

Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family

Robert A. Sharrock¹ and Peter H. Quail

Plant Gene Expression Center, Albany, California 94710 USA

Phytochrome is a plant regulatory photoreceptor that mediates red light effects on a wide variety of physiological and molecular responses. DNA blot analysis indicates that the *Arabidopsis thaliana* genome contains four to five phytochrome-related gene sequences. We have isolated and sequenced cDNA clones corresponding to three of these genes and have deduced the amino acid sequence of the full-length polypeptide encoded in each case. One of these proteins (*phyA*) shows 65–80% amino acid sequence identity with the major, etiolated-tissue phytochrome apoproteins described previously in other plant species. The other two polypeptides (*phyB* and *phyC*) are unique in that they have low sequence identity (~50%) with each other, with *phyA*, and with all previously described phytochromes. The *phyA*, *phyB*, and *phyC* proteins are of similar molecular mass, have related hydropathic profiles, and contain a conserved chromophore attachment region. However, the sequence comparison data indicate that the three *phy* genes diverged early in plant evolution, well before the divergence of the two major groups of angiosperms, the monocots and dicots. The steady-state level of the *phyA* transcript is high in dark-grown *A. thaliana* seedlings and is down-regulated by light. In contrast, the *phyB* and *phyC* transcripts are present at lower levels and are not strongly light-regulated. These findings indicate that the red/far red light-responsive phytochrome photoreceptor system in *A. thaliana*, and perhaps in all higher plants, consists of a family of chromoproteins that are heterogeneous in structure and regulation.

[Key Words: *Arabidopsis thaliana*; phytochrome; photomorphogenesis; plant gene family; angiosperm evolution; gene expression]

Received June 29, 1989; revised version accepted September 4, 1989.

Light is an important environmental factor controlling plant growth and development. Not only does light provide energy for photosynthesis, but plant growth patterns and a large number of plant developmental events, such as formation of leaf primordia, plastid development, and induction of flowering, are also responsive to light cues (Salisbury and Ross 1985). Physiological experiments suggest that two major plant regulatory photoreceptor systems are active in perception of light cues: one system sensing shorter wavelength blue and UV-A light, and the second sensing predominantly longer wavelength red/far red light (Shropshire and Mohr 1983). Up to this time, only the red/far red-responsive photoreceptor phytochrome has been isolated (Vierstra and Quail 1986). The quality and quantity of incident red and far red light influence a broad range of responses throughout the plant life cycle and in a wide variety of organs and tissues, indicating that the phytochrome re-

ceptor is one of the primary components of the regulatory system for higher plant photomorphogenesis.

The molecular properties of phytochrome have been determined most extensively for the abundant chromoprotein species purified from dark-grown (etiolated) oat tissue (Vierstra and Quail 1983). This species is a dimer of 124-kD subunits (Jones and Quail 1986), each of which contains a covalently attached linear tetrapyrrole chromophore (Rudiger and Scheer 1983). The complete amino acid sequences of the abundant, etiolated-tissue phytochrome from oat (Hershey et al. 1985), zucchini (Sharrock et al. 1986), pea (Sato 1988), rice (Kay et al. 1989a), and corn (Christensen and Quail 1989), have been derived from the corresponding nucleic acid sequences. The molecular mechanism of action of phytochrome is not known. No enzymatic or specific protein or nucleic acid-binding activity has been assigned to the chromoprotein. Nonetheless, it is clear that the unique photochromic properties of phytochrome are responsible for its regulatory function. Phytochrome, both in plants and as a purified chromoprotein, exists in ei-

¹Present address: Department of Biology, Montana State University, Bozeman, Montana 59717 USA.

ther of two spectrally distinct conformations: P_r (red light-absorbing, $\lambda_{\max} = 666$ nm) or P_{fr} (far red light-absorbing, $\lambda_{\max} = 730$ nm) (Vierstra and Quail 1983). These conformations are photointerconvertible; irradiation with red light converts P_r to P_{fr} and, conversely, irradiation with far red light converts P_{fr} back to P_r . For most phytochrome responses, conversion to P_{fr} induces the response, and conversion back to P_r cancels the induction (Shropshire and Mohr 1983). In this way, phytochrome functions as a reversible regulatory switch for plant photomorphogenic events.

In addition to simply being induced by red light, many phytochrome responses show sensitivity to the balance of red and far red light (Smith 1986) and to the intensity and duration of illumination (Mancinelli and Rubino 1978; Mandoli and Briggs 1981). The complex action spectra, variability of response and escape times, and differential level of far red reversibility of phytochrome responses have frequently been interpreted within the context of the known molecular properties of the apparently homogeneous phytochrome extracted from etiolated plant tissue (Kronenberg and Kendrick 1986). Nonetheless, for some time, the possibility has been recognized that less abundant forms of phytochrome might exist and that these might play important or even predominant roles in photoregulation. Indeed, some physiological observations, such as the *Zea* and *Pisum* 'paradoxes' (Hillman 1967) and *in vivo* spectroscopic data (Jabben and Holmes 1983), are very difficult to explain on the basis of one pool of homogeneous phytochrome. Tokuhsa and Quail (1983) presented the first direct evidence that extracts of fully green oat tissue contain two pools of phytochrome: a greatly reduced level of the etiolated-tissue form and a second, immunochemically distinct form. Further experiments confirmed these observations (Shimazaki and Pratt 1985; Tokuhsa et al. 1985) and extended them to a dicot species, pea (Abe et al. 1985). The green-tissue phytochrome from oats is of lower molecular mass, has different spectral properties, and is more stable *in vivo* in the presence of light when compared to the etiolated-tissue form (Tokuhsa et al. 1985).

The biochemical and physiological evidence for a second pool of phytochrome led us to screen for plant genomic sequences and cDNA clones that cross-hybridize under low stringency conditions with nucleic acid probes derived from the coding sequence of the abundant, etiolated-tissue phytochrome. This approach has been instrumental in the identification of novel components of receptor and signal transduction systems in animals. It has been used successfully to isolate members of several gene families such as steroid and thyroid hormone receptors (Thompson et al. 1987; Giguere et al. 1988), protein kinases (Ohno et al. 1988; Schaeffer et al. 1989), and potassium channels (Butler et al. 1989). We chose to screen for phytochrome-related sequences in *Arabidopsis thaliana*, a small cruciferous plant that exhibits typical photoresponsive characteristics and has many features that distinguish it as a model plant system for molecular genetic studies

(Meyerowitz 1987). In particular, we anticipated that the small genome size of *A. thaliana* might assist in establishing the minimum number of phytochrome genes present in higher plants. Here, we demonstrate that phytochrome in *A. thaliana* is encoded by a small gene family of at least three, but more likely four or five, members. We present the primary sequence of three of these gene products and their phylogenetic relationship to each other and to previously described phytochromes. In addition, we present preliminary studies on the apparent differential regulation of these genes.

Results

Isolation and sequence of phytochrome cDNA clones from A. thaliana

A probe generated by nick-translation of the zucchini phytochrome cDNA clone pFMD1 (Lissemore et al. 1987) was used initially to screen an *Arabidopsis* genomic library, and a single *A. thaliana* genomic phytochrome clone was isolated (R. Sharrock, C. Gatz, and P. Quail, unpubl.). Comparison of phytochrome sequences from distantly related plant species has shown previously that the highest conservation of amino acid and nucleic acid sequence occurs in the amino-terminal half of the receptor, around the chromophore attachment site (Sharrock et al. 1986). Therefore, a 0.9-kb single-stranded DNA (ssDNA) probe was prepared from the *A. thaliana* genomic clone covering this region of the phytochrome-coding sequence (0.9-kb probe; Fig. 1). Blots of total *A. thaliana* DNA digested with restriction enzymes show multiple bands of hybridization to this probe (Fig. 1A), indicating that the *Arabidopsis* genome contains multiple copies of phytochrome-related coding sequence.

The 0.9-kb probe was used to screen a λ gt10 size-fractionated cDNA library made from poly(A)⁺ RNA isolated from 3-week-old green *A. thaliana* leaves (Crawford et al. 1988). Several clones were isolated that hybridized to the probe under high-stringency wash conditions, and the insert from one of these clones, λ A2-3, was sequenced. The λ A2-3 insert (Fig. 2A) contains the complete amino acid-coding sequence for a 124-kD polypeptide (*phyA*; Fig. 2B), 5'- and 3'-noncoding sequence, and a poly(A) tail. Clones showing less stable hybridization to the 0.9-kb probe were separated into two classes on the basis of restriction enzyme analysis, and the inserts from representatives of these classes, λ A7-5 and λ A1-1, were sequenced (Fig. 2A). The λ A7-5 insert contains the complete amino acid-coding sequence for a 129-kD polypeptide (*phyB*; Fig. 2B), and λ A1-1 contains the complete coding sequence for a 124-kD polypeptide (*phyC*; Fig. 2B). Both of these inserts contain 5'- and 3'-noncoding sequence and poly(A) tails. Restriction enzyme analysis and hybridization of the three *A. thaliana* phytochrome cDNA clones to genomic DNA blots under high stringency conditions show that the central coding regions of the *phyA*, *phyB*, and *phyC* genes correspond to the bands indicated in the

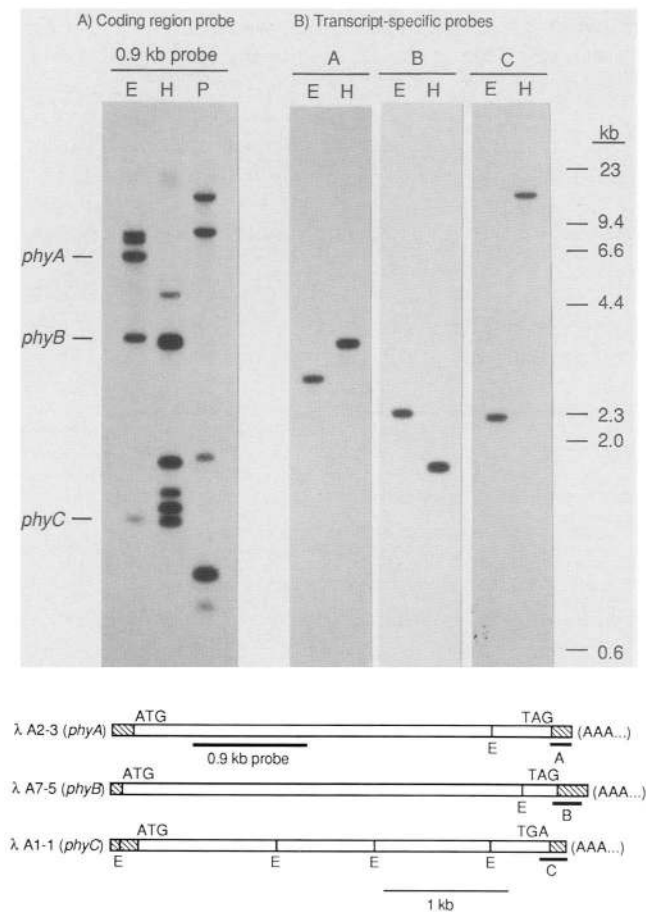


Figure 1. Southern blot analysis of total *A. thaliana* DNA hybridized with a phytochrome-coding region probe (0.9-kb probe) or with probes specific for the *phyA*, *phyB*, and *phyC* transcripts. Diagrams of the cDNA inserts from the λ gt10 *phy* clones are shown below the blots, and the internal *EcoRI* sites and the positions of the probes are indicated. Restriction enzymes are (E) *EcoRI*; (H) *HindIII*; (P) *PstI*.

