



ELSEVIER

Contents lists available at ScienceDirect

## Comptes Rendus Biologies

www.sciencedirect.com



Molecular biology and genetics/Biologie et génétique moléculaires

## Detection of phytochrome-like genes from *Rhazya stricta* (Apocynaceae) using *de novo* genome assembly



Jamal S.M. Sabir<sup>a</sup>, Nabih A. Baeshen<sup>a</sup>, Ahmed M. Shokry<sup>a,b</sup>, Nour O. Gadalla<sup>a,c</sup>, Sherif Edris<sup>a,d</sup>, Mohammed H. Mutwakil<sup>a</sup>, Ahmed M. Ramadan<sup>a,b</sup>, Ahmed Atef<sup>a</sup>, Magdy A. Al-Kordy<sup>a,c</sup>, Osama A. Abuzinadah<sup>a</sup>, Fotouh M. El-Domyati<sup>a,d</sup>, Robert K. Jansen<sup>a,e</sup>, Ahmed Bahieldin<sup>a,\*</sup>

<sup>a</sup> Department of Biological Sciences, Faculty of Science, King Abdulaziz University (KAU), PO Box 80141, Jeddah 21589, Saudi Arabia

<sup>b</sup> Agricultural Genetic Engineering Research Institute (AGERI), Agriculture Research Center (ARC), Giza, Egypt

<sup>c</sup> Genetics and Cytology Department, Genetic Engineering and Biotechnology Division, National Research Center, Dokki, Egypt

<sup>d</sup> Department of Genetics, Faculty of Agriculture, Ain Shams University, Cairo, Egypt

<sup>e</sup> Department of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA

## ARTICLE INFO

## Article history:

Received 22 June 2013

Accepted after revision 10 October 2013

Available online 13 November 2013

## Keywords:

*phyA**PhyB**phyC**phyE*

BLASTp

Alkaloids

## ABSTRACT

Phytochrome-like genes in the wild plant species *Rhazya stricta* Decne were characterized using a *de novo* genome assembly of next generation sequence data. *Rhazya stricta* contains more than 100 alkaloids with multiple pharmacological properties, and leaf extracts have been used to cure chronic rheumatism, to treat tumors, and in the treatment of several other diseases. Phytochromes are known to be involved in the light-regulated biosynthesis of some alkaloids. Phytochromes are soluble chromoproteins that function in the absorption of red and far-red light and the transduction of intracellular signals during light-regulated plant development. *De novo* assembly of the nuclear genome of *R. stricta* recovered 45,641 contigs greater than 1000 bp long, which were used in constructing a local database. Five sequences belonging to *Arabidopsis thaliana* phytochrome gene family (i.e., *AtphyABCDE*) were used to identify *R. stricta* contigs with phytochrome-like sequences using BLAST. This led to the identification of three contigs with phytochrome-like sequences covering *AtphyA*-, *AtphyC*- and *AtphyE*-like full-length genes. Annotation of the three sequences showed that each contig consists of one phytochrome-like gene with three exons and two introns. BLASTn and BLASTp results indicated that *RsphyA* mRNA and protein sequences had homologues in *Wrightia coccinea* and *Solanum tuberosum*, respectively. *RsphyC*-like mRNA and protein sequence were homologous to *Vitis vinifera* and *Vitis riparia*. *RsphyE*-like mRNA coding and protein sequences were homologous to *Ipomoea nil*. Multiple-sequence alignment of phytochrome proteins indicated a homology with 30 sequences from 23 different species of flowering plants. Phylogenetic analysis confirmed that each *R. stricta* phytochrome gene is related to the same phytochrome gene of other flowering plants. It is proposed that the absence of *phyB* gene in *R. stricta* is due to *RsphyA* gene taking over the role of *phyB*.

© 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

### 1. Introduction

Plant development and performance is strongly influenced by environment. Light is one of the most important factors affecting plants because it is involving in the regulation of seed germination, establishment of seedlings,

\* Corresponding author.

E-mail address: bahieldin55@gmail.com (A. Bahieldin).

determination of growth habit, and the transition to flowering [1]. Plants have a collection of photoreceptors perceiving information (light quality and quantity, and duration) about their light environment [2,3]. Red (R) and far-red (FR) light are important environmental signals in the regulation of plant development, with major roles in seedling de-etiolation, neighbor detection, and photoperiodism [4,5]. Plant R and FR photoreceptors, the phytochromes, are soluble chromoproteins consisting of a 120–130 kDa apoprotein and a linear tetrapyrrole chromophore. A small gene family *phy*, encoding the phytochrome apoproteins (chromoproteins), functions in the absorption of R and FR light and the transduction of intracellular signals during light-regulated plant development [6].

Plants contain multiple, distinct phytochrome species that are the products of a divergent gene family [7]. Physiological and biochemical experiments in earlier studies indicated that plants contain two different forms of phytochromes [8–10]. The first form is frequently referred to as type-I (or light labile) phytochrome, predominating in etiolated tissue, while the second form is referred to as type-II (or light-stable) phytochrome, prevalent in light-grown tissue [11–13].

The *phy* gene family was first identified in *Arabidopsis* and subsequently in many other plant species [1,7,14–16]. In *Arabidopsis*, there are five phytochrome genes, *phyABCDE* [17]. The product of the *phyA* gene, *phyA*, is light labile and predominates in etiolated seedlings where it accumulates to relatively high levels. *phyB* and *phyC* proteins are more light-stable, with *phyB* are common in light-grown tissues [18]. *Arabidopsis* may not be representative of flowering plants because additional *phy* loci related to *phyA* and *phyB* evolved independently several times in eudicots. Additionally, monocots may lack orthologs of *Arabidopsis phyD* and *phyE* [19]. Etiolated seedlings of *phyB* mutants are deficient in several responses to red light [20,21]. Light-grown seedlings of *phyB* mutants have an elongated, early flowering phenotype of the shade-avoidance syndrome of wild-type seedlings grown under low R/FR. The *phyB* mutant seedlings display attenuated responses to the low R/FR or to the end of day (EOD) far-red light, leading to the proposal that *phyB* may play a key role in the shade-avoidance response [22,23]. In addition, Hirschfeld et al. [13] indicated that the level of *phyC* apoprotein is strongly reduced in *Arabidopsis phyB* mutants. This observation suggests that phenotypes associated with *phyB* mutant genes result from the attenuation of *phyC* signaling. Some shade-avoidance responses of the *phyB* mutant to the low R/FR or to EOD far-red light are severely attenuated [22], although other responses, such as reduction in leaf area and acceleration of flowering, are clearly retained [24,25]. Devlin et al. [26] demonstrated that *phyA-phyB* double mutants respond to EOD far-red light by the acceleration of flowering and by the promotion of elongated internodes between rosette leaves. These responses are reversible by subsequent treatment with red light, indicating that one or more of *phyC*, *phyD*, or *phyE* control flowering time and internode elongation [26,27]. Unexpectedly, Hu et al. [28] revealed both light-dependent and -independent roles for phytochromes to regulate the *Arabidopsis* circadian clock,

indicating dual roles for phytochromes to arrest and/or promote the progression of plant development in response to the prevailing light environment. Phytochrome was also shown to function in circadian clock adjustment to plant iron status through a new retrograde pathway that involves a plastid-encoded protein [29].

