Gene Duplications and the Evolution of Phytochrome in Early Evolving Land Plants

Justyn Koenig and Katelyn

2021-12-07 22:57:59

[1]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA
[2]Center for Computation & Technology, Louisiana State University, Baton Rouge, LA, USA

* **Corresponding author**, email: jkoeni9@lsu.edu; 377 Life Science Building, Baton Rouge, LA 70803

Running headline: Gene Duplications and the Evolution of Phytochrome in Early Evolving Land Plants

Abstract: 450 million years ago land was dominated by small nonvascular plants that were exposed to sunlight. 300 million years ago, the evolution of vascular plants led to the origin of the first forest with complex canopies and deeply shaded habitats. Around 100 million years ago, complex canopy shade was a major factor in early land plant environments. The complex canopies formed by forest collected most of the available light, leaving only red and far-red wavelengths reaching the understory plants. Plants had to evolve complex phytochrome responses to the differences in light and dark environments. The two proteins that evolved were called Pfr (far red) absorbing light at 730nm and Pr (red) absorbing light at 670nm. The ratio between the presence of red and far-red proteins is crucial for plants survival in different light environments. Phytochrome responses to shade (low R/FR) include (1) inhibition of germination and (2) initiation of shade avoidance responses (less roots produced, early flowering). Phytochrome A (associated with far red light) and Phytochrome B (associated with red light) are the two phytochromes that will be of interest in this study. Phytochrome A and Phytochrome B have been found extensively in seed plants and ferns. Going back further in the plant evolutionary timeline, there is a knowledge gap regarding phytochromes A and B presence in primitive nonvascular plant species. The group of interest is Bryophyta, which has little to no insight into how Phytochrome A and Phytochrome B have evolved. Understanding the prevalence of PHY A and PHY B in the non-vascular clade of Bryophyta is important for gaining insight into the transition periods that cemented the phytochromes from early nonvascular plants to the vascular land plants we see today. Sarah Mathews and Fay Wei Li, as well as GenBank will be the source of the data sets used. GenBank will be used to get data for the organismal clade sequences. The alignments collected from GenBank will then be used in BLAST to find the specific genes of interest. The alignments will then be read, explored, and assessed. Gene trees will be generated in MAFT and inferred in R, APE, and IQ-TREE. Further analysis will be run in Datamonkey to test for positive selection among different branches of the moss phylogenetic tree.

Introduction

The origin of land plants dates to 470 – 515 million years ago (@Li_FW_2020). Coding sequences of genes found in plants can unlock the key to understanding the evolution of land plants at a molecular level (@Li_FW_2015). When plants first colonized land, their environment was much different than it is today. Shade was nowhere to be found, as large forest had not yet evolved. The earliest forms of terrestrial plants grew close to the ground, exposed to the sun's harsh rays. As time went on, large trees evolved, with thick canopies, casting shade over more ancestral lineages of plants such as the Bryophyta or moss. The adaptions needed for Bryophyta to adapt to the changes of light in these new environments has been well understudied and overlooked. Gene duplication found to occur in the evolution of this lineage could hold insight into how land plants have evolved to adapt to changing light environments. This study aims to understand how genes evolve after gene duplication events occur, with focus on Bryophyta. Phytochrome is the gene found in plants responsible for the detection of light. These photoreceptors are used by plants to sense if they are in light or shaded environments. They also are responsible for seed germination, seedling elongation, movement of leaves and the overall size of plants (@Li_FW_2015). Phytochrome therefore is a good candidate for studying how gene duplication events resulting in gene evolution have led to the large radiations of land plants present today. The earliest known mosses date back to around 400 million years ago. (@Liu_2019). Being one of the earliest land plants has allowed for Bryophyta to rapidly diversify and support a clade composed of more than 12,700 species (@Cox_2010). The long evolutionary history of Bryophyta makes them a good candidate for modeling the molecular evolution of genes and gene duplication events. Mosses have been found to fill almost any niche imaginable and growing on all seven continents. Adaptions to temperature extremes, water availability and light levels have given this clade of plants a one up in the evolutionary race of colonizing the land. Hornworts, Liverworts, and Mosses (Bryophyta) together compose the division of plants called Bryophytes (Figure 1). Bryophytes are a clade of early evolving non-vascular plants., with mosses be distinguished from other land plants by growing most of their life as gametophytes, that when mature always have their sporophyte present (@Liu_2019).

