Chapter 5

Phylogeny and species delimitations in European *Dicranum* (Dicranaceae, Bryophyta) inferred from nuclear and plastid DNA.

A. S. Lang, G. Bocksberger & M. Stech

Submitted to Molecular Phylogenetics and Evolution

Abstract

DNA sequences are increasingly used for taxonomy, inferring phylogenetic relationships and identifying species boundaries. Many specific methods to define species delimitation have appeared recently, with the generalized mixed Yule coalescent (GMYC) method being the most popular. However, only few studies on land plants have been published so far and GMYC analyses of bryophytes are largely missing. *Dicranum* is a large genus of mosses whose (morpho-) species are partly ill-defined and frequently confused. To infer molecular species delimitations, we reconstructed phylogenetic trees based on five chloroplast markers and nuclear ribosomal ITS sequences from 28 out of 30 species occurring in Europe. We further applied GMYC and PTP species delimitation methods in order to compare their discriminatory power with species boundaries inferred from the molecular phylogenetic reconstructions and with the morphological species concept. Phylogenetic circumscriptions were congruent with the morphological concept for 24 species, while three taxa were molecularly indistinguishable from other closely related species. Phylogenetic relationships between *Dicranum* species remained largely unsupported. Automated species delimitation achieved similar results but tended to overestimate the number of potential species and exposed several incongruences between the morphological concept and inference from molecular phylogenetic reconstructions. It is concluded that GMYC and PTP methods potentially provide a useful and objective way of delimiting bryophyte species, but studies on further bryophyte data sets are necessary to infer whether incongruences might ensue from evolutionary processes and to test the suitability of these approaches.

Introduction

DNA sequence data are widely used for inferring species delimitations and phylogenetic

relationships. Specific methods to analyze species boundaries based on molecular phylogenetic reconstructions, however, have appeared only recently (cf. Carstens et al. 2013 for review), with the generalized mixed Yule coalescent (GMYC) method (Fontaneto et al. 2007; Pons et al. 2006) being most popular. This method estimates the point of transition from the level of species to population evolutionary processes, i.e. it detects species boundaries based on differences in branching rates at both species and population levels. Automated species delimitation methods are therefore considered especially useful in organisms with unclear species boundaries, due to poor taxonomy knowledge and because processes such as lineage sorting and introgression can obscure the species tree signal (O'Meara 2010 and references therein). Most GMYC studies so far focused on different animal groups (e.g. Poulakakis *et al*. 2012; Zaldívar-Riverón *et al*. 2010) and very few examples of analyses of other organisms such as algae (e.g. Leliaert *et al*. 2009), fungi (e.g. Parnmen *et al*. 2012) and land plants (e.g. Hernández-León *et al*. 2013) have been published. GMYC analyses of plant, and especially bryophyte, species are hence still largely missing.

Bryophytes are an important component of terrestrial ecosystems and count up to 18,000 known species (*Goffinet & Shaw 2009*). Nevertheless, because of the limited number of morphological characters available, the morphological plasticity of species and the generally broad geographical distribution, the taxonomy of many bryophyte lineages is still ambiguous. Molecular data can facilitate the circumscription of species, especially in taxa with extreme morphological similarities (e.g. Dong *et al*. 2012; *Hedenäs & Eldenäs 2007*; Heinrichs *et al*. 2009; Stech *et al*. 2013).

Species circumscription and identification in the Holarctic moss genus *Dicranum* (Dicranaceae, Bryophyta) has been notoriously difficult. The genus counts more than 90 species (www.Tropicos.org; Frey & Stech 2009), many of which are broadly distributed and display a great range of morphological plasticity, and only few species are habitat-specific (Hedenäs & Bisang 2004). Moreover, *Dicranum* and related genera display little molecular variation, as shown in previous studies (Cox *et al*. 2010; La Farge *et al*. 2002; Stech 1999; Stech *et al*. 2012). Assessing species delimitations in *Dicranum* is thus challenging both at the morphological and molecular level. Our recent studies on the *Dicranum scoparium* and *D. acutifolium* species complexes (*Lang & Stech 2014*; Lang *et al*. *in press*) as well as on boreal-arctic *Dicranum* species (Lang *et al*. 2014) showed that in several cases conclusive species delimitations could only be obtained from combined analyses of several chloroplast markers and nuclear ribosomal ITS sequences.

The present study aims to elucidate species boundaries within *Dicranum* on a broader geographic scale, including 27 of the 29 *Dicranum* species occurring in Europe (Hedenäs & Bisang 2004) plus *D. septentrionale* Tubanova & Ignatova, a newly recorded species in Scandinavia (Lang *et al*. *in press*). Molecular phylogenetic reconstructions based on five chloroplast markers (*trnH*$_{GUG}$*-psbA*, *rps4-trnT*$_{UGU}$ and *trnL*$_{UAA}$*-trnF*$_{GAA}$ intergenic spacers, *rps19-rpl2, rpoB*) plus the nrITS1-5.8S-ITS2 region will be used to test, for the first time in bryophytes, the congruence of two automated species delineation approaches, the general mixed Yule-coalescent (GMYC) models and Poisson tree processes (PTP). Sequence-based species delimitations will furthermore be compared with morphologically recognized species.

