Land plant molecular phylogenetics: a review with comments on evaluating incongruence among phylogenies

Cymon J. Cox – Centro de Ciências do Mar, Universidade do Algarve, Faro 8005-139, Portugal.

Abstract

Land plants evolved from freshwater charophyte algal ancestors during a single transition to the terrestrial environment. The six major lineages of land plants are divided into two groups, the bryophytes (liverworts, mosses, and hornworts) and the tracheophytes (lycophytes, ferns, and seed plants), but while the tracheophytes are thought to be monophyletic, the bryophytes have typically been considered as the direct ancestors of tracheophytes and therefore an artificial, non-monophyletic, group. Here the molecular phylogenetic evidence for relationships is reviewed and evaluated especially in-light of large genome-level studies that have been completed in the last few years. Consideration is given to how to evaluate competing hypotheses with respect to the underlying evolutionary assumptions of the models used to analyse the molecular data, and the degrees of support for particular hypotheses. It is concluded that currently the two most-favourable hypotheses are that the bryophytes are a monophyletic group, or that a lineage consisting of liverworts and mosses branched first among land plants with the hornworts the most-closely related lineage to tracheophytes. Although hitherto rarely considered, the possible monophyly of bryophytes has important implications for the morphological reconstruction of the last common ancestor of all land plants. Indeed, it might suggest that the ancestor of land plants was vascularised and had alternating generations that were more isomorphic than is found in extant taxa. The evolution of the bryophytes might then have proceeded through elaboration of the gametophyte and reduction of the sporophyte, while the opposite being true of the tracheophyte lineage.

Introduction

The origin and diversification of terrestrial plants is one of the most important events in Earth history, but it has also proven to be one of the most challenging for phylogenetic systematics to reconstruct. Similarities between reproductive features of land plants, such as the possession of motile male gametes, and comparable cellular structures and pigmentation between land plants and algae, led as early as the late 19th century to the widely-held belief that land plants were likely derived from a single green algal ancestor (Kenrick and Crane 1997). Similarly, the tracheophytes, namely the seed plants, ferns, and lycophytes, a group with a conspicuous and dominant sporophyte generation that has a vascular system to transport water and metabolites through the plant body, were also recognised very early as a natural evolutionary group (figure by Haeckel 1876 depicted in Kenrick and Crane 1997). In today's terms, such a group is described as monophyletic and consists of all the descendants of a common ancestor. However, the monophly of a group comprising the other three major lineages of land plants, namely the gametophyte-dominant bryophytes comprising the hornworts, mosses, and liverworts, has never been accorded much credibility. From a historical perspective this may be a reflection of the observation that bryophytes share with freshwater algae a similar haploid-dominant alternation of generations, which enforces the view that the "lower" bryophytes are more primitive than the "higher", and implicitly more evolutionarily advanced, tracheophytes (Bower 1890). Although such an evolutionarily simplistic view of constantly increasing complexity is now widely discredited – besides which bryophytes are themselves complex and not in any way simplistic despite their small stature – the bryophytes are still most often regarded as the direct ancestors of tracheophytes, although the exact timings of the bryophyte diversification events is a matter of great contention.

The streptophyte green plants as a group include the terrestrial plants that develop via an embryo (embryophytes) and the freshwater algae (charophytes) from which they are descended (Palmer *et al*. 2004). Indeed, recognition that the embryophytes evolved from charophyte algae

settled a long-standing debate concerning the evolution of alternating generations in land plant life-cycles (for a recent review see Haig 2008). Charophyte algae have haplontic life-cycles with zygotic meiosis, suggesting that the diploid multicellular sporophyte generation of the embryophytes had evolved from interpolation of mitotic divisions after zygote formation: an observation that accords with the "antithetic theory" and contrasts with the "homologous theory" whereby the embryophyte ancestor had isomorphic, free-living, generations (Bennici 2008). However, recent molecular phylogenetic analyses (Turmel *et al*. 2006; Karol *et al*. 2010; Wodniak *et al*. 2011, Timme *et al*. 2012; Civáň *et al*. 2014, Wickett *et al*. 2014) have determined that the aflagellate, conjugating charophytes (Zygnematales) that are most often found as single cells, cellular colonies, or filaments, are the most-closely related group to embryophytes. This result is perhaps surprising given that the more-distantly related stonewort charophytes (Charales) are morphologically more complex, with a multicellular branching structure reminiscent of land plants and a suite of cellular characters similar to embryophytes (Karol *et al*. 2001; McCourt *et al.* 2004). This finding probably reflects a loss of many of these characters in Zygnematales as a consequence of their evolving conjugation as a means of reproduction (Timme *et al*. 2012).

It is now well-established that there was a single evolutionary transition to embryophyte land plants from an aquatic charophyte algal ancestor (e.g. Karol et al. 2001), however, the relationships among the major groups of embryophytes have proven difficult to determine. Embryophyte diversity is divided into two major groups, the bryophytes and the tracheophytes (also called vascular plants), each of which has three major divisions which represent monophyletic lineages, namely, the Bryophyta (mosses), Marchantiophyta (liverworts), and Anthocerotophyta (hornworts), and the Lycopodiophyta (lycophytes), Moniliformopses (ferns), and Spermatophyta (seed plants), respectively (Figure 1). The tracheophytes all have a diploid-dominant life-cycle with a vascularised sporophyte and typically form a statistically well-supported monophyletic group (they share a single common ancestor that is unique to them) in phylogenies. By contrast, the three bryophyte groups have haploid-dominant life-cycles where the unbranched, determinate sporophyte remains attached to, and is nourished by, the gametophyte. Despite these similarities in life-cycle, the three bryophyte lineages have most often been considered

"independently evolving" lineages that likely do not form a monophyletic group (Crandall-Stotler 1980).

An ancestral nature of bryophytes was supported by the earliest formal phylogenetic analyses of land plants conducted by Parenti (1980), who established that bryophytes were paraphyletic (that is, a group with a common ancestor but not all the descendants of that ancestor) based on morphological characters, with mosses and liverworts grouped together, but with hornworts most closely related to the vascular plants. By contrast, in a similar morphological cladistic analysis by Mishler and Churchill (1986), the liverworts alone were resolved as the earliest-branching lineage, with mosses most closely related to the tracheophytes, and the hornworts branching intermediately. These pioneering analyses in many ways confirmed previous assessments of land plant evolution based on less formal methods and cemented the idea of the bryophyte plant form as being ancestral to the more "advanced" vascular plants of the tracheophytes. Indeed, the seemingly elegant progression of increasing complexity in plant form, through single-celled algae to mulicellular thallose algae, to bryophytes, and then tracheophytes, is still commonly believed to be correct, although the exact details have changed.

Newer hypotheses may necessitate a re-think as to the earliest divisions among land plants and what form the ancestral land plants took. The novel, if slightly controversial at the time, morphological insight of Mishler and Churchill was that the conducting tissue (leptoids and hydriods) of mosses was homologous to the vascular tissue found in the tracheophytes (Mishler and Churchill 1986), and although this homology has since been challenged (Ligrone *et al*. 2000; Renzaglia et al. 2007), developmental genetics has also identify these tissues as homologous (Xu *et al*. 2014). Indeed, analyses of morphological and developmental characters have consistently supported a non-monophyletic bryophytes from which vascular plants are derived (Mishler and Churchill 1986; Bremer *et al*. 1987; Mishler *et al*. 1994; Garbary and Renzaglia 1998; Renzaglia et al. 2000), as have analyses that have included fossil plant taxa (Kenrick and Crane 1991; Kenrick and Crane 1987) and those employing supra-sequence genomic-level characters (Kelch *et al*. 2004; Mishler and Kelch 2004; Civáň *et al*. 2014). The only exceptions to this have been the analyses of ultrastructural traits of male gametogenesis which tend to resolve a monophyletic

bryophyte clade (monophyletic group), with mosses more closely related to liverworts than hornworts (Garbary et al. 1993; Renzaglia et al. 2000).

Molecular phylogenetics

The relative value of using morphological or molecular characters for the reconstruction of phylogeny have been debated vigorously, with little resolution or conclusion (Hillis 1987; Baker *et al*. 1998). While each data type has merits and demerits, it is undeniable that in a time where genomic DNA sequence data are so easily obtained, the focus of phylogenetic tree reconstruction has shifted almost wholly to the molecular realm. Nevertheless, ultimately a molecular phylogenetic tree in itself is of little value except for its ability to classify and predict change in an evolutionary framework, and its use to develop hypotheses of organismal trait evolution.