EcoRI digest on the Southern blot in Figure 1A (data not shown). No cDNA clones corresponding to the two highest molecular weight bands in this digest were recovered in this screen.

In all three cDNA clones, the initiator methionine codon for the large phytochrome open reading frame (ORF) is not the first AUG present at the 5' end of the mRNA. The upstream open reading frames (URFs) in each mRNA encode short peptides, 3–12 residues in length, before an in-frame translation termination codon is reached (Fig. 2A). The large ORF of *phyB* contains two methionines in its first 54 amino acids. Initiation of translation at the first methionine gives rise to the 129-kD protein shown in Figure 2B. Initiation at the second methionine would produce a 124-kD protein similar to the other phytochrome apoproteins; however, there is currently no reason to invoke preferential initiation at this site. To eliminate the possibility that the first in-frame ATG of the *phyB* sequence was introduced through a cDNA cloning artifact, the nucleotide se-

quence of the amino-terminal coding region of the *phyB* cDNA was confirmed by sequencing the same region of an *A. thaliana* genomic *phyB* clone (R. Sharrock and P. Quail, unpubl.).

Comparison of *A. thaliana* phytochromes A, B, and C

It has been reported previously that there are multiple phytochrome genes in oat (Hershey et al. 1985). However, oat is hexaploid and the genes described encode almost identical polypeptides (>98% identical). Genomic Southern blot analyses of pea and rice DNA have been interpreted to indicate the presence of only single phytochrome genes (Sato 1988; Kay et al. 1989b) and cDNA cloning from zucchini (Lissemore et al. 1987), pea (Sato 1988), and rice (Kay et al. 1989b) indicated the presence of only a single type of phytochrome transcript in etiolated tissue of these plant species. Our results in *Arabidopsis* (Figs. 1 and 2) contrast markedly with these previous reports. To compare the three *A. thaliana* phytochromes, the deduced amino acid sequences of *phyA*, *phyB*, and *phyC* have been aligned in Figure 2B in such a way as to minimize the number of gaps introduced. A consensus sequence of residues conserved in all three polypeptides is shown below the alignments. In all pairwise combinations, the *phyA*, *phyB*, and *phyC* polypeptides are approximately equally related to one another, 49–52% identical (Table 1). This level of sequence conservation indicates significant structural heterogeneity and distant evolutionary origins (see Phylogeny of phytochrome genes, below).

To better visualize the structural relatedness of the *phyA*, *phyB*, and *phyC* polypeptides, the distribution of amino acid sequence conservation along each pair of the aligned phytochrome sequences is presented in Figure 3 as a linear plot of percent identity within a 9-amino-acid moving window. Short conserved regions are observed over almost the entire lengths of *phyA*, *phyB*, and *phyC*. Similar regions of the three polypeptides are either highly conserved or prone to substitutions, and no large structural domains are conserved in two of the proteins and lost in the third. These data indicate that the overall structure of these proteins is conserved. Consistent with this interpretation, hydropathy profiles for *phyA*, *phyB*, and *phyC* are very similar over their entire lengths (Fig. 4). Notable deviation from this conserved phytochrome structure occurs at the ends of the *phyB* polypeptide in the form of amino- and carboxy-terminal extensions. The 35-residue *phyB* amino-terminal extension is unusual in its high glycine content (37%) and the presence of numerous amino acid doublet and triplet repeats (Fig. 2B), the functional significance of which is not known.

Phylogeny of phytochrome genes

The amino acid sequences of the three *A. thaliana* phytochromes and the published sequences of phytochromes from oat (Hershey et al. 1985), zucchini (Sharrock et al. 1986), pea (Sato 1988), rice (Kay et al. 1989a), and corn (Christensen and Quail 1989), have