*Rhazya stricta* Decne is a robust plant with erect stems and upright thick and smooth leaves placed close together on the stem [30]. It is common in the Arabian Peninsula and the Indian subcontinent. Leaves are used to make a tonic with a peculiar bitter taste. *Rhazya stricta* is used traditionally in Asia for the treatment of different types of diseases, such as skin diseases, stomach diseases, and hypertension [31]. The leaves, flowers and fruits are also used in joint infections and for cancer [32]. Phytochemical analyses have identified more than 100 alkaloids [33] with several pharmacological properties. There is considerable evidence that light can affect the production of alkaloids. For example, *Catharanthus roseus* L. is affected by light in the production of the dimeric indole alkaloids, vinblastine and vincristine. The biosynthesis and accumulation of vindoline in the intact plant is controlled by tissue-specific, development regulated, and light-dependent factors [34]. Several lines of evidence showed that phytochromes are involved in the light-regulated biosynthesis of these dimeric indole alkaloids. Also, Höft et al. [35] demonstrated that the alkaloid content in the leaves of *Tabernaemontana pachysiphon* Stapf. (Apocynaceae) is influenced by light intensity.

In this study, phytochrome-like genes of *phyA*, *phyC* and *phyE* were characterized from *de novo* assembled genome contigs from next generation sequencing in the wild plant species *Rhazya stricta*. Bioinformatic analyses of these data confirmed the recovery of three full-length phytochrome-like genes.

## 2. Materials and methods

### 2.1. Isolation of nuclear DNA

Extraction of total DNA was performed using the modified procedure of Gawel and Jarret [36]. Three samples of discs collected from the upper leaves of *R. stricta* were frozen in liquid nitrogen (approximately 500 mg of tissue each). To remove RNA contaminants, RNase A (10 mg/mL, Sigma, USA) was added to the DNA samples and incubated at 37 °C for 30 min. Estimation of the DNA concentration in different samples was done by measuring optical density at 260 nm according to the equation:

$$\text{DNA concentration (ug/mL)} = \text{OD}_{260} \times 50 \\ \times \text{dilutionfactor}$$

Purified DNA samples were sent to Beijing Genomics Institute (BGI), Shenzhen, China for sequencing.

### 2.2. Sequence filtering and bioinformatic analysis

The raw sequence data were obtained using the Illumina python pipeline v. 1.3. For obtained libraries,

Table 1

Accession number, description of the genes and organism (Latin name), for phytochrome genes used in this study.