Broad sampling of mosses was done in prior studies of phytochrome evolution in land plants. Phytochromes were found to have duplication events present in some species of moss (@Li_FW_2015). The data used from this study builds on the data collected by Fay-Wei Li and the 1KP Transcriptome Project (@Leebens-Mack_2019) and hopes to build a better understanding of the role of phytochrome gene duplication events over the evolutionary history of Bryophyta. Hornworts being the sister clade to Bryophyta, were used as the outgroup to root the phylogenetic tree created. This study aims to answer three fundamental questions; (1) how genes evolve after gene duplication, (2) what the molecular evolution of mosses shows us in relation to different gene lineages, and (3) are these different linages of genes evolving under different selective pressures. Sequences collected from both moss and hornworts were aligned in AliView. Aligned sequences were then run in the CIPRES portal through IQ-TREE to set parameters and generate a phylogenetic tree. The phylogenetic trees were then opened in DENDROSCOPE for viewing. A RELAXED, aBSREL and FEL test were then run in Datamonkey to test for selection pressure among different branches of the phylogenetic tree.

Methods

Data collected for the 1KP Transcriptome project (@Leebens-Mack_2019) focusing on the phylogeny of green plants was drawn from to acquire the phytochrome sequences of mosses and the outgroup liverworts for this study. AliView was used to code the gene sequences based on amino acid sequences and nucleotide sequences. The sequences derived from ALIVIEW were then analyzed in CIPRES for any possible errors. Within CIPRES, IQ-TREE was used to generate the phylogenetic tree with specific parameters needed. The bootstrap value was set to 1000 and the number of patterns specified was 10 for both the amino acid and the nucleotide sequences. The phylogenetic tree generated was then viewed in DENDROSCAPE, which is the application needed for opening phylogenetic tree files from CIPRES and IQ-TREE. RStudio was used to generate a smaller phylogenetic tree depicting the moss lineages in relation to liverworts and hornworts. This phylogenetic tree represents the Bryophytes (Figure 1).

Figure 1. Phylogenetic tree depicting the Bryophytes linages. Hepatophyta (Liverworts) are shown to be a sister clade to Bryophyta (Mosses) and Anthocerophyta (Hornworts). Three tests were run in Datamonkey to test for selection pressures throughout the Bryophyta lineage. RELAX analysis was run using the phylogenetic tree generated from IQ-TREE. RELAX analysis were used to identify if selection was relaxed or intensified on specific lineages (@Spielman_2017). Lineages were chosen for analysis labeled as either the Reference or Test lineages. aBSREL analysis were run to find possible evidence of lineage selection (@Spielman_2017). aBSREL was used to identify any branches or segments of branches that have been subjected to positive selection (@Spielman_2017). FEL analysis was conducted, which tests for the presence of selection among different sites using the likelihood ratio test. (@Spielman_2017). This analysis estimates synonymous and nonsynonymous substitutions in the nucleotides of the mosses. Omega is a parameter used for this analysis to test for selection pressures acting on branches of the phylogenetic tree. Omega is an estimation of the nonsynonymous to synonymous substitution rates. It represents a ratio of the two (@Wertheim_2014). Omega is used to identify any deviations from the neutral omega value of omega = 1 two (@Wertheim_2014). Deviations from neutral omega explain branches / lineages subjected to positive (greater than 1) or purifying (less than 1) selection. K is a parameter that measures intensity of selection. Selection is intensified if K is grater than 1, and relaxed if K is less than 1 (@Wertheim_2014).
R programing was used to generate Table 1. The rtree function was used in R to generate the phylogenetic trees represented in Figure 1 and Figure 4.

Results

Phylogenetic Tree The Phylogenetic tree created in IQ-TREE was composed of mosses representing all Bryophyta genera (Figure 3). The outgroup used was liverworts. Gene duplication events were found to occur for Andreaea. The duplication events were found at nodes supported by 100 and 82 percent Bootstrap values. Sphagnum and Takakia were not present at the first duplication event but were present along with Andrearea at a second gene duplication event. Figure 4 represents a phylogenetic tree showing that Sphagnum and Takakia, must have split first before Andreaea, for a gene duplication even to occur only for Andreaea.

