The application of molecular data to the phylogenetic delimitation of species in bryophytes: A note of caution

ALAIN VANDERPOORTEN¹ & A. JONATHAN SHAW²

¹University of Liège, Institute of Botany, B22 Sart Tilman, B-4000 Liège, Belgium. E-mail: a.vanderpoorten@ulg.ac.be
²Duke University, Department of Biology, Durham, NC 27708, USA. E-mail: shaw@duke.edu

Abstract

Molecular phylogenetics has been of prime importance in revisiting traditional taxonomic hypotheses, and this is especially true in taxa with reduced morphologies like bryophytes. Sequence identity at one or a few loci, as well as evidence for species para-or polyphyly, have been increasingly used to lump species. While sequence identity at loci that are usually variable within the group of interest can provide some incentive for additional study of such species, it does by no means alone provide sufficient evidence for synonymization. Similarly, the strict requirement that all species must be demonstrably monophyletic is equivalent to adopting an uncompromising view that reproductive isolation (i.e., the biological species concept) is the only valid evidence for species status, and that all species have to be 100% isolated. Some modes of speciation lead to paraphyletic species or even phylogenetic networks. We therefore encourage case by case evaluation of all available data rather than applying a single criterion such as monophyly. We make some suggestions about how to use molecular data in the circumscription of bryophyte species.

Key words: species concept; paraphyly; budding speciation; monophyly

Introduction

The species concept has long been a central issue in systematic biology. Mayden (1997) listed no less than 24 different, potentially conflicting species definitions. The reason that different criteria lead to incompatible species concepts is that various aspects of lineage divergence arise at different times during the process of speciation (De Queiroz 2007). Daughter species progressively diverge with time, but the acquisition of the different properties defining them (when they become phenotypically diagnosable, reciprocally monophyletic, reproductively incompatible, ecologically distinct, etc.) is not simultaneous. Before the acquisition of any one of those properties, everyone will agree that there is a single species, and after the acquisition of all, everyone will agree that there are two. In between, however, there will be disagreement.

In the large and taxonomically difficult moss genus Bryum Hedw. for instance, phylogenetic patterns uncovered by molecular analyses correspond poorly to traditional classifications based on morphology (Holyoak & Pedersen 2007). Arguing that molecular data provide a more accurate representation of phylogenetic history and relationships than do morphological characters, Holyoak & Pedersen (2007) concluded that any 'classification of the Bryaceae based on morphological characters alone cannot be defended'. Incongruence between inferences derived from molecules versus morphology in Bryum relates both to species delimitation and the resolution of multispecies clades. It has in fact become increasingly evident that bryophyte species, due to the limited availability of characters defining them, the focus on a few key-characters, and the influence of the environment in the evolution of those characters, render many morphologically defined species vulnerable to refutation by phylogenetic analyses (Vanderpoorten & Goffinet 2006). A large body of literature thus points to the sometimes severe incongruence between morpho-species concepts and molecular phylogenies (see Heinrichs et al. 2009a for review).
Table 1 lists examples of how taxonomic circumscriptions of bryophyte species have been modified with the use of molecular data. Molecular evidence has in some instances suggested novel species circumscriptions that subsequently served as a template for morphological re-assessments, leading to the discovery of new morphological features and descriptions of new species. The trend is, however, mostly towards a reduction in species numbers. During the period of active bryological exploration of extra-European regions during the nineteenth century, hundreds of new ‘geographical species’ were described based in large part on the assumption that populations from distant regions must represent species distinct from familiar European taxa (Shaw 2001), whereas recent taxonomic work suggests otherwise (e.g., Kruijer 2002; Burghardt & Gradstein 2008). Molecular data have mostly been used in two ways in support of those recent taxonomic revisions.

