Mugrabi de Kuppler AL, Fagúndez J, Bellstedt DU, Oliver EGH, Léon J, Pirie MD (2015) Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae). Molecular Phylogenetics and Evolution 88:121–131. http://dx.doi.org/10.1016/j.ympev.2015.04.005 (accepted version)

Testing reticulate versus coalescent origins of *Erica lusitanica* using a species phylogeny of the northern heathers (Ericeae, Ericaceae)

Mugrabi de Kuppler, A. L.,[1] Fagúndez, J.,[2] Bellstedt, D.U.,[3] Oliver, E.G.H.,[4] Léon, J., [1] and Pirie, M.D.[3,5,6]

[1]INRES Pflanzenzüchtung, Rheinische Friedrich-Wilhelms-Universität Bonn, Katzenburgweg 5, 53115 Bonn, Germany

[2]Department of Plant and Animal Biology and Ecology, University of A Coruña, Faculty of Science, 15001 A Coruña, Spain

[3] Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

[4] Department of Botany and Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

[5] Institut für Spezielle Botanik und Botanischer Garten, Johannes Gutenberg-Universität, Anselm-Franz-von-Bentzelweg 9a, 55099 Mainz, Germany

[6] Author for correspondence (pirie@uni-mainz.de)

Abstract

Whilst most of the immense species richness of heathers (*Calluna*, *Daboecia* and *Eric*a: Ericeae; Ericaceae) is endemic to Africa, particularly the Cape Floristic Region, the oldest lineages are found in the Northern Hemisphere. We present phylogenetic hypotheses for the major clades of Ericeae represented by multiple accessions of all northern *Erica* species and placeholder taxa for the large nested African/Madagascan clade. We identified consistent, strongly supported conflict between gene trees inferred from ITS and chloroplast DNA sequences with regard to the position of *Erica lusitanica*. We used coalescent simulations to test whether this conflict could be explained by coalescent stochasticity, as opposed to reticulation (e.g. hybridisation), given estimates of clade ages, generation time and effective population sizes (Ne). A standard approach, comparing overall differences between real and simulated trees, could not clearly reject coalescence. However, additional simulations showed that at the (higher) Ne necessary to explain conflict in *E. lusitanica*, further topological conflict would also be expected. Ancient hybridisation between ancestors of northern species is therefore a plausible scenario to explain the origin of *E. lusitanica*, and its morphological similarities to *E. arborea*. Assuming either process influences the results of species tree and further evolutionary inference. The coalescence scenario is equivocal with regard the standing hypothesis of stepping stone dispersal of *Erica* from Europe into Africa; whereas reticulate evolution in *E. lusitanica* would imply that the colonisation of Tropical East Africa by *E. arborea* instead occurred independently of dispersals within the rest of the African/Madagascan clade.

Keywords: *Calluna*, *Daboecia*, *Erica*, hybridisation, coalescent stochasticity, reticulate evolution

Abbreviations: Internal Transcribed Spacer regions of nuclear ribosomal DNA (ITS); Tropical East Africa (TEA); Cape Floristic Region (CFR).

1. Introduction

Erica L. is among the largest genera of flowering plants (Frodin, 2004) with 830-840 species (Oliver and Oliver, 2003; Oliver and Forshaw, 2012). Most of the immense richness of *Erica* is endemic to the Cape Floristic Region (CFR) of South Africa, but species of *Erica* and closely related *Calluna* Salisb. and *Daboecia* D.Don (Ericaceae; Ericeae, commonly referred to as 'heaths' or 'heathers') are archetypal elements of open landscapes of Europe and surrounding areas, in both the Temperate and Mediterranean biomes. These northern heathers (Nelson, 2012) have been the subject of various empirical studies addressing evolutionary and ecological questions including the environmental factors dictating species distributions (Gil-López et al., 2014; Ojeda et al., 1998), and patterns of dispersal and genetic diversity (investigated for individual species by e.g. Beatty and Provan, 2012; Désamoré et al., 2010; Désamoré et al., 2012). The tree heather, *Erica arborea*, in particular, has been investigated to infer dispersal patterns between Europe and Tropical East Africa (TEA) (Désamoré et al., 2010) and within TEA (Gizaw et al., 2013).

The geographically widespread *E. arborea* is very similar in gross morphology to another species found exclusively in Europe: *E. lusitanica*. They share a tall habit and white corollas, and have been grouped in different formal and informal classifications (e.g. Bentham, 1839; 'Tree heathers', Nelson, 2012). However, the interpretation of morphological variation in *Erica* is far from straightforward: there is evidence for extensive homoplasy of morphological characters (Oliver, 2000; Pirie et al., 2011). Floral characters may evolve rapidly, probably as adaptations to changing (pollinator) environments (Pirie et al., 2011; Rebelo et al., 1985; Van der Niet et al., 2014), and vegetative characters such as adaptations to recurrent fires (Ojeda et al., 2005) may have undergone similar shifts. As a result, the classic generic classification (Bentham, 1839; Hansen, 1950), based on such characters, has long been considered artificial (Hansen, 1950; Oliver, 2000). In *E. lusitanica*, micromorphological characters such as indumentum and seed coat sculpture appear more similar to

some other northern species than they are to *E. arborea* (Fagúndez et al., 2010; Fagúndez and Izco, 2010; Nelson, 2012).

The obvious means to assess this kind of morphological complexity is the molecular phylogeny. However, *E. lusitanica* has not previously been included in phylogenetic analyses, nor have several other European species, and current knowledge of the relationship of species of *Erica* L. in general is limited. Evidence available to date suggests that the heathers (tribe Ericeae; Ericaceae, including *Erica*, *Calluna* and *Daboecia*) comprise a basal grade of 'northern', largely European, species (McGuire and Kron, 2005; Pirie et al., 2011) subtending a single, much larger, 'southern' clade ('African/Malagasy *Erica*'; Pirie et al., 2011). Data that we collected in the course of ongoing work on the phylogeny of Ericeae confirmed this general pattern. The results that we report here showed much improved resolution between lineages of the northern grade, revealing strong conflict between phylogenetic trees based on plastid data and independent nuclear ITS with regard to the position of *E. lusitanica*.

This gene tree conflict raises an alternative hypothesis to explain homoplasy of morphological characters in *E. lusitanica*: instead of indicating parallel evolution of traits, it could be the result of hybridisation between morphologically dissimilar species (de Villiers et al., 2013). However, gene tree conflict can instead represent incomplete lineage sorting, being the result of coalescent stochasticity given a linear (rather than reticulate) species tree (Nichols, 2001).

As is the case in many empirical examples of gene tree conflict (Blanco-Pastor et al., 2012; de Villiers et al., 2013; Maureira-Butler et al., 2008; Pirie et al., 2009), both coalescent and reticulate scenarios are in principle plausible for *E. lusitanica*: The possibility of ancient hybridisation events cannot be ruled out since hybridisation between extant *Erica* species is documented: wild hybrids between European species include *Erica* × *stuartii* (MacFarl.) Mast. (*E. tetralix* L. × *E. mackayana* Bab.); *Erica* × *veitchii* Bean (*E. arborea* L. × *E. lusitanica* Rudolphi.); *Erica* × *watsonii* Benth. (*E. ciliaris* L. × *E.*

tetralix); *Erica* × *williamsii* Druce. (*E. vagans* L. × *E. tetralix*); and *Erica* × *nelsonii* Fagúndez (*E. tetralix* × *E. cinerea* L.) (Fagúndez, 2006; Fagúndez, 2012; Nelson, 2012; Rose, 2007), and various further crosses have been achieved in cultivation (Nelson, 2012). Moreover, there are no obvious karyological barriers to homoploid hybridisation: polyploidy has not been reported in *Erica* and chromosome counts are constant at n=12 for all studied species with the exception of the European species *E. spiculifolia* that is n=18 (Nelson and Oliver, 2005). Coalescent stochasticity on the other hand can generally be assumed to result in greater or lesser differences between inferred gene trees depending in particular on effective population sizes through time (Nichols, 2001).

In this paper, we attempt to discern whether gene tree conflict in the northern Ericeae with regards to *E. lusitanica* is the result of reticulate evolution or coalescent processes. We present two independent gene trees from DNA sequences 1) of multiple plastid markers and 2) of nuclear ribosomal ITS; from samples representing multiple accessions of all northern species of *Erica*, *Calluna vulgaris* (L.) Hull and *Daboecia cantabrica* (Huds.) K.Koch and exemplar sampling of the large African/Madagascan *Erica* clade. We use coalescent simulations and ancestral state reconstructions of selected morphological characters, and use data concatenation and coalescence based approaches in order to infer and test reticulate and linear species trees under each assumption separately. Finally, we use these trees to reassess the hypotheses concerning colonisation of Tropical East Africa by the putatively closely related tree heather, *Erica arborea*.

