Seeing the Light: the Origin and Evolution of Plant Photoreceptors

by

Fay-Wei Li

Department of Biology
Duke University

Date:_____
Approved:

_____
Kathleen M. Pryer, Supervisor

_____
Meng Chen

_____
Sönke Johnsen

_____
Corbin Jones

_____
Sarah Mathews

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2015

ABSTRACT

Seeing the Light: the Origin and Evolution of Plant Photoreceptors

by

Fay-Wei Li

Department of Biology
Duke University

Date:_____
Approved:

_____
Kathleen M. Pryer, Supervisor

_____
Meng Chen

_____
Sönke Johnsen

_____
Corbin Jones

_____
Sarah Mathews

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2015

Abstract

Plants use an array of photoreceptors to measure the quality, quantity, and direction of light in order to respond to ever-changing light environments. Photoreceptors not only determine how and when individual plants complete their life cycles, but they also have a profound and long-term macroevolutionary influence on species diversification. Despite their significances, very little is known about photoreceptors across plants as whole, and we lack a comprehensive view of photoreceptor evolution.

In my dissertation, I investigate the origin and evolution of three of the most prominent photoreceptor gene families in plants: **phytochromes**, **phototropins** and **neochromes**. Using newly available transcriptomic and genomic data, I completed the first in-depth survey of these photoreceptor families across land plants, green algae, red algae, glaucophytes, cryptophytes, haptophytes, and stramenopiles.

Phytochromes are red/far-red photoreceptors that play essential roles in seed germination, seedling photomorphogenesis, shade-avoidance, dormancy, circadian rhythm, phototropism, and flowering. Here, I show that the canonical plant phytochromes originated in a common ancestor of streptophytes (charophyte green algae plus land plants), and I identify the most likely sequence whereby the plant phytochrome structure evolved from its ancestral phytochrome. Phytochromes in charophyte algae are structurally diverse, including canonical and non-canonical forms, whereas in land plants, phytochrome structure is highly conserved. Liverworts, hornworts, and *Selaginella* apparently possess a single phytochrome gene copy, whereas independent gene duplications occurred within mosses, lycopods, ferns, and seed plants, leading to diverse phytochrome families in these clades. My detailed phylogeny

encompasses all of green plants and enables me to not only uncover new phytochrome lineages, but also to make links to our current understanding of phytochrome function in *Arabidopsis* and *Physcomitrella* (the major model organism outside of flowering plants). Based on this robust evolutionary framework, I propose new hypotheses and discuss future directions to study phytochrome mechanisms.

Phototropins are blue-light photoreceptors that regulate key adaptive physiological responses, including shoot-positive phototropism, root-negative phototropism, chloroplast accumulation/avoidance, stomatal opening, circadian rhythm, leaf expansion, and seedling elongation I show that phototropins originated in the common ancestor of Viridiplantae (all green algae [charophytes, chlorophytes, prasinophytes] plus land plants). Phototropins repeatedly underwent independent duplications in all major plant lineages (mosses, lycopods, ferns and seed plants), except for liverworts and hornworts, where phototropin is a single-copy gene. Following each major duplication event, phototropins subsequently differentiated in parallel, resulting in two specialized (yet partially overlapping) functional forms that primarily mediate either low- or high-light responses. My gene phylogeny further suggests that phototropins have co-evolved with phytochromes, as is evident from their molecular interactions and strikingly similar gene duplication patterns. I hypothesize that the co-evolution of phototropins with phytochromes, together with their subsequent convergent functional divergences in phototropic responses, contributed to the success of plants in adapting to diverse and heterogeneous habitats.

Neochromes are chimeric photoreceptors that, by fusing phytochrome and phototropin modules into a single protein, are able to use both red/far-red and blue light to modulate

phototropic responses. Neochromes were first discovered in ferns, and the evolution of neochromes was implicated as a key innovation that facilitated fern diversification under the low-light angiosperm canopies. Despite its significance from an evolutionary standpoint, the origin of neochromes has remained a mystery. Here I present the first 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 my 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 *Anthoceros punctatus* hornwort, I also discovered a novel phototropin gene that likely represents the ancestral lineage of the neochrome phototropin module. Thus, a neochrome originating in hornworts was horizontally transferred to ferns, where it may have played a significant role in the diversification of modern ferns.

In summary, my studies identified the molecular origins of phytochromes, phototropins and neochromes, and reconstructed their respective evolutionary histories. This new framework for photoreceptor evolution will stimulate new research linking ecology, evolution, and photochemistry to understand how plants adapt to variable light environments.

To my parents who showed me the wonder of ferns…

Contents

List of Tables

List of Figures

Acknowledgements

When I looked back the past five years, I realized how my graduate study would be impossible without the generous help from so many people. I would like to begin by thanking Carl Rothfels, who got me involved in the One Thousand Plant (1KP) project, from where I got most of my transcriptome data *for free* by, literally, writing a few lines of batch downloading scripts. I am indebted to all the 1KP contributors, and of course to the visionary leader Gane Ka-Shu Wong. I also want to thank Steve Kelly, Eftychios Frangedakis and Jane Langdale for sharing their hornwort genomic data, and to Josh Der for sharing his *Pteridium* transcriptome, both of which have been instrumental to my study.

Juan Carlos Villareal has been my best "hornwort buddy", who has shared every bit of his incredible hornwort knowledge with me. Dave Swofford taught me how to make phylogeny trees and the theories behind it, and saved my ass many times when we taught the phylogenetics course. Tom Mitchell-Olds introduced me to the world of Python, a language I cannot live without. Jon Shaw gave me a crash course on bryophyte phylogeny, so that I can pretend I know something when discussing gene duplications in mosses. Paul Manos offered me the opportunity to organize Duke Systematics Discussion Group (i.e. SDG), during which I got to know some incredible evolutionary biologists. Layne Huiet is a walking Wikipedia of molecular biology and gives the best tips ever in doing lab experiments. Karl Bates is the wizard of science communications, and made me feel like a "celebrity of ferns" for a day or two.

I appreciate the people, professional societies, and funding agencies that believed in my research potential and gave me the funding to pursue. These include: American Society of Plant Taxonomists, Duke Biology, National Science Foundation (DEB-1407158 and GRFP), Sigma Xi,

Introduction

"Light exerts a powerful influence on most vegetable tissues, and there can be no doubt that it generally tends to check their growth" – Charles Darwin, 1880

Light is the ultimate source of energy for much of life on earth, and inevitably governs the growth and physiology of photosynthetic organisms. Plants "see" light through photoreceptors (Figure 1). Five photoreceptor gene families are generally present in plants: phytochromes, phototropins, cryptochromes, Zeitlupes, and UVR8 (Möglich et al., 2010; Heijde and Ulm, 2012). Among them, **phytochromes** and **phototropins** are perhaps the most prominent, given their prevalence in controlling almost every aspect of the plant life cycle: from the dormancy and germination of seeds/spores, morphogenesis and growth, to flowering (Franklin and Quail, 2010; Christie, 2007). Ample evidence has shown that these two photoreceptors exert an adaptive significance on individual fitness (Galen et al., 2004), local adaptation (Ikeda et al., 2009; Ikeda and Setoguchi, 2010), and/or long-term macroevolutionary success (Mathews et al., 2003). Furthermore, their chimeric derivative **neochrome** (part phytochrome and part phototropin), was also implicated as a key innovation that facilitated fern radiation under low-light environments (Schneider et al., 2004; Schuettpelz and Pryer, 2009; Kawai et al., 2003).

To understand how plants adapted to, and thrived in, the diverse environments they inhabit, the roles of photoreceptors cannot be ignored. However, photoreceptor gene sequence data have been scarce and mostly limited to seed plants, thus impeding detailed reconstructions of photoreceptor evolutionary histories. The major goal of my dissertation research was to leverage recent genomic and transcriptomic data to investigate the origin and evolution of three photoreceptor families: **phytochromes** (Chapter 1), **phototropins** (Chapter 2), and **neochromes** (Chapter 3).

1

Chapter 3 on neochromes was published first (Li et al., 2014), along with preliminary data from Chapters 1 and 2. More comprehensive analyses on phytochromes (Chapter 1) and phototropins (Chapter 2) are being published in separate, dedicated papers.

Figure 1: Plants "see" light through photoreceptors. Domain structures of phytochromes, phototropins and neochromes are shown on the right. Neochrome is a chimeric photoreceptor fusing phytochrome and phototropin modules, and is able to respond to both blue and red/far-red light. Domain names: GAF (cGMP phosphodiesterase/adenylate cyclase/FhlA); KD (histidine-kinase-related-domain); PAS (Per/Arnt/Sim); PHY (Phytochrome); PKC (Protein Kinase C).

1. The origin and evolution of phytochromes

Li, F.-W., M. Melkonian, C.J. Rothfels, J.C. Villarreal, D.W. Stevenson, S.W. Graham, G.K.S. Wong, K.M. Pryer, and S. Mathews. **Novel phytochrome lineages and complex evolutionary histories revealed across extant plant diversity**. Nature Communications, *in review*

1.1 Introduction

Phytochromes are red/far-red light sensors, particularly prominent for their control of seed germination, seedling photomorphogenesis, shade-avoidance, dormancy, circadian rhythm, phototropism, and flowering (Möglich et al., 2010; Rockwell et al., 2006; Franklin and Quail, 2010). Because of their biological significance, phytochromes have been a major focus in plant research. Phytochrome photochemistry, function, and its associated signal transduction mechanisms have been investigated extensively, mostly using the model flowering plant *Arabidopsis thaliana* (Möglich et al., 2010; Rockwell et al., 2006; Franklin and Quail, 2010; Chen and Chory, 2011).

Canonical plant phytochromes comprise an N-terminal photosensory core module (PCM) and a C-terminal regulatory module (Franklin and Quail, 2010; Rockwell et al., 2006). The PCM contains three conserved domains in the linear sequence PAS, GAF, and PHY. It is essential for light reception and photoconversion between reversible conformations that absorb maximally in the red (650-670 nm) or far-red (705-740 nm) regions of the spectrum, referred to as Pr and Pfr, respectively. The C-terminal module consists of a PAS-PAS repeat followed by a histidine-kinase-related domain (HKRD). The HKRD resembles a histidine kinase domain but lacks the conserved histidine phosphorylation site, exhibiting serine/threonine kinase activity instead (Yeh and Lagarias, 1998; Fankhauser, 2000).

Plant phytochromes occur as a small nuclear-encoded gene family, and in seed plants they fall into three distinct clades: *PHYA*, *PHYB/E*, and *PHYC* (Mathews, 2010). The phylogenetic relationships among these clades are well resolved, allowing for the formulation of functional hypotheses for seed-plant phytochromes based on their orthology with *Arabidopsis* phytochromes (Mathews, 2010). Phytochrome diversity in non-seed plants, however, is very poorly understood, with the limited available data being derived from the *Physcomitrella* (moss) and *Selaginella* (lycophyte) genome projects (Banks et al., 2011; Rensing et al., 2008), and a few cloning studies (Schneider-Poetsch et al., 1994; Nozue et al., 1997; Okamoto et al., 1993; Pasentsis et al., 1998; Suzuki et al., 2001). The lack of a comprehensive phytochrome evolutionary framework for all land plants is an obstacle to understanding the evolution of phytochrome functional diversity, and makes it difficult, for example, to interpret correctly results from comparisons of function in *Arabidopsis thaliana* and *Physcomitrella patens*.

An especially remarkable plant phytochrome derivative is neochrome, a chimeric photoreceptor combining a phytochrome PCM and a blue light-sensing phototropin (Nozue et al., 1998). Neochromes have been detected only in zygnemetalean algae, ferns and hornworts (Suetsugu et al., 2005; Li et al., 2014). While it has been shown that the phototropin component of neochromes has two independent origins (one in zygnemetalean algae and the other in hornworts; see Chapter 3 and Li et al., 2014), the ancestry of the phytochrome portion remains unclear.

In addition to plants, phytochromes are present in prokaryotes, fungi, and several protistan and algal lineages (Ulijasz and Vierstra, 2011; Rockwell et al., 2014). These phytochromes share with canonical plant phytochromes the PCM domain architecture at the N-terminal, but they differ in their C-terminal regulatory modules. Prokaryotic and fungal phytochromes, for example, lack the PAS-PAS repeat, and have a functional histidine-kinase

domain with the conserved histidine residue. Recently, Rockwell et al. (2014) and Duanmu et al.

(2014) examined the phytochromes in several algal lineages (brown algae, cryptophytes,

glaucophytes, and prasinophytes), and discovered that some of them not only exhibit great

spectral diversity, but also have novel domain combinations within the C-terminal module.

Despite these important findings, phytochromes remain unreported from the majority of algal

lineages. Duanmu et al. (2014) proposed that the canonical plant phytochrome may have

originated among charophyte algae, but they were unable to confirm this.

In this study, I investigated newly available genomic and transcriptomic resources to

discover phytochrome homologs outside of seed plants. I examined a total of 300 genomes and

transcriptomes from ferns, lycophytes, bryophytes, charophytes, chlorophytes, and

prasinophytes (all in Viridiplantae), and from other plastid-bearing algal lineages, the

glaucophytes, cryptophytes, rhodophytes, haptophytes, and stramenopiles (Figure 2, Table 2). I

used these data to reconstruct the first detailed phytochrome phylogeny for the eukaryotic

branches of the tree of life, and to map all the major gene duplication events and domain

architecture transitions onto this evolutionary tree.

1.2 Results

I discovered a total of 350 phytochrome homologs in 148 transcriptome assemblies and

12 whole-genome sequences (Table 2, Table 3) spanning extant plant and algal diversity. In the

remaining 140 assemblies and genome sequences, I detected no phytochrome homologs. I

inferred a phytochrome phylogeny from an amino acid matrix that included the sequences I

discovered, together with previously published sequences from GenBank. To improve our

understanding of phytochrome and neochrome evolution, especially within ferns and

bryophytes, I also assembled three nucleotide matrices. The fern and bryophyte matrices

included 113 and 97 phytochrome sequences, respectively. The neochrome matrix included 16

neochromes and 95 phytochromes from selected bryophytes and charophytes.

Figure 2: The organismal lineages screened for phytochrome homologs. The phylogenetic relationships were derived from Wickett et al. (2014) and Marin (2012). Lineages that lack phytochrome are in grey. *Traditional Archaeplastida does not include cryptophytes.

The topologies of the phytochrome gene trees correspond well with published

organismal relationships (Wickett et al., 2014; Kuo et al., 2011; Cox et al., 2010; Gontcharov and

Melkonian, 2010; Cavalier-Smith et al., 2014; Burki et al., 2012; Grant and Katz, 2014; Marin, 2012;

Villarreal and Renner, 2012; Forrest et al., 2006), allowing me to pinpoint the phylogenetic

positions of gene duplication events and delineate novel phytochrome clades. Below I report

results on phytochrome diversity, phylogenetic structure, and domain architecture in the

stramenopiles, cryptophytes and Archaeplastida (or "Plantae": red algae + glaucophytes +

Viridiplantae; Adl et al., 2005).

6

1.2.1 Names for phytochrome gene lineages

The high diversity of phytochromes I discovered in charophytes, mosses and ferns—resulting from multiple, independent gene duplications (Figure 2)—demanded a sensible system for naming the gene lineages. Within each major organismal group of Archaeplastida (except seed plants, where a system for naming *PHY* has already been well established), I used numerical labels for the phytochrome clades that resulted from major gene duplication events (e.g., fern *PHY1-4* and charophyte *PHY1-2*). Subclades resulting from more local duplications were then named alphabetically within clades (e.g., Polypodiales *PHY4A-B* and Desmidiales *PHY2A-C*). It should be stressed that this alphanumeric system does not imply orthology across organismal groups; for example fern *PHY1* shares a lower degree of relatedness to charophyte *PHY1* than to fern *PHY2*. Charophyte *PHYX1* and *PHYX2* were so named here because they are not canonical plant phytochromes like charophyte *PHY1-2*, and their evolutionary origin is less clear. For the cryptophyte phytochromes with C-terminal serine/threonine kinase, I followed Duanmu et al. (2014) and called them phytochrome eukaryotic kinase hybrids (PEK).

1.2.2 Stramenopiles and haptophytes

Stramenopiles are a large eukaryotic clade that includes brown algae (such as kelps), golden algae, and diatoms, the latter being an important component of plankton. Within this group, phytochromes are known so far only from brown algae, some of their viruses, and diatoms. Their sequences form a clade that is sister to fungal phytochromes (Figure 3, Appendix Figure 17). Interestingly, the phytochrome from the brown algal virus EsV-1 (Delaroque et al., 2001) does not group with brown algae phytochromes, but instead is more closely related to those of diatoms. This relationship was not supported in a bootstrapping analysis (Appendix Figure 17); it was, however, also obtained by Duanmu et al. (2014) (but without support).

Figure 3: Phylogeny of phytochrome. Terminal clades are collapsed into higher taxonomical units (usually orders or classes) for display purposes; the detailed tree is shown in Supplementary Fig. 1. Orange circles indicate inferred gene duplications. Italicized letters within each circle corresponds to the duplication event mentioned in the text, and the numbers/letters adjacent to each orange circle are the names of the gene duplicates. Canonical plant phytochromes originated in the common ancestor of Streptophyta (green star), and some charophytes retain non-canonical phytochromes (PHYX1, PHYX2). Domain architectures are shown on the right. Domains that are not always present are indicated by dashed outlines. Domain names:

GAF, cGMP phosphodiesterase/adenylate cyclase/FhlA; H/KD, histidine phosphorylation site (H) in the histidine kinase domain (KD); PAS, Per/Arnt/Sim; PHY, Phytochrome; PKC, Protein Kinase C; REC, Response Regulator; and RING, Really Interesting New Gene. *Traditional Archaeplastida does not include cryptophytes (Adl et al., 2005).

Additional phytochrome data from stramenopiles will be necessary to clarify the origin of these viral phytochromes. I also examined haptophytes, a predominantly marine lineage of phytoplankton (their relationships with stramenopiles and other protists are unclear (Grant and Katz, 2014; Burki et al., 2012). No phytochrome could be found in the haptophyte transcriptomes.

1.2.3 Red algae

Red algae are mostly multicellular, marine species that includes many coralline reef-building algae. No phytochromes were found in the 28 red algal transcriptomes I examined, nor in the published genomes of *Porphyridium purpureum*, *Chondrus crispus*, *Cyanidioschyzon merolae*, *Galdieria sulphuraria,* and *Pyropia yezoensis* (Appendix Table 2). This result, based on data from all Rhodophyta classes (Yoon et al., 2006), provides compelling evidence for the absence of phytochromes from red algae (Figure 2).

1.2.4 Glaucophytes

Glaucophytes are a small clade of freshwater, unicellular algae with unusual plastids referred to as cyanelles, which, unlike plastids in rhodophytes and green plants, retain a peptidoglycan layer (Keeling, 2004). Phytochromes are present in glaucophytes, and when the tree is rooted on the branch to prokaryote/fungus/stramenopile phytochromes, glaucophyte phytochromes are resolved as sister to cryptophyte + Viridiplantae phytochromes (Figure 3, Appendix Figure 17). Glaucophyte phytochromes, in contrast with canonical plant phytochromes, have a single PAS domain in the C-terminal module, and the conserved histidine residue is present in the kinase domain, suggesting it retains histidine kinase activity (Duanmu et al., 2014).

Figure 4: The diversity and evolution of phytochrome C-terminal output module. The tree depicts the relationship of all the phytochrome-containing lineages. For each lineage, the domain architecture of the C-terminal regulatory module is shown on the right (connected by dashed lines). The N-terminal photosensory module is omitted. The substitution of the histidine phosphorylation site (H) in the histidine kinase domain (KD) occurred subsequent to the divergence of prasinophytes. The canonical plant phytochrome is restricted to streptophytes (in grey box), although Zygnematales and Coleochaetales also have non-canonical plant phytochromes. Domain names: PAS, Per/Arnt/Sim; PKC, Protein Kinase C; REC, Response Regulator; and RING, Really Interesting New Gene. *Traditional Archaeplastida does not include cryptophytes32. †Full length phytochrome from Charales is not available and the domain composition was inferred.

1.2.5 Cryptophytes

The phylogenetic position of cryptophytes remains controversial. They were once thought to be related to stramenopiles and haptophytes (belonging to the kingdom Chromalveolata), but some recent phylogenomic studies place them either as nested within, or sister to, Archaeplastida (Cavalier-Smith et al., 2014; Burki et al., 2012; Grant and Katz, 2014). In my analyses, cryptophyte + Viridiplantae phytochromes form a clade that is sister to glaucophyte phytochromes (Figure 3, Appendix Figure 17). Also, phytochromes from Viridiplantae and from some cryptophytes share the characteristic PAS-PAS repeat in the C-terminus (Figure 4). These cryptophyte phytochromes differ from the canonical phytochromes in their retention of the

conserved histidine phosphorylation site in the kinase domain (Figure 3, Figure 4). Some

cryptophyte phytochromes do not have the PAS-PAS repeat in the C-terminus, but instead

possess a single PAS followed by a serine/threonine kinase domain ("PKC" in Figure 3, Figure 4).

Despite this variation in the C-terminus, the N-terminal photosensory modules of all cryptophyte

phytochromes are monophyletic (Figure 3, Appendix Figure 17).

1.2.6 Viridiplantae

Viridiplantae comprise two lineages, Chlorophyta and Streptophyta. Chlorophyta

include chlorophytes (Trebouxiophyceae + Ulvophyceae + Chlorophyceae + Pedinophyceae) and

prasinophytes (Figure 2). Chlorophytes appear to lack phytochromes entirely; I did not find

homologs in any of the chlorophyte transcriptomes examined, including 14 Trebouxiophyceae, 21

Ulvophyceae, 59 Chlorophyceae, and 2 Pedinophyceae. This result is consistent with available

whole-genome sequence data; the genomes of *Chlamydomonas reinhardtii*, *Volvox carteri* and

Chlorella variabilis (Chlorophyceae) lack phytochromes. Prasinophytes, on the other hand, do have

phytochromes. Most of these have a PAS-PAS repeat, a histidine kinase domain, and a response-

regulator domain at the C-terminus (Duanmu et al., 2014). Prasinophyte phytochromes are

monophyletic and are the sister group to streptophyte phytochromes (Figure 3, Appendix Figure

17).

Streptophyta (or streptophytes) are an assemblage of the charophytes (a paraphyletic

grade of algae) and the land plants (Wickett et al., 2014) (Figure 2). I found phytochrome

homologs in all land plant clades, as well as in all charophyte lineages: Mesostigmatales

(including Chlorokybales), Klebsormidiales, Coleochaetales, Charales, Zygnematales, and

Desmidiales (Figure 1, Figure 2). The Charales phytochromes were not included in my final

phylogenetic analyses because the transcriptome contigs (and also the data currently available on

GenBank) are too short to be informative about their relationships. All streptophytes have

canonical plant phytochromes, including Mesostigmatales, the earliest-diverging charophyte lineage (Figure 2, Figure 3, Figure 4). This result suggests that the origin of the canonical plant phytochrome took place in the ancestor of extant streptophytes.

Within charophyte algae I identified several gene duplication events. I infer one duplication to have occurred after Mesostigmatales diverged ("A" in Figure 3), resulting in two clades: one is small and charophyte-specific (charophyte *PHY1*), whereas the other is large and includes charophyte *PHY2*, and the land plant phytochromes. Members of the charophyte *PHY1* clade are not common in the algal transcriptomes, and were found only in Desmidiales and in *Entransia* of the early-diverging Klebsormidiales (Appendix Figure 17). On the other hand, the charophyte *PHY2* homolog is found consistently across algal transcriptomes. It experienced additional duplications ("B" and "C" in Figure 3) that resulted in three phytochrome subclades within Desmidiales (Desmidiales *PHY2A-C*). Relationships recovered within each of these phytochrome subclades correspond well to species phylogenies for Desmidiales (Gontcharov and Melkonian, 2010).

I found that Zygnematales and Coleochaetales (charophytes) also have two non-canonical phytochrome clades (charophyte *PHYX1* and *PHYX2*, Figure 3). Some *PHYX1* has a response regulator domain at the C-terminus, similar to prasinophyte, cryptophyte, and glaucophyte phytochromes (Figure 2, Figure 3). Intriguingly, *PHYX1* lacks all the known conserved cysteine residues (CysA-D; Rockwell et al., 2014) in the PAS-GAF region of the N-terminus that bind bilin chromophores, indicating that either this protein may not bind a bilin, or that a non-conserved binding site is used.

Figure 5: Phylogenetic relationship of neochromes and phytochromes. The support values are shown for the neochrome branches only, in the following order: maximum likelihood bootstrap support (MLBS) from GTR nucleotide model / Bayesian posterior probabilities (PP) from GTR nucleotide model / aLRT support from codon model / maximum likelihood bootstrap values from JTT amino acid model / Bayesian posterior probabilities from JTT amino acid model. "*" indicates all the support values = 100 or 1.0. "-" denotes MLBS < 70, aLRT < 70, or PP < 0.95. Branches are thickened when MLBS > 70, aLRT > 70, and PP > 0.95.

1.2.7 Neochromes

My data suggest that the phytochrome module of neochrome had a single origin (Figure 3, Appendix Figure 17). Published data indicate that the phototropin module of neochromes, in contrast, had independent origins in algae and hornworts (Li et al., 2014; Chapter 3), implying two separate fusion events involving phytochromes that shared a common ancestor. To further explore this finding, I analyzed the neochrome nucleotide dataset (see above) using several nucleotide, codon and amino acid models, and performed a topology test. I consistently recovered the monophyly of the phytochrome module of neochromes, and usually with high support, from analyses using all models (Figure 5). Although *Anthoceros* (a hornwort) neochrome was resolved as sister to a Zygnematales (algal) neochrome, this relationship was not supported

(except in the MrBayes analysis of the nucleotide dataset). I then used the Swofford-Olsen-Waddell-Hillis (SOWH) test to compare the topology with all neochromes (the phytochrome module) forming a single clade, against an alternative in which neochromes of Zygnematales were forced to not group with hornworts + ferns. The alternative hypothesis was rejected ($P <$ 0.00001), and the monophyly of the phytochrome module of neochromes was favored.

1.2.8 Bryophytes

Phytochromes from mosses, liverworts, and hornworts each form a monophyletic group (Figure 6). I detected single phytochrome homologs in hornwort and liverwort transcriptomes. The gene phylogenies match the species relationships (Villarreal and Renner, 2012; Forrest et al., 2006), consistent with the presence of single orthologous genes in these taxa. Indeed, a single phytochrome has been identified via cloning methods in the liverwort, *Marchantia paleacea* var. *diptera* (Suzuki et al., 2001). I also searched the low-coverage draft genome of the hornwort *Anthoceros punctatus* (20X; Li et al., 2014, Chapter 3) and found only one phytochrome. To further evaluate gene copy number, I hybridized the *Anthoceros punctatus* genomic DNA with phytochrome RNA probes, and used Illumina MiSeq to sequence the captured DNA fragments. The same phytochrome contig (and only that contig) was recovered, suggesting that this hornwort does not harbor additional, divergent phytochrome copies.

In contrast, phytochromes in mosses are diverse, with at least four distinct clades resulting from three gene duplications (Figure 6). The phylogeny reveals those moss phytochromes that are orthologous to the previously named *Physcomitrella patens* phytochromes, *PpPHY1-5*. The *Physcomitrella* phytochromes and their orthologs form the following clades: moss *PHY1_3* (including *PpPHY1* and *PpPHY3*), moss *PHY2_4* (including *PpPHY2* and *PpPHY4*), and moss *PHY5* (including *PpPHY5A-C*). An ancient duplication ("D" in Figure 6) gave rise to moss *PHY1_3* and moss *PHY2_4* + *PHY5* clades. The timing of this duplication is dependent on the

14

phylogenetic position of the *Takakia* phytochrome, which was resolved here as sister to the moss *PHY2_4* + *PHY5* clade but without support (Figure 6). Because *Takakia* (Takakiopsida) represents the earliest-diverging lineage in the moss species phylogeny (Chang and Graham, 2011), the first phytochrome duplication probably predates the origin of all extant mosses. In the moss *PHY2_4* + *PHY5* clade, another duplication ("E" in Figure 6) occurred following the split of *Andreaea* (Andreaeopsida) but before *Atrichum* (Polytrichopsida) diverged, separating moss *PHY2_4* and *PHY5*. The moss *PHY5* clade had an additional duplication ("F" in Figure 6), probably after *Physcomitrella* diverged, that resulted in moss *PHY5D* and *PHY5E* subclades.

My results show that the phytochrome copies previously cloned from *Ceratodon purpureus*, which were named *CpPHY1-4* (Mittmann et al., 2009), have the following relationships with the moss phytochromes: *CpPHY1* and *CpPHY2* are each others closest relatives, and are members of the moss *PHY1_3* lineage; *CpPHY3* and *CpPHY4* are members of the moss *PHY5* lineage (Fig. 4). These results suggest that the four known *C. purpureus* phytochromes— "*CpPHY1*", "*CpPHY2*", "*CpPHY3*" and "*CpPHY4*" (Figure 6) should be renamed to *CpPHY1_3A*, *CpPHY1_3B*, *CpPHY5D* and *CpPHY5E*, respectively, and that the novel *C. purpureus* phytochrome discovered here should be designated as *CpPHY2_4*.