1. The association between sequence identity and species identity

One of the earliest uses of molecular data to assist with testing traditional species concepts was to search for fixed genetic differences among species. The absence of such differences served as evidence in support of synonymization (Table 1). It is extremely important, however, to keep in mind that a lack of molecular support for morphologically distinguishable species constitutes negative evidence. As such, a lack of molecular evidence that two species are different never provides definitive evidence that the samples belong to a single species, but rather just fail to provide positive evidence that they are different species. Molecular data may not resolve putative species for a variety of reasons, including (1) the morphological characters on which the species are based are plastic (i.e., non-genetic), (2) the morphological characters that distinguish species are genetically-based but convergent, (3) the species diverged so recently that there has been insufficient time to accumulate nucleotide sequence differences in the loci investigated even though the taxa might be reproductively isolated and on separate evolutionary trajectories, (4) the molecular markers are inappropriate to resolve the species because of little variation or low substitution rates; and (5) the studied molecular markers do not belong to the portions of the genome that code for the morphological differences seen among individuals.

Many cases can be cited where morphologically defined species have been combined because molecular data indicated that they were indistinguishable (Table 1). While we do not suggest that such decisions are never justified, it is important to be conservative in making such judgements. In the *Frullania tamarisci* (L.) Dumort. complex, for example, nucleotide sequences from two generally variable plastid loci, *trnL* and *trnG*, and the nuclear ribosomal ITS region, revealed virtually no variation and no phylogenetic structure within the eastern North American species, *F. asagrayana* Mont., sampled from North Carolina to Maine (Ramaiya et al. 2010). However, variation at 12 hypervariable microsatellite loci revealed two well defined groups of populations, one generally northern in distribution and the other southern. Analyses of microsatellite variation patterns indicate little or no interbreeding between these two groups, which therefore represent reproductively isolated biological species. If we had only the sequence data – and it is noteworthy that the data come from genomic regions generally thought to be relatively variable – we would conclude that there is only one species, yet the microsatellite data clearly show that this interpretation is incorrect.

A particularly striking non-botanical case of this sort of problem comes from research on two species of crows. Mitochondrial DNA sequences, AFLP markers, isozymes, and microsatellites all fail to reveal any differentiation between the black carrion crow and the hood crow, yet the two taxa differ in expression profiles when the transcriptomes were compared using pyrosequencing of mRNA (Wolf et al. 2010). Thus, the two morphologically defined taxa appear to differ primarily in regulatory genes that control the expression of other protein-coding genes, which themselves may not differ between species at the nucleotide level. This is an extreme case in which negative evidence from a lack of differentiation in a wide variety of molecular markers leads to the erroneous conclusion that the two crow taxa are not distinct, but it serves to emphasize that caution must be used when lumping species because they are not distinguished by molecular data. We do not suggest that attempts to test morphologically defined species with molecular data are not useful and important, but simply that the nature of negative evidence be kept in mind. We do suggest that taxonomic decisions should be avoided when the molecular data are limited. A number of studies, for example, have
lumped morphospecies together based on a lack of differentiation in sequences from one locus, or even part of a locus (e.g., ITS 2). This would be analogous to arguing that two species are not distinguishable based only on comparisons of the upper right margin of the perichaetial leaves! The question, of course, is how much molecular data, and of what type, is sufficient to make taxonomic conclusions.

