Correction

PLANT BIOLOGY. For the article “Amino acid polymorphisms in Arabidopsis phytochrome B cause differential responses to light,” by Daniele L. Filiault, Carolyn A. Wessinger, Jose R. Dinneny, Jason Lutes, Justin O. Borevitz, Detlef Weigel, Joanne Chory, and Julin N. Maloof, which appeared in issue 8, February 26, 2008, of Proc Natl Acad Sci USA (105:3157–3162; first published February 14, 2008; 10.1073/pnas.0712174105), the authors note that, in the Fig. 4 legend, the graph bars were identified incorrectly. This error does not affect the conclusions of the article. The figure and its corrected legend appear below.

![Graphs showing phenotypic association with PHYB polymorphisms](https://www.pnas.org/cgi/doi/10.1073/pnas.0804161105)
Amino acid polymorphisms in Arabidopsis phytochrome B cause differential responses to light

Daniele L. Filiault, Carolyn A. Wessinger, Jose R. Dinneny, Jason Lutes, Justin O. Borevitz, Detlef Weigel, Joanne Chory, and J. Noel Maloof

Plants have a sophisticated system for sensing and responding to their light environment. The light responses of populations and species native to different habitats show adaptive variation; understanding the mechanisms underlying photomorphogenic variation is therefore of significant interest. In Arabidopsis thaliana, phytochrome B (PHYB) is the dominant photoreceptor for red light and plays a major role in white light. Because PHYB has been proposed as a candidate gene for several quantitative trait loci (QTLs) affecting light response, we have investigated sequence and functional variation in Arabidopsis PHYB. We examined PHYB sequences in 33 A. thaliana individuals and in the close relative Arabidopsis lyrata. From 14 nonsynonymous polymorphisms, we chose 5 for further study based on previous QTL studies. In a larger collection of A. thaliana accessions, one of these five polymorphisms, 1143L, was associated with variation in red light response. We used transgenic analysis to test this association and confirmed experimentally that natural PHYB polymorphisms cause differential plant responses to light. Furthermore, our results show that allelic variation of PHYB activity is due to amino acid rather than regulatory changes. Together with earlier studies linking variation in light sensitivity to photoreceptor genes, our work suggests that photoreceptors may be a common target of natural selection.

Phytochromes exist in two photoconvertible forms: Pr, an inactive red-light absorbing form, and Pfr, the active far-red light absorbing form. In sunny conditions, which are characterized by high R/FR ratios, most phytochrome is in the active Pfr form. Shade causes a decrease in Pfr and a concomitant induction of shade-avoidance responses. Phytochromes signal through a web of downstream factors, including a family of related bHLH transcription factors, the PIFs and PILs (14), Gigantea (GIG) (15), and the bZIP transcription factors HY5 and HYH that integrate signals from multiple photoreceptor pathways (16). Light-regulated protein degradation, often mediated by the E3 ubiquitin ligase COP1 and associated proteins (17), is also important for phytochrome signaling.