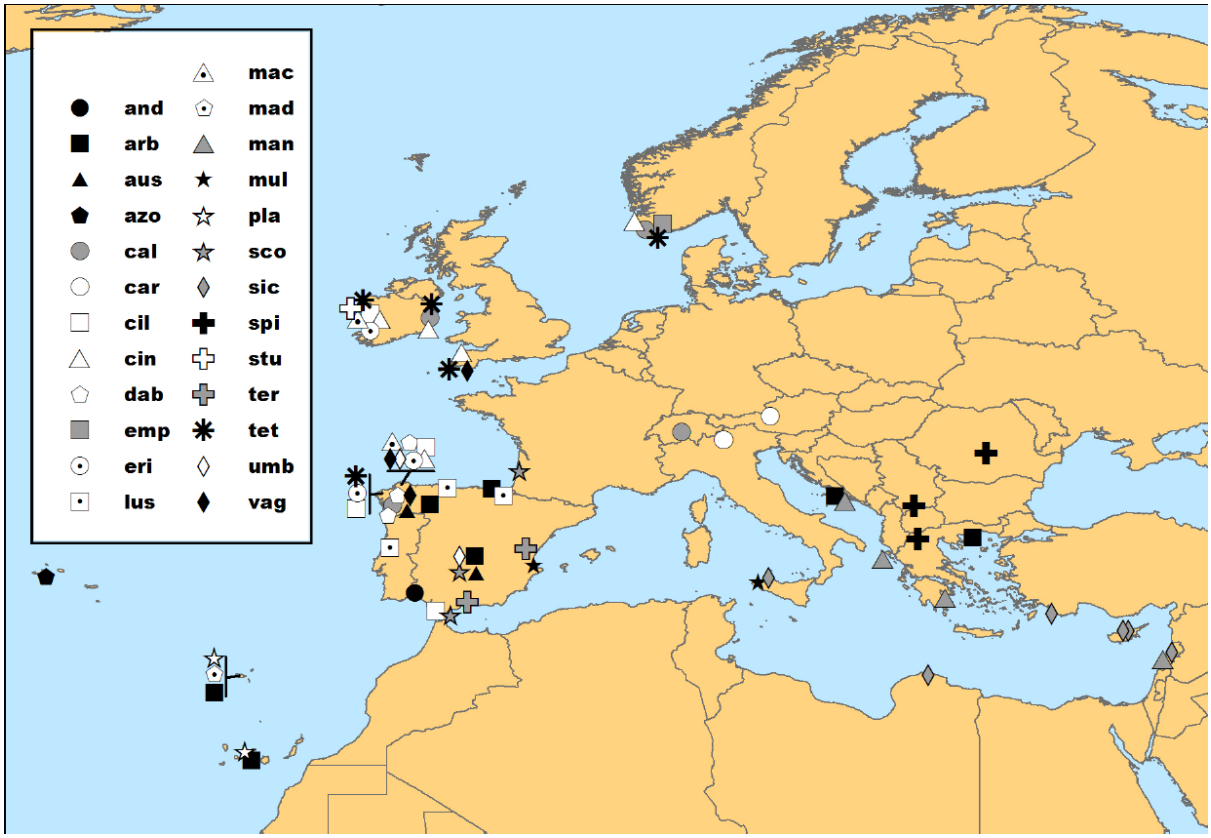
2. Materials and Methods

2.1 Taxon sampling

We sampled multiple populations from across the geographic distributions of all 21 non-hybrid species and two subspecies of *Erica* recognised by Nelson (2012) within the northern area (including for the first time *E. umbellata* and *E. maderensis*, as well as *E. platycodon* and *E. azorica* not

previously included in phylogenetic analyses of the genus), plus multiple accessions of *Calluna vulgaris* and *Daboecia cantabrica* (123 accessions in total). In addition, we sampled one naturally occurring northern *Erica* hybrid (*E.* x *stuartii*); seven *Erica* species from sub-Saharan Africa and Madagascar, representing the African/Madagascan clade (Pirie et al., 2011); and *Empetrum nigrum* L. (Ericoideae) as outgroup. A map illustrating the distribution of the northern samples is presented in Fig. 1, and full accessions details and authors of taxa (following the nomenclature of Nelson, 2012; Oliver and Forshaw, 2012) in Appendix A. Most samples and sequences were obtained newly for this study. Some ITS sequences were taken from Pirie et al. (2011) and further sequences were taken from van der Niet et al. (2014). Samples were obtained largely from field collections, but some were provided by the Bundesgarten-Belvedere Vienna in Austria, Botanic Gardens of the Rheinische Friedrich Wilhelms Universität Bonn in Germany, Gartenbauzentrum Straelen in Germany, and from private collections of known wild origin. Plant material was dried in silica gel and voucher specimens were deposited in herbaria (Appendix A).

Fig. 1. Samples. Collection sites of the studied populations in the northern area. and: *Erica andevalensis* arb: *E. arborea*, aus: *E. australis*, azo: *E. azorica*, cal: *Calluna vulgaris*, car: *E. carnea*, cil: *E. ciliaris*, cin: *E. cinerea*, dab: *Daboecia cantabrica*, emp: *Empetrum nigrum*, eri: *E. erigena*, lus: *E. lusitanica*, mac: *E. mackayana*, mad: *E. maderensis*, man: *E. manipuliflora*, mul: *E. multiflora*, pla: *E. platycodon*, sco: *E. scoparia*, sic: *E. sicula*, spi: *E. spiculifolia*, stu: *E. × stuartii*, ter: *E. terminalis*, tet: *E. tetralix*, umb: *E. umbellata*, vag: *E. vagans*. Tropical East African and South African populations are not shown.



2.2. Laboratory protocols

Two different lab protocols were employed: 1) Direct amplification (without DNA isolation) was performed using the method of Bellstedt et al. (2010); and 2) DNA isolation, (followed by separate PCR) was performed using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). In both cases leaf material was ground using a Qiagen Tissuelyser (Retsch GmbH & co., Haan, Germany), respectively: 1) at room temperature in grinding buffer; and 2) dry, having previously been frozen at -80°C for 24 hours. PCR was performed using standard protocols (Appendix B). Plastid regions *atpI-atpH* spacer, *trnK-matK* intron and *matK* gene, *psbM-trnH* spacer, *rbcL* gene, *rpl16* intron, *trnL-rpl32* spacer, *trnT-trnL-trnF-ndhJ* (including genes and intervening intron and spacer regions) and nuclear ribosomal ITS regions (including partial flanking 18S and 26S genes) were amplified and sequenced with primers listed in Appendix B. A targeted supermatrix strategy was employed (Pirie et al., 2008; Wiens, 2006), whereby more variable ITS and *trnL-trnF* spacer sequences were obtained for most samples, and the other, mostly less variable chloroplast markers were added for selected taxa in order to improve resolution of deeper nodes in the chloroplast tree (Appendix A). Sequences in general, and particularly ITS, were inspected for polymorphisms and apparent loss of function (e.g. indels in coding regions) that might be evidence of paralogy. In the absence of these phenomena, and in view of the minimal intra-individual ITS polymorphism across *Erica* and the largely consistent results obtained from cloning polymorphic amplicons reported in Pirie et al. (2011), we did not attempt to discover further ITS copies through cloning. For one sample of *E. x stuartii* the two overlapping traces in the ITS chromatograms were phased (by Sequiserve GmbH, Vaterstetten, Germany using 'sequencing analysis software'; Applied Biosystems) to obtain two different, unambiguous, sequence reads.

2.3. Alignment; matrix construction; model testing

Sequences were aligned by hand in Mesquite and matrices of individual markers were imported into SequenceMatrix (Vaidya et al., 2011) which was used to export concatenated matrices for analysis. The best fitting data partitioning strategies for those partitions (given models implemented in RAxML, BEAST, and MrBayes) were selected with PartitionFinder (Lanfear et al., 2012), using a heuristic search strategy ('greedy') and comparison of fit by means of the Bayesian information criterion. Individual markers, coding and non-coding regions within those markers, and codon positions within protein coding genes were all specified as potential data partitions (see results).