As with morphological analyses, the majority of molecular phylogenetic analyses of land plants have supported the non-monophyly of the bryophytes, although molecular-based trees have tended to place the liverworts alone as the earliest-diverging taxon (e.g. Lewis *et al.* 1997), a hypothesis that has come to be the most widely accepted. Despite this trend, almost all possible combinations of relationships among the three bryophyte groups and tracheophytes have been published in molecular phylogenetic trees (Table 1), including ones where the bryophytes are a monophyletic group (e.g. Nishiyama *et al*. 2004). The difficulty in determining the relationships among the six embryophyte groups is likely due in part to their rapid lineage diversification, perhaps as a response to niche adaptation, soon after terrestrial colonisation, leaving little time for accumulation of distinguishing characters (synapomorphies) to arise in early-diverging lineages (Karol *et al*. 2001; Liu *et al*. 2014). Such phylogenetic uncertainty is especially problematic as evolutionary analyses of the key genetic and morphological innovations underpinning land colonisation are important to our understanding the ecology of environments and early biotically-driven terrestrial environmental changes.

Despite the ambiguity among phylogenetic analyses of land plants, it is commonly assumed that bryophytes are paraphyletic and the liverworts are the earliest-diverging taxon (e.g. Bowman *et al*. 2007; Clarke *et al*. 2011). However, with the increased availability of data due to advances in

sequencing technology, as well as advances in phylogenetic modeling and gene tree evolutionary theory, bryophyte relationships have again come under close scrutiny and recent analyses have yet again challenged the widely-held belief of the ancestral nature of bryophytes in land plant history. How then to make sense of conflicting data sets, and even conflicting analyses of the same data? The arguments are often complex and nuanced, as no one aspect of evolutionary analyses is likely to be sufficient and alone necessary to guarantee that the result converges on the true phylogeny. Here I outline some considerations when assessing modern phylogenetic techniques and the results of those analyses.

Evaluating competing phylogenetic reconstructions

It is tempting to view phylogenetic reconstruction as fraught with difficulty and controversy, given the high-profile debates that exist on the origin of eukaryotes, animals, flowering plants, and of course many other relationships in the tree of Life. But such a view tends to overlook the remarkable advances that have been made: after all, if the phylogenetic evidence is consistent and points toward a certain relationship, it is eventually accepted as theory no matter how previously controversial such a hypothesis might have been. Indeed, the vast majority of phylogenetic relationships appear non-controversial even in land plant evolution: a) land plants are monophyletic and derived from charophyte algae, b) each of the six major land plant groups (mosses, liverworts, hornworts, lycophytes, ferns, and seed plants) are monophyletic, and c) vascular plants are monophyletic, etc. Nevertheless, despite much research, the relationships among land plant groups had proven difficult to reconstruct with any consistency or confidence. Indeed, the origin of land plants has become a pivotal question in phylogenetic systematics, prompting the use of new models and theory (Cox *et al*., 2014; Wickett *et al*., 2014) and helping drive the adoption of new methods (Sayyari and Mirarab, 2016).

How then to navigate a sea of competing, and often well supported, analyses? The concept of support for a phylogenetic hypothesis is important but should be seen only in the context of the particular data set being analysed and the assumptions implicit in the model. Here bootstrap support of >80%, or Bayesian posterior probabilities >0.95, are interpreted as being strong and

statistically significant support. The apparent, though sometimes weak, conflict between ML bootstrap and Bayesian analyses under similar models can be attributed to their different statistical frameworks: while bootstrap analyses measure the long-term probability of obtaining a similar result given additional data of the same kind, Bayesian analyses directly assess the probability of the model given the data (Shoemaker *et al*. 1999). When different methods and different data converge on a single species tree, we can have good confidence in having reconstructed the real phylogenetic history of the species. More specifically, analyses conducted using different genes and genomes (organellar and nuclear data) and different analytical methods (distance methods, maximum parsimony, likelihood and Bayesian methods) would ideally converge to the same, correct, species tree. However, some relationships are especially difficult to reconstruct, typically because lineage diversification occurred in a relatively short time-span, leaving few changes in support of relationships. In addition, gene (or contiguous non-recombining sequence) trees that conflict with the species tree due to population level processes (e.g. incomplete lineage sorting, hybridisation, introgression) can result in well-supported but incorrect phylogenies (Degnan and Rosenberg 2006 and 2009). Indeed, when faced with competing phylogenetic hypotheses, it may seem tempting to accept the result that is found most frequently as being correct, and ignore contradictory results as mere aberrations. But this type of consensus summarising of multiple analyses is often flawed: when a particular relationship is difficult to reconstruct, it can be easier to find an incorrect tree than it is the true species phylogeny. Indeed, in my view multiple congruent analyses do not out-weigh the importance of one "better" analysis that results in a different tree. The problem is determining which are the most appropriate analyses, and which are inadequate, in the context of some given data.

Until relatively recently, it has been assumed (often due to computational expediency) that the correct tree is most often than not an emergent property of concatenating large, so-called phylogenomic, amounts of data that negate statistical sampling (stochastic) errors (e.g. Rokas *et al*. 2003; Jeffroy *et al*. 2006). Indeed, older phylogenetic divergences that occur rapidly in time, that is, deep and short internal branches of the phylogeny, are often very problematic to reconstruct because they retain so few defining substitutions in extant descendants (Whitfield and Lockhart 2007). In such cases these relationships are often accurately resolved by employing larger

amounts of data. However, the expectation of a positive correlation between increased phylogenetic accuracy and larger data sets is not always met, as longer sequences can also amplify non-historical signals (biases) in the data that can result in systematic errors in tree reconstruction if inadequately accounted for by the substitution model (e.g. Phillips *et al*. 2004; Rodríguez-Ezpeleta *et al*. 2007). In the context of maximum parsimony (which implements an implicit model of change – Swofford *et al*. 1996), these biases lead to inaccurate phylogenies due to homoplasy (convergent similarity) being undetected, and erroneously reconstructed as ancestrally shared characters (synapomorphies) supporting the close relationship of taxa that are in reality more distantly related. For instance, biases in the rate of substitution among lineages can lead to the accumulation of sufficient amounts of homoplasy in fast evolving lineages to reach substitutional "saturation", which overwhelms the historical signal and causes the lineages to be joined by convergent similarity – the often cited phenomenon of "long-branch attraction" (Felsenstein 1978). Although in this case including additional taxa can increase phylogenetic accuracy when they segment long branches in the tree (Baurain *et al*. 2007; Hedtke *et al*. 2006), the additional taxa will not be beneficial if biases remain that are not adequately accounted for by the model (Poe and Swofford 1999; Heath *et al*. 2008). Better sampled trees are therefore not necessarily more accurate than sparsely sampled trees when biases exist, and in fact, additional taxa may introduce, or strengthen, biases not present in the original data (Kim, 1996; Rokas and Carroll 2005; Gatesy *et al*. 2007). Moreover, adding more data may even be detrimental if its inclusion precludes the use of more realistic models due to increased computational burden. Importantly, with the advent of next-generation sequencing, data is often no longer the limiting-factor and the emphasis in phylogenetics has shifted to constructing better-fitting and more realistic models (Baurain *et al*. 2007). But this means that although there may be sufficient data available for a specified phylogenetic question, there is still a computational balance to be struck between numbers of taxa, the amount of data, and model complexity that often leads to conflicting results and controversy surrounding difficult-to-determine relationships.

To obtain an accurate phylogenetic result, the evolutionary processes that require modeling can occur at the level of the gene (or smaller non-recombining sequence unit), the locus, and the

phylogenetic tree. At the level of the sequence data, variation in instantaneous substitution rates among nucleotides (or amino-acids) and their frequency (base composition) - which together specify the substitution rates - can vary across the data and across the tree (i.e. site- and time-heterogeneous processes). Variation in the absolute rate of substitution at sites in the data is commonly modeled by a mixture of rates drawn from a discrete gamma-distribution (Yang 1994), while variation across the branches of the tree is reflected in differing branch lengths. Among sites, variation among types of substitution processes is typically modeled by applying different substitution matrices to different data partitions, be they genes or some other division of the data (e.g. codon position), or by modeling variation in just the base (or amino acid) composition using a non-parametric (where the number of parameters of the model is *a priori* undetermined) model (i.e. CAT model: Lartillot and Philippe 2004). By contrast, substitution variation across the tree can be modeled by having instantaneous substitution rates (i.e. NDRH model: Foster *et al*. 2009) or the base composition (i.e. NDCH/NDCH2 model: Foster 2004; nhPhyML: Boussau and Gouy 2006) vary among branches. The older the phylogeny, the more likely it is to have evolved lineage-specific substitution processes, and hence these tree-heterogeneous substitution models are often a better fit to empirical data above the species-level than homogeneous substitution models.