Accession No.	Description	Latin Name
XM_637404	<i>Dictyostelium discoideum</i> AX4 prolyl 4-hydroxylase ( <i>phyA</i> ) mRNA, complete cds	<i>Dictyostelium discoideum</i> AX4
XM_004362186	<i>Dictyostelium fasciculatum</i> prolyl 4-hydroxylase ( <i>phyA</i> ) mRNA, complete cds	<i>Dictyostelium fasciculatum</i>
AF547224	Synthetic construct phytase ( <i>phyA</i> ) mRNA, complete cds	synthetic construct
EU786166	<i>Aspergillus japonicus</i> strain BCC18313 <i>PhyA</i> ( <i>phyA</i> ) mRNA, complete cds	<i>Aspergillus japonicus</i>
EU786167	<i>Aspergillus niger</i> strain BCC18081 <i>PhyA</i> ( <i>phyA</i> ) mRNA, complete cds	<i>Aspergillus niger</i>
XM_002561048	<i>Penicillium chrysogenum</i> Wisconsin 54-1255 phytase <i>phyA</i> from patent WO2003038111-A2- <i>Penicillium chrysogenum</i> ( <i>phyA</i> ) mRNA, complete cds	<i>Penicillium chrysogenum</i>
AJ310697	<i>Agrocybe pediades</i> mRNA for Phytase ( <i>phyA</i> gene)	<i>Agrocybe pediades</i>
AJ310700	<i>Trametes pubescens</i> mRNA for Phytase ( <i>phyA</i> gene)	<i>Trametes pubescens</i>
AB042805	<i>Aspergillus oryzae</i> <i>phyA</i> mRNA for phytase, complete cds	<i>Aspergillus oryzae</i>
AJ310696	<i>Peniophora lycii</i> mRNA for Phytase ( <i>phyA</i> gene)	<i>Peniophora lycii</i>
GU120223	<i>Aspergillus</i> sp. A25 phytase ( <i>phyA</i> ) mRNA, complete cds	<i>Aspergillus</i> sp. A25
AJ543399	<i>Trichoderma harzianum</i> <i>phyA</i> gene for phytase	<i>Trichoderma harzianum</i>
JF412664	<i>Amphicarpaea edgeworthii</i> phytochrome B ( <i>PhyB</i> ) mRNA, complete cds	<i>Amphicarpaea edgeworthii</i>
QJ771614	<i>Arabidopsis thaliana</i> ecotype Landsberg erecta (Ler-0) phytochrome C ( <i>PhyC</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
DQ208423	<i>Solanum tuberosum</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Solanum tuberosum</i>
AB743571	<i>Fragaria x ananassa</i> <i>phyA</i> mRNA for phytochrome A, complete cds	<i>Fragaria x ananassa</i>
XM_002271635	PREDICTED: <i>Vitis vinifera</i> phytochrome E ( <i>PhyE</i> ), mRNA	<i>Vitis vinifera</i>
GU994130	<i>Amblyopyrum muticum</i> cultivar Ae45 phytochrome A ( <i>PhyA</i> ) mRNA, complete cds	<i>Amblyopyrum muticum</i>
GU994114	<i>Triticum monococcum</i> cultivar BO1 phytochrome A ( <i>PhyA</i> ) mRNA, complete cds	<i>Triticum monococcum</i>
AB743572	<i>Fragaria x ananassa</i> <i>phyB</i> mRNA for phytochrome B, complete cds	<i>Fragaria x ananassa</i>
XM_003555718	<i>Glycine max</i> phytochrome type A-like (LOC100790763), mRNA	<i>Glycine max</i>
EU428746	<i>Glycine max</i> phytochrome A-1 ( <i>phyA</i> ) mRNA, complete cds	<i>Glycine max</i>
EU428748	<i>Glycine max</i> phytochrome A-3 ( <i>phyA</i> ) mRNA, complete cd.	<i>Glycine max</i>
EU428747	<i>Glycine max</i> phytochrome A-2 ( <i>phyA</i> ) mRNA, complete cds	<i>Glycine max</i>
NM_117721	<i>Arabidopsis thaliana</i> phytochrome D ( <i>PhyD</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
AY348568	<i>Orobanche minor</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Orobanche minor</i>
AY348569	<i>Monotropastrum globosum</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Monotropastrum globosum</i>
X75412	<i>P. crispum</i> <i>phyA</i> mRNA for phytochrome A	<i>Petroselinum crispum</i>
AF544028 AY078504	<i>Stellaria longipes</i> ecotype prairie phytochrome B ( <i>PhyB</i> ) mRNA, complete cds	<i>Stellaria longipes</i>
NM_001251357	<i>Glycine max</i> phytochrome A ( <i>PhyA</i> ), mRNA	<i>Glycine max</i>
L34844	<i>Soybean</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Glycine max</i>
NM_117923	<i>Arabidopsis thaliana</i> phytochrome E ( <i>PhyE</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
U93787	<i>Ipomoea nil</i> phytochrome E ( <i>PhyE</i> ) mRNA, complete cds	<i>Ipomoea nil</i>
NM_122975	<i>Arabidopsis thaliana</i> phytochrome C ( <i>PhyC</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
EU428751	<i>Glycine max</i> phytochrome B-3 ( <i>phyB</i> ) mRNA, complete cds	<i>Glycine max</i>
XM_003533109	<i>Glycine max</i> phytochrome B-like (LOC100799831), mRNA	<i>Glycine max</i>
NM_001123784	<i>Arabidopsis thaliana</i> phytochrome A ( <i>PhyA</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
U56731	<i>Sorghum bicolor</i> phytochrome C ( <i>PhyC</i> ) mRNA, complete cds	<i>Sorghum bicolor</i>
XM_002991073	<i>Selaginella moellendorffii</i> hypothetical protein ( <i>PhyB</i> ), mRNA	<i>Selaginella moellendorffii</i>
AF544029	<i>Stellaria longipes</i> ecotype prairie phytochrome C ( <i>PhyC</i> ) mRNA, complete cds	<i>Stellaria longipes</i>
U56729	<i>Sorghum bicolor</i> phytochrome A ( <i>PhyA</i> ) mRNA, complete cds	<i>Sorghum bicolor</i>
XM_002278574	PREDICTED: <i>Vitis vinifera</i> phytochrome A ( <i>PhyA</i> ), mRNA	<i>Vitis vinifera</i>
NM_100828	<i>Arabidopsis thaliana</i> phytochrome A ( <i>PhyA</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
XM_002318877	<i>Populus trichocarpa</i> phytochrome ( <i>phyA</i> ), mRNA	<i>Populus trichocarpa</i>
AJ001914	<i>Lycopersicon esculentum</i> mRNA for phytochrome A, type 1	<i>Solanum lycopersicum</i>
AB109891	<i>Oryza sativa</i> Japonica Group <i>PHYA</i> mRNA for phytochrome A, complete cds	<i>Oryza sativa</i> Japonica Group
AJ001318	<i>Populus tremula</i> x <i>Populus tremuloides</i> mRNA for phytochrome A	<i>Populus tremula</i> x <i>Populus tremuloides</i>
AB036762	<i>Armoracia rusticana</i> <i>phyA</i> mRNA for phytochrome A, complete cds	<i>Armoracia rusticana</i>
AJ001915	<i>Lycopersicon esculentum</i> mRNA for phytochrome A, type 2	<i>Solanum lycopersicum</i>
NM_001247561	<i>Solanum lycopersicum</i> alternative transcript type 3 ( <i>phyA</i> ), mRNA	<i>Solanum lycopersicum</i>
AJ001916	<i>Lycopersicon esculentum</i> mRNA for phytochrome A type 3	<i>Solanum lycopersicum</i>
AY348567	<i>Cuscuta pentagona</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Cuscuta pentagona</i>
EU428752	<i>Glycine max</i> phytochrome B-4 ( <i>phyB</i> ) mRNA, complete cds	<i>Glycine max</i>
AB018442	<i>Oryza sativa</i> Japonica Group mRNA for phytochrome C, complete cds	<i>Oryza sativa</i> Japonica Group
NM_127435	<i>Arabidopsis thaliana</i> phytochrome B ( <i>PhyB</i> ) mRNA, complete cds	<i>Arabidopsis thaliana</i>
AY345120	<i>Cyrtosia septentrionalis</i> phytochrome A ( <i>phyA</i> ) mRNA, complete cds	<i>Cyrtosia septentrionalis</i>
AY888046	<i>Triticum aestivum</i> putative phytochrome B ( <i>PhyB</i> ) mRNA, complete cds	<i>Triticum aestivum</i>
AB264087	<i>Lotus japonicus</i> <i>phyb</i> mRNA for phytochrome b, complete cds	<i>Lotus japonicas</i>

only high quality reads (quality > 20) were retained. Then, *de novo* assembly of the short single-end read dataset was performed using assembler Velvet, which has been developed for the assembly of short reads using the de Bruijn graph algorithm [37].

### 2.3. Basic local alignment search tool (BLAST)

BLAST finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases, and calculates the

**Table 2**Description of the three detected *Atphy*-like sequences in *Rhazya stricta* genome contigs greater than 1000 bp.

Subject	Query	Score	Identities	E-value
PhyA (NM_100828.3)	Contig no. 52,341	1266	1593/2037 (78%)	0.0
phyC (NM_122975.2)	Contig no. 6,028	893	1243/1737 (72%)	0.0
phyE (NM_117923.7)	Contig no. 6,070	850	1186/1654 (72%)	0.0

statistical significance of matches based on pair-wise alignment method. It can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families (<http://www.ncbi.nlm.nih.gov/BLAST>).

#### 2.4. Eukaryotic GeneMark.hmm program

GeneMark.hmm E-3.9 (<http://exon.gatech.edu/eukhmm.cgi>, A. Lukashin and M. Borodovsky, unpublished) was used for gene finding and annotation of *R. stricta* phytochrome-like genes. The statistical model of genomic sequence organization employed in the GeneMark.hmm algorithm is a HMM with duration [38] or a hidden semi-Markov model (HSMM).

#### 2.5. Sequence alignment and estimation of phylogenetic relationships

MUSCLE [39] was used for multiple-sequence alignment. GenBank accession numbers for phytochrome protein sequences data aligned and analyzed in this study are displayed in Table 1. Maximum likelihood was used to build a tree where the evolutionary rates are free to differ in different lineages. ML trees were constructed using amino acid sequences and the Jones–Taylor–Thornton (JTT) substitution model. To evaluate the reliability of the inferred trees, bootstrap analysis using 1000 replicates was carried out. All the phylogenetic analyses were performed using MEGA 5.2 [40].