Figure 4. Phylogenetic tree generated in R. Phylogeny depicting phytochrome gene alignments used in this study as well as Andreaea, Sphagnum and Takakia moss. Andreaea must have had a gene duplication event occur in between it and the Sphagnum and Takakia moss lineages. This would result in Andreaea having phytochrome 1-3 and phytochrome 2-4. RELAX The RELAX(ed) selection test was used to test for relaxed or intensified selection among different branches of the phylogenetic tree. 97 nucleotide sequences were used for this analysis, with data collected from 1140 sites on the phylogenetic tree. The selection intensity parameter, K, was found to be significant for the Test branch. The Test branch was shown to be subject to relaxed selection with a K value equal to 0.67. Any K value lower than 1 is said to be under relaxed selection. P was found to equal 0 and the Likelihood ratio (LR) was found to be equal to 68.78 (@Wertheim_2014). Figure 2, shows that the Reference branches moved closer to the Test Branches, towards an omega value of zero (zero being neutral selection). Figure 2. Graph depicting Test (in turquoise) and Reference branch (in black). The vertical dotted line represents neutrality (where omega = 1). The net movement od the Test and Reference branches is towards an omega value of 1. aBSREL aBSREL analysis were conducted to test for selection being present at specific lineages or branch sites, at the nucleotide level (@Spielman_2017). 97 sequences were run in this analysis, focusing on 1140 sites. aBSREL identified episodic diversifying selection acting on three of the 190 branches of the phylogenetic tree (@Smith_2015). 93 branches of the phylogenetic tree were tested for diversifying selection. Likelihood Ratio Test (LTR) with a p value less than or equal to 0.05 was used to identify significance (@Smith_2015). The three nodes found to be subject to diversifying selection were Node 34, 87, and 80 (see Table 1). The likelihood ratio test for the three nodes showed much larger values compared to the nodes that were not found to be under episodic diversifying selection.

Table 1. Table depicts the three nodes which were found to be subjected to episodic diversifying selection. LRT (second column) lists values generated by the likelihood ratio test. P value (third column) lists values estimated from the statistical test run.

FEL The FEL analysis test was run to show if there was any evidence of selection pressure on branches of

the phylogenetic tree at the codon level. The FEL analysis uses maximum likelihood ratio test to assign a model codon to each of the branch sites. (@Kosakovsky_Pond_Frost_2005) FEL estimates the occurrence of nonsynonymous and synonymous substitutions as each site to infer if selection acting on the branch sites was diversifying, purifying, neutral, or invariable. The P-value threshold used by the FEL analysis was set to 0.05. 97 sequences were used in this analysis, with 1140 codon sites. 190 branches were used for this analysis, with zero sites being under diversifying selection. 1059 sites were found to be subject to purifying selection. Diversifying positive selection is said to any P-value less than or equal to 0.05. Purifying selection is found to any P-value less than or equal to 0.05.

Tables Table: (#tab:byhand) Table 1. Table depicts the three nodes which were found to be subjected to episodic diversifying selection. LRT (second column) lists values generated by the likelihood ratio test. P value (third column) lists values estimated from the statistical test run.

```
Node = c("Node34", "Node87", "Node80")
LRT = c("20.3470", "15.2783", "13.1131")
Pvalue = c("0.0012", "0.0150", "0.0441")

ABtest <- data.frame(Node,
                     LRT,
                     Pvalue)
ABtest
```

```
##     Node     LRT Pvalue
## 1 Node34 20.3470 0.0012
## 2 Node87 15.2783 0.0150
## 3 Node80 13.1131 0.0441
```

Figures (ref:Figure 1. Phylogenetic tree depicting the Bryophytes linages. Hepatophyta (Liverworts) are shown to be a sister clade to Bryophyta (Mosses) and Anthocerophyta (Hornworts).)

```
library(ape)
oldtips <- c("t2", "t1", "t3")
oldtree <- rtree(n = 3, tip.label = oldtips)
plot(oldtree)
```

```
new_tiplabels <- c("Anthocerophyta", "Bryophyta", "Hepatophyta")
oldtree$tip.label <- new_tiplabels
plot(oldtree)
```