1.2.9 Lycophytes

Lycophyte phytochromes are resolved as monophyletic and are sister to the fern plus seed plant phytochromes (Figure 7). *Selaginella* and *Isoetes* (Isoetopsida) each have a single phytochrome, with the exception of *S. mollendorffii*, where two nearly identical phytochromes are apparent in the whole-genome sequence data. Their high degree of similarity suggests that they might be products of a species-specific gene duplication. In contrast, Lycopodiales have two distinct phytochrome clades that I name Lycopodiales *PHY1* and Lycopodiales *PHY2*. Because all the Lycopodiales lineages (Wikstrom, 2001) are represented in each phytochrome clade, I infer

that the duplication of Lycopodiales *PHY1/2* ("G" in Figure 1) predates the common ancestor of all

extant Lycopodiales.

Figure 6: Phytochrome phylogeny for bryophytes. Phytochromes previously identified are in bold. Support values associated with branches are maximum likelihood bootstrap values (BS) / Bayesian posterior probabilities (PP); these are only displayed (along with thickened branches) if BS > 70 and PP >

0.95. Thickened branches without numbers are 100/1.0. The position of orange circles estimates the origin of inferred gene duplications. Italic letters within each circle correspond to the duplication event mentioned in the text, and the numbers/letters adjacent to each circle indicate the names of the gene duplicates.

1.2.10 Ferns

Fern phytochromes form a clade that is sister to the seed-plant phytochromes (Figure 3, Figure 7, Appendix Figure 17). Within ferns I uncovered four phytochrome clades that I designate fern *PHY1*, *PHY2*, *PHY4A,* and *PHY4B*. The name *PHY3* was used previously to denote the chimeric photoreceptor that is now recognized as neochrome (Suetsugu et al., 2005; Li et al., 2014). The deep evolutionary split between the fern *PHY1* and *PHY2/4* clades predates the most recent ancestor of extant ferns ("H" in Figure 7). Fern *PHY2* and *PHY4* probably separated after Gleicheniales diverged ("I" in Figure 7), and the earliest-diverging fern lineages (i.e., Gleicheniales, Osmundales, Psilotales, Ophioglossales, Marattiales, and Equisetales) have the pre-duplicated *PHY2/4* copy. It should be noted that my broad-scale amino acid dataset resolved a slightly different topology, placing Gleicheniales *PHY2/4* closer to *PHY4* (Appendix Figure 17). However, the amino acid dataset included fewer sequences from ferns, which could reduce phylogenetic accuracy (Hillis, 1998). It is likely that that the phylogeny (Figure 7) inferred from rigorous analyses of nucleotide data more accurately reflects gene relationships.

I found that Ophioglossales and Osmundales each have two *PHY2/4* copies, which likely arose from independent gene duplications (Figure 7). The duplication of Ophioglossales *PHY2/4A* and *PHY2/4B* occurred either at the ancestor of Ophioglossales or of Ophioglossales + Psilotales, but the history of *PHY2/4* in Osmundales is unclear. The Osmundales *PHY2/4A* and *PHY2/4B* were not resolved as monophyletic, and the phylogenetic position of Osmundales *PHY2/4B* is incongruent with published fern species relationships (Kuo et al., 2011).

After the split of fern *PHY2* and *PHY4*, *PHY4* duplicated again, giving rise to fern *PHY4A* and *PHY4B* ("J" in Figure 7), and both are found in Polypodiales. I cannot precisely determine the

timing of this duplication event because the relationships among Polypodiales *PHY4A-B*,

Cyatheales *PHY4* and Salviniales *PHY4* are resolved without support. Interestingly, *PHY4A* was

previously known only from *Adiantum capillus-veneris* (as *AcPHY4*). Its first intron incorporated

an inserted Ty3/gypsy retrotransposon and the downstream exon sequence was unknown

(Nozue et al., 1997). I found full-length *PHY4A* transcripts in a wide range of Polypodiales,

suggesting that *PHY4A* likely is functional in most other species, if not in *A. capillus-veneris*.

PHY4B is a novel phytochrome clade that has not been documented before; it is not common in

the fern transcriptomes I examined.

1.2.11 Seed plants

Seed-plant phytochromes cluster into three clades (Appendix Figure 17) corresponding to

PHYA, *PHYB/E*, and *PHYC*, in accordance with previous studies (Mathews, 2010). Organismal

relationships within the gene subclades largely are consistent with those inferred in phylogenetic

studies of angiosperms (Bremer et al., 2009). Notably, however, support for the monophyly of

gymnosperms was low. I found two divergent transcripts of *PHYE* in Ranunculales, represented

by *Aquilegia* (Ranunculaceae; from whole genome data) and *Capnoides* (Papaveraceae; from

transcriptome data) (Figure 17 in Appendix A), suggesting that a gene duplication event occurred

deep in Ranunculales; however, more extensive sampling in Ranunculales is needed to resolve

the timing of this duplication.

Figure 7: Phytochrome phylogeny for ferns and lycophytes. Phytochromes that were previously reported are shown in bold. Support values associated with branches are maximum likelihood bootstrap values (BS) / Bayesian posterior probabilities (PP); these are only displayed (along with thickened branches) if BS > 70 and PP > 0.95. Thickened branches without numbers are 100/1.0. The position of orange circles estimates the origin of inferred gene duplications. Italic letters within each circle correspond to the duplication event mentioned in the text, and the numbers/letters adjacent to each circle indicate the names of the gene duplicates.

1.3 Discussions

My phylogenetic results refute previous hypotheses suggesting that plants acquired phytochrome from cyanobacteria via endosymbiotic gene transfer (Karniol et al., 2005; Herdman et al., 2000), because streptophyte and cyanobacterial phytochromes are not closest relatives in my phytochrome trees (Figure 3, Appendix Figure 17), a result also recently obtained by Duanmu et al. (2014). Instead, plant phytochromes evolved from a precursor shared with other Archaeplastida. I clearly placed the origin of canonical plant phytochromes in a common ancestor of extant streptophytes (Figure 3, Figure 4). My data, moreover, show that the origin of this structure required multiple steps. The gain of the internal PAS-PAS repeat took place first, in the ancestor of Viridiplantae, or of Viridplantae + cryptophytes (Figure 4). As noted above, the position of cryptophytes is uncertain, and its inclusion in Archaeplastida is not strongly supported in published studies (Burki et al., 2012; Grant and Katz, 2014; Cavalier-Smith et al., 2014). The topology of my phytochrome trees is consistent with a sister-group relationship between Viridiplantae and cryptophytes, but the topology also could result from endosymbiotic or horizontal gene transfer (EGT or HGT). The loss of the histidine phosphorylation site in the histidine kinase domain—hence the attainment of the canonical form—occurred later, in the ancestor of streptophytes, and seems to have been accompanied by a permanent dissociation with the response regulator at the C-terminal end (Figure 4). Some streptophytes have additional, non-canonical phytochromes. Charophyte *PHYX1* and *PHYX2*, both found in Zygnematales and Coleochaetales, have the conserved histidine residue, and some *PHYX1* also have a response regulator domain (Figure 3, Figure 4). The fact that charophyte *PHY1, PHYX1,* and *PHYX2* are not found in all streptophytes implies that duplications early in the history of streptophytes were followed by multiple losses of charophyte *PHY1* and the non-canonical charophyte *PHY*.

My findings highlight the different evolutionary histories of the phytochrome N- and C-terminal modules. The N-terminal photosensory module is deeply conserved across eukaryotes and prokaryotes, and the linear domain sequence of PAS-GAF-PHY is found in the majority of known phytochromes (Figure 3). In contrast, the evolution of the C-terminal regulatory module has been much more dynamic (Figure 4). For example, the C-terminal PAS may be absent, may occur singly, or may occur as a tandem repeat (Figure 4). Serine/threonine kinase or tyrosine kinase domains have also been independently recruited into the regulatory module in the cryptophyte and *Ceratodon purpureus* (moss) phytochromes (Thümmler et al., 1992) (Figure 3). The successful linkage of the phytochrome photosensory module with a variety of C-terminal modules has promoted phytochrome functional diversity. Certainly the most compelling example is that of the neochromes. It combines phytochrome and phototropin modules into a single protein to process blue and red/far-red light signals in the control of phototropism (Kanegae et al., 2006). Neochrome was first discovered in ferns (Nozue et al., 1998) and postulated to be a driver of the modern fern radiation under low-light, angiosperm-dominated forest canopies (Kawai et al., 2003; Schneider et al., 2004; Schuettpelz and Pryer, 2009). Suetsugu et al. (2005) later discovered a similar phytochrome-phototropin chimera in *Mougeotia scalaris* (zygnematalean alga), and proposed that neochrome had independently evolved twice. A recent study identified yet another neochrome from hornworts, and demonstrated that ferns acquired their neochromes from hornworts via horizontal gene transfer (Li et al., 2014; Chapter 3). By placing the phototropin portion of neochrome into a broad phylogeny of phototropins, Li et al. (2014) also showed that phototropin modules of neochromes had two separate origins, once in hornworts and once in zygnematalean algae. In contrast, the phytochrome portion of neochrome has had a different evolutionary history, with Zygnematales, hornworts, and ferns forming a single monophyletic group (Figure 5). This result is robust, and is supported by most of the

21

analyses and by a topology test. My results thus suggest that neochromes originated via two separate fusion events, involving two distinct sources of phototropin but the same phytochrome progenitor. This is a fascinating extension of the capacity and propensity of the phytochrome photosensory module to be linked with functionally distinct downstream domains.

The major clades of land plants differ markedly with respect to phytochrome gene diversity. It appears that phytochromes are single-copy in most liverworts, hornworts and Isoetopsida (Isoetaceae and Selaginellaceae), whereas they have independently diversified in Lycopodiales, mosses, ferns and seed plants (Figure 3). In ferns, a pattern of early gene duplication followed by gene losses could explain the phylogenetic positions of two Osmundales *PHY2/4*, which are incongruent with known species relationships in ferns (Figure 7). Interestingly, I observed a relationship between phytochrome copy number and species richness. For instance, the polypod ferns (Polypodiales), which account for 90% of extant fern diversity (Schuettpelz and Pryer, 2009), have four phytochrome copies, whereas other species-poor fern lineages have only two or three (Figure 7). Likewise, moss species belonging to the hyper-diverse Bryopsida— containing 95% of extant moss diversity—have experienced the highest number of phytochrome duplications compared to other bryophyte lineages (Figure 6). It is possible that the evolution of phytochrome structural and functional diversity enhanced the ability of polypod ferns and Bryopsida mosses to adapt to diverse light environments. Indeed, seed plants, ferns, and mosses each have at least one phytochrome duplicate that convergently evolved or retained the role of mediating high-irradiance responses (Cooke et al., 1993; Possart and Hiltbrunner, 2013; Mathews, 2006; Mathews and Tremonte, 2012), a trait likely to be important for surviving under deep canopy shade (Yanovsky et al., 1995) (see below). This "phytochrome-driven species diversification" hypothesis, however, needs rigorous testing by phylogenetic comparative

methods and functional studies in non-seed plants that identify the genetic bases of phytochrome functions.

The independent phytochrome diversification events in seed plants, ferns, mosses and Lycopodiales have significant implications for phytochrome functional studies. Moss phytochromes, for example, are more closely related to each other than to any of the seed-plant phytochromes (and the same is true, of course, for phytochromes from ferns, and those from Lycopodiales). Seed-plant phytochromes have undergone significant differentiation into two major types. One is represented by phyA of *Arabidopsis thaliana*, which is the primary mediator of red-light responses in deep shade and beneath the soil surface. It degrades rapidly in light, mediates very-low-fluence and high-irradiance responses, and depends on protein partners FHY1 (FAR-RED ELONGATED HYPOCOYTL 1) and FHL (FHY1 LIKE) for nuclear translocation. The other is represented by phyB-E of *A. thaliana*, which are the primary mediators of red-light responses in open habitats. They are light-stable, mediate low-fluence responses; and in the case of phyB, depend on an internal nuclear localization signal for nuclear translocation (Franklin and Quail, 2010; Chen and Chory, 2011). A similar partitioning of function has been documented in some fern and moss phytochrome duplicates (Sineshchekov et al., 2013; Possart and Hiltbrunner, 2013; Possart et al., 2014), demonstrating a case of convergent differentiation following independent gene duplications. In future studies, it would be of particular interest to infer the ancestral properties of land plant phytochrome: Was it light-labile or stable? What kinds of physiological responses did it mediate? How was nuclear translocation executed? Studies of liverwort, hornwort and *Selaginella* phytochromes, which exist as a single-copy gene, could serve as "baseline models" for understanding the genetic basis of phytochrome functional diversification.

Recent functional studies on a small but varied set of algal phytochromes revealed a surprising degree of spectral diversity, which might reflect adaptations to a range of marine and aquatic environments (Rockwell et al., 2014; Duanmu et al., 2014). For example, photoreversible phytochromes in prasinophyte algae include orange/far-red receptors as well as red/far-red receptors, and in algae outside of Viridiplantae, there also are blue/far-red and red/blue receptors (Duanmu et al., 2014). This sharply contrasts with the very limited spectral diversity in canonical plant phytochromes, all of which are red/far-red receptors. It appears then, that the transition of plants from marine or aquatic to terrestrial environments involved a centering of phytochrome evolution on a limited model. The novel algal phytochrome clades I uncovered here are a potential treasure trove for discovering the steps during this transition, and for characterization of new biochemical variants, some of which may have implications for understanding the role of phytochrome evolution in recolonization of marine and aquatic environments by terrestrial plants.

In summary, my study has revealed that the diversity of Viridiplantae phytochromes is far greater than was realized, and points to exciting opportunities to link this structural diversity with function and ecology.

1.4 Materials and Methods

1.4.1 Transcriptome- and genome-mining for phytochrome

The transcriptomes and genomes sampled in this study are listed in Appendix Table 2. I used the Python pipeline BlueDevil following Li et al. (2014) to mine transcriptomes. To search the whole-genome data, I used BLASTp implemented in Phytozome (Goodstein et al., 2012) or individual genome project portals (Appendix Table 2). The protein domain composition for each of the phytochrome sequence was determined by querying the NCBI Conserved Domain Database (Marchler-Bauer et al., 2011).

1.4.2 Sequence alignment

In addition to the phytochrome homologs mined from transcriptomes and genomes, I also gathered selected Genbank accessions and a sequence cloned from *Marattia howeana* (voucher: S.W.Graham & S.Mathews 15, deposited in NSW; primers: 110f – 5'GTNACNGCNTAYYTNCARCGNATG3', 788r – 5'GTMACATCTTGRSCMACAAARCAYAC3').

I assembled four sequence datasets, one was translated into an amino acid alignment and the others were analyzed as nucleotide matrices. The amino acid dataset included the majority of the sequences I have (423 sequences in total; Appendix Figure 17). The sequences were initially aligned using MUSCLE (Edgar, 2004), and I manually curated the alignment based on known domain boundaries and protein structures. The unalignable regions were excluded and the final alignment included 1,106 amino acid sites. The nucleotide datasets were assembled to provide higher phylogenetic resolution within fern + lycophyte phytochromes (113 sequences; Figure 7), bryophyte phytochromes (97 sequences; Figure 6), and neochromes (111 sequences; Figure 5).

25

Sequences were aligned as amino acids and then back-translated to nucleotides, and the alignment was refined by manual editing. The fern + lycophyte, and bryophyte phytochrome alignments contained 3,366 and 3,429 nucleotide sites, respectively. The neochrome alignment included only the N-terminal photosenory module (PAS-GAF-PHY domains; 1,920 nucleotide sites). All alignments are available from Dryad (http://dx.doi.org/10.5061/dryad.[NNNNN]).

1.4.3 Phylogenetic reconstruction

For the broad-scale amino acid alignment, JTT + I + G was selected as the best-fitting empirical model by ProTest under Akaike Information Criterion (Darriba et al., 2011). I used Garli v2.0 (Zwickl, 2006) to find the maximum likelihood tree, with ten independent runs and genthreshfortopoterm set to 100,000. The starting tree for Garli came from a RAxML (Stamatakis, 2006) run. To obtain bootstrap branch supports, RAxML was ran run with 1,000 replicates using JTT + G.

For the nucleotide alignments, I used PartitionFinder (Lanfear et al., 2012) to infer the best codon position partition schemes and substitution models, under Akaike Information Criterion. Maximum likelihood tree searching and bootstrapping (1,000 replicates) were done in RAxML. Bayesian inference was carried out in MrBayes (Ronquist et al., 2012), with two independent Markov chain Monte Carlo (MCMC) runs and four chains each. I unlinked the substitution parameters and set the rate prior to vary among partitions. The MCMC output was inspected using Tracer (Rambaut and Drummond, 2013) to ensure convergence and mixing (effective sample sizes all >200); 25% of the total generations were discarded as burn-in before analyzing the posterior distribution.

Additional analyses were applied to the neochrome dataset. First, I used CodonPhyML (Gil et al., 2013) to infer the tree topology and to assess support (SH-like aLRT branch support), using a codon substitution model. Four categories of non-synonymous/synonymous substitution rate ratios were drawn from a discrete gamma distribution, and codon frequencies were estimated from the nucleotide frequencies at each codon position (F3 × 4). Second, I translated the nucleotides into amino acids, and carried out maximum likelihood tree searching and bootstrapping (in RAxML), as well as Bayesian inference (in MrBayes) under the JTT + I + G model. Finally, I used the Swofford–Olsen–Waddell–Hillis (SOWH) test, implemented in SOWHAT (Church et al. 2015), to investigate whether the inferred tree topology (phytochrome portion of neochrome forming a clade) is significantly better than the alternative topology (neochrome not monophyletic). In SOWHAT, I used the default stopping criterion and applied a topological constraint forcing land plant and zygnematalean neochrome to be non-monophyletic.

1.4.4 Confirming gene copy number in hornworts by target enrichment

I used a target enrichment strategy to test whether hornworts have a single phytochrome locus. In this approach, specific RNA probes are hybridized to genomic DNA to enrich the representation of particular gene fragments. Target enrichment has several advantages over the traditional Southern blotting approach. In particular, it uses thousands of different hybridization probes (rather than just a few), and the end products are not DNA bands, but actual sequence data.

I designed a total of 7,502 120-mer RNA probes to target phytochrome, phototropin and neochrome genes, with a special focus on those of hornworts and ferns (probe sequences available from Dryad http://dx.doi.org/10.5061/dryad.[NNNNN]). The probes overlap every 60bp

(a 2X tiling strategy), and were synthesized and biotin-labeled by Mycroarray. Genomic DNA of the hornwort *Anthoceros punctatus* was extracted using a modified CTAB protocol, and sheared by Covaris with fragments peak at 300bp. Library preparation for Illumina sequencing was done using a KAPA Biosystem kit, in combination with NEBNext Multiplex Oligos. To enrich for potentially divergent homologs, I used the touchdown procedure of Li et al. (2013), in which the genomic DNA library and the probes were hybridized at 65 °C for 11 hours followed by 60 °C (11 hours), 55 °C (11 hours) and 50 °C (11 hours). The hybridized DNA fragments were captured by streptavidin beads and washed following the protocol of Mycroarray. The final product was pooled with nine other libraries in equimolar and sequenced on Illumina MiSeq (250bp paired-end). To process the reads, I used cutadapt (Martin, 2011) to remove the adaptor sequences, and used sickle (Joshi and Fass, 2011) to trim low-quality bases. The resulting reads were then assembled by SOAPdenovo2 (Luo et al., 2012), and the phytochrome contig was identified by tBLASTn (Camacho et al., 2009). The raw reads were deposited in NCBI SRA (SRP055877).

2. The origin and evolution of phototropins

Li, F.-W., C.J. Rothfels, M. Melkonian, J.C. Villarreal, D.W. Stevenson, S.W. Graham, G.K.S. Wong, S. Mathews, and K.M. Pryer. **On the origin and evolution of phototropins**. in review.

2.1 Introduction

Despite their sedentary nature, plants are not static and are capable of a surprising range of motion (Darwin and Darwin, 1880). Plants have evolved sophisticated phototropic responses—involving movement of shoots and/or chloroplasts—to optimize their exposure to light. Charles Darwin pioneered modern phototropism research by demonstrating that the shoot tip is the point of light perception from where "fluence" is transduced to initiate tropic movements (Darwin and Darwin, 1880). Subsequent studies soon led to the discovery of the plant hormone auxin (Darwin's "fluence"), and later the identification of the blue-light photoreceptors—phototropins (Liscum et al., 2014).

Phototropins regulate key adaptive physiological responses that are under light control, including shoot-positive phototropism, root-negative phototropism, chloroplast accumulation/avoidance, stomatal opening, circadian rhythm, leaf expansion, and seedling elongation (Christie, 2007). Our current understanding of the function and biochemistry of phototropins originates from basic research on *Arabidopsis thaliana* (a flowering plant), and to a much lesser extent on *Adiantum capillus-veneris* (a fern) and *Physcomitrella patens* (a moss). Despite the phylogenetic span encompassed by these model organisms, the orthology of phototropin genes has been ambiguous, confounding not only functional homology assignments, but also our

29

understanding of their role in allowing plants to adapt to heterogeneous light environments through time.

Figure 8: Organismal lineages screened for phototropin homologs. (A) Viridiplantae and algae. Lineages that lack phototropin are depicted in grey. Topology derived from Wickett et al. (2014) and Burki et al. (2012). Phototropin and phytochrome duplications are only shown on land plant branches (within grey box). (B) Ferns and lycophytes; topology derived from Wickett et al. (2014) and Kuo et al. (2011) (C) Mosses; topology derived from Cox et al. (2010). Capital letters above blue squares denote phototropin duplication events mentioned in the text and in Fig. 2. "†" indicates that the exact phylogenetic position of the gene duplication event is ambiguous.

2.2 Results

2.2.1 The origin of phototropins

To pinpoint the origin of phototropins and reconstruct their evolutionary history, I examined 215 transcriptomes and genomes spanning all extant plant and algal lineages (Appendix Table 4, Table 5). I show here that phototropins are present in all major land plant

30

lineages (seed plants, ferns, lycophytes, mosses, liverworts, and hornworts), as well as in green

algae (charophytes, chlorophytes, and prasinophytes; Figure 8). In contrast, I did not recover

phototropins from glaucophytes, red algae, cryptophytes, haptophytes and stramenopiles,

indicating that the origin of phototropin most likely took place in an ancestor of Viridiplantae

(green algae + land plants; Figure 8). Because the chlorophyte alga *Chlamydomonas reinhardtii* is

known to use phototropins to modulate its sexual processes (Huang and Beck, 2003), it is possible

that the function of early phototropins may not have involved phototropic responses.

Unfortunately, so little is known about phototropin function in green algal lineages that we

cannot determine when phototropins were recruited to direct trophic responses.

2.2.2 Phototropin phylogeny

Seed-plant phototropins form a monophyletic group that is sister to fern phototropins

(Figure 9). Here I infer a single gene duplication event in seed plants, one leading to *Arabidopsis*

PHOT1 and the other to *Arabidopsis PHOT2* (Christie, 2007). Because the *PHOT1* and *PHOT2*

clades each include angiosperms and gymnosperms, the duplication event that gave rise to these

two homologs predates the divergence of all extant seed plants ("A" in Figure 9). I also find strong

evidence for the monophyly of fern phototropins ("B" in Figure 9). Leptosporangiate ferns have

two phototropin homologs that we designate *PHOT1* and *PHOT2*, in reference to *Adiantum*

capillus-veneris PHOT1 and *PHOT2* (Kagawa et al., 2004), respectively. The earliest-diverging fern

lineages, Equisetales, Psilotales and Ophioglossales, each have one phototropin gene,

representing the pre-duplicated version of fern *PHOT1* and *PHOT2*. The exact phylogenetic

position as to where fern *PHOT1* and *PHOT2* diverged is ambiguous due to a lack of branch

support, although it probably was prior to leptosporangiate ferns diverging from Marattiales

("B" in Figure 9).

As with seed plants and leptosporangiate ferns, I infer a single duplication event in the lycophyte *Selaginella*, corresponding to *PHOT1* and *PHOT2* based on the genome annotation of *S. moellendorffii* (Banks et al., 2011). The phylogenetic position of this duplication is also unclear ("C" in Figure 10), but it must predate the common ancestor of extant *Selaginella* because the *PHOT1* clade contains all known major *Selaginella* lineages (Korall and Kenrick, 2002). For Isoetales and Lycopodiales, we found only one phototropin homolog, but whether or not it is indeed a single-copy gene in these lineages will require further confirmation.

All liverwort transcriptomes we examined contained only one phototropin (Figure 10), a result consistent with the recent demonstration that phototropin in *Marchantia polymorpha* is a single-copy gene (Komatsu et al., 2014). Hornwort phototropins also appear to be single-copy, based on our screening of hornwort transcriptomes and a low-coverage genome draft of *Anthoceros* (Li et al., 2014; Chapter 3). To further confirm the gene copy number in hornworts, we used a target enrichment strategy to sequence all phototropin-like genomic fragments in *Anthoceros punctatus*, and found no additional divergent copies.

Moss phototropins, on the other hand, have a significantly more complex evolutionary history. I determined that the phototropin annotations from the moss *Physcomitrella patens* genome (*PpPHOTA1-4*, *PpPHOTB1-3*) do not reflect gene orthology. Because "*PHOTAs*" and "*PHOTBs*" are intermingled, I reclassified the moss phototropins based on their phylogenetic relationships shown here (Table 1, Figure 10). Prior to the divergence of all extant mosses, a gene duplication event ("D" in Figure 10) took place, giving rise to moss *PHOT1* and *PHOT2*. In *PHOT1*, a second duplication occurred in the common ancestor of Bryopsida and Polytrichopsida

("E" in Figure 10) that split *PHOT1* into *PHOT1A* and *PHOT1B*. In *PHOT2*, two additional

duplications took place ("F" and "G" in Figure 10) subsequent to the divergence of Buxbaumiidae

(Bryopsida), resulting in *PHOT2A-C*. *PHOT2A* and *PHOT2B* are both present in Dicranidae and

Bryidae, whereas *PHOT2C* is only known in *Physcomitrella patens* (Funariidae). *Physcomitrella*

patens may also have lost the *PHOT2A* homolog. In green algae, most of the transcriptomes and

genomes revealed a single phototropin gene (Figure 11). The singular exception is Zygnematales,

where two phototropin homologs are present (*PHOTA* and *PHOTB*).

Table 1: Reclassification of *Physcomitrella patens* phototropins based on gene orthology.

Proposed new name	Previous annotation	Genbank No.
PpPHOT1A-1	*PpPHOTA1*	XM_001774204
PpPHOT1A-2	*PpPHOTA2*	XM_001774562
PpPHOT1A-3	*PpPHOTB3*	XM_001755269
PpPHOT1B	*PpPHOTA3*	XM_001765356
PpPHOT2B	*PpPHOTB2*	XM_001785674
PpPHOT2C-1	*PpPHOTB1*	XM_001766357
PpPHOT2C-2	*PpPHOTA4*	XM_001763052

Neochrome (NEO, Figure 10, Figure 11) is a unique phototropin variant that possesses

supplementary red/far-red-sensing domains from phytochromes (Nozue et al., 1998). Recent

studies have revealed two independent origins of neochromes, one in zygnematalean algae and

the other in hornworts (Suetsugu et al., 2005; Li et al., 2014; Chapter 3), and that the neochromes

found in ferns were derived from hornworts via horizontal gene transfer (Li et al., 2014; Chapter

3). Neochrome perceives both blue and red/far-red light to mediate phototropism and chloroplast

movement (Kanegae et al., 2006; Kawai et al., 2003) in ferns, and it appears to have played a

significant role in their diversification (Schneider et al., 2004; Schuettpelz and Pryer, 2009).

Neochrome function in zygnematalean algae, however, is still unclear. Because zygnematalean

algae have plate-like chloroplasts that rotate in response to both blue and red/far-red light

irradiation (Haupt and Scheuerlein, 1990), it was hypothesized that algal neochrome, originally

discovered in *Mougeotia scalaris*, is the gene candidate responsible (Suetsugu et al., 2005).

However, neochrome in *M. scalaris* responds only to red/far-red light and not to blue light

(Suetsugu et al., 2005). To explore whether *M. scalaris* might be an outlier among zygnematalean

algae in perhaps having a "defective" neochrome, I further investigated all the algal neochromes

that I mined. I discovered that none of them has the conserved cysteine residue in the LOV2

domain, that is essential for flavin mononucleotide chromophore adduct formation and blue light

signal transduction (Christie, 2007). Therefore, it is likely that all zygnematalean algae use

neochrome only for sensing red/far-red light, and use other blue-light photoreceptors

(phototropins or cryptochromes) to maneuver chloroplast rotations.