An evolutionary question of considerable interest is how selection acts on the different components of metabolic and developmental pathways. For example, in the case of the anthocyanin biosynthesis pathway, which is responsible for producing flower color pigments, it has been found that genes acting later in the pathway evolve more quickly than those upstream (18, 19). Interestingly, for interpretation of light signals, a number of changes have been found at the top of the pathway in the photoreceptors themselves. For example, reduced FR sensitivity of the A. thaliana strain Lm-2 was traced to a single amino acid change in PHYA that reduced photoconversion and autophosphorylation (20). Cloning of an A. thaliana flowering time quantitative trait loci (QTL) revealed that it was caused by an amino acid substitution that stabilized theCRY2 protein, thereby increasing activation of the photoperiod pathway (21). Variation in PHYC is responsible for differences in both flowering time and hypocotyl elongation among A. thaliana accessions (22), and a naturally occurring deletion in PHYD increases stem elongation (3). More generally, phytochrome genes evolve more quickly than the average plant gene (23), and there is evidence for positive selection early in the diversification of PHYA (24). Although phenotypic variation due to polymorphisms in A. thaliana PHYA (20), PHYC (22), and PHYD (3) has been described, PHYB plays by far the largest role for white-light responses in standard laboratory strains. In addition, PHYB is a candidate for a white and red light QTL in A. thaliana (25), for a flowering time QTL in tomato (26), and a bud set QTL in...
indicating each amino acid polymorphism. The table below indicates the
comprising of four Spanish accessions (Fig. 1), five Russian and Central Asian accessions; and the third clade
separation in PHYB and PHY domains (32; P = 0.4554, Fisher’s exact test, comparing
variation is limited to the GAF domain excluding the light-sensing
patterns of conservation were seen with both the DT and HMMI
region in the GAF domain. Using eudicot sequences, similar
conserved regions in four PHYB protein sequences from the
Markov model islands (HMMI) methods were used to identify
eShadow (35), both divergence threshold (DT) and hidden
Markov model islands were used to identify
evidence of nonneutral selection. Although there is an excess of
polymorphisms that are fixed between species compared with ones that segregate within species, this excess is
not significant [supporting information (SI) Table 1; P > 0.26],
so the underlying cause of the observed substitution pattern remains unresolved.

To examine whether the differential distribution of polymorphic
sites has persisted over a greater evolutionary distance, a
phylogenetic shadowing (34) approach was taken. Using
eShadow (35), both divergence threshold (DT) and hidden
Markov model islands (HMMI) methods were used to identify conserved regions in four PHYB protein sequences from the
Brassicales and 16 further eudicot sequences available in GenBank. Within the Brassicales, the HMMI algorithm predicted conservation in regions spanning most of the P2/PAS, GAF, and
PHY domains (Fig. 2), with one interesting exception, an ~50-aa region in the center of the GAF domain. This less conserved
region corresponds to the portion of the GAF domain in bacterial phytochrome that contributes to the light-sensing knot structure (36). Compared with the HMMI method, the DT
approach identified a slightly smaller conserved region from the
C-terminal third of the P2/PAS domain to approximately the first
50 aa of the PHY domain, again excluding the light-sensing knot region in the GAF domain. Using eudicot sequences, similar patterns of conservation were seen with both the DT and HMMI
models (SI Fig. 6). In these more divergent sequences, conservation is limited to the GAF domain excluding the light-sensing knot and the N-terminal end of the PHY domain. In summary, the patterns over longer time scales are predictive for the
variation seen within A. thaliana, suggesting that the PHY and
GAF domains are the most slowly-evolving domains during
18 accessions grouped into a large, poorly resolved clade. In
general, the pattern seen is consistent with previously reported
genomic patterns of isolation by distance (30, 31). A similar tree
was obtained by using neighbor-joining methods.
Because the 14 nonsynonymous polymorphisms all fell outside the functionally important GAF and PHY domains (Fig. 1 A), we asked whether this pattern could be explained by selection. We modeled sequence evolution, using maximum likelihood methods to assess whether substitution rates differed between the PHY and GAF domains, which are important for chromophore binding and photoconversion (32), and the rest of the protein. For codon position 3, where polymorphisms are usually silent, there was no evidence for different rates (P = 0.95). However, the substitution rate for codon positions 1 and 2, where polymorphisms frequently cause amino acid changes, was significantly lower in the GAF and PHY domains (P = 0.009) compared with the rest of the protein. We only found evidence for depressed substitution rates at positions 1 and 2, suggesting that the reduced sequence diversity in the PHY and GAF domains is due to selection on protein sequence and function, rather than local variation in mutation rate.

The observed distribution of nonsynonymous sites might reflect that the GAF and PHY domains are more highly constrained than the rest of the protein. However, we found no evidence that mutations induced in the laboratory and known to compromise PHYB function unevenly affected the PHY and GAF domains: 12 of 29 missense mutations are in the GAF or PHY domains (32; P = 0.4554, Fisher’s exact test, comparing mutation rate in the GAF and PHY domains with the rest of the protein). Thus, with respect to overall activity, the GAF and PHY domains are unlikely to be more highly constrained than the rest of the protein. An alternative explanation is that the observed pattern is a result of adaptive evolution, indicating that domains other than the PHY and GAF are better targets for selection. We used the McDonald–Kreitman test (33) to look for evidence of nonneutral selection. Although there is an excess of nonsynonymous polymorphisms that are fixed between species compared with ones that segregate within species, this excess is not significant [supporting information (SI) Table 1; P > 0.26], so the underlying cause of the observed substitution pattern remains unresolved.