(ref:Figure 4. Phylogenetic tree generated in R. Phylogeny depicting phytochrome gene alignments used in this study as well as Andreaea, Sphagnum and Takakia moss. Andreaea must have had a gene duplication event occur in between it and the Sphagnum and Takakia moss lineages. This would result in Andreaea having phytochrome 1-3 and phytochrome 2-4.)

```
library(ape)
labelstip <- c("t1", "t2", "t3", "t4", "t5", "t6", "t7")
firstt <- rtree(n = 7, tip.label = labelstip)
plot(firstt)
```

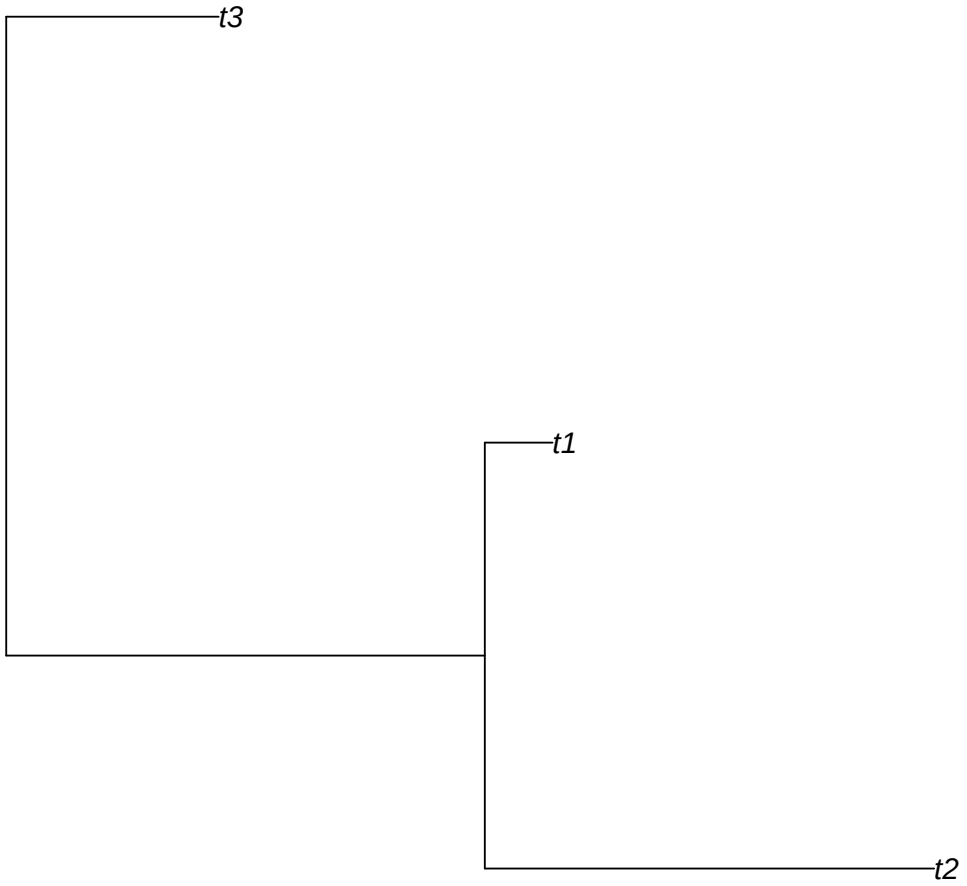

Figure 1: (ref:captionFig1)

