Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns

Fay-Wei Li[a,1], Juan Carlos Villarreal[b], Steven Kelly[c], Carl J. Rothfels[d], Michael Melkonian[e], Eftychios Frangedakis[c], Markus Ruhsam[f], Erin M. Sigel[a], Joshua P. Der[g,h], Jarmila Pittermann[i], Dylan O. Burge[j], Lisa Pokorny[k], Anders Larsson[l], Tao Chen[m], Stina Weststrand[l], Philip Thomas[f], Eric Carpenter[n], Yong Zhang[o], Zhijian Tian[o], Li Chen[o], Zhixiang Yan[o], Ying Zhu[o], Xiao Sun[o], Jun Wang[o], Dennis W. Stevenson[p], Barbara J. Crandall-Stotler[q], A. Jonathan Shaw[a], Michael K. Deyholos[n], Douglas E. Soltis[r,s,t], Sean W. Graham[u], Michael D. Windham[a], Jane A. Langdale[c], Gane Ka-Shu Wong[n,o,v,1], Sarah Mathews[w], and Kathleen M. Pryer[a]

[a]Department of Biology, Duke University, Durham, NC 27708; [b]Systematic Botany and Mycology, Department of Biology, University of Munich, 80638 Munich, Germany; [c]Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom; [d]Department of Zoology, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; [e]Botany Department, Cologne Biocenter, University of Cologne, 50674 Cologne, Germany; [f]Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, Scotland; [g]Department of Biology and [h]Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802; [i]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064; [j]California Academy of Sciences, San Francisco, CA 94118; [k]Real Jardín Botánico, 28014 Madrid, Spain; [l]Systematic Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden; [m]Fairy Lake Botanical Garden, Shenzhen, Guangdong 518004, China; [n]Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada T6G 2E9; [o]BGI-Shenzhen, Shenzhen 518083, China; [p]New York Botanical Garden, Bronx, NY 10458; [q]Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901; [r]Florida Museum of Natural History, [s]Department of Biology, and [t]Genetics Institute, University of Florida, Gainesville, FL 32611; [u]Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; [v]Department of Medicine, University of Alberta, Edmonton, AB, Canada T6G 2E1; and [w]Harvard University Herbaria, Cambridge, MA 02138

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved March 21, 2014 (received for review October 29, 2013)

Ferns are well known for their shade-dwelling habits. Their ability to thrive under low-light conditions has been linked to the evolution of a novel chimeric photoreceptor—neochrome—that fuses red-sensing phytochrome and blue-sensing phototropin modules into a single gene, thereby optimizing phototropic responses. Despite being implicated in facilitating the diversification of modern ferns, the origin of neochrome has remained a mystery. We present evidence for neochrome in hornworts (a bryophyte lineage) and demonstrate that ferns acquired neochrome from hornworts via horizontal gene transfer (HGT). Fern neochromes are nested within hornwort neochromes in our large-scale phylogenetic reconstructions of phototropin and phytochrome gene families. Divergence date estimates further support the HGT hypothesis, with fern and hornwort neochromes diverging 179 Mya, long after the split between the two plant lineages (at least 400 Mya). By analyzing the draft genome of the hornwort *Anthoceros punctatus*, we also discovered a previously unidentified phototropin gene that likely represents the ancestral lineage of the neochrome phototropin module. Thus, a neochrome originating in hornworts was transferred horizontally to ferns, where it may have played a significant role in the diversification of modern ferns.

phototropism | chloroplast movement

Plant growth and development are modulated by photoreceptor systems that provide information about the surrounding environment. Major peaks in the action spectra of these informational photoreceptors lie either in the UV-blue (e.g., cryptochromes and phototropins) or red/far-red (phytochromes) light regions (1). The chimeric photoreceptor neochrome is a remarkable exception. It fuses red-sensing phytochrome and blue-sensing phototropin modules into a single molecule (Fig. 1*A*) that mediates phototropic responses (1–4). Neochromes have a restricted occurrence in the plant tree of life and are hypothesized to have had two independent origins (5)—one in the green alga *Mougeotia scalaris* and another in ferns. The possession of neochrome may be evolutionarily advantageous, as evidenced by the greatly enhanced phototropic responses in ferns with neochrome (3, 4) and by its phylogenetic distribution within the fern lineage. The early-diverging fern orders Osmundales and Schizaeales do not possess neochrome (3). It has been reported only in Cyatheales (6) and Polypodiales (3, 6), lineages that mostly diversified following the Cretaceous/Tertiary establishment of low-light, angiosperm-dominated forest canopies (7, 8). As a result, it has been suggested that the evolution of neochrome was a key innovation that conferred a phototropic advantage on ferns growing under low-light conditions, facilitating their modern diversification in the "shadow of angiosperms" (3, 7, 8). Although potentially significant from an evolutionary standpoint, the origin of fern neochrome has remained a mystery.

In this study we investigated the origin of neochrome by searching for homologous sequences in 434 transcriptomes, and

Significance

Despite being one of the oldest groups of land plants, the majority of living ferns resulted from a relatively recent diversification following the rise of angiosperms. To exploit fully the new habitats created by angiosperm-dominated ecosystems, ferns had to evolve novel adaptive strategies to cope with the low-light conditions exerted by the angiosperm canopy. Neochrome, an unconventional photoreceptor that allows ferns to "see the light" better, was likely part of the solution. Surprisingly, we discovered that fern neochrome was derived from a bryophyte lineage via horizontal gene transfer (HGT). This finding not only provides the first evidence that a plant-to-plant HGT can have a profound evolutionary impact but also has implications for the evolution of photosensory systems in plants.

Author contributions: F.-W.L. designed research; F.-W.L., S.M., and K.M.P coordinated study; F.-W.L. performed research; F.-W.L., J.C.V., S.K., C.J.R., M.M., E.F., M.R., E.M.S., J.P.D., J.P., D.O.B., L.P., A.L., T.C., S.W., P.T., E.C., Y. Zhang, Z.T., L.C., Z.Y., Y. Zhu, X.S., J.W., D.W.S., B.J.C.-S., A.J.S., M.K.D., D.E.S., S.W.G., J.A.L., and G.K.-S.W. contributed new reagents/analytic tools; F.-W.L. and S.K. analyzed data; and F.-W.L., M.D.W., S.M., and K.M.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The DNA sequences reported here are deposited in the GenBank database (accession nos. KJ128382, KJ128383, KJ128384, and KJ194997–KJ195254). The 1KP transcriptomes can be accessed at www.bioinfodata.org/Blast4OneKP. The sequencing reads for *Anthoceros punctatus* draft genome and *Pteridium aquilinum* transcriptome are deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under SRA096687 and SRX423244, respectively. The assembled *P. aquilinum* transcriptome is deposited in NCBI Transcriptome Shotgun Assembly under GASP00000000. Alignments and tree files are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.fn2rg.

[1]To whom correspondence may be addressed. E-mail: fl43@duke.edu or gane@ualberta.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319929111/-/DCSupplemental.

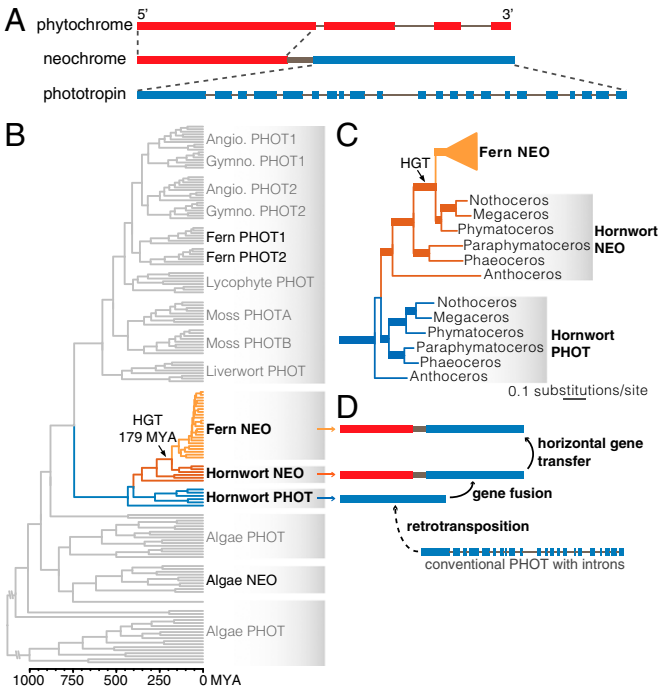

Fig. 1. The origin of fern neochrome. (*A*) Neochrome is a chimeric photoreceptor in which the N terminus consists of a phytochrome sensory module fused to an almost complete phototropin sequence at the C terminus. Thick and thin lines represent exons and introns, respectively (length not to scale). (*B*) Dated phylogeny of phototropin and neochrome, showing neochrome HGT from hornworts to ferns (details are given in *SI Appendix*, Fig. S5). The blue, brown, and yellow branches represent hornwort phototropin, hornwort neochrome, and fern neochrome, respectively. (*C*) Portion of the phototropin phylogeny showing relationships of fern neochrome (Fern NEO), hornwort phototropin (Hornwort PHOT), and hornwort neochrome (Hornwort NEO), with highly supported branches thickened (details are shown in *SI Appendix*, Figs. S1 and S2). (*D*) A schematic depicting the origin of fern neochrome involving retrotransposition of a phototropin gene (and hence the loss of introns), its fusion with a phytochrome, and HGT from hornworts to ferns.

40 whole or draft genomes of plants and algae (*SI Appendix*, Table S1) and surprisingly discovered neochrome homologs from hornworts (Fig. 1*B* and *SI Appendix*, Figs. S1–S3). Analyses of the hornwort draft genome (*Anthoceros punctatus*) suggest that neochrome originated in hornworts, independent from the green algae. Large-scale phylogenetic analyses and divergence time estimates further demonstrate that ferns acquired neochrome from hornworts via horizontal gene transfer (HGT).

Results and Discussion

Algal Neochrome. The only published algal neochrome is from a single species, *Mougeotia scalaris* (5). We identified homologs of neochrome in the transcriptomes of all 10 sampled members of the Zygnemataceae (9), including *Mougeotia*, *Mesotaenium*, *Cylindrocystis*, and *Zygnemopsis*, but in no other algal transcriptomes surveyed (*SI Appendix*, Table S1 and Fig. S1).

Neochrome in Hornworts. Among land plants, we documented the occurrence of neochrome in 25 additional fern species (*SI Appendix*, Figs. S1–S3). Surprisingly, we also discovered neochrome in hornworts, a small bryophyte lineage that diverged early in the history of land plants. The exact branching order among the three bryophyte lineages (hornworts, mosses, and liverworts) is not resolved with certainty; previous analyses have suggested that hornworts are sister to vascular plants (lycophytes, ferns, and seed plants) (10), but this relationship was challenged recently, and the monophyly of all bryophytes was proposed (11). We confirmed the presence of neochrome in hornworts through PCR and cloning, and isolated neochrome sequences from the genera *Nothoceros*, *Megaceros*, *Phymatoceros*, *Phaeoceros*, *Paraphymatoceros*, and *Anthoceros*, representing four of the five hornwort orders (namely, Dendrocerotales, Phymatocerotales, Notothyladales, and Anthocerotales). We were unable to obtain adequate material of the monotypic hornwort *Leiosporoceros* to test for the presence of neochrome in Leiosporocerotales. To confirm that our hornwort neochrome sequence data were indeed derived from the hornwort nuclear genome and not from contaminant algae or ferns, we performed genome-walking in *Nothoceros aenigmaticus* to obtain flanking genomic sequences. Downstream of neochrome we found a pseudogene for imidazoleglycerol-phosphate dehydratase (IGPD) and, because its sequence is most closely related to other hornwort IGPD genes (*SI Appendix*, Fig. S4), we are confident that neochrome is present in the hornwort genome.

Neochrome HGT from Hornworts to Ferns. The phylogenetic distribution of neochrome in land plants (present only in hornworts and ferns) could be explained by (*i*) an ancient origin along the branch that unites hornworts and tracheophytes, followed by losses from lycophytes and seed plants, (*ii*) independent origins in ferns and hornworts, or (*iii*) one or more instances of HGT between hornworts and ferns. To distinguish among these three possible scenarios, we compiled comprehensive sequence alignments of phototropin and phytochrome from across all land plants and algae, which included the corresponding domains from hornwort and fern neochromes, and evaluated the resultant gene phylogenies. Maximum likelihood and Bayesian estimates of phototropin and phytochrome phylogenies revealed that fern neochromes are embedded within hornwort neochromes with very strong branch support (Fig. 1 *B* and *C* and *SI Appendix*, Figs. S1–S3). This nested relationship indicates that neochrome was transferred horizontally from hornworts to ferns, along the stem lineage leading to *Phymatoceros* + *Nothoceros* + *Megaceros* (Fig. 1*C*, arrow, and *SI Appendix*, Figs. S1–S3). The alternative possibilities, suggesting either an ancient vertical transfer of neochrome (i.e., fern and hornwort neochromes were reciprocally monophyletic) or an independent origin of neochrome (i.e., fern neochromes were monophyletic with either fern phototropins or phytochromes) were both rejected ($P < 10^{-30}$) and were never observed in the Bayesian posterior tree samples.

We used estimates of divergence time to assess our HGT hypothesis further, reasoning that in a case of HGT the split between hornwort and fern neochrome should be younger than the split between the hornwort and fern lineages themselves. By integrating fossil calibrations (*SI Appendix*, Table S2) with a Bayesian relaxed molecular clock analysis, we estimated the divergence date between hornwort and fern neochrome to be ~179 Mya with a 95% highest posterior density interval of 133 and 229 Mya (Fig. 1*B* and *SI Appendix*, Fig. S5). This date is far more recent than published divergence estimates between ferns and hornworts (at least 400 Mya) (12) but is congruent with the date estimates for the stem branch leading to *Phymatoceros* + *Nothoceros* + *Megaceros* (85–244 Mya) (13). The disparity in divergence times rejects the hypothesis invoking multiple neochrome origins or losses and reinforces the HGT scenario.

