New Phytologist

Research

An approach to identify putative hybrids in the 'coalescent stochasticity zone', as exemplified in the African plant genus *Streptocarpus* (Gesneriaceae)

Margaret J. de Villiers[1], Michael D. Pirie[1], Mark Hughes[2], Michael Möller[2], Trevor J. Edwards[3] and Dirk U. Bellstedt[1]

[1]Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland, 7602, South Africa; [2]Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR, UK; [3]Botany Department, La Trobe University, Melbourne, Vic., Australia

Author for correspondence:
Margaret de Villiers
Tel: +27 021 808 5841
Email: margaret@sun.ac.za

Received: *25 October 2012*
Accepted: *29 November 2012*

New Phytologist (2013)
doi: 10.1111/nph.12133

Key words: genealogical sorting index, hybridization, incomplete lineage sorting, morphological optimizations, *Streptocarpus*.

Summary

• The inference of phylogenetic relationships is often complicated by differing evolutionary histories of independently-inherited markers. The causes of the resulting gene tree incongruence can be challenging to identify, often relying on coalescent simulations dependent on unverifiable assumptions. We investigated alternative techniques using the South African rosulate species of *Streptocarpus* as a study group.

• Two independent gene trees – from the nuclear ITS region and from three concatenated plastid regions (*trn*L-F, *rpl*20-*rps*12 and *trn*C-D) – displayed widespread, strongly supported incongruence. We investigated the causes by detecting genetic exchange across morphological borders using morphological optimizations and genetic exchange across species boundaries using the genealogical sorting index.

• Incongruence between gene trees was associated with ancestral shifts in growth form (in four species) but not in pollination syndrome, suggesting introgression limited by reproductive barriers. Genealogical sorting index calculations showed polyphyly of two additional species, while individuals of all others were significantly associated. In one case the association was stronger according to the internal transcribed spacer data than according to the plastid data, which, given the smaller effective population size of the plastid, may also indicate introgression.

• These approaches offer alternative ways to identify potential hybridization events where incomplete lineage sorting cannot be rejected using simulations.

Introduction

It has become standard practice to generate multiple gene trees to infer the phylogenetic relationships of organisms. From analyses of multilocus data sets, it has become increasingly apparent that gene trees more often than not differ to varying degrees, making the inference of a species tree from individual gene trees challenging. Incongruence among gene trees can be the result of analytical artefacts (e.g. paralogy or long-branch attraction), as well as biological phenomena (e.g. incomplete lineage sorting, horizontal gene transfer or hybridization).

Distinguishing hybridization from incomplete lineage sorting (ILC) is challenging because they can produce similar patterns in gene trees (Holder *et al.*, 2001). However, ILC is more likely when differences between gene trees represent alternative arrangements of relatively short branches. This is because with increasing elapsed time between cladogenic events genetic drift will remove many of the alleles that were originally shared among multiple daughter species and the remaining alleles will diverge to form newer alleles unique within each species (Degnan & Rosenberg, 2009). By contrast, hybridization can result in the transfer of genetic material between both closely- (Parrish *et al.*, 2003) and more distantly-related species (Alice *et al.*, 2001; Martinsen *et al.*, 2001; Lumaret & Jabbour-Zahab, 2009; Pirie *et al.*, 2009), and the resulting gene trees may therefore exhibit deeper incongruences.

Recently, many studies have attempted to distinguish hybridization from ILC as causes of incongruence (Kubatko & Degnan, 2007; Polihronakis, 2009) using, in particular, coalescent simulations (Buckley *et al.*, 2006; Holland *et al.*, 2008; Maureira-Butler *et al.*, 2008; Joly *et al.*, 2009; Pirie *et al.*, 2009; Willyard *et al.*, 2009; Pelser *et al.*, 2010). Coalescent simulations can indicate the conditions under which gene tree differences could be explained by ILC, but cannot be used to exclude the possibility of

hybridization. The conditions in question are effective population size and generation time, which are often difficult to estimate and are rarely consistent across all the lineages in question, and absolute ages, which are often bounded by wide confidence intervals.

Bayesian species tree inference methods (Drummond & Rambaut, 2007) can incorporate uncertainty in these parameters, but in turn generally work under the assumption that all gene tree conflict is the result of coalescent stochasticity, that is, that any hybrids have already been successfully identified and excluded. Alternative methods for inferring reticulate species trees have been proposed but these are subject to other limitations, for example requiring that both the species tree and the parentage of potential hybrids is known (Meng & Kubatko, 2009) and/or rely on the availability of a large number of independent loci (Bloomquist & Suchard, 2010; Yu *et al.*, 2011).

Further approaches to separate hybridization from ILC are therefore needed as alternative and corroborative tests for the many cases in which effective population sizes are large, timescales are short, and/or these factors are not known with confidence. These conditions define the challenging 'coalescent stochasticity zone' in which ILC cannot be rejected using simulations. There is a danger that under current methodological approaches hybridization will be systematically ignored when it occurs within this (potentially broad) zone.

The genealogical sorting index (*gsi*; Cummings *et al.*, 2008) is a measure of the degree of exclusive association of the members of a predefined group such as a species across a phylogenetic tree. The *gsi* is typically used to determine whether a group of organisms is different enough from its relatives to warrant recognition as a separate species (Cranston *et al.*, 2009; Bart *et al.*, 2010; Costanzo & Taylor, 2010; Weisrock *et al.*, 2010). However, Polihronakis (2009) applied it to detect taxa that show different degrees of monophyly in different gene trees, which might therefore have been involved in hybridization events. Previous expectations of species monophyly given coalescence differ for (selectively neutral) markers with different effective population sizes. Plastids are usually uniparentally inherited in angiosperms and thus have an effective population size that is smaller than that of biparentally inherited nuclear markers (Palumbi *et al.*, 2001; Hedrick, 2007). This means that lineage sorting of plastids occurs faster among interbreeding populations, and plastid markers are thus more likely to exhibit species monophyly in the absence of hybridization than are nuclear markers. Deviations from this expectation might be explained by hybridization (Chan & Levin, 2005).

Morphological optimizations have been used extensively to investigate the evolution of morphological characters in lineages by optimizing the characters on phylogenetic trees containing representative taxa (Palee *et al.*, 2006; Leclère *et al.*, 2009; Wright, 2009). Hybridization between morphologically distinct lineages can lead to an overestimation of the numbers of independent origins of morphological character states inferred assuming (incorrectly) a strictly bifurcating species tree (Pirie *et al.*, 2009). Conversely, gene tree conflict resulting in conflicting scenarios of morphological inheritance might be interpreted as evidence with which to identify putative hybrids. A species, monophyletic or not, that appears to have inherited a complex morphological trait from a direct ancestor according to one molecular marker but, by contrast, to have independently acquired that trait following divergence from a different common ancestor (or different common ancestors) according to another, is a plausible candidate for a hybridization scenario.

Both calculations of *gsi* values and optimizations of morphological characters are strongly dependent on the resolution of the phylogenetic tree. Lack of resolution weakens these calculations to the point at which they may no longer be meaningfully interpreted. Pirie *et al.* (2009) developed a concatenation approach in order to improve resolution without losing the information represented by conflict between gene trees. Taxa that exhibit conflicting phylogenetic positions according to different gene regions are duplicated in the matrix and the (different) conflicting gene regions recoded as missing data for each duplicate. The resulting supermatrix can be analysed using standard phylogenetic techniques to produce either a single tree (in which conflicting taxa are represented more than once; Pirie *et al.*, 2008, 2009) or, by excluding duplicates, two or more 'combined gene trees' (Pirie *et al.*, 2009). We apply the latter approach here to our empirical example, *Streptocarpus* Lindl. (Gesneriaceae).

Streptocarpus constitutes a good biological group for developing alternative approaches to unravel causes of incongruence among gene trees. This genus is native to parts of Africa and Madagascar. Hybridization appears to be part of the speciation processes in the genus, apparently entirely at the homoploid level (Hilliard & Burtt, 1971), as 58 chromosome counts of *c.* 140 described *Streptocarpus* species have thus far indicated diploidy with $2n = 2x = 30$ (for subgenus *Streptocarpella* Fritsch) or $2n = 2x = 32$ (for subgenus *Streptocarpus* and *Streptocarpus papangae* Humbert). Only three Madagascan species – *Streptocarpus hildebrandtii* Vatke ($2n = 6x = 128$), and *Streptocarpus perrieri* Humbert ($2n = 4x = 64$) and *Streptocarpus thompsonii* R.Br. ($2n = 4x = 60$) – show evidence of polyploidy (Jong & Möller, 2000; Möller & Cronk, 2001a; Möller *et al.*, 2002 onwards). In addition, the plastid genome has been shown to be transferred exclusively through the maternal line in *Streptocarpus* (Möller *et al.*, 2003). Möller & Cronk (2001a) reconstructed an internal transcribed spacer (ITS: region of nuclear ribosomal DNA) phylogeny of *Streptocarpus* and inferred that growth form is plastic within the genus, each state having originated more than once. Harrison *et al.* (1999) and later Hughes *et al.* (2006) optimized floral type onto ITS phylogenies of *Streptocarpus*, and found the six floral types to be plastic. This plasticity could reflect independent origins of similar morphologies (i.e. homoplasy) or might reflect the 'capture' of the molecular marker from a morphologically dissimilar parent species (Pirie *et al.*, 2009).

We seek to test these alternative hypotheses by focusing on one clade identified by Möller & Cronk (2001a,b) in their ITS analysis, the 'Cape primrose clade' (CPC). This clade consisted of 11 of the 14 South African rosulate species included in Möller & Cronk (2001a,b) and has a geographical distribution limited to South Africa, extending across most of eastern South Africa from

the eastern extremities of the Western Cape up through the Eastern Cape, KwaZulu/Natal, Mpumalanga and Limpopo Provinces. Our results show numerous incidences of conflict between the phylogenetic signals of nuclear ITS and plastid sequences of South African *Streptocarpus*. However, there is no direct evidence for the age of *Streptocarpus* (the estimate of Möller & Cronk, 2001b, 25–50 MY (million years), was based on assumed molecular evolutionary rates), which, in combination with uncertainty in effective population sizes, radically reduces the power of coalescent simulation approaches to explain this conflict. We demonstrate an alternative approach in proposing an improved phylogenetic hypothesis for the group and attempting to unravel the causes of this phylogenetic incongruence within it.