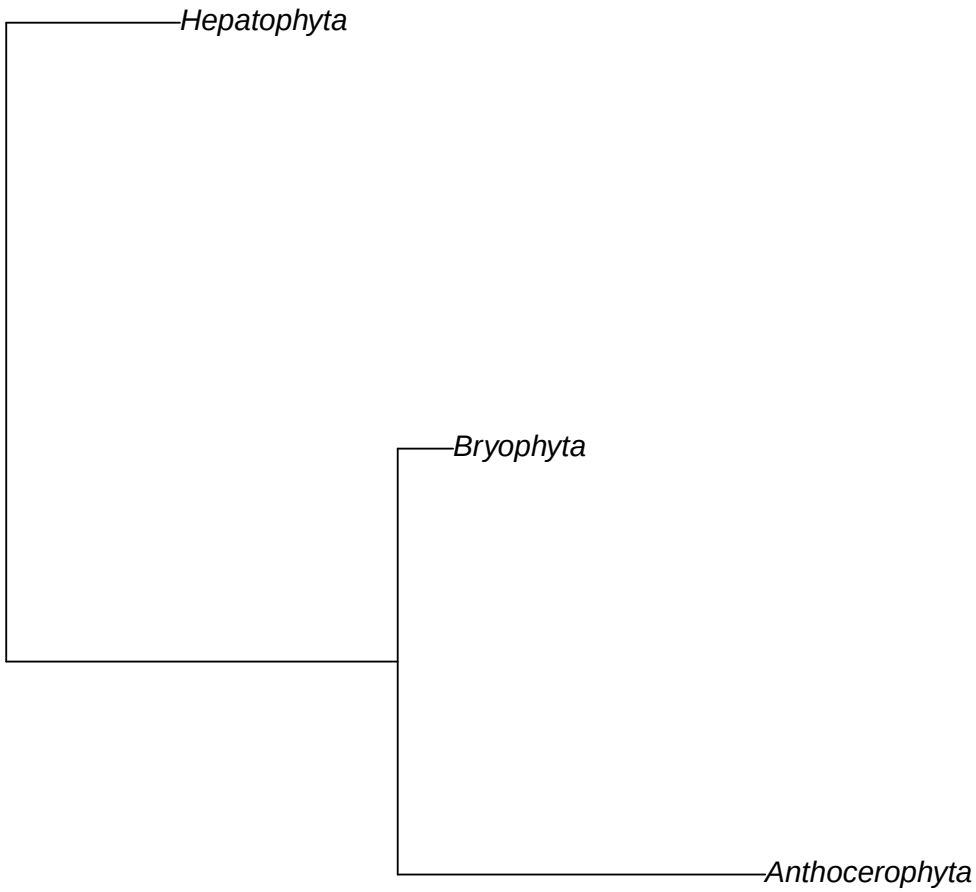

Figure 2: (ref:captionFig1)

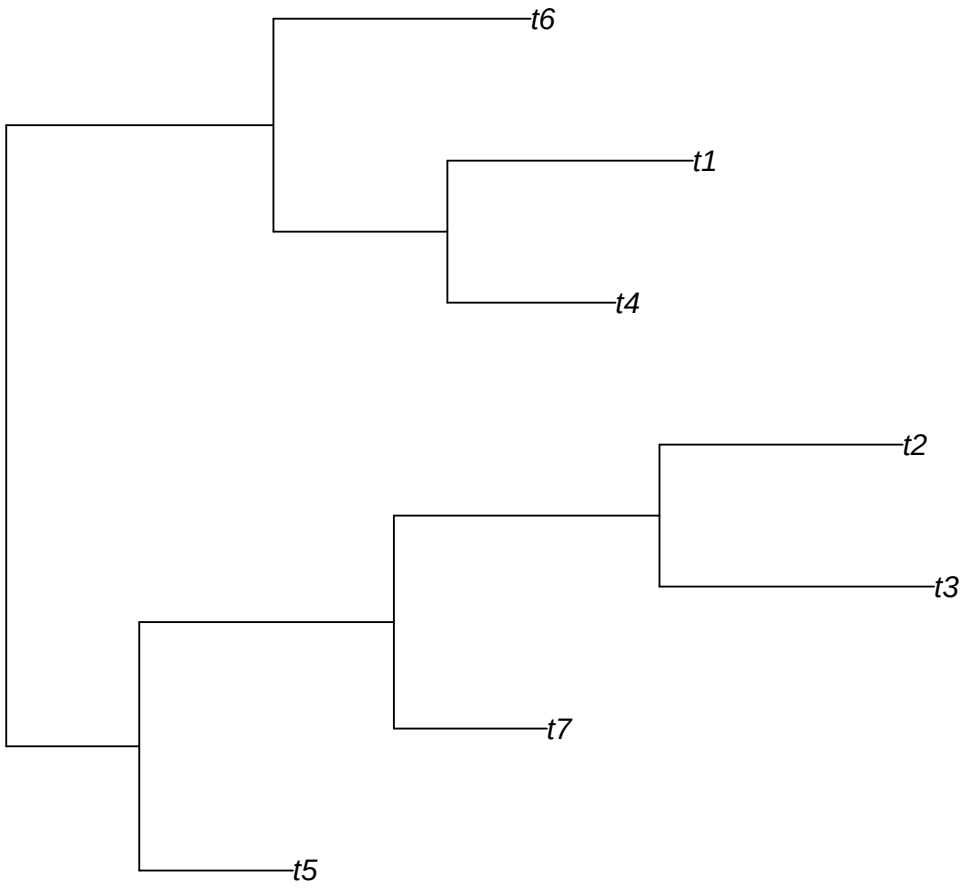

Figure 3: (ref:captionFig3)

```
brandnew <- c("Takakia", "Sphagnum", "Andreaea",  "Phytochrome 1_3", "Phytochrome 2_4", "Phytochrome 5_
firstt$tip.label <- brandnew
plot(firstt)
```

