCRY1a protein expression and flowering time in soybean, the present study found no association between polymorphisms of cryptochrome genes and photoperiodic sensitivity along latitudinal cline in *G. soja* (Fig. 3).

All cryptochrome genes, except for *CRY1a* 1, evolved in a neutral manner and most harbored a lower nucleotide diversity than those of the 102 functional genes of *G. soja* (Table 1). Although the coding regions of *CRY2a* 3 and *CRY2c* 1 harbored a high nucleotide diversity, nonsynonymous changes in these cryptochrome genes were suppressed because the π_a values of these genes were lower than those reported by Hyten et al. (2006) (Table 1).

Although *CRY1a* 1 was not neutral, neutrality in nonsynonymous and synonymous sites was detected, but silent sites were not (Table 1), which suggests that the intron of *CRY1a* 1 had balancing selection. Functional changes in this gene, however, may be suppressed because the nucleotide diversity of nonsynonymous sites was zero (Table 1).

These data suggest that polymorphisms in the exons of all cryptochrome genes exhibited a neutral evolutionary pattern; additionally, polymorphisms in functional regions were suppressed and no variation was associated with latitudinal cline. The cryptochrome genes reportedly might not be the primary determinants of time to flowering and maturity in soybean (Matsumura et al. 2009). This notion depends on the fact that the cryptochrome genes were mapped at different locations from the E genes that control time to flowering and maturity (Matsumura et al. 2009). This result does not conflict with the lack of a correlation between latitudinal cline and cryptochrome gene polymorphisms in *G. soja*. However, further studies of posttranscriptional and posttranslational regulation are needed to fully elucidate the function of cryptochrome genes. In addition, the function of the intron in *CRY1a* that exhibited balancing selection is unclear. Silvio et al. (2007) reported that a conserved noncoding sequence is associated with a flowering-time quantitative trait locus in maize. Tajima's *D* test indicated that *CRY1a* 1 experienced balancing selection as a past event. Furthermore, exon regions besides this intron may have also experienced balancing selection, but polymorphisms therein had been suppressed because of its functional importance. Thus, further study is needed to fully elucidate the function of the *CRY1a* intron.

Figure legends

Fig. 1   Collection sites of *Glycine soja* used in this study.

Fig. 2   Structure of cryptochrome genes and SNPs. White and gray bars represent exons and the 5' and 3' UTRs, respectively. The lines between white bars depict introns. Closed circles and triangles represent synonymous and nonsynonymous changes, respectively. Bold lines beneath gene structures indicate the regions sequenced. The numbers besides bold lines indicate the sequence number used in the neutral test and nucleotide diversity calculations.

Fig. 3   Haplotype relationships and distributions of the *CRY1* and *CRY2* genes of *Glycine soja* in the Japanese archipelago.