Material and methods

Sampling— A total of 202 *Dicranum* specimens were sampled (Appendix 1), representing 27 species of the 29 European species recognized by Hedenäs and Bisang (2004) and including the new European species record of *D. septentrionale*: six *Dicranum acutifolium* (Lindb. & Arnell)

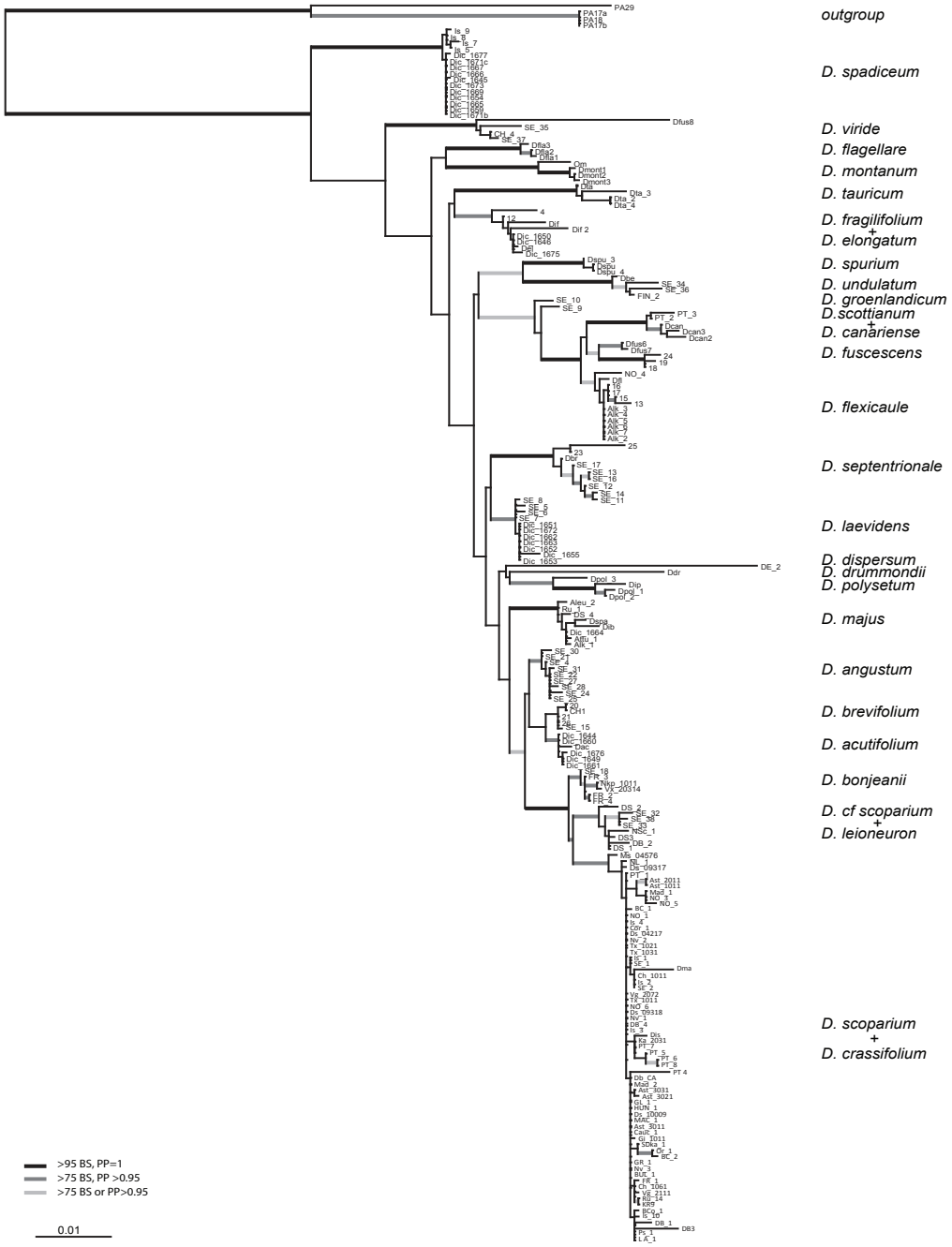

FIG. 1. Single optimal maximum likelihood phylogenetic reconstruction of 28 European *Dicranum* species inferred from the partitioned (non-coding chloroplast, chloroplast and ITS) nucleotide matrix including indels coded by simple indel coding (SIC). Branch thickness and colours indicate bootstrap support and posterior probabilities from ML and respectively Bayesian analyses.