Fig. 2a

Fig. 2b

Fig. 2c

Seed plants

PHOT1

A

PHOT2

Arabidopsis thaliana
Solanum lycopersicum
Fragaria vesca
Medicago truncatula
Citrus clementina
Vitis vinifera
Aquilegia coerulea
Austrobaileya scandens
Illicium floridanum
Amborella trichocarpa
Magnolia grandiflora
Goodyera pubescens
Zea mays
Smilax bona-nox

Angiosperm PHOT1

Cunninghamia lanceolata
Thuja plicata
Cephalotaxus harringtonia
Podocarpus rubens
Gnetum montanum
Welwitschia mirabilis
Stangeria eriopus

Gymnosperm PHOT1

Arabidopsis thaliana
Citrus clementina
Fragaria vesca
Medicago truncatula
Solanum lycopersicum
Vitis vinifera
Aquilegia coerulea
Goodyera pubescens
Smilax bona-nox
Zea mays
Magnolia grandiflora
Austrobaileya scandens
Illicium floridanum
Amborella trichocarpa

Angiosperm PHOT2

Thuja plicata
Cunninghamia lanceolata
Cephalotaxus harringtonia
Podocarpus rubens
Welwitschia mirabilis
Gnetum montanum
Stangeria eriopus

Gymnosperm PHOT2

Polystichum acrostichoides
Polypodium hesperium
Leucostegia immersa
Cystopteris reevesiana
Pteridium aquilinum
Adiantum capillus-veneris
Gaga arizonica
Ceratopteris thalictroides

Polypodiales PHOT1

Lonchitis hirsuta
Pilularia globulifera

Salviniales PHOT1

Leptosporangiate ferns
Dipteris conjugata
Osmunda sp.
Danaea nodosa
Danaea nodosa
Danaea nodosa

Gleicheniales PHOT1
Osmundales PHOT1
Marattiales PHOT1

Onoclea sensibilis
Blechnum spicant
Athyrium filix-femina
Woodsia scopulina
Homalosorus pycnocarpos
Gymnocarpium dryopteris
Cystopteris reevesiana
Polypodium hesperium
Davallia fejeensis
Leucostegia immersa
Polystichum acrostichoides
Asplenium platyneuron
Dennstaedtia davallioides
Pteridium aquilinum
Adiantum capillus-veneris
Adiantum aleuticum
Vittaria lineata
Gaga arizonica
Pityrogramma trifoliata
Pteris vittata
Pityrogramma trifoliata
Cryptogramma acrostichoides
Ceratopteris thalictroides
Lindsaea linearis
Lonchitis hirsuta

Polypodiales PHOT2

PHOT1

B ‡

PHOT2

Thyrsopteris elegans
Plagiogyria japonica
Pilularia globulifera
Azolla caroliniana
Anemia tomentosa
Lygodium japonicum
Dipteris conjugata
Osmunda sp.

Salviniales PHOT2
Cyatheales PHOT2
Schizaeales PHOT2
Gleicheniales PHOT2
Osmundales PHOT2

Leptosporangiate ferns

Ferns

Botrypus virginianus
Sceptridium dissectum
Ophioglossum
Psilotum nudum
Tmesipteris parva
Equisetum hyemale

Ophioglossales PHOT1/2
Psilotales PHOT1/2
Equisetales PHOT1/2

To Fig. 2b

0.1 substitution/site

35

Figure 9: Phylogeny of seed-plant and fern phototropins. Orange circles indicate inferred photropin duplication events. The italicized capital letter within each circle corresponds to the duplication event mentioned in the text, and the numbers/letters adjacent to each orange circle are the names of the gene duplicates. The values associated with branches are maximum likelihood bootstrap values / Bayesian posterior probabilities. "†" indicates that the exact phylogenetic position of the gene duplication event is ambiguous.

To Fig. 9

Fig. 9

Fig. 10

Fig. 11

96/.98

Selaginella moellendorffii 1-2
Selaginella moellendorffii 1-1
Selaginella willdenowii
Selaginella kraussiana
Selaginella acanthonota
Selaginella selaginoides

Selaginella
PHOT1

C+

96/+
Selaginella moellendorffii 2-1
Selaginella moellendorffii 2-2
Selaginella willdenowii
Selaginella kraussiana
Selaginella acanthonota

Selaginella
PHOT2

Lycophytes

Isoetes tegetiformans

Isoetes PHOT

84/+
96/+
Dendrolycopodium obscurum
Lycopodium deuterodensum
Diphasiastrum digitatum
Pseudolycopodiella caroliniana
Huperzia lucidula
Phylloglossum drummondii

**Lycopodiales
PHOT**

Liverworts

81/+
Porella pinnata
Radula lindenbergia
Scapania nemorosa
Bazzania trilobata
Schistochila sp
Metzgeria crassipilis
Pellia neesiana

**Liverwort
PHOT**

85/+

96/+
99/+
72/+
Marchantia polymorpha
Conocephalum conicum
Lunularia cruciata
Sphaerocarpos texanus

93/+
Rhynchostegium serrulatum
Neckera douglasii
Loeskeobryum brevirostre
Leucodon brachypus
Orthotrichum lyellii
Bryum argenteum

Bryidae

95/+
91/+
98/+
Scouleria aquatica
Ceratodon purpureus
Aulacomnium heterostichum

Dicranidae

**Bryopsida
PHOT1A**

74/+
82/+
Physcomitrella patens PHOTA2
Physcomitrella patens PHOTB3
Physcomitrella patens PHOTA1

Funariidae

Buxbaumiidae
Buxbaumia aphylla
Atrichum angustatum

**Polytrichopsida
PHOT1A**

PHOT1A

99/+ *E*

PHOT1B

Neckera douglasii
Rhynchostegium serrulatum
Leucodon brachypus
Loeskeobryum brevirostre
Aulacomnium heterostichum
Orthotrichum lyellii
Bryum argenteum

71/.98

Bryidae

**Bryopsida
PHOT1B**

99/+

Funariidae
Physcomitrella patens PHOTA3

93/+
Atrichum angustatum

**Polytrichopsida
PHOT1B**

Sphagnum lescurii

Sphagnopsida PHOT1

89/+
Neckera douglasii
Loeskeobryum brevirostre
Rhynchostegium serrulatum
Leucodon brachypus
Aulacomnium heterostichum
Orthotrichum lyellii
Bryum argenteum

Bryidae

81/+

99/+
Ceratodon purpureus
Scouleria aquatica

**Bryopsida
PHOT2A**

73/.98

Dicranidae

PHOT1

98/+ *D*

PHOT2

PHOT2A

94/+
Leucodon brachypus
Neckera douglasii
Rhynchostegium serrulatum
Orthotrichum lyellii
Scouleria aquatica
Fissidens adianthoides

71/+ *G*

Bryidae

Mosses

PHOT2B

87/+

Dicranidae

**Bryopsida
PHOT2B**

F

2C

Funariidae
Physcomitrella patens PHOTB2

Physcomitrella patens PHOTB1
Physcomitrella patens PHOTA4

Funariidae

**Bryopsida
PHOT2C**

Buxbaumiidae
Buxbaumia aphylla
Atrichum angustatum
Andreaea rupestris

Bryopsida PHOT2
Polytrichopsida PHOT2
Andreaeopsida PHOT2

97/+

Sphagnum lescurii

Sphagnopsida PHOT2

Takakia lepidozioides

Takakiopsida PHOT2

73/+
97/+
Dennstaedtia punctilobula
Plagiogyria distinctissima
Phegopteris hexag
Diplazium wichurae
Allantodia dilatata
Dipteris conjugata

95/+

Fern NEO

95/+
Adiantum capillus veneris
Hemidictyum marginatum
Adiantum tenerum

89/+

Neochrome

Nothoceros aenigmaticus
Megaceros flagellaris
Phaeomegaceros coriaceus
Phymatoceros phymatodes
Paraphymatoceros hallii
Phaeoceros carolinianus
Anthoceros punctatus

87/.99

99/+

Hornwort NEO

95/+

To Fig. 11

99/+
Anthoceros bhardwajii
Anthoceros puncatatus
Megaceros tosanus
Megaceros aenigmaticus
Phaeomegaceros coriaceus
Phymatoceros phymatodes
Phaeoceros carolinianus
Paraphymatoceros hallii

97/+
93/+

Hornworts

99/+

Hornwort PHOT

0.1 substitution/site

37

Figure 10: Phylogeny of lycophyte and bryophyte phototropins. Orange circles indicate inferred photropin duplication events. The italicized capital letter within each circle corresponds to the duplication event mentioned in the text, and the numbers/letters adjacent to each orange circle are the names of the gene duplicates. The values associated with branches are maximum likelihood bootstrap values / Bayesian posterior probabilities. "†" indicates that the exact phylogenetic position of the gene duplication event is ambiguous.

Figure 11: Phylogeny of algal phototropins. Orange circles indicate inferred photropin duplication events. The italicized capital letter within each circle corresponds to the duplication event mentioned in the text, and the numbers/letters adjacent to each orange circle are the names of the gene duplicates. The values associated with branches are maximum likelihood bootstrap values / Bayesian posterior probabilities. "†" indicates that the exact phylogenetic position of the gene duplication event is ambiguous.

39

2.3 Discussions

My phototropin phylogeny refutes the previous assertion that "*PHOT2*" is the ancestral phototropin and that "*PHOT1*" evolved later in seed plants (Galván-Ampudia and Offringa, 2007). The ancestral phototropin is neither "*PHOT1*" nor "*PHOT2*", because their paralogs in seed plants, lycophytes, ferns and mosses were derived from separate gene duplications that are confined to each organismal lineage. In other words, seed-plant *PHOT1* and *PHOT2* are more closely related to one another than to fern *PHOTs* or moss *PHOTs*. My revised gene orthology has important functional and evolutionary implications. Plants often respond differently under low- and high-light levels; chloroplasts, in particular, aggregate under weak light but retreat when the intensity is too high. Consequently, as shown in our phylogenetic reconstruction, phototropin paralogs have repeatedly, and convergently, specialized into mediating either low- or high-light responses, in seed plants, ferns, lycophytes and mosses, although some redundancies do exist (Christie, 2007). Of the two phototropins known in *Arabidopsis thaliana*, Atphot1 mediates phototropism under low-light intensity, and is more sensitive than Atphot2 in triggering chloroplast accumulation (Sakai et al., 2001). Atphot2, in contrast, responds predominantly to high-light intensity, and is solely responsible for chloroplast avoidance under strong light (Kagawa et al., 2001). A similar functional differentiation can also be seen in the fern *Adiantum capillus-veneris* Acphot1 and Acphot2 phototropins. Acphot2 controls chloroplast avoidance under high-light intensity, whereas Acphot1 plays a minor role in this response (Kagawa et al., 2004). Similarly, Kasahara et al. (2004) examined four phototropins in the moss *Physcomitrella patens*, and found that Ppphot1A-2 (see Table 1) is of primary importance in chloroplast avoidance behavior, whereas the others contribute to this response to a much less extent.

To understand how phototropin functional divergences (subfunctionalizations) repeatedly evolved in plants, the key is to reconstruct the function of ancestral phototropin that exists as a single-copy gene. In their recent study of the liverwort *Marchantia polymorpha* phototropin (single copy), Komatsu et al (2014) found that it encompasses all the functional characteristics of both Atphot1 and Atphot2. This finding suggests that the ancestral land plant phototropin was likely a "general-purpose" photoreceptor that responded to a wide range of light intensities. The subsequent and parallel specializations of phototropin into low- and high-light intensity functional responses may have played an important role in the adaptation of early land plants to Earth's changing landscapes. Since the formation of the earliest forests by extinct ferns and horsetails (cladoxylopsids) about 385 million years ago (Stein et al., 2007) through to today's angiosperm-dominated terrestrial ecosystems, light environments have become increasingly heterogeneous and dynamic. Possessing duplicated phototropin genes and co-opting them for different light intensities would be especially beneficial (Galen et al., 2004) and advantageous over the ancestral, general-purpose phototropin. Indeed, most of the land plant lineages that possess duplicated phototropin homologs (seed plants, ferns, lycophytes, and mosses) are more species rich than those that do not (liverworts and hornworts).

The evolutionary history pattern that I observe here for phototropins shows a striking resemblance to that for phytochromes. A similar sequence of convergent evolutionary events—following gene duplication—has also been reported for phytochromes across all major plant lineages (Chapter 1). Both photoreceptors (phytochromes and phototropins) duplicated repeatedly in seed plants, ferns, lycophytes and mosses, while they remained single-copy in liverworts and hornworts (Figure 9). Although this pattern of concerted gene family expansion

41

and stasis could be due to whole genome duplications (WGD), these two photoreceptors differ in the exact evolutionary positions of gene duplication events—they did not all happen along the same phylogenetic branches (Figure 9), suggesting that WGD is not solely responsible. I propose here that there has been a tight co-evolutionary relationship between phototropins and phytochromes. Recent studies have shown that these two photoreceptors not only share cross-talk in their signal transduction pathways (Lariguet et al., 2006; de Carbonnel et al., 2010; Demarsy et al., 2012), but also can physically interact (Jaedicke et al., 2012). In addition, the convergent evolution and horizontal gene transfer of neochromes (Suetsugu et al., 2005; Li et al., 2014 Chapter 3) further illustrate that, throughout plant evolutionary history, a tight partnership has resulted between the two photoreceptors. I hypothesize that the integration of both blue and red/far-red light information enabled plants to respond optimally to changing environments through time. Duplication of one photoreceptor may have prompted duplication in the other, and hence resulted in the rather parallel gene family evolutionary histories.

In summary, here I leveraged the recent surge in genomic and transcriptomic data to identify phototropins from across a broad repertoire of extant biodiversity. My study reveals that phototropins are unique to Viridiplantae, and that gene family expansion and stasis has operated uniquely within each of the various land plant lineages—a pattern similar to that of the phytochrome photoreceptor. Existing functional data for phototropins, interpreted in light of my gene phylogeny, suggests a history of repeated gene duplications followed by parallel functional divergences (subfunctionalizations). Our broad phylogenetic approach greatly complements ongoing photobiology research focused on select plant model organisms, and will enable future

research linking ecology, evolution, and photochemistry to understanding how plants adapt (and have adapted) to variable light environments.

2.4 Materials and Methods

2.4.1 Mining phototropins from transcriptomes and genomes

The transcriptomes and genomes I used are listed in Appendix Table 4, Table 5. To mine phototropin homologs, I used the BlueDevil python pipeline following Li et al (2014) for transcriptomes, and for genomes I used BLASTp implemented in Phytozome (Goodstein et al., 2012) or individual genome portal (Appendix Table 4, Table 5). A phototropin sequence from *Anthoceros bhardwajii* (voucher: Villarreal #6) was obtained by PCR and cloning (primers: photF1970 and photR4102, see Appendix Table 9).

2.4.2 Sequence alignment and phylogenetic reconstruction

I used MUSCLE (Edgar, 2004) to align the amino acid sequences, and then back-translated these to nucleotides. The resulting alignment was manually improved based on known domain and motif boundaries, and unalignable regions were excluded prior to phylogenetic analyses. I used PartitionFinder (Lanfear et al., 2012) to obtain the optimal data partition scheme (by codon position) and the associated nucleotide substitution models. Garli 2.0 (Zwickl, 2006) was employed to find the best maximum likelihood tree with "genthreshfortopoterm" set to 500,000 and 8 independent runs. I carried out bootstrapping to assess branch support, using RAxML (Stamatakis, 2006) with 1,000 replicates. The same partition scheme and models were used in MrBayes 3.2 (Ronquist et al., 2012) Bayesian inference. I carried out two independent MCMC runs, each with four chains and trees sampled every 1,000 generations. I unlinked substitution parameters and set the rate prior to vary among subsets. The resulting MCMC

statistics were inspected in Tracer (Rambaut and Drummond, 2013) to ensure convergence and

proper mixing; 25% of the total generations were discarded as burn-in before compiling the 50%

majority consensus tree. I also carried out phylogenetic reconstruction based on codon models.

CodonPhyML (Gil et al., 2013) was used, with Goldman-Yang codon substitution model

(Goldman and Yang, 1994), empirical codon frequency (F1X61) and three categories of non-

synonymous/synonymous substitution rate ratio.

2.4.3 Target enrichment for confirming phototropin copy number in hornworts

The target enrichment data were from Chapter 1, whereby a hornwort (*Anthoceros*

punctatus) DNA library was hybridized with 7,502 120mer RNA probes to enrich phototropin,

phytochrome and neochrome homologs. The captured fragments were sequenced on one-tenth of

a MiSeq (250bp PE) run. I used sickle (Joshi and Fass, 2011) and cutadapt (Martin, 2011) to clean

and trim the reads, respectively, and assembled using SOAPdenovo (Luo et al., 2012). The

phototropin contigs were identified by tBLASTn.

3. The origin and evolution of neochromes

3.1 Introduction

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 (Möglich et al., 2010). The chimeric photoreceptor, neochrome, is a remarkable exception. It fuses red-sensing phytochrome and blue-sensing phototropin modules into a single molecule (Figure 12A) that mediates phototropic responses (Nozue et al., 1998; Kawai et al., 2003; Kanegae et al., 2006). Neochromes have a restricted occurrence in the plant tree of life, and two independent origins (Suetsugu et al., 2005)—one in the green alga *Mougeotia scalaris* and another in ferns—suggesting that the possession of neochrome may be evolutionarily advantageous. This is consistent with evidence of greatly enhanced phototropic responses in ferns with neochrome (Kawai et al., 2003; Kanegae et al., 2006), as well as its phylogenetic distribution within the fern lineage. The early-diverging fern orders Osmundales and Schizaeales do not possess neochrome (Kawai et al., 2003). It has been reported only in Cyatheales (Yang et al., 2010) and Polypodiales (Kawai et al., 2003; Yang et al., 2010), lineages that mostly diversified following the Cretaceous/Tertiary establishment of low-light, angiosperm-dominated forest canopies (Schneider et al., 2004; Schuettpelz and Pryer, 2009).

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" (Schneider et al., 2004; Schuettpelz and Pryer, 2009; Kawai et al., 2003). Although potentially significant from an evolutionary standpoint, the origin of fern neochrome has remained a mystery, and no previous study has revealed how it might have evolved.

In this study, I investigated the origin of neochrome by searching for homologous sequences in 434 transcriptomes and 40 whole genomes of plants and algae (Appendix Table 6), and surprisingly discovered neochrome homologs from hornworts (Figure 12B, Appendix Table 6). 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 estimations further demonstrate that ferns acquired neochrome from hornworts via horizontal gene transfer (HGT).

3.2 Results and Discussions

3.2.1 Algal neochrome

The only published algal neochrome is from a single species, *Mougeotia scalaris* (Suetsugu et al., 2005). I identified homologs of neochrome in the transcriptomes of all 10 sampled members of the "Zygnemataceae" superclade [*sensu* Goncharov and Melkonian (2010)], including *Mougeotia*, *Mesotaenium*, *Cylindrocystis*, and *Zygnemopsis* but in no other algal transcriptomes surveyed (Figure 12, Appendix Table 6, Figure 18).

46

Figure 12: 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 in **Figure 21**). (C) Portion of the phototropin phylogeny, showing relationships of fern neochrome, hornwort phototropin and neochrome, with highly supported branches thickened (details in **Figure 13**). (D) A schematic depicting the origin of fern neochrome.

3.2.2 Novel neochrome in hornworts

Among land plants, I documented the occurrence of neochrome in 25 additional fern

species (Figure 13, Figure 14). Surprisingly, I also discovered neochrome in hornworts, a small

bryophyte lineage that diverged early in the history of land plants. Although the exact branching

order among the three bryophyte lineages (hornworts, mosses, liverworts) is not resolved with

certainty, some recent analyses have suggested that hornworts are sister to vascular plants

(lycophytes, ferns, and seed plants; Qiu et al., 2006). I empirically 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 out of the five hornwort orders (Dendrocerotales, Phymatocerotales,

Notothyladales, and Anthocerotales). I was unable to obtain adequate material of the monotypic

hornwort *Leiosporoceros* to test for the presence of neochrome in Leiosporocerotales. To confirm

that the hornwort neochrome sequence data were indeed derived from the hornwort nuclear

genome and not from contaminant algae or ferns, I performed genome-walking in *Nothoceros*

aenigmaticus to obtain flanking genomic sequences. Downstream of I found a pseudogene for

imidazoleglycerol-phosphate dehydratase (IGPD) and, because its sequence is most closely

related to other hornwort IGPD genes (Figure 20), I are confident that neochrome is present in the

hornwort genome.

3.2.3 Neochrome HGT from hornworts to ferns

The phylogenetic distribution of neochrome in land plants (present only in hornworts

and ferns) could be explained by 1) an ancient origin along the branch that unites hornworts and

tracheophytes, followed by losses from lycophytes and seed plants, 2) independent origins in

ferns and hornworts, or 3) one or more instances of horizontal gene transfer (HGT) between

hornworts and ferns. To distinguish among these three possible scenarios, I compiled

comprehensive sequence alignments of phototropin and phytochrome from across all land plants

48

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 (Figure 12B and C, Figure 13, Appendix Figure 18, Figure 19). This nested relationship indicates that neochrome was transferred horizontally from hornworts to ferns, along the stem lineage leading to *Phymatoceros + Nothoceros + Megaceros* (Figure 13, Appendix Figure 18, Figure 19). 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.

I used divergence time estimates to further test the HGT hypothesis, reasoning that, in a case of HGT, the split between hornwort and fern neochrome should be significantly younger than the split between the hornwort and fern lineages themselves. By integrating fossil calibrations (Appendix Table 7) with a Bayesian relaxed molecular clock analysis, I estimated the divergence date between hornwort and fern neochrome to be approximately 179 million years ago (MYA) with a 95% highest posterior density interval of 133 and 229 MYA (Figure 12, Appendix Figure 21). This date is far more recent than published divergence estimates between ferns and hornworts (at least 400 MYA; Hedges and Kumar, 2009), but is congruent with the date estimates for the stem branch leading to *Phymatoceros + Nothoceros + Megaceros* (85-244 MYA; Villarreal and Renner, 2012). 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 (Figure 12D), and thus closely resembles the C-terminal end of both fern and hornwort neochromes. I found this intron-free phototropin in all hornworts examined, by using PCR on genomic DNA. All other phototropins characterized to date, including those of ferns, contain more than twenty introns. I explored whether this might be a partial neochrome masquerading as a phototropin by using inverse PCR to obtain the 5′ upstream genomic region in *Nothoceros 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 do not have a canonical phototropin gene. Instead, hornwort phototropins are most closely related to fern and hornwort neochromes (Figure 12, Figure 13, Appendix Figure 19), implying that they likely represent the ancestral, retrotransposed phototropin lineage that gave rise to neochrome through fusion with the phytochrome module (Figure 12D).

3.2.4 Recurrent fern-to-fern HGT

I detected an extraordinary incongruence between my fern neochrome gene tree and the published phylogeny of ferns (Figure 14) (Schuettpelz and Pryer, 2007). By examining the entire Bayesian posterior tree sample, I found that none of the trees resolved neochromes from the same fern family to be monophyletic. This conflicting pattern is not observed in other fern phylogenies based on nuclear genes (Rothfels et al., 2013), and is not seen in the hornwort neochrome tree (Figure 13), which perfectly mirrors the published phylogeny of hornworts (Villarreal and Renner,

2012). Here I examine and discuss the possible causes of the incongruent gene tree/species tree in

ferns.

Figure 13: Phylogenetic relationships of fern neochrome (NEO), hornwort neochrome and phototropin (PHOT). Topology derived from the best maximum likelihood tree. Numbers above branches are maximum likelihood bootstrap values (BS) from Garli/BS from nhPhyML/aLRT SH-like supports under codon model (aLRT-SH)/Bayesian posterior probabilities (PP) from MrBayes/PP from BEAST; these are only displayed (along with thickened branches) when BS > 70, SH-aLRT > 70 and PP > 0.95. "+" denotes BS = 100, aLRT-SH = 100 or PP = 1.00; thickened branches without numbers are "+/+/+/+/+". Alphanumeric codes following species names are the four-letter 1KP transcriptome identifiers, 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.

Figure 14: 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.

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* (Nozue et al., 1998). This was corroborated by the cloning efforts that produced most of my sequence data (Appendix Table 8). Except for *Deparia* spp., where two divergent

sequences were found (Figure 14, arrowheads), I was only able to isolate a single neochrome for each fern species.

Next, I investigated whether an aberrant nucleotide substitution process may have misled the phylogenetic reconstruction. For example, pervasive positive selection or variation in GC content can obscure true phylogenetic signal (Sanderson and Shaffer, 2002; Kapralov and Filatov, 2007; Nabholz et al., 2011), thereby causing a gene tree to be incongruent with the species tree. Using codon models for tree inference can potentially accommodate complex selection profiles, by allowing different nonsynonymous/synonymous substitution rate ratios to fall into distinct classes (Gil et al., 2013). However, I 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 (Figure 15A). Similarly, inferences based on first + second codon positions, as well as on third codon positions only, also yielded topologies discordant with the species tree (Figure 15B,C).

I then used a random effects branch-site model to infer the dynamics of positive selection across the neochrome tree (Kosakovsky Pond et al., 2011). Only five fern branches were identified as having experienced significant episodic positive selection (Figure 15D), 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 (Figure 15E). 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.

Figure 15: 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 (**Figure 13**). 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 my study were deviant in base composition for neochrome. Each window is 400bp in size and the window slides every 50bp.

I therefore hypothesized that the incongruent tree could be the result of 1) multiple fern-to-fern HGT events, 2) an elevated gene turnover rate that may have been selected for after HGT (Lind et al., 2010; Näsvall et al., 2012), or 3) a combination of both. I have some evidence suggesting recurring fern-to-fern HGT might have been involved. For example, I discovered neochrome genes from two early-diverging fern orders [Gleicheniales (*Dipteris conjugata*) and Cyatheales (*Alsophila podophylla* and *Plagiogyria* spp.)] that were likely derived from secondary HGT events (Figure 14, arrows). These neochromes are not phylogenetically resolved as would be predicted based on published fern species relationships (Schuettpelz and Pryer, 2007), but instead are nested among Polypodiales (Figure 14). Furthermore, the split between these and other fern neochromes (81 MYA, 95% highest posterior density interval: 59-106 MYA; Appendix Figure 21) occurred long after the estimated organismal divergence dates for Gleicheniales (276 MYA) and Cyatheales (223 MYA) (Schuettpelz and Pryer, 2009), a pattern that may best be explained by fern-to-fern HGT.

My 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 (Cho et al., 1998; Sanchez-Puerta et al., 2008; 2011). Perhaps neochrome is similarly associated with mobile elements that may have facilitated its movement across species boundaries.

3.2.5 Evolutionary and physiological implications of neochrome in hornworts

My 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 are mediated by molecular interactions between physically separate phytochrome and phototropin proteins (Jaedicke et al., 2012). 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 years ago, Burr (1968) documented an unusual chloroplast photoresponse in *Megaceros* hornworts; she discovered that the large chloroplasts "contract" to form compact shapes under strong light. Although I confirmed this phenomenon in hornworts (Figure 16), future studies are needed to examine if neochrome is responsible for the contraction response and to explore other possible physiological roles.

Figure 16: Hornwort chloroplasts contract under strong light. (*A*) Before irradiation, chloroplasts of *Nothoceros aenigmaticus* (arrowhead) occupy most of the cellular space. (*B*) After irradiation with blue light (57 µmol m^{-2} s^{-1}) for 2 hours, chloroplasts evidently reduced in size. Scale bar = 40 µm.

3.2.6 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 horizontally transferred from hornworts to ferns. The life history of ferns may help to 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. By 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 (Huang, 2013).

To date, most documented examples of plant-to-plant HGT involve mitochondrial DNA and/or parasite-host transfers (Bergthorsson et al., 2003; 2004; Davis and Wurdack, 2004; Davis et al., 2005; Yoshida et al., 2010; Renner and Bellot, 2012; Xi et al., 2013); only a handful of cases include functional nuclear genes (Renner and Bellot, 2012; Xi et al., 2012; Zhang et al., 2013), and even fewer have possible adaptive implications (Christin et al., 2012). 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 (Schneider et al., 2004; Schuettpelz and Pryer, 2009; Kawai et al., 2003), my study has important implications for the macroevolutionary significance of plant-to-plant HGT.

3.3 Materials and Methods

3.3.1 Mining transcriptomes and whole genome sequences for homologs of neochrome, phototropin and phytochrome

All but one of the 434 transcriptomes utilized 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

(Appendix Table 6). Details on RNA extraction, sequencing and assembly for 1KP can be found in Johnson et al (2012). 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 (Grabherr et al., 2011). The sequencing reads were deposited in National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under experiment SRX423244.

For the 1KP transcriptomes, I used both SOAP *de novo* and SOAP *de novo trans* assemblies. For each assembly, a BLAST database was constructed using the BLAST+ package (Camacho et al., 2009). Neochrome, phototropin and phytochrome sequences were separately queried (by tBLASTn for 1KP and BLASTn for *Pteridium* assemblies) and the significant hits to transcriptome scaffolds were extracted. For each scaffold, the best open reading frame was identified, and the sequence was translated into amino acids and then BLASTp against the NCBI non-redundant 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 then merged using CAP3 (Huang, 1999). I carried out the above procedures using my Python pipeline BlueDevil (http://dx.doi.org/10.5061/dryad.fn2rg). I also searched and obtained photoreceptor homologs from 39 plant and algae whole genome sequences through Phytozome (Goodstein et al., 2012) and the *Amborella* Genome Database (http://www.amborella.org).