The origin of land plant neochrome within the hornwort lineage is supported by its relationship to hornwort phototropin. The single hornwort phototropin gene in the *Anthoceros punctatus* draft genome completely lacks introns (Fig. 1*D*) and thus closely resembles the C-terminal end of both fern and hornwort neochromes. We found this intron-free phototropin in all hornworts examined by using PCR on genomic DNA (*SI Appendix*, Fig. S2 and Table S3). All other phototropins characterized to date, including those of ferns, contain more than 20 introns. We

explored whether this gene might be a partial neochrome masquerading as a phototropin by using inverse PCR to obtain the 5′ upstream genomic region in *N. aenigmaticus*. Multiple stop codons were encountered upstream of the *Nothoceros* phototropin gene, and there was no indication of nearby phytochrome domains. These data suggest that hornworts might not have a canonical phototropin gene. Instead, hornwort phototropins are most closely related to fern and hornwort neochromes (Fig. 1 *B* and *C* and *SI Appendix*, Fig. S2), implying that they likely represent the ancestral, retrotransposed phototropin lineage that gave rise to neochrome through fusion with the phytochrome module (Fig. 1*D*). On the other hand, in the phytochrome phylogeny, hornwort phytochromes are not resolved as sister to hornwort and fern neochromes (*SI Appendix*, Fig. S3), although there is no branch support for this nonmonophyletic relationship. The phytochrome progenitor of neochrome therefore remains unclear.

Recurrent Fern-to-Fern HGT. We detected an extraordinary incongruence between our fern neochrome gene tree and the published phylogeny of ferns (*SI Appendix*, Fig. S6) (14). By examining the entire Bayesian posterior tree sample, we found that none of the trees resolved neochromes from the same fern family as being monophyletic. This conflicting pattern is not observed in other fern phylogenies based on nuclear genes (15) and is not seen in the hornwort neochrome tree (Fig. 1*C*), which perfectly mirrors the published phylogeny of hornworts (13). Here we investigate and discuss the possible causes of the incongruent gene tree/species tree in ferns.

Incomplete sampling of extant neochrome homologs is not likely to be the explanation, because neochrome has been shown by Southern blotting to be a single-copy gene in *Adiantum capillus-veneris* (2). This result was corroborated by the cloning efforts that produced most of our sequence data (*SI Appendix*, Table S3). Except for *Deparia* spp., in which two divergent sequences were found (*SI Appendix*, Fig. S6, arrowheads), we were able to isolate only a single neochrome from each fern species.

Next, we investigated whether an aberrant nucleotide substitution process may have misled the phylogenetic reconstruction. For example, pervasive positive selection or variation in guanine/cytosine (GC) content can obscure true phylogenetic signal (16–18), thereby causing a gene tree to be incongruent with the species tree. Using codon models for tree inference potentially can accommodate complex selection profiles by allowing different nonsynonymous/synonymous substitution rate ratios to fall into distinct classes (19). However, we found that incorporating codon models did not improve the incongruence between the gene tree and species tree; the resultant tree largely matches that from the nucleotide substitution model, with comparable branch support values (*SI Appendix*, Figs. S6 and S7). Similarly, inferences based on first + second-codon positions or only on third-codon positions also yielded topologies discordant with the species tree (*SI Appendix*, Fig. S7).

We then used a random effects branch-site model to infer the dynamics of positive selection across the neochrome tree (20). Only five fern branches were identified as having experienced significant episodic positive selection (*SI Appendix*, Fig. S7*D*), and the proportion of positively selected codon sites along each of these five branches is very low (<3%). These results suggest that positive selection operated on very few codons over a limited number of branches. Similarly, a sliding window analysis of GC content found none of the fern sequences to be deviant in base composition (*SI Appendix*, Fig. S7*E*). Taken together, the nucleotide substitution processes among fern neochromes appear to be unexceptional and are not likely to explain the incongruence between the gene tree and species tree.

We therefore hypothesized that the incongruent tree could be the result of (*i*) multiple fern-to-fern HGT events, (*ii*) an elevated gene turnover rate that may have been selected for after HGT (21, 22), or (*iii*) a combination of both factors. We have some evidence suggesting recurring fern-to-fern HGT might have been involved. For example, we discovered neochrome genes from two early-diverging fern orders [Gleicheniales (*Dipteris conjugata*) and Cyatheales (*Alsophila podophylla* and *Plagiogyria* spp.)] that likely were derived from secondary HGT events (*SI Appendix*, Fig. S6, arrows). These neochromes are not phylogenetically resolved, as would be predicted based on published fern species relationships (14), but instead are nested among Polypodiales (*SI Appendix*, Fig. S6). Furthermore, the split between these and other fern neochromes (81 Mya, 95% highest posterior density interval: 59–106 Mya; *SI Appendix*, Fig. S5) occurred long after the estimated organismal divergence dates for Gleicheniales (276 Mya) and Cyatheales (223 Mya) (8), a pattern that may be explained best by fern-to-fern HGT.

Our hypothesis of potentially recurrent HGT events within ferns is not unprecedented. In angiosperms, rampant HGTs have been documented for the mitochondrial *cox1* homing intron. This intron is believed to have experienced one initial "seed transfer" from fungi that was followed by at least 80 incidents of plant-to-plant HGT among 833 diverse angiosperm species (23–25). Perhaps neochrome is similarly associated with mobile elements that may have facilitated its movement across species boundaries.

Evolutionary and Physiological Implications of Neochrome in Hornworts. Our discovery of neochrome in hornworts is an important step toward understanding the evolution of photosensory systems in plants. In the moss *Physcomitrella patens*, both red and blue light can elicit directional chloroplast movements, and these movements are mediated by molecular interactions between physically separate phytochrome and phototropin proteins (26). The hornwort neochrome represents a strikingly different strategy for integrating these two photosensory systems, combining them into a single, chimeric gene. Light-induced directional chloroplast movement has not yet been observed in hornworts, probably because their epidermal cells usually contain only one chloroplast that occupies most of the cellular space. However, nearly 50 y ago, Burr (27) documented an unusual chloroplast photoresponse in *Megaceros* hornworts; she discovered that the large chloroplasts "contract" to form compact shapes under strong light. Although we confirmed this phenomenon in hornworts (*SI Appendix*, Fig. S8), future studies are needed to examine if neochrome is responsible for the contraction response and to explore other possible physiological roles.

Evolutionary Significance of Plant-to-Plant HGT. This study pinpoints the origin of land plant neochrome within the hornwort lineage and demonstrates that neochrome was transferred horizontally from hornworts to ferns. The life history of ferns may help explain their hypothesized susceptibility to HGT. Most land plants share a common sexual life cycle that alternates between a diploid sporophyte and a haploid gametophyte; only in ferns and lycophytes are the sporophytic and gametophytic phases both free-living and fully independent. Seed plants insulate their gametophytes from outside interactions with relatively impervious cell walls in microgametophytes and by embedding megagametophytes within protective sporophyte tissues. In contrast, almost all fern gametophytes are not enclosed and grow in direct, intimate contact with other fern and bryophyte gametophytes (including those of hornworts). These characteristics may facilitate the entrance of foreign genetic elements into fern germ lines (i.e., via the gamete-producing structures, the antheridia and archegonia, that are exposed to the environment) (28).

To date, most documented examples of plant-to-plant HGT involve mitochondrial DNA and/or parasite–host transfers (29–36); only a handful of cases include functional nuclear genes (34, 37, 38), and even fewer have possible adaptive implications

(39). Consequently, plant-to-plant HGT generally has been overlooked as a potentially significant factor in plant evolution. Given that neochrome may have played a major role in promoting the diversification of ferns under the Cretaceous/Tertiary angiosperm canopy (3, 7, 8), our study has important implications for the macroevolutionary significance of plant-to-plant HGT.

Materials and Methods

Mining Transcriptomes and Whole-Genome Sequences for Homologs of Neochrome, Phototropin, and Phytochrome. All but one of the 434 transcriptomes used were generated by the One Thousand Plants Project (1KP; www.onekp.com); these transcriptomes were derived from a diverse selection of brown algae, red algae, green algae, bryophytes, lycophytes, ferns, and seed plants (*SI Appendix*, Table S1). Details on RNA extraction, sequencing, and assembly for 1KP can be found in Johnson et al. (40). Additionally, a whole-plant normalized Illumina transcriptome library was constructed and sequenced for *Pteridium aquilinum* using pooled RNA from six sporophyte tissues (young sporeling leaf, rhizome tip, fiddlehead, mature sterile pinnae, and pinnae with developing and mature sporangia). The *Pteridium* transcriptome was assembled using default parameters in the Trinity RNA-seq pipeline version r2012-01–25p1 (41). The sequencing reads were deposited in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under experiment SRX423244.

For the 1KP transcriptomes, we used both Short Oligonucleotide Assembly Package (SOAP) de novo and SOAP de novo *trans* assemblies. For each assembly, a BLAST database was constructed using the BLAST+ package (42). Neochrome, phototropin, and phytochrome sequences were queried separately (by tBLASTn for 1KP and by BLASTn for *Pteridium* assemblies), and the significant hits to transcriptome scaffolds were extracted (e-value threshold of $<10^{-5}$). For each scaffold, the best ORF was identified, the sequence was translated into amino acids, and then BLASTp queried against the NCBI nonredundant protein database (nr). The scaffolds were discarded if they did not match neochrome, phototropin, or phytochrome homologs in the nr database with an e-value threshold of <0.001. For 1KP transcriptomes, the filtered scaffolds from SOAP de novo and SOAP de novo *trans* assemblies were merged using CAP3 (43). We carried out the above procedures using our Python pipeline BlueDevil (see http://dx.doi.org/10.5061/dryad.fn2rg). We also searched for and obtained photoreceptor homologs from 39 plant and algae whole-genome sequences through Phytozome (44) and the *Amborella* Genome Database (www.amborella.org) (45).

Assembling and Mining an *Anthoceros punctatus* Draft Genome for Homologs of Neochrome, Phototropin, and Phytochrome. To generate a draft genome for *Anthoceros punctatus*, genomic DNA was sheared into ~400-bp fragments and sequenced using Illumina HiSeq2000, giving a total of 25 million 90-bp paired-end reads (about 20× genome coverage). The reads were subjected to two cycles of read error correction using the ALLPATHS-LG FindError program (46) before being assembled using Velvet (47). Assemblies were generated for a range of kmer values (k = 21, 31, 41, 51, and 61) and then were combined. The redundant scaffolds were removed using Usearch (48), and overlapping contigs were subject to additional assembly using CAP3 to produce a draft genome assembly. The final assembly contains 29,582 contigs with a total combined assembly length of 99.5 Mb and the N50 contig length of 4,955 bp. The contig length ranges from 919 bp to 76.5 kb, with 1,643 contigs over 10 kb. The median and mean assembled contig coverage is 18.1× and 44×, respectively. The raw reads were deposited in NCBI under SRA096687.

This assembly was searched for homologs of neochrome, phototropin, and phytochrome using tBLASTn. Although phototropin and phytochrome genes were readily identified, no contig was found containing a putative neochrome sequence. To search for any *A. punctatus* neochrome gene that perhaps failed to be assembled, all sequencing reads were searched against a library of neochrome protein sequences using BLASTx. Reads obtaining an e-value of $\leq 10^{-10}$ were isolated and assembled using Velvet with liberal assembly parameters (-cov_cutoff 1 -min_pair_count 1 -edgeFractionCutoff 0.1 -scaffolding yes -min_contig_lgth 90) at five different values for kmer length (21, 31, 41, 51, and 61). The resulting assemblies were combined, redundant contigs were discarded using Usearch, and overlapping contigs were merged using CAP3. All sequencing reads then were mapped to these seed contigs using Bowtie2 (49) with the very-sensitive-local option. Paired-end reads where at least one read mapped to the seed contigs were selected. All the selected reads then were reassembled as above. This mapping and assembly process was repeated until no further reads could be identified and contigs could no longer be extended. The final assembly contained a single contig comprising a 797-bp fragment of neochrome. This fragment then was extended to include almost the entire ORF using a combination of PCR (see below) and additional read mapping and assembly.

Cloning of Neochrome, Phototropin, and Phytochrome. To verify empirically the presence of the hornwort photoreceptor genes found in the transcriptomes and to obtain intron/exon information, we cloned the genes from genomic DNA from five hornwort species (*SI Appendix*, Table S3). In addition, neochrome sequences were obtained from 25 fern species by PCR and cloning (*SI Appendix*, Table S3). Genomic DNA was extracted using the Qiagen DNAeasy Plant Mini Kit (Qiagen). The gene fragments were amplified using Phusion DNA polymerase (New England Biolabs) or Denville Choice Taq (Denville). The primers and detailed PCR conditions are summarized in *SI Appendix*, Tables S3 and S4. The amplified products were cloned into Promega pGEM-T (Promega) and sequenced.

Genome-Walking In Hornwort Phototropin and Neochrome. To rule out the possibility that the phototropin gene found in hornworts might be a partial neochrome, we used inverse PCR (50) to obtain the flanking genomic region. Genomic DNA of *N. aenigmaticus* was digested by apoI (New England Biolabs) and self-ligated using T4 DNA ligase (New England Biolabs). Then nested PCRs were conducted on the circularized DNA. The amplicons were cloned using Promega pGEM-T and sequenced. To search for the genes flanking neochrome in *N. aenigmaticus*, we used the Clontech GenomeWalker kit (Clontech) and followed the manufacturer's manual. The resulting PCR amplicons were cloned and sequenced. In total, we obtained 3,291-bp and 4,578-bp regions up- and down-stream, respectively, of neochrome. The primers for the above PCR reactions are listed in *SI Appendix*, Table S4.

Sequence Alignment for Neochrome, Phototropin, and Phytochrome. We built two large alignments for phototropin and phytochrome, with each alignment including the corresponding domains from hornwort and fern neochrome. The phototropin dataset contains 163 sequences from 106 species, and the phytochrome dataset includes 139 sequences from 76 species. To reduce ambiguities in sequence alignment, we included only the conserved domains (i.e., LOV1, LOV2, and STK for phototropins; PAS, GAF, PHY, PAS repeats, HisKA, and HATPase for phytochromes). The domain boundaries were identified by querying each scaffold against the NCBI Conserved Domain Database (51). Each domain was aligned separately (based on the amino acid sequences) using Muscle (52) and then was concatenated. We developed a Python script, DomainDivider (see http://dx.doi.org/10.5061/dryad.fn2rg), to automate these processes. We also generated a separate alignment for hornwort and fern neochromes. This alignment was based on entire neochrome sequences rather than on domains. All alignments were inspected manually, and ambiguously aligned regions were excluded before phylogenetic analyses. The phototropin, phytochrome, and neochrome alignments contain 1,716, 2,802, and 4,002 bp, respectively. The GenBank accession numbers are listed in *SI Appendix*, Figs. S1–S3.