C.E.O. Jensen, nine *D. angustum* Lindb., six *D. bonjeanii* De Not., five *D. brevifolium* (Lindb.) Lindb., three *D. canariense* Hampe ex Müll. Hal., five *D. crassifolium* Sérgio, Ochyra & Séneca, one *D. dispersum* Engelmark, one *D. drummondii* Müll. Hal., four *D. elongatum* Schleich. ex Schwägr., three *D. flagellare* Hedw., 11 *D. flexicaule* Brid., four *D. fragilifolium* Lindb.,six *D. fuscescens* Turner, two *D. groenlandicum* Brid., 11 *D. laevidens* R.S. Williams, three *D. leioneuron* Kindb., eight *D. majus* Turner, four *D. montanum* Hedw., four *D. polysetum* Sw., 65 *D. scoparium* Hedw., two *D. scottianum* Turner ex Robt. Scott, nine *D. septentrionale*, 15 *D. spadiceum* J.E. Zetterst., three *D. spurium* Hedw., four *D. tauricum* Sapjegin, four *D. undulatum* Schrad. ex Brid. and four *D. viride* (Sull. & Lesq.) Lindb. specimens. The sampling included 40 specimens newly sequenced for all six markers employed here, four specimens of which ITS sequences had already been generated by Tubanova *et al*. (2010) and Ignatova and Fedosov (2008) and 162 specimens of which chloroplast and ITS sequences were generated for previous studies (Lang & Stech 2014; Lang *et al*. 2014, *in press*; Stech 1999; Stech *et* al. 2006). As previous studies showed that *Holomitrium* is sister to *Dicranum* (La Farge *et al*. 2002; Stech *et* al. 2006), four samples, one *H. crispulum* Mart. and three *H. arboreum* Mitt, were chosen as outgroup representatives.

DNA extraction, amplification and sequencing— The greenest parts of single gametophyte stems were selected for DNA extraction and cleaned manually with demineralised water under a binocular. Total DNA extraction was carried out using the NucleoSpin® Extract II Kit (Macherey-Nagel, Düren, Germany). Six markers employed to delimit closely related *Dicranum* species in Lang and Stech (2014) and Lang *et al*. (2014 *in press*) were amplified and sequenced, i.e. five chloroplast regions (partial *rpoB* gene, $trnH_{GUG}$-*psbA*, *rps19-rpl2*, *rps4-*$trnT_{UGU}$ and $trnL_{UAA}$-$trnF_{GAA}$ intergenic spacer) and the nuclear ribosomal nrITS1-5.8S-ITS2 region. PCR amplifications were performed as described in Lang and Stech (2014). All PCR products were purified and sequenced at Macrogen Inc. (www.macrogen.com). GenBank accession numbers of all sequences are listed in Appendix 1.

Alignment and phylogenetic reconstruction— Sequences were aligned in Geneious v5.3.6 (Biomatters 2010) using 65% similarity matrix costs, and manually adjusted. Short hairpin-associated inversions in the *trnH-psbA* spacer, which can flip at the population level and may significantly reduce phylogenetic structure if undetected (Borsch & Quandt 2009; Quandt & Stech 2004; Whitlock *et al*. 2010), were positionally separated in the alignment and the corresponding indels were excluded.

Phylogenetic inferences were based on maximum likelihood (ML) and Bayesian inference (BI) analyses. Gaps were coded as informative by a simple indel coding strategy (SIC) (Simmons and Ochoterena 2000) implemented in SeqState (Müller 2004). To check for incongruence, phylogenetic reconstructions based on chloroplast and nuclear sequences were visually compared. In addition, an incongruence length difference test (ILD, Farris *et al*. 1994) as implemented in PAUP* 4.0b10 (Swofford 2002) was performed with 100 replicates. As both visual inspections and the ILD test indicated that the plastid and nuclear tree topologies were congruent (p=0.06), the two datasets were combined.

Three nucleotide partitions were used in ML and BI, namely the non-coding chloroplast markers (*rps4-trnT*, *trnL-trnF*, *trnH-psbA*, *rps19-rpl2*), the coding chloroplast region *rpoB* and the nuclear spacer nrITS. ML analyses were carried out with RAxML v.7.2.6 (Stamatakis 2006) employing the graphical user interface raxmlGUI v.0.93 (Silvestro & Michalak 2012) with the default GTR model