2.4. Phylogenetic inference

To test for experimental error and to infer and compare gene trees we performed preliminary phylogenetic analyses of individual markers separately and subsequently combined, partitioned analyses, including bootstrapping, under maximum likelihood using RAxML on CIPRES (Stamatakis, 2006; Stamatakis et al., 2008). Bootstrapping was halted automatically by RAxML following the majority-rule 'autoMRE' criterion and BS was presented on the best scoring ML tree. To assess the results for sensitivity to model assumptions, particularly given the non-random distribution of missing data in the matrix (Simmons, 2012), parsimony inference was performed using PAUP* (Swofford, 2003). To find the shortest trees, we employed a heuristic search strategy of 1,000 random addition sequences with TBR branch swapping and saving a maximum of 50 trees in each replicate. To assess clade support, we performed both bootstrapping and jackknifing, the former with 10,000 replicates with a single random addition sequence and TBR, saving a single tree each replicate following Müller (2005); the latter employing a jackknife proportion of 50%, with 1,000 replicates of 10 random addition sequences and TBR each, saving 10 trees each replicate following the general recommendations for supermatrix analyses of Simmons & Freudenstein (2011). Conflict between ITS and plastid trees was assessed by comparing nodes subject to 70% or higher bootstrap/jackknife support (BS/JS), illustrated by means of a tanglegram of 70% BS consensus trees using Dendroscope 3 (Huson and Scornavacca, 2012). Where such gene tree conflict was identified,

the taxa with conflicting phylogenetic signals were duplicated in a combined matrix following the approach of Pirie et al. (2008; 2009) in order to infer a single multi-labelled 'taxon duplication' tree using RAxML and PAUP*, as above, and MrBayes 3.2 (Ronquist et al., 2012) on CIPRES. For MrBayes analyses, the data were partitioned following PartitionFinder and two independent runs of 20 million generations each performed, sampling every 2,000 generations. Convergence and adequate sampling of runs was assessed using AWTY (Wilgenbusch et al., 2004) and Tracer version 1.5 (Rambaut and Drummond, 2003), and the post-burnin tree samples were summarised using the sumt command in MrBayes.

2.5. Ancestral state optimisations

To test whether alternative phylogenetic hypotheses for *E. lusitanica* would suggest independent origins of putatively homologous morphological characters, and thus suggest the 'capture' of a molecular marker by hybridisation, we used parsimony optimisations following de Villiers et al. (2013). To represent each phylogenetic hypothesis independently, we pruned *E. lusitanica* chloroplast and ITS taxa in turn from the parsimony strict consensus multi-labelled tree (in which the conflicting taxa were represented twice). The tree was further pruned to represent each species with a single terminal, also summarising the unresolved *E. scoparia*, *E. platycodon* and *E. azorica* complex as a single species. Given previous work (Fagúndez, 2006) and current results we accepted the hybrid status of *E.* x *stuartii* without further phylogenetic tests; it was also removed. We assumed that the radiation of the large clade that includes all exclusively Southern Hemisphere *Erica* species (Pirie et al., 2011) would have little or no impact on these analyses and removed all but *E. trimera* as representative of this clade, without attempting to infer its ancestral states.

We analysed 32 characters based on an ongoing study of the morphology, anatomy, palynology, phytochemistry and distribution of the European *Erica* species (Fagúndez et al in prep.; details presented in Appendix C). These characters were coded into unordered discrete states, either binary

or multistate and optimised over the trees under Fitch parsimony (Fitch, 1971) using Mesquite (Maddison and Maddison, 2009).

2.6. Coalescent simulations

To test whether gene tree differences could theoretically be explained by coalescent stochasticity we used a coalescent simulation approach in Mesquite, summarising the results in PAUP*. This approach assumes known clade ages (in the form of time calibrated trees), generation time and effective population size (the latter two assumed to be constant), and that there is free gene flow within populations (panmixis).

To obtain trees with branch-lengths proportional to time we used RELTIME (Tamura et al., 2012) to rate-smooth the multi-labelled ML tree from RAxML. We assumed local clocks and a root node age of 70 Mya, representing the mean estimate for the age of the genus derived from an Ericaceae-wide molecular dating analysis of Popp et al. (2011) that employed multiple fossil calibrations resulting in a range of 78-62 Mya for the most recent common ancestor of *Erica* and *Empetrum* (both Ericaceae subfamily Ericoideae). By pruning taxa from this ultrametric tree we then obtained two competing hypothetical "species trees" (each 'northern' species plus *E. trimera* represented as single taxa, as for the parsimony optimisations above) and two corresponding gene trees (all accessions of these species retained in the tree; the conflicting chloroplast and ITS sequences of *E.* x *stuartii* accessions were assigned to the appropriate parental species, *E. tetralix* and *E. mackayana*), representing two alternative phylogenetic hypotheses for *E. lusitanica*.

Generation time in *Erica* is difficult to assess, particularly given the ecology of the species and their adaptations to fire, reflected in markedly different life spans for seeders and resprouters (Ojeda et al., 1998; Segarra-Moragues and Ojeda, 2010). The time from germination to first flowering in *Erica* species can be as short as two years (*E. umbellata*, pers. obs.; *E. cinerea*, Bannister, 1965; *E. tetralix*, Bannister, 1966; *E. ciliaris*, Rose et al., 1996), or 4-6 years for *E. lusitanica* (Mather and Williams,

1990), whilst life spans have been reported as 19 years (*E. cinerea*, *E. tetralix*), 15-20 years (*E. ciliaris*) or 20 years (*Calluna*; Beijerinck, 1940). However, the latter are likely to be considerable underestimates, especially in species such as *E. lusitanica* and *E. arborea* that re-sprout after fire (Nelson, 2012; Silva et al., 2002). Growth rings measured in *E. carnea* provided a maximum age of 82 years, and 40 years for *Calluna* (Schweingruber and Poschlod, 2005), although a high variation can be found depending on the maturity stage of the community. In addition, seeds may remain viable in the seedbank of mature heathlands for over 100 years (Cumming and Legg, 1995; Thompson and Band, 1997), and following fire/disturbance, seeds of different ages germinate simultaneously. We therefore assumed two extremes for generation times: 2 (minimum) and 20 years.

The 'species' and gene trees were used to simulate samples of 1,000 gene trees using the "Coalescence Contained within Current Tree" module of Mesquite, further assuming constant effective population sizes (Ne) ranging from 10,000 to 200,000. We then calculated the differences between the two 'species trees' and the distributions of differences between each species tree and its corresponding simulated gene trees, using the partition metric (Penny and Hendy, 1985; implemented in PAUP* as the symmetric distance). Following Maureira-Butler et al. (2008), we would conclude that lineage sorting alone is unlikely to explain the difference between the two hypothetical species trees when this difference is higher than 95% of the difference simulated under coalescence. In order to explore further the implications of Ne for overall gene tree congruence we performed additional simulations ranging up to an Ne of 5,000,000 assuming generation time of 2 years. We calculated in each case the proportions of simulated trees in which nodes a) supported by both datasets independently; and b) defining the topological conflict in *E. lusitanica*, were recovered.

2.7. Species tree inference

To infer species trees for northern Ericeae assuming either that exclusively coalescence or that coalescence and reticulate evolution can explain gene tree differences we used *BEAST (Drummond and Rambaut, 2007) version 1.7.5 on CIPRES. Preliminary analyses employing the best fitting partitioning strategy and models inferred using PartitionFinder with the full sequence matrix and a lognormal relaxed clock model persistently failed to converge. The *BEAST analyses were therefore performed with simplified models and an alignment excluding chloroplast *rpl16*, *rpl32*, *atpI-atpH*, *psbM-trnH* and *rbcL* (representing the least informative markers and greatest proportion of missing data). The data was partitioned according to plastid and ITS data, with unlinked trees, substitution models (GTR+G; empirical base frequencies) and strict clock models. The age of the root node was calibrated in numbers of generations assuming two extremes of the uncertainty surrounding nodes ages (the confidence interval estimated by Popp et al., 2011) and generation time (as above): 39 and 3.1 million generations (corresponding to 78 Mya and generation time of 2 years and 62 Mya with generation time of 20 years, respectively). Calibrations were applied with normal distributed priors with minimal bounds. The ingroup (Ericeae; to the exclusion of *Empetrum*) was constrained to be monophyletic in order to root the tree. The same samples and species definitions were used as for the parsimony optimisations and coalescent simulations, except that *E. lusitanica* was treated either as a single species (standard species tree, coalescence hypothesis; “C”) or as two species represented by plastid and ITS trees independently (multi-labelled species tree, coalescence plus reticulate evolution hypothesis; “C+R”). We assumed a Yule process species tree demographic model. Default values were retained for other priors. Two independent MCMC runs of 100 million generations each were performed for each hypothesis. Convergence and adequate sampling of runs was assessed using AWTY (Wilgenbusch et al., 2004) and Tracer version 1.5 (Rambaut and Drummond, 2003). The post-burnin tree samples were summarised using LogCombiner and TreeAnnotator (Drummond et al., 2012).