Phylogenetic Analyses of Phototropin and Neochrome. Phototropin and neochrome phylogenies were inferred based on their nucleotide alignments. We used PartitionFinder (53) to identify the optimal data partition schemes and nucleotide substitution models under the Akaike Information Criterion. Based on this analysis, each codon position was treated as a distinct partition. For phototropin, the first, second, and third positions were assigned GTR+Γ+I substitution models; for neochrome, GTR+Γ+I, GTR+Γ+I, GTR+I models were applied to each codon position respectively. We used Garli (54) to obtain the maximum likelihood tree under the aforementioned models, with genthreshfortopoterm set to 1,000,000 and eight independent runs. Multiparametric bootstrapping was done using RAxML (55) with 1,000 replicates. For the neochrome alignment, we also carried out the same maximum likelihood analyses on the first + second-codon positions and on the third-codon positions separately. We used MrBayes (56) to conduct Bayesian tree inference under the same models, with two independent Markov chain Monte Carlo (MCMC) runs, four chains each, and trees sampled every 1,000 generations. Substitution parameters were unlinked, and the rate prior was set to vary among partitions. The MrBayes output was inspected using Tracer (57) to ensure proper convergence and mixing (effective sample sizes all >200), and 25% of the total generations were discarded as burn-in before making the 50% majority consensus tree. Because the stationary, homogeneous assumptions of GTR might be violated in cases associated with HGT and deep divergence (58), we also used a nonstationary, heterogeneous nucleotide substitution model implemented in nhPhyML (59) to infer the phototropin tree. The analysis was carried out

with 10 discrete categories of GC equilibrium frequencies, and the required starting tree was the best tree from the Garli analysis. To conduct bootstrapping in nhPhyML, we created a Python wrapper, and for each replicate, RAxML was used to input the starting tree. In addition to the nucleotide substitution model, we also used codon models to infer phylogenies, which were carried out in CodonPhyML (19) under a maximum likelihood framework. We used the Goldman-Yang (60) model with four categories of nonsynonymous/synonymous substitution rate ratios drawn from the discrete gamma distribution, and codon frequencies were estimated from the data under the F3 × 4 model (19). The tree topology search was done using the nearest neighbor interchange approach, and branch support was estimated using the SH-like aLRT (61, 62) method.

Phylogenetic Analyses of Phytochrome. For the phytochrome phylogeny, we used the protein alignment following the analytical strategy of Mathews et al. (63). Using ProtTest (64), JTT + F was found to be the best empirical substitution model under the Akaike Information Criterion. For the maximum likelihood analyses, we used Garli to search for the maximum likelihood tree, with genthreshfortopoterm set to 1,000,000 and eight independent runs, and RAxML to conduct the multiparametric bootstrapping with 1,000 replicates. For Bayesian tree inference, we used MrBayes with two independent MCMC runs, four chains each, and trees sampled every 1,000 generations. After 25% of the total generations were removed, the 50% majority consensus tree was calculated. Codon-based tree inference also was carried out as described above.

Topology Test. We used the Swofford–Olsen–Waddell–Hillis test (65) to compare the inferred HGT tree topology (i.e., fern neochromes embedded within hornworts) against the alternative topologies suggestive of vertical inheritance or independent origin, using the program sowhat (66) with RAxML and Seq-Gen (67). For testing the vertical inheritance topology, topological constraints forcing fern and hornwort neochromes to be reciprocally monophyletic were used; for independent origin, constraints were placed to have all fern genes to be monophyletic (i.e., monophyly either as neochrome + phototropin or neochrome + phytochrome). To calculate the posterior probability of the vertical transfer and independent origin topologies, we filtered the posterior tree samples from MrBayes and calculated the frequency of trees given the monophyly constraints. The filtering was done by PAUP* (68). We applied this same approach to examine the posterior distribution of fern neochrome gene trees. We searched for topologies that exhibited better congruence with the published species relationships (compared with the inferred gene tree). The constraint for tree filtering required that neochromes from the same fern family be monophyletic.

Phylogenetic Analysis of the IGPD Gene. As a result of genome-walking in *N. aenigmaticus*, we discovered an IGPD pseudogene downstream from neochrome. To place this pseudogene in phylogenetic context, we resolved an IGPD phylogeny for land plants. A subset of the transcriptomes and whole-genome sequences was mined for IGPD homologs (*SI Appendix*, Fig. S4) using BlueDevil, and an alignment of IGPD (624 bp in length) was constructed manually. We partitioned the data by codon position, with each partition given a GTR+Γ+I model as suggested by PartitionFinder under the Akaike Information Criterion. Maximum likelihood analyses were carried out in RAxML with 100 random starting trees, and multiparametric bootstrapping was done with 1,000 replicates.

Divergence Time Estimation of the Phototropin Gene Family. We used BEAST (69) to infer simultaneously the divergence times and phylogeny of the phototropin gene family. As recommended by PartitionFinder, the phototropin dataset was partitioned by codon position, each with the GTR+Γ+I substitution model. A total of 15 calibration priors were used (see *SI Appendix*, Table S2 for details) (8, 13, 70–77), and a birth–death speciation prior was used as the tree prior. We used the uncorrelated relaxed-clock model with rates drawn from a lognormal distribution. A starting tree was first estimated by r8s (78) and was provided to BEAST to initiate the run. Two independent MCMC runs were carried out, and the output was inspected in Tracer to ensure convergence and mixing (effective sample sizes all >200). The trees from the two runs were combined in LogCombiner (69) with a 25% burn-in and were summarized in TreeAnnotator (69). It should be noted that the stationary, homogeneous GTR model used here could be violated, especially in the case of HGT, and might affect the divergence estimates. However, there is no nonstationary, heterogeneous model that is currently implemented in divergence time analyses, and our results should be revisited in the future when more sophisticated methods are available.

Inferring Episodic Selection and GC Content Variation in Neochrome Evolution. To investigate whether fern neochromes had experienced pervasive episodic positive selection, we used the unrestricted, random effects branch-site model (20) implemented in the HyPhy package (20, 79). Branches with episodic positive selection were identified by the sequential likelihood ratio test (20). The neochrome alignment and the best maximum likelihood tree were used as the input data. The analyses were carried out on the Datamonkey server (79, 80). A GC content sliding window was constructed using a custom Python script; each window is 400 bp in size, and the window slides every 50 bp.

ACKNOWLEDGMENTS. We thank Y.-L. Qiu, C.-W. Li, J. Nelson, B. Shaw, J. Duff, D. Long, and L. Forrest for helping us obtain hornwort materials; L. Huiet for laboratory assistance; P. G. Wolf, C. W. dePamphilis, and P. E. Ralph for help in collecting the *Pteridium aquilinum* transcriptome data; J. Leebens-Mack, S. Joya, S. Ellis, and others for contributing plant tissues for 1KP transcriptome sequencing; D. L. Swofford and S. Wu for help with the phylogenetic analyses; and J. D. Palmer, M. Chen, J. Meireles, two anonymous reviewers, members of the K.M.P. laboratory, and members of the Duke Systematics Discussion Group for insightful comments on a draft manuscript. This work was supported by research grants from the Society of Systematic Biologists, Sigma Xi, American Society of Plant Taxonomists, and Duke University (to F.-W.L.), National Science Foundation Doctoral Dissertation Improvement Grant DEB-1407158 (to F.-W.L. and K.M.P.), DEB-1145614 (to K.M.P.), and a European Research Council Advanced Investigator Award (to J.A.L.). The 1000 Plants (1KP) initiative, led by G.K.-S.W., is funded by the Alberta Ministry of Enterprise and Advanced Education, Alberta Innovates Technology Futures, Innovates Centre of Research Excellence, Musea Ventures, and BGI-Shenzhen. This article is part of a doctoral dissertation in Biology at Duke University by F.-W.L., supported by a National Science Foundation Graduate Research Fellowship.

1. Möglich A, Yang X, Ayers RA, Moffat K (2010) Structure and function of plant photoreceptors. *Annu Rev Plant Biol* 61:21–47.
2. Nozue K, et al. (1998) A phytochrome from the fern *Adiantum* with features of the putative photoreceptor NPH1. *Proc Natl Acad Sci USA* 95(26):15826–15830.
3. Kawai H, et al. (2003) Responses of ferns to red light are mediated by an unconventional photoreceptor. *Nature* 421(6920):287–290.
4. Kanegae T, Hayashida E, Kuramoto C, Wada M (2006) A single chromoprotein with triple chromophores acts as both a phytochrome and a phototropin. *Proc Natl Acad Sci USA* 103(47):17997–18001.
5. Suetsugu N, Mittmann F, Wagner G, Hughes J, Wada M (2005) A chimeric photoreceptor gene, *NEOCHROME*, has arisen twice during plant evolution. *Proc Natl Acad Sci USA* 102(38):13705–13709.
6. Yang Y, Qi X, Sen L, Su Y, Wang T (2010) Cloning and sequence analysis of red/blue light chimeric photoreceptor genes from three fern species (*Coniogramme intermedia* var. *glabra, Plagiogyria distinctissima* and *Pronephrium lakhimpurnense*). *Am Fern J* 100(1):1–15.
7. Schneider H, et al. (2004) Ferns diversified in the shadow of angiosperms. *Nature* 428(6982):553–557.
8. Schuettpelz E, Pryer KM (2009) Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proc Natl Acad Sci USA* 106(27):11200–11205.
9. Gontcharov AA, Melkonian M (2010) Molecular phylogeny and revision of the genus *Netrium* (Zygnematophyceae, Streptophyta): *Nucleotaenium* gen. nov. *J Phycol* 46(2): 346–362.
10. Qiu Y-L, et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103(42):15511–15516.
11. Cox CJ, Li B, Foster PG, Embley TM, Civán P (2014) Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Syst Biol* 63(2):272–279.
12. Hedges SB, Kumar S (2009) *The Timetree of Life* (Oxford Univ Press, Oxford, UK).
13. Villarreal JC, Renner SS (2012) Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc Natl Acad Sci USA* 109(46):18873–18878.
14. Schuettpelz E, Pryer KM (2007) Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon* 56(4):1037–1050.
15. Rothfels CJ, et al. (2013) Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS ONE* 8(10):e76957.
16. Sanderson MJ, Shaffer HB (2002) Troubleshooting molecular phylogenetic analyses. *Annu Rev Ecol Syst* 33:49–72.
17. Kapralov MV, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol* 7:73.
18. Nabholz B, Künstner A, Wang R, Jarvis ED, Ellegren H (2011) Dynamic evolution of base composition: Causes and consequences in avian phylogenomics. *Mol Biol Evol* 28(8):2197–2210.
19. Gil M, Zanetti MS, Zoller S, Anisimova M (2013) CodonPhyML: Fast maximum likelihood phylogeny estimation under codon substitution models. *Mol Biol Evol* 30(6): 1270–1280.

20. Kosakovsky Pond SL, et al. (2011) A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol* 28(11):3033–3043.
21. Lind PA, Tobin C, Berg OG, Kurland CG, Andersson DI (2010) Compensatory gene amplification restores fitness after inter-species gene replacements. *Mol Microbiol* 75(5):1078–1089.
22. Näsvall J, Sun L, Roth JR, Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338(6105):384–387.
23. Cho Y, Qiu YL, Kuhlman P, Palmer JD (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc Natl Acad Sci USA* 95(24):14244–14249.
24. Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD (2008) Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. *Mol Biol Evol* 25(8):1762–1777.
25. Sanchez-Puerta MV, et al. (2011) Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. *BMC Evol Biol* 11:277.
26. Jaedicke K, Lichtenthäler AL, Meyberg R, Zeidler M, Hughes J (2012) A phytochrome-phototropin light signaling complex at the plasma membrane. *Proc Natl Acad Sci USA* 109(30):12231–12236.
27. Burr FA (1968) *Chloroplast Structure and Division in Megaceros Species*. PhD Dissertation (Univ of California, Berkeley).
28. Huang J (2013) Horizontal gene transfer in eukaryotes: The weak-link model. *Bioessays* 35(10):868–875.
29. Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424(6945):197–201.
30. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci USA* 101(51):17747–17752.
31. Davis CC, Wurdack KJ (2004) Host-to-parasite gene transfer in flowering plants: Phylogenetic evidence from Malpighiales. *Science* 305(5684):676–678.
32. Davis CC, Anderson WR, Wurdack KJ (2005) Gene transfer from a parasitic flowering plant to a fern. *Proc Biol Sci* 272(1578):2237–2242.
33. Yoshida S, Maruyama S, Nozaki H, Shirasu K (2010) Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328(5982):1128.
34. Renner SS, Bellot S (2012) in *Genomics of Chloroplasts and Mitochondria* (Springer Netherlands, Dordrecht), Vol 35, *Advances in Photosynthesis and Respiration*, pp 223–235.
35. Xi Z, et al. (2013) Massive mitochondrial gene transfer in a parasitic flowering plant clade. *PLoS Genet* 9(2):e1003265.
36. Rice DW, et al. (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342(6165):1468–1473.
37. Xi Z, et al. (2012) Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics* 13:227.
38. Zhang Y, et al. (2013) Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant *Phelipanche aegyptiaca* and related species. *BMC Evol Biol* 13:48.
39. Christin P-A, et al. (2012) Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. *Curr Biol* 22(5):445–449.
40. Johnson MTJ, et al. (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS ONE* 7(11):e50226.
41. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.
42. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
43. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9(9):868–877.
44. Goodstein DM, et al. (2012) Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40(Database issue):D1178–D1186.
45. Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science* 342(6165):1241089.
46. Maccallum I, et al. (2009) ALLPATHS 2: Small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol* 10(10):R103.
47. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
48. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.
49. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
50. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120(3):621–623.
51. Marchler-Bauer A, et al. (2011) CDD: A conserved domain database for the functional annotation of proteins. *Nucleic Acids Res* 39(Database issue):D225–D229.
52. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792–1797.
53. Lanfear R, Calcott B, Ho SYW, Guindon S (2012) Partitionfinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29(6):1695–1701.
54. Zwickl DJ (2006) *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion*. PhD dissertation (Univ of Texas at Austin, Austin).
55. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
56. Ronquist F, et al. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539–542.
57. Rambaut A, Drummond AJ (2009) Tracer v1.5. Available at http://tree.bio.ed.ac.uk/software/tracer/. Accessed June 2013.
58. Verbyla KL, Yap VB, Pahwa A, Shao Y, Huttley GA (2013) The embedding problem for markov models of nucleotide substitution. *PLoS ONE* 8(7):e69187.
59. Boussau B, Gouy M (2006) Efficient likelihood computations with nonreversible models of evolution. *Syst Biol* 55(5):756–768.
60. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11(5):725–736.
61. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55(4):539–552.
62. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.
63. Mathews S, Clements MD, Beilstein MA (2010) A duplicate gene rooting of seed plants and the phylogenetic position of flowering plants. *Philos Trans R Soc Lond B Biol Sci* 365(1539):383–395.
64. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
65. Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. *Syst Biol* 49(4):652–670.
66. Church SH, Ryan JF, Dunn CW (2014) Sowhat. Available from GitHub repository https://github.com/josephryan/sowhat. Accessed January 2014.
67. Rambaut A, Grassly NC (1997) Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci* 13(3): 235–238.
68. Swofford DL (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0a131* (Sinauer Associates, Sunderland, MA).
69. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.
70. Clarke JT, Warnock RCM, Donoghue PCJ (2011) Establishing a time-scale for plant evolution. *New Phytol* 192(1):266–301.
71. Hubers M, Kerp H (2012) Oldest known mosses discovered in Mississippian (late Visean) strata of Germany. *Geol* 40(8):755–758.
72. Guo C-Q, et al. (2012) *Riccardiothallus devonicus* gen. et sp. nov., the earliest simple thalloid liverwort from the Lower Devonian of Yunnan, China. *Rev Palaeobot Palynol* 176-177(C):35–40.
73. Prasad V, Strömberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. *Science* 310(5751):1177–1180.
74. Kotyk ME, Basinger JF, Gensel PG, de Freitas TA (2002) Morphologically complex plant macrofossils from the Late Silurian of Arctic Canada. *Am J Bot* 89(6):1004–1013.
75. Skog JE, Banks HP (1973) *Ibyka amphikoma*, gen. et sp. n., a new protoarticulate precursor from the late Middle Devonian of New York State. *Am J Bot* 60(4):366–380.
76. Galtier J, Wang S-J, Li C-S, Hilton J (2001) A new genus of filicalean fern from the Lower Permian of China. *Bot J Linn Soc* 137(4):429–442.
77. Trivett ML (1992) Growth architecture, structure, and relationships of *Cordaixylon iowensis* nov. comb. (Cordaitales). *Int J Plant Sci* 153(2):273–287.
78. Sanderson MJ (2003) r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19(2):301–302.
79. Kosakovsky Pond SL, Frost SDW, Muse SV (2005) HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.
80. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26(19): 2455–2457.