Fig. S1   Binding sites of PCR primers used to amplify the cryptochrome genes of *Glycine soja*. Closed and open triangles indicate forward and reverse primers, respectively. The type and number of primers are noted adjacent to the triangles.
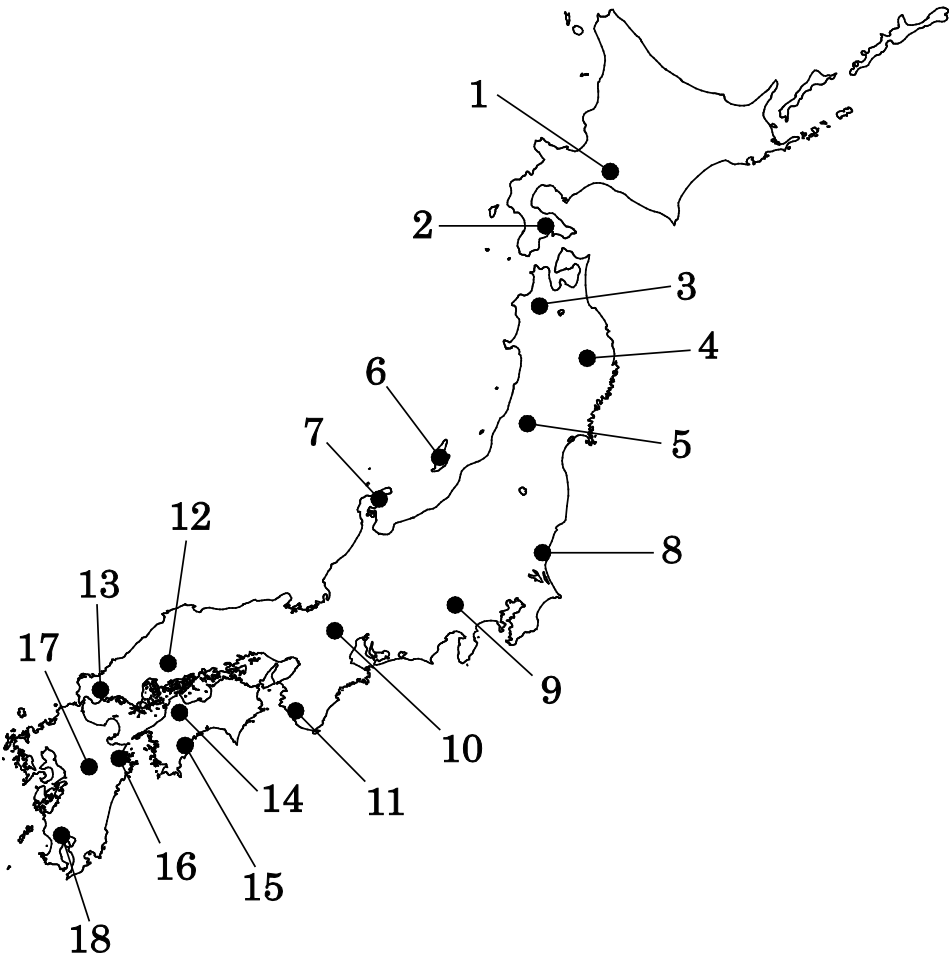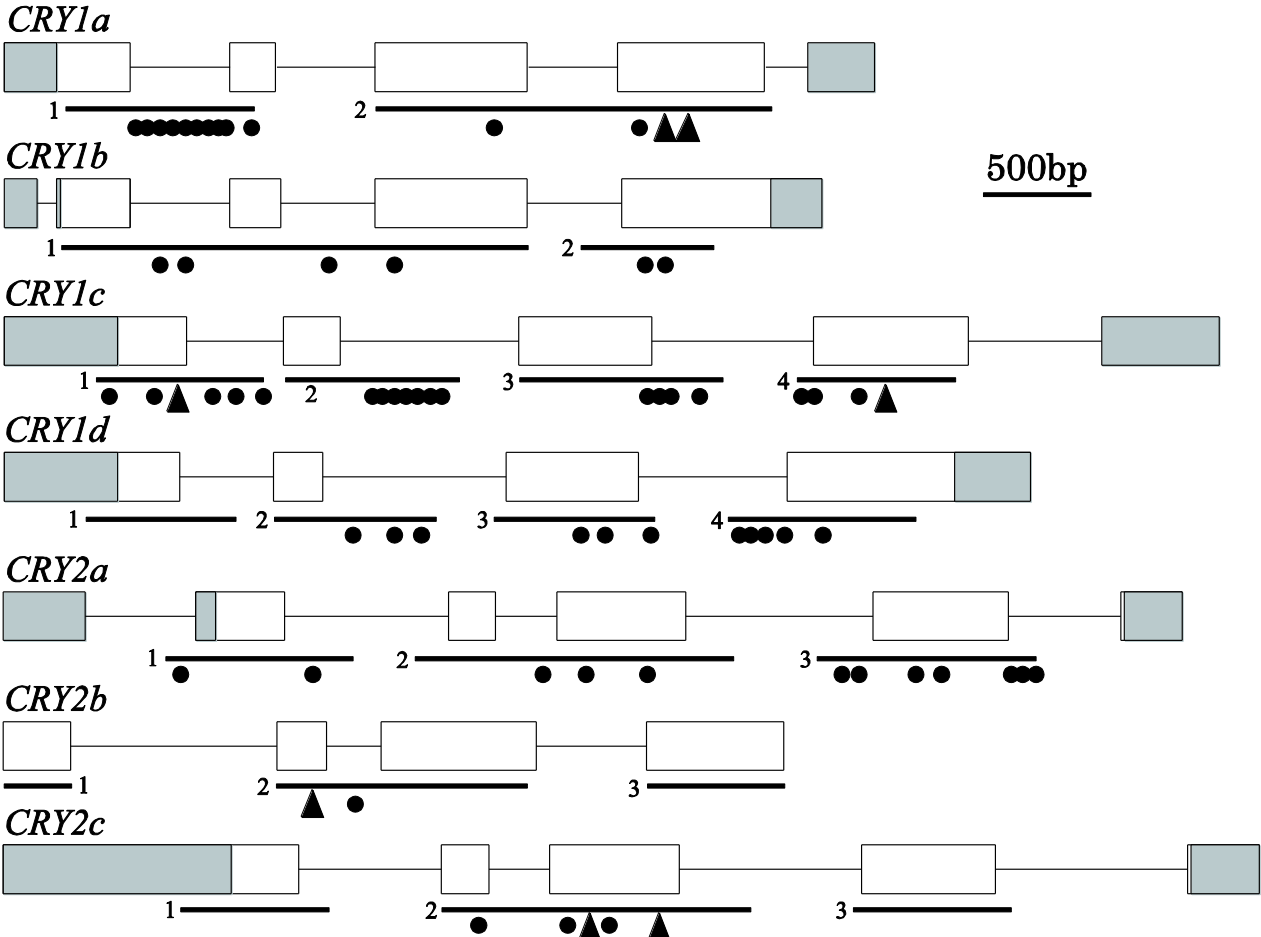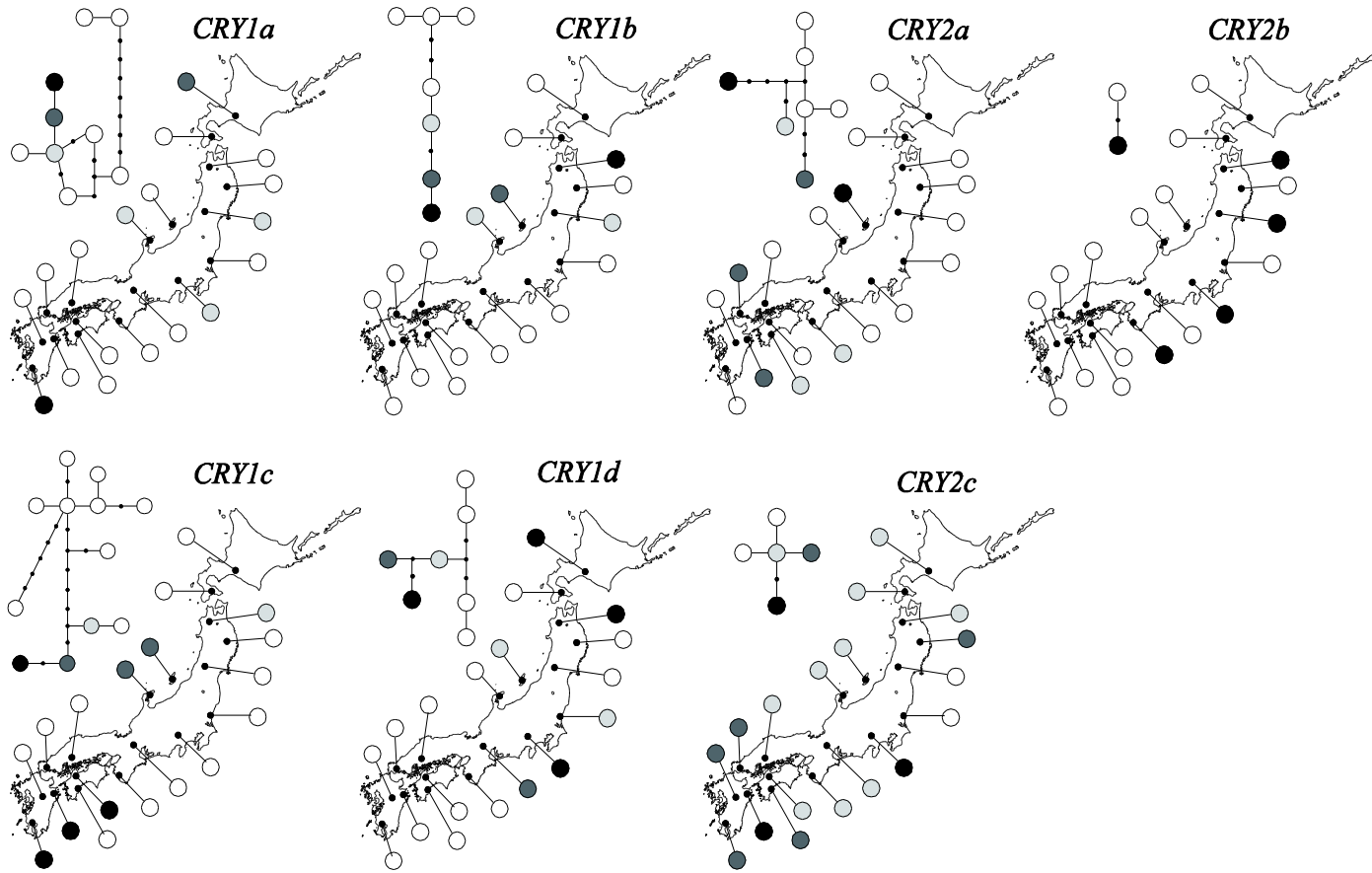
Fig. 1