**TABLE 1.** Taxonomic re-arrangements at the species level in bryophytes resulting from the use of molecular data

<table>
<thead>
<tr>
<th>Species</th>
<th>Reason for description of new species</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Homalothecium californicum</em> Hedenäs et al</td>
<td>Polyphyly of <em>H. megaptilium</em> (Sullivan) Robinson and morphological re-assessment</td>
<td>Hedenäs et al. 2009</td>
</tr>
<tr>
<td><em>Conocephalum salebrosum</em> Szweyk., Buczk. &amp; Odrzyk.</td>
<td>Identification of several lineages and subsequent morphological re-assessment</td>
<td>Szweykowski et al. 2005</td>
</tr>
<tr>
<td><em>Sphagnum beringiense</em> A.J.Shaw, R.E.Andrus &amp; B.Shaw</td>
<td>Identification of several lineages and subsequent morphological re-assessment</td>
<td>Shaw et al. 2008</td>
</tr>
<tr>
<td><em>Schizymenium shevockii</em> A.J.Shaw</td>
<td>Identification of several lineages and subsequent morphological re-assessment</td>
<td>Shaw 2000</td>
</tr>
<tr>
<td><em>Platyhypnidium mutatum</em> Ochyra &amp; Vanderp.</td>
<td>Sequence identity</td>
<td>Stech &amp; Frahm 1999</td>
</tr>
<tr>
<td><em>Brachytheciastrum</em> Ignatov &amp; Huttenun spp.</td>
<td>Polyphyly</td>
<td>Vanderpoorten &amp; Goffinet 2006</td>
</tr>
<tr>
<td><em>Hygroamblystegium</em> Loeske spp.</td>
<td>Polyphyly and morphological continuum owing to plasticity</td>
<td>Vanderpoorten et al. 2004</td>
</tr>
<tr>
<td><em>Leucobryum juniperoides</em> (Brid.) Müller</td>
<td>Polyphyly and morphological identity</td>
<td>Vanderpoorten et al. 2003</td>
</tr>
<tr>
<td><em>Leptoscyphus azoricus</em> (H. Buch &amp; Perss.) Grolle</td>
<td>Paraphyly and morphological identity</td>
<td>Vanderpoorten et al. 2010a</td>
</tr>
<tr>
<td><em>Brachythecium appleyardiae</em> McAdam &amp; A.J.E.Smith</td>
<td>Sequence and morphological identity</td>
<td>Blockeel et al. 2005</td>
</tr>
<tr>
<td><em>Brachythecium nelsonii</em> Grout, <em>B. gelidum</em> Bryhn</td>
<td>Sequence and morphological identity</td>
<td>Draper &amp; Hedenäs 2009</td>
</tr>
<tr>
<td><em>Timmia sibirica</em> Lindb. &amp; Arnell</td>
<td>Polyphyly</td>
<td>Budke &amp; Goffinet 2006</td>
</tr>
<tr>
<td><em>Thamnobryum maderense</em> (Kindb.) Hedenäs</td>
<td>Sequence identity</td>
<td>Stech et al. 2001</td>
</tr>
<tr>
<td><em>Hypopterygium tamarisci</em> (Sw.) Bridel ex Müller complex</td>
<td>Sequence and morphological identity</td>
<td>Pfeiffer et al. 2000</td>
</tr>
<tr>
<td><em>Hypnum heseleri</em> Ando &amp; Higuchi</td>
<td>Sequence identity</td>
<td>Hill et al. 2006</td>
</tr>
<tr>
<td><em>Plagiochila killarniensis</em> Pearson</td>
<td>Polyphyly</td>
<td>Heinrichs et al. 2004</td>
</tr>
<tr>
<td><em>Herbertus azoricus</em> P.W.Richards</td>
<td>Paraphyly and morphological identity</td>
<td>Feldberg et al. 2004</td>
</tr>
<tr>
<td><em>Plagiochila rubescens</em> (Lehman &amp; Lindenh.) Lindenh.</td>
<td>Sister relationship and morphological identity</td>
<td>Groth et al. 2004</td>
</tr>
<tr>
<td><em>Ditrichum plumbicola</em> Crundw.</td>
<td>Paraphyly</td>
<td>Frahm et al. 2008</td>
</tr>
<tr>
<td><em>Drepanoclados aduncus</em> (Hedwig) Warnst. complex</td>
<td>Polyphyly, morphological continuum, expected plasticity</td>
<td>Hedenäs 2008</td>
</tr>
<tr>
<td><em>Fissidens luisieri</em> P. de la Varde</td>
<td>Polyphyly and morphological identity</td>
<td>Werner et al. 2009</td>
</tr>
</tbody>
</table>
2. The application of the phylogenetic concept in bryophytes

Because of advances in molecular techniques and the ability to generate DNA sequences at progressively lower costs, phylogenies based upon multi-gene analyses have become increasingly available, allowing systematists to test species circumscriptions through application of phylogenetic species concepts. The main criterion of phylogenetic species concepts is monophyly. Any species, or group of species, found to be of para- or polyphyletic origin, should be merged within a single, monophyletic unit. This procedure can yield dramatic reductions to synonymy (see, e.g., Liede-Schumann & Hartmann 2009).