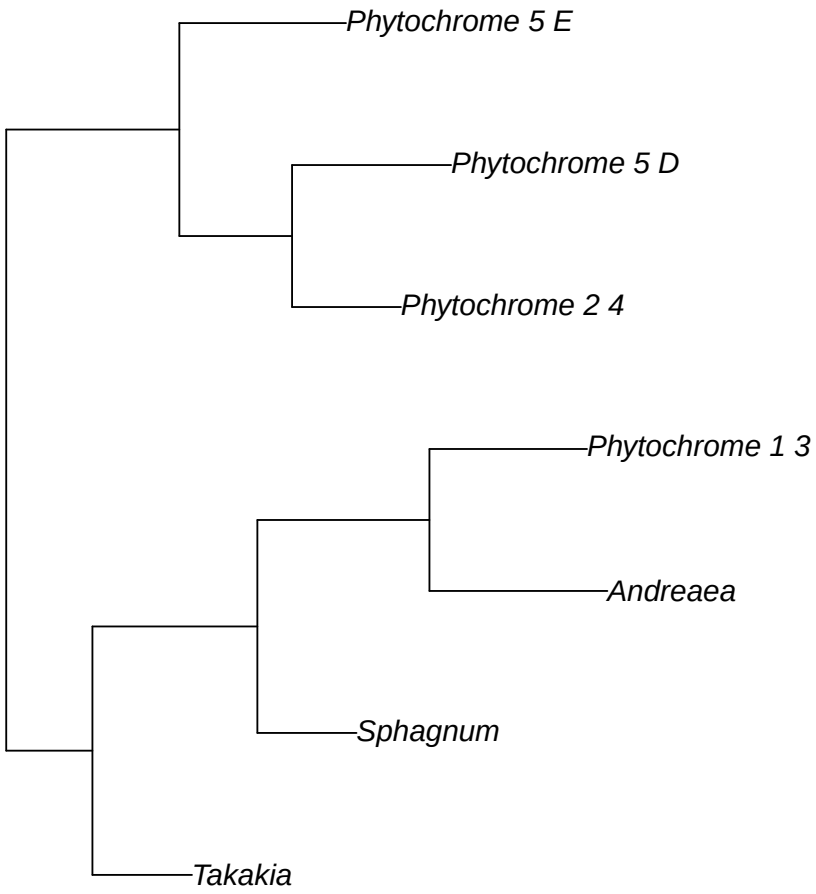

Figure 4: (ref:captionFig3)

(ref:Figure 2. Graph depicting Test (in turquoise) and Reference branch (in black). The vertical dotted line represents neutrality (where omega = 1). The net movement of the Test and Reference branches is towards an omega value of 1.)