of nucleotide substitution and +Γ rate heterogeneity for all partitions. Bootstrap analyses under ML were done using the thorough bootstrap heuristics algorithm with 20 runs and 1000 replicates. BI analyses were run on the CIPRES science gateway (Miller *et al*. 2010). Bayesian posterior probabilities were calculated based on the Markov chain Monte Carlo (MCMC) method, using MrBayes v3.2.1 x64 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). The a priori probabilities supplied were those specified in the default settings of the program. Best-fit models of nucleotide sequence evolution were selected according to the Akaike information criterion in MrModeltest (Posada and Crandall 1998) executed through PAUP*, namely HKY + Γ for the non-coding chloroplast partition, and HKY + I for coding and nuclear partitions. Sequence and indel data were treated as separate and unlinked partitions, employing the restriction site model ('F81') for the indel matrix as recommended by Ronquist *et al*. (2005). Two runs with four chains were run simultaneously (11×10^6 generations), with the temperature of the single heated chain set to 0.5. Chains were sampled every 1,000 generations and the respective trees written to a tree file. Fifty percent majority rule consensus trees and posterior probabilities of clades were calculated by combining the two runs and using the trees sampled after the chains converged. Trace plots generated in Tracer v1.5 (Rambaut & Drummond 2007) were used to check for convergence of the runs (plateaus of all runs at comparable likelihoods) and to infer the 'burnin', which was set to 25%.

Sequence-based species delimitation— Species boundaries were estimated using the GMYC (Fontaneto *et al*. 2007; Monaghan *et al*. 2009; Pons *et al*. 2006) and the PTP (Zhang *et al*. 2013) approaches. As GMYC requires a fully resolved topology with branch length estimates, we reconstructed an ultrametric tree with a strict molecular clock using parameters specified in BEAUti v. 2 and implemented in BEAST version 2.1.1 (Bouckaert *et al*. 2014). Branch lengths were estimated under a Yule prior with HKY nucleotide substitution model for each data partition. We included a gamma rate heterogeneity and no invariant sites for the chloroplast partition and both *rpoB* and ITS partitions included no gamma rate heterogeneity but estimated invariant sites. In the absence of fossil records, we applied a plastid substitution rate of 5.0×10^{-4} SD of 2.0-8.0 $\times 10^{-4}$ subst./site/My following Villarreal and Renner (2014) for chloroplast and *rpoB* partitions and a substitution rate of 1.35×10^{-3} subst./site/My for ITS as used in Heinrichs *et al*. (2006). The MCMC chains were run with 20×10^6 generations, saving the results every 2000th generation. The convergence of the runs was examined in Tracer v1.5. The maximum clade credibility tree was built from the combined runs after eliminating 25% of the trees for burnin in TreeAnnotator v1.7.2. The GMYC approach was carried out in R 2.15 (R Development Core Team 2013) using the splits (Ezard *et al*. 2009) and ape (Paradis *et al*. 2004) packages. The number of clusters and singletons were estimated by running both single and multiple threshold optimisations and using a multimodel Akaike information criterion with a model cutoff of deltaAICc = 7 (Monaghan *et al*. 2009; Pons *et al*. 2006; Powell 2012). On the contrary to the GMYC approach, PTP neither requires an ultrametric tree nor a sequence similarity threshold as input data because speciation rate is modelled by using the number of substitutions between branching and speciation events (Zhang *et al*. 2013). We therefore used the RaxML trees as input data, with 500,000 MCMC generations, thinning set to 100 and burnin at 25%. The calculations were conducted on the bPTP websever (http://species.h-its.org/ptp/).

Unbalanced sampling can affect the estimates of haplotypes and thus might overestimate the number of potential species (Bergsten *et al*. 2012; Zhang *et al*. 2013). Therefore, GMYC and PTP

analyses were additionally conducted on a reduced alignment containing only unique sequences (haplotypes). This reduced alignment, automatically obtained from the raxmlGUI interface, contained 145 sequences, with the strongest reduction in *D. scoparium* sequences retaining 18 out of the initial 65 sequences. The ultrametric and RaxML trees were reconstructed following the above-mentioned methods.

Results

Phylogenetic reconstruction— The total chloroplast alignment comprised 1914 positions, of which 222 were variable, and 132 of the variable characters were parsimony-informative. Of the 1142 positions in the ITS alignment, 124 ambiguous positions were removed from the subsequent calculations. The remaining 1019 positions comprised 217 variable characters, of which 139 were parsimony-informative. Simple indel coding of the combined dataset yielded 240 additional characters (excluding three corresponding to an inversion in *psbA-trnH*), of which 148 were parsimony-informative.

The single optimal ML tree of the combined markers is shown in Fig. 1, with bootstrap support (≥75% BS) from likelihood analyses and posterior probabilities (PP ≥95) from Bayesian inference indicated on the branches. The phylogenetic reconstruction resolved 23 clades that corresponded to morphological species, including the two species with only one sample (Fig. 1). While the clades of *D. acutifolium, D. angustum, D. bonjeanii, D. brevifolium, D. flagellare, D. fuscescens, D. laevidens, D. majus, D. montanum, D. polysetum*, *D. scoparium* s.l. (including *D. leioneuron*, *D.* cf. *scoparium*, and *D. scoparium* s.s., cf. Lang & Stech 2014), *D. septentrionale, D. spadiceum, D. spurium, D. tauricum, D. undulatum* and *D. viride* were strongly supported (≥81% BS, PP <0.97), *D. flexicaule* was supported only in the Bayesian reconstruction (62% BS, PP 0.99).