Two issues are associated with the application of the monophyletic species concept. First, paraphyletic relationships are often obtained owing to poorly resolved phylogenies, whereas in fact a monophyletic origin could not be significantly rejected based upon analyses constraining conspecific accessions to monophyly (see Harris 2008, for a review). Second, strict application of the monophyletic species concept, as opposed to allowing paraphyly under some circumstances, may be problematic, and, arguably, undesirable (see Zander 2007 and Brummitt 2008, for reviews). Indeed, it might not always be possible to resolve the ‘true’ history of speciation into a tree of monophyletic species because some modes of speciation lead to paraphyletic species or even phylogenetic networks (Barraclough & Nee 2001). Recognizing non-monophyletic species re-opens the door to the acceptance of poorly defined units based on vague combinations of characters, but many closely related species may simply not be, in reality, reciprocally monophyletic (see Funk & Omland 2003, for review). Moreover, reciprocal monophyly for closely related species may be extremely difficult to document even when it is in fact the case.

When constructing phylogenies based on nuclear genes that may be duplicated or part of larger gene families, paralogy can be an issue that must be considered since it can produce phylogenies in which taxa are not resolved as monophyletic. Because the nuclear ribosomal ITS region is typically present in high copy number within individual genomes, incomplete concerted evolution can occur such that multiple divergent ITS alleles co-exist within an individual. Under such circumstances, molecular analyses may resolve non-monophyly of species as an artifact of comparing paralogous gene copies. Many phylogenetic analyses of mosses rely on single copy plastid loci where paralogy is typically less of an issue. However, two other processes have been increasingly shown to play an important role in bryophyte molecular phylogenies.

A first process likely to render gene trees and species trees incongruent is hybridization. Although long suspected because of morphologically intermediate forms between putative parental species, hybridization has only fairly recently been evidenced in bryophytes (see Natcheva & Cronberg 2004 and McDaniel et al. 2010, for a review). One obvious example is in the case of allopolyploid speciation. Allopolyploids originate through reticulation rather than though “normal” divergent evolution, and for that reason do not fit the paradigm of phylogenetic species delimitation. Moreover, many or perhaps most allopolyploid species appear to have originated multiple times and are therefore polyphyletic (Soltis & Soltis 1999). Nevertheless, most allopolyploid taxa appear to function as biologically and ecologically meaningful units of biodiversity and populations that may have originated independently can interbreed and exchange genes. It is neither practical nor biologically accurate to consider each independent derivation of an allopolyploid a separate species. Moreover, independent origins of allopolyploids (and in fact, species formed by other mechanisms) raises the question of what exactly constitutes monophyly; a single origin involving one individual, a group of closely-related individuals, a single population, etc.

The second process involved in the gene tree/species tree problem is the retention of ancestral polymorphism. In particular, budding speciation is the process by which a population becomes spatially isolated and diverges following, for example, a long-distance dispersal event. The newly isolated population may be small and local, and would initially possess a restricted subset of parental alleles that will be lost under drift at a faster rate than in the larger parental population. Founder events may commonly yield a geographically restricted species whose monophyletic set of haplotypes is embedded within a more widely distributed and still paraphyletic parental species. This asymmetrically paraphyletic relationship will persist until gene coalescence renders the parental species monophyletic. In this case, the cause of paraphyly is incomplete lineage sorting, yet the gene tree accurately reflects the history of population divergence.
Although gene trees for different loci may in many cases be incongruent because of incomplete sorting, budding speciation is predicted to produce parallel patterns of paraphyly across nuclear and cytoplasmic loci (Funk & Omland 2003). Zander (2009) argued that many or even most new species arise out of more broadly distributed ancestral species that are phylogenetically structured, so the newly evolved species will be phylogenetically embedded within a larger ancestral taxon, rendering the latter paraphyletic (Zander 2007). Zander (2009) thus proposed that traditional taxa resolved as paraphyletic or polyphyletic based upon gene tree analyses may be interpreted as populations of surviving ancestors that are phenotypically static through processes like stabilizing selection, whereas the newly evolving species experienced faster morphological evolution, perhaps because of smaller effective population sizes.