In addition to modeling sequence substitutions, there has in recent years been an increased effort to accommodate processes acting above the level of the sequence in the phylogenetic model hierarchy that can also lead to systematic error (Lartillot 2015). Among loci, population-level variation in the number, fixation, and sorting of alleles can lead to incongruence among species trees and gene (or more specifically locus) trees due to retention and differential sorting of alleles through multiple speciation events, a process referred to as incomplete lineage sorting (ILS). Reconciliation of gene trees and species trees under the assumption of ILS can be achieved by modeling the probability of gene coalescences that are proportional to the effective population size of the branch and its length: large ancestral population sizes and shallow divergences are expected to result in greater species tree-locus tree discordance (Degnan and Rosenberg 2006 and 2009). Although fully hierarchical multi-species coalescent models that estimate both the species tree and gene trees simultaneously are available (e.g. *BEAST: Heled and Drummond 2010), their use is severely limited by their computational complexity, and more often than not, so-

called "short-cut" methods that start with pre-computed gene trees are used as a convenience.

Several implementations of short-cut multi-species coalescent models are available (e.g. MP-EST:

Liu *et al*. 2010; ASTRAL: Mirarab *et al*. 2014), however, despite their being computationally

tractable, simulations have shown that inaccurately computed gene trees are more problematic for

these methods than is ILS for concatenation methods (Gatesy and Springer 2014; Patel *et al*.

2013). Indeed, the statistical consistency of these methods can only be assured in the absence of

gene tree estimation error, or if error is present, then only if the tree conforms to a molecular clock

(Roch and Warnow 2015), which is very unlikely in phylogenies above the species level.

Nevertheless, such assurances are moot, as the length of non-recombining sequence data needed

to accurately estimate gene trees is sufficiently long that such sequences are typically an

unrealistic expectation of empirical data (Gatesy and Springer 2014; Roch and Warnow 2015).

Moreover, for ancient phylogenies where even short internal branches may span such

considerable time-scales that the expectation of ILS is greatly diminished due to allele fixation

during that period, consequently the use of short-cut multi-species coalescent models for

reconstructing ancient phylogenies has been strongly criticised (Gatesy and Springer 2014).

Assuming the absence of confounding deep gene coalescences, gene trees can differ from

species trees due to gene duplication, differential loss in subsequent lineages, and lateral gene

transfer (Szöllősi *et al.*2014). "Duplication-loss" (DL) models accounting for these processes are

now available but, again, they are computationally expensive in their fully hierarchical

implementations that assess both gene tree and species trees simultaneously (Szöllősi *et al*.

2014). The fundamental difference between ILS and DL models is that gene tree-species tree

discordance due to ILS occurs using completely orthologous gene sequences, while DL models

account for the same conflict among orthologous and parologous genes.

As a consequence of these varying degrees of model complexity, evaluating competing

phylogenetic hypotheses can be challenging. Putting aside the need for adequate sampling of both

taxa and data, the goodness-of-fit of a set of candidate models for a particular data set can be

assessed by computing a hierarchical likelihood ratio test (Posada and Crandall 1998) or by

calculating Bayes factors in a Bayesian statistical setting (Nylander *et al*. 2004). However, there

are currently no available models that are able to account for all known processes affecting

phylogenetic reconstruction – both data- and time-heterogeneous processes. At present, for a specific phylogenetic question we are often presented with an array of competing hypotheses based on different data and implementing a variety of different models. Thus analyses must be viewed in light of the data and model that they use, and the substitution and tree-evolutionary processes that are known to be present in the data and that are accounted for, and those that are not. For instance, if an analysis uses a site-homogeneous composition model and yet the data have not been shown to be composition homogeneous, then the extent to which the reconstruction is affected by potentially heterogeneous compositions is unknown, and weakens the credibility of the analysis. Likewise, concatenation methods that assume orthology among sequences ignore, due to their presumed absence, the tree-evolutionary processes of gene duplication and loss, and lateral gene transfer (and perhaps ILS, depending on the tree depth) and would naturally be less credible than models that accounted for these processes if the sequences were not orthologous. We might not reject a concatenated gene analysis that does not demonstrate that the genes are congruent, but we should be wary of assuming congruence if the analyses do not demonstrate as much, and our confidence in the analysis should be duly adjusted. In practice then, to gauge the value of a phylogenetic hypothesis we need to be able to evaluate the extent to which the model fits the data at hand – be it DNA, amino-acids, or some partition of such data - both with regards to known properties of that data, and to those properties which remain unassessed.

Current phylogenetic hypotheses of land plants

Table 1 is a list of formal phylogenetic studies that use explicit cladistic or molecular evolutionary models. Included are only studies with trees published in the main article that fully resolve relationships (the lineage or taxon has only one sister-group and is not placed at a polytomy) among the three bryophyte groups with respect to the tracheophytes, and where each group was monophyletic or consisted of a single taxon. Consequently, many analyses where only two bryophyte groups are included, even if their relationships are well supported, or where the three bryophyte lineages remain only partially unresolved (e.g. Malek *et al*. 1996), were not considered. However, studies are included even if relationships were not statistically supported, as high

statistical support is often not possible with a restricted number of morphological characters, and even high support can still be obtained for the wrong tree if systematic biases are present in molecular analyses. In Table 1 competing topologies are given names that typically describe the branching of bryophyte groups, using the common used terminology of "basal" groups. It should be noted that the term "basal" has no topologically or evolutionary meaningful usage and typically identifies the "smallest earliest-branching group sister to the rest of the clade" - liverworts in a "liverworts-basal" tree are no more "basal" to the rest of the land plants, than are the rest of the land plants "basal" to liverworts – it is just conventional usage, and it is used here for convenience.

Table 1 contains 44 unique studies, some which presented more than one main hypothesis. Perhaps the most striking aspect of the table is the number of different hypotheses generated. For five taxa - in this case, an algal outgroup, the three bryophytes groups, and the tracheophyte lineage - there are 15 possible unrooted topologies: of those a remarkable nine different hypotheses of land plant relationships have been published in studies. The first molecular study, based on 5S rDNA sequences (Hori *et al*. 1985), was published five years after the first morphology study (Parenti 1980). During the next decade and a half, an additional four studies of general morphology (Bremer 1985; Mishler and Churchill 1986; Bremer *et al*. 1987; Kenrick and Crane 1987), and four studies of male sperm ultrastructure (Garbary *et al*. 1993; Mishler *et al*. 1994; Garbary and Renzaglia 1998; Renzaglia *et al*. 2000) were published. Since the year 2000 no further morphological studies have been undertaken, although structural features of the plastid genome have been used (Kelch *et al*. 2004; see also Mishler and Kelch 2009; Civáň *et al*. 2014). Morphological characters support either a "liverworts-basal" topology (Parenti 1980; Bremer 1985; Mishler and Churchill 1986; Mishler *et al*. 1994) or a "hornworts-basal" topology (Bremer et al. 1987; Mishler *et al*. 1994; Renzaglia *et al*. 2000). By contrast, studies of sperm ultrastructural characters almost always resolve "monophyletic bryophytes" (Garbary *et al*. 1993; Mishler *et al*. 1994; Garbary and Renzaglia 1998; Renzaglia *et al*. 2000), with a single exception in a study where "hornworts-basal" was resolved (Garbary and Renzaglia 1998). The analysis of plastid structural characters resolved a "liverworts-basal" topology (Kelch *et al*. 2004).