Sharrock and Quail

A 1

URF-B

A CGTCTGCTCTGTTGTTAGGAGGACAAATAATAGAGAGGGCTAGCACAAGAGAGAAGTGGTGTTCGAGGCCAAAGAT

B GTCTCCGATAACTAGTGGGATGATGATTACCCCTAAATCTTCTGCTCAAGGTAATCTCTGAGAAATTTCTCAAATTCAAAATCAA

C CAACTTACAAGTTCTCTCAGCTTCTCTCCACCCTAAGGAGGAACGTTCCAAGATCCCTTCTCAGAGAATTCACGAAAAATCTTCAACAATG

*B URF-C

URF-A 200

A TTAGAATTTAACTATAACAAAAGCCTCTGCAGAGTGTGACTAGTCACAAGATCTGATCATGGCTTCTTGAAGTCTTCTTCTTCTTCTTCTTTAA

B CGGCATGCTTCCGGAGTCGGGGTAGTGGCGGTGGCGGTGGCCGTTGGCGGAGAAGAAGAACCGTCTCAAGTCACACTCCATAAACCGAAGA

C AAACCTTAAATGGAGAATCAATCCGATCCCTTGAACTCTTGGTGTGCTCTCACTCCATTTCTGAAATTTCAATGCTTTGTGATCTTCTGCAGATTCT

*A *C 300

A AGGAAAAATGTCAGGCTCAGGCCGACTCAGTCTCTGAGGGCTCAAGGCATCAAGGCACAGCGCTAGGATCATTCGCGAGACCCTGTAGATGCGA

B GGAGGAGAACAGCTCAATCGTCCGGAAACGAAATCTCTCAGACCAAGAAGCAACTGAATCAATGAGCAAAGCAATTCAACAGTACACCGTCGAGCAA

C TTGGAAGAGAAGAAAGAAATGTCATCGAACACTTCAGGAAGCTGTTCTACTAGATCTAGACAAAATCTCGAGTTTCTTCAAGTTCTGTCGAGCAA

400

A AACTCCATGCTGATTTGAGGAGTCAGGCAGCTCCTTTGATTACTCAACCTCAGTGGCGTCTCAGTGGCCGCTGGGAGAAACAGCCAC

B GACTCCAGCCGTTTCCGAAACATCCGGCAATCAGGAAATCAATCGACTACTCAACTCAAAACGACGAGCTACGGTTCTCTC

C AGCTACAGCAAACTTCGAAGAATCTGAGCGTATTATGACTATTCAGCTTCAATAAACTTGAACATGCAAGTCTTCTCTGTGAGATTCC

500

A AAGGCTGACAAAGTACCAGACTTATCTTCATCATATACAGAAGGAAAGCTGATTGAGCCTTCGGTGTGTTTACTTGCCTGGATGAGAAGACTTC

B GTACTGTAGCAACAGATCACAGCTTATCTCTCGAATTCAGCGAGGTGGTTACATTCAGCCTTCGGATGTATGATCGCGCTCGATGATCCAGTTCT

C TTCTTCAGCTGTCTCAACGTACTTACAGAAGTTCAGAGGAGGATGTTGATTCAACCTTTGGTGTGTTAATCGTGTGATGAGAAAACTT

600

A AAAGTATGTCATACAGCGAAGTGCATCTGAGCTGTGACAATGGCCAGTCTGACAGTCCTAGTGTGGCGAACCCCTGTCTAGGCATGGGACAG

B CGGATCATCGGTTACAGTAAAAACGCCAGAGAAATGTTAGGATATGCCTCAATCTGTTCTACTCTTGAAGAACCTGAGATTCTAGCTATGGGAACTG

C AAAGTCAATGCTTTAGTAAAACACTCAAGAGATGTTGGTGTGATTCCACATACAGTACCAAGTATGGAGCAGCGTGAAGCTTTGACTATAGAACTG

700

A ATATAAGAGCTTTTCACTGCCTAGTGGCTCTGCATGACAGAAGCCCTGGATTTGGAGATGTCTCTTGTGAATCCCATTTCTTGTCATGGCAG

B ATGTGAGATCTTTGTTCACCTCTCGAGCTCGATTCTACTCGAGCTCTGTTGGTCTGAGAGATFACCTGTTAAATCCGGTTGGATCCATTCGAA

C ATGTGAAATCAATGTTCTGCTCCAGGTTGTTCTGCTTTGGAGAAAGCTGTTGACTTGGTGAGATTAGTATTTGAAATCTTACAGCTTCAATGTAG

800

A GACTTCTGCAAAGCCCTTTTATGCGATTATCCACAGGGTTACAGGAGCATCATCGACTTTGAACCCGTGAAGCCTTATGAACTCCCATGACAGCT

B GAAACTGGTAAACCGTTTACGCCATTTCTCAGAGTTGATGTTGGTGTGTTATTGATTAGAGCCAGTAGAACCTGAGCCTTCGCTTCATTTATT

C GTCTTCAAGTAAACCTTTTATGCGATTCTGCATCGGATTGAGGAAGGCTTGTATAGATTGGAGCCTGTGAGTCTGTAGGTGCTGTGACTGCT

900

A GCTGGTCCCTTACAATCATAAGCTCGCTGCCAAAGCAATCACTAGGCTGCAATCTTACCAGCGGGAGTATGAAAAGCTTTGTGATACAATGGTTC

B GCTGGTCCCTTCAATCGCAGAACTCGCGGTTGGTGGATTTCTCAGTACAGGCTCTTCTGTTGGGAGATATAAGCTTTTGTGTGACACTGTCTGG

C GCCGGGCTTTAAGATCGTATAAGCTTGGCGGAAATCGATTTCAGAGTTCAGGACATGGCTAGTGGGAAATATGTTGTTGGTGTGATGCTTTGGTTA

1000

A AAGAGTCTTTGAACTCACGGGTATGACAGGGTGTGCTTATAAGTTTCAAGATGATGACGGTGAGGTGTCTCCGAGTTACAAAACCTGGGCT

B AAAGTGTAGGGACTTGAATGATGCTGTTATGAGTTTATAAGTTTCAAGATGAGCAGTGAAGAAGTTGAGCTGAGAGTAAACGAGACGATTT

C AGGAAGTTAGTAAATTAAGTGTATGATAGGGTGTGGTGTATAAGTCCATGAGGATGGCATGGGAAAGTATTGCTGAATGCTCCGCGAAATGAT

1100

A GGAGCTTATCTTGGGCTGCATTATCTGCCACCGACATCCCTCAAGCAGCCGTTTCTGTTTATGAAGAACAAGTCCGGATGATAGTGTGATGCAAT

B AGAGCTTATATTGGAGTGCATTTACTGCTAGTATCTTCAAGCTCAAGGCTCTGTTTAAAGCAGAACCGTCCGATGATAGTGTGATGCAAT

C GGAACCTTATCTTGGGTTGCATTACTCCGCTACTGATATACCGCAAGCTCCAGAGATTCTGTTTATGAGAAAAAGAGTATTGTGATTGTTCA

1200

A GCAAAACATGCTAGGGTCTTCAAGATGAAAAGCTTTTCTTTGACCTTACCCTTGTGGTCCACCTTAGAGCACCAGCAGCTGCAATTTGCATGACA

B GCCACACTGTCTTGTGGTCCAGGAGCATAGGCTAACTCAGTCTATGTGCTTGGTGGTCTACTCTTAGGGCTCCTCATGGTGTGCTCTCAGTATA

C GCGGTTCCGTTAAAGTCTTCAAGATTAAGAGTCTCTCACGCCAAATAAGTCTTTCTGGATCTACTTTGAGAGCTCCTCATGGTGTGACGACAGTATA

1300

A TGCCAAATGATGATTAATGATCTCTGGTTATGGCGGTTGAGTTAACGAGGAAGTGGAGAAGGGATGTCTCTGATGCTACTACACAGCTCAA

B TGCTTAACATGGGATCATTGCGCTCTTAGCAATGGCGGTTAATAATGAAGAAATGAAGATGATGGAGCAATGTAGTGTGGAAGAG

C TGAGTAAATAGGATCAGTGGCGTCTTTGTCATGTCTGTAATATCAATGGTGTAGTGTATGATGAGATGAACAGAGATTTACAGAC

1400

A GAGAAAGAGACTATGGGGTATTAGTGGTGTGTCACAATACGACTCCGAGGTTTGTTCATTTCTCTCAGGATGCTGCTGTGAGTTCTAGCTCAAGTGT

B TCTGATGAGCCTTTGGGTTTGGTGTGTTGCAATCACACTTCTTCTGCTGCTTACCCGTTCCGCTAAGGATATGCTTTGTGAGTTTGTGATGAGCTTT

C TGGCAGACACTTATGGGCTTTGGTGTGTTGCATCACGCAAGTCTTAGATTTGTTCCGTTCCATTACGATATGCTTTGAAATCTTGACTCAAGTATTT

1500

A GCCATACAGCTCAATAAGAGGTTGAACTCGATAACCAGATGGTGGAGAAGAACATTTTGGCCAGCCAGACACTCTGTGGGATATGCTGATGCTGATG

B GGTTTACAGTTAAACATGGAATTCAGTGTAGCTTTGCAATGTGACAGAAACGCGTTTGGAGAAGCAGACACTGTTATGTGATATGCTTCGCGTACT

C GCGGTGAGATCAACAAAGAAGCGGAATCAGCTGTTCTGTTGAAAGAGAAGCGTATTTGCAAACTCAGAGTGTGCTATGTGACATGCTTTCCGCAATG

1600

A CTCCACTGGGTATTGTGTGCAAAAGCCCCAACATAATGGACCTTGTGAAATGTGATGGAGCAGCTCTCTTGTATAAAGCAAGATATGAAAACCTGGGAAC

B CGCCTGTGAAATTTGTACACAGAGTCCAGTATCATGGACTTAGTGAATGTGACGGTGCAGCAATTTCTTTACCAGGGAAGTATACCAGCTGGGTGT

C CACCAATAGGATAGTCACTCAATCACCAAATAAATGATCTTGTAAATGTGATGGAGCAGCAATATATTACAGAGACAACTTCGGCTCTTAGGAT

1700

A AACTCCAAGTGAATCCACCTGCAGGATAGCTTCACTGTTGTGTAATACCACATGATGATCAACGGTGTGAGCAGTATGATGATGATGATGATGATG

B TGCTCCTAGTGAAGTTCAGATAAAGATGTTGGAGTGGTGTGCTTCGCAATCATGCGGATTAACCGGATTAAGCACTGATGTTTAGCGCATCGGG

C TACTCCCAGAGACACAAATTAGAGATCTAATGACTGGTTCCTAAAAGTGTAGGAGAAACACTGGCTTACCACGAAAGTCAATAAGGAGTGTG

1800

A TTTCTAGGGCTCTATCTCTCGGGATTCGGATGTTGGGATGGCAGCTGTGAGGATATCATCGAAAGACATGATTTTCTGGTTCCGTTCTCATAACGGCTG

B TATCCCGGTGCAAGTTCGGTGGGATGCTGTGGGCTATGGCAGTTGCATATATACAAAAGAGACTTTCTTTTGGTTTTCGATCTCACACTGGCA

C TATCCGGATGCTCTGTTCTTGGGAGTCAATATGTTGAAATGGCTGCCGTATATATTTCCGAAAAGATTTCTTTTCTGGTTCGGTCTGAGACTGCAA

1900

A GTGAAGTGAATGAGAGGTTGCGAAGCATGATCCAGATGATAGGATGATGCAAGGAGAATGCACCCAAGTGCATGTTCAAGGCTTTCTTGAAGTGTG

B AAGAAATCAAATGGGAGGCGCTAAGCATCATCCGAGGATAAAGATGATGGCAACCAAGATGCATCTCTGTTGCTTTCAGGCTTTCTTGAAGTGTG

C AACAGATCAAGTGGGTTGGTGAAGACAGATCCTAATGACAGA . . .GATGGAAGAGAAATGCATCTTAGATCCTCATTCAAGGCTTTATGAAAATAGT

2000

A CAAGACAAGGAGTTTACCTTGAAGGACTATGAGATGGATGCATACACTCCTTGAACCTATTTTGAAGAAATGCTTTCAAGGATAGTGAACACT . . .

B TAAGAGCCGAGTCAAGGAACTGGGAACTGCCGAAATGGATGCGATTACTGCTCCAGCTTATTTCTGAGAGACTCTTTTAAAGAAATCTGAGGCGGCTATG

C CAGGTGAAAAGTGTGCCCTGGGATGACATGAAAATGGATGCAATTAATCTCTGAGCTAATAATAAAGGCTCATTGCAAGAGGAGCAATCA

Figure 2. (See p. 1750 for legend.)

Novel phytochromes in *Arabidopsis*

2100
A GATGTGAATACAAAGGTCATTACTCGAAGCTAAATGATCTCAAAATGATGGTATACAAGAAGCTAGAAGCTGTGACCAAGTGAAGTGGTTTCGTT
B AACTCTAAAGTTGTGGATGGTGGTTACAGCATGTAGGGATATGGCGGGGAAACAGGGGATGATGAGTTAGGTGACAGTTCAGAGAGATGGTTAGGC
C AAGACTGTTGTGGATGTCCCACTTGTGGATAATAGGGTTCAGAAGGTAGATGAATGTTGTGTTATCGTGAATGAAATGGTGGCGT

2200
A TAATTGAGACTGCTACGGTGCCAAATATTGGCGGTTGATCTGATGGACTGGTTAATGGTTGGAACACGAAAATGCTGAGCTGACTGGTCTTTCGGTTGA
B TCATTGAGACTGCAACTGTTCTATATTCGCTGTGGATGCGGAGGCTGATCAATGGATGGAACGCTAAGATTGACAGTTCCTCAGTTGA
C TGATTGATACAGCAGCTGTCCCACTTTGCGGTTGATGCTCTGGTGTATAAATGGTTGGAATTCFAAGCGGCTGAGGTAAACAGGATGGCAGTTGA

2300
A TGAAGCAATCGGGAAGCAATTCCTCACA . . . CTGTTGAAGATCTCTAGTGGAAATCGTTAAAAGGATGCTAGAGAAGCAATGAGAGCAACTGAGGAG
B AGAAGCTATGGGGAAGTCTCTGGTTCTGATTTAATATACAAAGAGAATGAAGCAACTGTCAATAAGCTTCTTCTCGTCTTTGAGAGGGGACGAGGAA
C ACAAGCAATAGGCAAACT . . . STATCAGATCTCTGAGGACGATCTGTAGAAACCGTGAAGAACATGTTAGCCTTGGCTCTCGAAGGTAGTGAAGAA

2400
A CAGAATGTCCAGTTTGTAGATCAAGACACATCTGTCCAGGGCTGATGTGGGCCAAATAGTTTAGTTGTAATGCATGCGCAAGTGAAGATCTCCATGAAA
B AAGAATGTGGAGGTTAAGCTGAAACTTTTCAGCCCGAATCAAGGAAAGCAGTTTTTGTGGTTGTAATGCTTGTCCAGCAAGGACTACTTGAAACA
C CGTGTGCTGAGATCAGATCAGACGATTTGGTCTAAAGGAAAGCAGTCCGGTTGAGTTAGTTGTCACACTTGTGTAGCAGAGATATGACGAAATA

2500
A ACGTGGTTGGGGTGTGTTTTGTAGCCATGATCTACTGGCCAGAAGACTGTGATGGACAAGTTTACCGGGATTGAAAGGTGATTACAAGGCAATCATCCA
B ACATTGTCCGGCTTTGTTTTGTTGGACAAGCAGTTACTAGTCAGAAAATCGTAATGGATAAGTTTCATCAACATACAAGGAGATTACAAGGCTATTGTACA
C ATGTTCTTGGTGTATGCTTCATTGGACAAGATGTTACAGGCCAGAAAACGCTTACTGAAAATATAGCCCGGTGAAAGGAGATTATGCCCAATCATGTTG

2600
A AAATCCAAACCCGCTGATCCCGCAATATTGTTACCGATGAGTTTGGATGGTGCACAGAGTGAATCCAGCAATGTCAAAGTTAACCGGTTTGAAGCGA
B TAGCCAAACCCCTTAATCCCGCAATTTTGTCTGCTGACGAGAACAGCTGTGCTGGAATGGAACATGGCGATGGAAGAGCTTACCGGTTGGTCTCCG
C GAGCCCTCCACACTCATCCCAATTTTTATAACCAATGAAATGGGGTATGCTCAGAGTGGAAACACGCAATGCAGAAGCTCTCTGGGATAAAGAA