Fig. 2

Fig. 3

CRY1a   *CRY1b*   *CRY2a*   *CRY2b*

CRY1c   *CRY1d*   *CRY2c*

Table 1 Results of neutral test and nucleotide diversity in each cryptochrome gene

Locus	No.	D_{total}	D_a	D_s	D_{sil}	π_{total}	π_a	π_s	π_a/π_s	π_{cod}	π_{sil}
CRY1a	1	3.07 [*]	n.a.	1.63	3.07 [*]	0.00597	0	0.00542	0	0.00142	0.00885
	2	0.82	-0.69	2.00	2.00	0.00068	0.00027	0.00289	0.095	0.00088	0.00129
CRY1b	1	-0.03	n.a.	0.27	-0.03	0.00045	0	0.00100	0	0.00024	0.0008
	2	-1.07	n.a.	-1.07	-1.07	0.00037	0	0.00210	0	0.00050	0.00085
CRY1c	1	-0.06	-0.81	-0.81	0.18	0.00212	0.00046	0.00140	0.328	0.00069	0.00285
	2	0.12	n.a.	n.a.	0.12	0.00274	0	0	0	0	0.00355
	3	0.30	n.a.	-0.81	0.30	0.00127	0	0.00068	0	0.00015	0.00265
	4	1.47	1.67	-0.81	0.94	0.00206	0.00094	0.00063	1.496	0.00087	0.00513
CRY1d	1	1.08	n.a.	n.a.	1.08	0.00141	0	0	0	0	0.00179
	2	1.32	n.a.	0.64	1.32	0.00146	0	0.00405	0	0.00091	0.00484
	3	-0.64	n.a.	-0.23	-0.64	0.00117	0	0.00157	0	0.00037	0.00249
CRY2a	1	0.64	n.a.	n.a.	0.64	0.00102	0	0	0	0	0.00164
	2	-0.65	n.a.	-0.69	-0.65	0.00035	0	0.00154	0	0.00033	0.00067
	3	-0.56	n.a.	0.95	-0.56	0.00121	0	0.00501	0	0.00110	0.00224
CRY2b	1	0.92	0.70	n.a.	0.70	0.00065	0.00055	0	n.a.	0.00043	0.00081
CRY2c	1	-0.37	0.15	-0.65	-0.65	0.00069	0.00072	0.00254	0.283	0.00111	0.00066

Tajima's D and nucleotide diversity (π) are shown. Tajima's D and nucleotide diversity were estimated for the entire (D_{total} and π_{total}) sequence as well as nonsynonymous sites (D_a and π_a), synonymous sites (D_s and π_s) and silent sites (D_{sil} and π_{sil}) separately. Nucleotide diversity was also estimated for the coding region (π_{cod}). $P < 0.01$