To examine whether the differential distribution of polymorphic sites has persisted over a greater evolutionary distance, a phylogenetic shadowing (34) approach was taken. Using eShadow (35), both divergence threshold (DT) and hidden Markov model islands (HMMI) methods were used to identify conserved regions in four PHYB protein sequences from the Brassicales and 16 further eudicot sequences available in GenBank. Within the Brassicales, the HMMI algorithm predicted conservation in regions spanning most of the P2/PAS, GAF, and PHY domains (Fig. 2), with one interesting exception, an ~50-aa region in the center of the GAF domain. This less conserved region corresponds to the portion of the GAF domain in bacterial phytochrome that contributes to the light-sensing knot structure (36). Compared with the HMMI method, the DT approach identified a slightly smaller conserved region from the C-terminal third of the P2/PAS domain to approximately the first 50 aa of the PHY domain, again excluding the light-sensing knot region in the GAF domain. Using eudicot sequences, similar patterns of conservation were seen with both the DT and HMMI models (SI Fig. 6). In these more divergent sequences, conservation is limited to the GAF domain excluding the light-sensing knot and the N-terminal end of the PHY domain. In summary, the patterns over longer time scales are predictive for the variation seen within A. thaliana, suggesting that the PHY and GAF domains are the most slowly-evolving domains during evolution. That the GAF contribution to the phytochrome knot

Fig. 1. Sequence diversity in A. thaliana PHYB. (A) Amino acid changes in A. thaliana PHYB. A schematic diagram of PHYB domains is shown, with arrows indicating each amino acid polymorphism. The table below indicates the specific amino acid change. The amino acid to the left of the number indicates the residue in the reference sequence from the Columbia (Col-0) accession. (B) Maximum likelihood phylogenetic tree, based on nucleotide sequence, showing PHYB divergence in A. thaliana. Bold italic numbers indicate where amino acid variants separate different branches. Plain text numbers in smaller font indicate bootstrap support.

poplar (27, 28). Here, we examine the effect of natural variation in PHYB on differential light response in A. thaliana and find that polymorphisms in PHYB proteins contribute to variation in A. thaliana light sensitivity.

Results and Discussion

To determine whether polymorphisms in PHYB could contribute to variation in light response, we began by examining PHYB sequence diversity in Arabidopsis. We sequenced the PHYB gene from Ler and Cvi, two accessions known to segregate for LIGHT2, a QTL coincident with PHYB (25); 18 additional A. thaliana accessions; and the sister species A. lyrata. In our sequence analysis, we also included 14 A. thaliana accessions for which PHYB sequences were available in GenBank (29). In A. thaliana, we found 65 synonymous and 14 nonsynonymous polymorphisms. An additional 15 polymorphisms were identified in introns. Of the 14 nonsynonymous sites, 7 differentiated Ler and Cvi (Fig. 1 A and B). A maximum likelihood phylogenetic tree revealed three well supported clades: one clade of six accessions contained the standard laboratory strains Col and Ler, which originate from Central Europe; another one included five Russian and Central Asian accessions; and the third clade comprised of four Spanish accessions (Fig. 1 B). The remaining
is precisely the condition where variation in

associated with differences in hypocotyl elongation in red light. Polymorphisms 1, 3, and 12 were specifically found to be haplotype blocks are typically gene-sized (38). Consistent with whole-genome survey of LD in nonsingleton polymorphisms that distinguish L

Cri. A panel of 140 accessions that had been phenotyped and Cvi are associated with natural variation in light response. A panel of 140 accessions that had been phenotyped for seedling light responses (20) was genotyped for the five nonsingleton polymorphisms that distinguish Ler and Cvi. A whole-genome survey of LD in A. thaliana has shown that haplotype blocks are typically gene-sized (38). Consistent with this finding, there is strong LD across the PHYB gene (Fig. 3). Polymorphisms 1, 3, and 12 were specifically found to be associated with differences in hypocotyl elongation in red light (P < 0.01, P < 0.005, and P < 0.05, respectively; Fig. 4). Red light is precisely the condition where variation in PHYB would be most easily detected, suggesting that these associations are meaningful.