3.3.2 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

~400bp fragments, and sequenced using Illumina HiSeq2000, giving a total of 25 million 90bp

paired-end reads (about 20X genome coverage). The reads were subjected to two cycles of read

error correction using the ALLPATHS-LG FindError program (MacCallum et al., 2009) before

being assembled using Velvet (Zerbino and Birney, 2008). Assemblies were generated for a range

of kmer values (k=21, 31, 41, 51 and 61) and then combined. The redundant scaffolds were

removed using Usearch (Edgar, 2010) 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 99.5Mb and a N50 of 4955bp. Contig length

ranges from 919bp to 76.5kb, with 1643 contigs over 10kb. The median and mean assembled

contig coverage is 18.1X and 44X, respectively. The raw reads were deposited in NCBI under

SRA096687.

This assembly was searched for homologs of neochrome, phototropin and phytochrome

using translated BLAST searches. While phototropin and phytochrome genes were readily

identified, no contig was found containing putative neochrome sequence. To search for the *A.*

punctatus neochrome gene that failed to be assembled, all sequencing reads were searched against

a library of neochrome protein sequences using BLASTx. Reads obtaining an e-value of $\leq 1 \times 10^{-10}$

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 5 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 were then mapped to these "seed" contigs using Bowtie2 (Langmead and Salzberg, 2012) 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 were then re-assembled 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 797bp fragment of neochrome. This fragment was then extended to include almost the entire open reading frame using a combination of PCR (see below) and additional read mapping and assembly.

3.3.3 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, I cloned the genes from genomic DNA from five hornwort species (Appendix Table 8). In addition, neochrome sequences were obtained from 25 fern species by PCR and cloning (Appendix Table 8). Genomic DNA was extracted using 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 Appendix Table 8 and Table 9. The amplified products were cloned into Promega pGEM-T (Promega) and sequenced.

3.3.4 Genome walking in hornwort phototropin and neochrome

To rule out that the phototropin gene found in hornworts might be a partial neochrome, I used inverse PCR (Ochman et al., 1988) to obtain the flanking genomic region. Genomic DNA of *Nothoceros aenigmaticus* was digested by apoI (New England Biolabs) and self-ligated using T4 DNA ligase (New England Biolabs). Nested PCRs were then conducted on the circularized DNA.

60

The amplicons were cloned using Promega pGEM-T and sequenced. To search for the genes

flanking neochrome in *N. aenigmaticus*, I used the Clontech GenomeWalker kit (Clontech) and

followed the manufacturer's manual. The resulting PCR amplicons were cloned and sequenced.

In total, I obtained 3291 bp and 4578 bp regions up- and down-stream, respectively, of

neochrome. The primers for the above PCR reactions are listed in Appendix Table 9.

3.3.5 Sequence alignment for neochrome, phototropin and phytochrome

I 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, I only included 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 (Marchler-Bauer et al., 2011). Each

domain was separately aligned (based on the amino acid sequences) using Muscle (Edgar, 2004),

and then concatenated. I developed a Python script, DomainDivider

(http://dx.doi.org/10.5061/dryad.fn2rg), to automate these processes. I also generated a separate

alignment for hornwort and fern neochromes. This alignment was based on entire neochrome

sequences rather than domains. All alignments were manually inspected and ambiguously

aligned regions were excluded prior to 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 Figure 13, Figure 14, Appendix Figure 18, Figure 19.

3.3.6 Phylogenetic analyses of phototropin and neochrome

Phototropin and neochrome phylogenies were inferred based on their nucleotide alignments. I used PartitionFinder (Lanfear et al., 2012) 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, 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. I used Garli (Zwickl, 2006) to obtain the maximum likelihood tree under the aforementioned models, with genthreshfortopoterm set to 1,000,000 and 8 independent runs. Multiparametric bootstrapping was done using RAxML (Stamatakis, 2006) with 1000 replicates. For the neochrome alignment, I also carried out the same maximum likelihood analyses on the first + second codon positions, as well as on the third codon positions separately. I used MrBayes (Ronquist et al., 2012) to conduct Bayesian tree inference under the same models, with two independent MCMC runs, four chains each, and trees sampled every 1000 generations. Substitution parameters were unlinked and the rate prior was set to vary among partitions. The MrBayes output was inspected using Tracer (Rambaut and Drummond, 2013) to ensure proper convergence and mixing (effective sample sizes all > 200), and 25% of the total generations were discarded as burn-in prior to making the 50% majority consensus tree. Because the stationary, homogeneous assumptions of GTR might be violated in cases associated with HGT and deep divergence (Verbyla et al., 2013), I also employed a non-stationary, heterogeneous nucleotide substitution model implemented in nhPhyML (Boussau and Gouy, 2006) to infer the phototropin tree. The analysis was carried out with ten discrete categories of GC equilibrium frequencies, and the required starting tree was the best tree

from the Garli analysis. To conduct bootstrapping in nhPhyML, I created a Python wrapper, and

for each replicate, RAxML was used to input the starting tree. In addition to the nucleotide

substitution model, I also used codon models to infer phylogenies, which were carried out in

CodonPhyML (Gil et al., 2013) under a maximum likelihood framework. I used the GY (Goldman

and Yang, 1994) model with four categories of non-synonymous/synonymous substitution rate

ratios drawn from the discrete gamma distribution, and codon frequencies were estimated from

the data under the F3X4 model (Gil et al., 2013). The tree topology search was done using the

NNI approach, and branch support was estimated using the SH-like aLRT (Anisimova and

Gascuel, 2006; Guindon et al., 2010) method.

3.3.7 Phylogenetic analyses of phytochrome

For the phytochrome phylogeny, I used the protein alignment following the analytical

strategy of Mathews et al (2010). Using ProtTest (Abascal et al., 2005), JTT + F was found to be the

best empirical substitution model under the Akaike Information Criterion. For the maximum

likelihood analyses, I used Garli to search for the maximum likelihood tree, with

genthreshfortopoterm set to 1,000,000 and 8 independent runs, and RAxML to conduct the

multiparametric bootstrapping with 1000 replicates. For Bayesian tree inference, I used MrBayes

with two independent MCMC runs, four chains each, and trees sampled every 1000 generations.

After removing 25% of the total generations, the 50% majority consensus tree was calculated.

Codon-based tree inference was also carried out as described above.

3.3.8 Topology test

I used the SOWH test (Goldman et al., 2000) 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 (Church et al.)

with RAxML and Seq-Gen (Rambaut and Grassly, 1997). 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, I filtered the posterior tree samples from MrBayes and calculated the

frequency of trees given the monophyly constraints. The filtering was done by PAUP* (Swofford,

2002). I also applied this same approach to examine the posterior distribution of fern neochrome

gene trees. I searched for topologies that exhibited better congruence with the published species

relationships (as compared to the inferred gene tree). The constraint for tree filtering required

that neochromes from the same fern family be monophyletic.

3.3.9 Phylogenetic analysis of imidazoleglycerol-phosphate dehydratase gene (IGPD)

As a result of genome-walking in *Nothoceros aenigmaticus*, I discovered an IGPD

pseudogene downstream from neochrome. To place this pseudogene in phylogenetic context, I

resolved the first IGPD phylogeny for land plants. A subset of the transcriptomes and whole

genome sequences was mined for IGPD homologs (Fig. S4) using BlueDevil, and an alignment of

IGPD was manually constructed. I 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 1000 replicates.

3.3.10 Divergence time estimation of the phototropin gene family

I used BEAST (Drummond et al., 2012) to simultaneously infer 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 tmrca priors were employed as the calibration points (see Appendix Table 4; Clarke et al., 2011; Hubers and Kerp, 2012; Guo et al., 2012; Villarreal and Renner, 2012; Schuettpelz and Pryer, 2009; Prasad et al., 2005; Kotyk et al., 2002; Skog and Banks, 1973; Galtier et al., 2001; Trivett, 1992), and a birth-death speciation prior was used as the tree prior. I used the uncorrelated relaxed-clock model with rates drawn from a lognormal distribution. A starting tree was first estimated by r8s (Sanderson, 2003) and 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 (Drummond et al., 2012) with a 25% burn-in and summarized in TreeAnnotator (Drummond et al., 2012). 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 non-stationary, heterogeneous model that is currently implemented in divergence time analyses, and my results should be revisited in the future when more sophisticated methods are available.

3.3.11 Inferring episodic selection and GC content variation in neochrome evolution

To investigate whether fern neochromes had experienced pervasive episodic positive selection, I used the unrestricted, random effects branch-site model (Kosakovsky Pond et al., 2011) implemented in the HyPhy package (Kosakovsky Pond et al., 2011; 2005). Branches with

episodic positive selection were identified by the sequential likelihood ratio test (Kosakovsky

Pond et al., 2011). The neochrome alignment and the best maximum likelihood tree were used as

the input data. The analyses were carried out on the Datamonkey server (Delport et al., 2010;

Kosakovsky Pond et al., 2005). A GC content sliding window was constructed using a custom

Python script; each window is 400bp in size and the window slides every 50bp.

Appendix A: Supplementary Figures for Chapter One

Figure 17: The phylogeny of phytochromes reconstructed from 423 protein sequences. Topology derived from the best maximum likelihood tree. The values associated with branches are maximum likelihood bootstrap support. For better visualization, the complete tree is broken down into six subtrees (A-F) and scaled differently. The inset at the top-left corner shows the map to each subtree. (A) The seed plant phytochrome subtree. Phylogeny continues in B.

Figure 17: Continued. (B) Fern and lycophyte phytochrome subtree.

A. seed plant PHY

B. fern + lycophyte PHY

C. bryophyte PHY

D. neochrome

E. green algae PHY

F. prokaryote +
stramenopile + fungi PHY

**To B
(ferns + lycophytes)**

61 Claopodium rostratum E
Rhynchostegium serrulatum E
Loeskeobryum brevirostre E
74 60 Neckera douglasii E
57 Anomodon attenuatus E
54 Pseudotaxiphyllum elegans E
Leucodon julaceus E
55 Aulacomnium heterostichum E 2
96 Aulacomnium heterostichum E 1
63 Hedwigia ciliata E
Philonotis fontana E
Scouleria aquatica E
Racomitrium varium E
62 Syntrichia princeps E
Ceratodon purpureus "PHY4"
Leucodon brachypus E
92 Physcomitrella patens C
Physcomitrella patens B
61 Rhynchostegium serrulatum D
88 Claopodium rostratum D
51 Anomodon attenuatus D
69 Loeskeobryum brevirostre D
57 Pseudotaxiphyllum elegans D
52 Leucodon brachypus D 1
Leucodon brachypus D 2
Aulacomnium heterostichum D
Hedwigia ciliata D
Philonotis fontana D
Bryum argenteum D
81 Ceratodon purpureus "PHY3"
77 Syntrichia princeps D
61 Leucodon brachypus D
Racomitrium varium D
98 Scouleria aquatica D
Physcomitrella patens A
75
98
Atrichum angustatum
95 Claopodium rostratum
63 Rhynchostegium serrulatum
Stereodon subimponens
Leucodon brachypus
91 70 Leucodon julaceus
Anomodon attenuatus
99 96 Neckera douglasii
61 Loeskeobryum brevirostre
52 Pseudotaxiphyllum elegans
Aulacomnium heterostichum
74 Bryum argenteum
68 Hedwigia ciliata
77 Philonotis fontana
56 Scouleria aquatica
100 Physcomitrella patens 4
99 Physcomitrella patens 2
60 Leucobryum albidum
99 Racomitrium varium
Ceratodon purpureus
Atrichum angustatum
81
94
91
86 Andreaea rupestris
Sphagnum palustre
Takakia lepidozioides
100 Leucodon brachypus
Leucodon julaceus
Pseudotaxiphyllum elegans
Neckera douglasii
Claopodium rostratum
100 Rhynchostegium serrulatum
94 Stereodon subimponens
Loeskeobryum brevirostre
61 Aulacomnium heterostichum
100 Hedwigia ciliata
68 Bryum argenteum
Philonotis fontana
88 100 Ceratodon purpureus "PHY2"
Ceratodon purpureus "PHY1"
94 83 Leucobryum albidum
Racomitrium varium
98 Physcomitrella patens 1
Physcomitrella patens 3
100
93
91 Andreaea rupestris
98 Scapania nemorosa
63 Odontoschisma prostratum
64 Bazzania trilobata
Porella pinnata
85 Metzgeria crassipilis
Schistochila sp
95 Pellia neesiana
96 72 Conocephalum conicum
90 Ricciocarpos natans
100 Marchantia paleacea
Sphaerocarpos texanus
77 Lunularia cruciata
87 Paraphymatoceros hallii
54 Phaeoceros carolinianus
80 Megaceros flagellaris
63 Nothoceros aenigmaticus
Anthoceros punctatus
73

Moss PHY5

Moss PHY2_4

Moss PHY2_4/5

Moss PHY1_3

Liverwort PHY

Hornwort PHY

**To E
(green algae)**

**To Fig. 1D
(neochrome)**

0.1 substitutions/site

Figure 17: Continued. (C) Bryophyte phytochrome subtree.

69

Figure 17: Continued. (D) Neochrome subtree.

Figure 17: Continued. (E) Green algae phytochrome subtree.

Figure 17: Continued. (F) Prokaryote + stramenopile + fungi phytochrome subtree.

Appendix B: Supplementary Tables for Chapter One

Table 2: List of transcriptomes and genomes screened for phytochromes. "-": not applicable.

Taxa	Source	1KP[3] identififer
Angiosperm		
Amborella trichocarpa	Genome, Amborella Genome Project[1]	-
Aquilegia coerulea	Genome, Phytozome[2]	-
Arabidopsis thaliana	Genome, Phytozome	-
Buxus sempervirens	Transcriptome, 1KP[3]	IWMW
Capnoides sempervirens	Transcriptome, 1KP	AUGV
Cucumis sativus	Genome, Phytozome	-
Fragaria vesca	Genome, Phytozome	-
Illicium floridanum	Transcriptome, 1KP	VZCI
Laurelia sempervirens	Transcriptome, 1KP	WAIL
Oryza sativa	Genome, Phytozome	-
Peperomia fraseri	Transcriptome, 1KP	XSZI
Phaseolus vulgaris	Genome, Phytozome	-
Pistia stratioides	Transcriptome, 1KP	MFIN
Platanthera clavellata	Transcriptome, 1KP	MTHW
Solanum lycopersicum	Genome, Phytozome	-
Vitis vinifera	Genome, Phytozome	-
Gymnosperm		
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ
Encephalartos barteri	Transcriptome, 1KP	GNQG
Ginkgo biloba	Transcriptome, 1KP	SGTW
Gnetum montanum	Transcriptome, 1KP	GTHK
Pinus parviflora	Transcriptome, 1KP	IIOL
Podocarpus rubens	Transcriptome, 1KP	XLGK
Taiwania cryptomerioides	Transcriptome, 1KP	QSNJ
Thuja plicata	Transcriptome, 1KP	VFYZ
Welwitschia mirabilis	Transcriptome, 1KP	TOXE
Fern		
Adiantum tenerum	Transcriptome, 1KP	BMJR
Anemia tomentosa	Transcriptome, 1KP	CQPW
Argyrochosma nivea	Transcriptome, 1KP	XDDT
Asplenium platyneuron	Transcriptome, 1KP	KJZG
Athyrium filix-femina	Transcriptome, 1KP	AFPO
Azolla caroliniana	Transcriptome, 1KP	CVEG
Azolla filiculoides	Transcriptome, Brouwer et al.[4]	-
Blechnum spicant	Transcriptome, 1KP	VITX
Botrypus virginianus	Transcriptome, 1KP	BEGM
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW
Gaga arizonica	Transcriptome, 1KP	DCDT
Cryptogramma acrostichoides	Transcriptome, 1KP	WQML
Culcita macrocarpa	Transcriptome, 1KP	PNZO
Alsophila spinulosa	Transcriptome, 1KP	GANB
Cystopteris fragilis	Transcriptome, 1KP	LHLE
Cystopteris protrusa	Transcriptome, 1KP	YOWV
Danaea nodosa	Transcriptome, 1KP	DFHO
Deparia lobato-crenata	Transcriptome, 1KP	FCHS
Diplazium wichurae	Transcriptome, 1KP	UFJN
Dipteris conjugata	Transcriptome, 1KP	MEKP
Equisetum diffusum	Transcriptome, 1KP	CAPN
Gymnocarpium dryopteris	Transcriptome, 1KP	HEGQ
Homalosorus pycnocarpos	Transcriptome, 1KP	OCZL
Leucostegia immersa	Transcriptome, 1KP	WGTU
Lindsaea microphylla	Transcriptome, 1KP	YIXP
Lonchitis hirsuta	Transcriptome, 1KP	VVRN
Lygodium japonicum	Transcriptome, 1KP	PBUU
Marattia attenuata	Transcriptome, 1KP	UGNK
Myriopteris rufa	Transcriptome, 1KP	GSXD
Notholaena montieliae	Transcriptome, 1KP	YCKE
Osmunda javanica	Transcriptome, 1KP	VIBO
Osmunda sp	Transcriptome, 1KP	UOMY
Osmundastrum cinnamomeum	Transcriptome, 1KP	RFMZ
Pilularia globulifera	Transcriptome, 1KP	KIIX
Pityrogramma trifoliata	Transcriptome, 1KP	UJTT
Plagiogyria japonica	Transcriptome, 1KP	UWOD
Polypodium glycyrrhiza	Transcriptome, 1KP	CJNT
Polypodium hesperium	Transcriptome, 1KP	GYFU
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ
Pteridium aquilinum	Transcriptome, Der et al.[5]	-
Pteris ensiformis	Transcriptome, 1KP	FLTD
Sceptridium dissectum	Transcriptome, 1KP	EEAQ
Thyrsopteris elegans	Transcriptome, 1KP	EWXK
Vittaria appalachiana	Transcriptome, 1KP	NDUV
Woodsia ilvensis	Transcriptome, 1KP	YQEC
Woodsia scopulina	Transcriptome, 1KP	YJJY
Isoetopsida		
Selaginella cf pallescens	Transcriptome, 1KP	ABIJ
Selaginella moellendorffii	Genome, Phytozome	-
Selaginella willdenowii	Transcriptome, 1KP	KJYC
Selaginella kraussiana	Transcriptome, 1KP	ZFGK
Selaginella wallacei	Transcriptome, 1KP	JKAA
Selaginella apoda	Transcriptome, 1KP	LGDQ
Isoetes tegetiformans	Transcriptome, 1KP	PKOX
Lycopodiales		
Dendrolycopodium obscurum	Transcriptome, 1KP	XNXF
Huperzia selago	Transcriptome, 1KP	GTUO
Huperzia lucidula	Transcriptome, 1KP	GKAG
Pseudolycopodiella caroliniana	Transcriptome, 1KP	UPMJ
Moss		
Andreaea rupestris	Transcriptome, 1KP	WOGB
Anomodon attenuatus	Transcriptome, 1KP	QMWB
Claopodium rostratum	Transcriptome, 1KP	VBMM

Taxa	Source	1KP[3] identififer
Trebouxiophyceae		
Botryococcus braunii	Transcriptome, 1KP	ETGN
Botryococcus terribilis	Transcriptome, 1KP	QYXY
Coccomyxa pringsheimii	Transcriptome, 1KP	GXBM
Eremosphaera viridis	Transcriptome, 1KP	MNCB
Geminella sp.	Transcriptome, 1KP	PFUD
Leptosira obovata	Transcriptome, 1KP	ZNUM
Microthamnion kuetzigianum	Transcriptome, 1KP	DXNY
Nannochloris atomus	Transcriptome, 1KP	MFYC
Neochloris oleoabundans	Transcriptome, 1KP	EEJO
Parachlorella kessleri	Transcriptome, 1KP	AKCR
Prasiola crispa	Transcriptome, 1KP	WCLV
Prototheca wickerhamii	Transcriptome, 1KP	BILC
Stichococcus bacillaris	Transcriptome, 1KP	WXRI
Trebouxia arboricola	Transcriptome, 1KP	NKXU
Ulvophyceae		
Acrosiphonia sp.	Transcriptome, 1KP	JIWJ
Blastophysa cf. rhizopus	Transcriptome, 1KP	VHIJ
Bolbocoleon piliferum	Transcriptome, 1KP	LSHT
Bryopsis plumosa	Transcriptome, 1KP	JTIG
Cephaleuros virescens	Transcriptome, 1KP	YDCQ
Cladophora glomerata	Transcriptome, 1KP	VBLH
Codium fragile	Transcriptome, 1KP	GYBH
Desmochloris halophila	Transcriptome, 1KP	KSFK
Entocladia endozoica	Transcriptome, 1KP	OQON
Halochlorococcum marinum	Transcriptome, 1KP	ALZF
Helicodictyon planctonicum	Transcriptome, 1KP	AJAU
Ignatius tetrasporus	Transcriptome, 1KP	KADG
Ochlochaete sp.	Transcriptome, 1KP	CQQP
Oltmannsiellopsis viridis	Transcriptome, 1KP	PZBH
Oltmannsiellopsis viridis	Transcriptome, 1KP	QJYX
Percursaria percursa	Transcriptome, 1KP	OAEZ
Planophila laetevirens	Transcriptome, 1KP	CBNG
Planophila sp.	Transcriptome, 1KP	LETF
Pirula salina	Transcriptome, 1KP	NQYP
Pseudoneochloris marina	Transcriptome, 1KP	GJIY
Trentepohlia annulata	Transcriptome, 1KP	NATT
Chlorophyceae		
Ankistrodesmus sp.	Transcriptome, 1KP	OTQG
Aphanochaete repens	Transcriptome, 1KP	IJMT
Asteromonas gracilis	Transcriptome, 1KP	NTLE
Botryosphaerella sudetica	Transcriptome, 1KP	VJDZ
Brachiomonas submarina	Transcriptome, 1KP	GUBD
Carteria crucifera	Transcriptome, 1KP	VIAU
Carteria obtusa	Transcriptome, 1KP	RUIF
Chaetopeltis orbicularis	Transcriptome, 1KP	BAZF
Chlamydomonas reinhardtii	Genome, Phytozome	-
Chlamydomonas bilatus	Transcriptome, 1KP	MULF
Chlamydomonas cribrum	Transcriptome, 1KP	BCYF
Chlamydomonas moewusii	Transcriptome, 1KP	JRGZ
Chlamydomonas noctigama	Transcriptome, 1KP	VALZ
Chlamydomonas sp.	Transcriptome, 1KP	TSBQ
Chlamydomonas sp.	Transcriptome, 1KP	AOUJ
Chlorella variabilis	Genome, Blanc et al.[7]	-
Chloromonas oogama	Transcriptome, 1KP	IHOI
Chloromonas perforata	Transcriptome, 1KP	QRTH
Chloromonas reticulata	Transcriptome, 1KP	LBRP
Chloromonas rosae	Transcriptome, 1KP	AJUW
Chloromonas subdivisa	Transcriptome, 1KP	GFUR
Chloromonas tughillensis	Transcriptome, 1KP	UTRE
Cylindrocapsa geminella	Transcriptome, 1KP	DZPJ
Dunaliella salina	Transcriptome, 1KP	RHVC
Dunaliella tertiolecta	Transcriptome, 1KP	ZDIZ
Eudorina elegans	Transcriptome, 1KP	RNAT
Fritschiella tuberosa	Transcriptome, 1KP	VFIV
Golenkinia longispicula	Transcriptome, 1KP	BZSH
Gonium pectorale	Transcriptome, 1KP	KUJU
Haematococcus pluvialis	Transcriptome, 1KP	ODXI
Haematococcus pluvialis	Transcriptome, 1KP	AGIO
Hafniomonas reticulata	Transcriptome, 1KP	FXHG
Heterochlamydomonas inaequalis	Transcriptome, 1KP	IRYH
Lobochlamys segnis	Transcriptome, 1KP	OFUE
Lobomonas rostrata	Transcriptome, 1KP	JKKI
Microspora cf. tumidula	Transcriptome, 1KP	FOYQ
Neodesmus pupukensis	Transcriptome, 1KP	MWAN
Neochlorosarcina sp.	Transcriptome, 1KP	USIX
Oedogonium cardiacum	Transcriptome, 1KP	DVYE
Oedogonium foveolatum	Transcriptome, 1KP	SDPC
Oogamochlamys gigantea	Transcriptome, 1KP	XDLL
Pandorina morum	Transcriptome, 1KP	RYJX
Pediastrum duplex	Transcriptome, 1KP	XKWQ
Pediastrum duplex	Transcriptome, 1KP	XTON
Phacotus lenticularis	Transcriptome, 1KP	ZIVZ
Pleurastrum insigne	Transcriptome, 1KP	PRIQ
Pteromonas angulosa	Transcriptome, 1KP	LNIL
Pteromonas sp.	Transcriptome, 1KP	ACRY
Scenedesmus dimorphus	Transcriptome, 1KP	PZIF
Spermatozopsis exsultans	Transcriptome, 1KP	MXDS
Spermatozopsis similis	Transcriptome, 1KP	ENAU
Stephanosphaera pluvialis	Transcriptome, 1KP	ZLQE
Stigeoclonium helveticum	Transcriptome, 1KP	JMUI

Species	Source	Code
Atrichum angustatum	Transcriptome, 1KP	ZTHV
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH
Bryum argenteum	Transcriptome, 1KP	JMXW
Ceratodon purpureus	Transcriptome, 1KP	FFPD
Hedwigia ciliata	Transcriptome, 1KP	YWNF
Stereodon subimponens	Transcriptome, 1KP	LNSF
Leucobryum albidum	Transcriptome, 1KP	VMXJ
Leucodon brachypus	Transcriptome, 1KP	ZACW
Neckera douglasii	Transcriptome, 1KP	TMAJ
Philonotis fontana	Transcriptome, 1KP	ORKS
Physcomitrella patens	Genome, Phytozome	-
Pseudotaxiphyllum elegans	Transcriptome, 1KP	QKQO
Racomitrium varium	Transcriptome, 1KP	RDOO
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM
Leucodon julaceus	Transcriptome, 1KP	IGUH
Scouleria aquatica	Transcriptome, 1KP	BPSG
Sphagnum palustre	Transcriptome, 1KP	RCBT
Syntrichia princeps	Transcriptome, 1KP	GRKU
Takakia lepidozioides	Transcriptome, 1KP	SKQD
Liverwort		
Scapania nemorosa	Transcriptome, 1KP	IRBN
Porella pinnata	Transcriptome, 1KP	UUHD
Schistochila sp	Transcriptome, 1KP	LGOW
Metzgeria crassipilis	Transcriptome, 1KP	NRWZ
Pellia neesiana	Transcriptome, 1KP	JHFI
Odontoschisma prostratum	Transcriptome, 1KP	YBQN
Conocephalum conicum	Transcriptome, 1KP	ILBQ
Lunularia cruciata	Transcriptome, 1KP	TXVB
Sphaerocarpos texanus	Transcriptome, 1KP	HERT
Ricciocarpos natans	Transcriptome, 1KP	WJLO
Bazzania trilobata	Transcriptome, 1KP	WZYK
Hornwort		
Nothoceros aenigmaticus	Transcriptome, 1KP	DXOU
Megaceros flagellaris	Transcriptome, 1KP	UCRN
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB
Anthoceros punctatus	Genome, Li et al.[6]	-
Phaeomegaceros coriaceus	Transcriptome, 1KP	AKXB
Leiosporoceros dussii	Transcriptome, 1KP	ANON
Desmidiales		
Cosmarium granatum	Transcriptome, 1KP	MNNM
Cosmarium subtumidum	Transcriptome, 1KP	WDGV
Desmidium aptogonum	Transcriptome, 1KP	DFDS
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW
Onychonema laeve	Transcriptome, 1KP	GGWH
Penium exiguum	Transcriptome, 1KP	YSQT
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV
Planotaenium ohtanii	Transcriptome, 1KP	SNOX
Roya obtusa	Transcriptome, 1KP	XRTZ
Staurodesmus convergens	Transcriptome, 1KP	WCQU
Zygnematales		
Cylindrocystis brebissonii	Transcriptome, 1KP	YOXI
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ
Cylindrocystis sp	Transcriptome, 1KP	VAZE
Mesotaenium braunii	Transcriptome, 1KP	WSJO
Mesotaenium kramstei	Transcriptome, 1KP	NBYP
Netrium digitus	Transcriptome, 1KP	FFGR
Zygnemopsis sp	Transcriptome, 1KP	MFZO
Coleochaetales		
Coleochaete irregularis	Transcriptome, 1KP	QPDY
Chaetosphaeridium globosum	Transcriptome, 1KP	DRGY
Klebsormidiales		
Interfilum paradoxum	Transcriptome, 1KP	FPCO
Klebsormidium subtile	Transcriptome, 1KP	FQLP
Entransia fimbriata	Transcriptome, 1KP	BFIK
Mesostigmatales		
Chlorokybus atmophyticus	Transcriptome, 1KP	AZZW
Mesostigma viride	Transcriptome, 1KP	KYIO
Spirotaenia minuta	Transcriptome, 1KP	NNHQ
Charales		
Chara vulgaris	Transcriptome, 1KP	MWXT