Materials and Methods

Taxon sampling

In total, 85 populations representing 29 of 55 currently-recognized South African *Streptocarpus* species were sampled, including 10 of the 11 species assigned to the exclusively South African CPC. Multiple samples of widely distributed species were included. Where in previous studies the sequences were obtained from cultivated material the species in question were either resequenced from samples collected in the wild or were omitted entirely in order to avoid the risk of including plants of uncertain origin in the gene tree comparison. Two outgroup species from central Africa and Madagascar were sampled following the results of a global analysis of *Streptocarpus* (see later). The Supporting Information, Table S1, lists the accessions from which sequence data were generated for this study and Fig. 1 shows the collection localities of the South African taxa sampled.

DNA extraction

Total genomic DNA was extracted from 2 to 5 mg dried leaf squares using the cetyltrimethylammonium bromide (CTAB) extraction method (Doyle & Doyle, 1987), with modifications (*c.* 10 mg polyvinylpyrrolidone and doubled concentration of β-mercaptoethanol in the extraction buffer; use of a TissueLyzer (Qiagen); precipitation of DNA at −18°C for at least 30 min after adding 2.5 × volumes ice-cold isopropanol).

PCR amplification and sequencing

Sequences from four regions were generated: the nuclear ITS region, and the plastid regions *trn*L-F intron-spacer (*trn*L-F region), *rpl*20-*rps*12 and *trn*C-D.

The PCR amplifications for all loci were performed using final concentrations of 1× JMR-455 buffer, 2.5 mM $MgCl_2$, 200 μM dNTPs, 40 μg ml^{-1} bovine serum albumin, 0.5 μM each of the forward and reverse primers, 0.025 U μl^{-1} *Taq* polymerase and *c.* 2.4 ng DNA template in 100 μl reaction volumes.

The following primers were used to amplify the regions: AB101 (Douzery *et al.*, 1999) and ITS8P (Möller & Cronk, 1997) for ITS, c and f (Taberlet *et al.*, 1991) for *trn*L-F and *rpl*20 and 5′-*rps*12 (Hamilton, 1999) for *rpl*20-*rps*12. For the *trn*C-D region, primer pairs *trn*C-f and *trn*D-M (Demesure *et al.*, 1995) and *ycf*6f and *psb*Mr (Heinze, 2007) were used. In addition, *trn*Cf1 (5′ ACATGGCCGAGTGGTAAG 3′), *trn*CDf4 (5′ GC TAAGAGGCGGAAGTTG 3′), *trn*CDr1 (5′ GAAAGAGTCT GATTCATATGATAGA 3′), *trn*CDf3 (5′ AGAATAAGAGA TCGATAGTATGG 3′) and *trn*CDr4 (5′ TACTATTCAAGTC TCGACTACG 3′), designed on the *trn*C-D sequences of the *Streptocarpus* taxa sequenced during this study, were used to obtain sequences spanning the entire region.

The PCR conditions were as follows: for ITS, 94°C for 5 min, 30 cycles at 94°C, 55°C and 72°C for 1 min each with a final extension for 10 min at 72°C; for *trn*L-F, 35 cycles at 94°C, 54°C and 72°C for 45 s each with a final extension for 6 min at 72°C; for *rpl*20-*rps*12, 95°C for 1 min, 35 cycles at 95°C for 30 s, and 53°C and 72°C for 45 s each with a final extension for 10 min at 72°C; and for *trn*C-D, 95°C for 1 min, 35 cycles at 95°C for 30 s, 57°C for 30 s and 72°C for 2 min with a final extension for 10 min at 72°C. The PCR products were purified using the Wizard SV Gel and PCR Clean-up System (Promega Corporation, Madison, WI, USA).

Cycle sequencing was carried out in 10-μl reactions containing 1.3 μl Terminator Ready Reaction Mix (Applied Biosystems, Foster City, CA, USA), 2.7 μl Half-Dye™ Mix (Bioline Ltd, London, UK), 0.32 pmol μl^{-1} of each primer, *c.* 100 ng DNA, and made up to 10 μl with double-distilled H_2O. The primers used for sequencing were the same as for amplification, but the internal ITS2G (Möller & Cronk, 1997) was used for the sequencing of ITS when needed. The PCR conditions were 35 cycles for 10 s at 96°C, 30 s at 52°C and 4 min at 60°C.

Cycle sequencing products were run on an ABI Prism 3100 or 3130 XL 16-capillary Genetic Analyser (Applied Biosystems) at the Central Analytical Facility, University of Stellenbosch, South Africa. Sequence electropherograms were edited in CHROMAS 2.13 (Technelysium Pty, Tewantin, Queensland, Australia) and were submitted to GenBank (Table S1). The sequences were aligned visually in BIOEDIT 7 (Hall, 1999). Owing to alignment ambiguities of some bases in the ITS matrix of the South African taxa, a stretch of 10 bases was omitted from the analyses. Single polymorphic peaks in the ITS chromatograms were confirmed by resequencing where necessary and coded using standard degeneracy codes. Rather than attempting to phase polymorphic sequences, or conduct extensive cloning experiments in an attempt to sequence multiple ITS copies, where present, we simply excluded from the analyses any sequences including more than a single polymorphism.

Analysis of the sequence data

A global data matrix (the *Streptocarpus* ITS matrix) of all of the ITS sequences currently available from *Streptocarpus* (Möller & Cronk, 2001a,b; Hughes *et al.*, 2005; this study) was assembled in order to determine the relationships of the putatively CPC taxa within the genus (i.e. to test the delimitation of the ingroup) and identify appropriate outgroups. Polymorphic sequences were excluded (as earlier) and the data were assessed for potential

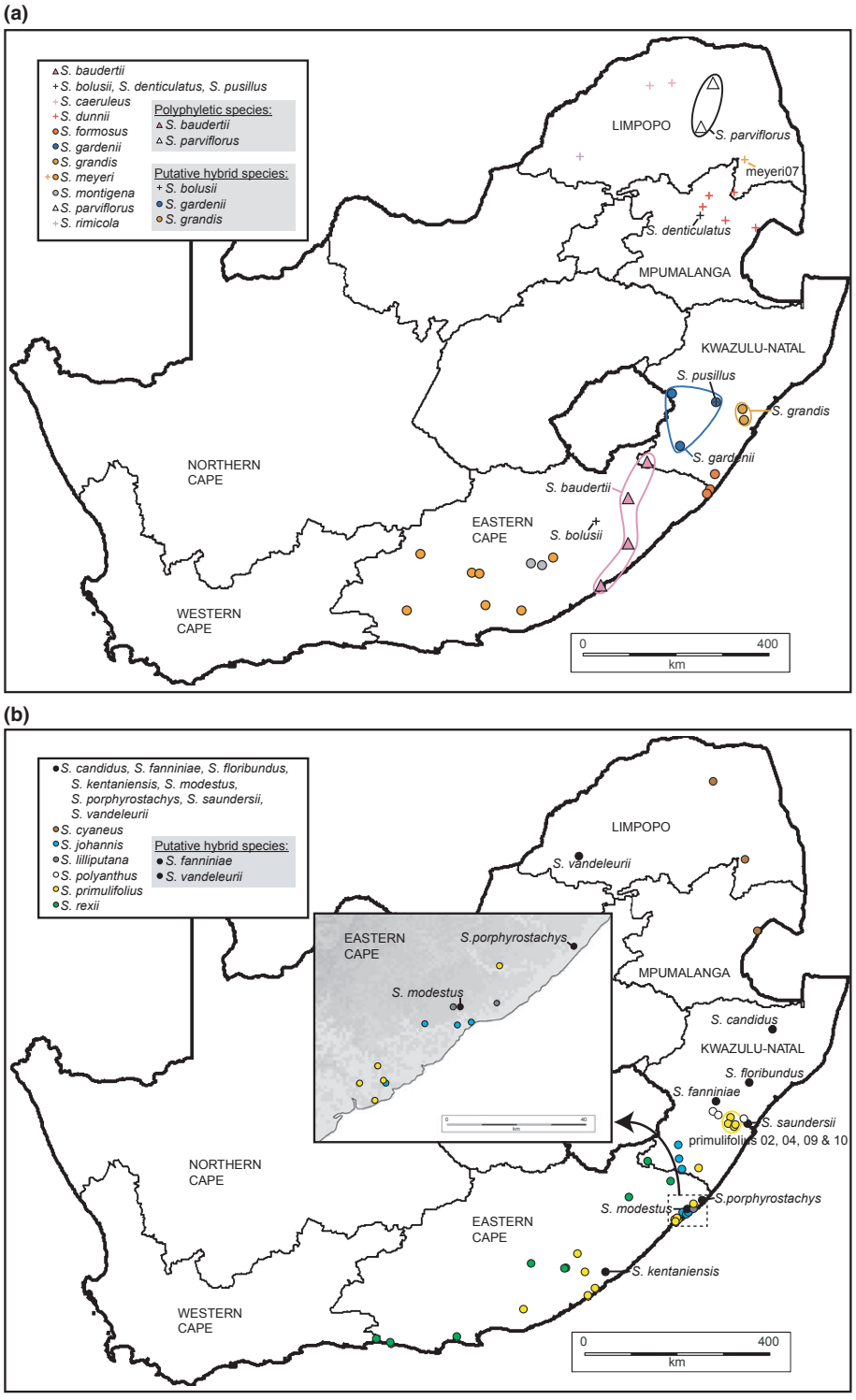

Fig. 1 Collection localities of all of the South African samples included in this study (see the Supporting Information, Table S1) divided between two maps (a and b) for clarity, with an enlargement of the Pondoland Centre area of the Eastern Cape also shown in (b). Species represented by only one population are in black; the rest of the populations are colour-coded according to the species to which they belong. Taxa that consistently emerge separately from the Cape Primrose clade taxa of Möller & Cronk (2001a,b) in all of the topologies are depicted with plus signs. *Streptocarpus baudertii* and *Streptocarpus parviflorus*, two putative polyphyletic species, are indicated. The putative hybrid species identified from the growth form optimizations and from the gsi_T (the genealogical sorting index value calculated from an ensemble of trees) calculations are also indicated.

recombination using both SIMPLOT (Lole *et al.*, 1999) and phylogenetic analyses of ITS1 and ITS2 separately (to test for phylogenetic conflict). These analyses showed no evidence for recombination (data not shown). This matrix included 185 samples from across the geographic range of the genus, representing both subgenera. Based on the results of this analysis, two further

targeted data sets of mostly South African taxa were assembled – one containing ITS sequences (Hughes *et al.*, 2005; this study; the SA ITS matrix), the other concatenated sequences from the three plastid regions (Hughes *et al.*, 2005, 2006; this study; the SA plastid matrix).