Because of the potentially confounding effects of population structure (39), the PHYB association can only be suggestive of being causal for differential light response across accessions. Therefore, we decided to compare the function of PHYB-Ler and PHYB-Cvi experimentally. We first asked whether differences in PHYB activity (if any) would be more likely due to coding or regulatory changes. Three findings suggested that regulatory differences might not be important. First, protein blots did not show any obvious differences in PHYB levels among accessions (data not shown). Second, an analysis of published microarray data (40) revealed that PHYB mRNA levels are similar across accessions, with no significant difference between Ler and Cvi (P > 0.8), and that there was no correlation between PHYB mRNA levels and hypocotyl elongation (Fig. 5A and B). Third, an eQTL study that used microarrays to examine expression differences in Ler, Cvi, and a derived mapping population found no evidence for differential PHYB expression (41). Therefore, we concluded that if PHYB was a source of variation in light sensitivity, then amino acid changes, rather than regulatory polymorphisms, between PHYB-Ler and PHYB-Cvi were most likely the cause.

To examine possible differences in protein activity, we linked PHYB cDNAs from both Ler and Cvi to the constitutive Cauliflower Mosaic Virus 35S promoter (35S), uncoupling PHYB function from any possible promoter differences. We transformed the null mutant phyB-9, which has greatly reduced light sensitivity and tall hypocotyls. Based on the allelic effects of the LIGHT2 QTL (25) and our association mapping results, 35S::PHYB-Ler was expected to produce a more active PHYB than 35S::PHYB-Cvi. Indeed, hypocotyls of phyb-9 35S::PHYB-Ler plants were significantly shorter than those of phyb-9 35S::PHYB-Cvi (P < 0.008 for T1 generation, and P < 0.0008 for T2 generation plants; Fig. 5C), indicating that PHYB-Cvi confers less light responsiveness than PHYB-Ler in a phyb-9 (Col) background. It is possible that this difference is due to genetic interactions between the Col background and the different PHYB alleles. However, because PHYB-Cvi is associated with reduced light response across accessions and no loci were found to be epistatic with LIGHT2 in Ler X Cvi QTL analysis, it is more likely that the reduced response of PHYB-Cvi is inherent to the protein. In summary, PHYB polymorphisms are likely responsible for the significant associations across accessions and at least partially responsible for the LIGHT2 QTL.

We used regression analysis to determine which of the three polymorphisms identified as significant in association mapping are most likely to be responsible for differences in PHYB activity. We regressed the polymorphisms both individually and in combination against the light responses. This process identified polymorphism 3 as the strongest candidate for the causative change. Specifically, models that included polymorphism 3 in

![Fig. 2. Phylogenetic shadowing. Shown is the percentage variation by amino acid position for four Brassicaceae PHYB sequences. Regions determined to be slow-evolving are indicated by orange bars (HMM) or green bars (DT). Functional domains are marked with red bars.](image)