The multi-labelled "C+R" species tree was summarised as a rooted network using the MUL to Network, Holm 2006 method in Dendroscope 3 (Huson and Scornavacca, 2012).

3. Results

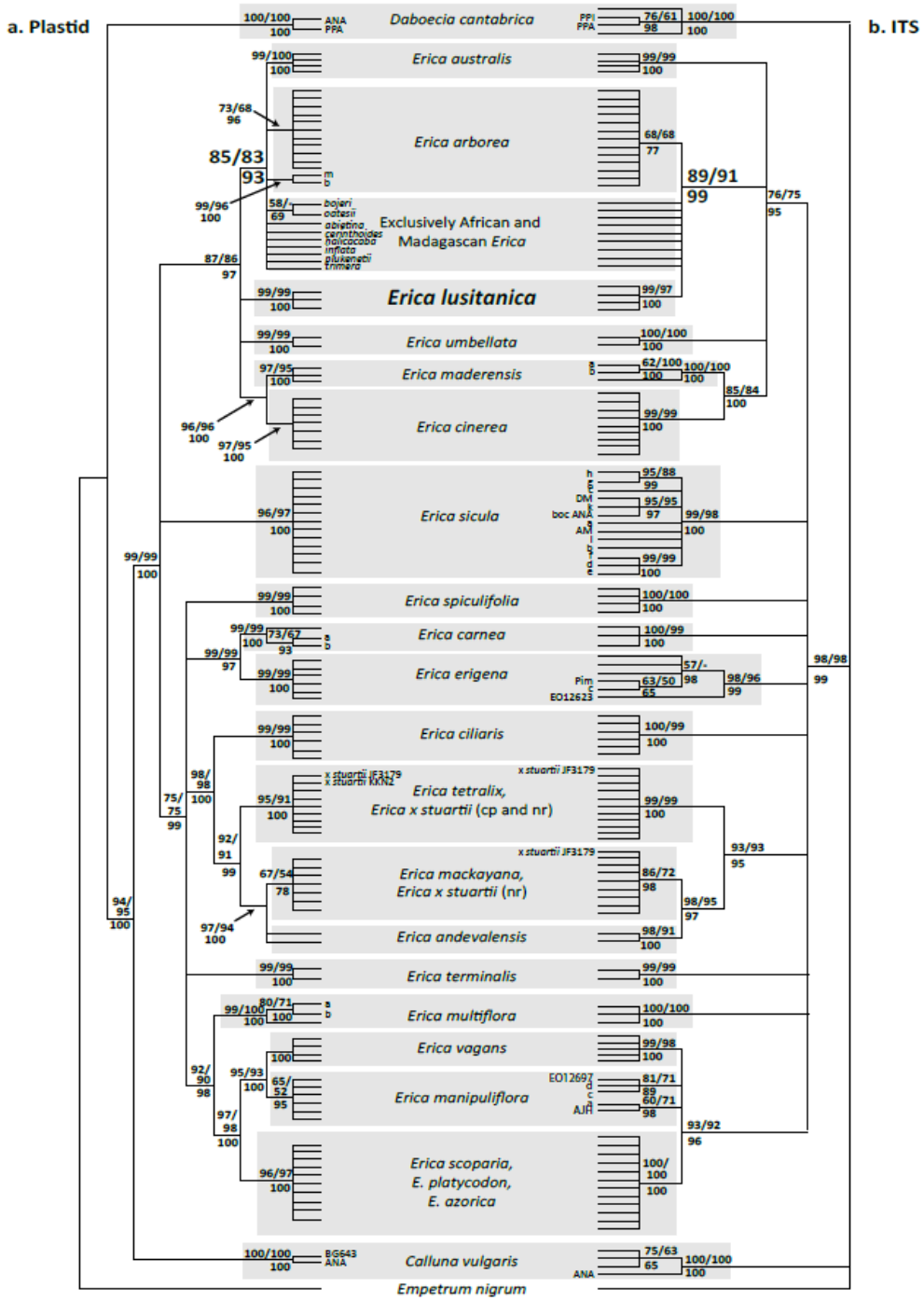
3.1. Phylogenetic inference

Trees inferred from individual chloroplast markers did not show supported topological conflict (data not shown), therefore the data were combined under the assumption of a single bifurcating chloroplast tree. For ITS, the total alignment length was 974, of which 31 positions were variable but parsimony uninformative and 271 parsimony informative (33 and 183 respectively, for *Erica* alone); CI=0.85, RI=0.89. For the concatenated plastid regions, the total alignment length was 10222 positions, 988 variable uninformative, 830 informative (507 and 557 respectively, for *Erica* alone); CI=0.59, RI=0.92. Following the targeted sampling strategy, a large proportion of the latter positions were coded as missing data for most taxa.

Twenty-four potential data partitions were assessed using PartitionFinder, which identified optimal partitioning strategies for RAxML analyses (employing GTR+G only) including just four partitions for the entire dataset and for BEAST/MrBayes analyses (employing models of differing complexity) including five partitions (Table 1).

Gene trees inferred from the chloroplast data and ITS respectively are represented in Fig. 2 (a tanglegram of plastid/ITS 70% ML BS consensus trees, presented using Dendroscope with modification by hand; parsimony BS/JK proportions indicated). Comparison of topology and BS/JS shows supported conflict involving *E. lusitanica* and the phased ITS sequences of *E. x stuartii* (Fig. 2).

Fig. 2. Gene trees. Tanglegram of 70% BS/JK consensus trees for a. plastid and b. ITS sequence data separately, inferred under parsimony and maximum likelihood, with parsimony BS/JK values above the nodes; ML BS below. Node support corresponding to conflict in *E. lusitanica* is indicated with larger type. Specific accessions are indicated where there are clades within species, otherwise terminals are summarised to species.