Species	Source	Code
Uronema sp.	Transcriptome, 1KP	ISGT
Uronema belkae	Transcriptome, 1KP	RAWF
Vitreochlamys sp.	Transcriptome, 1KP	QWRA
Volvox carteri	Genome, Phytozome	-
Volvox aureus	Transcriptome, 1KP	JWGT
Volvox aureus	Transcriptome, 1KP	WRSL
Volvox globator	Transcriptome, 1KP	ISPU
Pedinophyceae		
Pedinomonas minor	Transcriptome, 1KP	RRSV
Pedinomonas tuberculata	Transcriptome, 1KP	PUAN
Red Algae		
Betaphycus gelatinae	Transcriptome, 1KP	BWVJ
Ceramium kondoi	Transcriptome, 1KP	VZWX
Chondrus crispus	Genome, Collen et al.[8]	-
Chondrus crispus	Transcriptome, 1KP	UGPM
Chroodactylon ornatum	Transcriptome, 1KP	LLXJ
Cyanidioschyzon merolae	Genome, Matsuzaki et al.[9]	-
Dumontia simplex	Transcriptome, 1KP	IEHF
Eucheuma denticulatum	Transcriptome, 1KP	JEBK
Galdieria sulphuraria	Genome, Barbier et al.[10]	-
Glaucosphaera vacuolata	Transcriptome, 1KP	RSOF
Gloeopeltis furcata	Transcriptome, 1KP	SBLT
Gracilaria asiatica	Transcriptome, 1KP	VNAL
Gracilaria blodgettii	Transcriptome, 1KP	LIPN
Gracilaria chouae	Transcriptome, 1KP	FTRP
Gracilaria lemaneiformi	Transcriptome, 1KP	IKWM
Grateloupia filicina	Transcriptome, 1KP	ZJOJ
Grateloupia livida	Transcriptome, 1KP	IKIZ
Grateloupia turuturu	Transcriptome, 1KP	URSB
Grateloupia chiangii	Transcriptome, 1KP	PWKQ
Gymnogongrus ftabelliformis	Transcriptome, 1KP	CKXF
Heterosiphonia pulchra	Transcriptome, 1KP	YSBD
Kappaphycus alvarezii	Transcriptome, 1KP	IHJY
Mazzaella japonica	Transcriptome, 1KP	WEJN
Polysiphonia japonica	Transcriptome, 1KP	XAXW
Pyropia yezoensis	Genome, Nakamura et al.[11]	-
Porphyra yezoensis	Transcriptome, 1KP	ZULJ
Porphyridium cruentum	Transcriptome, 1KP	OBUY
Porphyridium purpureum	Genome, Bhattacharya et al.[12]	-
Porphyridium purpureum	Transcriptome, 1KP	PVGP
Rhodella violacea	Transcriptome, 1KP	RTLC
Rhodochaete parvula	Transcriptome, 1KP	JJZR
Sinotubimorpha guangdongensis	Transcriptome, 1KP	PYDB
Symphyocladia latiuscula	Transcriptome, 1KP	UYFR
Prasinophyte		
Monomastix opisthostigma	Transcriptome, 1KP	BTFM
Pyramimonas parkeae	Transcriptome, 1KP	TNAW
Tetraselmis cordiformis	Transcriptome, 1KP	DUMA
Cryptophyte		
Chroomonas sp	Transcriptome, 1KP	ROZZ
Cryptomonas curvata	Transcriptome, 1KP	BAKF
Guillardia theta	Genome, Curtis et al.[13]	-
Hemiselmis virescens	Transcriptome, 1KP	MJMQ
Rhodomonas sp	Transcriptome, 1KP	IAYV
Glaucophyte		
Cyanophora paradoxa	Genome, Price et al.[14]	-
Cyanoptyche gloeocystis	Transcriptome, 1KP	JKHA
Glaucocystis nostochinearum	Transcriptome, 1KP	POOW
Haptophytes		
Isochrysis sp.	Transcriptome, 1KP	BAJW
Pavlova lutheri	Transcriptome, 1KP	NMAK
Prymnesium parvum	Transcriptome, 1KP	LXRN
Stramenopiles - Chrysophytes		
Mallomonas sp.	Transcriptome, 1KP	BOGT
Ochromonas sp.	Transcriptome, 1KP	EBWI
Synura petersenii	Transcriptome, 1KP	DBYD
Synura sp.	Transcriptome, 1KP	VKVG
Stramenopiles - Brown algae		
Ishige okamurai	Transcriptome, 1KP	APTP
Petalonia fascia	Transcriptome, 1KP	VRGZ
Scytosiphon lomentaria	Transcriptome, 1KP	JCXF

[1]Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **342,** 1241089 (2013).

[2]Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40,** D1178–86 (2012).

[3]Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**, 17 (2014).

[4]Brouwer, P. *et al. Azolla* domestication towards a biobased economy? *New Phytol* **202,** 1069–1082 (2014).

[5]Der, J. P., Barker, M. S., Wickett, N. J., Depamphilis, C. W. & Wolf P. G. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12,** 99 (2011).

[6]Li, F.-W. *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. USA* **111,** 6672–6677 (2014).

[7]Blanc, G. *et al.* The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell* **22,** 2943–2955 (2010).

[8]Collen, J. et al. Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* **110,** 5247–5252 (2013).

[9]Matsuzaki, M. et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428,** 653–657 (2004).

[10]Barbier, G. *et al.* Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of Galdieria sulphuraria and significant differences in carbohydrate metabolism of both algae. *Plant Physiology* **137,** 460–474 (2005).

[11]Nakamura, Y. et al. The first symbiont-free genome sequence of marine red alga, Susabi-nori (*Pyropia yezoensis*). *PLoS ONE* **8,** e57122 (2013).

[12]Bhattacharya, D. *et al*. Genome of the red alga *Porphyridium purpureum*. *Nat Comms* **4**, 1941 (2013).

[13]Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65 (2012).

[14]Price, D. C. *et al. Cyanophora paradoxa ge*nome elucidates origin of photosynthesis in algae and plants. *Science* **335,** 843–847 (2012).

Table 3: Sources and GenBank accession numbers of the phytochromes used in phylogenetic analyses. "-": not applicable.

Phytochrome	Source	1KP[2] identififer	GenBank accession
Angiosperm PHYA			
Arabidopsis thaliana	Genome, Phytozome[1]	-	-
Fragaria vesca	Genome, Phytozome	-	-
Phaseolus vulgaris 1	Genome, Phytozome	-	-
Phaseolus vulgaris 2	Genome, Phytozome	-	-
Cucumis sativus 1	Genome, Phytozome	-	-
Cucumis sativus 2	Genome, Phytozome	-	-
Vitis vinifera	Genome, Phytozome	-	-
Solanum lycopersicum	Genome, Phytozome	-	-
Buxus sempervirens	Transcriptome, 1KP[2]	IWMW	XXXXXXXX
Aquilegia coerulea	Genome, Phytozome	-	-
Oryza sativa	Genome, Phytozome	-	-
Platanthera clavellata	Transcriptome, 1KP	MTHW	XXXXXXXX
Pistia stratioides	Transcriptome, 1KP	MFIN	XXXXXXXX
Laurelia sempervirens	Transcriptome, 1KP	WAIL	XXXXXXXX
Peperomia fraseri	Transcriptome, 1KP	XSZI	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Amborella trichocarpa	Genome, Amborella Genome Project[3]	-	-
Gymnosperm PHYN			
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ	XXXXXXXX
Podocarpus rubens	Transcriptome, 1KP	XLGK	KJ195161
Thuja plicata	Transcriptome, 1KP	VFYZ	XXXXXXXX
Taiwania cryptomerioides	Transcriptome, 1KP	QSNJ	KJ195162
Pinus parviflora	Transcriptome, 1KP	IIOL	KJ195163
Welwitschia mirabilis	Transcriptome, 1KP	TOXE	KJ195164
Gnetum montanum	Transcriptome, 1KP	GTHK	KJ195165
Ginkgo biloba	Transcriptome, 1KP	SGTW	XXXXXXXX
Encephalartos barteri	Transcriptome, 1KP	GNQG	KJ195166
Angiosperm PHYC			
Arabidopsis thaliana	Genome, Phytozome	-	-
Fragaria vesca	Genome, Phytozome	-	-
Cucumis sativus	Genome, Phytozome	-	-
Vitis vinifera	Genome, Phytozome	-	-
Solanum lycopersicum	Genome, Phytozome	-	-
Buxus sempervirens	Transcriptome, 1KP	IWMW	XXXXXXXX
Aquilegia coerulea	Genome, Phytozome	-	-
Oryza sativa	Genome, Phytozome	-	-
Platanthera clavellata	Transcriptome, 1KP	MTHW	XXXXXXXX
Pistia stratioides	Transcriptome, 1KP	MFIN	XXXXXXXX
Laurelia sempervirens	Transcriptome, 1KP	WAIL	XXXXXXXX
Peperomia fraseri	Transcriptome, 1KP	XSZI	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Amborella trichocarpa	Genome, Amborella Genome Project	-	-
Gymnosperm PHYO			
Podocarpus rubens	Transcriptome, 1KP	XLGK	KJ195167
Thuja plicata	Transcriptome, 1KP	VFYZ	XXXXXXXX
Taiwania cryptomerioides	Transcriptome, 1KP	QSNJ	KJ195168
Pinus parviflora	Transcriptome, 1KP	IIOL	KJ195169
Ginkgo biloba	Transcriptome, 1KP	SGTW	XXXXXXXX
Encephalartos barteri	Transcriptome, 1KP	GNQG	KJ195170
Angiosperm PHYB			
Arabidopsis thaliana D	Genome, Phytozome	-	-
Arabidopsis thaliana B	Genome, Phytozome	-	-
Fragaria vesca	Genome, Phytozome	-	-
Phaseolus vulgaris	Genome, Phytozome	-	-
Cucumis sativus	Genome, Phytozome	-	-
Vitis vinifera	Genome, Phytozome	-	-
Solanum lycopersicum 1	Genome, Phytozome	-	-
Solanum lycopersicum 2	Genome, Phytozome	-	-
Buxus sempervirens	Transcriptome, 1KP	IWMW	XXXXXXXX
Aquilegia coerulea	Genome, Phytozome	-	-
Oryza sativa	Genome, Phytozome	-	-
Platanthera clavellata	Transcriptome, 1KP	MTHW	XXXXXXXX
Pistia stratioides	Transcriptome, 1KP	MFIN	XXXXXXXX
Laurelia sempervirens	Transcriptome, 1KP	WAIL	XXXXXXXX
Peperomia fraseri	Transcriptome, 1KP	XSZI	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Amborella trichocarpa	Genome, Amborella Genome Project	-	-
Angiosperm PHYE			
Arabidopsis thaliana	Genome, Phytozome	-	-
Fragaria vesca 1	Genome, Phytozome	-	-
Fragaria vesca 2	Genome, Phytozome	-	-
Phaseolus vulgaris	Genome, Phytozome	-	-
Cucumis sativus	Genome, Phytozome	-	-
Solanum lycopersicum	Genome, Phytozome	-	-
Capnoides sempervirens 1	Transcriptome, 1KP	AUGV	XXXXXXXX
Aquilegia coerulea 1	Genome, Phytozome	-	-
Capnoides sempervirens 2	Transcriptome, 1KP	AUGV	XXXXXXXX
Aquilegia coerulea 2	Genome, Phytozome	-	-
Laurelia sempervirens	Transcriptome, 1KP	WAIL	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Gymnosperm PHYP			
Thuja plicata	Transcriptome, 1KP	VFYZ	XXXXXXXX
Taiwania cryptomerioides	Transcriptome, 1KP	QSNJ	KJ195155
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ	XXXXXXXX
Podocarpus rubens	Transcriptome, 1KP	XLGK	KJ195156
Pinus parviflora	Transcriptome, 1KP	IIOL	KJ195157
Gnetum montanum	Transcriptome, 1KP	GTHK	KJ195158
Welwitschia mirabilis	Transcriptome, 1KP	TOXE	KJ195159
Ginkgo biloba	Transcriptome, 1KP	SGTW	XXXXXXXX
Encephalartos barteri	Transcriptome, 1KP	GNQG	KJ195160
Fern PHY1			
Equisetum diffusum	Transcriptome, 1KP	CAPN	XXXXXXXX
Marattia howeana	Cloning, this study	-	XXXXXXXX
Osmunda sp	Transcriptome, 1KP	UOMY	KJ195150
Dipteris conjugata	Transcriptome, 1KP	MEKP	XXXXXXXX
Pilularia globulifera	Transcriptome, 1KP	KIIX	KJ195151
Plagiogyria japonica	Transcriptome, 1KP	UWOD	KJ195152
Adiantum capillus-veneris	GenBank	-	AB016151

Phytochrome	Source	1KP identififer	GenBank accession
Moss PHY2_4			
Atrichum angustatum	Transcriptome, 1KP	ZTHV	XXXXXXXX
Physcomitrella patens 4	Genome, Phytozome	-	XM_001773498
Physcomitrella patens 2	Genome, Phytozome	-	XM_001782287
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH	XXXXXXXX
Hedwigia ciliata	Transcriptome, 1KP	YWNF	KJ195190
Philonotis fontana	Transcriptome, 1KP	ORKS	XXXXXXXX
Ceratodon purpureus	Transcriptome, 1KP	FFPD	XXXXXXXX
Leucobryum albidum	Transcriptome, 1KP	VMXJ	XXXXXXXX
Racomitrium varium	Transcriptome, 1KP	RDOO	XXXXXXXX
Scouleria aquatica	Transcriptome, 1KP	BPSG	XXXXXXXX
Bryum argenteum	Transcriptome, 1KP	JMXW	KJ195191
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM	XXXXXXXX
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX
Cladopodium rostratum	Transcriptome, 1KP	VBMM	KJ195189
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX
Leucodon julaceus	Transcriptome, 1KP	IGUH	XXXXXXXX
Pseudotaxiphyllum elegans	Transcriptome, 1KP	QKQO	XXXXXXXX
Anomodon attenuatus	Transcriptome, 1KP	QMWB	XXXXXXXX
Stereodon subimponens	Transcriptome, 1KP	LNSF	XXXXXXXX
Moss PHY5			
Atrichum angustatum	Transcriptome, 1KP	ZTHV	XXXXXXXX
Physcomitrella patens A	Genome, Phytozome	-	XM_001761093
Physcomitrella patens B	Genome, Phytozome	-	XM_001767172
Physcomitrella patens C	Genome, Phytozome	-	XM_001754314
Rhynchostegium serrulatum D	Transcriptome, 1KP	JADL	XXXXXXXX
Aulacomnium heterostichum D	Transcriptome, 1KP	WNGH	XXXXXXXX
Hedwigia ciliata D	Transcriptome, 1KP	YWNF	XXXXXXXX
Philonotis fontana D	Transcriptome, 1KP	ORKS	XXXXXXXX
Leucobryum albidum D	Transcriptome, 1KP	VMXJ	XXXXXXXX
Ceratodon purpureus D	GenBank	-	AY123149
Scouleria aquatica D	Transcriptome, 1KP	BPSG	XXXXXXXX
Racomitrium varium D	Transcriptome, 1KP	RDOO	XXXXXXXX
Syntrichia princeps D	Transcriptome, 1KP	GRKU	XXXXXXXX
Bryum argenteum D	Transcriptome, 1KP	JMXW	KJ195187
Leucodon brachypus D	Transcriptome, 1KP	ZACW	XXXXXXXX
Anomodon attenuatus D	Transcriptome, 1KP	QMWB	XXXXXXXX
Cladopodium rostratum D	Transcriptome, 1KP	VBMM	KJ195185
Loeskeobryum brevirostre D	Transcriptome, 1KP	WSPM	XXXXXXXX
Pseudotaxiphyllum elegans D	Transcriptome, 1KP	QKQO	XXXXXXXX
Racomitrium varium E	Transcriptome, 1KP	RDOO	XXXXXXXX
Scouleria aquatica E	Transcriptome, 1KP	BPSG	XXXXXXXX
Syntrichia princeps E	Transcriptome, 1KP	GRKU	XXXXXXXX
Hedwigia ciliata E	Transcriptome, 1KP	YWNF	KJ195186
Philonotis fontana E	Transcriptome, 1KP	ORKS	XXXXXXXX
Leucobryum albidum E	Transcriptome, 1KP	VMXJ	XXXXXXXX
Ceratodon purpureus E	GenBank	-	EU122393
Rhynchostegium serrulatum E	Transcriptome, 1KP	JADL	XXXXXXXX
Aulacomnium heterostichum E 1	Transcriptome, 1KP	WNGH	XXXXXXXX
Aulacomnium heterostichum E 2	Transcriptome, 1KP	WNGH	XXXXXXXX
Anomodon attenuatus E	Transcriptome, 1KP	QMWB	XXXXXXXX
Neckera douglasii E	Transcriptome, 1KP	TMAJ	XXXXXXXX
Pseudotaxiphyllum elegans E	Transcriptome, 1KP	QKQO	XXXXXXXX
Cladopodium rostratum E	Transcriptome, 1KP	VBMM	KJ195184
Loeskeobryum brevirostre E	Transcriptome, 1KP	WSPM	XXXXXXXX
Leucodon julaceus E	Transcriptome, 1KP	IGUH	XXXXXXXX
Leucodon brachypus E	Transcriptome, 1KP	ZACW	XXXXXXXX
Hornwort PHY			
Nothoceros aenigmaticus	Cloning, Li et al.[6]	DXOU	KJ128384
Megaceros tosanus	Transcriptome, 1KP	UCRN	KJ195197
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB	KJ195198
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB	KJ195199
Anthoceros punctatus	Genome, Li et al.[6]	-	KJ195038
Phaeomegaceros coriaceus	Transcriptome, 1KP	AKXB	XXXXXXXX
Leiosporoceros dussii	Transcriptome, 1KP	ANON	XXXXXXXX
Hornwort NEO			
Nothoceros aenigmaticus	Transcriptome, 1KP	DXOU	KJ128382
Megaceros flagellaris	Transcriptome, 1KP	UCRN	KJ195132
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB	KJ194999
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB	KJ194998
Anthoceros punctatus	Cloning, Li et al.	-	KJ194997
Fern NEO			
Adiantum capillus-veneris	GenBank	-	AB012082
Dipteris conjugata	Cloning, Li et al.	-	KJ195011
Blechnum spicant	Cloning, Li et al.	-	KJ195007
Hemidictyum marginatum	Cloning, Li et al.	-	KJ195037
Zygnematales NEO			
Mougeotia scalaris 1	GenBank	-	AB206961
Mougeotia scalaris 2	GenBank	-	AB206962
Cylindrocystis brebissonii 1	Transcriptome, 1KP	YOXI	KJ195042
Cylindrocystis sp 1	Transcriptome, 1KP	VAZE	KJ195043
Cylindrocystis brebissonii 2	Transcriptome, 1KP	YOXI	KJ195044
Cylindrocystis sp 2	Transcriptome, 1KP	VAZE	KJ195041
Zygnemopsis sp	Transcriptome, 1KP	MFZO	KJ195039
Desmidiales PHY2C			
Desmidium aptogonum	Transcriptome, 1KP	DFDS	XXXXXXXX
Onychonema laeve	Transcriptome, 1KP	GGWH	XXXXXXXX
Cosmarium granatum 1	Transcriptome, 1KP	MNNM	XXXXXXXX
Staurodesmus convergens 1	Transcriptome, 1KP	WCQU	XXXXXXXX
Cosmarium subtumidum	Transcriptome, 1KP	WDGV	XXXXXXXX
Staurodesmus convergens 2	Transcriptome, 1KP	WCQU	XXXXXXXX
Cosmarium granatum 2	Transcriptome, 1KP	MNNM	XXXXXXXX
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV	XXXXXXXX
Penium exiguum	Transcriptome, 1KP	YSQT	XXXXXXXX
Gonatozygon kinahanii 1	Transcriptome, 1KP	KEYW	XXXXXXXX
Gonatozygon kinahanii 2	Transcriptome, 1KP	KEYW	XXXXXXXX
Desmidiales PHY2B			
Desmidium aptogonum	Transcriptome, 1KP	DFDS	XXXXXXXX

Species	Source	Code	Accession
Pteridium aquilinum	Transcriptome, Der et al.[4]	-	SRX423244
Cystopteris protrusa	Transcriptome, 1KP	YOWV	XXXXXXXX
Blechnum spicant	Transcriptome, 1KP	VITX	KJ195153
Polypodium glycyrrhiza	Transcriptome, 1KP	CJNT	KJ195154
Asplenium platyneuron	Transcriptome, 1KP	KJZG	XXXXXXXX
Alsophila spinulosa	Transcriptome, 1KP	GANB	XXXXXXXX
Woodsia scopulina	Transcriptome, 1KP	YJJY	XXXXXXXX
Deparia lobato-crenata	Transcriptome, 1KP	FCHS	XXXXXXXX
Vittaria appalachiana	Transcriptome, 1KP	NDUV	XXXXXXXX
Danaea nodosa	Transcriptome, 1KP	DFHO	XXXXXXXX
Azolla filiculoides	Transcriptome, Brouwer et al.[5]	-	-
Thyrsopteris elegans	Transcriptome, 1KP	EWXK	XXXXXXXX
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW	XXXXXXXX
Lonchitis hirsuta	Transcriptome, 1KP	VVRN	XXXXXXXX
Fern PHY2/4			
Equisetum diffusum A	Transcriptome, 1KP	CAPN	KJ195136
Equisetum diffusum B	Transcriptome, 1KP	CAPN	XXXXXXXX
Botrypus virginianus A	Transcriptome, 1KP	BEGM	KJ195137
Sceptridium dissectum A	Transcriptome, 1KP	EEAQ	XXXXXXXX
Botrypus virginianus B	Transcriptome, 1KP	BEGM	XXXXXXXX
Sceptridium dissectum B	Transcriptome, 1KP	EEAQ	XXXXXXXX
Psilotum nudum	GenBank	QVMR	X74930
Marattia attenuata	Transcriptome, 1KP	UGNK	XXXXXXXX
Osmunda sp A	Transcriptome, 1KP	UOMY	KJ195138
Dipteris conjugata	Transcriptome, 1KP	MEKP	KJ195139
Danaea nodosa	Transcriptome, 1KP	DFHO	XXXXXXXX
Osmundastrum cinnamomeum A	Transcriptome, 1KP	RFMZ	XXXXXXXX
Osmunda javanica A	Transcriptome, 1KP	VIBO	XXXXXXXX
Osmundastrum cinnamomeum B	Transcriptome, 1KP	RFMZ	XXXXXXXX
Osmunda javanica B	Transcriptome, 1KP	VIBO	XXXXXXXX
Fern PHY2			
Anemia tomentosa	Transcriptome, 1KP	CQPW	KJ195140
Pilularia globulifera	Transcriptome, 1KP	KIIX	KJ195142
Plagiogyria japonica	Transcriptome, 1KP	UWOD	KJ195141
Adiantum capillus-veneris	GenBank	-	AB016232
Pteridium aquilinum	Transcriptome, Der et al.	-	SRX423244
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ	XXXXXXXX
Cystopteris fragilis	Transcriptome, 1KP	LHLE	XXXXXXXX
Blechnum spicant	Transcriptome, 1KP	VITX	KJ195143
Leucostegia immersa	Transcriptome, 1KP	WGTU	XXXXXXXX
Polypodium hesperium	Transcriptome, 1KP	GYFU	KJ195144
Gymnocarpium dryopteris	Transcriptome, 1KP	HEGQ	XXXXXXXX
Homalosorus pycnocarpos	Transcriptome, 1KP	OCZL	XXXXXXXX
Athyrium filix-femina	Transcriptome, 1KP	AFPO	XXXXXXXX
Diplazium wichurae	Transcriptome, 1KP	UFJN	XXXXXXXX
Deparia lobato-crenata	Transcriptome, 1KP	FCHS	XXXXXXXX
Asplenium platyneuron	Transcriptome, 1KP	KJZG	XXXXXXXX
Woodsia scopulina	Transcriptome, 1KP	YJJY	XXXXXXXX
Woodsia ilvensis	Transcriptome, 1KP	YQEC	XXXXXXXX
Vittaria appalachiana	Transcriptome, 1KP	NDUV	XXXXXXXX
Pityrogramma trifoliata	Transcriptome, 1KP	UJTT	XXXXXXXX
Pteris ensiformis	Transcriptome, 1KP	FLTD	XXXXXXXX
Myriopteris rufa	Transcriptome, 1KP	GSXD	XXXXXXXX
Gaga arizonica	Transcriptome, 1KP	DCDT	XXXXXXXX
Argyrochosma nivea	Transcriptome, 1KP	XDDT	XXXXXXXX
Notholaena montieliae	Transcriptome, 1KP	YCKE	XXXXXXXX
Cryptogramma acrostichoides	Transcriptome, 1KP	WQML	XXXXXXXX
Thyrsopteris elegans	Transcriptome, 1KP	EWXK	XXXXXXXX
Culcita macrocarpa	Transcriptome, 1KP	PNZO	XXXXXXXX
Lindsaea microphylla	Transcriptome, 1KP	YIXP	XXXXXXXX
Lonchitis hirsuta	Transcriptome, 1KP	VVRN	XXXXXXXX
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW	XXXXXXXX
Fern PHY4A			
Plagiogyria japonica	Transcriptome, 1KP	UWOD	KJ195145
Pilularia globulifera	Transcriptome, 1KP	KIIX	KJ195146
Adiantum tenerum	Transcriptome, 1KP	BMJR	KJ195147
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ	XXXXXXXX
Cystopteris fragilis	Transcriptome, 1KP	LHLE	XXXXXXXX
Blechnum spicant	Transcriptome, 1KP	VITX	KJ195148
Leucostegia immersa	Transcriptome, 1KP	WGTU	XXXXXXXX
Polypodium hesperium	Transcriptome, 1KP	GYFU	KJ195149
Adiantum capillus-veneris	GenBank	-	AB003364
Asplenium platyneuron	Transcriptome, 1KP	KJZG	XXXXXXXX
Gaga arizonica	Transcriptome, 1KP	DCDT	XXXXXXXX
Myriopteris rufa	Transcriptome, 1KP	GSXD	XXXXXXXX
Argyrochosma nivea	Transcriptome, 1KP	XDDT	XXXXXXXX
Vittaria appalachiana	Transcriptome, 1KP	NDUV	XXXXXXXX
Pteris ensiformis	Transcriptome, 1KP	FLTD	XXXXXXXX
Cryptogramma acrostichoides	Transcriptome, 1KP	WQML	XXXXXXXX
Alsophila spinulosa	Transcriptome, 1KP	GANB	XXXXXXXX
Azolla caroliniana	Transcriptome, 1KP	CVEG	XXXXXXXX
Lonchitis hirsuta	Transcriptome, 1KP	VVRN	XXXXXXXX
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW	XXXXXXXX
Anemia tomentosa A	Transcriptome, 1KP	CQPW	XXXXXXXX
Anemia tomentosa B	Transcriptome, 1KP	CQPW	XXXXXXXX
Lindsaea microphylla	Transcriptome, 1KP	YIXP	XXXXXXXX
Lygodium japonicum	Transcriptome, 1KP	PBUU	XXXXXXXX
Thyrsopteris elegans	Transcriptome, 1KP	EWXK	XXXXXXXX
Fern PHY4B			
Pteridium aquilinum	Transcriptome, Der et al.	-	SRX423244
Cystopteris fragilis	Transcriptome, 1KP	LHLE	XXXXXXXX
Blechnum spicant	Transcriptome, 1KP	VITX	XXXXXXXX
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ	XXXXXXXX
Gymnocarpium dryopteris	Transcriptome, 1KP	HEGQ	XXXXXXXX
Woodsia ilvensis	Transcriptome, 1KP	YQEC	XXXXXXXX
Athyrium filix-femina	Transcriptome, 1KP	AFPO	XXXXXXXX
Diplazium wichurae	Transcriptome, 1KP	UFJN	XXXXXXXX
Isoetopsida PHY			
Selaginella cf pallescens	Transcriptome, 1KP	ABIJ	XXXXXXXX
Selaginella moellendorffii	Genome, Phytozome	-	XM_002991595