Molecular studies of land plants are far more numerous that those of morphology; 44 as compared to 10 studies. Inevitably it is difficult to summarise an already selective set of studies and find commonalities when so many variables influence phylogenetic reconstruction, e.g. taxon sampling, sequence data characteristics, model complexity, and methods. However some general trends are apparent even if exceptions can be found to all: 1) plastid amino-acids tend to support a monophyletic bryophytes (Nishiyama *et al*. 2004; Rodríguez-Ezpeleta *et al*. 2006; Lemieux *et al*. 2007; Cox *et al*. 2014; Civáň *et al*. 2014; Wickett *et al*. 2014; Puttick *et al*. 2018; Gitzendanner *et al.* 2018), 2) plastid nucleotides tend to support paraphyletic bryophytes, especially "liverwort, moss, hornwort" (Lewis *et al*. 1997; Delwiche *et al*. 2002; Nishiyama *et al*. 2004; Wolf *et al*. 2005; Smith *et al*. 2009; Gao *et al*. 2010; Chang and Graham 2011; Fiz-Palacios *et al*. 2011; Magallón *et al*. 2013; Kim *et al*. 2014; Ruhfel *et al*. 2014; Zhong *et al*. 2014), or 3) plastid nucleotides support "moss plus liverwort-basal" with hornworts sister to tracheophytes (Nishiyama *et al*. 2004; Turmel *et al*. 2006; Lemieux *et al*. 2007; Qiu *et al*. 2006; Gao *et al*. 2010; Karol *et al*. 2010), and 4) the results of nuclear rDNA studies depend on the molecule, with 5S rDNA supporting "monophyletic bryophytes" embedded in a paraphyletic tracheophytes (Hori *et al*. 1985; Steele *et al*. 1991; Quandt *et al*. 2004), and 18S (and or 26S) supporting "hornworts-basal, moss plus liverworts" (Hedderson *et al*. 1996, 1998) or "liverworts-basal, moss plus hornwort" (Waters *et al*. 1992; Mishler *et al*. 1994). Few analyses of nuclear and mitochondrial data sets have been conducted, but neither data type gives entirely consistent results (e.g. Duff and Nickrent 1999; Wickett *et al*. 1014; Liu *et al*. 2014). Combining various data sources has also led to inconsistent results, for instance analyses of nuclear, mitochondrial, and plastid nucleotides and nuclear rDNA (Nickrent *et al*. 2000) resulted in "hornworts-basal, moss plus liverworts" while combined plastid and mitochondrial nucleotides, and nuclear, plastid and mitochondrial rDNA resulted in "liverworts, moss, hornwort" (Qiu *et al*. 2006).

Typically, more recent studies have contained better taxon selections, larger quantities of sequence data, and employed more complex models or methods, but not always. Since 2010, ten plastid (49 loci - Karol *et al*. 2010, reanalysed by Cox *et al*. 2014; 79 loci - Gao *et al*. 2010; 83 loci - Civáň *et al*. 2014; 78 loci - Ruhfel *et al*. 2014; 72 loci - Zhong *et al*. 2014; 35 loci - Kim *et al*. 2014; 79 and 88 loci - Lemieux *et al*. 2014, 2016; 78 loci - Gitzendanner *et al.* 2018), two mitochondrial

(40 loci - Turmel *et al*. 2013; 41 loci - Liu *et al.* 2014), and two nuclear studies (674 loci - Wickett *et al*. 2014, reanalysed by Puttick *et al*. 2018) have been published with data sets containing more than 20 loci (which here are arbitrarily considered to be "genome-level" studies). Five studies of plastid nucleotide data have resulted in topologies that favour either "liverwort, moss, hornwort" and "liverwort plus moss-basal" (Gao *et al*. 2010; Karol *et al*. 2010; Ruhfel *et al*. 2014; Zhong *et al*. 2014; Kim *et al*. 2014). However, Cox *et al*. (2014) demonstrated that the support for non-monophyletic bryophytes in the study of Karol *et al*. (2010) (and in the study of plastid nucleotides in Qiu *et al*. 2006) was a result of a systematic bias in composition among lineages due to synonymous substitutions, and moreover, the authors suggested that this bias is potentially a common problem in any analysis of protein-coding nucleotides where synonymous substitutions occur at rates orders of magnitude greater than those of non-synonymous substitutions. Consequently, the authors argued that the analyses of amino acids that resulted in "monophyletic bryophytes" should be preferred (see also Civáň et al. 2014). Not all genomic-level analyses of plastid proteins, however, support "monophyletic bryophytes"; two analyses resolve "liverwort plus moss-basal" (Ruhfel *et al*. 2014; Lemieux *et al*. 2016), and one "moss, liverwort, hornwort" (Lemieux *et al*. 2014). Strong statistical support for non-monophyly of bryophytes was lacking (i.e. 54%) in the study of Ruhfel *et al*. (2014), but in the two other studies strong support was obtained for the sister-group relationship of hornworts and tracheophytes (89% bootstrap support under a GTR+Γ_4 model and 1.0 posterior probability under a CATGTR+Γ_4 (there was no support for mosses-basal) - Lemieux *et al*. 2014; 1.0 posterior probability under a CATGTR+Γ_4 model - Lemieux et al. 2016). By contrast, analyses of similar data resulted in "monophyletic bryophytes" (1.0 posterior probability, under both gcpREV+Γ_4 and gcpREV+I+Γ_4+NDCH - Cox *et al*. (2014) (gcpREV is an empirical matrix specifically for green plants; Cox and Foster 2013)), and CAT+gcpREV+Γ_4 (0.89 posterior probability - Civáň *et al*. 2014). Besides differences in taxon complements, the main difference between these analyses is the use of an empirical substitution matrix (gcpREV) versus estimation of the rates directly from the data (GTR) – in theory, the latter fits the data better as the substitution rates are estimated directly from the data at hand. So although there is good reason to believe the support from plastid nucleotide data for a "hornworts plus tracheophytes" clade is due solely to synonymous substitutions that accumulate composition

biases leading to systematic error, one of the current best-fitting amino-acid model (CAT) results in support for hornworts sister to tracheophytes over a "monophyletic bryophytes" tree. What remains to be tested is whether modeling among-site composition bias (i.e. CAT model) and modeling among-lineage composition bias (i.e. NDCH model) simultaneously reconstructs the bryophytes as paraphyletic or monophyletic.

Despite this impasse, the plastid analyses of Cox *et al*. (2014) and Lemieux *et al*. (2014 and 2016) contained only limited numbers of bryophytes - five, three, and four respectively, due to computational limitations. Additional taxa could help to reduce branch lengths and aid resolution among the groups. Indeed, it has been argued that analyses with large taxon complements are a indispensable test for accurate phylogenetic reconstruction when using sparsely sampled but data-dense analyses from genome-level studies (Qiu *et al*. 2006). However, such a proposition would only be useful if the results of studies with large taxon complements were the final arbiter of conflicting phylogenetic hypotheses: an argument I would strongly reject. For instance, the phylogenetic analyses of Qiu *et al*. (2006) were explicitly aimed at resolving land plant relationships, and in one data set contained 193 taxa for five loci (two plastid protein-coding genes, two plastid rRNA genes, one mitochondrial rRNA gene, and one nuclear rRNA gene). The resulting topology strongly supported the "liverworts-basal" hypothesis (Fig. 1, Qiu *et al*. 2006). However, the same result can be obtained, albeit with slightly lower support values, from the same data with similar methods but using only 27 taxa (14 land plants; six bryophytes, eight tracheophytes) when taxa are chosen to sample both lineages from the first divergence in each of the three bryophyte lineages (Fig S.21 Cox *et al*. 2014). Consequently it is not large taxon numbers *per se* that are necessary but "judicious" sampling that is relevant to the question, and especially sampling either side of the first divergence of study clades, to reduce branches between major clades to a minimum. Indeed, if your question is land plant relationships, then including large numbers of taxa in the major clades may not be the optimal sampling strategy if model complexity is necessarily limited by computing times and superfluous taxa introduce additional biases.