2700
A GAGGAAGTATTGACAAAATGCTCTTAGGAGAAGTATTGGGACCGAGAAGTCAATGTTGCTCTAAAGAATCAAGAAGCCTTTGTAACCTTGGGATG
B AGTGAAGTATTGGGAAATGATTGTCCGGGAAGTGTGGG AGCTGTGCATGCTAAAGGGTCTGATGCTTAAACCAAGTTCATGATTTG
C GAAGAAGTTGCAATAAATCTTCTCGGGAGGTTTTTACCACAGATGATTTAGTTGCTGCTTAAAGACCATGACACTTAAACGAGTGTAGAAATG

2800
A TGCTGAACAATGCTGTGACCAGTCAA . . . GATCCAGATAAAGTATCGTTTGTCTTTTACAAGAGGTGGCAAGTATGTGGAGTGTCTGTTGTGTGAG
B TATTGCATAATGCGATTGGTGGCCAA . . . GATACGGATAAGTTCCCTTTCCCATCTTTGACCGCAATGGGAAGTTGTTCCAGGCTCTATTGACTGCAAA
C GTTTCATGCTGTGATTCTGGCCAAAAGAACATAGAGAAGCTTTTATTGGCTTTTACCATCGTGTAGTTAGCTTATCGAGGCAATGCTTTCTGCAAA

2900
A TAAGAACTGGACAGGAAAGGTGTAGTGACAGGTGCTTCTGTTTCTGCAACTTGCAGCCATGAGCTGCAGCAAGCGCTCCATGTTCAACGTTTAGCT
B CAAGCGGTTAGCCTCGAGGAAAGGTTATGGGGCTTTCTGTTTCTGCAAAATCCGAGCCCTGAGCTGCAGCAAGCTTTAGCAGTCCACGAGGCGAG
C CAAAAGGACTGATATTGAGGAAAGGTTTACCGGGTTTTATGCTTTTTGCAAGTACCTAGTCCAGAACTCCAATATGCTCTACAGGTTACAGAAATATCA

3000
A GAGCGAACCCGAGTGAAGAGACTAAAGGCTTAGCATAACAAAAGACAGATCAGGAATCCGCTATCTGGGATCATGTTTACAAGGAAAATGATAGAGG
B GACACAGAGTGTTCACGAAGGCAAAAGAGTGGCTTATATTGTCAGGTGATAAAGAATCCTTTGAGCGGTATGCGTTTCGCAAACTCATGTTGGAG
C GAGCATGCAATGCTGTGCCCTCAACAAATGGCATATCTCCGCCATGAAGTGAAGGACCCGAAAAGGCAATATCCTTCTTCAAGATTGCTCCATT

3100
A GTACTGAATTAGGACCAGAGCAAAAGCGGATTTGCAAACTAGCGCGTTATGTGAGAAGCAACTAAGCAAGATCCTCGATGATTGGATCTTGAAGCAT
B CCACAGACTTGAACGAGGACAGAAAGCAGTTACTTGAAACAAGTGTCTTCTGCGAGAAACAGATCTCAAGGATCGTCCGGGACATGGATCTTGAAGCAT
C CATCTGGATTAAGTGAAGACAAAAGCGGCTCCTGAGGACAAGCGTTTTATGTCAGGGAGCAGTTAGCCAAAGTCAAAAGCAGCTCAGCATAGAGGAAAT

3200
A CATTGAAGGATGCTTGGATTGGAAATGAAAGAATCACCTTAAATGAAGTGTGACTGCTTCCACAAGTCAAGTAAATGATGAAGAGTAAACGAAAGAGT
B TGAAGCTGATCATTTGTCTTAAAGAGGGAAGTTTTTCTTGGAAAGTGTCTAAACCGGATTTGTAAGTCAAGCGATGTTCTTATTAAGGCAAGGAGT
C CGAAGAGGCTATGTGAACTGATTGACAGCAATTCGGCTGCAGGAATCCCTGGAAGCAGTTGTAAGCAAGTGTGAGGCTGAGCATAGAACGTTAA

3300
A GTTCGGATAACAAATGAGACCGGAGAAGAAGTAAATGCTGACACTTTGATGGAGACAGTATTAGGCTTCAACAAGTCTTGGCAGATTTTATGCTGATGG
B CTTGAGCTGATCCGTCATTCGCAAGAGATCAATCAATAGAGGTTTTTGGAGACAGATAAGGATTCACAGCTCCTGGCTGAGTTTCTGCTGAGTA
C GTACAAATCAGTGGGATTTATCCTCAAGAAGTTTCAATCAATGATGTTGATGGAGCAACTTAAAGCTTACAGCAATCCTTTGAGACACTATTAAGCA

3400
A CTGTAACCTTACACCATCCGAGGTCAGCTAACTGTTTCAGCTTCCCTG AGGAAGGATCAGCTCGGGCTTCTGTGCATCTTGTAAATCTAGA
B TAATCCGGTATGCACCATCTCAAGAGTGGTGGAGATCCATTTAAGCCAACTTCAAGCAAAATGGCTGATGGA TTCGCCCATCCGACAGA
C GCATACGCTTCCAGCTGCATTTAGAGGATTTGTGTCTCATTTCAAGGTAATTCACGGATAGAAGCTATAGGAAAAGAAATGAAAGAGTGAAGTCTGA

3500
A GATCAGGTTAACGCATACCGGAGCTGGGATACCTGAGTTTACTAAACAAATGTTGGGACT . . . GAGGAAGATGTGTGAGAAGAAGGATTGAGCTTA
B ATTCAGAAATGGCGTGTCCAGGTAAGGTTGCTCCAGAGCTAGTCCGAGACATGTTCCATAGCAGCAGGTTG . . . ACAAGCCTGAAGGTTTAGGTTA
C GTTCAGGATAATACCCGGCACCAGGACTGCTGAGGATCTGGTAAGAGAGATGTTTACGCTTGTGAGAAAGGAAACATCAAGGGAAGGTTTGGGATTA

3600
A ATGGTTAGCCGGAAACTGTTGAAGCTGATGAAT . . . GGAGATCTCAGTACTTGAGACAAGCTGGGAAATCAAGTTTCAATATCACTGCGGAACCTCGCTG
B AGCGTATGTCGAAAGATTTTAAAGCTAATGAAC . . . GGTGAGTTCAATACATCCGAGAAATCAGAAACGGTCTATTTCTCATCATCTTGAACCTCCCTG
C CACATTACCCAGAAGCTGGTGAACCTCATGGAGAGAGAACATTGAGATACCTCAGAGAGTCTGAAATGTCAGCCTTGTGATCTCACAGAATTTCCCT

3700
A CAGCAAAACAGTATGTTCCCAAAAAGAAAAGGGGCTGCGCTGATATAAAAATAGTCACTGGTGTCTTTGCTTGTAACTTTCTTATCGCTTTTGTGTTTCG
B TACCTCGAAAGCGACCATTTGCTCAACTGCTAGTGGAGTGGTGACATGATGCTGATGATGCCATAATGATCACACTTCAGTTGGTATGAGAGTTTGTATCA
C TGAATTTGAAGCTGAAGACCTGTCTACAAGATTACATTTTATTGAAATGATGTTGGTGTGTTTTGAAAGTCTTCAATTTTTTTTCTCAGATATATAAATGTA

PRIMER C PRIMER A 3800
A TTTTCAAATTCAGTAACGATGAAATATCCATCCATTTACATCTTCTGTTGAACTCTTTTCTGAAAGCTGTAATATGGATGCATATCTAATCTC (AAA . .)
B TTGTATGAGTGTGTTGTGCTTAACGACGCTCGGAGGAGGATAGAAAAGTTTTTTTTTCTCCGGTGAGATTAGTAGAGAAGAGGAGATTATTGCGGTT
C TGTTTAGTTACCGTAACCTAATACATATATATATAGGAAAGTTAATTTG (AAA . .)

3900
PRIMER B
B CAGCTCAGCTCGCCGAAAAAAGCTAACAGTGTGTAGAGAATTTCAAGACTTTGTTTGTGCTGTGTAATGACAACTCCGAGAGAAACAAAAACA
B ATGAGAT (AAA . .)

Figure 2. (See p. 1750 for legend.)