All data sets were analysed in PAUP* 4.0b10 (Swofford, 2002) using Fitch Parsimony (FP; Fitch, 1971). The heuristic search strategy involved 1000 random taxon additions, with tree-bisection-reconnection (TBR) branch swapping on and MulTrees on, saving a maximum of 10 trees per replicate. Support for clades was estimated by performing 10 000 nonparametric bootstrap (BS) pseudoreplicates with random taxon addition, TBR branch swapping on and MulTrees off. Branches with BS $\geq$ 75% were considered to be strongly supported.

The separate SA ITS and SA plastid data sets were also each analysed using Bayesian inference (BI). The best-fitting model of nucleotide substitution for the nuclear region and the combined plastid regions were each determined with the Akaike Information Criterion (Akaike, 1974) test as implemented in MRMODELTEST 2.2 (Nylander, 2004). The BI analyses were performed in MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003). Runs of 5 000 000 generations for the SA ITS data set and 15 000 000 generations for the SA plastid data set were conducted using two independent runs of four chains each and sampling every 1000 generations. Convergence (both topological and sampling of model parameters) and the length of the burn-in for each analysis were assessed using TRACER v1.5.0 (Rambaut & Drummond, 2003) and ARE WE THERE YET (Wilgenbusch *et al.*, 2004). For each of the two analyses, the trees generated during the burn-ins were discarded and the two tree files were combined using LOGCOMBINER v1.4.7 (Drummond & Rambaut, 2006). The SA ITS and SA plastid 50% majority-rule consensus trees were constructed in PAUP* from the tree files. Clades with posterior probability values (PP) $\geq$ 0.95 were considered to be strongly supported.

The BS consensus trees resulting from the FP analyses and the 50% majority-rule consensus trees resulting from the BI analyses of the SA ITS and plastid data sets were then compared to assess congruence between the phylogenetic signals of the two markers. Taxa were identified as incongruent (the 'conflicting' taxa) when they emerged in strongly supported (BS $\geq$ 75% and/or PP $\geq$ 0.95) different positions between the two trees relative to the other ('congruent') taxa. First we excluded the conflicting taxa and performed separate and combined analyses of the SA ITS and SA plastid sequence data (the reduced ITS, reduced plastid and reduced combined matrices). Second, we performed combined analyses including the congruent taxa, one analysis including the ITS sequences of the conflicting taxa and the plastid data coded as missing (the combined + ITS matrix), and a separate analysis including the plastid sequences of the conflicting taxa and the ITS data coded as missing (the combined + plastid matrix), according to the method of Pirie *et al.* (2009; also used in Antonelli *et al.*, 2010; Pelser *et al.*, 2010; Humphreys *et al.*, 2011). The reduced ITS, reduced plastid, reduced combined, combined + ITS and combined + plastid matrices were analysed using both FP and BI, with the reduced ITS matrix, the reduced plastid and the reduced combined each run for 10 000 000 generations, and the combined + ITS and the combined + plastid each for 15 000 000 generations in the BI analyses, and the same settings as above for the FP analyses. A strict consensus tree of the FP analysis and both 50% and 95% majority-rule consensus trees of the BI analysis (the former was used to display the results and the latter used for the growth form optimizations) were constructed in each case. Subsets of the trees resulting from the BI analyses of the reduced combined, combined + ITS and combined + plastid matrices were used in all subsequent analyses.

Morphological optimizations

In order to assess the extent of potential genetic exchange across morphological boundaries, morphological character optimizations were conducted. Reconstructions of morphological character evolution given the combined + ITS and the combined + plastid phylogenies were compared to give an indication of the extent to which each marker has been involved in genetic exchange between lineages possessing different morphological characteristics.

For the character optimizations, two morphological characters – growth form and floral form – were analysed. Caulescent taxa possess a typical shoot with opposite leaves. By contrast, because of a modified development of the shoot apical meristem, acaulescent taxa either consist of a single cotyledon with one or two additional leaves arising from its base (the unifoliates and plurifoliates), or of a single cotyledon with many additional leaves arising from its base in the shape of a rosette (the rosulates; Hilliard & Burtt, 1971; Jong & Burtt, 1975). The growth forms of taxa were coded as caulescent, unifoliate/plurifoliate and rosulate as used in Möller & Cronk (2001a). Floral form was coded as pouch, bird, open-tubed and keyhole, as in Harrison *et al.* (1999) and Hughes *et al.* (2006). A new floral form, the Acanth type, was specified for the representatives of *Streptocarpus lilliputana* Bellstedt & T.J. Edwards. The flower of *S. lilliputana* is very similar to that of two Acanthaceae species, *Mackaya bella* Harv. and *Asystasia varia* N.E.Br.: the tube curves downwards, widening abruptly in the upper third of the tube.

The optimizations were performed in MESQUITE 2.72 (Maddison & Maddison, 2009) using the FP criterion on trees from the BI analyses of the reduced combined, the combined + ITS and the combined + plastid matrices, thinned down to 1000 trees in each case using LOGCOMBINER v1.4.7 (Drummond & Rambaut, 2006) to speed up the analyses. For each of the matrices, the results of the optimizations were summarized on the 95% majority-rule consensus tree from the BI analyses.

Assessment of monophyly of the species

Finally, to determine the degree of potential genetic exchange across species boundaries, the extent of association among the members of the 20 species represented by more than one individual was calculated by means of the *gsi* (Cummings *et al.*, 2008) implemented at http://www.genealogicalsorting.org. A *gsi* of 1 indicates that the species forms a monophyletic group, whereas a *gsi* approaching 0 indicates that the members of the species are to a greater or lesser extent randomly dispersed among members of

other species in the trees. A permutation test gives the probability of retrieving a *gsi* value equal to or greater than the observed *gsi* value by chance under the null hypothesis that there is no significant association among the members of each species (Cummings *et al.*, 2008). We used the gsi_T (the *gsi* value calculated from an ensemble of trees) calculated from a sample of 100 BI trees each from the analyses of the reduced combined, the combined + ITS and the combined + plastid matrices and the significance of the gsi_T values were calculated using 10 000 permutations on 20 of the BI trees from each matrix.

Results

Fitch parsimony analysis of the genus *Streptocarpus* using ITS sequence data

The global *Streptocarpus* ITS matrix included 749 characters, of which 443 (59.1%) were variable and 341 (45.5%) were parsimony-informative. The FP analysis resulted in 530 equally parsimonious trees of 1473 steps with a consistency index (CI) of 0.486 and a retention index (RI) of 0.817.

This analysis yielded a similar topology (Fig. S1) to that obtained by Möller & Cronk (2001a,b). The Southern African ITS2 deletion clade (BS = 71%) and South African CPC (BS = 54%) of Möller & Cronk (2001a,b) both emerged within a strongly supported clade (BS = 80%) otherwise containing all of the South African taxa. Based on this analysis, *S. papangae* (subgenus *Streptocarpella*, from Madagascar) and *S. montanus* Oliv. (subgenus *Streptocarpus*, from Kenya/Tanzania) were selected as outgroups for the subsequent analyses focused on the South African taxa, especially the CPC.

Nuclear and plastid sequence data of the South African *Streptocarpus* taxa

The ITS sequence data of the South African taxa were much more divergent than the corresponding plastid sequence data, with *c.* 26% variable characters among the ITS sequences as opposed to only *c.* 7% among the concatenated plastid sequences (Table 1). Nevertheless, as there were more plastid characters, the plastid sequences yielded more variable characters than the ITS sequences (173 variable characters in the 674-character SA ITS matrix, 280 variable characters in the 3955-character SA plastid matrix).

Phylogenetic analyses of the ITS, plastid and combined data matrices of the South African taxa

The FP analysis of the SA ITS matrix yielded 9720 equally parsimonious trees of 252 steps with a CI of 0.774 and an RI of 0.926 (Fig. S2a, Table 1). The BS analysis resulted in 22 strongly supported clades (BS $\geq$ 75%). For the BI analysis (Fig. 2a), the SYM + G model was selected as best fitting the ITS data by MRMODELTEST. Trees from the first 500 000 generations (10%) were discarded as burn-in. The BI analysis yielded 32 strongly supported clades (PP $\geq$ 0.95).

The concatenated SA plastid matrix yielded slightly more resolved trees than the ITS matrix (Fig. 2b). The plastid FP trees (Fig. S2b) had a higher CI (0.893) and RI (0.945) than did the ITS FP trees, indicating lower levels of homoplasy among the plastid characters. For the BI analysis, the AIC test indicated GTR + I + G as the best-fitting model for the plastid data. Trees from the first 399 000 generations (2.7%) were discarded as burn-in.

A comparison of the ITS and plastid Bayesian 50% majority-rule consensus trees in Fig. 2 revealed widespread incongruence between them, with only 11 identical clades (indicated by asterisks in Fig. 2). Results derived from the reduced ITS and reduced plastid matrices are presented in Fig. S3, and those from the reduced combined matrix in Figs S4 and S5. The trees from the first 399 000 (4%) generations were discarded as the burn-in in the reduced ITS, the reduced plastid and the reduced combined BI analyses. The combined analyses resulted in a greater number of variable and parsimony informative characters and a greater number of more robustly supported clades than the separate reduced ITS and reduced plastid analyses (Table 1).

The trees from the first 499 000 generations (3.3%) were discarded from the BI analyses of the combined + ITS and combined + plastid matrices. Parsimony and BI analyses of these matrices resulted in increased numbers of clades with BS $\geq$ 75% (both combined + ITS and combined + plastid trees) and PP $\geq$ 0.95 (combined + plastid trees), but the loss of one node of PP $\geq$ 0.95 in the combined + ITS tree compared with the plastid and ITS matrices analysed separately (Figs 3, S6, Table 1). Given this overall substantial improvement in support, we regard the single node PP $\geq$ 0.95 lost from the ITS BI analysis as questionable and use the trees resulting from the combined matrices for the subsequent analyses.