Species	Source	Code	Accession
Onychonema laeve	Transcriptome, 1KP	GGWH	XXXXXXXX
Cosmarium subtumidum	Transcriptome, 1KP	WDGV	XXXXXXXX
Staurodesmus convergens	Transcriptome, 1KP	WCQU	XXXXXXXX
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV	XXXXXXXX
Penium exiguum	Transcriptome, 1KP	YSQT	XXXXXXXX
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW	XXXXXXXX
Desmidiales PHY2B			
Desmidium aptogonum	Transcriptome, 1KP	DFDS	XXXXXXXX
Cosmarium subtumidum	Transcriptome, 1KP	WDGV	XXXXXXXX
Cosmarium granatum	Transcriptome, 1KP	MNNM	XXXXXXXX
Onychonema laeve	Transcriptome, 1KP	GGWH	XXXXXXXX
Staurodesmus convergens	Transcriptome, 1KP	WCQU	XXXXXXXX
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV	XXXXXXXX
Penium exiguum	Transcriptome, 1KP	YSQT	XXXXXXXX
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW	XXXXXXXX
Roya obtusa	Transcriptome, 1KP	XRTZ	KJ195200
Planotaenium ohtanii	Transcriptome, 1KP	SNOX	KJ195201
Zygnematales PHY2			
Mougeotia scalaris	Genbank	-	AB206965
Mesotaenium kramstei	Transcriptome, 1KP	NBYP	XXXXXXXX
Mesotaenium caldariorum	Genbank	-	U31283
Zygnemopsis sp	Transcriptome, 1KP	MFZO	KJ195202
Cylindrocystis sp	Transcriptome, 1KP	VAZE	KJ195203
Cylindrocystis brebissonii	Transcriptome, 1KP	YOXI	KJ195204
Mesotaenium braunii	Transcriptome, 1KP	WSJO	KJ195205
Netrium digitus	Transcriptome, 1KP	FFGR	XXXXXXXX
Coleochaetales PHY2			
Coleochaete irregularis	Transcriptome, 1KP	QPDY	KJ195206
Chaetosphaeridium globosum	Transcriptome, 1KP	DRGY	KJ195207
Klebsormidiales PHY2			
Interfilum paradoxum	Transcriptome, 1KP	FPCO	KJ195208
Klebsormidium subtile	Transcriptome, 1KP	FQLP	KJ195209
Entransia fimbriata	Transcriptome, 1KP	BFIK	KJ195210
Desmidiales PHY1			
Cosmarium granatum	Transcriptome, 1KP	MNNM	XXXXXXXX
Cosmarium subtumidum	Transcriptome, 1KP	WDGV	XXXXXXXX
Staurodesmus convergens	Transcriptome, 1KP	WCQU	XXXXXXXX
Onychonema laeve	Transcriptome, 1KP	GGWH	XXXXXXXX
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV	XXXXXXXX
Penium exiguum	Transcriptome, 1KP	YSQT	XXXXXXXX
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW	XXXXXXXX
Roya obtusa	Transcriptome, 1KP	XRTZ	XXXXXXXX
Planotaenium ohtanii	Transcriptome, 1KP	SNOX	XXXXXXXX
Klebsormidiales PHY1			
Entransia fimbriat	Transcriptome, 1KP	BFIK	XXXXXXXX
Mesostigmatales PHY1/2			
Mesostigma viride	Transcriptome, 1KP	KYIO	XXXXXXXX
Chlorokybus atmophyticus 1	Transcriptome, 1KP	AZZW	XXXXXXXX
Chlorokybus atmophyticus 2	Transcriptome, 1KP	AZZW	XXXXXXXX
Spirotaenia minuta	Transcriptome, 1KP	NNHQ	XXXXXXXX
Zygnematales PHYX2			
Zygnemopsis sp	Transcriptome, 1KP	MFZO	XXXXXXXX
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ	XXXXXXXX
Mesotaenium kramstei	Transcriptome, 1KP	NBYP	XXXXXXXX
Netrium digitus	Transcriptome, 1KP	FFGR	XXXXXXXX
Coleochaetales PHYX2			
Coleochaete irregularis	Transcriptome, 1KP	QPDY	XXXXXXXX
Chaetosphaeridium globosum	Transcriptome, 1KP	DRGY	XXXXXXXX
Zygnematales PHYX1			
Zygnemopsis sp	Transcriptome, 1KP	MFZO	XXXXXXXX
Mesotaenium kramstei	Transcriptome, 1KP	NBYP	XXXXXXXX
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ	XXXXXXXX
Cylindrocystis brebissonii	Transcriptome, 1KP	YOXI	XXXXXXXX
Cylindrocystis sp	Transcriptome, 1KP	VAZE	XXXXXXXX
Netrium digitus	Transcriptome, 1KP	FFGR	XXXXXXXX
Coleochaetales PHYX1			
Coleochaete irregularis	Transcriptome, 1KP	QPDY	XXXXXXXX
Prasinophyte PHY			
Monomastix opisthostigma	Transcriptome, 1KP	BTFM	-
Pyramimonas parkeae	Transcriptome, 1KP	TNAW	-
Tetraselmis cordiformis	Transcriptome, 1KP	DUMA	-
Dolichomastix tenuilepsis	GenBank	-	KF876183
Tetraselmis astigmatica	GenBank	-	KF876182
Prasinoderma coloniale	GenBank	-	KF876180
Nephroselmis pyriformis	GenBank	-	KF876181
Micromonas pusilla	GenBank	-	KF754357
Cryptophyte PHY			
Hemiselmis virescens	Transcriptome, 1KP	MJMQ	XXXXXXXX
Cryptomonas curvata	Transcriptome, 1KP	BAKF	XXXXXXXX
Rhodomonas sp	Transcriptome, 1KP	IAYV	XXXXXXXX
Guillardia theta 1	Genome, Curtis et al.[7]	-	-
Guillardia theta 2	Genome, Curtis et al.	-	-
Chroomonas sp	Transcriptome, 1KP	ROZZ	XXXXXXXX
Guillardia theta PEK1	Genome, Curtis et al.	-	-
Guillardia theta PEK2	Genome, Curtis et al.	-	-
Guillardia theta PEK3	Genome, Curtis et al.	-	-
Glaucophyte PHY			
Cyanophora paradoxa 1	Genome, Price et al.[8]	-	KF597305
Cyanophora paradoxa 2	Genome, Price et al.	-	KF615870
Cyanophora paradoxa 4	Genome, Price et al.	-	
Gloeochaete wittrockiana GPS2	Transcriptome, Duanmu et al.[9]	-	-
Gloeochaete wittrockiana GPS4	Transcriptome, Duanmu et al.	-	-
Gloeochaete wittrockiana GPS3	Transcriptome, Duanmu et al.	-	-
Gloeochaete wittrockiana GPS1	Transcriptome, Rockwell et al.[10]	-	KF894953
Cyanoptyche gloeocystis 1	Transcriptome, 1KP	JKHA	XXXXXXXX
Cyanoptyche gloeocystis 2	Transcriptome, 1KP	JKHA	XXXXXXXX
Cyanoptyche gloeocystis 3	Transcriptome, 1KP	JKHA	XXXXXXXX
Glaucocystis nostochinearum 1	Transcriptome, 1KP	POOW	XXXXXXXX
Glaucocystis nostochinearum 2	Transcriptome, 1KP	POOW	XXXXXXXX
Glaucocystis nostochinearum 3	Transcriptome, 1KP	POOW	XXXXXXXX

Selaginella willdenowii	Transcriptome, 1KP	KJYC	XXXXXXXX	Glaucocystis nostochinearum 4	Transcriptome, 1KP	POOW	XXXXXXXX
Selaginella kraussiana	Transcriptome, 1KP	ZFGK	XXXXXXXX	**Diatom PHY**			
Selaginella wallacei	Transcriptome, 1KP	JKAA	KJ195172	Thalassiosira pseudonana	GenBank	-	XM_002290775
Selaginella apoda	Transcriptome, 1KP	LGDQ	XXXXXXXX	Phaeodactylum tricornutum	GenBank	-	XM_002179026
Isoetes tegetiformans	Transcriptome, 1KP	PKOX	KJ195171	**Brown algae PHY**			
Lycopodiales PHY1				Petalonia fascia	Transcriptome, 1KP	VRGZ	XXXXXXXX
Dendrolycopodium obscurum	Transcriptome, 1KP	XNXF	KJ195173	Scytosiphon lomentaria	Transcriptome, 1KP	JCXF	XXXXXXXX
Huperzia selago	Transcriptome, 1KP	GTUO	KJ195174	Ishige okamurai	Transcriptome, 1KP	APTP	XXXXXXXX
Pseudolycopodiella caroliniana	Transcriptome, 1KP	UPMJ	XXXXXXXX	Ectocarpus siliculosus	GenBank	-	FN649030
Lycopodiales PHY2				Feldmannia species virus isolate FsV	GenBank	-	EU916176
Dendrolycopodium obscurum	Transcriptome, 1KP	XNXF	KJ195175	Ectocarpus siliculosus virus isolate EsV	GenBank	-	AF204951
Huperzia lucidula	Transcriptome, 1KP	GKAG	KJ195176	**Fungi PHY**			
Pseudolycopodiella caroliniana	Transcriptome, 1KP	UPMJ	XXXXXXXX	Aspergillus fumigatus	GenBank	-	DS499603
Liverwort PHY				Ustilago maydis	GenBank	-	AACP01000210
Scapania nemorosa	Transcriptome, 1KP	IRBN	KJ195177	Cochliobolus heterostrophus	GenBank	-	AY456024
Porella pinnata	Transcriptome, 1KP	UUHD	KJ195178	Neurospora crassa 1	GenBank	-	BK004087
Schistochila sp	Transcriptome, 1KP	LGOW	KJ195179	Neurospora crassa 2	GenBank	-	DQ128077
Metzgeria crassipilis	Transcriptome, 1KP	NRWZ	KJ195180	Penicillium marneffei	GenBank	-	XM_002144684
Pellia neesiana	Transcriptome, 1KP	JHFI	KJ195181	**Cyanobacteria PHY**			
Odontoschisma prostratum	Transcriptome, 1KP	YBQN	XXXXXXXX	Microcystis aeruginosa	GenBank	-	AP009552
Marchantia paleacea	GenBank	-	AB022917	Nodularia spumigena	GenBank	-	AAVW01000039
Conocephalum conicum	Transcriptome, 1KP	ILBQ	KJ195182	Synechocystis sp	GenBank	-	BA000022
Lunularia cruciata	Transcriptome, 1KP	TXVB	XXXXXXXX	Anabaena variabilis	GenBank	-	NC_007413
Sphaerocarpos texanus	Transcriptome, 1KP	HERT	KJ195183	Cyanothece sp	GenBank	-	ABVE01000001
Ricciocarpos natans	Transcriptome, 1KP	WJLO	XXXXXXXX	Nostoc sp	GenBank	-	BA000019
Bazzania trilobata	Transcriptome, 1KP	WZYK	XXXXXXXX	Tolypothrix sp 1	GenBank	-	AF309559
Moss PHY1_3				Tolypothrix sp 2	GenBank	-	AF309560
Andreaea rupestris	Transcriptome, 1KP	WOGB	XXXXXXXX	**Cyanobacteria PAS-less PHY**			
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX	Microcoleus chthonoplastes	GenBank	-	DS989848
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH	XXXXXXXX	Lyngbya sp	GenBank	-	AAVU01000002
Hedwigia ciliata	Transcriptome, 1KP	YWNF	KJ195194	Nostoc punctiforme	GenBank	-	CP001037
Philonotis fontana	Transcriptome, 1KP	ORKS	XXXXXXXX	Anabaena variabilis	GenBank	-	CP000117
Ceratodon purpureus A	GenBank	-	U72993	Synechococcus sp	GenBank	-	ABD00881
Ceratodon purpureus B	GenBank	-	U87632	**Bacteria PHY**			
Leucobryum albidum	Transcriptome, 1KP	VMXJ	XXXXXXXX	Agrobacterium tumefaciens	GenBank	-	C58AE007869
Physcomitrella patens 1	Genome, Phytozome	-	XM_001765983	Pseudomonas aeruginosa	GenBank	-	CH482383
Physcomitrella patens 3	Genome, Phytozome	-	XM_001778103	Pseudomonas syringae	GenBank	-	AE016853
Racomitrium varium	Transcriptome, 1KP	RDOO	XXXXXXXX	Ralstonia pickettii	GenBank	-	CP001068
Leucodon julaceus	Transcriptome, 1KP	IGUH	XXXXXXXX	Stenotrophomonas maltophilia	GenBank	-	AM743169
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX	Rhodopseudomonas palustris	GenBank	-	BX572602
Pseudotaxiphyllum elegans	Transcriptome, 1KP	QKQO	XXXXXXXX	Oceanicola granulosus	GenBank	-	AAOT01000028
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX	Methylocella silvestris	GenBank	-	CP001280
Stereodon subimponens	Transcriptome, 1KP	LNSF	XXXXXXXX	Sclerotinia sclerotiorum	GenBank	-	CH476645
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM	XXXXXXXX	Magnetospirillum magnetotacticum	GenBank	-	AAAP01003476
Cladopodium rostratum	Transcriptome, 1KP	VBMM	KJ195193				
Bryum argenteum	Transcriptome, 1KP	JMXW	KJ195195				
Moss PHY2_4/5							
Andreaea rupestris	Transcriptome, 1KP	WOGB	XXXXXXXX				
Sphagnum palustre	Transcriptome, 1KP	RCBT	XXXXXXXX				
Takakia lepidozioides	Transcriptome, 1KP	SKQD	XXXXXXXX				

[4]Brouwer, P. *et al. Azolla* domestication towards a biobased economy? *New Phytol* **202,** 1069–1082 (2014).

[5]Der, J. P., Barker, M. S., Wickett, N. J., Depamphilis, C. W. & Wolf P. G. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12,** 99 (2011).

[6]Li, F.-W. *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. USA* **111,** 6672–6677 (2014).

[7]Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492,** 59–65 (2012).

[8]Price, D. C. *et al. Cyanophora paradoxa g*enome elucidates origin of photosynthesis in algae and plants. *Science* **335,** 843–847 (2012).

[9]Duanmu, D. *et al.* Marine algae and land plants share conserved phytochrome signaling systems. *Proc. Natl. Acad. Sci. USA* **111,** 15827–15832 (2014).

[10]Rockwell, N. C. *et al.* Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl. Acad. Sci. USA* **111,** 3871–3876 (2014).

Appendix C: Supplementary Tables for Chapter Two

Table 4: List of transcriptomes and genomes screened for phototropins. "-": not applicable.

Taxa	Source	1KP[3] identifier
Angiosperm		
Amborella trichocarpa	Genome, Amborella Genome Project[1]	-
Aquilegia coerulea	Genome, Phytozome	-
Arabidopsis thaliana	Genome, Phytozome[2]	-
Austrobaileya scandens	Transcriptome, 1KP	FZJL
Citrus clementina	Genome, Phytozome	-
Fragaria vesca	Genome, Phytozome	-
Goodyera pubescens	Transcriptome, 1KP[3]	VTUS
Illicium floridanum	Transcriptome, 1KP	VZCI
Magnolia grandiflora	Transcriptome, 1KP	WBOD
Medicago truncatula	Genome, Phytozome	-
Smilax bona-nox	Transcriptome, 1KP	MWYQ
Solanum lycopersicum	Genome, Phytozome	-
Vitis vinifera	Genome, Phytozome	-
Zea mays	Genome, Phytozome	-
Gymnosperm		
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ
Cunninghamia lanceolata	Transcriptome, 1KP	OUOI
Gnetum montanum	Transcriptome, 1KP	GTHK
Podocarpus rubens	Transcriptome, 1KP	XLGK
Stangeria eriopus	Transcriptome, 1KP	KAWQ
Thuja plicata	Transcriptome, 1KP	VFYZ
Welwitschia mirabilis	Transcriptome, 1KP	TOXE
Ferns		
Adiantum aleuticum	Transcriptome, 1KP	WCLG
Adiantum capillus-veneris	GenBank	-
Anemia tomentosa	Transcriptome, 1KP	CQPW
Asplenium platyneuron	Transcriptome, 1KP	KJZG
Athyrium filix-femina	Transcriptome, 1KP	URCP
Azolla caroliniana	Transcriptome, 1KP	CVEG
Blechnum spicant	Transcriptome, 1KP	VITX
Botrypus virginianus	Transcriptome, 1KP	BEGM
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW
Cryptogramma acrostichoides	Transcriptome, 1KP	WQML
Cystopteris reevesiana	Transcriptome, 1KP	RICC
Danaea nodosa A	Transcriptome, 1KP	DFHO
Davallia fejeensis	Transcriptome, 1KP	OQWW
Dennstaedtia davallioides	Transcriptome, 1KP	MTGC
Dipteris conjugata	Transcriptome, 1KP	MEKP
Equisetum hymale	Transcriptome, 1KP	JVSZ
Gaga arizonica	Transcriptome, 1KP	DCDT
Gymnocarpium dryopteris	Transcriptome, 1KP	HEGQ
Homalosorus pycnocarpos	Transcriptome, 1KP	OCZL
Leucostegia immersa	Transcriptome, 1KP	WGTU
Lindsaea linearis	Transcriptome, 1KP	NOKI
Lonchitis hirsuta	Transcriptome, 1KP	VVRN
Lygodium japonicum	Transcriptome, 1KP	PBUU
Onoclea sensibilis	Transcriptome, 1KP	HTFH
Ophioglossum vulgatum	Transcriptome, 1KP	QHVS
Osmunda sp.	Transcriptome, 1KP	UOMY
Pilularia globulifera	Transcriptome, 1KP	KIIX
Pityrogramma trifoliata A	Transcriptome, 1KP	UJTT
Plagiogyria japonica	Transcriptome, 1KP	UWOD
Polypodium hesperium	Transcriptome, 1KP	GYFU
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ
Psilotum nudum	Transcriptome, 1KP	QVMR
Pteridium aquilinum	Transcriptome, Der et al[4]	-
Pteris vittata	Transcriptome, 1KP	POPJ
Sceptridium dissectum	Transcriptome, 1KP	EEAQ
Thyrsopteris elegans	Transcriptome, 1KP	EWXK
Tmesipteris parva	Transcriptome, 1KP	ALVQ
Vittaria lineata	Transcriptome, 1KP	SKYV
Woodsia scopulina	Transcriptome, 1KP	YJJY

Taxa	Source	1KP[3] identifier
Zygnematales		
Mesotaenium endlicherianum	Transcriptome, 1KP	WDCW
Mesotaenium kramstei	Transcriptome, 1KP	NBYP
Zygnemopsis sp	Transcriptome, 1KP	MFZO
Desmidiales		
Roya obtusa	Transcriptome, 1KP	XRTZ
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW
Planotaenium ohtanii	Transcriptome, 1KP	SNOX
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV
Penium exiguum	Transcriptome, 1KP	YSQT
Desmidium aptogonum	Transcriptome, 1KP	DFDS
Staurodesmus convergens	Transcriptome, 1KP	WCQU
Cosmarium tinctum	Transcriptome, 1KP	BHBK
Coleochaetales		
Coleochaete irregularis	Transcriptome, 1KP	QPDY
Coleochaete scutata	Transcriptome, 1KP	VQBJ
Chaetosphaeridium globosum	Transcriptome, 1KP	DRGY
Klebsormidioales		
Interfilum paradoxum	Transcriptome, 1KP	FPCO
Klebsormidium subtile	Transcriptome, 1KP	FQLP
Entransia fimbriat	Transcriptome, 1KP	BFIK
Mesostigmatales		
Chlorokybus atmophyticus	Transcriptome, 1KP	AZZW
Mesostigma viride	Transcriptome, 1KP	KYIO
Spirotaenia minuta	Transcriptome, 1KP	NNHQ
Prasinophyte		
Ostreococcus tauri	Genome, Phytozome	-
Ostreococcus lucimarinus	Genome, Phytozome	-
Micromonas pusilla	Genome, Phytozome	-
Dolichomastix tenuilepi	Transcriptome, 1KP	XOAL
Pyramimonas parkeae	Transcriptome, 1KP	TNAW
Scourfieldia sp	Transcriptome, 1KP	EGNB
Nephroselmis olivace	Transcriptome, 1KP	MMKU
Tetraselmis cordiformis	Transcriptome, 1KP	DUMA
Scherffelia dubia	Transcriptome, 1KP	FMVB
Pycnococcus provasolii	Transcriptome, 1KP	MXEZ
Ulvophyceae		
Bolbocoleon piliferum	Transcriptome, 1KP	LSHT
Persursaria percursa	Transcriptome, 1KP	OAEZ
Helicodictyon planctonicum	Transcriptome, 1KP	AJAU
Entocladia endozoica	Transcriptome, 1KP	OQON
Trebouxiophyceae		
Coccomyxa pringsheimii	Transcriptome, 1KP	GXBM
Botryococcus terribilis	Transcriptome, 1KP	QYXY
Prasiola crispa	Transcriptome, 1KP	WCLV
Trebouxia arboricola	Transcriptome, 1KP	NKXU
Chlorophyceae		
Oedogonium foveolatum	Transcriptome, 1KP	SDPC
Uronema sp.	Transcriptome, 1KP	ISGT
Aphanochaete repens	Transcriptome, 1KP	IJMT
Fritschiella tuberosa	Transcriptome, 1KP	VFIV
Hafniomonas reticulata	Transcriptome, 1KP	FXHG
Carteria obtusa	Transcriptome, 1KP	RUIF
Scenedesmus dimorphus	Transcriptome, 1KP	PZIF
Cylindrocapsa geminella	Transcriptome, 1KP	DZPJ
Pediastrum duplex	Transcriptome, 1KP	XTON
Brachiomonas submarina	Transcriptome, 1KP	GUBD
Heterochlamydomonas inaequa	Transcriptome, 1KP	IRYH
Volvox carteri	Genome, Phytozome	-
Chlamydomonas reinhardtii	Genome, Phytozome	-
Oogamochlamys gigantea	Transcriptome, 1KP	XDLL
Oedogonium cardiacum	Transcriptome, 1KP	DVYE
Chloromonas tughillensi	Transcriptome, 1KP	UTRE

Selaginella

Species	Source	Code
Selaginella acanthonota	Transcriptome, 1KP	ZYCD
Selaginella kraussiana	Transcriptome, 1KP	ZFGK
Selaginella moellendorffii	Genome, Phytozome	-
Selaginella selaginoides	Transcriptome, 1KP	KUXM
Selaginella willdenowii	Transcriptome, 1KP	KJYC

Isoetales

Species	Source	Code
Isoetes tegetiformans	Transcriptome, 1KP	PKOX

Lycopodiales

Species	Source	Code
Dendrolycopodium obscurum	Transcriptome, 1KP	XNXF
Diphasiastrum digitatum	Transcriptome, 1KP	WAFT
Huperzia lucidula	Transcriptome, 1KP	GKAG
Lycopodium deuterodensum	Transcriptome, 1KP	PQTO
Phylloglossum drummondii	Transcriptome, 1KP	ZZEI
Pseudolycopodiella caroliniana	Transcriptome, 1KP	UPMJ

Liverwort

Species	Source	Code
Bazzania trilobata	Transcriptome, 1KP	WZYK
Conocephalum conicum	Transcriptome, 1KP	ILBQ
Lunularia cruciata	Transcriptome, 1KP	TXVB
Marchantia polymorpha	Transcriptome, 1KP	JPYU
Metzgeria crassipilis	Transcriptome, 1KP	NRWZ
Pellia neesiana	Transcriptome, 1KP	JHFI
Porella pinnata	Transcriptome, 1KP	UUHD
Radula lindenbergia	Transcriptome, 1KP	BNCU
Scapania nemorosa	Transcriptome, 1KP	IRBN
Schistochila sp	Transcriptome, 1KP	LGOW
Sphaerocarpos texanus	Transcriptome, 1KP	HERT

Moss

Species	Source	Code
Andreaea rupestris	Transcriptome, 1KP	WOGB
Atrichum angustatum	Transcriptome, 1KP	ZTHV
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH
Bryum argenteum	Transcriptome, 1KP	JMXW
Buxbaumia aphylla	Transcriptome, 1KP	HRWG
Ceratodon purpureus	Transcriptome, 1KP	FFPD
Fissidens adianthoides	Transcriptome, 1KP	NWQC
Leucodon brachypus	Transcriptome, 1KP	ZACW
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM
Neckera douglasii	Transcriptome, 1KP	TMAJ
Orthotrichum lyellii	Transcriptome, 1KP	CMEQ
Physcomitrella patens	Genome, Phytozome	-
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL
Scouleria aquatica	Transcriptome, 1KP	BPSG
Sphagnum lescurii	Transcriptome, 1KP	GOWD
Takakia lepidozioides	Transcriptome, 1KP	SKQD

Hornwort

Species	Source	Code
Anthoceros puncatatus	Genome, Li et al[5]	-
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB
Phaeomegaceros coriaceus	Transcriptome, 1KP	AKXB
Megaceros flagellaris	Transcriptome, 1KP	UCRN
Nothoceros aenigmaticus	Transcriptome, 1KP	DXOU

Zygnematales

Species	Source	Code
Cylindrocystis brebissonii	Transcriptome, 1KP	YOXI
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ
Cylindrocystis sp	Transcriptome, 1KP	VAZE
Mesotaenium braunii	Transcriptome, 1KP	WSJO
Mesotaenium caldariorum	Transcriptome, 1KP	HKZW

Red Algae

Species	Source	Code
Betaphycus gelatinae	Transcriptome, 1KP	BWVJ
Ceramium kondoi	Transcriptome, 1KP	VZWX
Chondrus crispus	Genome, Collen et al.[6]	-
Chondrus crispus	Transcriptome, 1KP	UGPM
Chroodactylon ornatum	Transcriptome, 1KP	LLXJ
Cyanidioschyzon merolae	Genome, Matsuzaki et al.[7]	-
Dumontia simplex	Transcriptome, 1KP	IEHF
Eucheuma denticulatum	Transcriptome, 1KP	JEBK
Galdieria sulphuraria	Genome, Barbier et al.[8]	-
Glaucosphaera vacuolata	Transcriptome, 1KP	RSOF
Gloeopeltis furcata	Transcriptome, 1KP	SBLT
Gracilaria asiatica	Transcriptome, 1KP	VNAL
Gracilaria blodgettii	Transcriptome, 1KP	LJPN
Gracilaria chouae	Transcriptome, 1KP	FTRP
Gracilaria lemaneiformi	Transcriptome, 1KP	IKWM
Grateloupia filicina	Transcriptome, 1KP	ZJOJ
Grateloupia livida	Transcriptome, 1KP	IKIZ
Grateloupia turuturu	Transcriptome, 1KP	URSB
Grateloupia chiangii	Transcriptome, 1KP	PWKQ
Gymnogongrus ftabelliformis	Transcriptome, 1KP	CKXF
Heterosiphonia pulchra	Transcriptome, 1KP	YSBD
Kappaphycus alvarezii	Transcriptome, 1KP	IHJY
Mazzaella japonica	Transcriptome, 1KP	WEJN
Polysiphonia japonica	Transcriptome, 1KP	XAXW
Pyropia yezoensis	Genome, Nakamura et al.[9]	-
Porphyra yezoensis	Transcriptome, 1KP	ZULJ
Porphyridium cruentum	Transcriptome, 1KP	OBUY
Porphyridium purpureum	Genome, Bhattacharya et al.[10]	-
Porphyridium purpureum	Transcriptome, 1KP	PVGP
Rhodella violacea	Transcriptome, 1KP	RTLC
Rhodochaete parvula	Transcriptome, 1KP	JJZR
Sinotubimorpha guangdongensi	Transcriptome, 1KP	PYDB
Symphyocladia latiuscula	Transcriptome, 1KP	UYFR

Glaucophyte

Species	Source	Code
Cyanophora paradoxa	Genome, Price et al.[11]	-
Cyanoptyche gloeocystis	Transcriptome, 1KP	JKHA
Glaucocystis nostochinearum	Transcriptome, 1KP	POOW

Cryptophyte

Species	Source	Code
Chroomonas sp	Transcriptome, 1KP	ROZZ
Cryptomonas curvata	Transcriptome, 1KP	BAKF
Guillardia theta	Genome, Curtis et al.[12]	-
Hemiselmis virescens	Transcriptome, 1KP	MJMQ
Rhodomonas sp	Transcriptome, 1KP	IAYV

Haptophytes

Species	Source	Code
Isochrysis sp.	Transcriptome, 1KP	BAJW
Pavlova lutheri	Transcriptome, 1KP	NMAK
Prymnesium parvum	Transcriptome, 1KP	LXRN

Stramenopiles - Chrysophytes

Species	Source	Code
Mallomonas sp.	Transcriptome, 1KP	BOGT
Ochromonas sp.	Transcriptome, 1KP	EBWI
Synura petersenii	Transcriptome, 1KP	DBYD
Synura sp.	Transcriptome, 1KP	VKVG

Stramenopiles - Brown algae

Species	Source	Code
Ishige okamurai	Transcriptome, 1KP	APTP
Petalonia fascia	Transcriptome, 1KP	VRGZ
Scytosiphon lomentaria	Transcriptome, 1KP	JCXF

[1] Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **342,** 1241089 (2013).

[2] Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40,** D1178–86 (2012).

[3] Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *GigaScience* **3,** 17 (2014).

[4] Der, J. P., Barker, M. S., Wickett, N. J., Depamphilis, C. W. & Wolf P. G. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12,** 99 (2011).

[5] Li, F.-W. *et al*. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. USA* **111,** 6672–6677 (2014).

[6] Collen, J. et al. Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* **110,** 5247–5252 (2013).

[7] Matsuzaki, M. et al. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* **428,** 653–657 (2004).

[8] Barbier, G. *et al.* Comparative genomics of two closely related unicellular thermo-acidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of Galdieria sulphuraria and significant differences in carbohydrate metabolism of both algae. *Plant Physiology* **137,** 460–474 (2005).