### 3. Results

#### 3.1. Next generation sequencing and BLAST searches for phytochrome-like gene family

A total of 73,841,902 single-end short DNA sequence reads was generated for *R. stricta* using the HiSeq 2000 Illumina platform (Illumina, San Diego, CA). The raw sequencing reads have been deposited at EMBL (accession number PRJEB4739). *De novo* assembly using Velvet generated 714,083 contigs and the 45,641 contigs greater than 1000 bp were used in constructing a local database. Five sequences belonging to *Arabidopsis thaliana* phytochrome gene family (i.e. *AtphyABCDE*, accession nos. NM\_100828.3, NM\_127345.3, NM\_122975.2, NM\_117721.1 and NM\_117923.7, respectively) were obtained from GenBank and used for searching the local database to identify contigs with phytochrome-like sequences using BLAST. The BLAST searches resulted in the identification of three phytochrome-like contigs in *R. stricta*. The first contig (number 152,341) was 4955 bp long and had an average depth of coverage of 15.32 and corresponds to *AtphyA*-like sequence. The second contig

(number 6,028) was 7216 bp long with an average coverage of 13.25 and represents an *AtphyC*-like. The third contig (number 66,070) was 7000 bp long with an average coverage of 12.61 and corresponds to an *AtphyE*-like sequence (Table 2).

#### 3.2. Annotation of the three phytochrome gene sequences

GeneMark.hmm was used to identify exon-, intron- and protein-like structures in three phytochrome-like sequences of the *R. stricta* contigs. Each contig consisted of one phytochrome-like gene comprised of three exons and two introns (Fig. 1, Table 3). Contigs 152,341, 6,028 and 66,070 were submitted to GenBank (accession numbers: phyA-HG380749; phyC-HG380750; phyE-HG380751). BLASTn and BLASTp of the GenBank nucleotide and protein databases were performed using the predicted mRNA and protein sequence of each contig to search for homologous sequences.

*Rhazya stricta* *phyA*-like (or *RsphyA*-like) mRNA sequence (contig 152,341) had the highest homology to the *phyA* coding sequence of *Wrightia coccinea* with query coverage equal to 49% and an E-value of 0.0. Protein sequence searches indicated that *RsphyA*-like sequence is most similar to *phyA* of *Solanum tuberosum* with query coverage equal to 96% and an E-value of 0.0. Both results confirmed that contig 152,341 likely contains a full-length copy of the *RsphyA*-like gene (Table 3 and Fig. 1a). Searches of mRNA sequence predicted from the second contig (6,028) identified a *Rhazya stricta* *phyC*-like (or *RsphyC*-like) sequence that had the highest sequence identity to the *phyC* coding sequence of *Vitis vinifera* with query coverage equal to 98% and an E-value of 0.0. The protein sequence query identified the highest identity to *phyC* of *Vitis riparia* with query coverage equal to 99% and an E-value of 0.0. These results confirmed that contig 6,028 likely contains a full-length *RsphyC*-like gene (Table 3 and Fig. 1b). Finally, using the mRNA sequence predicted from the third contig (66,070), *R. stricta* *phyE*-like (or *RsphyE*-like) sequence was found to have high sequence identity to the *phyE* coding sequence of *Ipomoea nil* with query coverage equal to 96% and an E-value of 0.0. Protein sequence queries indicated that this contig had the highest identity with *phyE* of *Ipomoea nil* with query coverage equal to 97% and an E-value of 0.0. Both results confirmed that contig 66,070 likely contains a full-length copy of the *RsphyE*-like gene (Table 3 and Fig. 1c).

Protein sequences of the three identified *R. stricta* *phy*-like genes were analyzed against the pfam database (<http://pfam.sanger.ac.uk/>) to detect the protein family and conserved domains. The analysis revealed the presence of a phytochrome protein family (accession number PF00360). The domains summary of the predicted proteins is shown in Tables 4–6 and Fig. 2.

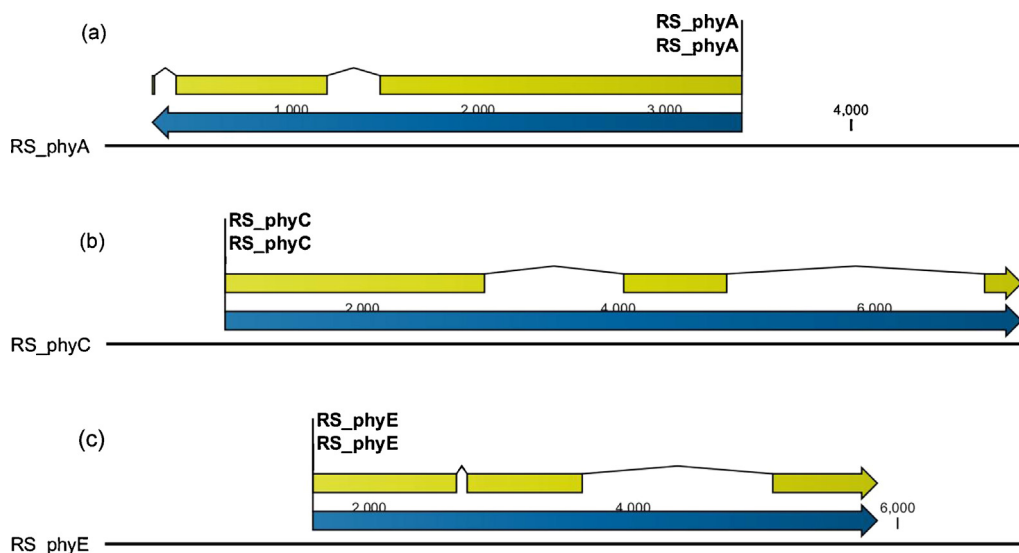


Fig. 1. The relative position for the three phytochrome-like genes (blue), i.e. *RspHYA*-like (a) *RspHYC*-like (b) and *RspHYE*-like (c), and exons (yellow) within the three contigs. Color online.

Table 3

Annotation results of the three contigs with *RspHY*-like sequences showing number, type, range and length of exons (bp).

Gene	Contig No.	Exon No.	Strand	Exon Type	Exon range (bp)		Exon length (bp)
					Start	End	
<i>phyA</i>	15,2341	3	–	Terminal	252	270	19
<i>phyA</i>	15,2341	2	–	Internal	379	1195	817
<i>phyA</i>	15,2341	1	–	Initial	1372	3418	2047
<i>phyC</i>	6,028	1	+	Initial	923	2960	2038
<i>phyC</i>	6,028	2	+	Internal	4035	4851	817
<i>phyC</i>	6,028	3	+	Terminal	6854	7154	301
<i>phyE</i>	66,070	1	+	Initial	1570	2667	1098
<i>phyE</i>	66,070	2	+	Internal	2737	3619	883
<i>phyE</i>	66,070	3	+	Terminal	5050	5852	803

### 3.3. Multi-sequence alignment (MSA) and phylogenetic analysis

The best BLASTp search hits were used to perform multi-sequence alignment. This resulted in 30 sequences from 23 different species. Alignment of the 30 sequences was obtained by gap open penalty of 10 and gap extension penalty of one. Alignment of *R. stricta* phytochromes-like proteins and the 30 amino acid sequences from the BLASTp search showed high sequence identity (Table 7). Phylogenetic relationships among these sequences showed that each phytochrome-like gene of the *R. stricta* grouped with the same type of subunit in the other plant species (Fig. 3).