Morphological optimisations

Optimizations of growth form and floral form are presented in Figs 4 and 6(a) and Figs 5 and 6(b), respectively. The ancestral states of four conflicting species (indicated by arrows in Fig. 4) differed between the two topologies. Although the Acanth floral form appears to have originated once (in *S. lilliputana*, which is monophyletic in all of the analyses), floral form also appears to have been subject to multiple shifts (from an ancestral open-tubed form) according to optimizations over both the combined + ITS and the combined + plastid trees (Fig. 5). The most dispersed is the keyhole form, which is present in three species whose representatives were scattered across large portions of the topologies in both optimizations. Nevertheless, in contrast to the growth form optimizations, there is no evidence to suggest that the conflicting samples are associated with shifts in floral form.

The growth form and floral form optimizations of the reduced combined topology (Fig. 6) represent the evolution of these two characters based on relationships among only the congruent taxa (i.e. in the absence of potentially confounding effects of e.g. chloroplast capture). Removal of conflicting taxa resulted in fewer shifts in growth form (Fig. 6a) and indicated that the CPC was

Table 1 Statistics from the South African (SA) nuclear internal transcribed spacer (ITS) and plastid sequence data and the Fitch parsimony (FP) and Bayesian (BI) analyses of the seven matrices

	SA ITS matrix	SA plastid matrix	Reduced ITS matrix	Reduced plastid matrix	Reduced combined matrix	Combined + ITS matrix	Combined + plastid matrix
Total characters	674	3955	674	3955	4629	4629	4629
Variable characters	173 (25.67%)	280 (7.08%)	128 (18.99%)	170 (4.30%)	298 (6.44%)	343 (7.41%)	408 (8.81%)
Variable characters/taxon	173/98 taxa = 1.77	280/98 taxa = 2.86	128/51 taxa = 2.51	170/51 taxa = 3.33	298/51 taxa = 5.84	343/98 taxa = 3.50	408/98 taxa = 4.16
Parsimony-informative characters	117 (17.36%)	141 (3.57%)	68 (10.09%)	61 (1.54%)	129 (2.79%)	178 (3.85%)	209 (4.52%)
Parsimony-informative characters/taxon	117/98 taxa = 1.19	141/98 taxa = 1.44	68/51 taxa = 1.33	61/51 taxa = 1.20	129/51 taxa = 2.53	178/98 taxa = 1.82	209/98 taxa = 2.13
MP analyses							
Equally parsimonious trees	9720	9700	42	9820	9900	8520	8930
Length	252	328	158	188	353	447	494
CI	0.774	0.893	0.886	0.947	0.901	0.834	0.877
RI	0.926	0.945	0.940	0.939	0.925	0.921	0.935
Clades with BS $\geq$ 75%/taxon	22 clades/ 98 taxa = 0.22	22 clades/98 taxa = 0.22	11 clades/51 taxa = 0.22	7 clades/51 taxa = 0.14	16 clades/51 taxa = 0.31	24 clades/98 taxa = 0.24	27 clades/98 taxa = 0.28
BI analyses							
AIC model	SYM + G	GTR + I + G	SYM + G	GTR + I + G	ITS: SYM + G plastid: GTR + I + G	ITS: SYM + G plastid: GTR + I + G	ITS: SYM + G plastid: GTR + I + G
PP $\geq$ 0.95/taxon	32 clades/98 taxa = 0.33	30 clades/98 taxa = 0.31	16 clades/51 taxa = 0.31	9 clades/51 taxa = 0.18	19 clades/51 taxa = 0.37	31 clades/98 taxa = 0.32	38 clades/98 taxa = 0.39

CI, consistency index; RI, retention index; BS, bootstrap; AIC, Akaike information criterion; PP, posterior probability.

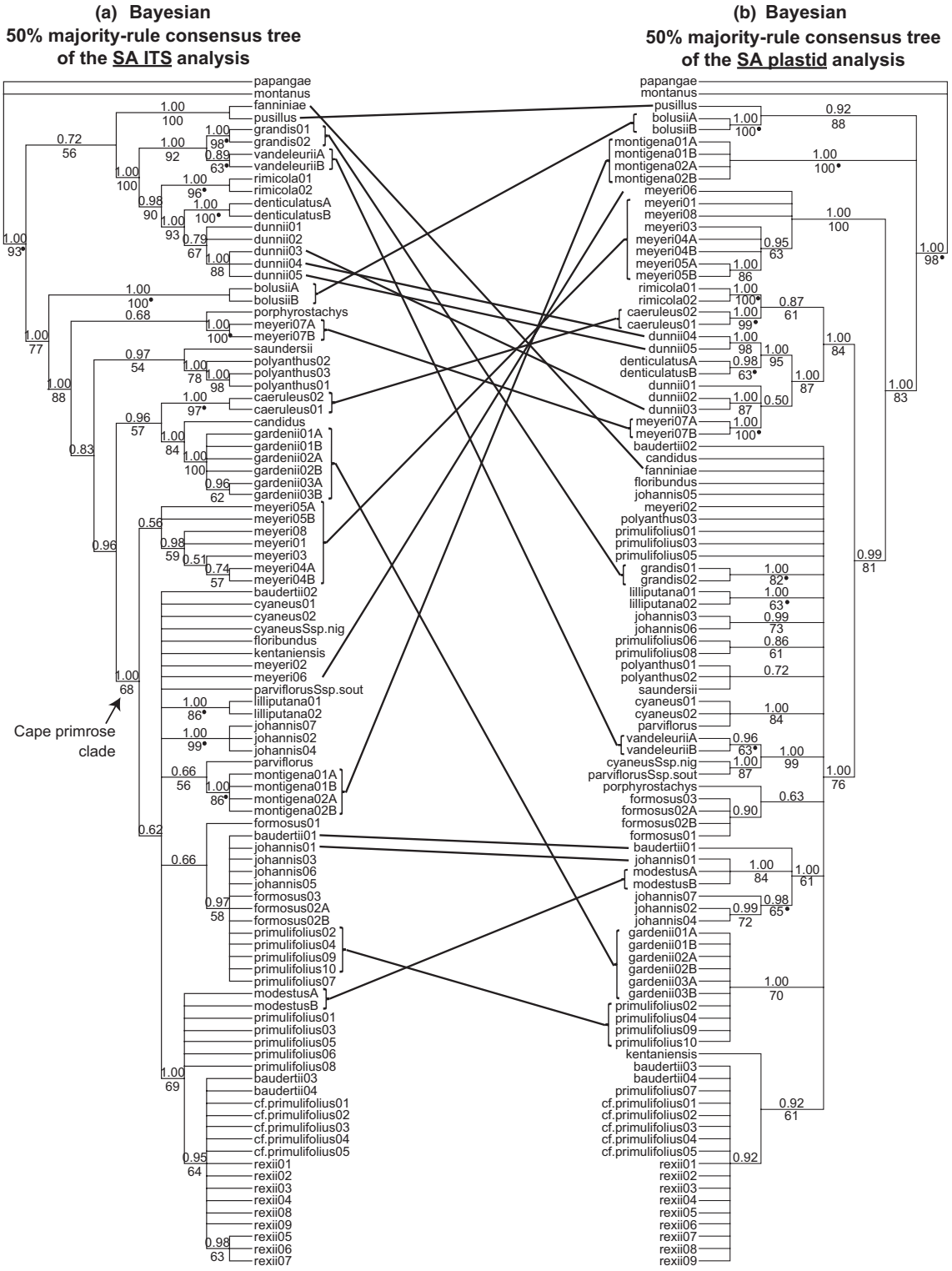

Fig. 2 Fifty per cent majority-rule consensus trees reconstructed from the Bayesian analyses of (a) the South African (SA) internal transcribed spacer (ITS) and (b) the SA plastid sequence data matrices. Posterior probabilities (PP) are given above the branches, and Fitch parsimony bootstrap values (BS) are given below the corresponding branches. Clades that contain the same taxa between the two trees are marked by an asterisk below the subtending branches in each topology. Lines between the two topologies indicate the alternative positions of the taxa that emerge in strongly supported but conflicting positions between the two trees. The clade congruent with the Cape primrose clade of Möller & Cronk (2001a,b) is indicated in the SA ITS topology. Refer to Supporting Information Table S1 for locality data of the taxa.