Dicranum groenlandicum did not form a monophyletic clade. Six species were molecularly indistinguishable from other closely related species: *D. fragilifolium* and *D. elongatum* formed a highly supported clade (95% BS, PP 0.99). While *D. crassifolium* was intermingled with *D. scopariums* s.s., *D leioneuron* clustered with North American specimens of *D.* cf. *scoparium* in a highly supported clade (92% BS, PP 1). Finally, *D. scottianum* and *D. canariense* formed a highly supported clade (100% BS, PP 1). However, the samples of both *D. scottianum* and *D. canariense* clustered in supported subclade (94, 97% BS, PP 1, respectively).

Sequence- based species delimitation— The lineage through-time plot (Fig. 2b, c) indicated an exponential increase in branching rate near the tip of the tree. The single threshold GMYC model using the ultrametric phylogenetic tree created in BEAST resulted in the identification of 24 *Dicranum* clusters with high probabilities (CI= 23-26, lnL of null model= 741.079, ML of GMYC model= 748.162, p= 0.00269**) and 10 additional lineages consisting of single sequences, resulting in a total of 34 entities, excluding the outgroup (Fig. 2 a, b). The multiple threshold method gave four threshold times, resulting in a total of 58 entities that consisted of 38 clusters (CI= 30-39, lnL of null model= 741.079, ML of GMYC model= 752.849, p=0.000634***) and 20 singletons, excluding the outgroup (Fig. 2 c; Appendix 2). Although the multiple-threshold option was statistically preferred over the single-threshold option (deltaAIC=2.944), neither model was significantly different (Chi-square= 9.375, d.f.= 9, p= 0.40339). An inspection of the results obtained from both analyses revealed that the multiple-threshold GMYC model considered a higher number of clusters from samples that belonged to single lineages (Fig. 2a). Therefore, we