Discussion

The Bryophyta lineage has been around for an estimated 400 million years, giving moss a very large amount of time to diverge and speciate. Mosses have been shown to diverge and adapt to many different environments and stresses, including in response to light levels (@Liu_2019). Over evolutionary time Bryophyta has had gene duplication events which have resulted in the presence of phytochrome throughout all its genera. This study shows the possibility of gene duplication events occurring before the evolution of phytochrome 1-3 in mosses, but after the split of Takakia moss and Sphagnum moss from the rest of the Bryophyta clade. This duplication event allowed for Andreaea to have been duplicated after the phytochrome 1-3 lineage. This finding needs to be further explored, as to pinpoint where exactly the gene duplication event occurred before the phytochrome 1-3 branch in the phylogenetic tree. More insight into this question could hold answers as to why Andreaea had gene duplication, but Takakia and Sphagnum did not. Through the phylogenetic tree created from IQ-TREE, Sphagnum and Takakia were shown to be sister taxa to a large clade of moss. Andreaea was shown to be a sister taxon to many of the other moss genera as well, with the addition of a gene duplication event being present. The RELAX analysis testing found branches within the phylogenetic tree to be statistically significant for relaxed selection (@Wertheim_2014). K, the parameter used to measure selection intensity was shown to be 0.67. Any K value less than or equal to 1 represents branches or lineages that were subjected to relaxed selection. The presence of relaxed selection means that genes and genomes are subjected to higher probabilities of evolving. This molecular evolution could be the driving force behind the evolution and presence of phytochrome throughout Bryophyta. This finding is important because it adds a piece to the puzzle of understanding how gene duplication events have resulted in the evolution of phytochrome throughout land plants. The aBSREL analysis was run to test specific branches and lineages across the moss phylogenetic tree for diversifying selection at the nucleotide level. (Spielman et al., 2017). The data found from this analysis showed high probability of three nodes along the phylogenetic tree being subjected to episodic diversifying selection. Nodes 34, 87, and 80 had high p values (all above 12.00) meaning that diversifying selection was occurring. Diversifying selection occurs when values that are extreme are favored for selection. This form of selection drives speciation (lineage splitting). This data shows that at the nucleotide level, episodic diversifying selection is present. This data does not represent a large amount of diversifying selection, as only three of 93 nodes were found to be under this selection pressure. Future work could be done, focusing on the taxa present after each node to find any correlations between diversifying selection and the species of moss it acts on. FEL analysis testing was run to test for selection pressure along the branches of the phylogenetic tree at the codon level. Using maximum likelihood analysis, this test found there to be 1059 sites where branches were under purifying selection. A significant and unexpected finding was that there was no diversifying selection pressure found at the codon level. This shows that while nucleotides are undergoing diversifying selection, the nucleotides evolution has not been great enough yet to change (for evolution to occur) at the codon level. More analysis could be run to see if moss species not represented in the data set have any disruptive selection at the codon level. Similar experiments run by Fay-Wei Li on moss phytochrome phylogeny have shown similar results with some dissimilarities. Fay-Wei Li's analysis was not similar as it found a gene duplication event to occur before the splitting of Sphagnum, Takakia and Andreaea in one of their phylogenetic trees depicting all representative clades of Viridiplantae. In a phylogenetic tree of phytochrome presence in moss, their results matched ours, with the gene duplication event occurring at the node where phytochrome 1-3 split from phytochrome 2-5. This study was limited by the availability of moss gene alignments and sequences. Though moss species from all genera were represented, it would be more beneficial to have more species diversity present in the study. In addition, future work on this topic could also include hornwort gene alignments into its analysis, thereby represent all three lineages of Bryophytes.

References

Cox, C. J., Goffinet, B., Wickett, N. J., Boles, S. B., & Shaw, & A. J. (2010). Phytotaxa, Moss diversity: A molecular phylogenetic analysis of genera. Phytotaxa, 9, 175–195. www.mapress.com/phytotaxa/
Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: A comparison of meth-

ods for detecting amino acid sites under selection. Molecular Biology and Evolution, 22(5), 1208–1222. https://doi.org/10.1093/molbev/msi105 Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M., Rensing, S. A., Soltis, D. E., Soltis, P. S., Stevenson, D. W., Ullrich, K. K., Wickett, N. J., DeGironimo, L., ... Wong, G. K. S. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. Nature, 574(7780), 679–685. https://doi.org/10.1038/s41586-019-1693-2 Liu, Y., Johnson, M. G., Cox, C. J., Medina, R., Devos, N., Vanderpoorten, A., Hedenäs, L., Bell, N. E., Shevock, J. R., Aguero, B., Quandt, D., Wickett, N. J., Shaw, A. J., & Goffinet, B. (2019). Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-09454-w Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular Biology and Evolution, 32(5), 1342–1353. https://doi.org/10.1093/molbev/msv022 Spielman, S. J., Weaver, S., Shank, S. D., Rife Magalis, B., Li, M., & Kosakovsky Pond, S. L. (n.d.). Evolution of viral genomes: Interplay between selection, recombination and other forces. Wertheim, J. O., Murrell, B., Smith, M. D., Pond, S. L. K., & Scheffler, K. (2015). RELAX: Detecting relaxed selection in a phylogenetic framework. Molecular Biology and Evolution, 32(3), 820–832. https://doi.org/10.1093/molbev/msu400