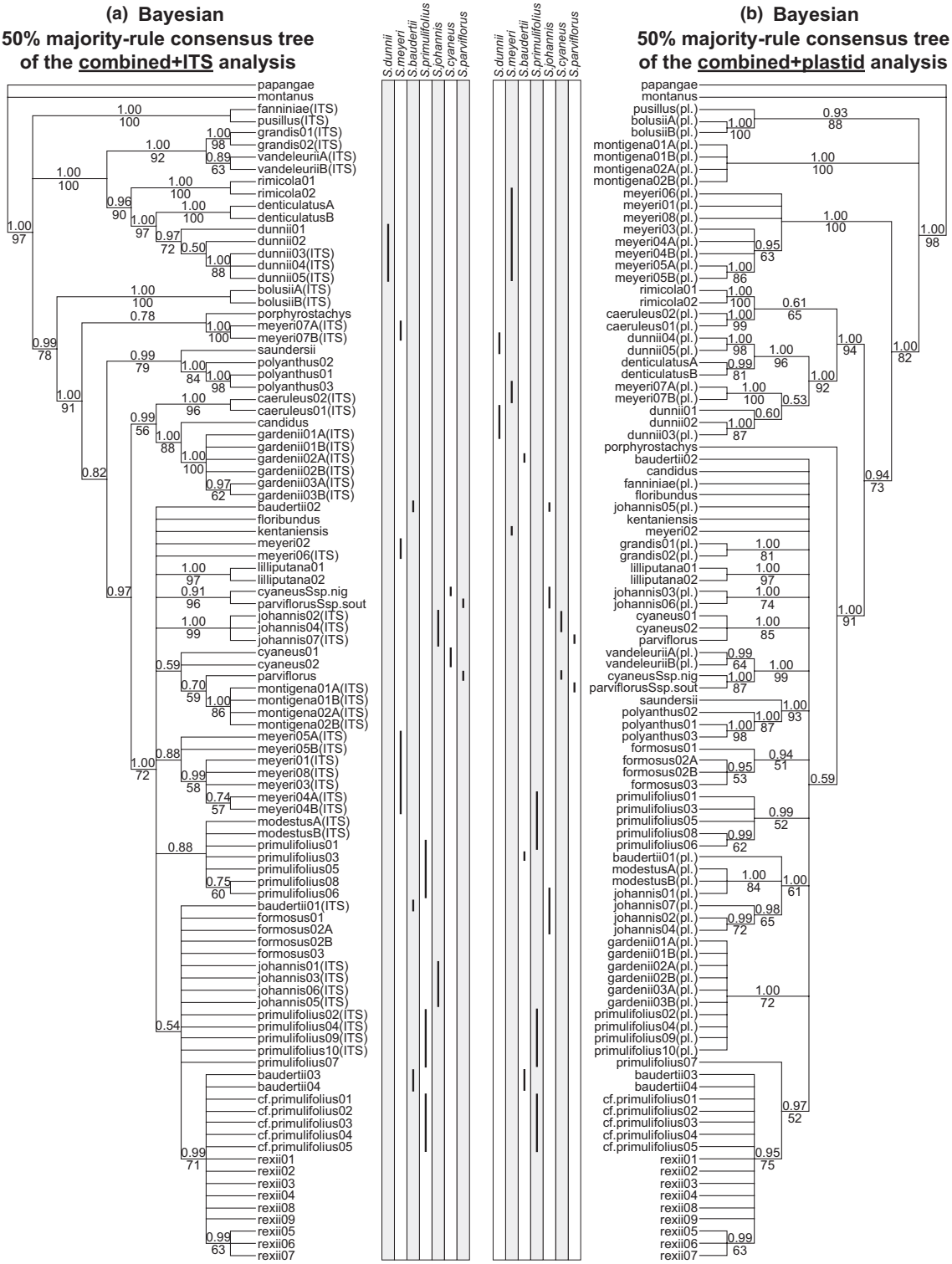

Fig. 3 Fifty per cent majority-rule consensus trees reconstructed from the Bayesian analyses of (a) the combined internal transcribed spacer (ITS) and plastid sequence data only including the ITS sequences of the conflicting taxa (combined + ITS) and (b) the combined ITS and plastid sequence data only including the plastid sequences of the conflicting taxa (combined + plastid). Posterior probabilities (PP) are given above the branches, while Fitch parsimony bootstrap values (BS) are given below the corresponding branches. The samples of species that are scattered across the separate trees are indicated by bars to the right of the combined + ITS tree and to the left of the combined + plastid tree, with the corresponding species name indicated at the top of the bar. Please refer to Supporting Information Table S1 for locality data of the taxa.

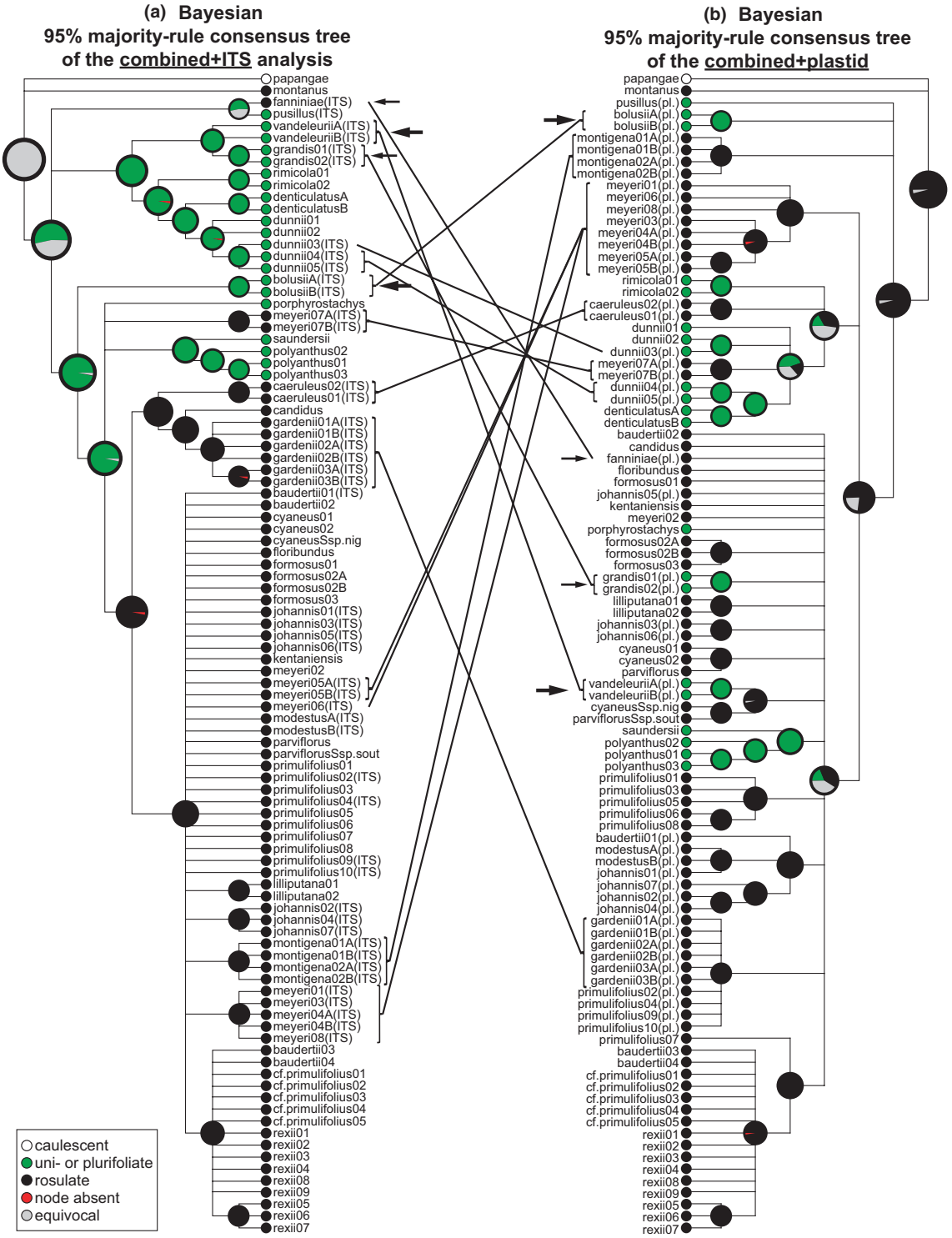

Fig. 4 Optimization of the three growth forms depicted on the 95% majority-rule consensus trees reconstructed from the Bayesian analyses of the combined internal transcribed spacer (ITS) and plastid sequence data (a) only including the ITS and (b) only including the plastid sequences of the conflicting taxa. Arrows indicate taxa that arise in clades with conflicting ancestral optimizations between the two trees. Lines between the two topologies indicate the alternative positions of the samples that emerge in strongly supported but conflicting positions between the two trees. Please refer to Supporting Information Table S1 for locality data of the taxa.

ancestrally rosulate. The exclusion of conflicting taxa did not have a similar effect on the distribution of the different floral forms (Fig. 6b): similar to the combined + ITS and the combined + plastid optimizations of floral form (Fig. 5), all but the Acanth and bird types, which are only present in one species each, were scattered across the tree (Fig. 6b). The ancestral floral form of the CPC was nevertheless inferred consistently across these analyses to have been open-tubed.