While nuclear ribosomal gene data has a long tradition of usage in land plant phylogenetics (Hori et al. 1985; Steele et al. 1991; Waters et al. 1992; Hedderson et al. 1996 and 1998; Nishiyama and Kato 1999; Nickrent et al. 2000), until recently the use of nuclear protein-coding

genes has been limited to a single study of a Class III HD–Zip gene (Floyd et al. 2006) which support the as-then novel "moss, hornwort, liverwort" phylogeny. However, nuclear protein-coding genes came to the fore when Wickett et al. (2014) published analyses of 852 nuclear genes for 103 land plants: a pioneering study far exceeding the numbers of genes available from plastids. Nevertheless, despite the extensive data set, analyses of the data provided conflicting hypotheses. The many data treatments and analyses provided varying results and levels of support, but the authors chose to highlight two results: firstly, ML analyses of first and second codon position nucleotides, which supported "hornworts-basal", and secondly, short-cut coalescent analyses of amino-acid gene trees which supported "monophyletic bryophytes". In general, as with plastid data, analyses of nucleotides supported a non-monophyletic bryophytes whereas amino-acid data supported bryophyte monophyly, although again there were many exceptions. However, strong support for the hornworts-basal topology when using 1st and 2nd codon position nucleotides can be attributed solely to synonymous substitutions (at ARG, LEU, and SER codons) at these sites (Sousa and Cox, unpublished) as was similarly observed in plastid data (Cox et al. 2014). Reanalysis of these data by Puttick et al. (2018), using a better-fitting site-heterogeneous composition model (CAT) and branch-heterogeneous composition model (NDCH2; Foster 2004), also supported "monophyletic bryophytes", but other less well-supported analyses found some support for either "hornwort-basal" or "liverwort plus moss-basal" (Puttick et al. 2018). The strength of support for the liverwort plus moss clade in their analyses, as well as congruence with many other analyses that recover the same clade, led the authors to formally name the clade "Setaphyta": the name recognises that both mosses and liverworts have a sporophytic seta (although developmentally different) which is absent in hornworts.

Analyses of mitochondrial protein-coding genes are similarly limited to two recent studies. CAT-model analyses by Turmel et al. (2013) of the nucleotide positions (identified after removal of the fastest-evolving 20% of sites) of 40 mitochondrial gene resulted in a "liverwort, moss, hornwort" tree, however while support for hornworts sister to tracheophytes was strong, the support for liverworts-basal, as opposed to liverworts plus mosses, was considerably weaker. In analyses by Liu et al. (2014), strong support for "mosses-basal" from nucleotide analyses of 41 protein-coding genes was again shown to be due to among-lineage composition bias, a similar observation to

those of Cox et al. (2014) for plastid nucleotides. However, strong support was obtained for hornworts as the sister group to tracheophytes, although the absence of ferns, one of the major lineages of tracheophytes, and the zygnematales, the closest outgroup, from the analyses may have been detrimental to the accuracy of the results.

Conclusions

In my view there are currently two supported hypotheses of land plant relationships: firstly "monophyletic bryophytes", and secondly "liverworts plus mosses-basal", with hornworts sister to tracheophytes (Fig. 1). Both of these hypotheses assume that mosses and liverworts are sister taxa, a hypothesis which is strongly supported by most recent genome-level phylogenetic analyses and validates their recognition as a formal group, the Setaphyta. These two hypotheses are favoured because of two suppositions drawn from the arguments laid out in the preceding discussion: 1) that analyses of slower evolving amino-acids are to be preferred over those of nucleotides, which due to their rate of change accumulate larger amounts of homoplasy at the deep phylogenetic level of land plant diversification (~450-480MYA), and 2) better-fitting models, especially those that account for composition heterogeneity either among lineages or among sites, tend to support (or at least not conflict with) these hypotheses over those that favour either one of the bryophyte groups alone as the earliest-diverging lineage. For analyses of both plastid and nuclear protein-coding gene nucleotide sequences, it can be shown that conflict with amino-acid translations of the same data is a result of biases introduced by synonymous substitutions. Moreover, phylogenies of non-synonymous substitutions tend to be congruent (or not conflicting) with amino-acid trees, and indeed, analyses of both plastid and nuclear gene proteins support the monophyly of bryophytes using some composition heterogeneous models (NDCH and CAT). The important exception here is, however, the CAT+GTR plastid protein analysis, which supports "liverwort plus moss-basal". Mitochondrial protein data also strongly support a similar topology; but these data are limited by the smaller number of genes and currently are the most poorly sampled taxonomically (lacking major plant lineages). It is important to note that both organellar genomes, the plastid and mitochondrion, are thought only rarely to undergo recombination and are therefore

likely single linkage groups and representative of single gene trees, although they are not often conceptually treated as such. With this in mind, although there is conflicting evidence, especially between the mitochondrial data and the nuclear plus plastid data, the results from the mitochondrial analyses are less convincing, and therefore the monophyly of bryophytes should be considered the better supported hypothesis at this time.

Land plant phylogeny is crucial to our understanding the development of plant form and life history, as well determining the place of plants in the paleo-environment. It seems evident, given the analytically short branches separating clades in phylogenies, and consequently the difficulty in reconstructing relationships, that the four major lineages, liverworts, mosses, hornworts, and tracheophytes, evolved over a relatively short evolutionary time-span, with lineages accumulating few diagnostic changes. It seems plausible that once the charophyte algal ancestor of land plants had evolved the necessary form and physiology to survive out of water, for instance tolerance to UV and the mechanisms to maintain cell turgor pressure, a panorama of ecological niches became available for exploitation, which ultimately drove a relatively rapid diversification of new lineages. The paucity of bryophyte macro-fossils before the Silurian-Devonian boundary (~420MYA) when tracheophytes are first recorded has vexed botanists, because of the obvious implication of phylogenies that indicate tracheophytes are derived from bryophytes. Indeed this absence has been attributed to the poor preservation attributes of bryophytes, a group that do not unequivocally appear in the fossil record until much later (liverworts 393MYA; mosses 100MYA; hornworts 34MYA – Clarke et al. 2011); although some Cambrian (~510MYA) trilete spores, and "cryptospores" first occurring in the mid-Ordivician (~470MYA), are described as bryophyte-like (though they are equally tracheophyte-like as well), thereby filling the void left by the absence of evidence of bryophyte meso-fossils. Modern phylogenetics does, however, pose an alternative explanation: if bryophytes are monophyletic (or in fact even if they are paraphyletic, with liverworts and mosses-basal) they may have evolved much later than assumed and diversified alongside the early tracheophytes. Moreover, the possible monophyly of bryophytes has important implications for the morphological reconstruction of the last common ancestor of all land plants, and by extension the *gestalt* body-plan needed to recognise early stem-lineage land plants in the fossil record. For instance, if, as now seem likely, the water conducting vessels of bryophytes (leptoids

and hydriods) are developmentally homologous to xylem and phloem (Xu *et al.* 2014), and given that there is now very strong evidence mosses and liverworts are sister-groups, it seems likely that *all* embryophytes are descended from a vascularised most-recent common ancestor of land plants, unless hornworts, which lack specialised conducting vessels, are the earliest diverging lineage (a hypothesis that is not well supported). This would imply that the vascular system in mosses, and liverworts such as *Pallavicinia* and *Jensenia* (Crandall-Stotler pers. comm.; Ligrone *et al.* 2002), are likely reduced to-a-degree from that of the most-recent common land plant ancestor, and if bryophytes are monophyletic that hornworts have likely lost vascularity entirely through reduction. This in turn, raises the intriguing possibility that plant fossil assemblages such as the Rhynie Chert (~407MYA) are indicative of stem-group land plants from which both bryophytes and tracheophytes could have evolved, perhaps from taxa that had isomorphic generations and from which bryophytes evolved by elaboration of the gametophyte, and tracheophytes via elaboration of the sporophyte, generation. Of course, bryophyte monophyly does not preclude the original explanation that assemblages such as the Rhynie Chert are purely stem-tracheophytes (Kenrick 2017) and that bryophytes have an older history obscured by lack of fossil preservation or our ability to distinguish early bryophytes, but currently the best-supported phylogenetic hypotheses of land plants are not incompatible with a range of novel interpretations of the fossil record.