[a] 1 and 2 correspond to sequence regions represented in Fig. 2

[b] represent the significant values of Tajima's D ($p < 0.01$)
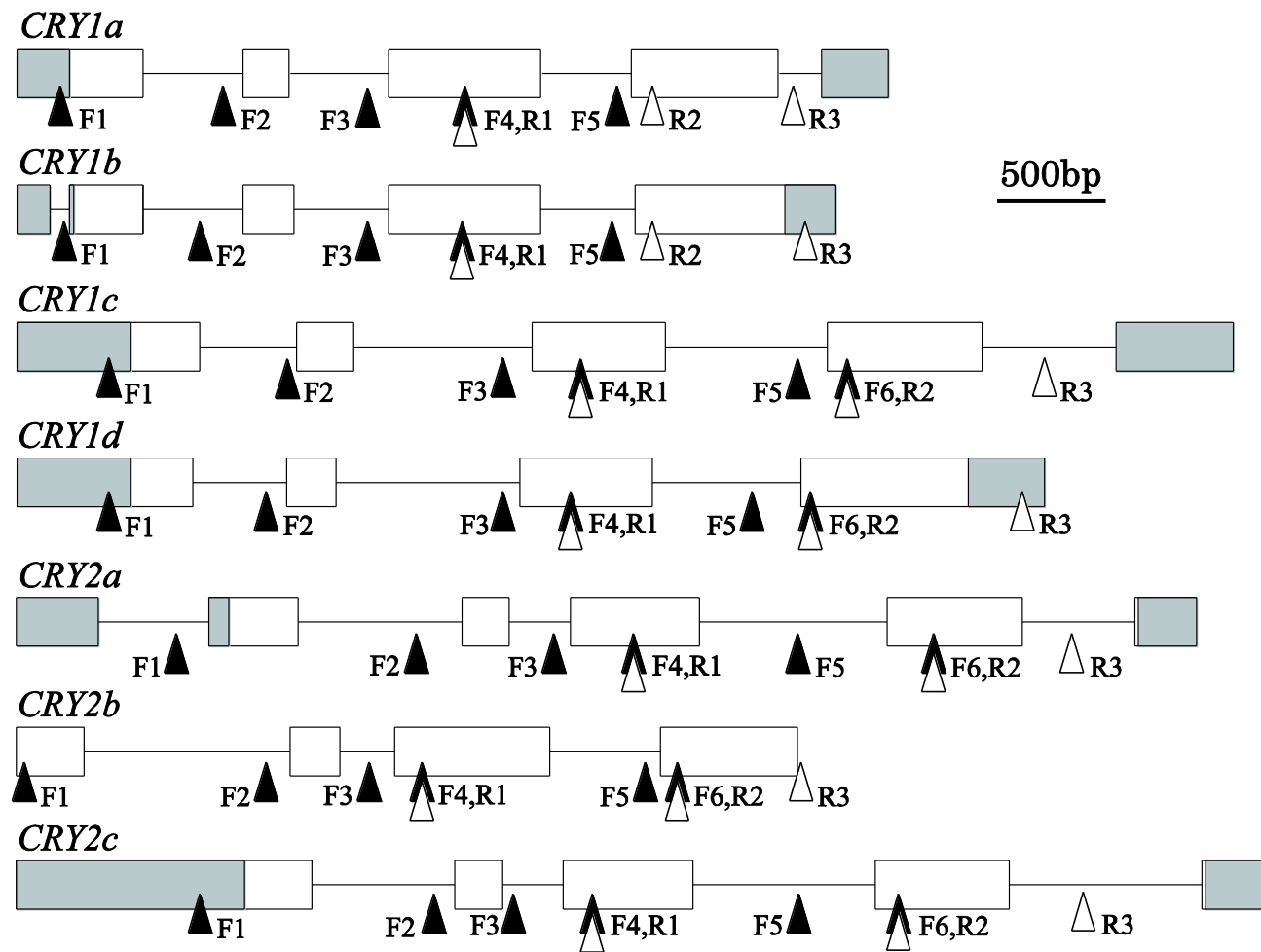
Fig. S1

Table S1 The geographic information of *Glycine soja* used in this study.

Sample No.	Accession No.	Latitude/Longitude	Prefecture
1	B01166	42°77'N/141°83'E	Hokkaido
2	B01134	41°83'N/140°64'E	Hokkaido
3	B02086	40°64'N/140°50'E	Aomori
4	B02112	39°67'N/141°19'E	Iwate
5	B02134	38°70'N/140°23'E	Yamagata
6	B03058	38°03'N/138°40'E	Niigata
7	B03034	37°40'N/137°24'E	Ishikawa
8	B04096	36°38'N/140°48'E	Ibaraki
9	B04134	35°66'N/138°64'E	Yamanashi
10	B06083	35°25'N/136°16'E	Shiga
11	B06053	33°65'N/135°40'E	Wakayama
12	B08064	34°67'N/132°69'E	Hiroshima
13	B08040	34°47'N/131°44'E	Yamaguchi
14	B07162	33°89'N/133°11'E	Ehime
15	B07161	33°33'N/133°29'E	Kochi
16	B09051	32°95'N/131°86'E	Ohita
17	B09080	32°96'N/131°08'E	Kumamoto
18	B09088	31°82'N/130°31'E	Kagoshima

Table S2 PCR primer sequences used for cryptochrome genes in *Glycine soja*.