Sharrock and Quail

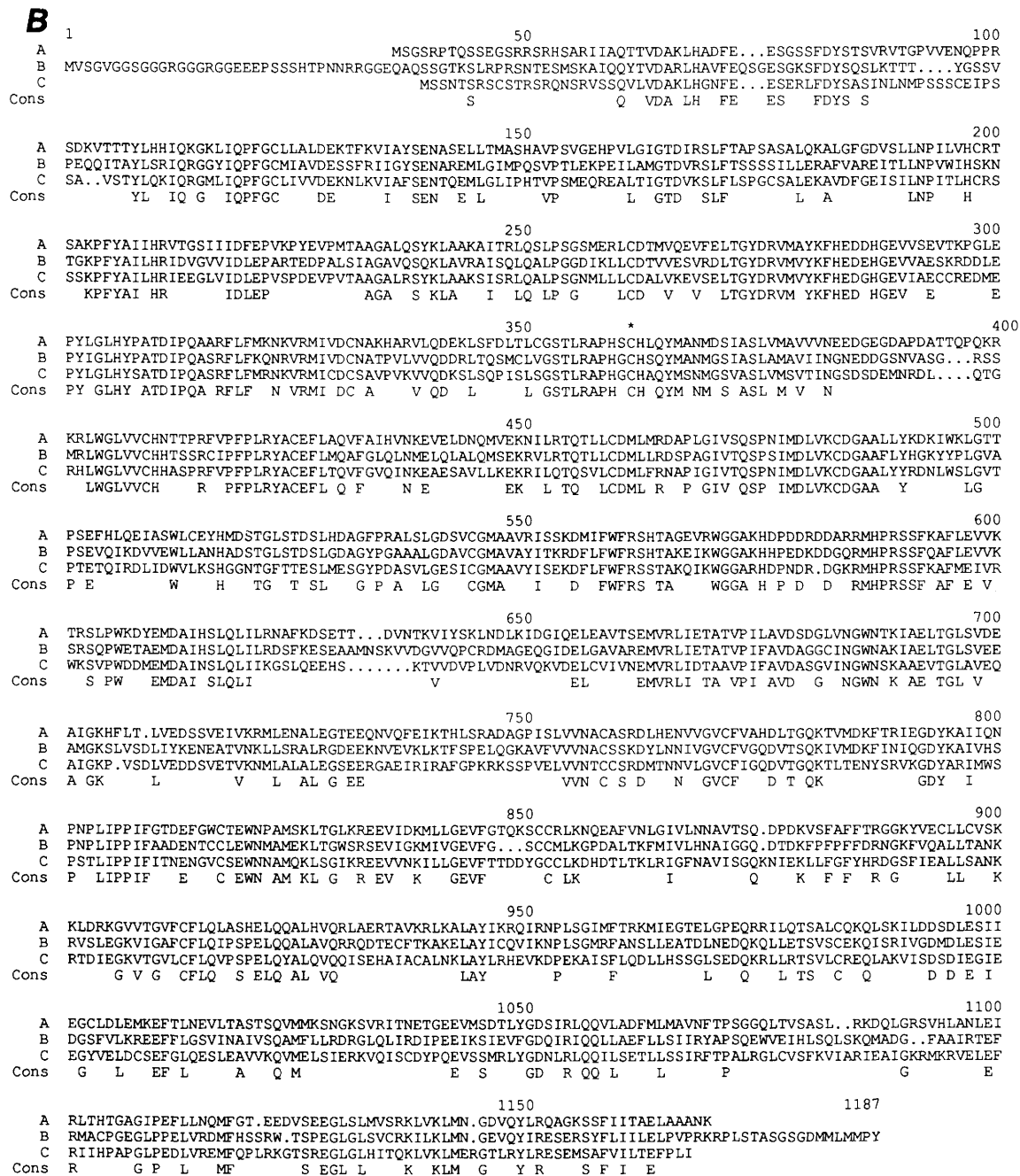


Figure 2. (A) Nucleotide sequences of the cDNA inserts from λ A2-3(*phyA*), λ A7-5(*phyB*), and λ A1-1(*phyC*). The coding regions of the three sequences have been aligned so that they correspond to the peptide sequence alignment in B. Because the initiator ATG codons for the three sequences do not line up in this alignment, the first nucleotide of the *phyB* cDNA sequence has arbitrarily been given position 1. The initiator ATG codons for the phytochrome ORFs are boxed and marked above the alignment with an asterisk. Termination codons are also boxed and marked. Sequences at the 3' ends of the cDNAs, to which antisense synthetic oligonucleotides were annealed and elongated to make transcript-specific hybridization probes, are underlined and labeled. Primer A was elongated to a *Pst*I site at position 3598, primer B to an *Rsa*I site at position 3601, and primer C to a *Pfl*MI site at position 3512 (see Materials and methods). URFs at the 5' ends of the cDNAs are underlined and labeled. (B) Polypeptide sequences of *phyA*, *phyB*, and *phyC* were deduced from the nucleotide sequences in A. The polypeptides have been aligned in such a way as to maximize homology. Residues conserved in all three polypeptides are shown below the alignment. The cysteine residue (361) that corresponds to the chromophore attachment site identified in purified oat phytochrome is indicated by an asterisk.

Table 1. Percent amino acid sequence identity among phytochromes from various plant species and *A. thaliana* *phyA*, *B*, and *C*

	<i>phyA</i> phytochromes						<i>phyB</i>	<i>phyC</i>
	monocot			dicot				
	oat	rice	corn	zucchini	pea	<i>phyA</i>		
Oat	—	89	88	64	65	64	48	49
Rice	89	—	88	64	65	64	49	49
Corn	88	88	—	64	65	64	48	49
Zucchini	64	63	63	—	78	79	52	51
Pea	64	64	64	78	—	79	51	52
<i>phyA</i>	63	63	63	79	79	—	52	52
<i>phyB</i>	46	46	46	49	48	49	—	51
<i>phyC</i>	48	48	48	51	51	52	49	—

For each pair of aligned sequences, the number of identical residues was divided either by the total number of positions in the alignment, including gaps and extensions, (below the diagonal) or by the number of amino acids in the shorter of the two sequences (above the diagonal) and expressed as a percent. Values derived for comparisons within monocot or dicot *phyA* sequences are boxed; values for comparisons between monocot and dicot sequences are enclosed by dashed lines. References for previously published phytochrome sequences are oat (Hershey et al. 1985), rice (Kay et al. 1989a), corn (Christensen and Quail 1989), zucchini (Sharrock et al. 1986), and pea (Sato 1988).

been aligned, and the percent identical residues in each pairwise combination has been calculated in two ways. First, the number of identical amino acids has been divided by the total number of positions occupied by res-

idues in the alignment, including gaps and amino- and carboxy-terminal extensions. These values are displayed in the lower left of Table 1, below the diagonal, and are an index of the degree of structural relatedness of the paired polypeptides. Second, the number of identical amino acids has been divided by the number of residues in the shorter of the two sequences, ignoring gaps and extensions (Table 1, upper right above the diagonal). Values calculated in this way provide an index of the extent of evolutionary relatedness of the paired sequences (Doolittle 1981; Feng et al. 1985). These two methods of calculation yield only minor differences in the values obtained, and these differences are not considered further here. The values determined (Table 1) indicate that all previously described phytochrome polypeptides are significantly more related to the *A. thaliana phyA* protein than to *phyB* and *phyC*. These previously characterized sequences correspond to the abundant etiolated-tissue phytochromes in these species and are very likely functional homologs of one another. We propose that the designation *phyA* be extended to these sequences. Within this group, the three monocot *phyA* sequences are highly related to each other (88–89% identical), the three dicot *phyA* sequences are less related to each other (78–79% identity), and comparisons across monocot/dicot lines show 63–65% identity. These results are similar, in general, to results obtained for phylogenetic comparison of glyceraldehyde-3-phosphate dehydrogenase and chalcone synthase sequences from various angiosperm plant species (Martin et al. 1989). In contrast, the *Arabidopsis phyB* and *phyC* sequences are unique in that they are equivalently and highly divergent from each other and from all of the *phyA* phytochromes (Table 1).

The data in Table 1 can be interpreted most readily as

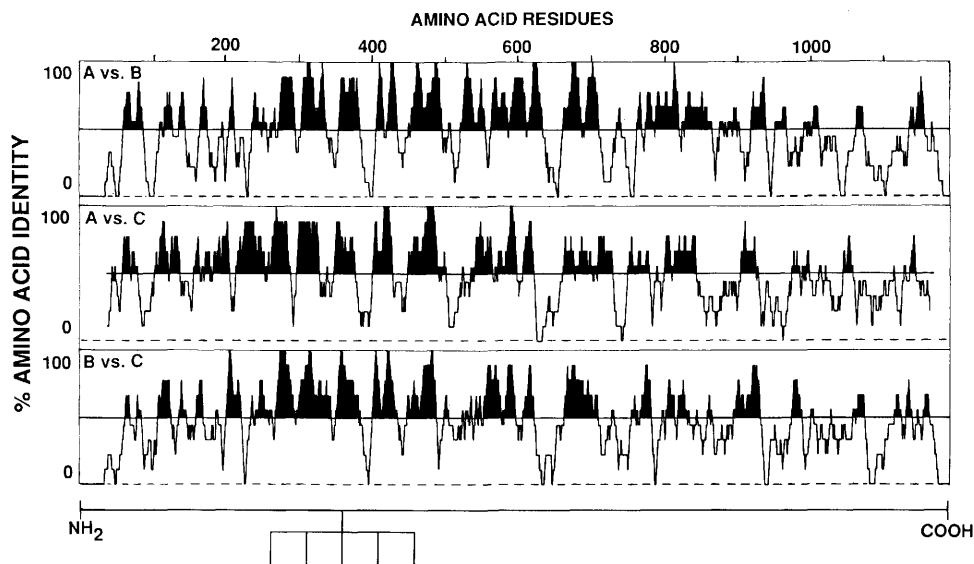


Figure 3. Local level of amino acid sequence identity for the three pairwise alignments of *phyA*, *phyB*, and *phyC*. The number of identical residues within a window of 9 amino acids is expressed as percent identity and plotted at the middle position of the window. Gaps in the alignment in Fig. 2B are counted as mismatches. Shaded areas correspond to regions of >50% identity. A schematic representation of the longest polypeptide, *phyB*, is shown below the plots, indicating the position of the chromophore attachment site.

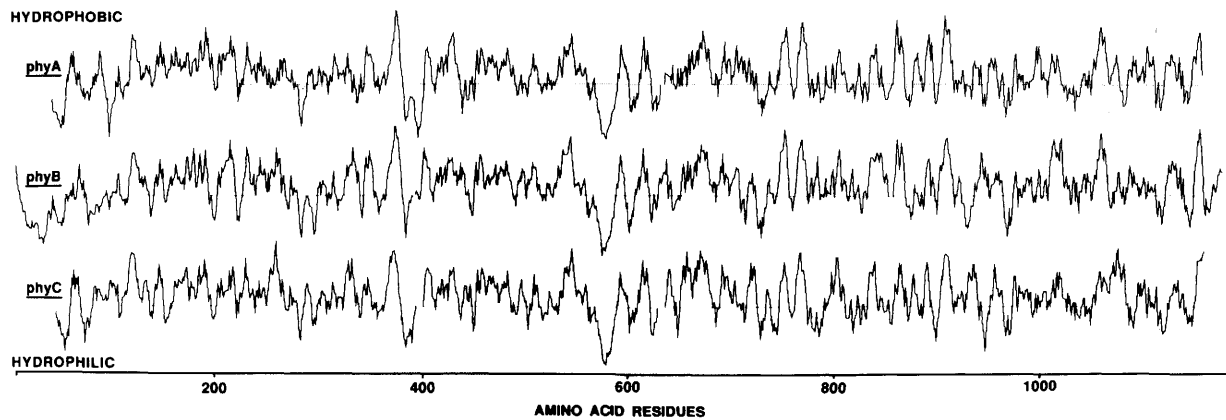


Figure 4. Hydropathy profiles of *phyA*, *phyB*, and *phyC*. Hydropathy analysis was performed according to Kyte and Doolittle (1982), using a window of 9 amino acids. The profiles have been aligned in the same way as the sequences in Fig. 2B, with gaps introduced at the same position.

defining a phylogenetic tree containing a tripartite branching of the three major phytochrome types (A, B, and C) from a precursor gene followed by subsequent divergence of the *phyA* genes (Fig. 5). Parsimony analysis using the PHYLIP programs of Felsenstein (1985) of the nucleotide sequences for the eight phytochrome genes yields an unrooted phylogenetic tree, consistent with that shown in Figure 5 (R. Sharrock and P. Quail, unpubl.). The tree indicates that the trifurcation of the phytochrome gene family into types A, B, and C is an ancient evolutionary event. If the branch point for divergence of the monocot and dicot *phyA* sequences in Figure 5 corresponds to the divergence of the monocots and dicots during angiosperm evolution, 100–300 million years ago (Lidgard and Crane 1988; Martin et al. 1989), the gene duplication events that gave rise to *phyA*, *phyB*, and *phyC* occurred before that, much earlier in the evolution of vascular plants.