## 4. Discussion

Using local BLASTn searches, three *phy*-like genes (i.e., *RspHYA*-, *RspHYC*- and *RspHYD*-like) were identified in *Rhazya stricta*. By applying more sensitive protein search tools, such as GeneMark.hmm, it was possible to identify the start and stop codons as well as the exon and intron boundaries of *phy*-like homologues in *R. stricta* (Fig. 1). Domain searches confirmed that both the C-terminal and

N-terminal domains are conserved for all three *Rhazya phy*-like genes (Fig. 2). *phyA* and *phyB* proteins have been shown to be the primary mediators of the phytochrome-mediated development (reviewed in Mathews [41]).

The fact that *phyB* gene was not detected in *R. stricta* could be the result of the protein sequence being too divergent from *Arabidopsis* to be detected by BLAST. This explanation seems unlikely because of the high amino acid sequence identity between the three detected phytochrome genes of *Rhazya* and other angiosperms (96–99%). Alternatively, *phyB* protein may be absent from *Rhazya* and one of the other subunits has taken over its function. *Rhazya stricta* grows in extremely high light intensities in the deserts of western Asia and these conditions may have affected the evolutionary history of phytochrome genes in this plant species. Earlier studies of single and double mutants for *phyA* and *phyB* genes have demonstrated redundancy of function between both genes [42], supporting the hypothesis that that *phyA* may have taken over the function of *phyB* in *Rhazya*. Other studies on transcriptional profiling of etiolated *phyB* mutants subjected to red (R) wavelengths showed similar response to that of the wild-type controls [43,44]. These results suggest that other

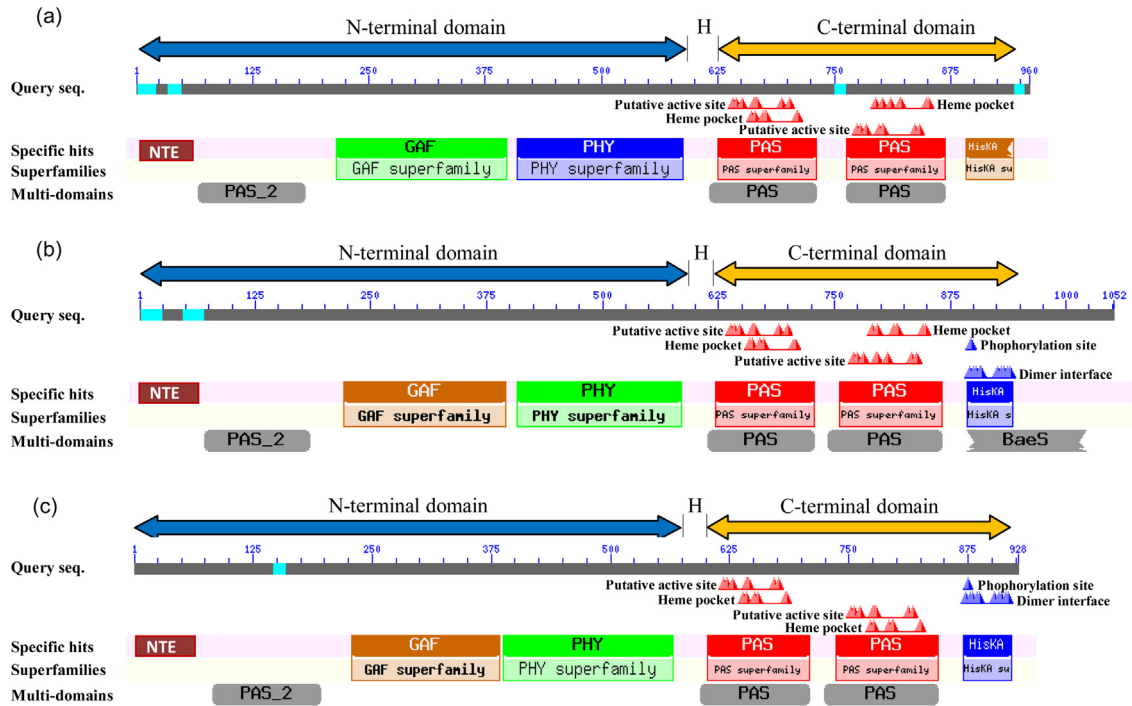


Fig. 2. Domain structure of the three phytochrome-like proteins of *R. stricta* displayed as a Pfam diagram. The figure shows relative position for the identified protein family and N-terminal and C-terminal domains within the three identified proteins, i.e. RsphyA-like (a) RsphyC-like (b) and RsphyE-like (c). Domains are connected by a flexible hinge region (H). Color online.

Table 4

Pfam search results show relative position, bit score and E-value for the identified protein family and domains within *R. stricta* phyA-like identified protein.

	Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
					Start	End	Start	End	From	To			
N-terminal domain	NTE	N-terminal extension	Region	N/A	1	65	1	65	N/A	N/A	65	60.8	4.00E <sup>-09</sup>
	PAS_2	PAS fold	Domain	CL0183	66	182	66	182	2	110	110	135.8	7.30E <sup>-40</sup>
	GAF	GAF domain	Domain	CL0161	215	398	215	398	1	154	154	142.4	1.20E <sup>-41</sup>
	PHY	Phytochrome region	Family	CL0161	409	588	409	586	1	181	183	234.8	3.60E <sup>-70</sup>
C-terminal domain	PAS	PAS fold	Domain	CL0183	616	731	617	731	2	113	113	78.8	2.50E <sup>-22</sup>
	PAS	PAS fold	Domain	CL0183	746	871	747	869	2	111	113	82.0	2.40E <sup>-23</sup>
	HisK	Histidine kinase-related (phospho-acceptor)	Domain	CL0025	891	953	894	948	4	61	68	28.5	1.10E <sup>-06</sup>

Table 5

Pfam search results shows relative position, bit score and E-value for the identified protein family and domains within *R. stricta* phyC-like identified protein.

	Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
					Start	End	Start	End	From	To			
N-terminal domain	NTE	N-terminal extension	Region	N/A	1	67	1	67	N/A	N/A	67	137	1.00E <sup>-35</sup>
	PAS_2	PAS fold	Domain	CL0183	68	184	68	184	1	110	110	145	1.00E <sup>-42</sup>
	GAF	GAF domain	Domain	CL0161	218	396	221	395	5	153	154	133.7	5.70E <sup>-39</sup>
	PHY	Phytochrome region	Family	CL0161	407	586	407	584	1	181	183	234.9	3.30E <sup>-70</sup>
C-terminal domain	PAS	PAS fold	Domain	CL0183	613	728	614	728	2	113	113	67.2	9.80E <sup>-19</sup>
	PAS	PAS fold	Domain	CL0183	743	868	745	866	3	111	113	58.5	4.70E <sup>-16</sup>
	HisK	Histidine kinase-related (phospho-acceptor)	Domain	CL0025	892	952	893	950	6	66	68	27.7	1.90E <sup>-06</sup>

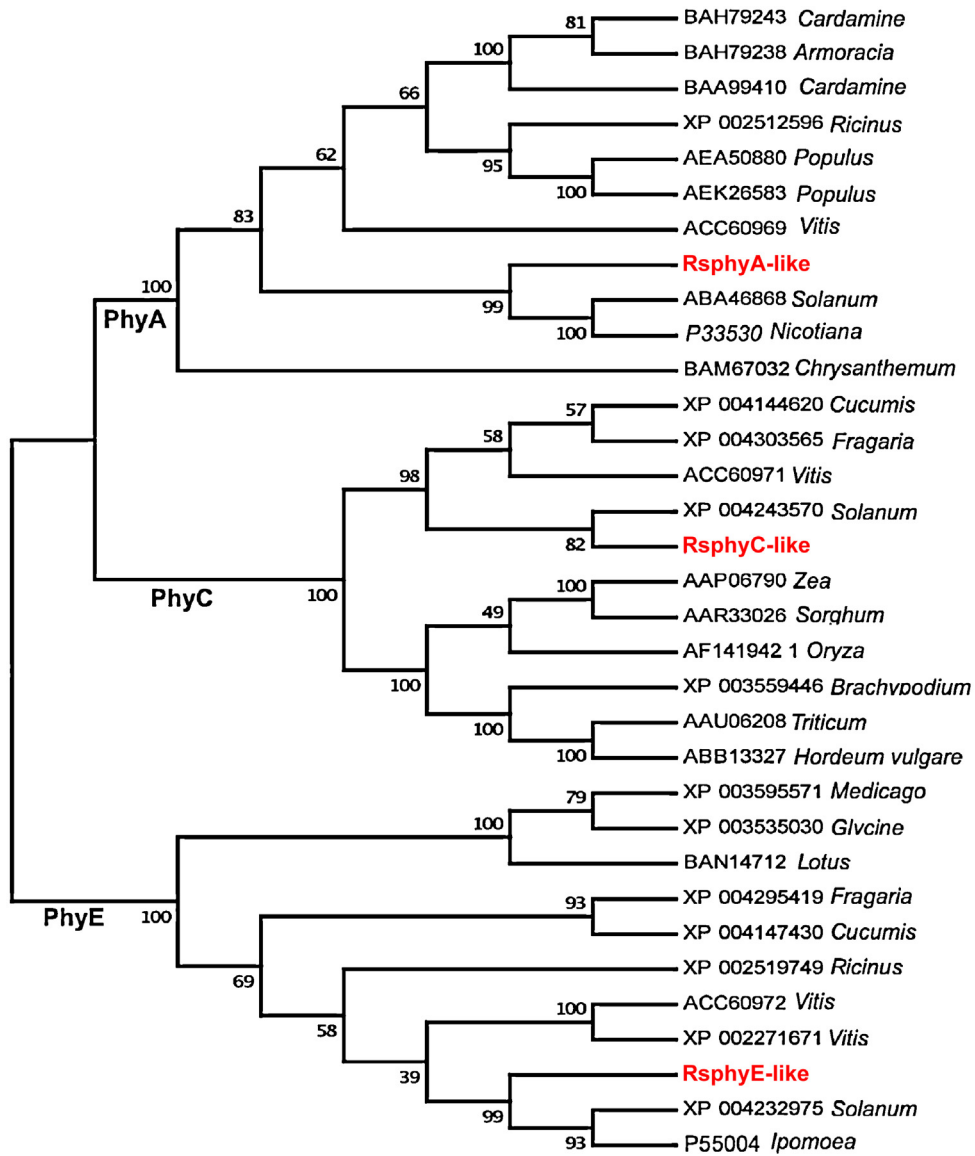


Fig. 3. Phylogenetic analysis of 33 phytochromes including those of *R. stricta* phytochrome proteins. Color online.

Table 6

Pfam search results shows relative position, bit score and E-value for the identified protein family and domains within *R. stricta* phyE-like identified protein.

	Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value
					Start	End	Start	End	From	To			
N-terminal domain	NTE	N-terminal extension	Region	N/A	1	67	1	67	N/A	N/A	67	138	3.00E <sup>-36</sup>
	PAS_2	PAS fold	Domain	CL0183	83	196	83	196	1	110	110	136.7	3.70E <sup>-40</sup>
	GAF	GAF domain	Domain	CL0161	229	371	231	368	3	126	154	104.8	4.30E <sup>-30</sup>
	PHY	Phytochrome region	Family	CL0161	387	566	387	563	1	180	183	237.3	6.10E <sup>-71</sup>
C-terminal domain	PAS	PAS fold	Domain	CL0183	593	709	594	709	2	113	113	85.3	2.30E <sup>-24</sup>
	PAS	PAS fold	Domain	CL0183	724	846	725	844	2	111	113	57	1.40E <sup>-15</sup>
	HisK	Histidine kinase-related (phospho-acceptor)	Domain	CL0025	865	924	869	921	5	59	68	27.9	1.70E <sup>-06</sup>