Gene	Type of primer	number	primer sequence (5'-3')
CRY1a	F	1	GAAGTTTTTTAGCTTTGTAGAT
		2	ATGCATTTTTTTAGCTTGATTTGT
		3	CATGTGCATAATTTACCAATAGGC
		4	ACGAAGGAAACAAAGCCGGC
		5	TACCTGCCCGTGTCTGTGTT
	R	1	GCCGGCTTTGTTTCCTTCGT
		2	CACTCGGTTGGTAATCTTGC
		3	ACAATTGGTTTGATGTTCCTC
CRY1b	F	1	AGTTTTTAGCTTTAGCTTTGC
		2	GAACAAGACATATCACTTAAGCAT
		3	CAGATATAGATATTCAAGAAGAATTC
		4	ACGAAGGAAACAAAGCTGGT
		5	ATATATGATCGATCTTAAGTTCATAA
	R	1	ACCAGCTTTGTTTCCTTCGT
		2	CACTCAGTGGGTAATCTTGA
		3	ACAACTAATTCGATGTTACTG
CRY1c	F	1	GTTGTGGTGTTCTTGAGGTG
		2	CTTCAGCTTCATGATGATGA
		3	GGTCTTACTTTGGTACATTGTACATA
		4	TAAACGGTGCTCTAATCGAG
		5	GATATTGACATTTTGAGACTCGAG
		6	TGCGCTACAAATGTGACCCA
	R	1	CTCGATTAGAGCACCGTTTA
		2	TGGGTCACATTTGTAGCGCA
		3	AAGCACGAGTGGTTGTAGAT
CRY1d	F	1	TGTTCTTGAGGCTCAGTGGT
		2	AAGTTCATCCCTCATGATGAT
		3	CTTTTAATGAAGCATCACTCTTGAC
		4	TAAACGGTGCTTTGATCGAA
		5	CTTTTCAGTGTTTGAAAGTTAAC
		6	AGCGGTACAAATGTGACCCA
	R	1	TTCGATCAAAGCACCGTTTA
		2	TGGGTCACATTTGTACCGCT
		3	AAGCACGAGTGGTTGTAGAT

Gene	Type of primer	number	primer sequence (5'-3')
CRY2a	F	1	TTGAGGCTCTTATCCCATGT
		2	TTGCATACTCTTGGACATTG
		3	GTAATCTGATCAACTTCAACTTC
		4	ATTTTGGAGAATTAAGTGCA
		5	CTGTCTAACCAAAGTATATCTG
		6	CCAGAATTACCCAAAGCCAA
	R	1	TGCACTTAATTCTCCAAAAT
		2	TTGGCTTTGGGTAATTCTGG
		3	TTACTGTATCTCTGCTAGCC
CRY2b	F	1	TGCTGCTGCTGCTGCCAA
		2	GAAAAATTTCTGATATAGCTTCTC
		3	AGTCGAGATATATAAAAACCAATGAG
		4	GTATCCAGGTTGGAACAAGA
		5	TAATGAGGAATAAATACTACTCAC
		6	ACGCCAATGGTTGCCTGAGT
	R	1	TCTTGTTCCAACCTGGATAC
		2	ACTCAGGCAACCATTGGCGT
		3	TCCTGAGTTTGATAAGGTTC
CRY2c	F	1	TTCCCCAAAAATCCAGATCC
		2	ATTTCTTCACAGTTCAATGTTAG
		3	GGTATGATATGACTATATTAACACA
		4	CGTGGAGCAGCATCTACTAG
		5	GAAGTTCTGTTTCATGTATTAGTG
		6	GCTTACTGTTCTTAGAGCAG
	R	1	CTAGTAGATGCTGCTCCACG
		2	CTGCTCTAAGAACAGTAAGC
		3	TGTCAGAATGTGTATGGGCC