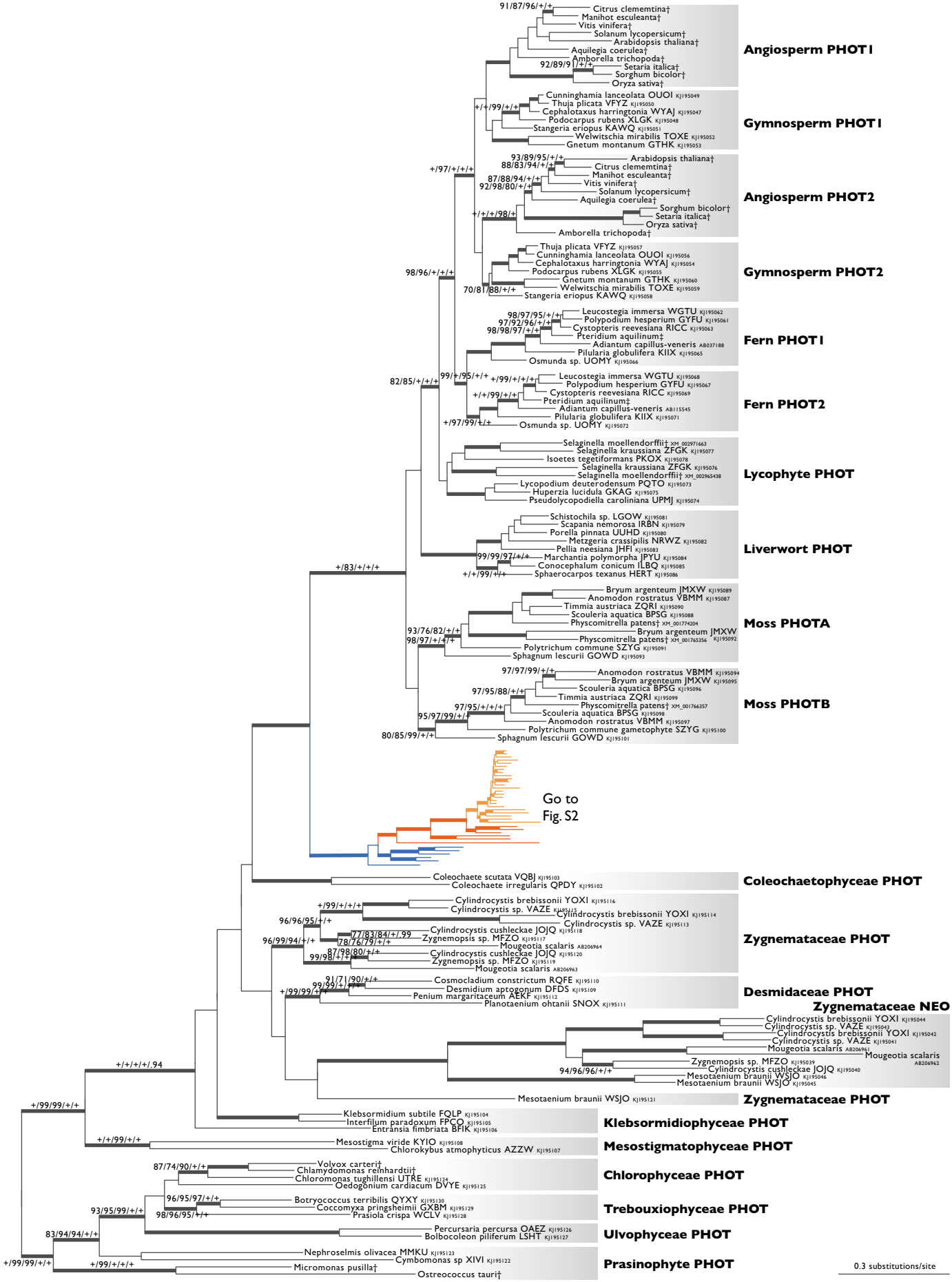

Fig. S1. Phylogenetic relationships of land plant and algal phototropin (PHOT) and the corresponding domains from hornwort, fern, and algal neochrome (NEO). Topology derived from the best maximum likelihood tree. The five support values associated with branches are maximum likelihood bootstrap values (BS) from Garli / BS from nhPhyML / aLRT supports under codon model (aLRT) / Bayesian posterior probabilities (PP) from MrBayes / PP from BEAST; these are only displayed (along with thickened branches) when BS > 70, aLRT > 70 and PP > 0.95. "+" denotes BS = 100, aLRT = 100 or PP = 1.00; thickened branches without numbers are "+/+/+/+/+". Alphanumeric codes following species names are the four-letter 1KP transcriptome identifiers, or Genbank accessions, or both; "†" indicates the sequence came from genome sequence data, and "‡" from *Pteridium aquilinum* transcriptome. The blue, orange and yellow branches represent hornwort phototropin, hornwort neochrome and fern neochrome, respectively.

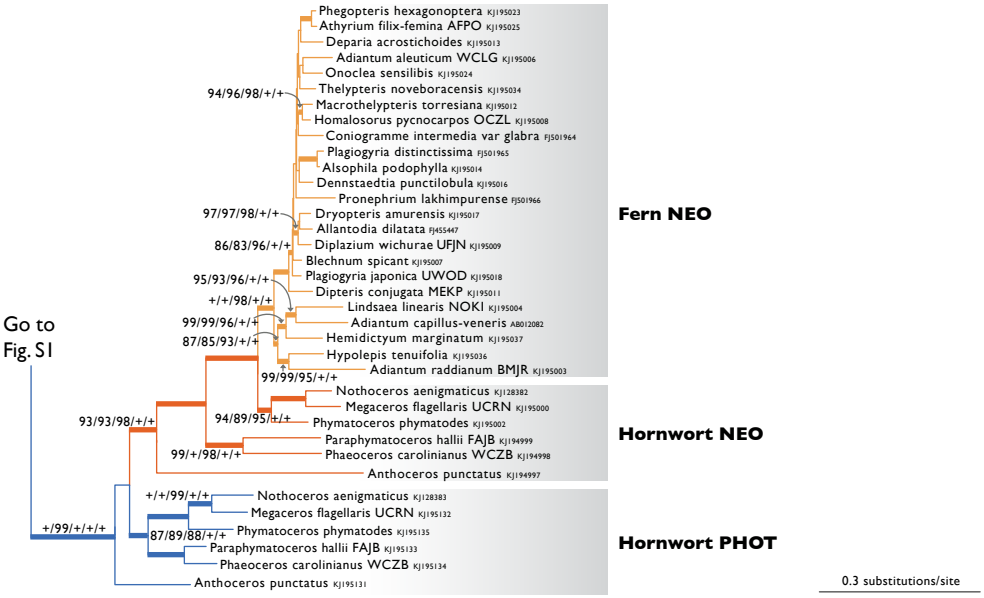

Fig. S2. Phylogenetic relationships of fern neochrome (NEO), hornwort neochrome and phototropin (PHOT). This figure is continued from Fig. S1 and follows the same conventions.

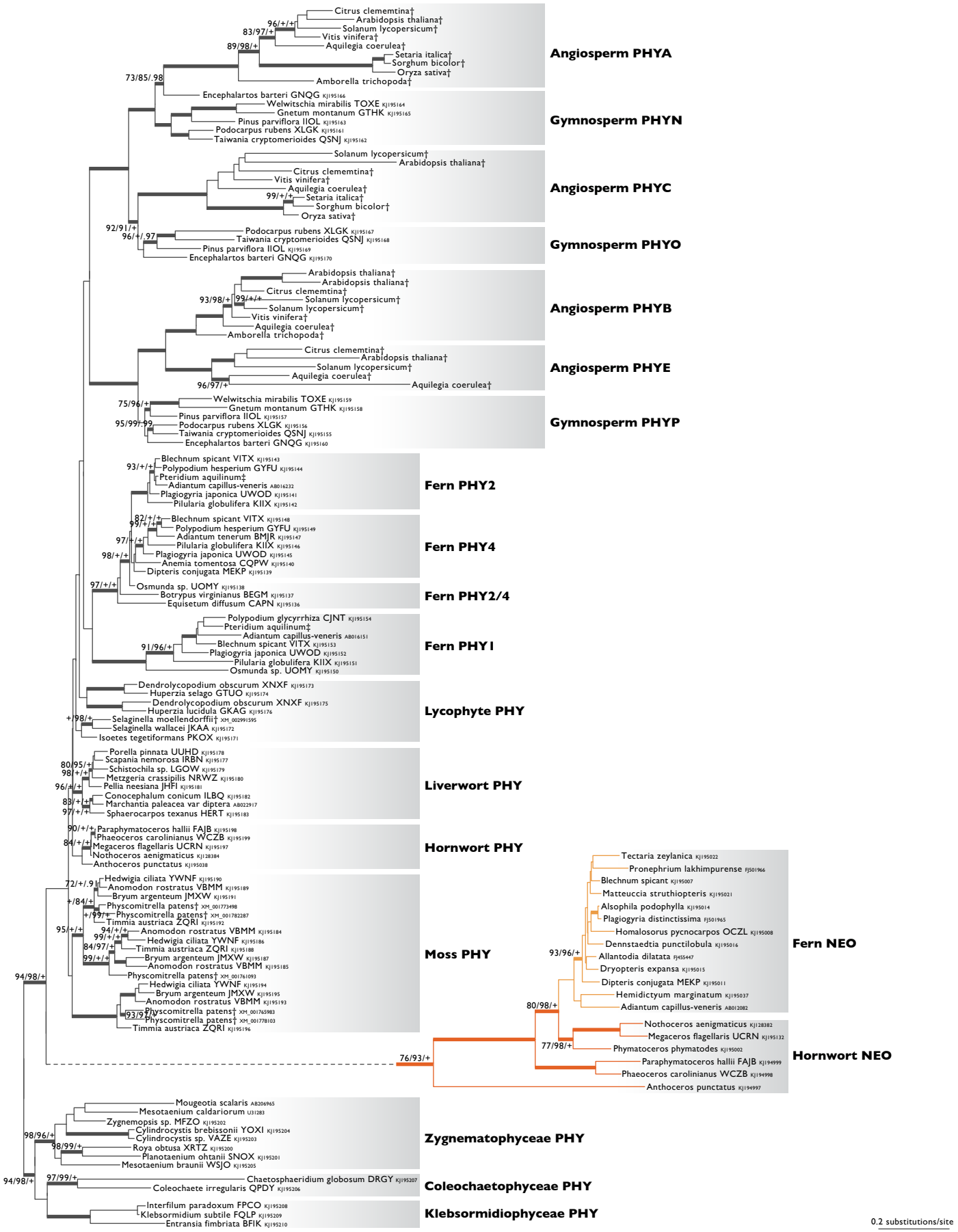

Fig. S3. Phylogenetic relationships of land plant and algal phytochrome (PHY) and the corresponding domains from hornwort and fern neochrome (NEO). Topology derived from the best maximum likelihood tree. The three support values associated with branches are maximum likelihood bootstrap values (BS) / aLRT supports under codon model (aLRT) / Bayesian posterior probabilities (PP) from MrBayes; these are only displayed (along with thickened branches) if BS > 70, aLRT > 0.95 and PP > 0.95. "+" denotes BS = 100, aLRT = 100 or PP = 1.00; thickened branches without numbers are "+/+/+". Alphanumeric codes following species names are the four-letter 1KP transcriptome identifiers, or Genbank accessions, or both; "†" indicates the sequence came from whole genome sequence data, and "‡" from *Pteridium aquilinum* transcriptome. For space considerations, the dashed line artificially extends the NEO clade and does not reflect true branch length. The orange and yellow branches represent hornwort neochrome and fern neochrome, respectively.