Expression of the *phyA*, *phyB*, and *phyC* mRNAs

We used transcript-specific ssDNA hybridization probes derived from sequences at the 3' ends of the cDNA

clones (see Materials and methods, Figs. 1 and 2A) to determine the patterns of expression of the three phytochrome genes. The transcript-specific probes detect unique bands on total *A. thaliana* DNA Southern blots (Fig. 1B). When hybridized to Northern blots of total RNA isolated from 5-day-old *A. thaliana* seedlings grown under various light conditions, all three probes detect RNAs 4.0–4.4 kb in length (Fig. 6), consistent with the sizes of the cDNAs that were isolated (3.6–3.8 kb).

The level of *phyA* mRNA is high in dark-grown tissue, is not strongly affected by a pulse of red light, but is markedly reduced within a few hours after transfer to white light (Fig. 6). In addition, the *phyA* probe hybridizes to more than one RNA transcript. The lower molecular weight (4.0 kb) *phyA* transcript appears to be strongly down-regulated by white light, whereas the higher molecular weight (4.4 kb) transcript is clearly visible only in the white light-irradiated sample (Fig. 6). A situation similar to this has been described for the gene encoding the *phyA* homolog in pea, where the multiple transcripts were shown to be the result of transcription initiation at multiple start sites within a complex pro-

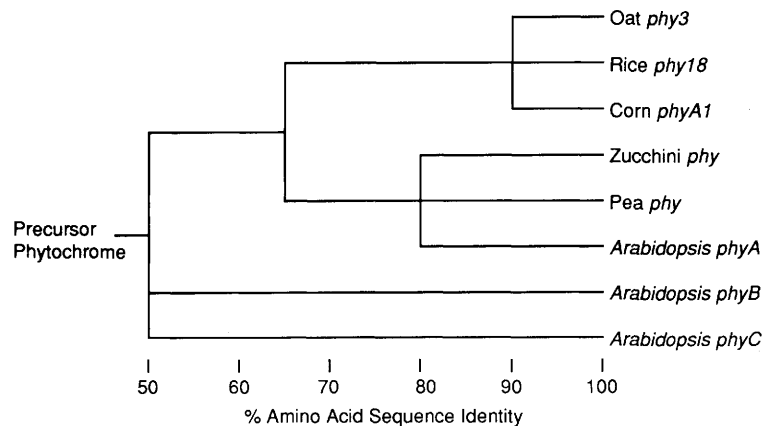


Figure 5. Deduced phylogeny of phytochrome polypeptides. Percent amino acid sequence identity between each pairwise combination of phytochromes (Table 1, values above the diagonal) was used to group related sequences and as a measure of phylogenetic distance in constructing the tree.

moter (Sato 1988). A complex promoter structure analogous to the pea *phyA* promoter is located upstream of the *A. thaliana phyA* gene (R. Sharrock and P. Quail, unpubl.). The *A. thaliana phyB* and *phyC* mRNAs are less abundant than the *phyA* mRNA in dark-grown tissue (5–10%) and are not strongly regulated by the light conditions tested, although the *phyB* transcript level shows a small transient increase following transfer to white light (Fig. 6).

Discussion

We used the technique of low stringency hybridization to identify and isolate *A. thaliana* cDNA clones that are related to the red light-responsive photoreceptor phytochrome and presented the sequences of three such cDNAs and their derived polypeptides. These sequences include the *A. thaliana* homolog of the previously characterized abundant etiolated-tissue phytochrome (*phyA*) and identify two new classes of phytochrome apoproteins (*phyB* and *phyC*). In addition to these three proteins, there is evidence from DNA blot analysis for the presence of one or two additional related sequences in the *A. thaliana* genome. We propose that *A. thaliana* contains a family of at least three, and potentially five, phytochrome genes whose products are diverse, red light-responsive regulatory photoreceptors. This proposal has implications for the mechanism of regulatory light perception in all plants. Hormone and growth factor receptors in animal systems are frequently members of gene families or superfamilies that encode numerous related receptor structures (Evans 1988; Yarden and Ullrich 1988). Structural homology within these families reflects general similarity in their mode of action, such as conserved DNA-binding or protein kinase domains, whereas differential activities of individual members of the families likely result from variations in ligand specificity, restricted tissue or cell-type localization, or different pathways of cellular activation. By analogy, the family of homologous but markedly divergent phytochrome polypeptides that we have described share structural features but may perform widely different roles in the regulation of higher plant photomorphogenesis.

Analysis of the previously published sequences of phytochrome from oat, rice, corn, zucchini, and pea indicates that these proteins are all homologs of the *A. thaliana phyA* gene product. The relatively abundant *phyA* phytochrome from dark-grown plant tissue is a chromoprotein that exists in two spectrally distinct conformations, P_r and P_{fr} , and photoconversion between these two conformations underlies the role of phytochrome as a regulatory molecule. The *phyB* and *phyC* gene products contain regions of high sequence similarity to *phyA* phytochrome, notably at and around the chromophore attachment site, and the calculated hydrophobic properties of the three proteins are very similar. Though this strongly suggests that *phyB* and *phyC* are indeed apoproteins for red/far red light-responsive photoreceptors, rigorous proof of this awaits spectral charac-

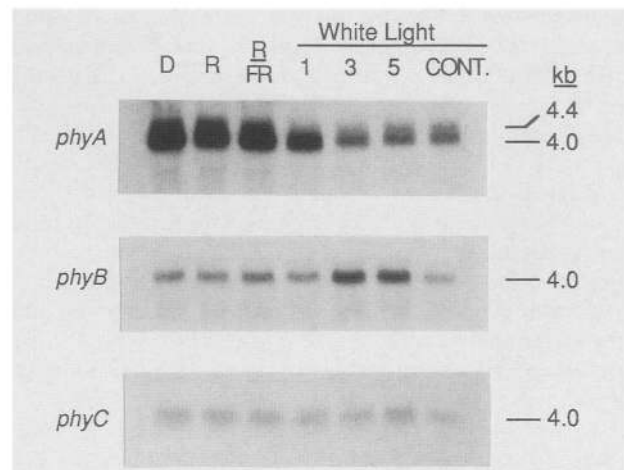


Figure 6. Blot hybridization analysis of the *phy* mRNAs present in total RNA from *A. thaliana* seedlings. RNA was isolated from seedlings that were grown for 5 days completely in the dark (D), given a pulse of red (R) or red followed by far red light (R/FR) 3 hr prior to harvest, transferred to white light for 1, 3, or 5 hr prior to harvest, or grown for 5 days in continuous white light (CONT). Blots were probed with the ssDNA transcript-specific probes for the *phyA*, *phyB*, and *phyC* mRNAs (Fig. 1). Transcript sizes were estimated from mobility relative to RNA molecular weight standards.

terization of the chromophore-containing forms of these proteins. The levels of *phyB* and *phyC* mRNA detected in etiolated *A. thaliana* seedlings indicate that the *phyB* and *phyC* proteins are likely to be less abundant than *phyA* phytochrome in this tissue. Previously, there have been several reports of low-abundance, green-tissue forms of phytochrome in oat and pea that are immunologically and spectrally distinct from the major, etiolated-tissue form (Abe et al. 1985; Shimazaki and Pratt 1985; Tokuhisa et al. 1985). It is possible that *phyB* for *phyC* corresponds to this low abundance phytochrome or, alternatively, that the partially purified chromoprotein fractions from oat and pea are mixtures of the homologs of *phyA*, *phyB*, and *phyC* in these plant species. These questions can now be approached using antisera specific to the A, B, and C forms of phytochrome.

Phytochrome has been detected spectrally in all angiosperm and gymnosperm plants that have been examined, in ferns and bryophytes, and in some species of algae (Correll et al. 1977). Currently, sequence information is available only for phytochrome from a few angiosperm genera. Comparison of these sequences shows that the origins of the *phyA*, *phyB*, and *phyC* genes were very likely gene duplications that occurred early in higher plant evolution, long before the divergence of the two major groups of angiosperms, the monocots and dicots. This expansion and divergence of the phytochrome-coding capacity may have accompanied emergence of novel mechanisms of regulation of photomorphogenesis in early plants. The ancient origins of the phytochrome gene family also indicate the *phyA*, *phyB*,

and *phyC* homologs are likely to be present in all angiosperms and, perhaps, all higher plants. Though published data have been interpreted as indicating the presence of only a single phytochrome gene in pea (Sato 1988) and in rice (Kay et al. 1989b), these studies were performed using hybridization conditions that would not favor detection of distantly related sequences. Previously, we presented evidence for multiple phytochrome-related sequences in DNA blots of tomato (Sharrock et al. 1988). If the A, B, and C forms of phytochrome are conserved elements of higher plant photoregulation systems, the roles of these receptors in the regulation of photomorphogenesis in *A. thaliana* may also be conserved across a wide range of plant genera.

Amino acid identity profiles of the *phyA*, *phyB*, and *phyC* sequences aligned in the three possible pairwise combinations indicate that similar regions of the three polypeptides have been conserved and that no deletion, replacement, or rearrangement of large structural domains has occurred. Notable deviations from this conserved structure are the 35-residue amino-terminal and 11-residue carboxy-terminal extensions of the *phyB* polypeptide. As was observed in comparison of *phyA* sequences from monocot and dicot plant species (Sharrock et al. 1986), the largest blocks of amino acid sequence conserved among *phyA*, *phyB*, and *phyC* (Figs. 2B and 3) are at and around the cysteine residue (position 361 in Fig. 2B), which serves as the chromophore attachment site (Lagarias and Rapoport 1980). Native *phyA* phytochrome is purified from the soluble fraction of plant extracts (Vierstra and Quail 1983) and appears, by immunocytochemistry, to be distributed evenly in the cytoplasm of etiolated-tissue cells (McCurdy and Pratt 1986; Saunders et al. 1983). From consideration of the hydrophobic properties of the proteins (Fig. 4), *Arabidopsis phyB* and *phyC* also appear to be soluble proteins, and although phytochrome modulation of membrane properties and physical association of phytochrome with membranes or plastids have been reported (for review, see Roux 1986), no membrane-spanning or transit peptide sequences are predicted within the *phyB* and *phyC* proteins.

In addition to encoding divergent polypeptides, the *A. thaliana phyA*, *phyB*, and *phyC* genes are expressed in different ways, quantitatively and qualitatively. Transcripts for all three phytochromes are detected in RNA from 5-day-old seedlings and are represented in a cDNA library made from RNA from 3-week-old rosette leaves. The *phyA* mRNA is the most abundant of the three in dark-grown (etiolated) tissue and is down-regulated in the presence of white light. Light-induced down-regulation of *phyA* mRNA abundance has been described in several plant species. In monocots, this effect is rapid and pronounced, is mediated by the phytochrome system itself (Colbert et al. 1985), and is, in large part, due to reduced transcription of *phyA* genes (Lissemore and Quail 1988). In dicots, down-regulation of *phyA* mRNA is less pronounced (Lissemore et al. 1987; Sato 1988) and, in the case of *Arabidopsis*, appears to require the presence of continuous white light. In contrast to

phyA, the *phyB* and *phyC* mRNAs are present at low levels and are not significantly light-regulated. The predominance of *phyA* mRNA in dark-grown tissue indicates that the *phyA* receptor is likely to be the most abundant phytochrome in etiolated seedlings and may play a specific role in the de-etiolation process. In fully green plant tissue, the levels of the mRNAs for the three photoreceptors are within severalfold of each other, making it less likely that one phytochrome is physically or functionally predominant.