**Table 7**  
Information for 30 plant phytochrome proteins used for multi-sequence alignment.

No.	Accession	Description	Latin name	Length (aa)
1	AAF66603	phytochrome C	<i>Oryza sativa</i> Indica Group	1137
2	AAP06790	phytochrome C1 apoprotein	<i>Zea mays</i>	1135
3	AAR33026	phytochrome C	<i>Sorghum bicolor</i> subsp. <i>Verticilliflorum</i>	1135
4	AAU06208	phytochrome C	<i>Triticum aestivum</i>	1139
5	ABA46868	phytochrome A	<i>Solanum tuberosum</i>	1123
6	ABB13327	phytochrome C	<i>Hordeum vulgare</i> subsp. <i>vulgare</i>	1139
7	ACC60969	phytochrome A	<i>Vitis riparia</i>	1124
8	ACC60971	phytochrome C	<i>Vitis riparia</i>	1123
9	ACC60972	phytochrome E	<i>Vitis riparia</i>	1124
10	AEA50880	phytochrome A	<i>Populus tremula</i>	958
11	AEK26583	phytochrome A	<i>Populus tremula</i>	1109
12	BAA99410	phytochrome A	<i>Armoracia rusticana</i>	1122
13	BAH79238	phytochrome A	<i>Cardamine nipponica</i>	1122
14	BAH79243	phytochrome A	<i>Cardamine resedifolia</i>	1122
15	BAM67032	phytochrome A	<i>Chrysanthemum seticuspe</i> f. <i>boreale</i>	1121
16	BAN14712	phytochrome E	<i>Lotus japonicas</i>	943
17	P33530	Phytochrome A1	<i>Nicotiana tabacum</i>	1124
18	P55004	Phytochrome E	<i>Ipomoea nil</i>	1115
19	XP_002271671	PREDICTED: phytochrome E	<i>Vitis vinifera</i>	1124
20	XP_002512596	phytochrome A	<i>Ricinus communis</i>	1124
21	XP_002519749	phytochrome B	<i>Ricinus communis</i>	1131
22	XP_003535030	PREDICTED: phytochrome E-like	<i>Glycine max</i>	1120
23	XP_003559446	PREDICTED: phytochrome C-like	<i>Brachypodium distachyon</i>	1140
24	XP_003595571	Phytochrome E	<i>Medicago truncatula</i>	1122
25	XP_004144620	PREDICTED: phytochrome C-like	<i>Cucumis sativus</i>	1119
26	XP_004147430	PREDICTED: phytochrome E-like	<i>Cucumis sativus</i>	1134
27	XP_004232975	PREDICTED: phytochrome E-like	<i>Solanum lycopersicum</i>	1137
28	XP_004243570	PREDICTED: phytochrome C-like	<i>Solanum lycopersicum</i>	1118
29	XP_004295419	PREDICTED: phytochrome E-like	<i>Fragaria vesca</i> subsp. <i>vesca</i>	1162
30	XP_004303565	PREDICTED: phytochrome C-like	<i>Fragaria vesca</i> subsp. <i>vesca</i>	1122

phytochrome family member(s) may be predominantly responsible for perception and transduction of R light in the absence of phyB protein. Recently, phyA protein was shown to play a dominant role in regulating rapid gene expression responses to R light [45]. These and other more recent studies indicate a shuttle role for phyA protein in response to R light, which is normally masked in the presence of phyB [46]. Thus, we propose that absence of phyB gene in *R. stricta* has resulted in *RspHYA* taking over the role of phyB.

Phytochromes play both unique and overlapping roles across the life stages of plants to regulate a range of developmental processes especially at seed germination and seedling stages. The latter stage is known for accumulating the higher levels of alkaloids useful for different pharmaceutical purposes in another medicinally important plant *Pinellia ternata* (Araceae) [47]. The present study provides valuable information for future manipulations of phytochrome gene expression at the seedling stage to enhance resources for better understanding alkaloids and their potential medicinal/pharmaceutical use in *Rhazya stricta*.

#### Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

#### Acknowledgments

The authors gratefully acknowledge the financial support from King Abdulaziz University (KAU) Vice

President for Educational Affairs Prof. Dr. Abdulrahman O. Alyoubi and the KAU Deanship of Scientific Research, Jeddah, Saudi Arabia, represented by the Unit of Strategic Technologies Research through the Project number 431/008-D for the Project entitled: "Environmental Meta-Genomics and Biotechnology of *Rhazya stricta* and its Associated Microbiota".

#### References

- [1] J. Li, G. Li, H. Wang, X. Wang, X. Wang, Phytochrome signaling mechanisms, in: *The Arabidopsis Book* 9, 2011, p. e0148 doi/10.1199/tab.0148.
- [2] J.M. Christie, Phototropin blue-light receptors, *Annu. Rev. Plant Biol.* 58 (2007) 21–45.
- [3] L. Rizzini, J.J. Favory, C. Cloix, D. Faggionato, A. O'Hara, E. Kaiserli, R. Baumeister, E. Schäfer, F. Nagy, G.I. Jenkins, R. Ulm, Perception of UV-B by the *Arabidopsis* UVR8 protein, *Science* 332 (2011) 103–106.
- [4] H. Smith, Sensing the light environment: The functions of the phytochrome family, in: R.E. Kendrick, G.H.M. Kronenberg (Eds.), *Photomorphogenesis in Plants*, 2nd ed, Kluwer, Dordrecht, The Netherlands, 1994, pp. 377–416.
- [5] X. Zheng, W. Suowei, H. Zhai, P. Zhou, M. Song, L. Su, Y. Xi, Y. Li, Y. Cai, F. Meng, L. Yang, H. Wang, J. Yang, *Arabidopsis* phytochrome B promotes SPA1 nuclear accumulation to repress photomorphogenesis under far-red light, *Plant Cell* 25 (2013) 115–133.
- [6] R.A. Sharrock, T. Clack, Patterns of expression and normalized levels of the five *Arabidopsis* phytochromes, *Plant Physiol.* 130 (2002) 442–456.
- [7] S. Mathews, R.A. Sharrock, Phytochrome gene diversity, *Plant Cell Environ.* 20 (1997) 666–671.
- [8] H. Abe, K.T. Yamamoto, A. Nagatani, M. Furuya, Characterization of green tissue-specific phytochrome isolated immunochemically from pea seedlings, *Plant Cell Physiol.* 26 (1985) 1387–1399.
- [9] Y. Shimazaki, L.H. Pratt, Immunochemical detection with rabbit polyclonal and mouse monoclonal antibodies of different pools of phytochrome from etiolated and green *Avena* shoots, *Planta* 164 (1985) 333–344.
- [10] J.G. Tokuhisa, S.M. Daniels, P.H. Quail, Phytochrome in green tissue: spectral and immunochemical evidence for two distinct molecular