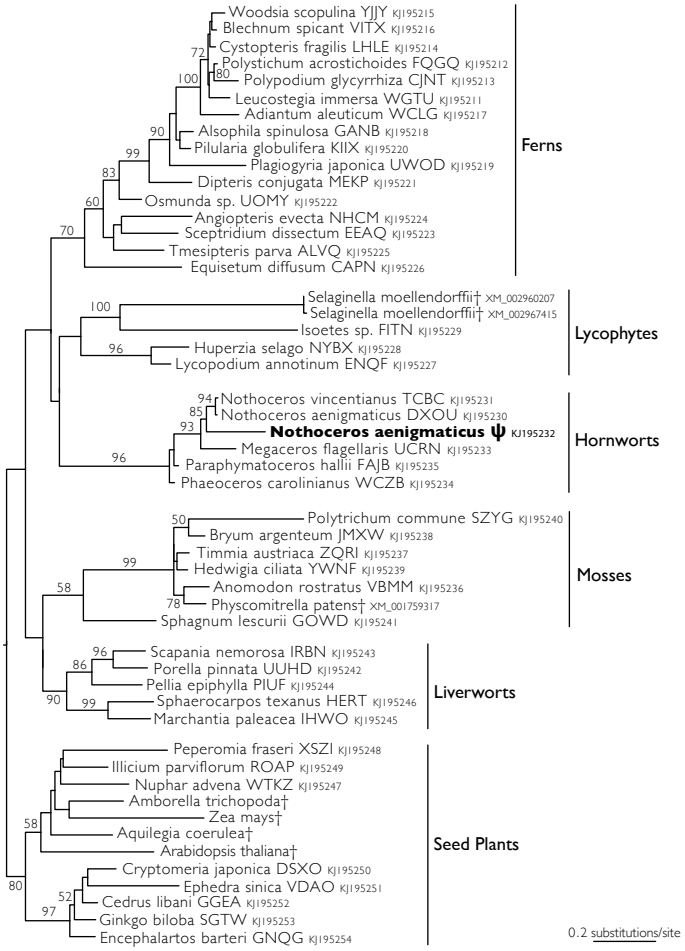

Fig. S4. Phylogenetic relationships of land plant imidazoleglycerol-phosphate dehydratase (IGPD). In the hornwort *Nothoceros aenigmaticus*, we conducted genome-walking downstream of neochrome and found a IGPD pseudogene (denoted by ψ). In a land plant phylogeny of IGPD our *N. aenigmaticus* pseudogene is most closely related to other hornwort IGPD. This relationship confirms that our hornwort neochrome sequence data were indeed derived from the hornwort genome, and not from symbiotic algae or fungi. Numbers associated with branches are maximum likelihood bootstrap support values. Alphanumeric codes following species names are the four-letter 1KP transcriptome identifiers, or Genbank accessions, or both; "†" indicates the sequence came from whole genome sequence data.

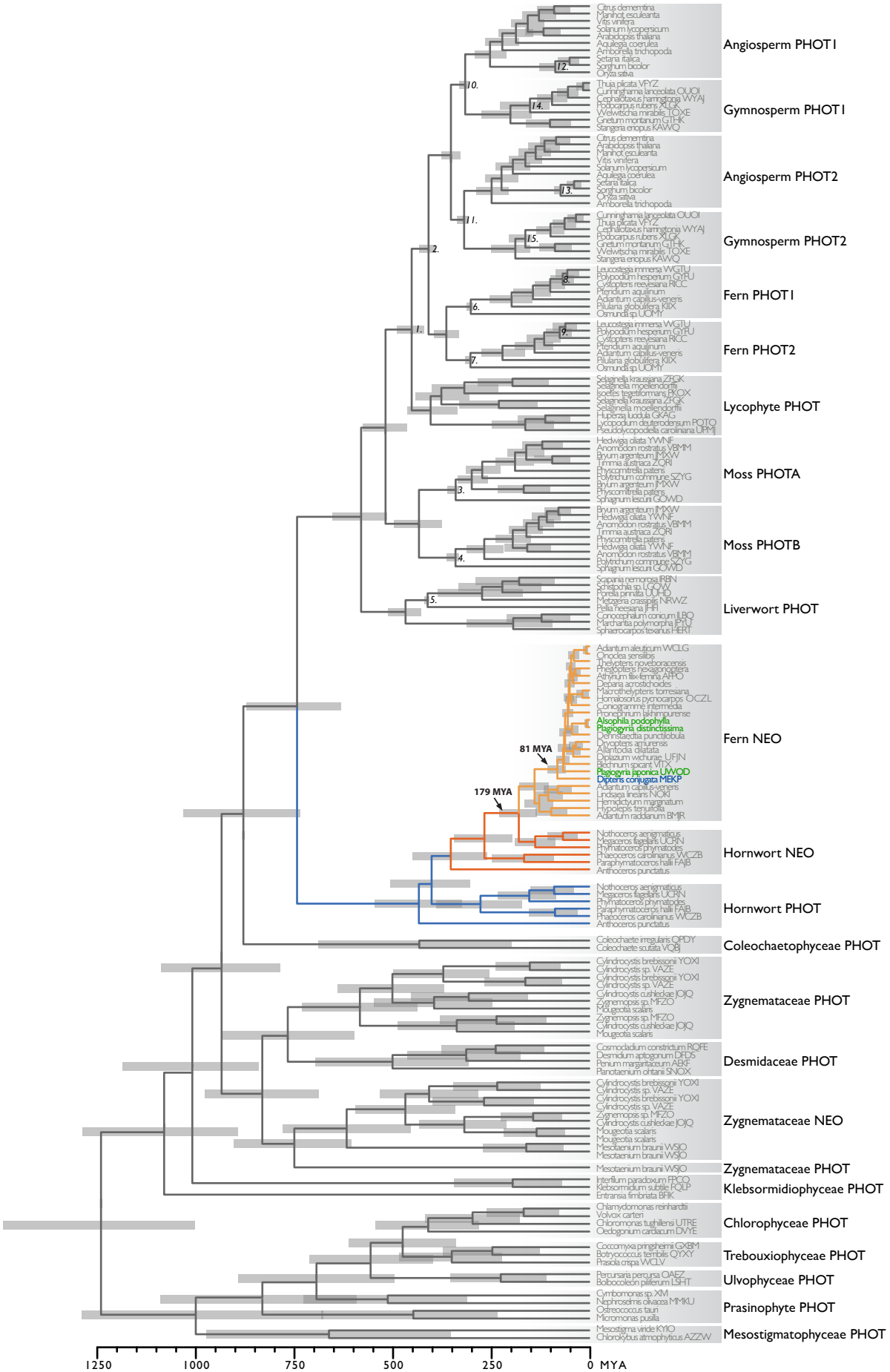

Fig. S5. Chronogram of land plant and algal phototropin (PHOT) and the corresponding domains from hornwort, fern, and algal neochrome (NEO). A simplified version of this figure is shown in Fig. 1B. Grey bars represent 95% highest posterior density intervals of the age estimates. Italicized numbers adjacent to nodes refer to the fossil or secondary time calibrations detailed in Table S2. Two divergence time estimates are highlighted: one marks the HGT event (179 MYA) and the other marks the split of Gleicheniales (blue taxon), Cyatheales (green taxa) and other neochromes (81 MYA).

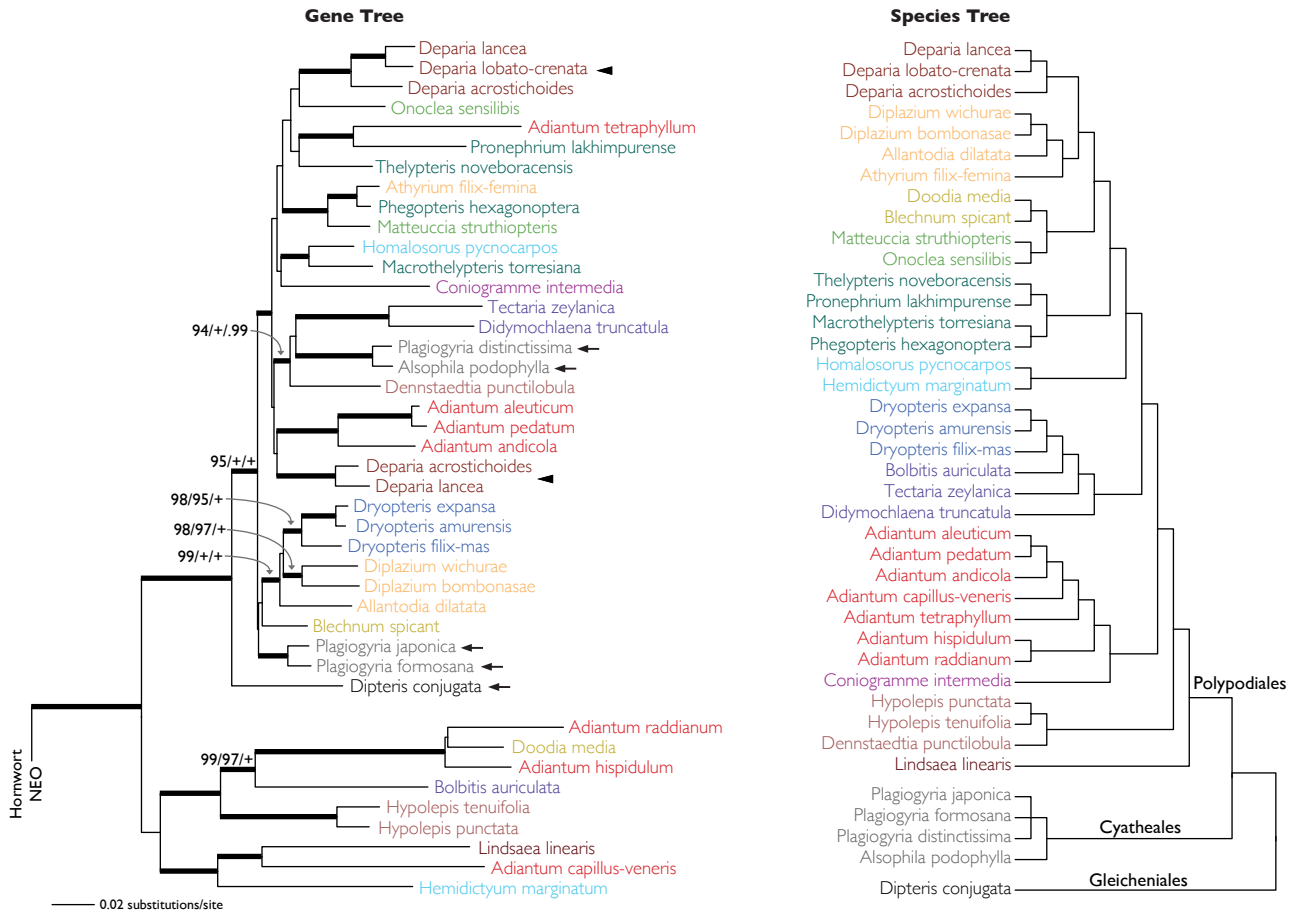

Fig. S6. Phylogenetic incongruence between fern neochrome gene tree and fern species tree. The gene tree topology is derived from the best maximum likelihood tree based on the nucleotide dataset, and the species tree summarized from Schuettpelz and Pryer (1), Kuo et al (2), Rothfels and Schuettpelz (3), and Rothfels et al (4). Tree inference based on codon models, 1st + 2nd and 3rd codon positions yielded similar topologies (Fig. S7). Closely related species/genera are coded with the same color. The neochrome gene tree is rooted with hornwort neochromes (not shown). Numbers above branches are maximum likelihood bootstrap values (BS) / aLRT supports under codon model (aLRT) / Bayesian posterior probabilities from MrBayes (PP), and are only displayed (along with thickened branches) if BS > 70, aLRT > 70 and PP > 0.95. "+" denotes BS = 100, aLRT = 100 or PP = 1.00; thickened branches without numbers are "+/+/+". Arrowheads point to the two divergent neochrome copies found in *Deparia* spp. Arrows point to neochromes from Gleicheniales and Cyatheales that appear nested among Polypodiales neochromes.

1. Schuettpelz E, Pryer KM (2007) Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon* 56:1037–1050.
2. Kuo LY, Li FW, Chiou WL, Wang CN (2011) First insights into fern matK phylogeny. *Mol Phylogenet Evol* 59:556–566.
3. Rothfels CJ, Schuettpelz E (2013) Accelerated rate of molecular evolution for vittarioid ferns is strong but not driven by selection. *Syst Biol* 63:31–54.
4. Rothfels CJ et al. (2013) Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS ONE* 8:e76957.