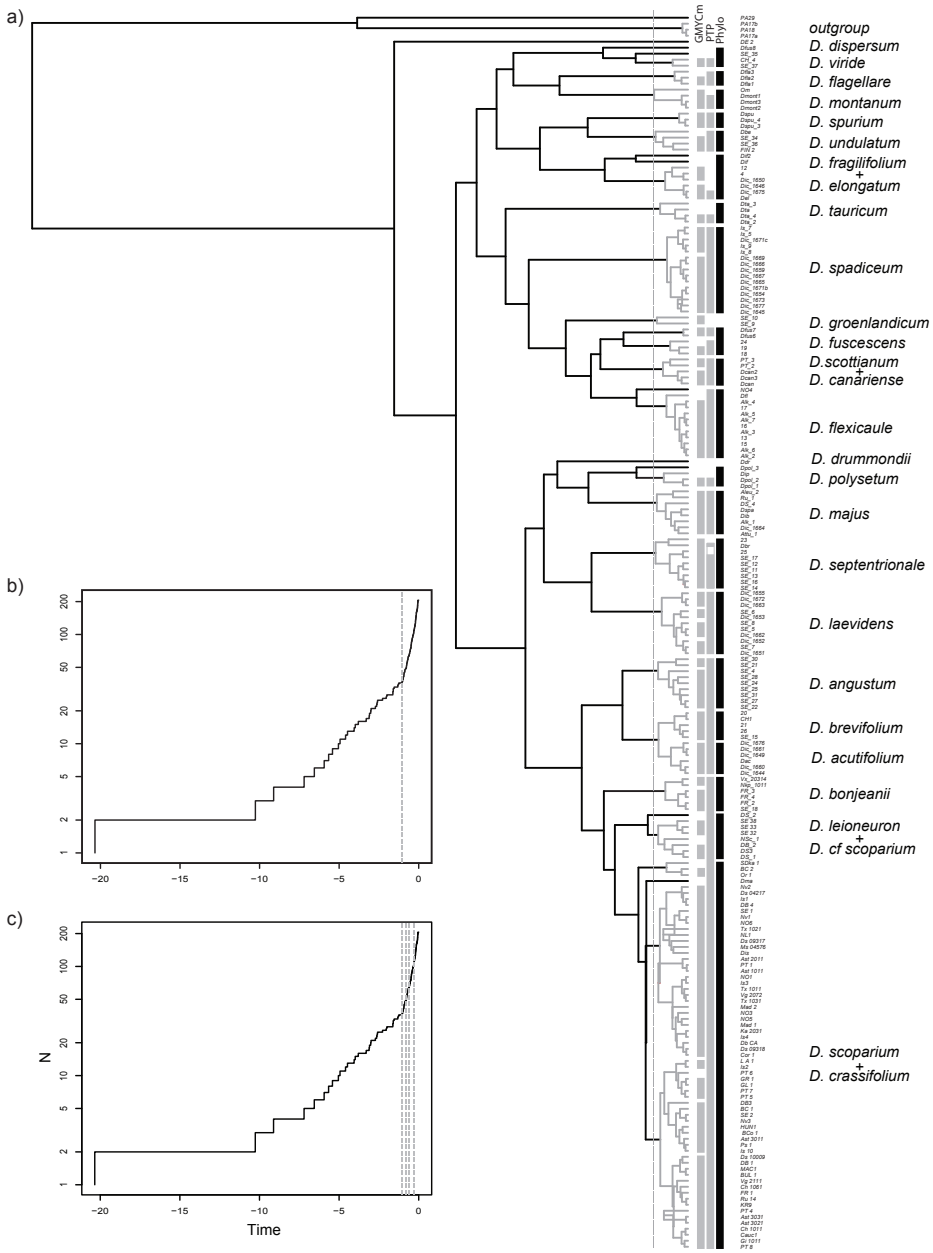

FIG. 2. Ultrametric tree depicting the relationship of *Dicranum* based on Bayesian analysis using a Yule model in BEAST and with fit of the general mixed Yules coalescent (GMYC) single threshold model from plastid and nuclear data. Branch length fitted a strict molecular clock (a). Lineages-through-time plot for single (b) and multiple (c) GMYC thresholds are illustrated. Estimated entities are indicated in grey. The vertical dashed line represent the timing of the earliest coalescent event. The two grey columns indicate the estimated entities from the multiple-threshold model (GMYCm) and the maximum likelihood solution of the Poisson tree processes (PTP), respectively. The black column indicates the phylogenetic boundaries of the *Dicranum* taxa (Phylo).

took a more conservative approach and discussed only the results of the single-threshold method.

The trees resulting from PTP gave similar results to GMYC (Fig. 2a). The number of estimated species varied between 25 and 116, excluding the outgroup (acceptance rate= 0.593), with 37 partitions supported by the ML search, excluding the outgroup (Fig. 2a, Appendix 3).

GMYC results based on the reduced alignment were similar to the results based on the extended alignment. The single-threshold model indicated the presence of 23 clusters and 31 entities while the multiple models resulted in four threshold times and resulted in 30 clusters and 48 entities, excluding the outgroup (CI= 9-40/ 38-68, lnL of null model= 259.1778/ 259.1778, ML of GMYC model= 263.0662/ 266.5076, p=0.051/ 0.023*, respectively; Table 1). The number of estimated species obtained from PTP method ranged between 49-102 entities, excluding the outgroup (acceptance rate= 0.716), with 42 partitions supported by the ML search, excluding the outgroup (Appendix 4).

Discussion

Phylogenetic reconstruction versus morphological species— The present study comprises the largest molecular dataset of *Dicranum* available so far, including all but two *Dicranum* species occurring in Europe following Hedenäs and Bisang (2004), plus *D. septentrionale*, recently described from Russia and newly identified in Scandinavia (Tubanova *et al*. 2010; Lang *et al*. 2014). The majority of the analysed species (23 out of 28, including two singletons), were molecularly recognisable based on the combined analysis of five chloroplast markers and nuclear ribosomal ITS sequences (Fig. 1), albeit not all with significant statistical support. The results support our recent phylogenetic studies on *Dicranum* species complexes and Arctic *Dicranum* species (Lang & Stech 2014; Lang *et al*. 2014; Lang et al. in press) in that a combination of molecular markers data can clarify species circumscriptions in *Dicranum*, and that the low resolution and clade support within *Dicranum* in earlier analyses (e.g. La Farge *et al*. 2002; Stech et al. 2006; Tubanova *et al*. 2010; Tubanova & Ignatova 2011) was a result of too few molecular markers analysed (cf. also Stech & Quandt 2010). Furthermore, the present study shows that, at least for Europe, the molecular data to a large extent support the morphological species concept, despite morphological confusions and subtle diagnostic characters in several species (e.g. Lang *et al*. in press; Tubanova *et al*. 2010).