Table 1. Phylogenetic studies of land plant relationships. Analyses are only included if all three bryophyte groups are represented, each is monophyletic if comprising multiple taxa, and all three bryophyte groups are fully resolved among themselves but not necessarily statistically supported. (Modified from Puttick *et al.* 2018)

	Embryophyte sister-group[10]	Data	Reference
"bryophyte monophyly"[1]	Charales	Nuclear rDNA[11]	Hori *et al.* 1985
	Charales	Nuclear rDNA[11]	Steele *et al.* 1991
	Coleochaetales/Charales	Sperm ultrastructure	Garbary *et al.* 1993
	Coleochaetales	Sperm ultrastructure	Mishler *et al.* 1994
	Charales	Sperm ultrastructure	Renzaglia *et al.* 2000
	charophytes/chlorophytes	Chloroplast amino acids	Nishiyama *et al.* 2004
	Chlorophyta/Coleochaetales	Chloroplast nucleotides[11]	Quandt *et al.* 2004
	Coleochaetales	Chloroplast nucleotides[12]	Goremykin and Hellwig 2005
	Coleochaetales	Chloroplast amino acids[11]	Rodríguez-Ezpeleta *et al.* 2007
	Zygnematales	Chloroplast amino acids and nucleotides[11]	Turmel *et al.* 2006
	Zygnematales	Chloroplast amino acids	Lemieux *et al.* 2007
	Zygnematales	Chloroplast amino acids	Cox *et al.* 2014
	Zygnematales	Chloroplast amino acids	Civáň *et al.* 2014
	Zygnematales	Nuclear amino acids	Wickett *et al.* 2014
	Zygnematales	Chloroplast amino acids	Gitzendanner et al. 2018
	Zygnematales	Nuclear amino acids	Puttick *et al.* 2018
"hornworts-basal, moss plus liverworts"[2]	Coleochaetales	Nuclear rDNA	Hedderson *et al.* 1996
	Coleochatales/Klebsormidiales	Nuclear rDNA	Hedderson *et al.* 1998
	Coleochaetales	Sperm ultrastructure	Garbary and Renzaglia 1998
	chlorophytes	Mitochondrial rDNA	Duff and Nickrent 1999
	Chlorophyta/Coleochaetales	Nuclear and chloroplast nucleotides[12]	Nishiyama and Kato 1999
	Charales/Zygnematales	Nuclear, mitochondrial, chloroplast nucleotides and nuclear rDNA[12]	Nickrent *et al.* 2000
	Coleochaetales	Morphology, nuclear and mitochondrial nucleotides	Renzaglia *et al.* 2000
	Zygnematales	Nuclear nucleotides[12]	Wickett *et al.* 2014
	Zygnematales	Nuclear amino acids	Puttick *et al.* 2018
"hornwort, liverwort, moss"[3]	Charales	Morphology	Bremer *et al.* 1987
	Coleochaetales	Morphology and nuclear rDNA	Mishler *et al.* 1994
"liverwort plus moss-basal"[4]	Coleochaetales	Chloroplast nucleotides[13]	Nishiyama *et al.* 2004
	Zygnematales	Chloroplast nucleotides	Turmel *et al.* 2006
	Zygnematales	Chloroplast nucleotides	Lemieux *et al.* 2007
	Zygnematales	Chloroplast nucleotides[12]	Qiu *et al.* 2006
	Zygnematales	Chloroplast nucleotides[12]	Gao *et al.* 2010
	Zygnematales	Chloroplast nucleotides	Karol *et al.* 2010
	Zygnematales	Chloroplast amino acids	Ruhfel *et al.* 2014
	Zygnematales	Chloroplast amino acids	Lemieux *et al.* 2016
	Zygnematales	Nuclear amino acids	Puttick *et al.* 2018
"liverwort, moss, hornwort"[5]	"Charophceae"	Morphology	Parenti 1980
	Coleochaetes/Charales	Chloroplast nucleotides	Lewis *et al.* 1997
	Charales	Chloroplast nucleotides	Delwiche *et al.* 2002
	Coleochaetales	Chloroplast genome structure	Kelch *et al.* 2004
	Coleochaetales	Chloroplast nucleotides[12]	Nishiyama *et al.* 2004
	Coleochaetales	Chloroplast nucleotides	Wolf *et al.* 2005
	Charales/Coleochaetales	Mitochondrial nucleotides	Groth-Malonek *et al.* 2005
	Charales or Zygnematales	Chloroplast nucleotide and rDNA, and mitochondrial and nuclear rDNA	Qiu *et al.* 2006

	Charales	Chloroplast and mitochondrial nucleotides, and nuclear, chloroplast, and mitochondrial rDNA	Qiu *et al.* 2007
	Charales	Chloroplast nucleotides	Smith *et al.* 2009
	Zygnematales	Chloroplast nucleotides	Gao *et al.* 2010
	Zygnematales	Chloroplast nucleotides	Chang and Graham 2011
	chlorophytes/charophytes	Chloroplast nucleotides and nuclear rDNA	Fiz-Palacios *et al.* 2011
	Coleochaetales	Chloroplast nucleotides	Magallón *et al.* 2013
	Charales or Zygnematales	Mitochondrial nucleotides	Turmel *et al.* 2013
	Charales/Coleochaetales	Chloroplast nucleotides	Kim *et al.* 2014
	Charales	Mitochondrial nucleotides and amino acids	Liu *et al.* 2014
	Zygnematales	Chloroplast nucleotides	Ruhfel *et al.* 2014
	Zygnematales	Chloroplast nucleotides	Zhong *et al.* 2014
"liverwort, hornwort, moss"[6]	"charophytes"	Morphology	Mishler and Churchill 1986
	Charales/Coleochaetales	Morphology	Bremer 1985
	Coleochaetales	Morphology	Bremer *et al.* 1987
	Coleochaetales	Morphology and chloroplast and nuclear rDNA	Mishler *et al.* 1994
	Charales/Coleochaetales	Morphology	Kenrick and Crane 1997
	Charales	Chloroplast and mitochondrial nucleotides, nuclear rDNA	Karol *et al.* 2001
"liverwort-basal, moss plus hornwort"[7]	Klebsormidiales/Coeleochaetales	Nuclear rDNA	Waters *et al.* 1992
	Klebsormidiales	Chloroplast and nuclear rDNA	Mishler *et al.* 1994
	chlorophytes	Mitochondrial rDNA	Duff and Nickrent 1999
"moss, liverwort, hornwort"[8]	Charales	Mitochondrial nucleotides	Liu *et al.* 2014
	Coleochaetales	Chloroplast amino acids	Lemieux *et al.* 2014
"moss, hornwort, liverwort"[9]	Charales	Nuclear amino acids	Floyd *et al.* 2006

[1] (algal outgroup, ((liverwort, moss, hornwort), tracheophytes))

[2] (algal outgroup, (hornwort, ((liverwort, moss), tracheophytes))

[3] (algal outgroup, (hornwort, (liverwort, (moss, tracheophytes))))

[4] (algal outgroup, ((liverwort, moss), (hornwort, tracheophytes)))

[5] (algal outgroup, (liverwort, (moss, (hornwort, tracheophytes))))

[6] (algal outgroup, (liverwort (hornwort (moss, tracheophytes))))

[7] (algal outgroup, (liverwort ((moss, hornwort), tracheophytes)))

[8] (algal outgroup, (moss, (liverwort, (hornwort, tracheophytes))))

[9] (algal outgroup, (moss, (hornwort, (liverwort, tracheophytes))))

[10] either the designated outgroups or, if identified, the most closely related taxon to land plants

[11] note that although bryophytes were monophyletic they were embedded in a paraphyletic tracheophyte clade

[12] protein coding-genes 1st and 2nd codon positions only: Goremykin and Hellwig 2005; Nishiyama *et al*. 2004; Wickett *et al*. 2014; Nickrent *et al.* 2000; Gao *et al*. 2010

[13] a) excluding the nucleotide sites coding leucine and third codon positions, and b) fourfold degenerate sites

References:

Baker, R.H., Yu, X. and DeSalle, R., 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Molecular Phylogenetics and Evolution*, *9*: 427-436.

Baurain, D., Brinkmann, H. and Philippe, H., 2006. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? Molecular Biology and Evolution **24**: 6-9.

Bennici, A., 2008. Origin and early evolution of land plants: Problems and considerations. *Communicative & integrative biology* **1**: 212-218.

Boussau, B. and Gouy, M., 2006. Efficient likelihood computations with nonreversible models of evolution. *Systematic biology* **55**: 756-768.

Bower, F.O., 1890. On antithetic as distinct from homologous alternation of generations in plants. Annals of Botany **4**: 347-370.

Bowman, J.L., Floyd, S.K. and Sakakibara, K., 2007. Green genes - comparative genomics of the green branch of life. *Cell* **129**: 229-234.

Bremer, K., 1985. Summary of green plant phylogeny and classification. *Cladistics* **1**: 369-385.

Bremer, K., Humphries, C.J., Mishler, B.D. and Churchill, S.P., 1987. On cladistic relationships in green plants. *Taxon* **36**: 339-349.

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

Cháň, P., Foster, P.G., Embley, M.T., Seneca, A. and Cox, C.J., 2014. Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. *Genome biology and evolution* **6**: 897-911.