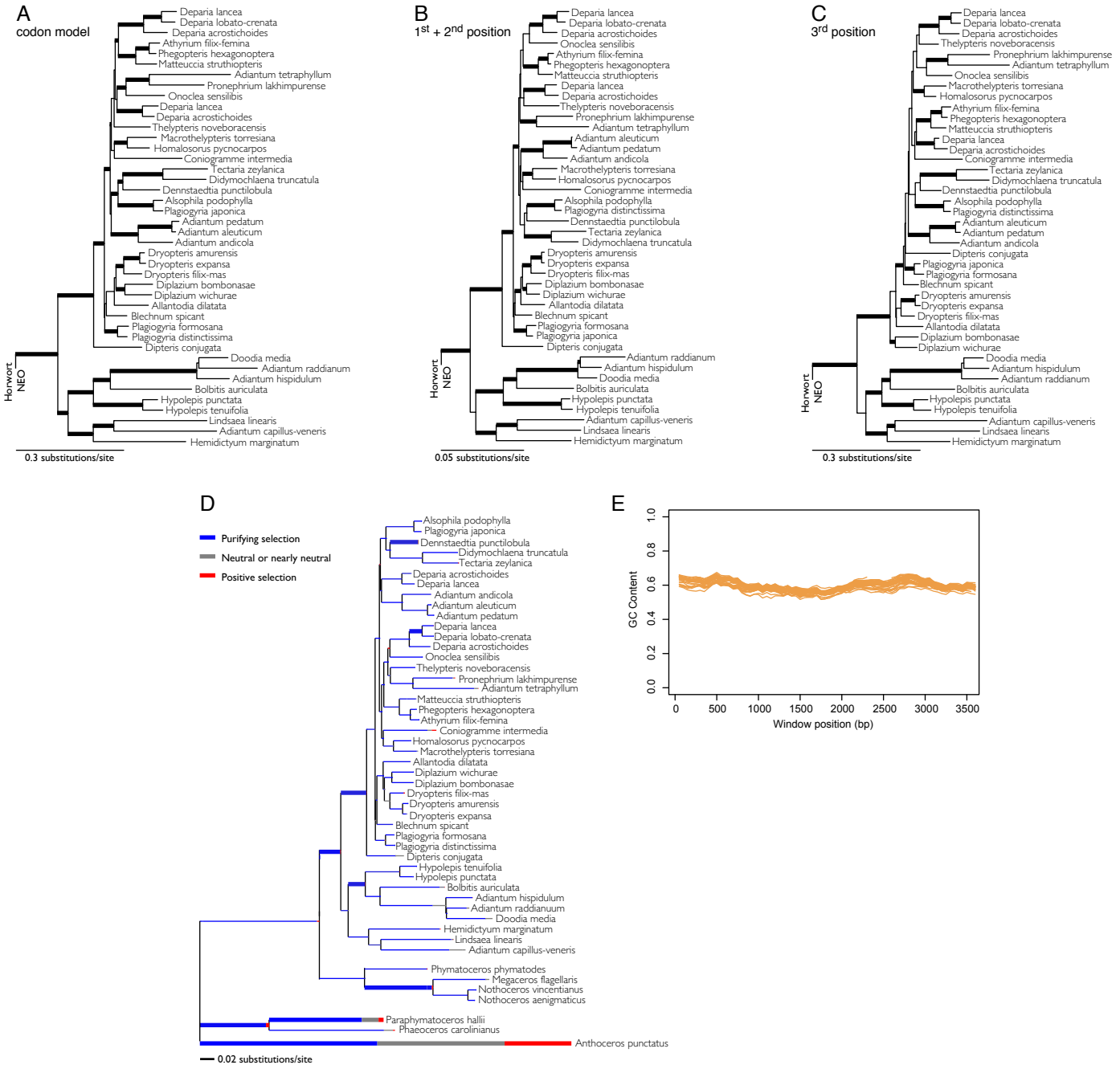

Fig. S7. Phylogeny, selection profile and GC content of fern neochromes. Maximum likelihood reconstructions of gene phylogeny based on (*A*) codon model, (*B*) first and second codon positions, and (*C*) third codon position. Thickened branches indicate aLRT supports (in *A*) or bootstrap supports (in *B*, *C*) > 70. (*D*) Selection profile displayed along phylogenetic branches for fern and hornwort neochromes. Tree topology derived from the best maximum likelihood tree (Fig. S6). The width of each color along a branch is proportional to the number of codon sites in the corresponding selection class. Thickened branches have experienced significant episodic positive selection ($P < 0.05$). (*E*) Sliding window analysis of GC content for fern neochrome. Each line displays the GC content for each neochrome sequence. None of the ferns in our study were deviant in base composition for neochrome. Each window is 400bp in size and the window slides every 50bp.

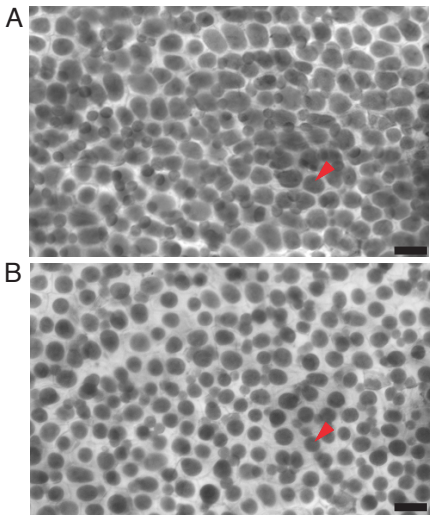

Fig. S8. Hornwort chloroplasts (arrowhead) contract under strong light. (*A*) Before irradiation, chloroplasts of *Nothoceros aenigmaticus* occupy most of the cellular space. (*B*) After irradiation of blue light (57 μmol m^{-2} s^{-1}) for 2 hours, chloroplasts evidently reduced in size. Scale bar = 40 μm.

Table S1. List of transcriptomes and genome sequences screened for neochrome, phototropin and phytochrome genes. All the transcriptomes were from 1KP (www.onekp.com), except for *Pteridium aquilium*. The four letter codes following species names are the 1KP transcriptome identifiers. Details of transcriptome tissue type and specimen voucher can be found at www.onekp.com. "*" denotes whole or draft genome sequences.

Angiosperms
- Amborella trichopoda*
- Aquilegia coerulea*
- Arabidopsis lyrata*
- Arabidopsis thaliana*
- Brachypodium distachyon*
- Capsella rubella*
- Carica papaya*
- Citrus clementina*
- Citrus sinensis*
- Cucumis sativus*
- Eucalyptus grandis*
- Fragaria vesca*
- Glycine max*
- Gossypium raimondii*
- Linum usitatissimum*
- Malus domestica*
- Manihot esculenta*
- Medicago truncatula*
- Mimulus guttatus*
- Oryza sativa*
- Panicum virgatum*
- Phaseolus vulgaris*
- Populus trichocarpa*
- Prunus persica*
- Ricinus communis*
- Setaria italica*
- Solanum lycopersicum*
- Solanum tuberosum*
- Sorghum bicolor*
- Theobroma cacao*
- Vitis vinifera*
- Zea mays*

Gymnosperms
- Austrotaxus spicata BTTS
- Callitris macleayana RMMV
- Cathaya agryrophylla NPRL
- Cedrus libani GGEA
- Cephalotaxus harringtonia WYAJ
- Cryptomeria japonica DSXO
- Cunninghamia lanceolata OUOI
- Cupressus dupreziana QNGJ
- Cycas micholitzii XZUY
- Dioon edule WLIC
- Encephalartos barteri GNQG
- Ephedra sinica VDAO
- Ginkgo biloba SGTW
- Glyptostrobus pensilis OXGJ
- Gnetum montanum GTHK
- Juniperus scopulorum XMGP
- Keteleeria evelyniana JUWL
- Larix speciosa WVWN
- Nothotsuga longibracteata AREG
- Phyllocladus hypohyllus JRNA
- Picea engelmanii AWQB
- Pinus jeffreyi MFTM
- Pinus parviflora IIOL
- Pinus ponderosa JBND
- Pinus radiata DZQM
- Podocarpus coriaceus SCEB
- Podocarpus rubens XLGK
- Pseudolarix amabilis AQFM
- Pseudotaxus chienii YLPM
- Pseudotsuga menziesii IOVS
- Sciadopitys verticillata YFZK
- Stangeria eriopus KAWQ
- Taiwania cryptomerioides QSNJ
- Taxus baccata WWSS
- Taxus cuspidata ZYAX
- Thuja plicata VFYZ
- Thujopsis dolabrata NKIN
- Torreya nucifera HQOM
- Torreya taxifolia EFMS
- Tsuga heterophylla GAMH
- Welwitschia mirabilis TOXE
- Widdringtonia cedarbergensis AUDE
- Wollemia nobilis RSCE

Ferns
- Adiantum aleuticum WCLG
- Adiantum raddianum BMJR
- Anemia tomentosa CQPW
- Angiopteris evecta NHCM
- Argyrochosma nivea XDDT
- Asplenium nidus PSKY
- Asplenium platyneuron KJZG
- Azolla caroliniana CVEG
- Athyrium filix-femina URCP
- Athyrium filix-femina AFPO
- Blechnum spicant VITX
- Bolbitis repanda JBLI
- Botrypus virginianus BEGM
- Cibotium glaucum ORJE
- Crepidomanes venosum TWFZ
- Cryptogramma acrostichoides WQML
- Culcita macrocarpa PNZO
- Cyathea spinulosa GANB
- Cystopteris fragilis XXHP
- Cystopteris fragilis LHLE
- Cystopteris protrusa YOWV
- Cystopteris reevesiana RICC
- Cystopteris utahensis HNDZ
- Davallia fejeensis OQWW
- Dennstaedtia davallioides MTGC
- Deparia lobato-crenata FCHS
- Didymochlaena truncatula RFRB
- Diplazium wichurae UFJN
- Dipteris conjugata MEKP
- Equisetum diffusum CAPN
- Equisetum hyemale JVSZ
- Gaga arizonica DCDT
- Gymnocarpium dryopteris HEGQ
- Hemionitis arifolia ZXJO
- Homalosorus pycnocarpos OCZL
- Hymenophyllum bivalve QIAD
- Hymenophyllum cupressiforme TRPJ
- Leucostegia immersa WGTU
- Lindsaea linearis NOKI
- Lindsaea microphylla YIXP
- Lygodium japonicum PBUU
- Marattia sp. UXCS
- Myriopteris eatonii GSXD
- Nephrolepis exaltata NWWI
- Notholaena montieliae YCKE
- Onoclea sensibilis HTFH
- Ophioglossum petiolatum QHVS
- Ophioglossum petiolatum WTJG
- Osmunda javanica VIBO
- Osmunda regalis YKSS
- Osmunda regalis UOMY
- Osmundastrum cinnamomeum BIVQ
- Pilularia globulifera KIIX
- Pityrogramma trifoliata UJTT
- Plagiogyria japonica UWOD
- Pleopeltis polypodioides UJWU
- Polypodium amorphum YLJA
- Polypodium glycyrrhiza CJNT
- Polypodium hesperium GYFU
- Polypodium hesperium IXLH
- Phlebodium pseudoaureum ZQYU
- Polystichum acrostichoides FQGQ
- Psilotum nudum QVMR
- Pteridium aquilinum[1]
- Pteris ensiformis FLTD
- Pteris vittata POPJ
- Sceptridium dissectum EEAQ
- Sticherus lobatus XDVM
- Thelypteris acuminata MROH
- Thyrsopteris elegans EWXK
- Tmesipteris parva ALVQ
- Vittaria appalachiana NDUV
- Vittaria lineata SKYV
- Woodsia ilvensis YQEC
- Woodsia scopulina YJJY

Lycophytes
- Dendrolycopodium obscurum XNXF
- Diphasiastrum digitatum WAFT
- Huperzia lucidula GKAG
- Huperzia myrisinites CBAE
- Huperzia selago GTUO
- Huperzia selago NYBX
- Huperzia squarrosa GAON
- Lycopodiella apressa ULKT
- Lycopodium annotinum ENQF
- Lycopodium deuterodensum PQTO
- Phylloglossum drummondii ZZEI
- Pseudolycopodiella caroliniana UPMJ
- Selaginella moellendorffii*
- Selaginella acanthonota ZYCD
- Selaginella apoda LGDQ
- Selaginella kraussiana ZFGK
- Selaginella lepidophylla ABIJ
- Selaginella selaginoides KUXM
- Selaginella stauntoniana ZZOL
- Selaginella wallacei JKAA
- Selaginella willdenowii KJYC
- Isoetes sp. PYHZ
- Isoetes sp. FITN
- Isoetes tegetiformans PKOX

Hornworts
- Anthoceros punctatus*[2]
- Megaceros flagellaris UCRN
- Nothoceros vincentianus TCBC
- Nothoceros aenigmaticus DXOU
- Paraphymatoceros hallii FAJB
- Phaeoceros carolinianus WCZB

Liverworts
- Bazzania trilobata WZYK
- Blasia sp. AEXY
- Calypogeia fissa RTMU
- Conocephalum conicum ILBQ
- Frullania sp. CHJJ
- Frullania sp. TGKW
- Lunularia cruciata TXVB
- Marchantia emarginata TFYI
- Marchantia paleacea IHWO
- Marchantia paleacea HMHL
- Marchantia polymorpha JPYU
- Metzgeria crassipilis NRWZ
- Odontoschisma prostratum YBQN
- Pallavicinia lyellii YFGP
- Pellia epiphylla PIUF
- Pellia neesiana JHFI
- Plagiochila asplenioides NWQC
- Porella navicularis KRUQ
- Porella pinnata UUHD
- Radula lindenbergia BNCU
- Riccia berychiana WJLO
- Scapania nemorosa IRBN
- Schistochila sp. LGOW
- Sphaerocarpos texanus HERT

Mosses
- Anomodon attenuatus QMWB
- Anomodon rostratus VBMM
- Atrichum angustatum ZTHV
- Aulacomnium heterostichum WNGH
- Bryum argenteum JMXW
- Buxbaumia aphylla HRWG
- Ceratodon purpureus FFPD
- Dicranum scoparium NGTD
- Funaria sp. XWHK
- Hedwigia ciliata YWNF
- Hypnum subimponens LNSF
- Leucobryum albidum VMXJ
- Leucobryum glaucum RGKI
- Leucodon sciuroides ZACW
- Neckera douglasii TMAJ
- Orthotrichum lyellii CMEQ
- Philonotis fontana ORKS
- Physcomitrella patens*
- Physcomitrium pyrimorfe YEPO
- Plagiomnium insigne BGXB
- Polytrichum commune SZYG
- Pseudotaxiphyllum elegans QKQO
- Racomitrium varium RDOO
- Rhynchostegium serrulatum JADL
- Rhytidiadelphus loreus WSPM
- Scouleria aquatica BPSG
- Schwetschkeopsis fabronia IGUH
- Sphagnum lescurii GOWD
- Sphagnum palustre RCBT
- Sphagnum recurvatum UHLI
- Syntrichia princeps GRKU
- Takakia lepidozioides SKQD
- Thuidium delicatulum EEMJ
- Timmia austriaca ZQRI

Charophytes
- Bambusina borreri QWFV
- Chaetosphaeridium globosum DRGY
- Chara vulgaris MWXT
- Chlorokybus atmophyticus AZZW
- Closterium lunula DRFX
- Coleochaete irregularis QPDY
- Coleochaete scutata VQBJ
- Cosmarium broomei HIDG
- Cosmarium granatum MNNM
- Cosmarium ochthodes HJVM
- Cosmarium ochthodes STKJ
- Cosmarium subtumidum WDGV
- Cosmarium tinctum BHBK
- Cosmocladium cf. constrictum RQFE
- Cylindrocystis brebissonii YOXI
- Cylindrocystis brebissonii RPGL
- Cylindrocystis cushleckae JOJQ
- Cylindrocystis sp. VAZE
- Desmidium aptogonum DFDS
- Entransia fimbriata BFIK
- Euastrum affine GYRP
- Gonatozygon kinahanii KEYW
- Interfilum paradoxum FPCO
- Klebsormidium subtile FQLP
- Mesostigma viride KYIO
- Mesotaenium braunii WSJO
- Mesotaenium caldariorum HKZW
- Mesotaenium endlicherianum WDCW
- Mesotaenium kramstei NBYP
- Micrasterias fimbriata MCHJ
- Mougeotia sp. ZRMT
- Netrium digitus FFGR
- Nucleotaenium eifelense KMNX
- Onychonema laeve GGWH
- Penium exiguum YSQT
- Penium margaritaceum AEKF
- Phymatodocis nordstedtiana RPQV
- Planotaenium ohtanii SNOX
- Pleurotaenium trabecula MOYY
- Roya obtusa XRTZ
- Spirogyra sp. HAOX
- Spirotaenia minuta NNHQ
- Spirotaenia sp. TPHT
- Staurastrum sebaldi ISHC
- Staurodesmus convergens WCQU
- Staurodesmus omearii RPRU
- Xanthidium antilopaeum GBGT
- Zygnemopsis sp. MFZO