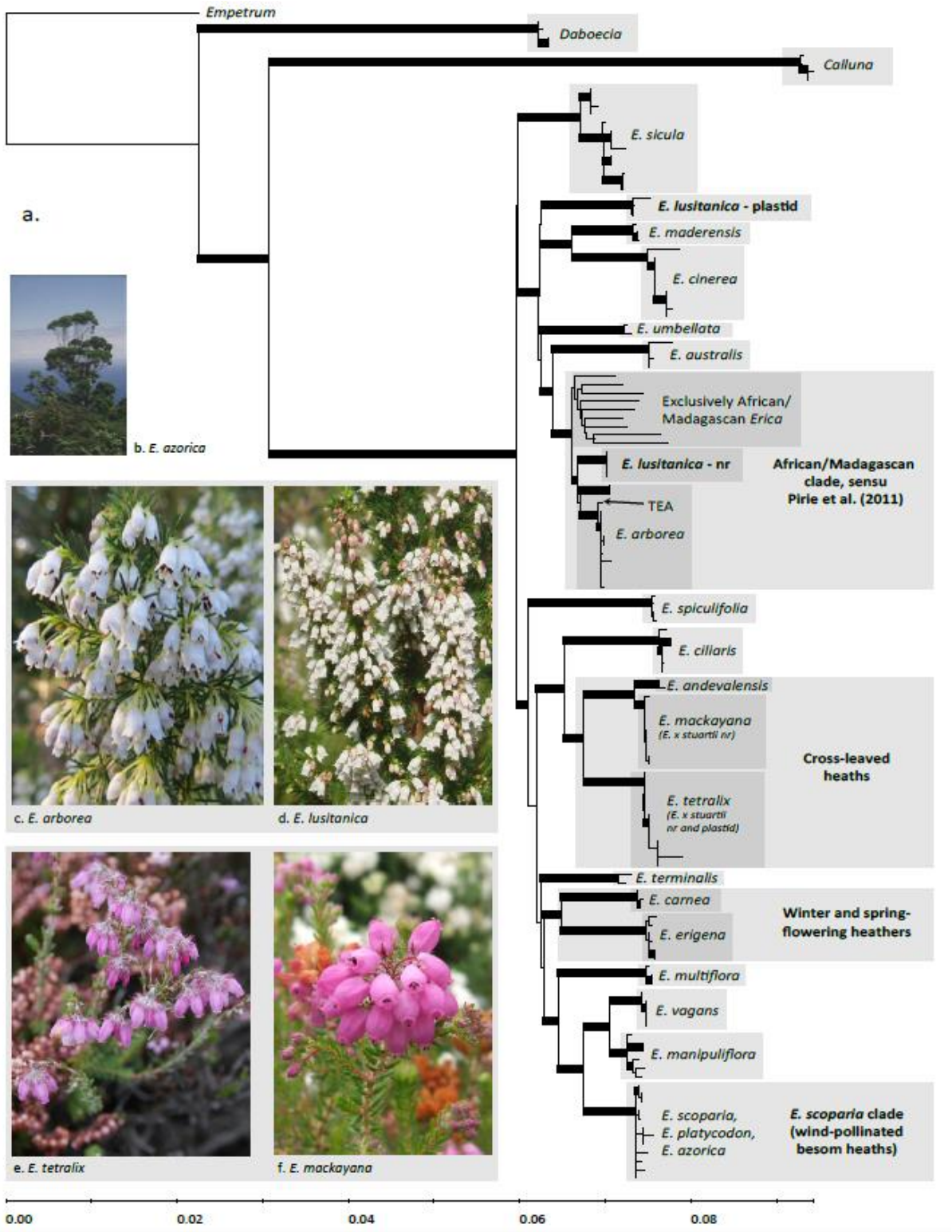


Combined analysis of all data using the taxon duplication approach resulted in a multi-labelled tree representing these incidences of conflict. The best tree from RAxML analysis, visualised using TRED (http://www.reelab.net/tred) is presented in Fig. 3, with parsimony BS/JK (PAUP*) and posterior probability (PP; MrBayes) clade support presented in Appendix D. The multi-labelled tree showed a clear increase in numbers of nodes subject to support ≥70% BS/JK compared to either gene tree separately, although the improvement relative to the better resolved plastid tree was mostly limited to within species (Figs. 2 and 3).

3.2. Character optimisations

Of the parsimony optimisations of 32 binary or multistate coded characters, the minority (6) showed only a single step, with the mean number of steps being 3.1/2.8 (plastid/ITS hypotheses); maximum 6/6; and total 98/89. None of the characters showed a greater number of steps given the ITS hypothesis for *E. lusitanica* compared to the plastid hypothesis, whereas 9 showed an additional step given the chloroplast hypothesis.

Fig. 3. Multi-labelled "taxon duplication" phylogeny. a. RAxML phylogram of the combined chloroplast and ITS data with *E. lusitanica* duplicated to represent the conflicting signal of the markers (see text). ML bootstrap (BS) proportions ≥70% are represented as thicker branches (presented using TRED). Detailed terminal names, BS values and posterior probability clade support are presented on an identical topology in Appendix D. Further clades referred to in the text are indicated. The scale bar represents substitutions per site. b-f. Photos of selected species referred to in the text, by courtesy of The Heather Society; © Barry Sellers (e); © Charles Nelson (b, c, d, f)


Empetrum
Daboecia
Calluna
E. sicula
a.
E. lusitanica - plastid
E. maderensis
E. cinerea
E. umbellata
E. australis
Exclusively African/
Madagascan Erica
E. lusitanica - nr
African/Madagascan
clade, sensu
Pirie et al. (2011)
TEA
E. arborea
b. E. azorica
E. spiculifolia
E. ciliaris
E. andevalensis
E. mackayana
(E. x stuartii nr)
Cross-leaved
heaths
E. tetralix
(E. x stuartii
nr and plastid)
c. E. arborea
d. E. lusitanica
E. terminalis
E. carnea
Winter and spring-
flowering heathers
E. erigena
E. multiflora
E. vagans
E. manipuliflora
E. scoparia,
E. platycodon,
E. azorica
E. scoparia clade
(wind-pollinated
besom heaths)
e. E. tetralix
f. E. mackayana
0.00
0.02
0.04
0.06
0.08

3.3. Coalescent simulations

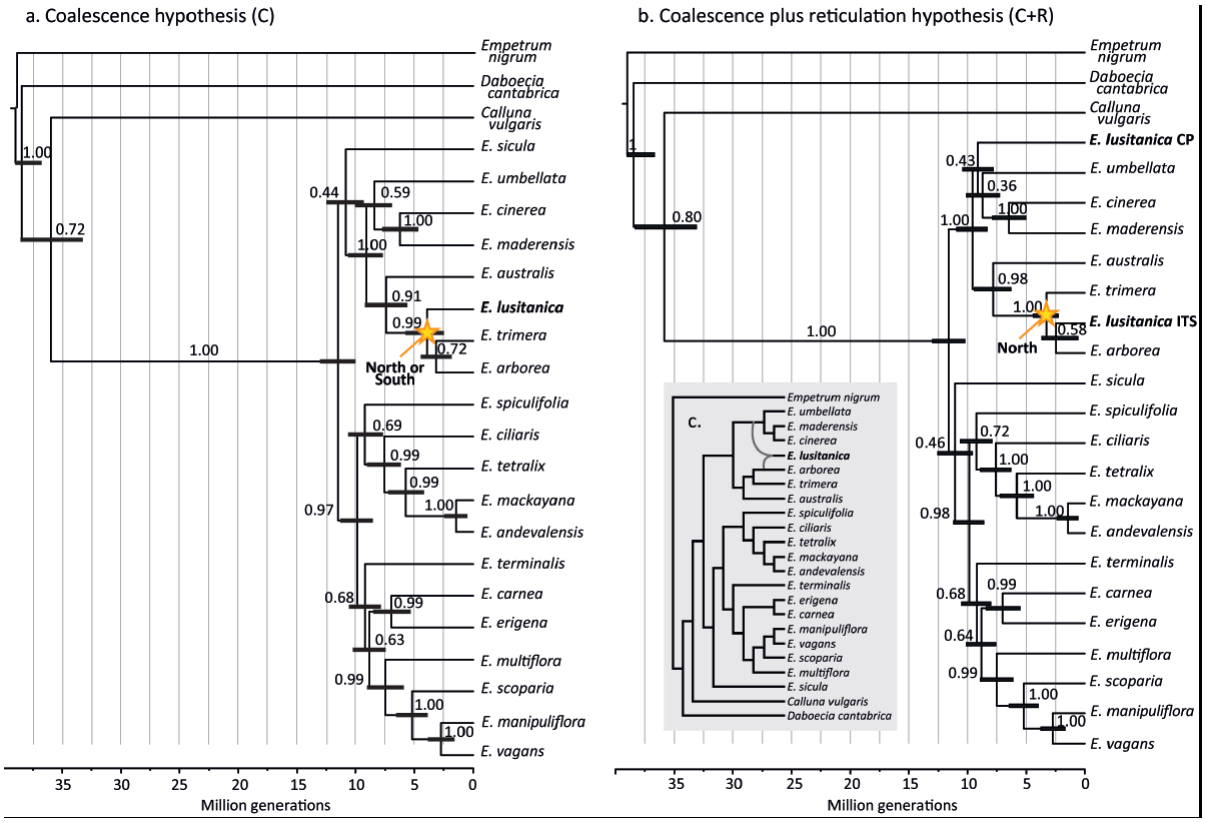
The symmetric distance between chloroplast and ITS 'species trees' was 6; the symmetric distances between these trees and the trees simulated from them under coalescence (up to Ne=200,000) are reported in Table 2. Values of Ne and generation time at which the distance between species trees was not significantly different from the distance between species and simulated trees were 20,000/20 years and 200,000/2 years. Further trees were simulated assuming 2 year generation time and values for Ne up to 5,000,000. We assessed these trees for the presence of three nodes that receive strong support in both individual gene trees and two further nodes that define the topological conflict in *E. lusitanica*. These nodes and the proportions of trees in which they were recovered are reported in Table 3. At Ne=200,000 all nodes were recovered in >99% of trees, whereas at Ne=1,000,000 all except the ITS *E. arborea*/*E. lusitanica* node were recovered in <95% of trees.

3.4. Species tree inference

Coalescence ('C') and coalescence plus reticulate evolution ('C+R') species tree hypotheses (the latter also summarised as a rooted phylogenetic network) are illustrated in Fig. 4. Topologies were identical, and support almost so, irrespective of the assumed age of the root node. Posterior probability clade support was generally lower than that inferred using MrBayes, including <0.95 PP support for the relationships between *E. lusitanica*, *E. arborea* and *E. trimera*. However, these three species form a clade (contradicting the plastid gene tree) with PP=0.99 and 1.00 respectively. The maximum clade credibility (MCC) 'C+R' species tree shows topology consistent with the standard multilabelled tree, with *E. lusitanica* ITS sister to *E. arborea* (PP=0.59 [root age of 3.1 million generations]; PP=0.58 [39 million generations]) and *E. lusitanica* plastid uncertain but more closely related to other European species; it is also represented as a rooted network (Fig. 4c). The MCC 'C'

species tree, by contrast, differs from both independent gene trees in showing *E. lusitanica* as sister to *E. arborea* plus *E. trimera*, albeit also poorly supported (PP=0.72).