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

Cox, C.J., Li, B., Foster, P.G., Embley, T.M. and Cháň, P., 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. *Systematic Biology* **63**: 272-279.

Cox, C.J. and Foster, P.G., 2013. A 20-state empirical amino-acid substitution model for green plant chloroplasts. Molecular phylogenetics and evolution **68**: 218-220.

Crandall-Stotler, B., 1980. Morphogenetic designs and a theory of bryophyte origins and divergence. *BioScience* **30**: 580-585.

Degnan, J.H. and Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. *PLoS genetics* **2**: 68.

Degnan, J.H. and Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**: 332-340.

Delwiche, C.F., Karol, K.G., Cimino, M.T. and Sytsma, K.J., 2002. Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene *rbcL*. *Journal of Phycology* **38**: 394-403.

Duff, R.J. and Nickrent, D.L., 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* **86**: 372-386.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

Fiz-Palacios, O., Schneider, H., Heinrichs, J. and Savolainen, V., 2011. Diversification of land plants: insights from a family-level phylogenetic analysis. *BMC Evolutionary Biology* **11**: 341.

Floyd, S.K., Zalewski, C.S. and Bowman, J.L., 2006. Evolution of class III homeodomain–leucine zipper genes in streptophytes. *Genetics* **173**: 373-388.

Foster, P.G., 2004. Modeling compositional heterogeneity. *Systematic Biology* **53**: 485-495.

Foster, P.G., Cox, C.J. and Embley, T.M., 2009. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**: 2197-2207.

Gao, L., SU, Y.J. and Wang, T., 2010. Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *Journal of Systematics and Evolution* **48**: 77-93.

Garbary, D.J., Renzaglia, K.S. and Duckett, J.G., 1993. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. *Plant Systematics and Evolution* **188**: 237-269.

Garbary, D.J, and Renzaglia K.S., 1998. Bryophyte phylogeny and the evolution of land plants: evidence from development and ultrastructure. In Bryology for the Twenty-First Century, Eds. Bates, J.W., Ashton, N.W., and Duckett, J.G. British Bryological Society, UK. Pp. 45-63.

Gatesy, J., DeSalle, R. and Wahlberg, N., 2007. How many genes should a systematist sample? Conflicting insights from a phylogenomic matrix characterized by replicated incongruence. *Systematic Biology* **56**: 355-363.

Gatesy, J. and Springer, M.S., 2014. Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution* **80**: 231-266.

Gitzendanner, M.A., Soltis, P.S., Wong, G.K.-S., Ruhfel, B.R., and Soltis, D.E. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* **105**: 291-301.

Goremykin, V.V. and Hellwig, F.H., 2005. Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. *Plant Systematics and Evolution* **254**: 93-103.

Groth-Malonek, M., Pruchner, D., Grewe, F. and Knoop, V., 2004. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Molecular Biology and Evolution* **22**: 117-125.

Haig, D., 2008. Homologous versus antithetic alternation of generations and the origin of sporophytes. *The Botanical Review* **74**: 395-418.

Hedderson, T.A., Chapman, R.L. and Cox, C.J., 1998. Bryophytes and the origins and diversification of land plants: new evidence from molecules. In Bryology for the Twenty-First Century, Eds. Bates, J.W., Ashton, N.W., and Duckett, J.G. British Bryological Society, UK. Pp. 65-77.

Hedderson, T.A., Chapman, R.L. and Rootes, W.L., 1996. Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Systematics and Evolution* **200**: 213-224.

Hedtke, S.M., Townsend, T.M. and Hillis, D.M., 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* **55**: 522-529.

Heled, J. and Drummond, A.J., 2009. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* **27**: 570-580.

Hillis, D.M., 1987. Molecular versus morphological approaches to systematics. *Annual review of Ecology and Systematics* **18**: 23-42.

Hori, H., Lim, B.L. and Osawa, S., 1985. Evolution of green plants as deduced from 5S rRNA sequences. *Proceedings of the National Academy of Sciences* **82**: 820-823.

Jeffroy, O., Brinkmann, H., Delsuc, F. and Philippe, H., 2006. Phylogenomics: the beginning of incongruence? *TRENDS in Genetics* **22**: 225-231.

Karol, K.G., Arumuganathan, K., Boore, J.L., Duffy, A.M., Everett, K.D., Hall, J.D., Hansen, S.K., Kuehl, J.V., Mandoli, D.F., Mishler, B.D. and Olmstead, R.G., 2010. Complete plastome sequences of Equisetum arvense and Isoetes flaccida: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evolutionary Biology* **10**: 321.

Karol, K.G., McCourt, R.M., Cimino, M.T. and Delwiche, C.F., 2001. The closest living relatives of land plants. *Science* **294**: 2351-2353.

Kelch, D.G., Driskell, A. and Mishler, B.D., 2004. Inferring phylogeny using genomic characters: a case study using land plant plastomes. In Molecular Systematics of Bryophytes, Eds., Goffinet, B., Hollowell, V., and Magill, R., Monographs in Systematic Botany from the Missouri Botanical Garden, 98. Missouri Botanical Garden Press, U.S.A. Pp. 3-12.

Kenrick, P., 2018. Changing expressions: a hypothesis for the origin of the vascular plant life cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**: 20170149.

Kenrick, P. and Crane, P.R., 1997. The origin and early diversification of land plants

Kenrick, P. and Crane, P.R., 1997. The origin and early diversification of land plants. A cladistic study (Vol. 560). Smithsonian Institute Press Washington DC. U.S.A. Pp 1-441.

Kim, J., 1996. General inconsistency conditions for maximum parsimony: effects of branch lengths and increasing numbers of taxa. *Systematic Biology* **45**: 363-374.

Kim, H.T., Chung, M.G. and Kim, K.J., 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. *Molecules and Cells* **37**: 372.

Lartillot, N., 2015. Probabilistic models of eukaryotic evolution: time for integration. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 20140338.

Lartillot, N. and Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Molecular Biology and Evolution **21**: 1095-1109.

Lewis, L.A., Mishler, B.D. and Vilgalys, R., 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbc*L. *Molecular Phylogenetics and Evolution* **7**: 377-393.

Ligrone, R., Duckett, J.G. and Renzaglia, K.S., 2000. Conducting tissues and phyletic relationships of bryophytes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**: 795-813.

Lemieux, C., Otis, C. and Turmel, M., 2007. A clade uniting the green algae Mesostigma viride and Chlorokybus atmophyticus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology* **5**: 2.

Lemieux, C., Otis, C. and Turmel, M., 2014. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. BMC Genomics **15**: 857.

Lemieux, C., Otis, C. and Turmel, M., 2007. A clade uniting the green algae Mesostigma viride and Chlorokybus atmophyticus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biology **5**: 2.

Liu, Y., Cox, C.J., Wang, W. and Goffinet, B., 2014. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* **63**: 862-878.

Liu, L., Yu, L. and Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* **10**: 302.

Magallón, S., Hilu, K.W. and Quandt, D., 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* **100**: 556-573.

Malek, O., Lättig, K., Hiesel, R., Brennicke, A. and Knoop, V., 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. *The EMBO Journal* **15**: 1403-1411.

McCourt, R.M., Delwiche, C.F. and Karol, K.G., 2004. Charophyte algae and land plant origins. *Trends in Ecology & Evolution* **19**: 661-666.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. and Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**: i541-i548.

Mishler, B.D. and Churchill, S.P., 1984. A cladistic approach to the phylogeny of the "bryophytes". *Brittonia* **36**: 406-424.

Mishler, B.D. and Kelch, D.G., 2009. Phylogenomics and early land plant evolution. In Bryophyte Biology, Eds. Goffinet, B. and Shaw, A.J., 2nd Edition. Cambridge University Press, Cambridge, U.K. Pp.173-197.

Mishler, B.D., Lewis, L.A., Buchheim, M.A., Renzaglia, K.S., Garbary, D.J., Delwiche, C.F., Zechman, F.W., Kantz, T.S. and Chapman, R.L., 1994. Phylogenetic relationships of the" green algae" and" bryophytes". *Annals of the Missouri Botanical Garden* **81**: 451-483.

Nickrent, D.L., Parkinson, C.L., Palmer, J.D. and Duff, R.J., 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* **17**: 1885-1895.

Nishiyama, T. and Kato, M., 1999. Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Molecular Biology and Evolution* **16**: 1027-1036.

Nishiyama, T., Wolf, P.G., Kugita, M., Sinclair, R.B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K. and Ueda, K., 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Molecular Biology and Evolution* **21**: 1813-1819.