Trebouxiophyceae
- Botryococcus braunii ETGN
- Botryococcus sudeticus VJDZ
- Botryococcus terribilis QYXY
- Chlorella minutissima MWAN
- Coccomyxa pringsheimii GXBM
- Eremosphaera viridis MNCB
- Geminella sp. PFUD
- Leptosira obovata ZNUM
- Microthamnion kuetzigianum DXNY
- Nannochloris atomus MFYC
- Parachlorella kessleri AKCR
- Pedinomonas minor RRSV
- Pedinomonas tuberculata PUAN
- Prasiola crispa WCLV
- Prototheca wickerhamii BILC
- Stichococcus bacillaris WXRI
- Trebouxia arboricola NKXU

Chlorophyceae
- Ankistrodesmus sp. OTQG
- Aphanochaete repens IJMT
- Asteromonas gracilis NTLE
- Brachiomonas submarina GUBD
- Carteria crucifera VIAU
- Carteria obtusa RUIF
- Chaetopeltis orbicularis BAZF
- Chlamydomonas reinhardtii*
- Chlamydomonas bilatus MULF
- Chlamydomonas cribrum BCYF
- Chlamydomonas moewusii JRGZ
- Chlamydomonas noctigama VALZ
- Chlamydomonas sp. TSBQ
- Chlamydomonas sp. AOUJ
- Chloromonas oogama IHOI
- Chloromonas perforata QRTH
- Chloromonas reticulata LBRP
- Chloromonas rosae AJUW
- Chloromonas subdivisa GFUR
- Chloromonas tughillensis UTRE
- Chlorosarcinopsis halophila KSFK
- Dunaliella salina RHVC
- Dunaliella tertiolecta ZDIZ
- Eudorina elegans RNAT
- Fritschiella tuberosa VFIV
- Golenkinia longispicula BZSH
- Gonium pectorale KUJU
- Haematococcus pluvialis ODXI
- Haematococcus pluvialis AGIO
- Hafniomonas reticulata FXHG
- Helicodictyon planctonicum AJAU
- Heterochlamydomonas inaequalis IRYH
- Lobochlamys segnis OFUE
- Lobomonas rostrata JKKI
- Microspora cf. tumidula FOYQ
- Neochloris oleoabundans EEJO
- Neochloris sp. GJIY
- Neochlorosarcina sp. USIX
- Oedogonium cardiacum DVYE
- Oedogonium foveolatum SDPC
- Oogamochlamys gigantea XDLL
- Pandorina morum RYJX
- Pediastrum duplex XKWQ
- Pediastrum duplex XTON
- Phacotus lenticularis ZIVZ
- Pirula salina NQYP
- Pleurastrum insigne PRIQ
- Pteromonas angulosa LNIL
- Pteromonas sp. ACRY
- Scenedesmus dimorphus PZIF
- Scourfieldia sp. EGNB
- Spermatozopsis exsultans MXDS
- Spermatozopsis similis ENAU
- Stephanosphaera pluvialis ZLQE
- Stigeoclonium helveticum JMUI
- Uronema sp. ISGT
- Uronema belkae RAWF
- Vitreochlamys sp. QWRA
- Volvox carteri*
- Volvox aureus JWGT
- Volvox aureus WRSL
- Volvox globator ISPU

Ulvophyceae
- Acrosiphonia sp. JIWJ
- Blastophysa cf. rhizopus VHIJ
- Bolbocoleon piliferum LSHT
- Bryopsis plumosa JTIG
- Cephaleuros virescens YDCQ
- Cladophora glomerata VBLH
- Codium fragile GYBH
- Cylindrocapsa geminella DZPJ
- Entocladia endozoica OQON
- Halochlorococcum marinum ALZF
- Ignatius tetrasporus KADG
- Ochlochaete sp. CQQP
- Oltmannsiellopsis viridis PZBH
- Oltmannsiellopsis viridis QJYX
- Percursaria percursa OAEZ
- Planophila laetevirens CBNG
- Planophila terrestris LETF
- Trentepohlia annulata NATT

Prasinophytes
- Bathycoccus prasinos MCPK
- Cymbomonas sp. XIVI
- Dolichomastix tenuilepis XOAL
- Mantoniella squamata QXSZ
- Micromonas pusilla*
- Monomastix opisthostigma BTFM
- Nephroselmis olivacea MMKU
- Nephroselmis pyriformis ISIM
- Ostreococcus tauri*
- Ostreococcus lucimarinus*
- Picocystis salinarum TGNL
- Prasinococcus capsulatus XMCL
- Prasinoderma coloniale HYHN
- Pseudoscourfieldia marina JMTE
- Pycnococcus provasolii MXEZ
- Pyramimonas parkeae TNAW
- Scherffelia dubia FMVB
- Tetraselmis chui HVNO
- Tetraselmis cordiformis DUMA
- Tetraselmis striata HHXJ
- coccoid prasinophyte XJGM

Glaucophyta
- Cyanophora paradoxa QFND
- Cyanophora paradoxa YTYU
- Cyanoptyche gloeocystis JKHA
- Glaucocystis cf. nostochinearum POOW
- Gloeochaete wittrockiana PQED

Red Algae
- Betaphycus gelatinae BWVJ
- Ceramium kondoi VZWX
- Chondrus crispus UGPM
- Chroodactylon ornatum LLXJ
- Dumontia simplex IEHF
- Eucheuma denticulatum JEBK
- Glaucosphaera vacuolata RSOF
- Gloeopeltis furcata SBLT
- Gracilaria asiatica VNAL
- Gracilaria blodgettii LJPN
- Gracilaria chouae FTRP
- Gracilaria lemaneiformi IKWM
- Grateloupia filicina ZJOJ
- Grateloupia livida IKIZ
- Grateloupia turuturu URSB
- Grateloupia chiangii PWKQ
- Gymnogongrus ftabelliformis CKXF
- Heterosiphonia pulchra YSBD
- Kappaphycus alvarezii IHJY
- Mazzaella japonica WEJN
- Polysiphonia japonica XAXW
- Porphyra yezoensis ZULJ
- Porphyridium cruentum OBUY
- Porphyridium purpureum PVGP
- Rhodella violacea RTLC
- Rhodochaete parvula JJZR
- Sinotubimorpha guangdongensis PYDB

Chromista
- Chroomonas sp. ROZZ
- Colpomenia sinuosa QLMZ
- Cryptomonas curvata BAKF
- Desmarestia viridis FSQE
- Dictyopteris undulata LIRF
- Hemiselmis virescens MJMQ
- Ishige okamurai APTP
- Isochrysis sp. BAJW
- Kjellmaniella crassifolia RAPY
- Laminaria japonica OGZM
- Laminaria japonica QDTV
- Mallomonas sp. BOGT
- Nannochloropsis oculata JQFK
- Ochromonas sp. EBWI
- Pavlova lutheri LLEN
- Pavlova lutheri NMAK
- Pavlova lutheri RFAD
- Petalonia fascia VRGZ
- Prorocentrum micans TZJQ
- Proteomonas sulcata IRZA
- Prymnesium parvum LXRN
- Punctaria latifolia ASZK
- Rhodomonas sp. IAYV
- Sargassum horneri RWXW
- Sargassum thunbergii YRMA
- Sargassum vachellianum HFIK
- Sargassum fusiforme LDRY
- Sargassum hemiphyllum VYER
- Sargassum henslowianum FIKG
- Sargassum integerrimum FOMH
- Sargassum muticum JGGD
- Scytosiphon lomentaria JCXF
- Scytosiphon dotyo ULXR
- Symphyocladia latiuscula UYFR
- Synura petersenii DBYD
- Synura sp. VKVG
- Undaria pinnatifida FIDQ

Excavata
- Euglena sp. UNBZ

[1]P.G. Wolf 923 (UTC); Norwich, UK.

[2]D.G. Long s.n. (OXF); Edinburgh, UK

Table S2. The calibrations used in dating the divergences within the phototropin gene family.

No.	Clade	Calibration	Date (MYA)	Prior	Reference	Justification
1	Tracheophyta	*Zosterophyllum sp.*	416	lognormal (mean: 3.5, STD: 1, offset: 416)	70, 74	Oldest unequivocal record of total group of lycoposid; see Ref. 72 for detailed justifications
2	Euphyllophyta	*Ibyka sp.*	388.2	lognormal (mean: 3.5, STD: 1, offset: 388.2)	70, 75	Oldest unequivocal record of monilophyte based on protoxylem morphology; see Ref. 72 for detailed justifications
3	Bryophyta PHOTA	"type III" fragment	330.9-346.7	lognormal (mean: 1.5, STD: 1, offset: 330.9)	71	Oldest unequivocal record of crown Bryophyta; the fibrils and pores similar to those of *Sphagnum* water-storage cells
4	Bryophyta PHOTB	"type III" fragment	330.9-346.7	lognormal (mean: 1.5, STD: 1, offset: 330.9)	71	Oldest unequivocal record of crown Bryophyta; the fibrils and pores similar to those of *Sphagnum* water-storage cells
5	Jungermanniopsida	*Riccardiothallus devonicus*	407-411	lognormal (mean: 1.5, STD: 1, offset: 407)	72	Oldest unequivocal record of crown Jungermanniopsida; gross morphology similar to the extant *Riccardia* species
6	Polypodiopsida PHOT1	*Rastropteris pirtgquanensis*	296	lognormal (mean: 1.5, STD: 1, offset: 296)	8, 76	Oldest unequivocal record of Osmundaceae stem; see Ref. 8 for detailed justifications
7	Polypodiopsida PHOT2	*Rastropteris pirtgquanensis*	296	lognormal (mean: 1.5, STD: 1, offset: 296)	8, 76	Oldest unequivocal record of Osmundaceae stem; see Ref. 8 for detailed justifications
8	Eupolypod PHOT1	imported secondary date	116.7	Normal (mean: 116.7, STD: 35.01)	8	A well-established time estimate for the divergence of Eupolypods
9	Eupolypod PHOT2	imported secondary date	116.7	Normal (mean: 116.7, STD: 35.01)	8	A well-established time estimate for the divergence of Eupolypods
10	Spermatophyta PHOT1	*Cordaixylon iowensis*	306.2	lognormal (mean: 2.5, STD: 1, offset: 306.2)	70, 77	Oldest unequivocal record of Acrogymnospermae; see Ref. 72 for detailed justifications
11	Spermatophyta PHOT2	*Cordaixylon iowensis*	306.2	lognormal (mean: 2.5, STD: 1, offset: 306.2)	70, 77	Oldest unequivocal record of Acrogymnospermae; see Ref. 72 for detailed justifications
12	Grass PHOT1	phytoliths in dinosaur coprolites	65-67	lognormal (mean: 1.5, STD: 1, offset: 65)	73	Oldest unequivocal record of PACMAD or BEP of grass; phytoliths morphology similar to subclades in PACMAD or in BEP
13	Grass PHOT2	phytoliths in dinosaur coprolites	65-67	lognormal (mean: 1.5, STD: 1, offset: 65)	73	Oldest unequivocal record of PACMAD or BEP of grass; phytoliths morphology similar to subclades in PACMAD or in BEP
14	Coniferae PHOT1	*Araucaria mirabilis*	147	lognormal (mean: 1.5, STD: 1, offset: 147)	70	Oldest unequivocal record of Cupressophyta crown; see Ref. 72 for detailed justifications
15	Coniferae PHOT2	*Araucaria mirabilis*	147	lognormal (mean: 1.5, STD: 1, offset: 147)	70	Oldest unequivocal record of Cupressophyta crown; see Ref. 72 for detailed justifications

Table S3. The primers and PCR protocols used in this study. See Table S4 for primer sequences.