[9] Nakamura, Y. et al. The first symbiont-free genome sequence of marine red alga, Susabi-nori (*Pyropia yezoensis*). *PLoS ONE* **8,** e57122 (2013).

[10] Bhattacharya, D. *et al*. Genome of the red alga *Porphyridium purpureum*. *Nat Comms* **4,** 1941 (2013).

[11] Price, D. C. *et al. Cyanophora paradoxa g*enome elucidates origin of photosynthesis in algae and plants. *Science* **335,** 843–847 (2012).

[12] Curtis, B. A. *et al.* Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492,** 59–65 (2012).

Table 5: Sources and GenBank accession numbers of the phototropins used in phylogenetic analyses. "-": not applicable.

Taxa	Source	1KP[3] identififer	GenBank accession
Angiosperm PHOT1			
Arabidopsis thaliana	Genome, Phytozome[1]	-	-
Medicago truncatula	Genome, Phytozome	-	-
Citrus clementina	Genome, Phytozome	-	-
Fragaria vesca	Genome, Phytozome	-	-
Solanum lycopersicum	Genome, Phytozome	-	-
Vitis vinifera	Genome, Phytozome	-	-
Aquilegia coerulea	Genome, Phytozome	-	-
Zea mays	Genome, Phytozome	-	-
Goodyera pubescens	Transcriptome, 1KP[2]	VTUS	XXXXXXXX
Smilax bona-nox	Transcriptome, 1KP	MWYQ	XXXXXXXX
Magnolia grandiflora	Transcriptome, 1KP	WBOD	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Austrobaileya scandens	Transcriptome, 1KP	FZJL	XXXXXXXX
Amborella trichocarpa	Genome, Amborella Genome Project[3]	-	-
Gymnosperm PHOT1			
Cunninghamia lanceolata	Transcriptome, 1KP	OUOI	KJ195049
Thuja plicata	Transcriptome, 1KP	VFYZ	KJ195050
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ	KJ195047
Podocarpus rubens	Transcriptome, 1KP	XLGK	KJ195048
Gnetum montanum	Transcriptome, 1KP	GTHK	KJ195053
Welwitschia mirabilis	Transcriptome, 1KP	TOXE	KJ195052
Stangeria eriopus	Transcriptome, 1KP	KAWQ	KJ195051
Angiosperm PHOT2			
Arabidopsis thaliana	Genome, Phytozome	-	-
Medicago truncatula	Genome, Phytozome	-	-
Citrus clementina	Genome, Phytozome	-	-
Fragaria vesca	Genome, Phytozome	-	-
Solanum lycopersicum	Genome, Phytozome	-	-
Vitis vinifera	Genome, Phytozome	-	-
Aquilegia coerulea	Genome, Phytozome	-	-
Zea mays	Genome, Phytozome	-	-
Goodyera pubescens	Transcriptome, 1KP	VTUS	XXXXXXXX
Smilax bona-nox	Transcriptome, 1KP	MWYQ	XXXXXXXX
Magnolia grandiflora	Transcriptome, 1KP	WBOD	XXXXXXXX
Illicium floridanum	Transcriptome, 1KP	VZCI	XXXXXXXX
Austrobaileya scandens	Transcriptome, 1KP	FZJL	XXXXXXXX
Amborella trichocarpa	Genome, Amborella Genome Project	-	-
Gymnosperm PHOT2			
Cunninghamia lanceolata	Transcriptome, 1KP	OUOI	KJ195056
Thuja plicata	Transcriptome, 1KP	VFYZ	KJ195057
Cephalotaxus harringtonia	Transcriptome, 1KP	WYAJ	KJ195054
Podocarpus rubens	Transcriptome, 1KP	XLGK	KJ195055
Gnetum montanum	Transcriptome, 1KP	GTHK	KJ195060
Welwitschia mirabilis	Transcriptome, 1KP	TOXE	KJ195059
Stangeria eriopus	Transcriptome, 1KP	KAWQ	KJ195058
Fern PHOT1			
Polypodium hesperium	Transcriptome, 1KP	GYFU	KJ195061
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ	XXXXXXXX
Leucostegia immersa	Transcriptome, 1KP	WGTU	KJ195062
Cystopteris reevesiana	Transcriptome, 1KP	RICC	KJ195063
Adiantum capillus-veneris	GenBank	-	AB037188
Gaga arizonica	Transcriptome, 1KP	DCDT	XXXXXXXX
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW	XXXXXXXX
Pteridium aquilinum	Transcriptome, Der et al[4]	-	KJ195064
Lonchitis hirsuta	Transcriptome, 1KP	VVRN	XXXXXXXX
Pilularia globulifera	Transcriptome, 1KP	KIIX	KJ195065
Dipteris conjugata	Transcriptome, 1KP	MEKP	XXXXXXXX
Osmunda sp.	Transcriptome, 1KP	UOMY	KJ195066
Danaea nodosa A	Transcriptome, 1KP	DFHO	XXXXXXXX
Danaea nodosa B	Transcriptome, 1KP	DFHO	XXXXXXXX
Danaea nodosa C	Transcriptome, 1KP	DFHO	XXXXXXXX
Fern PHOT2			
Polypodium hesperium	Transcriptome, 1KP	GYFU	KJ195067
Davallia fejeensis	Transcriptome, 1KP	OQWW	XXXXXXXX
Leucostegia immersa	Transcriptome, 1KP	WGTU	KJ195068
Polystichum acrostichoides	Transcriptome, 1KP	FQGQ	XXXXXXXX
Athyrium filix-femina	Transcriptome, 1KP	URCP	XXXXXXXX
Blechnum spicant	Transcriptome, 1KP	VITX	XXXXXXXX
Onoclea sensibilis	Transcriptome, 1KP	HTFH	XXXXXXXX
Woodsia scopulina	Transcriptome, 1KP	YJJY	XXXXXXXX
Homalosorus pycnocarpos	Transcriptome, 1KP	OCZL	XXXXXXXX
Cystopteris reevesiana	Transcriptome, 1KP	RICC	KJ195069
Gymnocarpium dryopteris	Transcriptome, 1KP	HEGQ	XXXXXXXX
Asplenium platyneuron	Transcriptome, 1KP	KJZG	XXXXXXXX
Pteridium aquilinum	Transcriptome, Der et al	-	KJ195070
Dennstaedtia davallioides	Transcriptome, 1KP	MTGC	XXXXXXXX
Adiantum capillus-veneris	GenBank	-	AB115545
Adiantum aleuticum	Transcriptome, 1KP	WCLG	XXXXXXXX
Vittaria lineata	Transcriptome, 1KP	SKYV	XXXXXXXX

Taxa	Source	1KP[3] identififer	GenBank accession
Moss PHOT1B			
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH	XXXXXXXX
Ceratodon purpureus	Transcriptome, 1KP	FFPD	XXXXXXXX
Orthotrichum lyellii	Transcriptome, 1KP	CMEQ	XXXXXXXX
Bryum argenteum	Transcriptome, 1KP	JMXW	KJ195092
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM	XXXXXXXX
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX
Atrichum angustatum	Transcriptome, 1KP	ZTHV	XXXXXXXX
Moss PHOT2			
Takakia lepidozioides	Transcriptome, 1KP	SKQD	XXXXXXXX
Sphagnum lescurii	Transcriptome, 1KP	GOWD	KJ195101
Andreaea rupestris	Transcriptome, 1KP	WOGB	XXXXXXXX
Atrichum angustatum	Transcriptome, 1KP	ZTHV	XXXXXXXX
Buxbaumia aphylla	Transcriptome, 1KP	HRWG	XXXXXXXX
Moss PHOT2A			
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH	XXXXXXXX
Scouleria aquatica	Transcriptome, 1KP	BPSG	KJ195096
Ceratodon purpureus	Transcriptome, 1KP	FFPD	XXXXXXXX
Orthotrichum lyellii	Transcriptome, 1KP	CMEQ	XXXXXXXX
Bryum argenteum	Transcriptome, 1KP	JMXW	KJ195095
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM	XXXXXXXX
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX
Moss PHOT2B			
Physcomitrella patens	Genome, Phytozome	-	XM 001785674
Orthotrichum lyellii	Transcriptome, 1KP	CMEQ	XXXXXXXX
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX
Scouleria aquatica	Transcriptome, 1KP	BPSG	KJ195098
Fissidens adianthoides	Transcriptome, 1KP	NWQC	XXXXXXXX
Moss PHOT2C			
Physcomitrella patens 2C-1	Genome, Phytozome	-	XM 001766357
Physcomitrella patens 2C-2	Genome, Phytozome	-	XM 001763052
Hornwort PHOT			
Anthoceros puncatatus	Genome, Li et al[5]	-	KJ195131
Anthoceros bhardwajii	Cloning, This study	-	XXXXXXXX
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB	KJ195133
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB	KJ195134
Phymatoceros phymatodes	Cloning, Li et al[5]	-	KJ195135
Phaeomegaceros coriaceus	Transcriptome, 1KP	AKXB	XXXXXXXX
Megaceros flagellaris	Transcriptome, 1KP	UCRN	KJ195132
Nothoceros aenigmaticus	Cloning, Li et al.	DXOU	KJ128383
Hornwort NEO			
Anthoceros punctatus	Genome, Li et al	-	KJ194997
Phaeoceros carolinianus	Transcriptome, 1KP	WCZB	KJ194998
Paraphymatoceros hallii	Transcriptome, 1KP	FAJB	KJ194999
Phymatoceros phymatodes	Cloning, Li et al.	-	KJ195002
Phaeomegaceros coriaceus	Transcriptome, 1KP	AKXB	XXXXXXXX
Megaceros flagellaris	Transcriptome, 1KP	UCRN	KJ195000
Nothoceros aenigmaticus	Cloning, Li et al.	DXOU	KJ128382
Fern NEO			
Hemidictyum marginatum	GenBank	-	KJ195037
Adiantum capillus-veneris	GenBank	-	AB012082
Adiantum raddianum	Transcriptome, 1KP	BMJR	KJ195003
Allantodia dilatata	GenBank	-	FJ455447
Dipteris conjugata	Transcriptome, 1KP	MEKP	KJ195011
Diplazium wichurae	GenBank	-	KJ195009
Phegopteris hexagonoptera	GenBank	-	KJ195023
Plagiogyria distinctissima	GenBank	-	FJ501965
Dennstaedtia punctilobula	GenBank	-	KJ195016
Zygnematales PHOTA			
Mougeotia scalaris	GenBank	-	AB206963
Mesotaenium caldariorum	Transcriptome, 1KP	HKZW	XXXXXXXX
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ	KJ195120
Zygnemopsis sp	Transcriptome, 1KP	MFZO	KJ195119
Mesotaenium kramstei	Transcriptome, 1KP	NBYP	XXXXXXXX
Zygnematales PHOTB			
Mougeotia scalaris	GenBank	-	AB206964
Mesotaenium caldariorum	Transcriptome, 1KP	HKZW	XXXXXXXX
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ	KJ195118
Zygnemopsis sp	Transcriptome, 1KP	MFZO	KJ195117
Mesotaenium kramstei	Transcriptome, 1KP	NBYP	XXXXXXXX
Cylindrocystis brebissonii 1	Transcriptome, 1KP	YOXI	KJ195114
Cylindrocystis sp 1	Transcriptome, 1KP	VAZE	KJ195113
Cylindrocystis brebissonii 2	Transcriptome, 1KP	YOXI	KJ195116
Cylindrocystis sp 2	Transcriptome, 1KP	VAZE	KJ195115

Species	Source	Code	Accession
Gaga arizonica	Transcriptome, 1KP	DCDT	XXXXXXXX
Pityrogramma trifoliata A	Transcriptome, 1KP	UJTT	XXXXXXXX
Pityrogramma trifoliata B	Transcriptome, 1KP	UJTT	XXXXXXXX
Pteris vittata	Transcriptome, 1KP	POPJ	XXXXXXXX
Cryptogramma acrostichoides	Transcriptome, 1KP	WQML	XXXXXXXX
Ceratopteris thalictroides	Transcriptome, 1KP	PIVW	XXXXXXXX
Lindsaea linearis	Transcriptome, 1KP	NOKI	XXXXXXXX
Lonchitis hirsuta	Transcriptome, 1KP	VVRN	XXXXXXXX
Plagiogyria japonica	Transcriptome, 1KP	UWOD	XXXXXXXX
Thyrsopteris elegans	Transcriptome, 1KP	EWXK	XXXXXXXX
Azolla caroliniana	Transcriptome, 1KP	CVEG	XXXXXXXX
Pilularia globulifera	Transcriptome, 1KP	KIIX	KJ195071
Lygodium japonicum	Transcriptome, 1KP	PBUU	XXXXXXXX
Anemia tomentosa	Transcriptome, 1KP	CQPW	XXXXXXXX
Dipteris conjugata	Transcriptome, 1KP	MEKP	XXXXXXXX
Osmunda sp.	Transcriptome, 1KP	UOMY	KJ195072
Fern PHOT1/2			
Equisetum hymale	Transcriptome, 1KP	JVSZ	XXXXXXXX
Ophioglossum vulgatum	Transcriptome, 1KP	QHVS	XXXXXXXX
Sceptridium dissectum	Transcriptome, 1KP	EEAQ	XXXXXXXX
Botrypus virginianus	Transcriptome, 1KP	BEGM	XXXXXXXX
Psilotum nudum	Transcriptome, 1KP	QVMR	XXXXXXXX
Tmesipteris parva	Transcriptome, 1KP	ALVQ	XXXXXXXX
Selaginella PHOT1			
Selaginella moellendorffii 1-1	Genome, Phytozome	-	XM 002965438
Selaginella moellendorffii 1-2	Genome, Phytozome	-	XM 002982913
Selaginella willdenowii	Transcriptome, 1KP	KJYC	XXXXXXXX
Selaginella kraussiana	Transcriptome, 1KP	ZFGK	KJ195076
Selaginella acanthonota	Transcriptome, 1KP	ZYCD	XXXXXXXX
Selaginella selaginoides	Transcriptome, 1KP	KUXM	XXXXXXXX
Selaginella PHOT2			
Selaginella moellendorffii 2-1	Genome, Phytozome	-	XM 002971663
Selaginella moellendorffii 2-2	Genome, Phytozome	-	XM 002991543
Selaginella willdenowii	Transcriptome, 1KP	KJYC	XXXXXXXX
Selaginella kraussiana	Transcriptome, 1KP	ZFGK	KJ195077
Selaginella acanthonota	Transcriptome, 1KP	ZYCD	XXXXXXXX
Isoetales PHOT			
Isoetes tegetiformans	Transcriptome, 1KP	PKOX	KJ195078
Lycopodiales PHOT			
Pseudolycopodiella caroliniana	Transcriptome, 1KP	UPMJ	KJ195074
Diphasiastrum digitatum	Transcriptome, 1KP	WAFT	XXXXXXXX
Dendrolycopodium obscurum	Transcriptome, 1KP	XNXF	XXXXXXXX
Lycopodium deuterodensum	Transcriptome, 1KP	PQTO	KJ195073
Phylloglossum drummondii	Transcriptome, 1KP	ZZEI	XXXXXXXX
Huperzia lucidula	Transcriptome, 1KP	GKAG	KJ195075
Liverwort PHOT			
Sphaerocarpos texanus	Transcriptome, 1KP	HERT	KJ195086
Conocephalum conicum	Transcriptome, 1KP	ILBQ	KJ195085
Lunularia cruciata	Transcriptome, 1KP	TXVB	XXXXXXXX
Marchantia polymorpha	Transcriptome, 1KP	JPYU	KJ195084
Pellia neesiana	Transcriptome, 1KP	JHFI	KJ195083
Metzgeria crassipilis	Transcriptome, 1KP	NRWZ	KJ195082
Bazzania trilobata	Transcriptome, 1KP	WZYK	XXXXXXXX
Porella pinnata	Transcriptome, 1KP	UUHD	KJ195080
Radula lindenbergia	Transcriptome, 1KP	BNCU	XXXXXXXX
Scapania nemorosa	Transcriptome, 1KP	IRBN	KJ195079
Schistochila sp	Transcriptome, 1KP	LGOW	KJ195081
Moss PHOT1			
Sphagnum lescurii	Transcriptome, 1KP	GOWD	KJ195093
Moss PHOT1A			
Physcomitrella patens 1A-1	Genome, Phytozome	-	XM 001774204
Physcomitrella patens 1A-2	Genome, Phytozome	-	XM 001774562
Physcomitrella patens 1A-3	Genome, Phytozome	-	XM 001755269
Aulacomnium heterostichum	Transcriptome, 1KP	WNGH	XXXXXXXX
Buxbaumia aphylla	Transcriptome, 1KP	HRWG	XXXXXXXX
Orthotrichum lyellii	Transcriptome, 1KP	CMEQ	XXXXXXXX
Bryum argenteum	Transcriptome, 1KP	JMXW	KJ195089
Loeskeobryum brevirostre	Transcriptome, 1KP	WSPM	XXXXXXXX
Leucodon brachypus	Transcriptome, 1KP	ZACW	XXXXXXXX
Rhynchostegium serrulatum	Transcriptome, 1KP	JADL	XXXXXXXX
Scouleria aquatica	Transcriptome, 1KP	BPSG	KJ195088
Neckera douglasii	Transcriptome, 1KP	TMAJ	XXXXXXXX
Moss PHOT1B			
Atrichum angustatum	Transcriptome, 1KP	ZTHV	XXXXXXXX
Physcomitrella patens	Genome, Phytozome	-	XM 001765356

Species	Source	Code	Accession
Zygnematales NEO			
Mesotaenium endlicherianum	Transcriptome, 1KP	WDCW	XXXXXXXX
Mesotaenium braunii	Transcriptome, 1KP	WSJO	KJ195046
Mesotaenium braunii	Transcriptome, 1KP	WSJO	KJ195045
Cylindrocystis cushleckae	Transcriptome, 1KP	JOJQ	KJ195040
Zygnemopsis sp	Transcriptome, 1KP	MFZO	KJ195039
Cylindrocystis brebissonii 2	Transcriptome, 1KP	YOXI	KJ195044
Cylindrocystis sp 2	Transcriptome, 1KP	VAZE	KJ195043
Cylindrocystis brebissonii 1	Transcriptome, 1KP	YOXI	KJ195042
Cylindrocystis sp 1	Transcriptome, 1KP	VAZE	KJ195041
Mesotaenium caldariorum	Transcriptome, 1KP	HKZW	XXXXXXXX
Mougeotia scalaris NEO2	GenBank	-	AB206962
Mougeotia scalaris NEO1	GenBank	-	AB206961
Zygnematales PHOT?			
Mesotaenium endlicherianum	Transcriptome, 1KP	WDCW	XXXXXXXX
Mesotaenium braunii	Transcriptome, 1KP	WSJO	XXXXXXXX
Desmidiales PHOT			
Roya obtusa	Transcriptome, 1KP	XRTZ	XXXXXXXX
Gonatozygon kinahanii	Transcriptome, 1KP	KEYW	XXXXXXXX
Planotaenium ohtanii	Transcriptome, 1KP	SNOX	KJ195111
Phymatodocis nordstedtiana	Transcriptome, 1KP	RPQV	XXXXXXXX
Penium exiguum	Transcriptome, 1KP	YSQT	XXXXXXXX
Desmidium aptogonum	Transcriptome, 1KP	DFDS	KJ195109
Staurodesmus convergens	Transcriptome, 1KP	WCQU	XXXXXXXX
Cosmarium tinctum	Transcriptome, 1KP	BHBK	XXXXXXXX
Coleochaetales PHOT			
Coleochaete irregularis	Transcriptome, 1KP	QPDY	KJ195102
Coleochaete scutata	Transcriptome, 1KP	VQBJ	KJ195103
Chaetosphaeridium globosum	Transcriptome, 1KP	DRGY	XXXXXXXX
Klebsormidioles PHOT			
Interfilum paradoxum	Transcriptome, 1KP	FPCO	KJ195105
Klebsormidium subtile	Transcriptome, 1KP	FQLP	KJ195104
Entransia fimbriat	Transcriptome, 1KP	BFIK	KJ195106
Mesostigmatales PHOT			
Chlorokybus atmophyticus	Transcriptome, 1KP	AZZW	KJ195107
Mesostigma viride	Transcriptome, 1KP	KYIO	KJ195108
Spirotaenia minuta	Transcriptome, 1KP	NNHQ	XXXXXXXX
Prasinophyte PHOT			
Ostreococcus tauri	Genome, Phytozome	-	-
Ostreococcus lucimarinus	Genome, Phytozome	-	-
Micromonas pusilla	Genome, Phytozome	-	-
Dolichomastix tenuilepi	Transcriptome, 1KP	XOAL	XXXXXXXX
Pyramimonas parkeae	Transcriptome, 1KP	TNAW	XXXXXXXX
Scourfieldia sp	Transcriptome, 1KP	EGNB	XXXXXXXX
Nephroselmis olivace	Transcriptome, 1KP	MMKU	KJ195123
Tetraselmis cordiformis	Transcriptome, 1KP	DUMA	XXXXXXXX
Scherffelia dubia	Transcriptome, 1KP	FMVB	XXXXXXXX
Pycnococcus provasolii	Transcriptome, 1KP	MXEZ	XXXXXXXX
Ulvophyceae PHOT			
Bolbocoleon piliferum	Transcriptome, 1KP	LSHT	KJ195127
Persursaria percursa	Transcriptome, 1KP	OAEZ	KJ195126
Helicodictyon planctonicum	Transcriptome, 1KP	AJAU	XXXXXXXX
Entocladia endozoica	Transcriptome, 1KP	OQON	XXXXXXXX
Trebouxiophyceae PHOT			
Coccomyxa pringsheimii	Transcriptome, 1KP	GXBM	KJ195129
Botryococcus terribilis	Transcriptome, 1KP	QYXY	KJ195130
Prasiola crispa	Transcriptome, 1KP	WCLV	KJ195128
Trebouxia arboricola	Transcriptome, 1KP	NKXU	XXXXXXXX
Chlorophyceae PHOT			
Oedogonium foveolatum	Transcriptome, 1KP	SDPC	XXXXXXXX
Uronema sp.	Transcriptome, 1KP	ISGT	XXXXXXXX
Aphanochaete repens	Transcriptome, 1KP	IJMT	XXXXXXXX
Fritschiella tuberosa	Transcriptome, 1KP	VFIV	XXXXXXXX
Hafniomonas reticulata	Transcriptome, 1KP	FXHG	XXXXXXXX
Carteria obtusa	Transcriptome, 1KP	RUIF	XXXXXXXX
Scenedesmus dimorphus	Transcriptome, 1KP	PZIF	XXXXXXXX
Cylindrocapsa geminella	Transcriptome, 1KP	DZPJ	XXXXXXXX
Pediastrum duplex	Transcriptome, 1KP	XTON	XXXXXXXX
Brachiomonas submarina	Transcriptome, 1KP	GUBD	XXXXXXXX
Heterochlamydomonas inaequalis	Transcriptome, 1KP	IRYH	XXXXXXXX
Volvox carteri	Genome, Phytozome	-	-
Chlamydomonas reinhardtii	Genome, Phytozome	-	-
Oogamochlamys gigantea	Transcriptome, 1KP	XDLL	XXXXXXXX
Oedogonium cardiacum	Transcriptome, 1KP	DVYE	KJ195125
Chloromonas tughillensi	Transcriptome, 1KP	UTRE	KJ195124

[1]Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40,** D1178–86 (2012).

[2]Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *GigaScience* **3,** 17 (2014).

[3]Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* **342,** 1241089 (2013).

[4]Der, J. P., Barker, M. S., Wickett, N. J., Depamphilis, C. W. & Wolf P. G. De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics* **12,** 99 (2011).

[5]Li, F.-W. *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. USA* **111,** 6672–6677 (2014).

Appendix D: Supplementary Figures for Chapter Three

Go to
Figure 13

Figure 18: 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. Numbers above branches are maximum likelihood bootstrap values (BS) from Garli/BS from nhPhyML/aLRT SH-like supports under codon model (aLRT-SH)/Bayesian posterior probabilities (PP) from MrBayes/PP from BEAST; these are only displayed (along with thickened branches) when BS > 70, SH-aLRT > 70 and PP > 0.95. "+" denotes BS = 100, aLRT-SH = 100 or PP = 1.00; thickened branches without numbers are "+/+/+/+/+". Alphanumeric codes following species names are the four-letter 1KP transcriptome identifiers, 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. See **Figure 13** for the relationships of fern neochrome (NEO), hornwort neochrome and phototropin (PHOT).

Angiosperm PHYA

Gymnosperm PHYN

Angiosperm PHYC

Gymnosperm PHYO

Angiosperm PHYB

Angiosperm PHYE

Gymnosperm PHYP

Fern PHY2

Fern PHY4

Fern PHY2/4

Fern PHY1

Lycophyte PHY

Liverwort PHY

Hornwort PHY

Moss PHY

Fern NEO

Hornwort NEO

Zygnematophyceae PHY

Coleochaetophyceae PHY

Klebsormidiophyceae PHY

0.2 substitutions/site

84

Figure 19: 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.

Figure 20: Phylogeny of land plant imidazoleglycerol-phosphate dehydratase (IGPD). *N. aenigmaticus* pseudogene (denoted by ψ) is most closely related to other hornwort IGPD, confirming that hornwort neochrome is indeed in the hornwort genome, not from symbiotic algae or fungi. Numbers associated with branches are maximum likelihood bootstrap support. Alphanumeric codes following species names are the 1KP transcriptome identifiers, or Genbank accessions, or both; "†" indicates the sequence came from whole genome sequence.

Figure 21: 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).

Appendix E: Supplementary Tables for Chapter Three

Table 6: 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 name 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 argyrophylla NPRL
Cedrus libani GGEA
Cephalotaxus harringtonia WYAJ
Cryptomeria japonica DSXO
Cunninghamia lanceolata OUOI
Cupressus dupreziana GNGJ
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 hypophyllus 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
Botrypus virginianus BEGM
Cibotium glaucum ORIE
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 YLIA
Polypodium glycyrrhiza CJNT
Polypodium hesperium VBMM
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 appressa 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 RGXI
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 sp. 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
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 lemaneiformii 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 7: The calibrations used in dating the divergence of phototropin gene family

No.	Clade	Calibration	Date (MYA)	Prior	Reference	Justification
1	Tracheophyta	*Zosterophyllum sp.*	416	lognormal (mean: 3.5, STD: 1, offset: 416)	71, 75	Oldest unequivocal record of total group of lycoposid; see Ref. 71 for detailed justifications
2	Euphyllophyta	*Ibyka sp.*	388.2	lognormal (mean: 3.5, STD: 1, offset: 388.2)	71, 76	Oldest unequivocal record of monilophyte based on protoxylem morphology; see Ref. 71 for detailed justifications
3	Bryophyta PHOTA	"type III" fragment	330.9-346.7	lognormal (mean: 1.5, STD: 1, offset: 330.9)	72	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)	72	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)	73	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, 77	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, 77	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)	71, 78	Oldest unequivocal record of Acrogymnospermae; see Ref. 71 for detailed justifications
11	Spermatophyta PHOT2	*Cordaixylon iowensis*	306.2	lognormal (mean: 2.5, STD: 1, offset: 306.2)	71, 78	Oldest unequivocal record of Acrogymnospermae; see Ref. 71 for detailed justifications
12	Grass PHOT1	phytoliths in dinosaur coprolites	65-67	lognormal (mean: 1.5, STD: 1, offset: 65)	74	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)	74	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)	71	Oldest unequivocal record of Cupressophyta crown; see Ref. 71 for detailed justifications
15	Coniferae PHOT2	*Araucaria mirabilis*	147	lognormal (mean: 1.5, STD: 1, offset: 147)	71	Oldest unequivocal record of Cupressophyta crown; see Ref. 71 for detailed justifications

Table 8: The primers and PCR protocols used in this study.

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 9: The primer sequences used in PCR.

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

References

Abascal, F., Zardoya, R., and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**: 2104–2105.

Adl, S.M. et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. **52**: 399–451.

Anisimova, M. and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. Syst. Biol. **55**: 539–552.

Banks, J.A. et al. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. Science **332**: 960–963.

Bergthorsson, U., Adams, K.L., Thomason, B., and Palmer, J.D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature **424**: 197–201.

Bergthorsson, U., Richardson, A.O., Young, G.J., Goertzen, L.R., and Palmer, J.D. (2004). Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc. Natl. Acad. Sci. USA **101**: 17747–17752.

Boussau, B. and Gouy, M. (2006). Efficient likelihood computations with nonreversible models of evolution. Syst. Biol. **55**: 756–768.

Bremer, B. et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc. **161**: 105–121.

Burki, F., Okamoto, N., Pombert, J.F., and Keeling, P.J. (2012). The evolutionary history of haptophytes and cryptophytes: phylogenomic evidence for separate origins. Proc. R. Soc. Lond. B **279**: 2246–2254.

Burr, F.A. (1968). *Chloroplast Structure and Division in Megaceros Species, PhD dissertation, University of California at Berkeley*.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics **10**: 421.