Though the *phyA* gene is the only member of the *Arabidopsis* phytochrome family that has been shown to be regulated at the level of mRNA abundance, all three *phy* mRNAs contain URFs in their 5'-leader regions, preceding the initiator AUG for the phytochrome-coding sequence. These URFs contain 3–12 in-frame codons before a translation termination codon is reached. Most eukaryotic mRNAs do not contain URFs (Kozak 1987), and, in some cases, mRNAs that contain URFs show complex translational regulation. For example, Mueller and Hinnebusch (1986) demonstrated that the four 5'-leader URFs in the yeast *GCN4* mRNA have strong regulatory effects on translation of the downstream coding sequences. All dicot phytochrome 5' leaders that have been sequenced, including *phyA* from zucchini (Sharrock et al. 1986), pea (Sato 1988), and *Arabidopsis* and *phyB* and *phyC*, contain at least one URF. Moreover, the multiple *phyA* mRNA 5' ends encoded by the complex *phyA* promoters of pea (Sato 1988) and *Arabidopsis* (R. Sharrock and P. Quail, unpubl.) contain differing numbers of URFs, from one URF in the shortest 5'-untranslated sequence to four URFs in the longest. Monocot phytochrome 5' leaders, including *phyA* from oat (Hershey et al. 1985), rice (Kay et al. 1989a), and corn (Christensen and Quail 1989), do not contain URFs. Hence, it is possible that regulation of dicot phytochrome gene expression, including all members of the gene family, includes a component of translational regulation that is not present in monocots.

The biology of receptor systems in plants is poorly understood as compared to such systems in animals. The mechanisms through which physical and chemical signals are perceived and transmitted in plants may be related to mechanisms that operate in animal receptor systems or they may be unique. In the case of the phytochrome system, absorption of a photon of light by a soluble chromoprotein receptor triggers a diverse and complex array of growth and developmental responses. The discovery of multiple phytochrome genes in *A. thaliana* suggests that, as in many animal receptor systems, diversity of response in plant light perception may reflect heterogeneity of receptor structure and differential patterns of expression of a family of receptor genes. With the identification and characterization of the *phyA*, *phyB*, and *phyC* phytochromes, we are now in a position to develop peptide-specific antisera, determine the tissue localization of the various phytochromes, and begin analysis of the functions of these gene products using classical and molecular genetic approaches in *Arabidopsis*.

Materials and methods

Plant material and growth conditions

All experiments were performed using *A. thaliana* land race Columbia. Seeds were purchased from Guhy's Nursery (Tucson, Arizona). Plants used for preparation of DNA were grown for 3 weeks in soil at 24°C under fluorescent light. For preparation of RNA, ~5000 seeds were sterilized for 60 min in 20% bleach/0.2% SDS and washed six times in sterile water. Sterile seeds were distributed onto a Whatman No. 1 filter paper disk overlaying 0.8% agar/0.5 × Murashige-Skoog salts (Difco) in a 150 × 25-mm petri dish and placed at 4°C in complete darkness. After 7 days, the seeds were brought to room temperature and irradiated for 10 min with white light. The plates were sealed, wrapped in foil, and incubated for 5 days at 24°C in the dark. Under these conditions, seed germination was >90% and etiolated seedlings were ~1 cm tall. Irradiation of 5-day-old seedlings was performed at 24°C for periods of 5 min with red or far red light (Lissemore et al. 1987). Seedlings were harvested 3 hr after irradiation and frozen in liquid N₂. White light irradiation was performed under fluorescent bulbs at 24°C for the indicated times. The outputs of the red, far red, and white light sources were 8 W/m², 11 W/m², and 28 W/m², respectively, as measured by an EG&G Gamma Scientific model 550-1 radiometer/photometer.

Plant DNA preparation and analysis

Total *A. thaliana* DNA was prepared by the method of Ausubel et al. (1989). Samples containing 3 µg of DNA were digested with restriction enzymes, separated on 0.8% agarose gels, and transferred to GeneScreen Plus (Dupont) hybridization membranes according to procedures recommended by the manufacturer. Blots probed with the 0.9-kb coding-region probe (Fig. 1) were prehybridized for 12 hr at 42°C in 30% formamide, 5 × SSC, 5 × Denhardt's solution, 40 mM NaPO₄ (pH 6.8), 0.5% BSA, 1% SDS, and 100 µg/ml sonicated denatured salmon testes DNA and subsequently hybridized to the probe for 18 hr at 42°C in 30% formamide, 5 × SSC, 40 mM NaPO₄ (pH 6.8), 10% dextran sulfate (*M_r* = 500,000), and 100 µg/ml sonicated denatured salmon testes DNA. These membranes were washed under low-stringency conditions: twice for 20 min at room temperature and once for 1 hr at 60°C in 2 × wash solution (2 × SSC, 5 mM EDTA, 1.5 mM sodium pyrophosphate, 0.5% SDS). Blots hybridized with transcript-specific probes (probes A, B, C, Fig. 1) were treated in the same way, except the prehybridization and hybridization buffers contained 50% formamide and the final wash was for 1 hr at 65°C in 0.1 × wash solution (high-stringency conditions). Autoradiography was done at -70°C with an intensifying screen.

The coding-region hybridization probe (0.9-kb probe; Fig. 1) was a ³²P-labeled, ssDNA synthesized from a recombinant M13mp18 (Yanisch-Perron et al. 1985) phage containing a 925-bp *Pst*I fragment of *A. thaliana* phytochrome-coding sequence. Preparation of this probe has been described (Sharrock et al. 1988). Transcript-specific ssDNA probes were prepared using as template sense-strand ssDNA of M13mp18 clones of the most 3' *Eco*RI fragments of the *phyA*, *phyB*, and *phyC* cDNA inserts (Fig. 1). Synthetic oligonucleotides, annealed to these templates just upstream of the poly(A) tails (Fig. 2A), were extended with Klenow fragment in the presence of [³²P]dCTP, truncated with *Pst*I(*phyA*), *Rsa*I(*phyB*), or *Pf*MI(*phyC*), and the labeled single-stranded DNAs were purified from denaturing gels as described (Sharrock et al. 1988).

Library screening

An amplified λgt10 library of size-fractionated (3- to 6-kb fraction) *A. thaliana* cDNAs was supplied by N. Crawford (University of California, San Diego). The cDNA was prepared from poly(A)⁺ RNA isolated from rosette leaves of plants grown 16–20 days under continuous illumination (Crawford et al. 1988). Approximately 250,000 recombinant phage were plated, and the plaques were transferred to nitrocellulose using standard methods (Maniatis et al. 1982). Filters were prehybridized and hybridized with the 0.9-kb coding-region probe in 30% formamide buffers and were washed at low stringency, as described above for DNA blots. Prospective phytochrome-related clones were plaque-purified, and phage DNA was prepared for subcloning (Grossberger 1987).

RNA preparation and analysis

All RNA preparation procedures were carried out on ice, using baked glassware and solutions treated with diethylpyrocarbonate. Phenol/isoamyl alcohol/chloroform 25 : 1 : 25 (PIC) was equilibrated four times with 100 mM Tris-HCl (pH 8.3) and 10 mM EDTA. RNA was prepared from 1–3 grams fresh weight whole plant *A. thaliana* tissue. Frozen tissue was ground to a powder in liquid N₂ in a mortar and pestle, transferred to a 30-ml Corex tube containing 5 ml of extraction buffer [50 mM Tris-HCl (pH 8.3), 150 mM NaCl, 10 mM EDTA, 1% lauryl sarcosine], and mixed with a Polytron at low speed. An equal volume of PIC was then added, and the sample was mixed with the Polytron at high speed for 1 min. After centrifugation, the aqueous phase was extracted three more times with PIC, and the nucleic acids precipitated with Na acetate/ethanol. The precipitates were dissolved in 10 mM Tris-HCl (pH 8.3) and 10 mM EDTA, and RNA was precipitated twice with 2 M LiCl, the first time overnight on ice in a total volume of 4 ml and the second time for 5–6 hr in a volume of 2 ml. The RNA pellet was dissolved in 0.27 ml H₂O and ethanol-precipitated. RNA yields were 80–100 µg of RNA per gram of tissue for etiolated seedlings and 150–300 µg of RNA per gram of tissue for green seedlings.

Samples of total RNA (5 µg) were separated by electrophoresis in 0.8% agarose/3% formaldehyde gels containing 20 mM NaPO₄ (pH 6.8) and transferred to GeneScreen hybridization membranes (Dupont), using procedures recommended by the manufacturer. Membranes were prehybridized and hybridized to the gene-specific ssDNA probes A, B, and C (Fig. 1) in 50% formamide buffers and were washed at high stringency as described above for DNA blots.

DNA sequencing and sequence analysis

The cDNA inserts from clones λA2-3, λA1-1, and λA7-5 were subcloned into M13mp18 and sequenced completely in both directions by the dideoxy method (Sanger et al. 1977), using synthetic oligonucleotides as primers. DNA and polypeptide sequence analysis and alignment were performed using the programs of the UWGCG Software Package (Devereux et al. 1984). Gaps in the aligned polypeptide sequences (Fig. 2B) were introduced according to the BESTFIT program and by eye and are parsimonious in that the number of gaps introduced to optimize sequence similarity have been kept to a minimum. Gaps in the nucleotide sequence (Fig. 2A) were introduced at positions corresponding to the gaps in the peptide sequences. Alignment of the eight available phytochrome polypeptide sequences (Table 1) was done using the BESTFIT and LINEUP

Sharrock and Quail

programs and by eye. Parsimony analysis of 325 phylogenetically informative sites in the phytochrome nucleic acid sequences was performed using the programs of Felsenstein (1985).

Acknowledgments

We thank Dr. Nigel Crawford for his gift of the *A. thaliana* cDNA library, Jo Anne Welsch for excellent technical assistance, Susan DeSimoni and Dr. Christiana Gatz for their contributions to the early phase of this work, and Drs. Sheila McCormick, Alan Christensen, and Tim Caspar for critical reading of the manuscript. This work was supported by Department of Energy grant DE-FG03-87ER13742 and U.S. Department of Agriculture grant 85-CRCR-1-1578 to P.H.Q. R.A.S. was the recipient of postdoctoral fellowship DMB-8508836 from the National Science Foundation.