Fig. 4. Species trees/network. Maximum clade credibility species trees inferred using *BEAST, with clade posterior probabilities presented next to nodes; a. the coalescence only hypothesis ('C'); b. the coalescence plus reticulation hypothesis ('C+R'). The ancestral areas ('north' versus 'south') for nodes referred to in the text are indicated with stars. c. (insert) the 'C+R' hypothesis represented as a rooted network summarised using Dendroscope; a reticulation in *E. lusitanica* is denoted with grey branches.



4. Discussion

4.1. A robust and representative phylogenetic hypothesis for northern Ericeae

The phylogenetic results presented here represent the first analysis including all northern species of *Erica* and a considerable improvement in resolution compared to previous work (McGuire and Kron, 2005; Pirie et al., 2011). Our results indicate the monophyly of most currently accepted species. A notable exception is the *E. scoparia* clade, or wind-pollinated 'besom heaths' (Nelson, 2012), formed by the south-western Mediterranean *E. scoparia* and its relatives of the Macaronesian islands (*E. azorica*, Fig. 3b; and *E. platycodon*). Species delimitation in this group is still the subject of debate and low sequence divergence between the samples analysed here leaves this situation effectively unresolved. However, the combined morphological and molecular similarity of currently delimited taxa, which is in contrast particularly to the greater sequence divergence of the other species, provides circumstantial evidence in favour of treating this clade as a single (complex) species, as did Désamoré *et al.* (2012).

A number of species groupings that were proposed in previous work (Bayer, 1993; Bentham, 1839; Hansen, 1950; Nelson, 2012) are apparent in the trees presented here. For example, *E. tetralix* (Fig. 3e), *E. mackayana* (Fig. 3f) and *E. andevalensis* (the 'cross-leaved heaths', Nelson, 2012); and the morphologically similar *E. carnea* and *E. erigena* (sect. Callicodon, Bentham, 1839; 'winter- and spring-flowering heathers', Nelson, 2012), are monophyletic. Phylogenetic relationships of some species remain unresolved, particularly those of *E. terminalis* and *E. spiculifolia*. This uncertainty is the result of low sequence divergence apparent at the base of the tree, meaning that the major lineages within *Erica* probably diverged in relatively quick succession. Increased character sampling, for example of plastid markers, might yet resolve the remaining basal uncertainty. However, with shorter intervals between speciation events coalescent stochasticity can cause larger differences between gene trees. Hence, to address this uncertainty meaningfully it would be important to

obtain and compare two or more independent, resolved, gene trees. Beyond the lack of basal support, the two independent gene trees presented here yielded generally consistent results, and the combined phylogenetic hypothesis showed increased support both for relationships overall, and for the gene tree conflict shown in *E. lusitanica* (Fig. 3a, d).

4.2. Reticulate or parallel evolution of European tree heathers?

The standard interpretation of our coalescent simulation results would be that a degree of gene tree conflict comparable to that observed in *E. lusitanica* could be explained by coalescence given constant Ne of 200,000 or greater with a 2 year generation time, or 20,000 or greater with a 20 year generation time. These values appear modest compared to current population sizes: most northern species of Ericeae are dominant in their habitats, growing in large populations. However, further scrutiny of the simulated trees suggests that there would only be an appreciable chance of any individual gene tree contradicting the positions of *E. lusitanica* when Ne is considerably higher than these values. Furthermore, under these conditions, more extensive topological conflict between gene trees would be expected (Table 3), which is not evident here. These results imply that the standard interpretation of coalescent simulations represents a bias against rejecting coalescence, particularly when gene tree conflict is limited, and that in this case despite the overall uncertainty the coalescence scenario may be less likely than e.g. a single ancient hybridisation event.

Similar bias in such tests is already recognised: given inevitable uncertainty in estimates of ages, generation times, and (effective) sizes of past populations, gene tree conflict in *E. lusitanica* could still fall within the 'coalescent stochasticity zone', in which coalescent stochasticity effectively cannot be rejected (de Villiers et al., 2013). Furthermore, hybridisation can, and often may, take place between closely related species (Abbott et al., 2013) irrespective of our means to prove it using coalescent simulations.

We used an approach suggested by de Villiers et al. (2013) for identifying putative hybrids, based on the assumption that inheritance of morphological characteristics is generally parsimonious (independent origins, particularly of complex structures, is in principle unlikely) and that less parsimonious scenarios implied by a given gene tree but contradicted by another are evidence for reticulate evolution (i.e. hybridisation). Floral morphology seems to be evolutionarily labile in *Erica* overall (Pirie et al., 2011) and might be expected to shift irrespective of linear or reticulate inheritance. Many of the characters that we analysed were vegetative, but these too appeared to have evolved in parallel or have been lost secondarily, resulting in multiple steps across the tree. However, all characters that showed a difference in number of steps between gene trees (which differ only in the position of *E. lusitanica*) – that is, almost a third of the total – were less parsimonious in the plastid tree. This includes parallel evolution of tall habit and presence of a lignotuber that, following the ITS topology, would be regarded as synapomorphies of *E. lusitanica* and *E. arborea* (albeit independently evolved in *E. scoparia* and *E. australis*). Taken together with the lack of evidence for wider gene tree conflict, we interpret this as evidence for reticulate evolution, potentially 'chloroplast capture', resulting from hybridisation.

Further tools for assessing putative hybrids include co-dominant markers, such as microsatellites, as used by Segarra-Moragues and Ojeda (2010); or single copy nuclear genes, such as at103 used by Desamore et al. (2011), for which alleles from different parents might be identified. This approach is more likely to be effective at identifying recent, rather than more ancient, hybridisation and gene flow. Polymorphism in ITS can also be used as evidence for hybrids and the identity of their parents. The absence of polymorphism in *E. lusitanica* is consistent with the timing of divergence in the gene trees inferred here, suggesting a relatively ancient hybridisation event followed by concerted evolution with no evidence for subsequent gene flow. Additional independent nuclear markers will be needed to test and further characterise this scenario.

4.3. Assuming reticulate evolution versus exclusively coalescence influences the results of species tree inference

The multi-labelled phylogeny presented in Fig. 3 represents the conflict between gene trees with regard *E. lusitanica* under the assumption that the data is otherwise congruent (Pirie et al., 2009). As an alternative approach, we employed a coalescence based species tree method implemented in *BEAST (Drummond and Rambaut, 2007) given two theoretical scenarios: 1) assuming gene tree differences to be the result of coalescent stochasticity (the 'C' hypothesis); and 2) specifying *E. lusitanica* as reticulate (the 'C + R' hypothesis), using the "taxon duplication" approach (Pirie et al., 2008; Pirie et al., 2009) in a coalescence framework (Blanco-Pastor et al., 2012; Pimentel et al., 2013). The latter approach assumes that in a priori identified hybrid or otherwise reticulate taxa, specified markers have been inherited from distinct, potentially unknown, parental lineages. Within those lineages, a coalescence process is still assumed.

Similar to the coalescent simulation approach, the species tree analyses had to incorporate considerable further uncertainty in assumptions, particularly the combined age and generation time needed to calibrate the tree. However, these sources of uncertainty did not appear to result in differences between species trees or support values, despite the two extremes of potential age/generation times (3.1 compared to 39 million generations) differing by one order of magnitude. Differences were instead apparent between analyses invoking the 'C' versus 'C+R' hypotheses, i.e. depending on our assumptions regarding gene tree conflict in *E. lusitanica*. The 'C+R' hypothesis multi-labelled species tree derived using *BEAST (Fig. [4]) was consistent with the multi-labelled tree obtained with concatenation of non-conflicting data (Fig. [3]), whereas the 'C' hypothesis delivered a topology that differed both from this and from both gene trees individually, albeit with low support.

Support values for the *BEAST species trees were generally lower than those inferred under the standard phylogenetic approach. This might be partly explained by the reduction in sequence data

and model complexity that was necessary in order to achieve convergence with this dataset. However, it can also be expected from a method in which gene trees are independent and can differ, compared to one in which they are the sum of the evidence for a single underlying tree. Given the consistent species monophyly, apparent insensitivity of results to the assumed age of the group, and, with the exception of *E. lusitanica* and the recent hybrid *E. x stuartii*, the notable lack of gene tree incongruence, it is arguable that the complex coalescence-based model may not be necessary for inferring a (reticulate) species tree from this dataset. Irrespective of this opinion, it is apparent from these results that assumptions regarding the processes underlying gene tree conflict, whilst difficult to justify, are nevertheless important, even (or particularly) in a coalescent framework in which such conflict is integrated into the model.