3. Differences in rates of morphological and molecular evolution

It is often assumed that rates of morphological and molecular evolution are highly correlated (Barraclough & Savolainen 2001; Soltis et al. 2002), yielding compatible species definitions from the two sources of information. Issues arise, however, when rates of molecular and morphological variation are uncoupled. Heterogeneity in rates of morphological evolution within monophyletic groups of bryophytes is evidenced by actual measures of morphological transition rates. In the liverwort genus Leptoscyphus Mitt. for example, L. cuneifolius (Hook.) Mitt. has retained a more or less constant morphology during about six million years, whereas other species in the genus, despite their younger ages, evolved many more morphological novelties (Devos & Vanderpoorten 2009). In the pleurocarpous moss sub-family Helicodontioideae, the monotypic Hedenasiastrum percurrents (Hedenäs) Ignatov & Vanderp. has been considered a living fossil that shares with its most recent ancestor, dated at 40 million years, all of its morphological traits, whereas one third of the morphological features of Rhynchostegiella macilenta (Renauld & Cardot) Cardot differ from those of its one million year old most recent ancestor (Aigoin et al. 2009). Retention of a constant morphology despite genetic divergence over millions, or even tens of millions of years, has been termed ‘cryptic speciation’ and is increasingly reported among bryophyte species (see Heinrichs et al. 2009a, b for review). In fact, fossils of mosses and liverworts from the Tertiary and even Secondary eras are, in general, similar to the modern flora (see Vanderpoorten et al. 2010b for review).

The fact that some lineages accumulate morphological transformations at a much faster rate than others suggests that many differences in complex morphological traits (Brakefield 2006) do not result from accumulated mutations in multiple genes, but are rather based on one or a few point mutations, or even changes in the mechanisms of gene regulation (see the crow example above). In the beach mouse, for example, a single amino acid substitution contributes to adaptive colour patterns (Hoekstra et al. 2006). Hedenäs & Eldenäs (2008) similarly evoked the possibility that a single or a few genes may be responsible for dramatic morphological modifications in some mosses, while the rest of the genome has had no time to sort out. Such an interpretation definitely contrasts with the view that rates of molecular and morphological evolution are generally correlated (Barraclough & Savolainen 2001; Soltis et al. 2002). Fast and dramatic morphological changes owing to single or few genetic changes, or changes in gene expression, can explain why characters often appear to shift states so easily along moss phylogenies. Such an explanation might also apply to other species of mosses that exhibit striking morphological difference, and yet share identical non-coding sequences with the common species they derive from (e.g., Platypnuidium mutatum Ochyra & Vanderp.; Stech & Frahm 1999; P. torrenticola (Ochyra, C. Schmidt & Bültmann) Ochyra & Bednarek-Ochyra; Werner et al. 2007; Thamnobryum angustifolium Nieuwland; Olsson et al. 2009; Leptodon corsicus Enroth et al.; Sotiaux et al. 2009). As emphasised by Hedenäs & Eldenäs (2008), we still know little about molecular mechanisms of morphogenesis in mosses. With the advances of population genomics and quantitative genetics, we will soon have the toolkits needed to discover how such bizarre morphologies evolved and to find the genes underlying ecologically important traits.
Conclusion

Molecular phylogenetics has been extremely powerful for revisiting traditional taxonomic hypotheses, and this is especially true in taxa with reduced morphologies like bryophytes. In particular, mono-or stenotypic genera have often been given an inflated taxonomic rank owing to their peculiar morphologies. More often than not, such morphologically distinct genera are found to be phylogenetically nested within larger groups. A good example of that is the genus *Ephemerum* Hampe, which almost all bryologists now accept as a member of the Pottiaceae based on molecular phylogenetic results despite the differentiated morphology. The reduction of the apparently thalloid liverwort genera *Mizutania* Furuki & Z. Iwatsuki (Masuzaki et al. 2010) and *Metzgeriopsis* Goebel (Gradstein et al. 2006) to extremely reduced expressions of the leafy liverwort genera *Calypogeia* Raddi and *Cololejeunea* (Spruce) Schiffn., respectively, provide additional examples of improved understanding gained by the analysis of sequence data.