In contrast to these results, nine species showed discrepancies between their morphological concepts and their molecular circumscription, namely *D. groenlandicum*, *D. elongatum*, *D. fragilifolium*, *D. scottianum, D. canariense*, *D. leioneuron* and *D.* cf *scoparium*. *Dicranum groenlandicum* was resolved as paraphyletic but without significant statistical support (Fig. 1). This arctic species is morphologically very similar to *D. laevidens* and, in absence of sporophytes, both species are essentially differentiated based on the growth form. However, recent molecular studies on arctic *Dicranum* suggested that both species represent two separate entities (Lang *et al*. 2014). The present phylogenetic reconstruction confirms the separation of *D. groenlandicum* from *D. laevidens* and further confirms the delineation of the latter species. Nevertheless, additional sequences of *D. groenladicum* are necessary to infer its delimitation. *Dicranum elongatum* and *D. fragilifolium* are morphologically different and occupy different habitats (Ireland 2007). Moreover, *Dicranum elongatum* is frequently confused with *D. groenlandicum*, while *D. fragilifolium* shares morphological similarities with *D. tauricum* (Hedenäs & Bisang 2004; Ireland 2007). Despite their clear morphological distinctions, the present molecular phylogenetic reconstruction indicates that both *D. elongatum* and *D. fragilifolium* belong to the same taxon (Fig. 1). The two Macaronesian-Atlantic European species *D. canariense* and *D. scottianum* were resolved in one well-supported clade.

Because of their morphological resemblance, *D. canariense* has been considered as a subspecies or variety of *D. scottianum* (Tropicos.org). In the current concept, *D. canariense* differs from the latter by its strongly denticulate margins and thick and denticulate costa (Hedenäs & Bisang 2004). The sampling included in this study confirms their close relationship and indicates that both taxa should be distinguished at subspecies level, however a larger sampling would be necessary to confirm these results. Morphological and ecological characters of *D. leioneuron* have been discussed several times, as it is frequently confused with either *D. bonjeanii* or *D. scoparium* (Ahti & Isoviita 1962; Corley 1991). Consequently, *D. leioneuron* has been sometimes considered as an ecotype of *D. scoparium* or a variety of *D. bonjeanii* (Ahti & Isoviita 1962), a hypothesis that is rejected by the present phylogenetic reconstructions (Fig. 1), which in turn confirm the observations of Corley (1991). Despite being molecularly separated from *D. bonjeanii* and *D. scoparium* s.str., the *D. leioneuron* specimens included in this study clustered in a well-supported lineage together with North American samples, named as *D.* cf. *scoparium* in Lang and Stech (2014). Morphology and habitat of these two groups are, however, clearly different: the North American specimens have falcate-secund leaves that are serrate on the margins and a lamellate costa. The *D. leioneuron* specimens, on the other hand, have all the characteristics of this species, i.e. small and erect-patent leaves; very thin nerve and without dorsal lamellae. Additionally, flagellary shoots are common in this species. Although the present data does not indicate any hybridization processes, the use of other molecular methods or more variable markers could bring new insights in understanding the relationship between *D. leioneuron* and *D. scoparium*. Finally, *D. crassifolium* is a species that has been described recently (Sérgio *et al*. 1995) and that has been found only in few places in Europe. This species resembles *D. scoparium* but is most similar to *D. transylvanicum* (not included here) due to a bi- or even tristratose leaf lamina and denticulate leaf margins. The present molecular phylogenetic inferences, however, show that this species actually corresponds to *D. scoparium*. *Dicranum scoparium* is known to be very plastic morphologically and occurs in a very broad range of habitats (Hedenäs & Bisang 2004; Ireland 2007; Lang & Stech 2014; Smith 2004), including soil or humus, as well as on rocks or tree bases, in open and shady places where *D. crassifolium* grows as well (Sérgio *et al*. 1995). What triggers the deviating leaf lamina morphology of *D. crassifolium*, and how *D. transylvanicum* relates to *D. crassifolium* and *D. scoparium*, remais to be tested.

Various factors such as the environment or polyploidisation may account for the observed morphological variability of *D. crassifolium*, *D. scottianum* and *D. canariense,* for exemple. Deviating morphologies are frequently observed in bryophytes, especially in species growing in stressful environmental conditions (Buryová & Shaw 2005; Hedenäs *et al*. 2006; Pereira *et al*. 2013; Såstad 1998; Såstad *et al*. 1999; Spitale & Petraglia 2010). Most of the *Dicranum* species are widespread and found in a great range of habitats. Hence, local adaptation could partly explain the morphological differences of genetically similar taxa, such as observed in *D. fragilifolium* and *D. elongatum* or *D. leioneurion*. Although the present data does not indicate any hybridisation events, this genetic process is known to influence the morphology (Draper & Hedenäs 2009; Hedenäs 2008; Natcheva & Cronberg 2004; Sotiaux *et al*. 2009). Moreover, the consequences of the special sexual reproduction of *Dicranum*, i.e. dwarf males growing on the branch of a female plant (pseudomonoicy), are largely unknown and would deserve further investigations, in order to explain genetic relationship of closely related species.