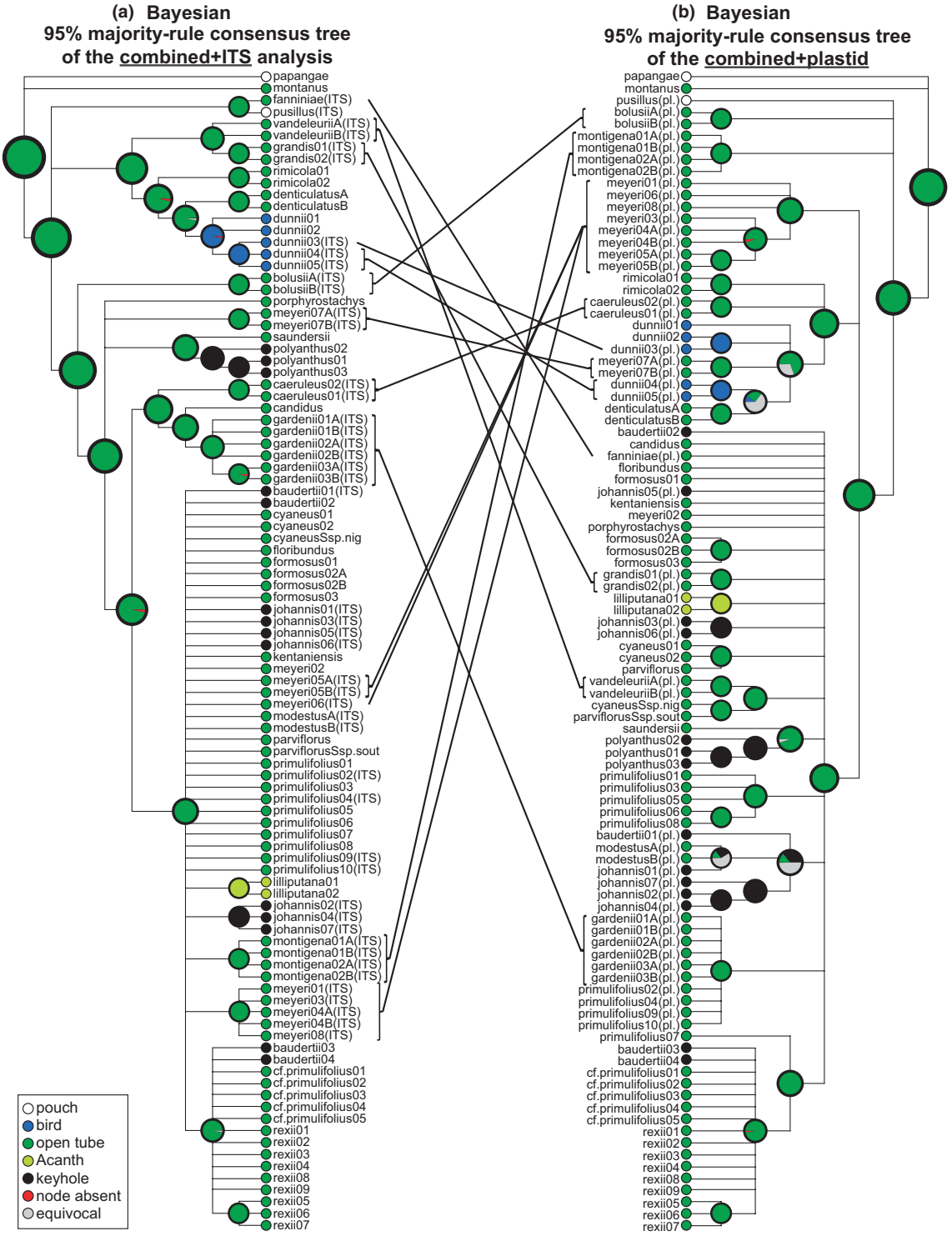

Fig. 5 Optimization of the five floral forms depicted on the 95% majority-rule consensus trees reconstructed from the Bayesian analyses of the combined internal transcribed spacer (ITS) and plastid sequence data, (a) only including the ITS and (b) only including the plastid sequences of the conflicting taxa. Lines between the two topologies indicate the alternative positions of the samples that emerge in strongly supported but conflicting positions between the two trees. Please refer to Supporting Information Table S1 for locality data of the taxa.

Monophyly of the species

The gsi_T values calculated from the topologies ranged from 0.114 to 1.000 for the combined + ITS analysis, from 0.121 to 1.000 for the combined + plastid analysis and from 0.095 to 1.000 for the reduced combined analysis (Table 2). The gsi_T values of 18 of the species were significant, indicating that their members were associated more with one another in the topologies than would be expected by chance. The remaining two species, *Streptocarpus baudertii* L.L. Britten and *Streptocarpus parviflorus* Hook.f., had

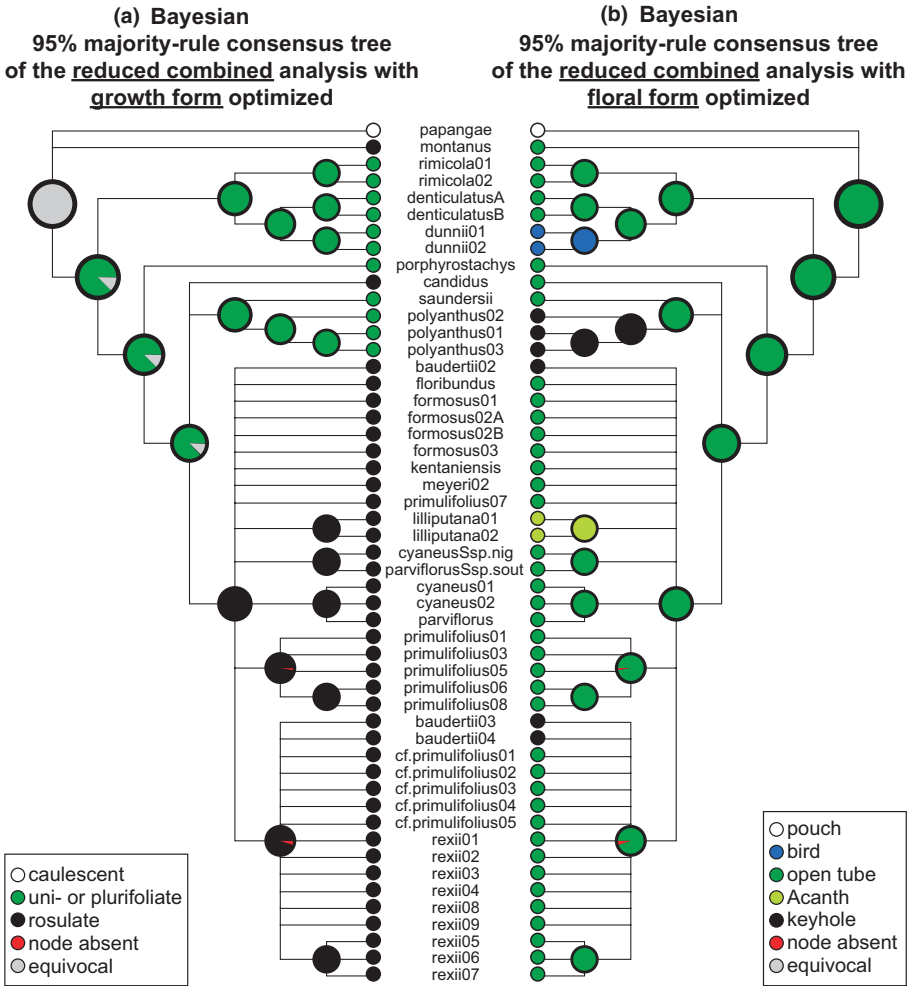

Fig. 6 Optimizations of (a) the three growth forms and (b) the five floral forms depicted on the 95% majority-rule consensus tree reconstructed from the Bayesian analyses of the combined internal transcribed spacer (ITS) and plastid sequence data excluding the samples that emerge in strongly supported conflicting positions between the ITS and plastid topologies. Please refer to Supporting Information Table S1 for locality data of the taxa.

nonsignificant gsi_T values based on the topologies of all three matrices, indicating that their members were not significantly associated with one another according to either of the markers.

Streptocarpus bolusii C.B. Clarke, *Streptocarpus caeruleus* Hilliard & B.L. Burtt, *Streptocarpus denticulatus* Turrill, *Streptocarpus grandis* N.E.Br., *S. lilliputana*, *Streptocarpus montigena* L.L. Britten, *Streptocarpus polyanthus* Hook., *Streptocarpus rimicola* Story and *Streptocarpus vandeleurii* Baker f. & S. Moore each had gsi_T values of, or very close to, 1, indicating that they were monophyletic in all or nearly all of the topologies from each of the matrices. The gsi_T values of most of the species were very similar to each other between the topologies from the combined + ITS and combined + plastid matrices, indicating that most of the species had similar degrees of exclusivity based on both markers. The gsi_T of *S. gardenii* Hook. was higher in the combined + ITS matrix than the combined + plastid matrix, indicating that the members of this species were associated more with one another in the combined + ITS topologies than in the combined + plastid topologies. The opposite was the case for *Streptocarpus formosus* (Hilliard & B.L. Burtt) T.J. Edwards, the gsi_T values of which indicated that its members were associated more with one another in the combined + plastid than in the combined + ITS topologies.

Discussion

Phylogenetic relationships and morphological evolution in South African *Streptocarpus*

The plastid and ITS gene trees (Fig. 2) are largely incongruent, and therefore could not be combined directly to infer a species tree. Despite this, relationships among the more stable taxa could be reconstructed by omitting the conflicting taxa, thereby allowing the congruent nuclear and plastid data of the remaining taxa to be combined to provide a more resolved phylogeny (Fig. 3). Following Pirie *et al.* (2009), combining the ITS and plastid sequence data of the stable taxa also provided a generally more resolved framework into which the ITS and plastid lineages of the conflicting taxa could be placed to determine their differing positions more effectively.

The resulting phylogenetic trees are still far from fully resolved, but the improvement in both phylogenetic resolution and

Table 2 Genealogical sorting index (gsi_T) calculations and *P*-values of the permutations performed on the topologies resulting from the combined + ITS, combined + plastid and the reduced combined matrices

Species	Number of individuals sampled	Combined + ITS topologies		Combined + plastid topologies		Reduced combined topologies		
		gsi_T	*P*	gsi_T	*P*	Number of individuals sampled	gsi_T	*P*
Streptocarpus baudertii	4	0.1135236	0.0938 ns	0.1212403	0.1382 ns	3	0.0952875	0.4276 ns
S. bolusii	2	1.0000000	0.0005 *	1.0000000	0.0003 *	–	–	–
S. caeruleus	2	1.0000000	0.0006 *	1.0000000	0.0005 *	–	–	–
S. cyaneus	3	0.3064541	0.0028 *	0.2586031	0.0040 *	3	0.3448661	0.0061 *
S. denticulatus	2	1.0000000	0.0003 *	0.9949480	0.0006 *	2	1.0000000	0.0026 *
S. dunnii	5	0.9916560	0.0001 *	0.7166984	0.0001 *	2	0.9897960	0.0027 *
S. formosus	4	0.4731963	0.0001 *	0.9922606	0.0001 *	4	0.9654255	0.0001 *
S. gardenii	6	1.0000000	0.0001 *	0.5893273	0.0001 *	–	–	–
S. grandis	2	1.0000000	0.0006 *	1.0000000	0.0002 *	–	–	–
S. johannis	7	0.2993816	0.0002 *	0.3360540	0.0001 *	–	–	–
S. lilliputana	2	1.0000000	0.0007 *	1.0000000	0.0001 *	2	1.0000000	0.0040 *
S. meyeri	11	0.4457665	0.0001 *	0.3428398	0.0001 *	–	–	–
S. modestus	2	0.3601923	0.0175 *	0.6514062	0.0040 *	–	–	–
S. montigena	4	1.0000000	0.0001 *	1.0000000	0.0001 *	–	–	–
S. parviflorus	2	0.1466390	0.0969 ns	0.1342533	0.1042 ns	2	0.1890185	0.0984 ns
S. polyanthus	3	1.0000000	0.0001 *	1.0000000	0.0001 *	3	1.0000000	0.0001 *
S. primulifolius	15	0.3136237	0.0001 *	0.3859147	0.0001 *	11	0.4216254	0.0001 *
S. rexii	9	0.5549498	0.0001 *	0.5586325	0.0001 *	9	0.5138421	0.0001 *
S. rimicola	2	1.0000000	0.0006 *	1.0000000	0.0003 *	2	1.0000000	0.0030 *
S. vandeleurii	2	0.9595833	0.0005 *	0.9949480	0.0002 *	–	–	–

Significance: ns, not significant; *, $P \leq 0.05$.

representation of gene trees allows us to understand better the relationships of South African *Streptocarpus*, including the troublesome delimitation of the CPC. The ITS phylogeny of this study is broadly consistent with that of Möller & Cronk (2001a,b), with the addition of the South African rosulate species *Streptocarpus floribundus* Weigend & T.J. Edwards, *S. lilliputana* and *S. parviflorus* to a purely rosulate CPC, but with the exclusion of the most isolated population (from Mariepskop) of the (otherwise CPC) species *Streptocarpus meyeri* B.L. Burtt. However, the plastid phylogeny indicates different relationships, with the three unifoliate/plurifoliate species, *S. grandis*, *S. polyanthus* and *S. porphyrostachys* Hilliard, and the rosulate species, *Streptocarpus fanniniae* (Harvey ex) C.B. Clarke, emerging among the CPC species to the exclusion of the rosulate *S. montigena* and most of the *S. meyeri* populations.