Negative molecular evidence has been increasingly used to lump species with identical sequences at one or a few loci, but while sequence identity at loci that are usually variable within the groups can provide incentive for additional morphological and molecular investigation, it does by no means alone provide sufficient evidence for synonymization. Similarly, evidence for paraphyly does not necessarily provide strong evidence for combining the paraphyletic and nested species, because some speciation mechanisms lead to such patterns among recently evolved sister species. We argue that there are events of major biological import that occur when a new divergent taxon is “budded off” from within an ancestral widespread species; however, the point at which both species become reciprocally monophyletic can simply reflect the stochastic process of gene coalescence and is of no real biological significance in and of itself. Reproductive isolation through one mechanism or another is necessary, though not necessarily sufficient, for the development of reciprocal monophyly. Thus, the evolution of reproductive isolation is of critical importance evolutionarily, whereas the development of reciprocal monophyly is biologically trivial. It is for this reason that most evolutionary biologists who study the process of speciation focus on reproductive isolation, whereas those (systematists) primarily concerned with the delimitation of species tend to focus on the more philosophical concept of reciprocal monophyly. As Zander (2007) noted, insistence on avoiding paraphyly can result in a species concept that does not reflect the actual speciation process.

An important goal of this essay is to voice a word of caution, and to encourage case by case evaluation of all available data rather than simply applying a single criterion such a monophyly. In our view, the strict requirement that all species must be demonstrably monophyletic is rather like adopting an uncompromising view that reproductive isolation (i.e., the biological species concept) is the only valid evidence for species status, and that it must be complete for two species to be distinguished. We do wish to emphasize that molecular data, and phylogenetic analyses, have and will continue to provide invaluable information for making species-level systematic decisions. Given that an absence of molecular differentiation between putative taxa is negative evidence, one might argue that such information is of no value since it can always be that additional data would demonstrate the putative species to be distinct. How much negative data, and what sort of negative data, are sufficient for making conclusions? Of course we do not have the answer to that question. Nevertheless, we offer the following observations that can be kept in mind when applying molecular data to species-level problems. 1. Nucleotide sequence data from one or a few loci are often insufficient at the species level. The oft-stated criticism of molecular systematics in general that conclusions are based on an extremely limited portion of the genome (one or a few loci) and are therefore inadequate, is rather empty because of the high levels of inter-locus congruence in support of deep phylogenetic patterns that have repeatedly been observed over the last 20 years. At the species level, however, where stochastic processes of lineage sorting are more often in-progress, limited sampling of the genome can be a more serious problem. There is simply no way to defend making taxonomic decisions to lump species based on a lack of differentiation in sequence at one or a few loci. 2. The choice of loci is critical. It is well known that nuclear loci are typically more variable than plastid loci. Moreover, sampling multiple nuclear loci provides the opportunity to directly assess recombination (including interspecific interbreeding). 3. Sometimes other types
of molecular data should be brought to bear on species-level taxonomic problems. In particular, we suggest that DNA fingerprinting methods including RFLPs, ISSRs, and microsatellites, are especially useful for many species-level systematic problems. Although microsatellites can be difficult and/or expensive to develop for a particular genus, these markers have the advantage of being taxon-specific (reducing or eliminating the danger of amplifying artifactual “alleles” from fungal endophytes or other contaminants), co-dominant in expression, and highly variable. Moreover, microsatellite (and to a lesser degree, other fingerprinting methods) permit explicit analyses of interbreeding between putative species because multiple loci scattered across the genome are typically analyzed. We suggest that if species delimitation, and species differentiation, are the primary goals in a research program, nucleotide sequence data should be complemented with approaches that focus on larger numbers of unlinked loci that are more variable. As technology continues to improve, it may well be possible to obtain sequence data for very large numbers of unlinked loci and at that point we would change our recommendation to focus on sequence data, since the conceptual framework including substitution models and tree-building methods are best developed for these sorts of data. But until that time is upon us, we encourage bryophyte systematists working at the species level to supplement sequence data with information from other kinds of markers that are better suited to recently diverged taxa.

References


Hedenäs, L. (2008) Molecular variation in Drepanocladus aduncus s.l. does not support recognition of more than one


(Bryopsida) as inferred from molecular data. *Nova Hedwigia* 72: 251–257.