TABLE 1. Type of alignment, species delimitation method and number of estimated entities obtained for Dicranum. LR and LR test of the GMYC single- threshold (GMYCs) and multiple-threshold (GMYCm) analyses are also mentioned. Significant values are indicated with an asterisk. The species delimitation results are compared with the number of supported phylogenetic entities (phylo) obtained from maximum likelihood analyses.

Alignment	method	Number of sequences	Number of estimated clusters	entities	LR	LR test
extended	GMYCs	206	24	33	9.723	0.021*
	GMYCm		35	58	19.778	0.011*
	PTP			37		
	phylo		21	23		
reduced	GMYCs	145	23	31	7.778	0.051
	GMYCm		30	47	14.660	0.023*
	PTP			42		
	phylo		20	22		

Species delineation using GMYC and PTP— The definition of boundaries between species clusters is essential, as it will influence the interpretation of the phylogenetic reconstructions (Powell 2012). However, one of the major drawbacks of molecular taxonomy is putting a non-arbitrary threshold for delineating species. The main advantage of general mixed Yule-coalescent (GMYC) or Poisson tree processes (PTP) methods is the objective estimation of phylogenetic entities and the circumscription of taxa based on branch length dynamics rather than sequences similarities (Monaghan *et al*. 2009; Pons *et al*. 2006). Although GMYC and PTP performances have been proven to be stable under a wide range of conditions, the accuracy of species delimitation methods will principally depend on the singularities of the data set and the initial species concept used (Talavera *et al*. 2013; Zhang *et al*. 2013). In this study, 34 species were recovered by GMYC single threshold methods, which corresponds generally well with the phylogenetic reconstruction.

However, disagreements were observed, such as in *D. scoparium* but also *D. viride*, *D. fragilifolium- D. elongatum*, *D. flexicaule D. fuscescens* and *D. polysetum*, where overestimations in the number of entities occurred compared to the molecular and morphological delimitations (Fig. 2a). Each of these species counted one additional entity when compared to the phylogenetic tree, except for *D. fragilifolium- D. elongatum* and *D. scoparium* which counted a total of three, respectively four entities. Simultaneously, GMYC calculations considered both sample of *D. groenlandicum* as one species and both *D. brevifolium* and *D. acutifolium* were considered as belonging to the same lineage. The number of ML estimates obtained from the PTP of the extended dataset were relatively similar to the results obtained from GMYC methods (Table 1). However, the number of PTP estimates based on the reduced dataset was slightly higher (Table 1). Simulations have shown that an unbalanced sampling are likely to increase the estimates of haplotypes of the oversampled species (Bergsten *et al*. 2012; Zhang *et al*. 2013) and each specimen of an undersampled species might be counted as separate entity (Zhang *et al*. 2013). In our study, the reduced sampling of *D. scoparium* did not decrease the number of potential species. On the contrary, most of the haplotypes or unique sequences, in particular within *D. scoparium*, were considered as single lineages (Appendix 4). The effect of unbalanced sampling in our dataset has probably less impact on the species delimitation due to the generally low variability in *Dicranum*. Indeed, weak signals and high levels of uncertainty can explain the large range of estimated

species in both PTP estimations (J. Zhang, pers. communication).

Overestimations in the GMYC have been observed in previous studies (e.g. Miralles and Vences 2013; Puillandre *et al*. 2012; Talavera *et al*. 2013) and were often related with errors in the GMYC methods or in the construction of the ultrametric, rather than to taxonomical knowledge gaps (Talavera *et al*. 2013; Zhang *et al*. 2013). As our PTP estimates, obtained from a RaxML tree, were relatively close to the phylogenetic clades and not substantially different from the GMYC results, we considered that errors in the ultrametric tree construction had little effects on the species delimitation. As for now, the GMYC and PTP analysis revealed multiple lineages within species in *Dicranum* that lack morphological and ecological support. Simultaneously, these methods showed an absence of DNA divergences between *D. acutifolium* and *D. brevifolium* as well as between *D. scottianum* and *D. canariense*, which indicates that these four morpho-species might belong to two single taxa.

Conclusions

Biodiversity assessments rely on the correct delimitation of species. The identification of bryophyte species is largely based on morphological characters, which are often subtle and difficult to apply, or prone to plasticity induced by environmental conditions. Phylogenetic species delimitations, on the other hand, also rely on a certain degree of subjectivity. Automated methods such as GMYC and PTP may provide a more objective approach to molecular species delineation based on maximum likelihood inferences, although inferred boundaries are only putative. Our results showed that DNA-based circumscriptions were generally congruent with morphological species delimitations. Nevertheless, GMYC and especially PTP methods exposed several incongruences between morphological concepts and inference from molecular phylogenetic reconstructions. These incongruences might ensue from evolutionary processes, but also display the need for further testing on other bryophyte data sets to infer the suitability of GMYC and PTP methods for species delimitation in bryophytes.