![Fig. 3. PHYB linkage disequilibrium. Pairwise linkage disequilibrium (r²) between single nucleotide polymorphisms in the A. thaliana PHYB gene. Polymorphic sites are arranged along the diagonal from lower left (5') to upper right (3'). Numbered polymorphisms highlighted by darker lines correspond to the numbered amino acid polymorphisms in Fig. 1. To read the LD between any two sites, trace a column upwards from the 5' site and a row leftward from the 3' site; the square in which the column and row intersect indicates the LD.](image)
combination with either 1 or 12 fit the data significantly better than models with polymorphism 1 or 12 alone. However, the combination of polymorphisms 3 plus 1 or 3 plus 12 did not provide any improvement over polymorphism 3 as the only explanatory variable (SI Table 2). Polymorphism 3 falls within a moderately conserved region of Brassicaceae PHYB. Comparison of the *A. thaliana* PHYB and PHYA sequences with the bacterial phytochrome DrBphP sequence indicated that this polymorphism lies within the P2/PAS domain of the photosensory core (36). By superimposing secondary structural elements of bacterial phytochrome on *Arabidopsis* PHYB sequence, we found that polymorphism 3 was predicted to be on the surface in a span of three α-helices, away from the chromophore binding pocket. This surface is speculated to be involved in protein–protein interactions (36), suggesting a possible mode-of-action for polymorphism 3.

Determining the genes responsible for natural variation and identifying the underlying QTL remains an important challenge (42). We have shown that amino acid polymorphisms in the major photoreceptor for red light contribute to variation in photomorphogenesis across *A. thaliana* accessions. There is a long-standing debate whether developmental variation is primarily due to changes in gene expression or in protein activity. Our findings with PHYB, together with those for PHYA, PHYC, and CRY2 (20–22), weigh in on the side of protein changes. Second, it has been proposed that in some pathways variants are more likely to be found downstream, rather than upstream, because of relatively relaxed constraints on downstream genes (18, 19) or selective sweeps acting on upstream genes (43). Here, we report that there is functional variation at the top of the PHYB pathway, complementing previous studies of other light-response pathways that also pointed to the photoreceptors themselves being responsible for variation in light sensing (20–22,44). Similarly, another study has shown evidence for selection upstream in a floral development pathway (43).

Changes in PHYA and CRY2 are limited to individual accessions (20, 21), making it difficult to know whether they confer adaptive advantages or are simply only mildly deleterious mutations that have not yet been purged from nature. In contrast, alternative variants that cause differential light sensitivity are more common in the case of PHYC (22) and PHYB (this work), suggesting more strongly that they may be important in adaptation. If any of these changes are indeed adaptive, it could provide any improvement over polymorphism 3 as the only explanatory variable (SI Table 2). Polymorphism 3 falls within a moderately conserved region of Brassicaceae PHYB. Comparison of the *A. thaliana* PHYB and PHYA sequences with the bacterial phytochrome DrBphP sequence indicated that this polymorphism lies within the P2/PAS domain of the photosensory core (36). By superimposing secondary structural elements of bacterial phytochrome on *Arabidopsis* PHYB sequence, we found that polymorphism 3 was predicted to be on the surface in a span of three α-helices, away from the chromophore binding pocket. This surface is speculated to be involved in protein–protein interactions (36), suggesting a possible mode-of-action for polymorphism 3.

Determining the genes responsible for natural variation and identifying the underlying QTL remains an important challenge (42). We have shown that amino acid polymorphisms in the major photoreceptor for red light contribute to variation in photomorphogenesis across *A. thaliana* accessions. There is a long-standing debate whether developmental variation is primarily due to changes in gene expression or in protein activity. Our findings with PHYB, together with those for PHYA, PHYC, and CRY2 (20–22), weigh in on the side of protein changes. Second, it has been proposed that in some pathways variants are more likely to be found downstream, rather than upstream, because of relatively relaxed constraints on downstream genes (18, 19) or selective sweeps acting on upstream genes (43). Here, we report that there is functional variation at the top of the PHYB pathway, complementing previous studies of other light-response pathways that also pointed to the photoreceptors themselves being responsible for variation in light sensing (20–22,44). Similarly, another study has shown evidence for selection upstream in a floral development pathway (43).

Changes in PHYA and CRY2 are limited to individual accessions (20, 21), making it difficult to know whether they confer adaptive advantages or are simply only mildly deleterious mutations that have not yet been purged from nature. In contrast, alternative variants that cause differential light sensitivity are more common in the case of PHYC (22) and PHYB (this work), suggesting more strongly that they may be important in adaptation. If any of these changes are indeed adaptive, it could
indicate that overall light sensitivity rather than a particular downstream process is more prone to selection. It will therefore be interesting to study the molecular evolution of the different light-sensing pathways as comprehensive whole-genome variation data become available.