Taxa	Gene	Primary PCR primers	Secondary PCR primers[1]	PCR program[2]	Specimen voucher
Hornworts:					
Phymatoceros phymatodes	neochrome	neoF65 + neoR2818	neoF430 + neoR2776	a/a	J. Pittermann s.n. (DUKE)
Phymatoceros phymatodes	neochrome	neoF65 + neoR4110	neoF2367 + neoR3456	a/a	J. Pittermann s.n. (DUKE)
Phymatoceros phymatodes	neochrome	neoF65 + neoR4110	neoF3230 + neoR4110	a/a	J. Pittermann s.n. (DUKE)
Phymatoceros phymatodes	phototropin	neoF65 + neoR4110	neoF2367 + neoR3456	a/a	J. Pittermann s.n. (DUKE)
Phymatoceros phymatodes	phototropin	photF1856 + photR2508	photF1970 + photR2245	a/a	J. Pittermann s.n. (DUKE)
Phymatoceros phymatodes	phototropin	photF2774 + photR4339	-	a	J. Pittermann s.n. (DUKE)
Megaceros flagellaris	neochrome	neoF65 + neoR902	-	b	B. Crandall-Stotler s.n. (ABSH)
Megaceros flagellaris	neochrome	neoF649 + neoR1950	-	b	B. Crandall-Stotler s.n. (ABSH)
Megaceros flagellaris	neochrome	neoF1844 + neoR2361	-	c	B. Crandall-Stotler s.n. (ABSH)
Megaceros flagellaris	neochrome	neoF2239 + neoR3300	-	b	B. Crandall-Stotler s.n. (ABSH)
Megaceros flagellaris	neochrome	neoF2361 + neoR4110	-	c	B. Crandall-Stotler s.n. (ABSH)
Megaceros flagellaris	phototropin	photF1856 + photR4339	photF1970 + photR4339	a/a	B. Crandall-Stotler s.n. (ABSH)
Nothoceros aenigmaticus	neochrome	F5 + R1_T1	F565 + R1_T1	d/d	F.W. Li 1291 (DUKE)
Nothoceros aenigmaticus	neochrome[3]	neoF4018 + AP1	neoF4110 + AP2	e/f	F.W. Li 1569 (DUKE)
Nothoceros aenigmaticus	neochrome[3]	neoR429 + AP1	R3re_phyN + AP2	e/f	F.W. Li 1569 (DUKE)
Nothoceros aenigmaticus	neochrome[3]	NaNEO_3-1_GM1 + AP1	NaNEO_3-1_GM2 + AP2	e/f	F.W. Li 1569 (DUKE)
Nothoceros aenigmaticus	phototropin	5upF1 + R7	5upF2 + R7	c/g	F.W. Li 1291 (DUKE)
Nothoceros aenigmaticus	phototropin[4]	F565 + I_R1	I_F2 + I_R2	d/d	F.W. Li 1291 (DUKE)
Nothoceros aenigmaticus	phytochrome	F-200_Maphy + R4850_Maphy	F-3_Maphy + R4450_Maphy	h/h	F.W. Li 1291 (DUKE)
Phaeoceros carolinianus	neochrome	neoF65 + neoR877	-	c	B. Crandall-Stotler s.n. (ABSH)
Phaeoceros carolinianus	neochrome	neoF649 + neoR1950	-	b	B. Crandall-Stotler s.n. (ABSH)
Phaeoceros carolinianus	neochrome	neoF1576 + neoR4104	-	a	B. Crandall-Stotler s.n. (ABSH)
Anthoceros punctatus	neochrome	neoF67 + neoR832	-	c	D. Chamberlain s.n. (E)
Anthoceros punctatus	neochrome	neoF428 + neoR3049	neoF812 + neoR2938	a/g	D. Chamberlain s.n. (E)
Anthoceros punctatus	neochrome	neoF2938 + neoR4104-2	neoF3049 + neoR4104-2	a/a	D. Chamberlain s.n. (E)
Ferns:					
Adiantum andicola	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	C.J. Rothfels 2641, DB5549[5] (DUKE)
Adiantum hispidulum	neochrome	neoF58 + neoR4238	neoF651 + neoR3718	c/c	L. Huiet s.n., DB9529 (DUKE)
Adiantum hispidulum	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	L. Huiet s.n., DB9529 (DUKE)
Adiantum pedatum	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	C.J.Rothfels 3839, DB7517 (DUKE)
Adiantum pedatum	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	C.J.Rothfels 3839, DB7517 (DUKE)
Adiantum tetraphyllum	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	L. Huiet 105, DB2505 (UC)
Adiantum tetraphyllum	neochrome	neoF20 + neoR4242	neoF20 + neoR2236	i/c	L. Huiet 105, DB2505 (UC)
Adiantum tetraphyllum	neochrome	neoF1108 + neoR3065	-	k	L. Huiet 105, DB2505 (UC)
Alsophila podophylla	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/l	E. Schuettpelz 1201A, DB4948 (DUKE)
Alsophila podophylla	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/l	E. Schuettpelz 1201A, DB4948 (DUKE)
Alsophila podophylla	neochrome	neoF20 + neoR4242	neoF538 + neoR4000	i/l	E. Schuettpelz 1201A, DB4948 (DUKE)
Bolbitis auriculata	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	F. Rakotondrainibe, DB3504 (P)
Dennstaedtia punctilobula	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	C.J. Rothfels 4167, DB8975 (DUKE)
Dennstaedtia punctilobula	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/l	C.J. Rothfels 4167, DB8975 (DUKE)
Dennstaedtia punctilobula	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/l	C.J. Rothfels 4167, DB8975 (DUKE)
Deparia acrostichoides	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	C.J. Rothfels 3894, DB7797 (DUKE)
Deparia acrostichoides	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	C.J. Rothfels 3894, DB7797 (DUKE)
Deparia acrostichoides	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	C.J. Rothfels 3894, DB7797 (DUKE)
Deparia lancea	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	E. Schuettpelz 298, DB2558 (DUKE)
Deparia lancea	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	E. Schuettpelz 298, DB2558 (DUKE)
Didymochlaena truncatula	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	E. Schuettpelz 267, DB2435 (DULE)
Diplazium bombonasae	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	R.C. Moran 7493, DB3764 (DUKE)
Doodia media	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	E. Schuettpelz 295, DB2555 (DUKE)
Dryopteris amurensis	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	A. Uchida 1392, DB7982 (TNS)
Dryopteris amurensis	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	A. Uchida 1392, DB7982 (TNS)
Dryopteris expansa	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	A. Ebihara TH2007-507, DB7977 (TNS)
Dryopteris expansa	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	A. Ebihara TH2007-507, DB7977 (TNS)
Hemidictyum marginatum	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/j	M. Christenhusz 2476, DB3054 (CAY)
Hemidictyum marginatum	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/j	M. Christenhusz 2476, DB3054 (CAY)
Hemidictyum marginatum	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/j	M. Christenhusz 2476, DB3054 (CAY)
Hemidictyum marginatum	neochrome	neoF1108 + neoR3065	-	k	M. Christenhusz 2476, DB3054 (CAY)
Hypolepis tenuifolia	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	E. Schuettpelz 286, DB2574 (DUKE)
Macrothelypteris torresiana	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	Schuettpelz 335, DB2980 (DUKE)
Macrothelypteris torresiana	neochrome	neoF20 + neoR4242	neoF20 + neoR2336	i/c	Schuettpelz 335, DB2980 (DUKE)
Macrothelypteris torresiana	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	Schuettpelz 335, DB2980 (DUKE)
Matteuccia struthiopteris	neochrome	neoF20 + neoR786	-	b	A. Larsson 258, DB7946 (DUKE)
Matteuccia struthiopteris	neochrome	neoF649 + neoR1950	-	b	A. Larsson 258, DB7946 (DUKE)
Matteuccia struthiopteris	neochrome	neoF1530 + neoR2300	-	m	A. Larsson 258, DB7946 (DUKE)
Matteuccia struthiopteris	neochrome	neoF2239 + neoR3300	-	m	A. Larsson 258, DB7946 (DUKE)
Matteuccia struthiopteris	neochrome	neoF2935 + neoR3720	-	m	A. Larsson 258, DB7946 (DUKE)
Matteuccia struthiopteris	neochrome	neoF58 + neoR4238	neoF651 + neoR3718	c/c	A. Larsson 258, DB7946 (DUKE)
Onoclea sensilibis	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	E. Schuettpelz 353, DB2998 (DUKE)
Onoclea sensilibis	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	E. Schuettpelz 353, DB2998 (DUKE)
Phegopteris hexagonoptera	neochrome	neoF20 + neoR4242	neoF651 + neoR3718	i/c	M. Christenhusz 3844, DB2731 (TUR)
Phegopteris hexagonoptera	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/c	M. Christenhusz 3844, DB2731 (TUR)
Plagiogyria formosana	neochrome	neoF20 + neoR786	-	b	E. Schuettpelz 1083A, DB4826 (DUKE)
Plagiogyria formosana	neochrome	neoF649 + neoR1950	-	b	E. Schuettpelz 1083A, DB4826 (DUKE)
Plagiogyria formosana	neochrome	neoF1530 + neoR2300	-	m	E. Schuettpelz 1083A, DB4826 (DUKE)
Plagiogyria formosana	neochrome	neoF2935 + neoR3720	-	m	E. Schuettpelz 1083A, DB4826 (DUKE)
Tectaria zeylanica	neochrome	neoF20 + neoR4242	-	i	E. Schuettpelz 514, DB3569 (GOET)
Thelypteris noveboracensis	neochrome	neoF20 + neoR4242	neoF2115 + neoR4242	i/l	C.J. Rothfels 4164, DB8972 (DUKE)

[1]The primer pair for secondary PCR in nested PCR reaction. "-" indicates no nested PCR was conducted.

[2]The PCR program used (primary PCR/secondary PCR, if nested PCR was used).

[3]Genome walking using Clontech Genome Walker kit.

[4]Genome walking using Inverse PCR.

[5]Fern DNA Database number (http://fernlab.biology.duke.edu).

a 98:30s, (98:10s, 70:30s,72:90s)x35, 72:600s
b 98:30s, (98:10s, 60:30s,72:90s)x35, 72:600s
c 98:30s, (98:10s, 72:120s)x35, 72:600s
d 98:30s, (98:10s, 72:180s)x35, 72:600s
e (94:25s, 72:180s)x7, (94:25s, 67:180s)x32, 67:420s
f (94:25s, 72:180s)x5, (94:25s, 67:180s)x20, 67:420s
g 98:30s, (98:10s, 67:30s,72:90s)x35, 72:600s
h 98:30s, (98:10s, 67:30s,72:150s)x35, 72:600s
i 98:30s, (98:10s, 68:30s,72:120s)x35, 72:600s
j 94:300s, (94:60s, 60:60,72:120s)x35, 72:600s
k 94:300s, (94:60s, 56:60,72:240s)x35, 72:600s
l 98:30s, (98:10s, 70:30s,72:120s)x35, 72:600s
m 98:30s, (98:10s, 55:30s,72:90s)x35, 72:600s

Table S4. The primer sequences.

Primer	Sequence (5'-3')
5upF1	ATTCACAAATGTTGCCCGATGTGC
5upF2	CTGCACTCCTACTCGTTACCG
AP1	GTAATACGACTCACTATAGGGC
AP2	ACTATAGGGCACGCGTGGT
F-200_Maphy	AGCGTGTAGCCTTGTCCTGTAC
F-3_Maphy	GCGACAGCGGCAAAGTTGAAG
F5	GCGGCAGGCTGCTCAACTACAG
F565	TACACCGAAGGCTACAAGGCTAATG
I_F2	CAAGTGCAATCCAATGATGCCGC
I_R1	TTCTGTAGTTGAGCAGCCTGCC
I_R2	GAGGAGTAGCCGGTCATGGTGAAG
NaNEO_3-1_GM1	TGTGGAACAAAGGCAACTTGGGACGAA
NaNEO_3-1_GM2	ATGTGAAGCCTCAAGCAAATGTTACAAGT
neoF1108	GTGCAGCTCAACATKGAGCTGGA
neoF1530	TCBTRTTTTGGTTYAGGTCRCAYACTGC
neoF1576	CTGGACAGGGACGACGACTCTCG
neoF1844	CATTGAGGGACAAGGAGGAGTTACCAGG
neoF20	CCAAGACGAAGCACAGCGTG
neoF2115	GGAGGTGATTGGAGSCAACTGC
neoF2239	AGGAAAGATGGYAGCWRYTTYTGGAA
neoF2300	GCTRGAGGTDASCAAGTACACDGAGGG
neoF2361	CGGCACCAGGACAAGGTTTCTG
neoF2367	CAGTCSCTCATCAAGTACGAYGT
neoF2935	GTKCAGCTYATCCGAGATGCAGT
neoF2938	CTGTCCCTGGAGATCGTGAACTACACC
neoF3049	CAACAGAAGGTGGCGGATTATGTTCC
neoF3230	CAGACCATCTATGGGTGCGGCATTC
neoF4018	ATCTTGCTCTACGAGATGCTCTATGGC
neoF4110	TACATTCCCAACCAGCATCCCAGTGAG
neoF428	GYACGGATSTGCGGATGCTCTTCAC
neoF430	ACGGATSTGCGGATGCTCTTCAC
neoF58	AGBGCNGATGCMAGRCTYCATGC
neoF649	GATCGDGTGATGGCBTACAARTTYCA
neoF649	GATCGDGTGATGGCBTACAARTTYCA
neoF65	ATGCGAGGCTKAATGCGGCGTTTGAG
neoF651	TCGGGTGATGGCCTACAAGTTCCA
neoF67	GCGAGGCTGMATGCGGYGTTYGAG
neoF812	ACAAGTTYCAYGAGGACGAGCACG
neoR1950	CCYCGAAYNGCYTCCATCCAYTCCTG
neoR2236	AGAAGYTGCTGCCRTCTTTCCTGTA
neoR2361	CRGAAACCTTGTCCTGGTGCCG
neoR2776	GCGAAGATGATGGGGTTGTCCG
neoR2818	GCACCTCCTCCCTGCTGTACTCTGTCAG
neoR2938	GGTGTAGTTCACGATCTCCAGGGACAG
neoR3049	GGAACATAATCCGCCACCTTCTGTTG
neoR3065	CTGHACTCCGATGAAGTACTGGA
neoR3300	GYARCCTSGGATCTGWGATCAC
neoR3456	AGCATCATSGCCTTGTCCATG
neoR3718	TGACVCCCATGCAGTGGAGGTACTC
neoR3720	GTTCTCBGGCTTSAGRTCBCGGTAGATG
neoR4104	ATGCTGGTKGGGAATGTRAGCTCCTTG
neoR4104-2	AYGCTGSTSGGGAAKGTGAGCTCCTTG
neoR4110	AGGCTCACTGGGATGCTGGTTGGG
neoR4238	CGGATRAGAGGCCAGTYGATKYCTYGGA
neoR4242	CGGATGAGAGGCCAGTCGATKYCT
neoR429	GAGTGAACAGCATCCGCACATCCGTG
neoR786	GGTARTGCARGCCVAGRTAHGGCTCC
neoR832	GAGGCTGATCGGCTGGTGGAGC
neoR877	ATGTACTGCGTGTGGCAACCGTGC
neoR902	GACGAGACGGAGCCCATGTTGC
photF1856	CTGGTGSTCAAGGAGGAGCTGG
photF1970	GCTCTCCWCCTTCCAGCAGACG
photF2645	CTTCGCCTCYGACCAYTTCCTGG
photF2774	GGAGAGACGGGACATCACTGTGC
photR2508	AGCAGCGACAGAAATCCCGAGGAC
photR2901	GCTCGTACTCGCTSCCRTCCAG
photR4102	ATGCTGSTSGGRAATGTGAGCTCCTTGTT
photR4339	TCYKCCTCGTCCCACTCCAGRTC
R1_T1	ACCCAGGATCAAAACACATCGCTG
R3re	GACGCATTCTCGCTCATTGCCAGGAT
R4450_Maphy	CCATCCACCACAGGTTCTGAACAC
R4850_Maphy	AAAATGTCCAGGACCGTCAGGTTC
R7	AGAGTGGTGGCCAAGTCAATTCC