- species of phytochrome in light-grown *Avena sativa* L, *Planta* 164 (1985) 321–332.
- [11] M. Furuya, Phytochromes: their molecular species, gene families, and functions, *Annu. Rev. Plant Physiol.* 44 (1993) 617–645.
- [12] L.H. Pratt, M.M. Cordonnier-Pratt, B. Hauser, M. Caboche, Tomato contains two differentially expressed genes encoding B-type phytochromes, neither of which can be considered an ortholog of *Arabidopsis* phytochrome B, *Planta* 197 (1995) 203–206.
- [13] M. Hirschfeld, J.M. Tepperman, T. Clack, P.H. Quail, R.A. Sharrock, Coordination of phytochrome levels in *phyB* mutants of *Arabidopsis* as revealed by apoprotein-specific monoclonal antibodies, *Genetics* 149 (1998) 523–535.
- [14] S. Mathews, M. Lavin, R.A. Sharrock, Evolution of the phytochrome gene family and its utility for phylogenetic analyses of angiosperms, *Anal. Missouri Bot. Garden* 82 (1995) 296–321.
- [15] R. Alba, P.M. Kelmenson, M.M. Cordonnier-Pratt, L.H. Pratt, The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms, *Mol. Biol. Evol.* 17 (2000) 362–373.
- [16] G. Bae, G. Choi, Decoding of light signals by plant phytochromes and their interacting proteins, *Annu. Rev. Plant Biol.* 59 (2008) 281–311.
- [17] T. Clack, S. Mathews, R.A. Sharrock, The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: the sequences and expression of *PHYD* and *PHYE*, *Plant Mol. Biol.* 25 (1994) 413–427.
- [18] D.E. Somers, R.A. Sharrock, J.M. Tepperman, P.H. Quail, The *hy3* long hypocotyl mutant of *Arabidopsis* is deficient in phytochrome B, *Plant Cell* 3 (1991) 1263–1274.
- [19] S. Mathews, R.A. Sharrock, The phytochrome gene family in grasses (Poaceae): a phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms, *Mol. Biol. Evol.* 13 (1996) 1141–1150.
- [20] M. Koomneef, E. Rolf, C.J.P. Spruit, Genetic control of light inhibited hypocotyl elongation in *Arabidopsis thaliana* (L.) Heynh, *Z. Pflanzenphysiol.* 100 (1980) 147–160.
- [21] J.W. Reed, P. Nagpal, D.S. Poole, M. Furuya, J. Chory, Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development, *Plant Cell* 5 (1993) 147–157.
- [22] A. Nagatani, J. Chory, M. Furuya, Phytochrome B is not detectable in the *hy3* mutant of *Arabidopsis*, which is deficient in responding to end-of-day far-red light treatments, *Plant Cell Physiol.* 32 (1991) 1119–1122.
- [23] G.C. Whitelam, H. Smith, Retention of phytochrome mediated shade-avoidance responses in phytochrome deficient mutants of *Arabidopsis*, cucumber and tomato, *J. Plant Physiol.* 139 (1991) 119–125.
- [24] P.R.H. Robson, G.C. Whitelam, H. Smith, Selected components of the shade-avoidance syndrome are displayed in a normal manner in mutants of *Arabidopsis thaliana* and *Brassica rapa* deficient in phytochrome B, *Plant Physiol.* 102 (1993) 1179–1184.
- [25] K.J. Halliday, M. Koomneef, G.C. Whitelam, Phytochrome B and at least one other phytochrome mediate the accelerated flowering response of *Arabidopsis thaliana* L. to low red/far-red ratio, *Plant Physiol.* 104 (1994) 1311–1315.
- [26] P.F. Devlin, K.J. Halliday, N.P. Harberd, G.C. Whitelam, The rosette habit of *Arabidopsis thaliana* is dependent upon phytochrome action: novel phytochromes control internode elongation and flowering time, *Plant J.* 10 (1996) 1127–1134.
- [27] I.C. Jang, R. Henriques, H.S. Seo, A. Nagatani, N.H. Chua, *Arabidopsis* phytochrome interacting factor proteins promote phytochrome B polyubiquitination by COP1 E3 ligase in the nucleus, *Plant Cell* 22 (2010) 2370–2383.
- [28] W. Hu, K.A. Franklin, R.A. Sharrock, M.A. Jones, S.L. Harmer, J.C. Lagarias, Unanticipated regulatory roles for *Arabidopsis* phytochromes revealed by null mutant analysis, *Proc. Natl. Acad. Sci. USA* 110 (2013) 1542–1547.
- [29] P.A. Salomé, M. Oliva, D. Weigel, U. Krämer, Circadian clock adjustment to plant iron status depends on chloroplast and phytochrome function, *EMBO J.* 32 (2013) 511–523.
- [30] S. Nazimuddin, M. Qaiser, Apocynaceae, in: E. Nisir, S.I. Ali (Eds.), *Flora of Pakistan*, 148, 1983, pp. 26–27.
- [31] S. Mukhopadhyay, G.A. Handy, S. Funayama, G.A. Cordell, Anticancer indole alkaloids of *Rhazya stricta*, *J. Nat. Prod.* 44 (1981) 696–700.
- [32] A.U. Rahman, M.M. Qureshi, Applications of modern 2D-NMR techniques in structure elucidation of natural products, *Pure Appl. Chem.* 62 (1990) 1385–1388.
- [33] S.A. Gilani, A. Kikuchi, Z.K. Shinwari, Z.I. Khattak, K.N. Watanab, Phytochemical, pharmacological and ethnobotanical studies of *Rhazya stricta* Decne, *Phytother. Res.* 21 (2006) 301–307.
- [34] R.J. Aerts, V. De Luca, Phytochrome is involved in the light-regulation of vindoline biosynthesis in *Catharanthus*, *Plant Physiol.* 100 (1992) 1029–1032.
- [35] M. Höft, R. Verpoorte, E. Beck, Growth and alkaloid contents in leaves of *Tabernaemontana pachysiphon* Stapf (Apocynaceae) as influenced by light intensity, water and nutrient supply, *Oecologia* 107 (1996) 160–169.
- [36] N.L. Gawel, R.L. Jarret, A modified CTAB DNA extraction procedure of *Musa* and *Ipomoea*, *Plant Mol. Biol. Rep.* 9 (1991) 262–266.
- [37] D.R. Zerbino, E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Res.* 18 (2008) 821–829.
- [38] L.R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* 77 (1989) 257–286.
- [39] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucl. Acids Res.* 32 (2004) 1792–1797.
- [40] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.* 28 (2011) 2731–2739.
- [41] S. Mathews, Phytochrome evolution in green and nongreen plants, *J. Heredity.* 96 (2005) 197–204.
- [42] G.C. Whitelam, P.F. Devlin, Roles of different phytochromes in *Arabidopsis* photomorphogenesis, *Plant Cell Environ.* 20 (1997) 752–758.
- [43] L. Ma, J. Li, L. Qu, J. Hager, Z. Chen, H. Zhao, X.W. Deng, Light control of *Arabidopsis* development entails coordinated regulation of genome expression and cellular pathways, *Plant Cell* 13 (2001) 2589–2607.
- [44] J.M. Tepperman, M.E. Hudson, R. Khanna, T. Zhu, S.H. Chang, X. Wang, P.H. Quail, Expression profiling of *phyB* mutant demonstrates substantial contribution of other phytochromes to red-light-regulated gene expression during seedling de-etiolation, *Plant J.* 38 (2004) 725–739.
- [45] J.M. Tepperman, Y.S. Hwang, P.H. Quail, *phyA* dominates in transduction of red-light signals to rapidly responding genes at the initiation of *Arabidopsis* seedling de-etiolation, *Plant J.* 48 (2006) 728–742.
- [46] J. Li, W. Terzaghi, X.W. Deng, Genomic basis for light control of plant development, *Protein Cell* 3 (2012) 106–116.
- [47] J. Zeng, Z. Peng, J. Song, X. Ma, S. Wei, Y. Luo, The dynamic change of total alkaloids content from *Pinellia ternate*, *Zhong Yao Cai.* 27 (2004) 471–473.