The variable positions of the conflicting taxa in our analyses are reflected in ancestral states for growth form that differ between the combined + ITS and the combined + plastid topologies (Fig. 4). The reduced combined analysis (Fig. 6a) excludes the potentially confounding influence of non-linear inheritance in phylogenetically conflicting taxa identified in the current analyses. It suggests that, with the possible exception of *S. candidus* Hilliard, the rosulate growth form arose once among the South African species. Thus a rosulate growth form can be regarded as a synapomorphy for the CPC, despite clear conflict as to the delimitation of this clade.

In contrast to the growth form optimizations, the ancestral floral form of the CPC is consistent irrespective of the gene tree used to infer it. While the pouch form is plesiomorphic in the genus (Harrison *et al.*, 1999; Hughes *et al.*, 2006), the ancestral state of the CPC is an open-tubed flower. With the exception of the Acanth form, the other forms included here have each arisen many times. The most scattered form in the optimizations is the keyhole, which is present in *Streptocarpus johannis* L.L. Britten, *S. baudertii* and *S. polyanthus*, and has also evolved in *Streptocarpus saxorum* Engl. (subgenus *Streptocarpella*) from the pouch flower form (Harrison *et al.*, 1999; Hughes *et al.*, 2006). The keyhole form is unique among the floral forms found in *Streptocarpus* in having a narrow, S-shaped tube and a strongly laterally compressed mouth that is slightly constricted in the centre, resembling a keyhole (Hilliard & Burtt, 1971).

Despite the extensive conflict between the two gene trees, the keyhole form is present almost exclusively in stable taxa whose positions are not in conflict, that is *S. johannis*, *S. baudertii* and *S. polyanthus*. There is therefore no evidence from these analyses to suggest that this form has been transferred among distantly related lineages through hybridization. This lends further credence to its role in enhancing reproductive isolation. This is not, in retrospect, an ideal character for identifying hybrids in the manner employed here. It suggests that selection pressure (such as from a specialized pollinator, albeit one that has yet to be identified) may instead have driven the evolution of more or less the same floral shape on separate occasions. A microsatellite study of genetic diversity in populations of different species possessing the keyhole form revealed significant inbreeding in some of these species (De Villiers, 2008). This suggests a plausible mechanism by

which such shifts might have occurred – allele fixation. In general, multiple such shifts in floral shape are not unusual, having been detected in many different plant groups (Roalson *et al.*, 2003; Tripp & Manos, 2008; Martén-Rodríguez *et al.*, 2010; Pirie *et al.*, 2011), and may indeed be directly implicated in the speciation process.

Monophyly of the species

The topologies reconstructed from the different matrices and the gsi_T calculations reveal different degrees of association within species. Nine of the species constitute monophyletic entities for the markers analysed. The representatives of a further nine of the species do not constitute monophyletic groups, but are nevertheless significantly associated. These species are therefore probably either relatively young and their lineages have not yet sorted to completion (van der Niet & Linder, 2008; Degnan & Rosenberg, 2009), or some of their populations have been involved in hybridizations to a limited extent.

In contrast to those of the rest of the species, the gsi_T values of *S. baudertii* and *S. parviflorus* are nonsignificant. This could have a number of causes: first, that the species is young and its gene lineages have just begun to sort; second, that its members have been extensively involved in hybridization, causing them to emerge in scattered positions in the phylogenies; and third that the species boundaries have been incorrectly defined, resulting in polyphyletic taxa.

Streptocarpus parviflorus is represented by both of its two subspecies in our analyses. It is part of a larger species complex of morphologically intergrading populations that includes *Streptocarpus cyaneus* S. Moore and five other species not included in the current analyses (Hilliard & Burtt, 1971; Edwards *et al.*, 1992, 2008; Weigend & Edwards, 1994; Edwards, 2003). The two subspecies emerged with different, geographically proximate *S. cyaneus* in both the combined + plastid and the reduced combined topologies. In the absence of supported gene tree conflict with regard to *S. parviflorus* there is no need to invoke either hybridization or lineage sorting artefacts to explain its lack of monophyly. It is more likely that species boundaries within the complex are incorrectly defined.

Streptocarpus baudertii is morphologically more distinct. Its rosette can only be confused with that of *S. meyeri* and it has the keyhole floral form that is only present in a minority of *Streptocarpus* species (of which *S. johannis* and *S. polyanthus* are included in the current analyses). Nevertheless, *S. baudertii* is variable in terms of both vegetative and floral morphology across its geographical range. The leaves of the Hillsdrift population (baudertii02), the southernmost population (referred to by Hilliard & Burtt, 1971), are held much more erect than in the rest of the species. Moreover, the populations are florally quite different from one another. For example, the flowers of the northern populations (baudertii01, baudertii03 and baudertii04) are *c.* 1.5 times larger than those of the southern Hillsdrift population (baudertii02; D.U. Bellstedt, pers. obs.). While Hilliard & Burtt (1971) proposed that *S. baudertii* originated as a hybrid (between *S. polyanthus* and *S. meyeri*) there is no evidence to support this in our results. The lack of association of its representatives in the phylogenies instead suggests that *S. baudertii* is also polyphyletic, and that the morphological similarities between its populations represent convergence.

Distinguishing hybridization from incomplete lineage sorting

The morphological optimizations of growth form and the gsi_T calculations revealed some potential hybrid taxa among the South African *Streptocarpus* species. Based on the growth form optimizations (Fig. 4, in which potential hybrids identified from these optimizations are indicated by arrows), the most convincing candidates are *S. vandeleurii* and *S. bolusii*. However, cases can also be made for hybrid origins of *S. grandis* and *S. fanniniae*. These species show similar patterns of incongruence between the combined + ITS and combined + plastid topologies. The combined + ITS topologies tend to reflect relationships surmised by Hilliard & Burtt (1971), that is *S. vandeleurii*, *S. grandis* and *S. fanniniae* group with morphologically similar species (although a more stringent test particularly with regard to *S. fanniniae* could be performed given denser sampling of non-CPC southern African species). By contrast, the combined + plastid topologies tend to reflect geographical patterns more closely, with *S. vandeleurii*, *S. grandis* and *S. fanniniae* all emerging in a predominantly rosulate clade with species of overlapping geographical ranges – a pattern suggestive of 'chloroplast capture'. *Streptocarpus bolusii* emerged in less densely sampled parts of the combined + ITS and combined + plastid topologies, and the species with which it could have hybridized are not represented in the analyses.

The gsi_T calculations (Table 2) also point to further species that might have been involved in hybridization. In most cases, the gsi_T values derived from combined + ITS vs combined + plastid topologies are similar. However, those of *S. gardenii* and *S. formosus* differ markedly. The plastid genome is inherited exclusively maternally in *Streptocarpus* (Möller *et al.*, 2003), as opposed to the biparentally inherited nuclear genome, and plastid markers therefore have smaller effective population sizes than do nuclear markers. For this reason, plastid markers should become fixed in genetically isolated groups faster than nuclear markers. A note of caution is warranted given the potential impact of concerted evolution on ITS copies, but this expectation is otherwise consistent with the higher gsi_T values for the *S. formosus* plastid, but not with the lower values for *S. gardenii*. The latter might therefore reflect past hybridization events between *S. gardenii* and the northernmost populations of *Streptocarpus primulifolius* Gand. Geographically, these two taxa/population groups are located very closely together, making introgression (chloroplast capture) a plausible scenario. There is, however, no obvious way in which to test the significance of differences between significant *gsi* values (as opposed to significant vs nonsignificant ones). This conclusion should therefore be treated as a hypothesis to be tested with further independent data.

Conclusions

Gene tree conflict in Southern African *Streptocarpus* is rampant. As such, the CPC cannot be defined consistently according to both plastid and ITS data. Nevertheless, in taking this conflict into account we have identified strong support for a CPC of variable membership that descends from a common ancestor with rosulate growth form and open-tubed flowers. By inference from either gene tree, it would appear that shifts in growth and floral form are common in the group. A number of apparent growth form shifts (in ancestors of *S. bolusii*, *S. vandeleurii*, *S. grandis* and *S. fanniniae*) are associated with gene tree conflict and might be interpreted as evidence for hybridization, as opposed to ILC. By contrast, shifts in floral form are not associated with conflict amongst the ITS and plastid data. This might be expected in a reticulate evolutionary scenario given the role of floral form in maintaining reproductive isolation. From our calculations of *gsi* for each of the species, one further putative hybrid was identified: *S. gardenii*, the plastid sequences of which would otherwise be inferred to have sorted more slowly than those of nuclear ITS. This is contrary to expectations given the smaller effective population size of plastids. Two further species, *S. baudertii* and *S. parviflorus*, were shown to be polyphyletic, illustrating the further taxonomic complexities of South African *Streptocarpus*.

The power of the methods presented here is limited so they cannot be used to categorically confirm or refute putative hybrids. However, the current standard approach using coalescent simulations is also limited – it can only be used to discern between conflict that could or could not be caused by coalescent stochasticity – and even with optimal data regarding clade ages, generation times or ancestral population sizes hybridization cannot be ruled out. As our approach is independent of these factors it offers a means to identify plausible cases of hybridization within the 'coalescent stochasticity zone' in which ILC cannot be rejected using simulations. This may be useful in other cases similar to our *Streptocarpus* example for which this zone is too broad to be a basis for useful inference. It could also serve as an additional independent test in groups for which the factors driving coalescence might be modelled with greater confidence, for example as a means to exclude putative hybrids (or to model them appropriately; Blanco-Pastor *et al.*, 2012) in coalescent-based inferences of species trees.