Materials and Methods

DNA Sequencing and Assembly. Detailed sequencing methods are described in SI Text. All polymorphisms were confirmed in multiple sequencing reads. PHYB sequence for Arabidopsis lyrata subspecies lyrata was assembled from whole genome shotgun reads, using the Staden package (45). Sequences have been deposited in GenBank (29) as accession nos. EU352775–EU352793.

Phylogenetic Analysis. PHYB sequences were obtained either from the above sequencing or from GenBank (SI Table 3) and were aligned with ClustalX (46, 47). PHYLIPI (48) was used to bootstrap the dataset 100 times, determine maximum likelihood and neighbor joining trees, and find majority rule consensus trees. PAUP* software, Version 4.0b10 (49), was used to determine whether there were different rates of nucleotide substitution in the PI/PAS, GAF, and PHY domains relative to the rest of the protein within A. thaliana by comparing different substitution models, as described in SI Text.

Phylogenetic Shadowing. Sequences obtained from GenBank (SI Table 3) were used as input for eShadow (http://eshadow.cdfcc.org). HHMM analysis of Brassicales used the following probabilities: eS = 0.85, eF = 0.80, and T = 0.2. Brassicales DT analysis used a maximum percent variation of 5% and a minimum length of 80 aa. For eudicot sequences, HHMM analysis used probabilities of eS = 0.75, eF = 0.60, and T = 0.1, whereas DT analysis used a maximum percent variation of 20% and a minimal length of 80 aa.

Association Mapping. Simple sequence length polymorphism, cleavable amplified polymorphic sequence (CAPS) and derived CAPS assays were designed and used to genotype > 100 A. thaliana accessions for the five nonsingleton nonsynonymous polymorphisms between Ler and Cvi (SI Tables 4 and 5). A permutation-based approach was used to determine association with pheno-
typical differences. First, a t statistic for correlation between genotype and phenotype was calculated for each polymorphism. To establish the significance of these associations, 10,000 permuted datasets were analyzed, and, for each permutation, the highest t statistic across all polymorphisms was recorded. An association was deemed significant if its t statistic was larger than the appropriate quantile t statistic from the permuted dataset.

LD Analysis. Ninety-two SNPs were identifed from a MAFFT multiple sequence alignment of 33 genomic PHYB sequences (SI Tables 3 and 6). One was removed from the analysis because of incomplete data. The LDheatmap package (50) in the R statistical environment (51) was used to determine and plot pairwise linkage disequilibrium, using allelic correlation (r²).

Transgenic Analysis. The null phyB-9 allele was transformed with PHYB-Cvi and PHYB-Ler as described in SI Text. Two independent transformations were performed with a total of 483 PHYB-Cvi and 662 PHYB-Ler T1 transplants assayed in nine independent experiments. Significance was determined by using a linear mixed-effects model with PHYB construct as a fixed effect and transformation plus experiment (within transformation) as random effects. To confirm these results, kanamycin resistant T2 progeny from 40 PHYB-Cvi and 44 PHYB-Ler T1 plants were analyzed in two independent experiments. Experiment was not a significant factor, so the mixed-effects model contained PHYB construct (fixed effect), plate (random effect), and T1 parent (random effect).

Microarray Analysis. Microarray data (40) were imported into R (51) and Bioconductor (52) and normalized by RMA (53). Because we were only genotyping a single gene, a simple t test was used to determine whether there was evidence for differential PHYB expression between Ler and Cvi (n = 2 and 3 replicates, respectively).

ACKNOWLEDGMENTS. We thank Andread Wallace and Gabriel Trainer for technical assistance. This work was funded by National Science Foundation Grant DBI-0227103 (to J.N.M. and Cynthia Weinig), National Institutes of Health Grant GM62932 (to J.C. and D.W.), The Howard Hughes Medical Institute, and National Institutes of Health Training Grant T32 GM070377 (to D.L.F.).