4.4. The biogeographic implications of coalescence versus hybridisation in *E. lusitanica*

The conclusion of McGuire & Kron (2005) of a northern ancestral area for *Erica* is supported here by the paraphyly of the northern Ericeae, in which the African/Madagascan clade is deeply nested. They further interpreted the widespread *E. arborea* as representing the first in a stepwise southwards expansion of the genus. Our plastid and *BEAST 'C' hypothesis trees place *E. lusitanica* (not included in the analysis of McGuire & Kron, 2005) within the European basal grade. The results of parsimony optimisation of geographic areas ('northern' versus 'southern') over these trees are nevertheless equivocal with regard the ancestral area of *E. arborea* and the rest of the African/Madagascan clade (the latter represented by *E. trimera*; Fig. 4). However, our (single) sample of *E. arborea* from TEA is nested within a clade of northern *E. arborea* samples with a relatively ancient common ancestor (as indicated by sequence divergence in Fig. 3), and the ITS and *BEAST 'C+R' hypothesis trees indicate *E. lusitanica* (which is northern) as sister to this clade. Parsimony optimisations over the latter trees imply independent colonisations of Tropical East Africa (TEA) from Europe by *E. arborea* and by the rest of the African/Madagascan clade (Fig. 4).

An independent origin of *E. arborea* in TEA is still consistent with the scenario of Désamoré et al. (2010) who, on the basis of much denser population-level sampling inferred a Pleistocene (hence likely more recent) migration from TEA to the Mediterranean. The general pattern and processes underlying the *Erica* expansion and radiation in Africa requires testing with representative sampling of the species-rich African/Madagascan clade (Pirie et al., ms. in prep.). An implication of both the results of Désamoré et al. (2010) and those presented here is that North-South dispersal events may have occurred more often than previously was assumed.

Acknowledgements—The authors gratefully thank Mr. Kurt Kramer, Ms. Bettina Banse, the plant breeders and producers of the Azerca group, Mr. Michael Knaack and his team from Belvedere Garten in Vienna, the gardener's team of the Botanical Garden in Bonn, and the Gartenbauzentrum team in Straelen for spending their valuable time and making available plant material for this study. Further material was kindly provided by Charles Nelson, various members of The Heather Society, and Fernando Ojeda. For financial support we thank the Ministerium für Klimaschutz, Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des Landes Nordrhein-Westfalen, the Faculty of Agriculture Lehr- und Forschungsschwerpunkt „Umweltverträgliche und Standortgerechte Landwirtschaft", Bonn University, and the Landgard foundation. This work was further supported by a postdoctoral fellowship from the Claude Leon Foundation to MDP and is based on research partially funded by the South African National Research Foundation (DUB and MDP). Any opinion, finding and conclusion or recommendation expressed in this material is that of the authors and the NRF does not accept liability in this regard.

Appendices

Appendix A: Accessions details for samples analysed, including Genbank accession numbers.

Appendix B: Lab protocols.

Appendix C: Characters and character states included in the character optimisation analysis (from Fagúndez et al in prep.).

Appendix D: RAxML tree resulting from the combined analysis of ITS and plastid data, using the taxon duplication approach with *E. lusitanica*. The topology and branch lengths are identical to that of Fig. 3, here including detailed terminal names and clade support values (format: RaxML/parsimony bootstrap support (BS) above; Baysian posterior probability/parsimony jackknife proportion below).

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, C., Jones, J., Keller, B., Marczewski, T., Mallet, J., Martinez-Rodriguez, P., Möst, M., Mullen, S., Nichols, R., Nolte, A.W., Parisod, C., Pfennig, K., Rice, A.M., Ritchie, M.G., Seifert, B., Smadja, C.M., Stelkens, R., Szymura, J.M., Väinölä, R., Wolf, J.B.W., Zinner, D., 2013. Hybridization and speciation. J. Evol. Biol. 26, 229-246.

Bannister, P., 1965. *Erica cinerea* L. J. Ecol. 53, 527-542.

Bannister, P., 1966. *Erica tetralix* L. J. Ecol. 54, 795-813.

Bayer, E., 1993. Erica L. (Ericaceae). In: Castroviejo, S. (Ed.), Flora Ibérica IV, Madrid, pp. 485-506.

Beatty, G.E., Provan, J., 2012. Post-glacial dispersal, rather than in situ glacial survival, best explains the disjunct distribution of the Lusitanian plant species *Daboecia cantabrica* (Ericaceae). J. Biogeogr. 40, 335-344.

Beijerinck, W., 1940. Calluna: A monograph on the Scotch heather. Verhandl. Kon. Ned. Akad. Wetensch. Nat. 38.

Bellstedt, D.U., Pirie, M.D., Visser, J.C., de Villiers, M.J., Gehrke, B., 2010. A rapid and inexpensive method for the direct PCR amplification of DNA from plants. Amer. J. Bot. 97, e65-68.

Bentham, G., 1839. Ericaceae. In: De Candolle, A.P. (Ed.), Prodromus systematis naturalis regni vegetabilis. Treutel & Weiss, Paris, pp. 580-733.

Blanco-Pastor, J.L., Vargas, P., Pfeil, B.E., 2012. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. PLoS ONE 7, e39089.

Cumming, G., Legg, C., 1995. Longevity of the *Calluna vulgaris* seed bank determined from history of lead smelting at Leadhills and Wanlockhead, Scotland. In: Thompson, D.B.A., Hester, A.J., Usher, M.B. (Eds.), Heaths and Moorland: Cultural Landscapes. HMSO, Edinburgh, pp. 135-139.

de Villiers, M.J., Pirie, M.D., Hughes, M., Möller, M., Edwards, T., Bellstedt, D.U., 2013. An approach to identify putative hybrids in the 'coalescent stochasticity zone', as exemplified in the African plant genus *Streptocarpus* (Gesneriaceae). New Phytol. 198, 284-300.

Désamoré, A., Laenen, B., Devos, N., Popp, M., González-Mancebo, J.M., Carine, M.A., Vanderpoorten, A., 2010. Out of Africa: north-westwards Pleistocene expansions of the heather *Erica arborea*. J. Biogeogr. 38, 164-176.

Désamoré, A., Laenen, B., González-Mancebo, J.M., Jaén Molina, R., Bystriakova, N., Martinez-Klimova, E., Carine, M.A., Vanderpoorten, A., 2012. Inverted patterns of genetic diversity in continental and island populations of the heather *Erica scoparia* s.l. J. Biogeogr. 39, 574-584.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molec. Biol. Evol. 29, 1969-1973.

Fagúndez, J., 2006. Two wild hybrids of *Erica* L.(Ericaceae) from northwest Spain. Botanica Complutensis 30, 131-135.

Fagúndez, J., Izco, J., 2010. Seed morphology of the European species of *Erica* L. sect. Arsace Salisb. ex Benth. (Ericaceae). Acta Botanica Gallica 157, 45-54.

Fagúndez, J., Juan, R., Fernández, I., Pastor, J., Izco, J., 2010. Systematic relevance of seed coat anatomy in the European heathers (Ericeae, Ericaceae). Pl. Syst. Evol. 284, 65-76.

Fagúndez, J., 2012. Study of some European wild hybrids of *Erica* L. (Ericaceae), with descriptions of a new nothospecies: *Erica* × *nelsonii* Fagúndez and a new nothosubspecies: *Erica* × *veitchii* nothosubsp. *asturica* Fagúndez. Candollea 67, 51-57.

Fitch, W.M., 1971. Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. 20, 406-416.

Frodin, D.G., 2004. History and concepts of big plant genera. Taxon 53, 753-776.

Gil-López, M.J., Segarra-Moragues, J.G., Ojeda, F., 2014. Population Genetic Structure of a Sandstone Specialist and a Generalist Heath Species at Two Levels of Sandstone Patchiness across the Strait of Gibraltar. PLoS ONE 9, e98602.

Gizaw, A., Kebede, M., Nemomissa, S., Ehrich, D., Bekele, B., Mirré, V., Popp, M., Brochmann, C., 2013. Phylogeography of the heathers *Erica arborea* and *E. trimera* in the afro-alpine 'sky islands' inferred from AFLPs and plastid DNA sequences. Flora 208, 453-463.

Hansen, I., 1950. Die europäischen Arten der Gattung *Erica* L. Bot. Jahrb. 75, 1-81.

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Syst. Biol. 61, 1061-1067.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Molec. Biol. Evol. 29, 1695-1701.

Maddison, W.P., Maddison, D.R., 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72. Available from http://mesquiteproject.org.

Mather, L.J., Williams, P.A., 1990. Phenology, seed ecology, and age structure of Spanish health (*Erica lusitanica*) in Canterbury, New Zealand. N.Z. J. Bot. 28, 207-216.

Maureira-Butler, I.J., Pfeil, B.E., Muangprom, A., Osborn, T.C., Doyle, J.J., 2008. The Reticulate History of *Medicago* (Fabaceae). Syst. Biol. 57, 466-482.