Nylander, J.A., Ronquist, F., Huelsenbeck, J.P. and Nieves-Aldrey, J., 2004. Bayesian phylogenetic analysis of combined data. *Systematic biology*,= **53**: 47-67.

Patel, S., Kimball, R.T. and Braun, E.L., 2013. Error in phylogenetic estimation for bushes in the tree of life. *Journal of Phylogenetics & Evolutionary Biology* **1**: 2.

Parenti, L.R., 1980. A phylogenetic analysis of the land plants. *Biological Journal of the Linnean Society* **13**: 225-242.

Palmer, J.D., Soltis, D.E. and Chase, M.W., 2004. The plant tree of life: an overview and some points of view. *American Journal of Botany* **91**: 1437-1445.

Phillips, M.J., Delsuc, F. and Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. *Molecular biology and evolution* **21**: 1455-1458.

Poe, S. and Swofford, D.L., 1999. Taxon sampling revisited. *Nature* **398**: 299.

Puttick, M.N., Morris, J., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D., Donoghue, P.C.J., 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology* (in press).

Posada, D. and Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J. and Estabrook, G.F., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences* **103**: 15511-15516.

Qiu, Y.L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R.W., Hendry, T.A. and Taylor, D.W., 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* **168**: 691-708.

Quandt, D., Müller, K., Stech, M., Frahm, J.P., Frey, W., Hilu, K.W. and Borsch, T., 2004. Molecular evolution of the chloroplast trnL-F region in land plants. In Molecular Systematics of Bryophytes, Eds. Goffinet, B., Hollowell, V., and Magill, B., Monographs in Systematic Botany from the Missouri Botanical Garden, 98. Missouri Botanical Garden Press, U.S.A. Pp. 13-37.

Renzaglia, K.S., Duff, R.J., Nickrent, D.L. and Garbary, D.J., 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **355**: 769-793.

Renzaglia, K.S., Schuette, S., Duff, R.J., Ligrone, R., Shaw, A.J., Mishler, B.D. and Duckett, J.G., 2007. Bryophyte phylogeny: advancing the molecular and morphological frontiers. The Bryologist **110**: 179-213.

Roch, S. and Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Systematic Biology **64**: 663-676.

Rodríguez-Ezpeleta, N., Philippe, H., Brinkmann, H., Becker, B. and Melkonian, M., 2006. Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of Mesostigma in the Streptophyta. *Molecular Biology and Evolution* **24**: 723-731.

Rokas, A., Williams, B.L., King, N. and Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**: 798.

Rokas, A. and Carroll, S.B., 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* **22**: 1337-1344.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E. and Burleigh, J.G., 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* **14**: 23.

Sayyari, E. and Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution **33**: 1654-1668.

Shoemaker, J.S., Painter, I.S. and Weir, B.S., 1999. Bayesian statistics in genetics: a guide for the uninitiated. *Trends in Genetics* **15**: 354-358.

Smith, S.A., Beaulieu, J.M. and Donoghue, M.J., 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* **9**: 37.

Steele, K.P., Holsinger, K.E., Jansen, R.K. and Taylor, D.W., 1991. Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Molecular Biology and Evolution **8**: 240-248.

Swofford, D.L., Olsen, G.J., Waddell, P.J. and Hillis, D.M., 1996. Phylogenetic inference. In Molecular systematics, Eds. Hillis, D.M., Moritz, C. and Mable, B.K. Sinauer Associates, Sunderland, Massachusetts, U.S.A.

Szöllősi, G.J., Tannier, E., Daubin, V. and Boussau, B., 2014. The inference of gene trees with species trees. *Systematic Biology* **64**: 42-62.

Timme, R.E., Bachvaroff, T.R. and Delwiche, C.F., 2012. Broad phylogenomic sampling and the sister lineage of land plants. *PLoS One* **7**: e29696.

Turmel, M., Otis, C. and Lemieux, C., 2006. The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution* **23**: 1324-1338.

Turmel, M., Otis, C. and Lemieux, C., 2013. Tracing the evolution of streptophyte algae and their mitochondrial genome. Genome Biology and Evolution **5**: 1817-1835.

Waters, D.A., Buchheim, M.A., Dewey, R.A. and Chapman, R.L., 1992. Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. *American Journal of Botany* **79**: 459-466.

Whitfield, J.B. and Lockhart, P.J., 2007. Deciphering ancient rapid radiations. *Trends in Ecology & Evolution* **22**: 258-265.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P, Graham, S.W., Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Barbara, S., Villarreal, J.C., Roure, B., Philippe, H., dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang, J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong G.K-S., Leebens-Mack, J., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences* 111: 4859-4868.

Wodniok, S., Brinkmann, H., Glöckner, G., Heidel, A.J., Philippe, H., Melkonian, M. and Becker, B., 2011. Origin of land plants: do conjugating green algae hold the key? *BMC Evolutionary Biology*, **11**: 104.

Wolf, P.G., Karol, K.G., Mandoli, D.F., Kuehl, J., Arumuganathan, K., Ellis, M.W., Mishler, B.D., Kelch, D.G., Olmstead, R.G. and Boore, J.L., 2005. The first complete chloroplast genome sequence of a lycophyte, Huperzia lucidula (Lycopodiaceae). *Gene* **350**: 117-128.

Xu, B., Ohtani, M., Yamaguchi, M., Toyooka, K., Wakazaki, M., Sato, M., Kubo, M., Nakano, Y., Sano, R., Hiwatashi, Y. and Murata, T., 2014. Contribution of NAC transcription factors to plant adaptation to land. *Science* **343**: 1505-1508.

Yang, Z., 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306-314.

Zhong, B., Xi, Z., Goremykin, V.V., Fong, R., Mclenachan, P.A., Novis, P.M., Davis, C.C. and Penny, D., 2013. Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Molecular Biology and Evolution* **31**: 177-183.
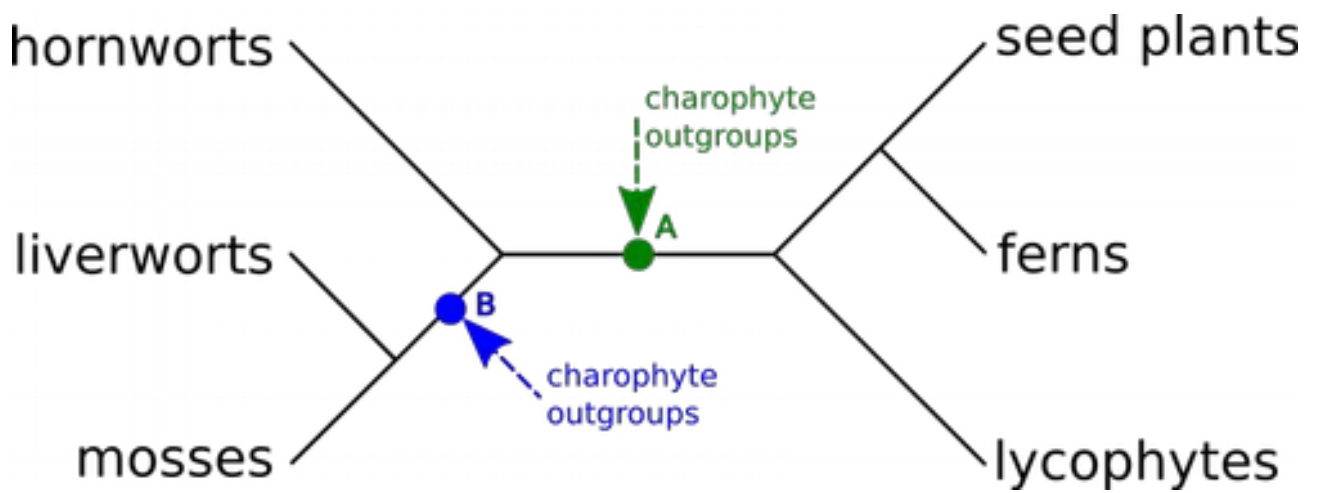
Figure 1. Unrooted phylogeny of land plants. The unrooted relationship among land plants are well supported by recent studies. Indicated on the phylogeny are the two most probable rooting positions of the charophyte algal outgroups as argued in these analyses: A) monophyletic bryophytes, B) paraphyletic bryophytes, with liverwort plus mosses-basal, and hornworts the sister-group to tracheophytes.