Cavalier-Smith, T., Chao, E.E., Snell, E.A., Berney, C., Fiore-Donno, A.M., and Lewis, R. (2014). Multigene eukaryote phylogeny reveals the likely protozoan ancestors of

opisthokonts (animals, fungi, choanozoans) and Amoebozoa. Mol. Phylogenet. Evol. **81C**: 71–85.

Chang, Y. and Graham, S.W. (2011). Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. Am. J. Bot **98**: 839–849.

Chen, M. and Chory, J. (2011). Phytochrome signaling mechanisms and the control of plant development. Trends Cell Biol. **21**: 664–671.

Cho, Y., Qiu, Y.L., Kuhlman, P., and Palmer, J.D. (1998). Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA **95**: 14244–14249.

Christie, J.M. (2007). Phototropin blue-light receptors. Annu. Rev. Plant. Biol. **58**: 21–45.

Christin, P.-A., Edwards, E.J., Besnard, G., Boxall, S.F., Gregory, R., Kellogg, E.A., Hartwell, J., and Osborne, C.P. (2012). Adaptive evolution of C_4 photosynthesis through recurrent lateral gene transfer. Curr Biol **22**: 445–449.

Church, S.H., Ryan, J.F., and Dunn, C.W. (2014) Sowhat, Available from GitHub repository https://github.com/josephryan/sowhat.

Clarke, J.T., Warnock, R.C.M., and Donoghue, P.C.J. (2011). Establishing a time-scale for plant evolution. New Phytol **192**: 266–301.

Cooke, T.J., Hickok, L.G., Van der Woude, W.J., Banks, J.A., and Scott, R.J. (1993). Photobiological characterization of a spore germination mutant *dkg1* with reversed photoregulation in the fern *Ceratopteris richardii*. Photochem. Photobiol. **57**: 1032–1041.

Cox, C.J., Goffinet, B., Wickett, N.J., Boles, S.B., and Shaw, A.J. (2010). Moss diversity: A molecular phylogenetic analysis of genera. Phytotaxa **9**: 175–195.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics **27**: 1164–1165.

Darwin, C. and Darwin, S.F. (1880). The power of movement in plants (John Murray).

Davis, C.C. and Wurdack, K.J. (2004). Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science **305**: 676–678.

Davis, C.C., Anderson, W.R., and Wurdack, K.J. (2005). Gene transfer from a parasitic flowering plant to a fern. Proc. R. Soc. Lond. B **272**: 2237–2242.

de Carbonnel, M., Davis, P., Roelfsema, M.R.G., Inoue, S.-I., Schepens, I., Lariguet, P., Geisler, M., Shimazaki, K.-I., Hangarter, R., and Fankhauser, C. (2010). The Arabidopsis PHYTOCHROME KINASE SUBSTRATE2 protein is a phototropin signaling element that regulates leaf flattening and leaf positioning. Plant Physiol. **152**: 1391–1405.

Delaroque, N., Müller, D.G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001). The complete DNA sequence of the *Ectocarpus siliculosus* Virus EsV-1 genome. Virology **287**: 112–132.

Delport, W., Poon, A.F.Y., Frost, S.D.W., and Kosakovsky Pond, S.L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics **26**: 2455–2457.

Demarsy, E., Schepens, I., Okajima, K., Hersch, M., Bergmann, S., Christie, J., Shimazaki, K.-I., Tokutomi, S., and Fankhauser, C. (2012). Phytochrome Kinase Substrate 4 is phosphorylated by the phototropin 1 photoreceptor. Embo J **31**: 3457–3467.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. **29**: 1969–1973.

Duanmu, D. et al. (2014). Marine algae and land plants share conserved phytochrome signaling systems. Proc. Natl. Acad. Sci. USA **111**: 15827–15832.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**: 1792–1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**: 2460–2461.

Fankhauser, C. (2000). Phytochromes as light-modulated protein kinases. Semin. Cell Dev. Biol. **11**: 467–473.

Forrest, L.L., Davis, E.C., Long, D.G., Crandall-Stotler, B.J., Clark, A., and Hollingsworth, M.L. (2006). Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. Bryologist **109**: 303–334.

Franklin, K.A. and Quail, P.H. (2010). Phytochrome functions in Arabidopsis development. J. Exp. Bot. **61**: 11–24.

Galen, C., Huddle, J., and Liscum, E. (2004). An experimental test of the adaptive

evolution of phototropins: blue-light photoreceptors controlling phototropism in Arabidopsis thaliana. Evolution **58**: 515–523.

Galtier, J., Wang, S.-J., Li, C.-S., and Hilton, J. (2001). A new genus of filicalean fern from the Lower Permian of China. Bot. J. Linn. Soc. **137**: 429–442.

Galván-Ampudia, C.S. and Offringa, R. (2007). Plant evolution: AGC kinases tell the auxin tale. Trends Plant Sci. **12**: 541–547.

Gil, M., Zanetti, M.S., Zoller, S., and Anisimova, M. (2013). CodonPhyML: fast maximum likelihood phylogeny estimation under codon substitution models. Mol. Biol. Evol. **30**: 1270–1280.

Goldman, N. and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**: 725–736.

Goldman, N., Anderson, J.P., and Rodrigo, A.G. (2000). Likelihood-based tests of topologies in phylogenetics. Syst. Biol. **49**: 652–670.

Gontcharov, A.A. and Melkonian, M. (2010). Molecular phylogeny and revision of the genus *Netrium* (Zygnematophyceae, Streptophyta): *Nucleotaenium* gen. nov. J. Phycol. **46**: 346–362.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. **40**: D1178–86.

Grabherr, M.G. et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29**: 644–652.

Grant, J.R. and Katz, L.A. (2014). Building a phylogenomic pipeline for the eukaryotic tree of life – addressing deep phylogenies with genome-scale data. PLoS Curr. **6**.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**: 307–321.

Guo, C.-Q., Edwards, D., Wu, P.-C., Duckett, J.G., Hueber, F.M., and Li, C.-S. (2012). *Riccardiothallus devonicus* gen. et sp. nov., the earliest simple thalloid liverwort from the Lower Devonian of Yunnan, China. Rev. Palaeobot. Palynol. **176-177**: 35–40.

Haupt, W. and Scheuerlein, R. (1990). Chloroplast movement. Plant Cell Environ. **13**:

595–614.

Hedges, S.B. and Kumar, S. (2009). The Timetree of Life (Oxford University Press, Oxford).

Heijde, M. and Ulm, R. (2012). UV-B photoreceptor-mediated signalling in plants. Trends Plant Sci. **17**: 230–237.

Herdman, M., Coursin, T., Rippka, R., Houmard, J., and de Marsac, N.T. (2000). A new appraisal of the prokaryotic origin of eukaryotic phytochromes. J. Mol. Evol. **51**: 205–213.

Hillis, D.M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. **47**: 3–8.

Huang, J. (2013). Horizontal gene transfer in eukaryotes: the weak-link model. BioEssays **35**: 868–875.

Huang, K. and Beck, C. (2003). Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **100**: 6269.

Huang, X. (1999). CAP3: A DNA Sequence Assembly Program. Genome Res **9**: 868–877.

Hubers, M. and Kerp, H. (2012). Oldest known mosses discovered in Mississippian (late Visean) strata of Germany. Geol **40**: 755–758.

Ikeda, H. and Setoguchi, H. (2010). Natural selection on *PHYE* by latitude in the Japanese archipelago: insight from locus specific phylogeographic structure in *Arcterica nana* (Ericaceae). Mol. Ecol. **19**: 2779–2791.

Ikeda, H., Fujii, N., and Setoguchi, H. (2009). Molecular evolution of phytochromes in *Cardamine nipponica* (Brassicaceae) suggests the involvement of *PHYE* in local adaptation. Genetics **182**: 603–614.

Jaedicke, K., Lichtenthaler, A.L., Meyberg, R., Zeidler, M., and Hughes, J. (2012). A phytochrome-phototropin light signaling complex at the plasma membrane. Proc. Natl. Acad. Sci. USA **109**: 12231–12236.

Johnson, M.T.J. et al. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS ONE **7**: e50226.

Joshi, N.A. and Fass, J.N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) Available at https://github.com/najoshi/sickle.

Kagawa, T., Kasahara, M., Abe, T., Yoshida, S., and Wada, M. (2004). Function analysis of phototropin2 using fern mutants deficient in blue light-induced chloroplast avoidance movement. Plant Cell Physiol. **45**: 416–426.

Kagawa, T., Sakai, T., Suetsugu, N., Oikawa, K., Ishiguro, S., Kato, T., Tabata, S., Okada, K., and Wada, M. (2001). Arabidopsis NPL1: A phototropin homolog controlling the chloroplast high-light avoidance response. Science **291**: 2138–2141.

Kanegae, T., Hayashida, E., Kuramoto, C., and Wada, M. (2006). A single chromoprotein with triple chromophores acts as both a phytochrome and a phototropin. Proc. Natl. Acad. Sci. USA **103**: 17997–18001.

Kapralov, M.V. and Filatov, D.A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. BMC. Evol. Biol. **7**: 73.

Karniol, B., Wagner, J.R., Walker, J.M., and Vierstra, R.D. (2005). Phylogenetic analysis of the phytochrome superfamily reveals distinct microbial subfamilies of photoreceptors. Biochem. J. **392**: 103–116.

Kasahara, M., Kagawa, T., Sato, Y., Kiyosue, T., and Wada, M. (2004). Phototropins mediate blue and red light-induced chloroplast movements in Physcomitrella patens. Plant Physiol. **135**: 1388–1397.

Kawai, H., Kanegae, T., Christensen, S., Kiyosue, T., Sato, Y., Imaizumi, T., Kadota, A., and Wada, M. (2003). Responses of ferns to red light are mediated by an unconventional photoreceptor. Nature **421**: 287–290.

Keeling, P.J. (2004). Diversity and evolutionary history of plastids and their hosts. Am. J. Bot **91**: 1481–1493.

Komatsu, A., Terai, M., Ishizaki, K., Suetsugu, N., Tsuboi, H., Nishihama, R., Yamato, K.T., Wada, M., and Kohchi, T. (2014). Phototropin encoded by a single-copy gene mediates chloroplast photorelocation movements in the liverwort *Marchantia polymorpha*. Plant Physiol. **166**: 411–427.

Korall, P. and Kenrick, P. (2002). Phylogenetic relationships in Selaginellaceae based on *rbcL* sequences. Am. J. Bot **89**: 506–517.

Kosakovsky Pond, S.L., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. Bioinformatics **21**: 676–679.

Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., and Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. **28**: 3033–3043.

Kotyk, M.E., Basinger, J.F., Gensel, P.G., and de Freitas, T.A. (2002). Morphologically complex plant macrofossils from the Late Silurian of Arctic Canada. Am. J. Bot **89**: 1004–1013.

Kuo, L.-Y., Li, F.-W., Chiou, W.-L., and Wang, C.-N. (2011). First insights into fern *matK* phylogeny. Mol. Phylogenet. Evol. **59**: 556–566.

Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. **29**: 1695–1701.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth **9**: 357–359.

Lariguet, P., Schepens, I., Hodgson, D., Pedmale, U., Trevisan, M., Kami, C., De Carbonnel, M., Alonso, J., Ecker, J., and Liscum, E. (2006). PHYTOCHROME KINASE SUBSTRATE 1 is a phototropin 1 binding protein required for phototropism. Proc. Natl. Acad. Sci. USA **103**: 10134–10139.

Li, C., Hofreiter, M., Straube, N., Corrigan, S., and Naylor, G.J.P. (2013). Capturing protein-coding genes across highly divergent species. BioTechniques **54**: 321–326.

Li, F.-W. et al. (2014). Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proc. Natl. Acad. Sci. USA **111**: 6672–6677.

Lind, P.A., Tobin, C., Berg, O.G., Kurland, C.G., and Andersson, D.I. (2010). Compensatory gene amplification restores fitness after inter-species gene replacements. Mol. Microbiol. **75**: 1078–1089.

Liscum, E., Askinosie, S.K., Leuchtman, D.L., Morrow, J., Willenburg, K.T., and Coats, D.R. (2014). Phototropism: growing towards an understanding of plant movement. Plant Cell **26**: 38–55.

Luo, R. et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience **1**: 18.

MacCallum, I. et al. (2009). ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol **10**: R103.

Marchler-Bauer, A. et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. **39**: D225–9.

Marin, B. (2012). Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. Protist **163**: 778–805.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal **17**: 1–3.

Mathews, S. (2010). Evolutionary studies illuminate the structural-functional model of plant phytochromes. Plant Cell **22**: 4–16.

Mathews, S. (2006). Phytochrome-mediated development in land plants: red light sensing evolves to meet the challenges of changing light environments. Mol. Ecol. **15**: 3483–3503.

Mathews, S. and Tremonte, D. (2012). Tests of the link between functional innovation and positive selection at phytochrome A: the phylogenetic distribution of far-red high-irradiance responses in seedling development. Int. J. Plant Sci. **173**: 662–672.

Mathews, S., Burleigh, J.G., and Donoghue, M.J. (2003). Adaptive evolution in the photosensory domain of phytochrome A in early angiosperms. Mol. Biol. Evol. **20**: 1087–1097.

Mathews, S., Clements, M.D., and Beilstein, M.A. (2010). A duplicate gene rooting of seed plants and the phylogenetic position of flowering plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. **365**: 383–395.

Mittmann, F., Dienstbach, S., Weisert, A., and Forreiter, C. (2009). Analysis of the phytochrome gene family in *Ceratodon purpureus* by gene targeting reveals the primary phytochrome responsible for photo- and polarotropism. Planta **230**: 27–37.

Möglich, A., Yang, X., Ayers, R.A., and Moffat, K. (2010). Structure and function of plant photoreceptors. Annu. Rev. Plant. Biol. **61**: 21–47.

Nabholz, B., Künstner, A., Wang, R., Jarvis, E.D., and Ellegren, H. (2011). Dynamic evolution of base composition: causes and consequences in avian phylogenomics.

Mol. Biol. Evol. **28**: 2197–2210.

Näsvall, J., Sun, L., Roth, J.R., and Andersson, D.I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. Science **338**: 384–387.

Nozue, K., Kanegae, T., and Wada, M. (1997). A full length *Ty3/gypsy-type* retrotransposon in the fern *Adiantum*. J. Plant Res. **110**: 495–499.

Nozue, K., Kanegae, T., Imaizumi, T., Fukuda, S., Okamoto, H., Yeh, K., Lagarias, J., and Wada, M. (1998). A phytochrome from the fern *Adiantum* with features of the putative photoreceptor NPH1. Proc. Natl. Acad. Sci. USA **95**: 15826–15830.

Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics **120**: 621–623.

Okamoto, H., Hirano, Y., Abe, H., Tomizawa, K., FURUYA, M., and Wada, M. (1993). The deduced amino acid sequence of phytochrome from *Adiantum* includes consensus motifs present in phytochrome B from seed plants. Plant Cell Physiol. **34**: 1329–1334.

Pasentsis, K., Paulo, N., Algarra, P., Dittrich, P., and Thümmler, F. (1998). Characterization and expression of the phytochrome gene family in the moss *Ceratodon purpureus*. Plant J. **13**: 51–61.

Possart, A. and Hiltbrunner, A. (2013). An evolutionarily conserved signaling mechanism mediates far-red light responses in land plants. Plant Cell **25**: 102–114.

Possart, A., Fleck, C., and Hiltbrunner, A. (2014). Shedding (far-red) light on phytochrome mechanisms and responses in land plants. Plant Sci. **217-218**: 36–46.

Prasad, V., Strömberg, C.A.E., Alimohammadian, H., and Sahni, A. (2005). Dinosaur coprolites and the early evolution of grasses and grazers. Science **310**: 1177–1180.

Qiu, Y.-L. et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA **103**: 15511–15516.

Rambaut, A. and Drummond, A.J. (2013). Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/.

Rambaut, A. and Grassly, N.C. (1997). Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. **13**: 235–238.

Renner, S.S. and Bellot, S. (2012). Horizontal Gene Transfer in Eukaryotes: Fungi-to-Plant and Plant-to-Plant Transfers of Organellar DNA. In Genomics of Chloroplasts and Mitochondria, Advances in Photosynthesis and Respiration. (Springer Netherlands: Dordrecht), pp. 223–235.

Rensing, S.A. et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64–69.

Rockwell, N.C., Duanmu, D., Martin, S.S., Bachy, C., Price, D.C., Bhattacharya, D., Worden, A.Z., and Lagarias, J.C. (2014). Eukaryotic algal phytochromes span the visible spectrum. Proc. Natl. Acad. Sci. USA **111**: 3871–3876.

Rockwell, N.C., Su, Y.-S., and Lagarias, J.C. (2006). Phytochrome structure and signaling mechanisms. Annu. Rev. Plant. Biol. **57**: 837–858.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. **61**: 539–542.

Rothfels, C.J., Larsson, A., Li, F.-W., Sigel, E.M., Huiet, L., Burge, D.O., Ruhsam, M., Graham, S.W., Stevenson, D.W., Wong, G.K.-S., Korall, P., and Pryer, K.M. (2013). Transcriptome-mining for single-copy nuclear markers in ferns. PLoS ONE **8**: e76957.

Sakai, T., Kagawa, T., Kasahara, M., Swartz, T.E., Christie, J.M., Briggs, W.R., Wada, M., and Okada, K. (2001). Arabidopsis nph1 and npl1: blue light receptors that mediate both phototropism and chloroplast relocation. Proc. Natl. Acad. Sci. USA **98**: 6969–6974.

Sanchez-Puerta, M., Abbona, C., Zhuo, S., Tepe, E., Bohs, L., Olmstead, R., and Palmer, J. (2011). Multiple recent horizontal transfers of the *cox1* intron in Solanaceae and extended co-conversion of flanking exons. BMC. Evol. Biol. **11**: 277.

Sanchez-Puerta, M.V., Cho, Y., Mower, J.P., Alverson, A.J., and Palmer, J.D. (2008). Frequent, phylogenetically local horizontal transfer of the *cox1* group I Intron in flowering plant mitochondria. Mol. Biol. Evol. **25**: 1762–1777.

Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution, divergence times in the absence of a molecular clock. Bioinformatics **19**: 301–302.

Sanderson, M.J. and Shaffer, H.B. (2002). Troubleshooting molecular phylogenetic

analyses. Annu Rev Ecol Syst **33**: 49–72.

Schneider, H., Schuettpelz, E., Pryer, K.M., Cranfill, R., Magallón, S., and Lupia, R. (2004). Ferns diversified in the shadow of angiosperms. Nature **428**: 553–557.

Schneider-Poetsch, H.A.W., Marx, S., Kolukisaoglu, U., Hanelt, S., and Birgit, B. (1994). Phytochrome evolution: Phytochrome genes in ferns and mosses. Physiol. Plantarum **91**: 241–250.

Schuettpelz, E. and Pryer, K.M. (2009). Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. Proc. Natl. Acad. Sci. USA **106**: 11200–11205.

Schuettpelz, E. and Pryer, K.M. (2007). Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon **56**: 1037–1050.

Sineshchekov, V., Koppel, L., Okamoto, H., and Wada, M. (2013). Fern *Adiantum capillus-veneris* phytochrome 1 comprises two native photochemical types similar to seed plant phytochrome A. J. Photochem. Photobiol. B **130C**: 20–29.

Skog, J.E. and Banks, H.P. (1973). *Ibyka amphikoma*, gen. et sp. n., a new protoarticulate precursor from the late Middle Devonian of New York State. Am. J. Bot **60**: 366–380.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**: 2688–2690.

Stein, W.E., Mannolini, F., Hernick, L.V., Landing, E., and Berry, C.M. (2007). Giant cladoxylopsid trees resolve the enigma of the Earth's earliest forest stumps at Gilboa. Nature **446**: 904–907.

Suetsugu, N., Mittmann, F., Wagner, G., Hughes, J., and Wada, M. (2005). A chimeric photoreceptor gene, *NEOCHROME*, has arisen twice during plant evolution. Proc. Natl. Acad. Sci. USA **102**: 13705–13709.

Suzuki, T., Takio, S., Yamamoto, I., and Satoh, T. (2001). Characterization of cDNA of the liverwort phytochrome gene, and phytochrome involvement in the light-dependent and light-independent protochlorophyllide oxidoreductase gene expression in *Marchantia paleacea* var. *diptera*. Plant Cell Physiol. **42**: 576–582.

Swofford, D.L. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0a131 (Sinauer Associates, Sunderland, MA).

Thümmler, F., Dufner, M., Kreisl, P., and Dittrich, P. (1992). Molecular cloning of a

novel phytochrome gene of the moss *Ceratodon purpureus* which encodes a putative light-regulated protein kinase. Plant Mol. Biol. **20**: 1003–1017.

Trivett, M.L. (1992). Growth architecture, structure, and relationships of *Cordaixylon iowensis* nov. comb. (Cordaitales). Int. J. Plant Sci. **153**: 273–287.

Ulijasz, A.T. and Vierstra, R.D. (2011). Phytochrome structure and photochemistry: recent advances toward a complete molecular picture. Curr. Opin. Plant. Biol. **14**: 498–506.

Verbyla, K.L., Yap, V.B., Pahwa, A., Shao, Y., and Huttley, G.A. (2013). The embedding problem for Markov models of nucleotide substitution. PLoS ONE **8**: e69187.

Villarreal, J.C. and Renner, S.S. (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**: 18873–18878.

Wickett, N.J. et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA **111**: E4859–E4868.

Wikstrom, N. (2001). Evolution of Lycopodiaceae (Lycopsida): estimating divergence times from *rbcL* gene sequences by use of nonparametric rate smoothing. Mol. Phylogenet. Evol. **19**: 177–186.

Xi, Z., Bradley, R.K., Wurdack, K.J., Wong, K., Sugumaran, M., Bomblies, K., Rest, J.S., and Davis, C.C. (2012). Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics **13**: 227.

Xi, Z., Wang, Y., Bradley, R.K., Sugumaran, M., Marx, C.J., Rest, J.S., and Davis, C.C. (2013). Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genetics **9**: e1003265.

Yang, Y., Qi, X., Sen, L., Su, Y., and 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**: 1–15.

Yanovsky, M.J., Casal, J.J., and Whitelam, G.C. (1995). Phytochrome A, phytochrome B and Hy4 are involved in hypocotyl growth responses to natural radiation in *Arabidopsis*: weak de-etiolation of the *phyA* mutant under dense canopies. Plant Cell Environ. **18**: 788–794.

Yeh, K.C. and Lagarias, J.C. (1998). Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. Proc. Natl. Acad. Sci. USA **95**: 13976–13981.

Yoon, H.S., Muller, K.M., Sheath, R.G., Ott, F.D., and Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta). J. Phycol. **42**: 482–492.

Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K. (2010). Horizontal gene transfer by the parasitic plant *Striga hermonthica*. Science **328**: 1128.

Zerbino, D.R. and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18**: 821–829.

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.

Zwickl, D.J. (2006). *Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets Under the Maximum Likelihood Criterion*, PhD dissertation, University of Texas.

Biography

I was born on September 28, 1987, in Taipei, Taiwan. I pursued my undergraduate education at Department of Life Sciences, National Taiwan University and earned a Bachelor of Science in 2009. I then served in ROC Army for one year as a tank platoon leader and ammunition officer before starting graduate school at Duke.

The papers I have authored include:

As first author—

Li, F.W., M. Melkonian, C.J. Rothfels, J.C. Villarreal, D.W. Stevenson, S.W. Graham, G.K.S. Wong, K.M. Pryer, and S. Mathews. Novel phytochrome lineages and complex evolutionary histories revealed across extant plant diversity. **Nature Communications**, *in review*

Li, F.W., K.M. Pryer. 2014. Crowdfunding the *Azolla* fern genome project: a grassroots approach. **GigaScience** 3: 16.

Li, F.W., *…32 co-authors…*, S. Mathews & K.M. Pryer. 2014. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. **Proceedings of the National Academy of Sciences USA**, 111: 6672-6677.

Li, F.W., K.M. Pryer & M.D. Windham. 2012. *Gaga*, a new fern genus segregated from *Cheilanthes* (Pteridaceae). **Systematic Botany** 37: 845-860.

Li, F.W., L.Y. Kuo, C.J. Rothfels, A. Ebihara, W.L. Chiou, M.D. Windham & K.M. Pryer. 2011. *rbcL* and *matK* earn a thumbs up as the core DNA barcode for ferns. **PLoS ONE** 6: e26597.

Li, F.W. 2011. Book review: Knapp, Ralf. 2011. Ferns and Fern Allies of Taiwan. **Taxon** 60: 1233-1234.

Kuo, L.Y.*, **F.W. Li***, W.L. Chiou & C.N. Wang. 2011. The first insight into fern *matK* phylogeny. **Molecular Phylogenetics and Evolution** 59: 556-566.

*Equal contributions

Li, F.W.*, L.Y. Kuo*, Y.M. Huang, W.L. Chiou & C.N. Wang. 2010. Tissue-Direct PCR, a rapid and extraction-free method for barcoding of ferns. **Molecular Ecology Resources** 10: 92-95.

*Equal contributions

Li, F.W., B.C. Tan, V. Buchbender, R.C. Moran, G. Rouhan, C.N. Wang & D. Quandt. 2009. Identifying a mysterious aquatic fern gametophyte. **Plant Systematics and Evolution** 281: 77-86.

As co-author—

Windham M.D., J.B. Beck, **F.W. Li**, L. Allphin, J.G. Carmen, D.A. Sherwood, C.A. Rushworth, E.M. Sigel, P.J. Alexander, C.D. Bailey, and I.A. Al-Shehbaz. Searching for diamonds in the apomictic rough. I: Do rare sexual populations of *Boechera lignifera* (Brassicaceae) represent a distinct species? **Systematic Botany**, *in review*.

Wolf, P.G., E.B. Sessa, D.B. Marchant, **F.W. Li**, C.J. Rothfels, E.M. Sigel, M.A. Gitzendanner, C.J. Visger, J.A. Banks, D.E. Soltis, P.S. Soltis, K.M. Pryer, J.P. Der. An exploration into fern genome space. **Genome Biology and Evolution**, *in review.*

Rothfels C.J., **F.W. Li**, E.M. Sigel, L. Huiet, A. Larsson, D.O. Burge, M. Ruhsam, M. Deyholos, D. Soltis, N. Stewart, S. Shaw, L.M. Pokorny, T. Chen, C. dePamphilis, L. DeGironimo, D.W. Stevenson, S.W. Graham, G.K.-S. Wong, and K.M. Pryer. The evolutionary history of ferns inferred from 25 single-copy nuclear genes. **American Journal of Botany**, *in review*.

Pryer, K.M., L. Huiet, **F.W. Li**, C.J. Rothfels, E. Schuettpelz. Maidenhair ferns—*Adiantum*—are indeed monophyletic, and sister to to the shoestring ferns—vittaroids (Pteridaceae). **American Journal of Botany**, *in review*.

Sessa, E.B., J.A. Banks, M.S. Barker, J.P. Der, A.M. Duffy, S.W. Graham, M. Hasebe, J. Langdale, **F.W. Li**, D.B. Marchant, K.M. Pryer, C.J. Rothfels, S.J. Roux, M.L. Salmi, E.M. Sigel, D.E. Soltis, P.S. Soltis, D.W. Stevenson, P.G. Wolf. 2014. Between two fern genomes. **GigaScience** 3: 15.

Zhang, W.Y., L.Y. Kuo, **F.W. Li,** C.N. Wang & W.L. Chiou. The hybrid origin of *Adiantum meishanianum* (Pteridaceae): a rare and endemic species in Taiwan. **Systematic Botany,** 39.

Rothfels, C.J., A. Larsson, **F.W. Li**, E.M. Sigel, L. Huiet, D.O. Burge, M. Ruhsam, S. Graham, D. Stevenson, G.K.S. Wong, P. Korall & K.M. Pryer. 2013. Transcriptome-minging for fern single-copy nuclear regions. **PLoS ONE** 8: e76957.

Scholarships, fellowships, and academic honors that I have received since my undergraduate education include: **Edgar T. Wherry Award**, Botanical Society of America (2014); **Graduate Student Research Fellowship**, Torrey Botanical Society (2014; US $2,500); **NSF Doctoral Dissertation Improvement Grant**, National Science Foundation (2014; US $13,000); **Duke Biology Grant-in-Aid**, Department of Biology, Duke University (2012, 2013, 2014; US $1,000 each year); **Shirley and Alan Graham Graduate Student Research Grant**, American Society of Plant Taxonomists (2013; US $1,000); **NSF Graduate Research Fellowship**, National

Science Foundation (2012; US $30,000/year for 3 years); **Sigma Xi Matching Grant**, Duke

Graduate School (2012; US $1,000); **Sigma Xi Grant-in-Aid of Research**, Sigma Xi (2012; US

$1,000); **Awards for Graduate Student Research**, Society of Systematic Biologists (2011; US

$1,700); **Dr. AT Chen Research Scholarship**, National Taiwan University (2009; US $2,400); **Dean

Award**, College of Life Science, National Taiwan University (2009); **Scientific Research Award**,

College of Life Science, National Taiwan University (2009); **Undergraduate Research Grant**,

National Science Council, Taiwan (2007, 2008 US $1,400 each year); **Presidential Award**, National

Taiwan University (2007).