McGuire, A.F., Kron, K.A., 2005. Phylogenetic relationships of European and African ericas. Int. J. Pl. Sci. 166, 311-318.

Müller, K., 2005. The efficiency of different search strategies in estimating parsimony jackknife, bootstrap, and Bremer support. BMC Evol. Biol. 5, 58.

Nelson, E.C., Oliver, E.G.H., 2005. Chromosome numbers in *Erica*-an updated checklist. Heathers 2, 57-58.

Nelson, E.C., 2012. Hardy heathers from the Northern Hemisphere. Royal Botanic Gardens, Kew, Richmond, United Kingdom.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358-364.

Ojeda, F., Arroyo, J., Maranon, T., 1998. The Phytogeography of European and Mediterranean Heath Species (Ericoideae, Ericaceae): A Quantitative Analysis. J. Biogeogr. 25, 165-178.

Ojeda, F., Brun, F.G., Vergara, J.J., 2005. Fire, rain and the selection of seeder and resprouter life-histories in fire-recruiting, woody plants. New Phytol. 168, 155-165.

Oliver, E.G.H., 2000. Systematics of Ericaceae (Ericeae-Ericoideae): species with indehiscent and partially dehiscent fruits. Contrib. Bolus Herb 19.

Oliver, E.G.H., Oliver, I., 2003. Ericaceae. In: Germishuizen, G., Meyer, N.L. (Eds.), Plants of southern Africa: an annotated checklist. Strelitzia 19. South African National Biodiversity Institute, Pretoria, pp. 424-451.

Oliver, E.G.H., Forshaw, N., 2012. Genus *Erica* An Identification Aid Version 3.00. Contributions to the Bolus Herbarium 22.

Penny, D., Hendy, M.D., 1985. The use of tree comparison metrics. Syst. Zool. 34, 75-82.

Pimentel, M., Sahuquillo, E., Torrecilla, Z., Popp, M., Catalán, P., Brochmann, C., 2013. Hybridization and long-distance colonization at different time scales: towards resolution of long-term controversies in the sweet vernal grasses (*Anthoxanthum*). Ann. Bot. 112, 1015-1030.

Pirie, M.D., Humphreys, A.M., Galley, C., Barker, N.P., Verboom, G.A., Orlovich, D., Draffin, S.J., Lloyd, K., Baeza, C.M., Negritto, M., Ruiz, E., Cota Sanchez, J.H., Reimer, E., Linder, H.P., 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. Mol. Phylogenet. Evol. 48, 1106-1119.

Pirie, M.D., Humphreys, A.M., Barker, N.P., Linder, H.P., 2009. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). Syst. Biol. 58, 612-628.

Pirie, M.D., Oliver, E.G.H., Bellstedt, D.U., 2011. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macro-morphology. Mol. Phylogenet. Evol. 61, 593-601.

Popp, M., Mirré, V., Brochmann, C., 2011. A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). Proc. Natl. Acad. Sci. USA.

Rambaut, A., Drummond, A.J., 2003. Tracer v. 1.5. Available from http://beast.bio.ed.ac.uk/Tracer.

Rebelo, A.G., Siegfried, W.R., Oliver, E.G.H., 1985. Pollination syndromes of *Erica* species in the south-western Cape. S. Afr. J. Bot. 51, 270-280.

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

Rose, R., 2007. The effects of hybridization on the small-scale variation in seed-bank composition of a rare plant species, *Erica ciliaris* L. Seed Sci. Res. 17, 201-210.

Rose, R.J., Bannister, P., Chapman, S.B., 1996. Biological Flora of the British Isles. *Erica ciliaris* L. J. Ecol. 84, 617-628.

Schweingruber, F.H., Poschlod, P., 2005. Growth rings in herbs and shrubs: life span, age determination and stem anatomy. Swiss Federal Research Institute WSL.

Segarra-Moragues, J.G., Ojeda, F., 2010. Postfire response and genetic diversity in *Erica coccinea*: connecting population dynamics and diversification in a biodiversity hotspot. Evolution 64, 3511-3524.

Silva, J.S., Rego, F.C., Martins-Loução, M.A., 2002. Belowground traits of Mediterranean woody plants in a Portuguese shrubland. Ecologia Mediterranea 28, 5-13.

Simmons, M.P., Freudenstein, J.V., 2011. Spurious 99% bootstrap and jackknife support for unsupported clades. Mol. Phylogenet. Evol. 61, 177-191.

Simmons, M.P., 2012. Radical instability and spurious branch support by likelihood when applied to matrices with non-random distributions of missing data. Mol. Phylogenet. Evol. 62, 472-484.

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

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Syst. Biol. 57, 758-771.

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4. Sinauer Associates, Sunderland, Mass.

Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipski, A., Kumar, S., 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. USA 109, 19333-19338.

Thompson, K., Band, S.R., 1997. Survival of a lowland heathland seed bank after a 33-year burial. Seed Sci. Res. 7, 409-411.

Vaidya, G., Lohman, D.J., Meier, R., 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27, 171-180.

Van der Niet, T., Pirie, M.D., Shuttleworth, A., Johnson, S.D., Midgley, J.J., 2014. Do pollinator distributions underlie the evolution of pollination ecotypes in the Cape shrub *Erica plukenetii*? Ann. Bot. 113, 301-316.

Wiens, J.J., 2006. Missing data and the design of phylogenetic analysis. J. Biomed. Inform. 39, 34-42.

Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2004. AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available from: http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php.

Tables

Table 1. Best fitting partitioning strategies and substitution models inferred using PartitionFinder for RAxML (GTR only) and BEAST/MrBayes (models as indicated). Partitions correspond to plastid unless specified as rDNA

Partitions (RAxML)	**Partitions (BEAST/MrBayes)**	**Model (BEAST/MrBayes)**
18S, 28S, and 5.8S (rDNA); *ndhJ* gene, *rbcL* codon positions 1 and 2, *trnF* gene, *trnL* gene	18S (rDNA); *rbcL* codon positions 1 and 2, *trnL* gene	K80+G (NST=1)
	28S, 5.8S (rDNA); *ndhJ* gene, *trnF* gene	TrNef+G (NST=2)
atpI-atpH, *matK* codon positions 1, 2 and 3, *psbM-trnH*, *rbcL* codon position 3, *rbcL* spacer, *rpl16* intron, *trnL-rpl32*, *trnK-matK* spacer, *trnL* intron	*atpI-atpH*, *matK* codon positions 1, 2 and 3, *psbM-trnH*, *rbcL* codon position 3, *rbcL* spacer, *rpl16* intron, *trnL-rpl32*, *trnK-matK* spacer, *trnL* intron	GTR+G (NST=6)
trnT-trnL, *trnL-trnF*, *trnF-ndhJ*	*trnT-trnL*, *trnL-trnF*, *trnF-ndhJ*	GTR+G (NST=6)
ITS1, ITS2 (rDNA)	ITS1, ITS2 (rDNA)	TrNef+G (NST=2)

Table 2. The symmetric distances between empirical 'species' trees and corresponding simulated trees (excluding 2.5% at either extreme), assuming a coalescence process with Ne and generation time as indicated. In two sets of comparisons *E. lusitanica* is represented according to CP and ITS signal separately. Values ≥6 (the distance between empirical CP and ITS trees) are indicated in bold: these represent the conditions under which the degree of gene tree difference could be explained by coalescent stochasticity.

Ne	Generation time (years)	Distance plastid	Distance ITS
10,000	2 / 20	0-2 / 0-4	0-2 / 0-4
20,000	2 / 20	0-2 / **0-8**	0-2 / **0-6**
100,000	2 / 20	0-4 / **0-20**	0-4 / **0-18**
200,000	2	**0-8**	**0-6**

Table 3. The percentage of simulated gene trees (assuming generation time of 2 years and increasing effective population sizes as indicated) that include 1) three clades that are strongly supported by both plastid and ITS datasets; and 2) two clades that define the topological conflict in *E. lusitanica* (the sister group relationship to *E. arborea* in the ITS tree, and *E. arborea*/*E. australis* clade in the plastid tree).

	Plastid/ITS			**plastid**	**ITS**
Ne	maderensis/ cinerea	mackayana/ andevalensis/ tetralix	vagans/ manipuliflora/ scoparia	arborea/ australis	arborea/ lusitanica
200,000	100/100	100/100	100/100	100	100
500,000	100/99	95/97	99/98	91	100
1,000,000	92/94	84/85	90/92	77	97
2,000,000	74/76	68/66	75/74	55	85
5,000,000	<50/<50	<50/<50	<50/<50	<50/<50	<50/<50