References

- Abe, H., K.T. Yamamoto, A. Nagatani, and M. Furuya. 1985. Characterization of green tissue-specific phytochrome isolated immunochemically from pea seedlings. *Plant Cell Physiol.* **26**: 1387–1399.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1989. *Current protocols in molecular biology*, pp. 2.3.1–2.3.3. Greene Publishing Associates and Wiley-Interscience, New York.
- Butler, A., A. Wei, K. Baker, and L. Salkoff. 1989. A family of putative potassium channel genes in *Drosophila*. *Science* **243**: 943–947.
- Christensen, A.H. and P.H. Quail. 1989. Structure and expression of a maize phytochrome-encoding gene. *Gene* (in press).
- Colbert, J.T., H.P. Hershey, and P.H. Quail. 1985. Phytochrome regulation of phytochrome mRNA abundance. *Plant Mol. Biol.* **5**: 1–101.
- Correll, D.L., J.L. Edwards, and W. Shropshire, eds. 1977. *Phytochrome: A bibliography with author, biological materials, taxonomic, and subject indexes of publications prior to 1975*. Smithsonian Institution Press, Washington, D.C.
- Crawford, N.M., M. Smith, D. Bellissimo, and R.W. Davis. 1988. Sequence and nitrate regulation of the *Arabidopsis thaliana* mRNA encoding nitrate reductase, a metalloflavo-protein with three functional domains. *Proc. Natl. Acad. Sci.* **85**: 5006–5010.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387–395.
- Doolittle, R.F. 1981. Similar amino acid sequences: Chance or common ancestry? *Science* **214**: 149–159.
- Evans, R.M. 1988. The steroid and thyroid hormone receptor superfamily. *Science* **240**: 889–895.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **38**: 783–791.
- Feng, D.F., M.S. Johnson, and R.F. Doolittle. 1985. Aligning amino acid sequences: Comparison of commonly used methods. *J. Mol. Evol.* **21**: 112–125.
- Giguere, V., N. Yang, P. Segui, and R.M. Evans. 1988. Identification of a new class of steroid hormone receptors. *Nature* **331**: 91–94.
- Grossberger, D. 1987. Minipreps of DNA from bacteriophage lambda. *Nucleic Acids Res.* **15**: 6737.
- Hershey, H.P., R.F. Barker, K.B. Idler, J.L. Lissemore, and P.H. Quail. 1985. Analysis of cloned cDNA and genomic sequences for phytochrome: Amino acid sequences for two gene products expressed in etiolated *Avena*. *Nucleic Acids Res.* **13**: 8543–8559.
- Hillman, W.S. 1967. The physiology of phytochrome. *Annu. Rev. Plant Physiol.* **18**: 301–324.
- Jabben, M. and M.G. Holmes. 1983. Phytochrome in light-grown plants. In *Photomorphogenesis*. Encyclopedia of plant physiology, new series, vol. 16B (ed. W. Shropshire and H. Mohr), pp. 704–722. Springer-Verlag, Berlin.
- Jones, A.M. and P.H. Quail. 1986. Quaternary structure of 124-kilodalton phytochrome from *Avena sativa* L. *Biochemistry* **25**: 2987–2995.
- Kaufman, L.S., W.F. Thompson, and W.R. Briggs. 1984. Different red light requirements for phytochrome-induced acclimation of cab RNA and rbcS RNA. *Science* **226**: 1447–1449.
- Kay, S.A., B. Keith, K. Shinozaki, and N.-H. Chua. 1989a. The sequence of the rice phytochrome gene. *Nucleic Acids Res.* **17**: 2865–2866.
- Kay, S.A., B. Keith, K. Shinozaki, M.-L. Chye, and N.-H. Chua. 1989b. The rice phytochrome gene: Structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5' upstream region. *Plant Cell* **1**: 351–360.
- Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **20**: 8125–8148.
- Kronenberg, G.H.M. and R.E. Kendrick. 1986. The physiology of action. In *Photomorphogenesis in plants* (ed. R.E. Kendrick and G.H.M. Kronenberg), pp. 99–114. Martinus Nijhoff Publishers, Dordrecht.
- Kyte, J. and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105–132.
- Lagarias, J.C. and H. Rapoport. 1980. Chromopeptides from phytochrome. The structure and linkage of the P_r form of the phytochrome chromophore. *J. Am. Chem. Soc.* **102**: 4821–4828.
- Lidgard, S. and P.R. Crane. 1988. Quantitative analysis of the early angiosperm radiation. *Nature* **331**: 344–346.
- Lissemore, J.L. and P.H. Quail. 1988. Rapid transcriptional regulation by phytochrome of the genes for phytochrome and chlorophyll a/b binding protein in *Avena sativa*. *Mol. Cell. Biol.* **8**: 4840–4850.
- Lissemore, J.L., J.T. Colbert, and P.H. Quail. 1987. Cloning of cDNA for phytochrome from etiolated *Cucurbita* and coordinate photoregulation of the abundance of two distinct phytochrome transcripts. *Plant Mol. Biol.* **8**: 485–496.
- Mancinelli, A.L. and I. Rubino. 1978. The 'high irradiance' responses of plant photomorphogenesis. *Bot. Rev.* **44**: 129–150.
- Mandoli, D.F. and W.R. Briggs. 1981. Phytochrome control of two low-irradiance responses in etiolated oat seedlings. *Plant Physiol.* **67**: 733–739.
- Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Martin, W.F., A. Gierl, and H. Saedler. 1989. Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* **339**: 46–48.
- McCurdy, D.W. and L.H. Pratt. 1986. Immunogold electron microscopy of the phytochrome in *Avena*: Identification of intracellular sites responsible for phytochrome sequestering and enhanced pelletability. *J. Cell. Biol.* **103**: 2541–2550.
- Meyerowitz, E.M. 1987. *Arabidopsis thaliana*. *Annu. Rev. Genet.* **21**: 93–111.
- Mueller, P.P. and A.G. Hinnebusch. 1986. Multiple upstream AUG codons mediate translational control of *GCN4*. *Cell* **45**: 201–207.

- Ohno, S., Y. Akita, Y. Konno, S. Imajoh, and K. Suzuki. 1988. A novel phorbol ester receptor/protein kinase, nPKC, distantly related to the protein kinase C family. *Cell* **53**: 731–741.
- Roux, S.J. 1986. Phytochrome and membranes. In *Photomorphogenesis in plants* (ed. R.E. Kendrick and G.H.M. Kronenberg), pp. 115–136. Martinus Nijhoff Publishers, Dordrecht.
- Rudiger, W. and H. Scheer. 1983. Chromophores in photomorphogenesis. In *Photomorphogenesis*. Encyclopedia of plant physiology, new series, vol. 16A (ed. W. Shropshire and H. Mohr), pp. 119–151. Springer-Verlag, Berlin.
- Salisbury, F.B. and C.W. Ross. 1985. *Plant physiology*. Wadsworth Publishing Company, Belmont.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Sato, N. 1988. Nucleotide sequence and expression of the phytochrome gene in *Pisum sativum*: Differential regulation by light of multiple transcripts. *Plant Mol. Biol.* **11**: 697–710.
- Saunders, M.J., M.-M. Cordonnier, B.A. Palevitz, and L.H. Pratt. 1983. Immunofluorescence visualization of phytochrome in *Pisum sativum* L. epicotyls using monoclonal antibodies. *Planta* **159**: 545–553.
- Schaeffer, E., D. Smith, G. Mardon, W. Quinn, and C. Zuker. 1989. Isolation and characterization of two new *Drosophila* protein kinase C genes, including one specifically expressed in photoreceptor cells. *Cell* **57**: 403–412.
- Sharrock, R.A., J.L. Lissemore, and P.H. Quail. 1986. Nucleotide and amino acid sequence of a *Cucurbita* phytochrome cDNA clone: Identification of conserved features by comparison with *Avena* phytochrome. *Gene* **47**: 287–295.
- Sharrock, R.A., B.M. Parks, M. Koornneef, and P.H. Quail. 1988. Molecular analysis of the phytochrome deficiency in an *aurea* mutant of tomato. *Mol. Gen. Genet.* **213**: 9–14.
- Shimazaki, Y. and L.H. Pratt. 1985. Immunochemical detection with rabbit polyclonal and mouse monoclonal antibodies of different pools of phytochrome from etiolated and green *Avena* shoots. *Planta* **164**: 333–344.
- Shropshire W. and H. Mohr, eds. 1983. *Photomorphogenesis*. Encyclopedia of plant physiology, new series, vol. 16A and 16B. Springer-Verlag, Berlin.
- Smith, H. 1986. The perception of light quality. In *Photomorphogenesis in plants* (ed. R.E. Kendrick and G.H.M. Kronenberg), pp. 187–217. Martinus Nijhoff Publishers, Dordrecht.
- Thompson, C.C., C. Weinberger, R. Lebo, and R.M. Evans. 1987. Identification of a novel thyroid hormone receptor expressed in the mammalian central nervous system. *Science* **237**: 1610–1614.
- Tokuhsa, J.G. and P.H. Quail. 1983. Spectral and immunochemical characterization of phytochrome isolated from light-grown *Avena sativa*. (abstr.) *Plant Physiol. Suppl.* **72**: 85.
- Tokuhsa, J.G., S.M. Daniels, and P.H. Quail. 1985. Phytochrome in green tissue: Spectral and immunochemical evidence for two distinct molecular species of phytochrome in light-grown *Avena sativa* L. *Planta* **164**: 321–332.
- Vierstra, R.D. and P.H. Quail. 1983. Purification and initial characterization of 124-kilodalton phytochrome from *Avena*. *Biochemistry* **22**: 2498–2504.
- . 1986. The protein. In *Photomorphogenesis in plants* (ed. R.E. Kendrick and G.H.M. Kronenberg), pp. 35–60. Martinus Nijhoff Publishers, Dordrecht.
- Yanisch-Perron, C., J. Viera, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **22**: 103–119.
- Yarden, Y. and A. Ullrich. 1988. Growth factor receptor tyrosine kinases. *Annu. Rev. Biochem.* **57**: 443–478.



Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family.

R A Sharrock and P H Quail

Genes Dev. 1989, **3**:

Access the most recent version at doi:[10.1101/gad.3.11.1745](https://doi.org/10.1101/gad.3.11.1745)

References

This article cites 43 articles, 11 of which can be accessed free at:
<http://genesdev.cshlp.org/content/3/11/1745.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

The advertisement features a dark blue background with a glowing DNA double helix structure on the left. The text 'horizon' is in white, with 'a PerkinElmer company' in smaller white text below it. To the right, the text 'Streamline your research with Horizon Discovery's ASO tool' is displayed in white, with 'Horizon Discovery's ASO tool' in a larger, bold font.