Acknowledgements

We are indebted to David Styles, Louise Badenhorst, Jan Buring, Benny Bytebier, Ernst van Jaarsveld, Liesl van der Walt and Cameron McMaster for providing plant material. We also thank the University of Stellenbosch DNA sequencing facility. This research was supported in part by funding granted to D.U.B. by the National Research Foundation, South Africa, by a Claude Leon Foundation Postdoctoral Fellowship to M.D.P., and to M.M. from the Leverhulme Trust (grant no. F/00 771 A-0735). Fieldwork was supported by the Carnegie Trust for the Universities in Scotland. The Royal Botanic Garden Edinburgh is funded by the Rural and Environment Science and Analytical Services division (RESAS) of the Scottish Government.

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716–723.

Alice LA, Eriksson T, Eriksen B, Campbell CS. 2001. Hybridization and gene flow between distantly related species of *Rubus* (Rosaceae): evidence from nuclear ribosomal DNA internal transcribed spacer region sequences. *Systematic Botany* **26**: 769–778.

Antonelli A, Humphreys AM, Lee WG, Linder HP. 2010. Absence of mammals and the evolution of New Zealand grasses. *Proceedings of the Royal Society B: Biological Sciences* **278**: 675–701.

Bart HL Jr, Clements MD, Blanton RE, Piller KR, Hurley DL. 2010. Discordant molecular and morphological evolution in buffalofishes (Actinopterygii: Catostomidae). *Molecular Phylogenetics and Evolution* **56**: 808–820.

Blanco-Pastor JL, Vargas P, Pfeil BE. 2012. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PLoS ONE* **7**: e39089.

Bloomquist EW, Suchard MA. 2010. Unifying vertical and nonvertical evolution: a stochastic ARG-based framework. *Systematic Biology* **59**: 27–41.

Buckley TR, Cordeiro M, Marshall DC, Simon C. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (Maoricicada Dugdale). *Systematic Biology* **55**: 411–425.

Chan KMA, Levin SA. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* **59**: 720–729.

Costanzo KS, Taylor DJ. 2010. Rapid ecological isolation and intermediate genetic divergence in lacustrine cyclic parthenogens. *BMC Evolutionary Biology* **10**: 166.

Cranston KA, Hurwitz B, Ware D, Stein L, Wing RA. 2009. Species trees from highly incongruent gene trees in rice. *Systematic Biology* **58**: 489–500.

Cummings MP, Neel MC, Shaw KL. 2008. A genealogical approach to quantifying lineage divergence. *Evolution* **62**: 2411–2422.

De Villiers MJ. 2008. *Phylogenetic and population genetic studies in the genus* Streptocarpus *Lindl. (Gesneriaceae DC.).* PhD thesis, University of Stellenbosch, Stellenbosch, South Africa.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* **24**: 332–340.

Demesure B, Sodzi N, Petit RJ. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* **4**: 129–131.

Douzery EJP, Pridgeon AM, Kores P, Linder HP, Kurzweil H, Chase MW. 1999. Molecular phylogenetics of Diseae (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. *American Journal of Botany* **86**: 887–899.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Drummond AJ, Rambaut A. 2006. *BEAST.* [WWW document] URL http://beast.bio.ed.ac.uk/ [accessed on 27 January 2013].

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Edwards TJ. 2003. Two new species of *Streptocarpus* (Gesneriaceae) from South Africa. *Novon* **13**: 185–188.

Edwards T, Hughes M, Möller M, Bellstedt D. 2008. New *Streptocarpus* species (Gesneriaceae) from South Africa. *Botanical Journal of the Linnean Society* **158**: 743–748.

Edwards TJ, Kunhardt C, Venter S. 1992. Gesneriaceae: notes on the genus *Streptocarpus*. *Bothalia* **22**: 192–194.

Fitch WM. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Zoology* **20**: 406–416.

Hall TA. 1999. *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.* Nucleic Acids Symposium Series, 95–98.

Hamilton MB. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* **8**: 521–523.

Harrison CJ, Möller M, Cronk QCB. 1999. Evolution and development of floral diversity in *Streptocarpus* and *Saintpaulia*. *Annals of Botany* **84**: 49–60.

Hedrick PW. 2007. Sex: differences in mutation, recombination, selection, gene flow and genetic drift. *Evolution* **61**: 2750–2771.

Heinze B. 2007. A data base of PCR primers for the study of the chloroplast genome in plants. *Plant Methods* **3**: 4.

Hilliard OM, Burtt BL. 1971. *Streptocarpus: an African plant study*. Pitermaritzburg, South Africa: University of Natal Press.

Holder MT, Anderson JA, Holloway AK. 2001. Difficulties in detecting hybridization. *Systematic Biology* **50**: 978–982.

Holland BR, Benthin S, Lockhart PJ, Moulton V, Huber KT. 2008. Using supernetworks to distinguish hybridization from lineage-sorting. *BMC Evolutionary Biology* **8**: 202.

Hughes M, MacMaster G, Möller M, Bellstedt DU, Edwards TJ. 2006. Breeding system of a plesiomorphic floral type: an investigation of small flowered *Streptocarpus* (Gesneriaceae) species. *Plant Systematics and Evolution* **262**: 13–24.

Hughes M, Möller M, Bellstedt DU, Edwards TJ, Villiers Md. 2005. Refugia, dispersal and divergence in a forest archipelago: a study of *Streptocarpus* in eastern South Africa. *Molecular Ecology* **14**: 4415–4426.

Humphreys AM, Antonelli A, Pirie MD, Linder HP. 2011. Ecology and evolution of the diaspore 'burial syndrome'. *Evolution* **65**: 1163–1180.

Joly S, McLenachan Patricia A, Lockhart Peter J. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* **174**: E54–E70.

Jong K, Burtt BL. 1975. The evolution of morphological novelty exemplified in the growth patterns of some Gesneriaceae. *New Phytologist* **75**: 297–311.

Jong K, Möller M. 2000. New chromosome counts in *Streptocarpus* (Gesneriaceae) from Madagascar and the Comoro Islands and their taxonomic significance. *Plant Systematics and Evolution* **224**: 173–182.

Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**: 17–24.

Leclère L, Schuchert P, Cruaud C, Couloux A, Manuel M. 2009. Molecular phylogenetics of *Thecata* (Hydrozoa, Cnidaria) reveals long-term maintenance of life history traits despite high frequency of recent character changes. *Systematic Biology* **58**: 509–526.

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology* **73**: 152–160.

Lumaret R, Jabbour-Zahab R. 2009. Ancient and current gene flow between two distantly related Mediterranean oak species, *Quercus suber* and *Q. ilex*. *Annals of Botany* **104**: 725–736.

Maddison WP, Maddison DR. 2009. *Mesquite: a modular system for evolutionary analysis. Version 2.72.* [WWW document] URL http://mesquiteproject.org [accessed on 27 January 2013].

Martén-Rodríguez S, Fenster CB, Agnarsson I, Skog LE, Zimmer EA. 2010. Evolutionary breakdown of pollination specialization in a Caribbean plant radiation. *New Phytologist* **188**: 403–417.

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325–1335.

Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ. 2008. The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* **57**: 466–482.

Meng C, Kubatko LS. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. *Theoretical Population Biology* **75**: 35–45.

Möller M, Brooks KJ, Hughes M. 2003. Plastid inheritance in *Streptocarpus* (Gesneriaceae) and an inferred hybrid origin for a population of *S. aff. primulifolius* from Igoda River, South Africa. *Edinburgh Journal of Botany* **60**: 389–408.

Möller M, Cronk QCB. 1997. Origin and relationships of *Saintpaulia* (Gesneriaceae) based on ribosomal DNA internal transcribed spacer (ITS) sequences. *American Journal of Botany* **84**: 956.

Möller M, Cronk QCB. 2001a. Evolution of morphological novelty: a phylogenetic analysis of growth patterns in *Streptocarpus* (Gesneriaceae). *Evolution* **55**: 918–929.

Möller M, Cronk QCB. 2001b. Phylogenetic studies in *Streptocarpus* (Gesneriaceae): reconstruction of biogeographic history and distribution patterns. *Systematics and Geography of Plants* **71**: 545–555.

Möller M, Pullan M, Kiehn M, Skog LE. 2002 onwards. *RBGE WebCyte – Gesneriaceae cytology database*. [WWW document] URL http://elmer.rbge.org.uk/webcyte/webcyteintro.php [accessed 6 October 2013].

van der Niet T, Linder HP. 2008. Dealing with incongruence in the quest for the species tree: a case study from the orchid genus *Satyrium*. *Molecular Phylogenetics and Evolution* **47**: 154–174.

Nylander JAA. 2004. *MrModeltest, version 2.3*. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. [WWW document] URL http://www.abc.se/~nylander/ [accessed 15 August 2008].

Palee P, Denduangboripant J, Anusarnsunthorn V, Möller M. 2006. Molecular phylogeny and character evolution of *Didymocarpus* (Gesneriaceae) in Thailand. *Edinburgh Journal of Botany* **63**: 231–251.

Palumbi SR, Cipriano F, Hare MP. 2001. Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution* **55**: 859–868.

Parrish TL, Koelewijn HP, Dijk PJv, Kruijt M. 2003. Genetic evidence for natural hybridization between species of dioecious *Ficus* on island populations. *Biotropica* **35**: 333–343.

Pelser PB, Kennedy AH, Tepe EJ, Shidler JB, Nordenstam B, Kadereit JW, Watson LE. 2010. Patterns and causes of incongruence between plastid and nuclear Senecioneae (Asteraceae) phylogenies. *American Journal of Botany* **97**: 856–873.

Pirie MD, Humphreys AM, Barker NP, Linder HP. 2009. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic Biology* **58**: 612–628.

Pirie MD, Humphreys AM, Galley C, Barker NP, Verboom GA, Orlovich D, Draffin SJ, Lloyd K, Baeza CM, Negritto M *et al.* 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution* **48**: 1106–1119.

Pirie MD, Oliver EGH, Bellstedt DU. 2011. A densely sampled ITS phylogeny of the Cape flagship genus *Erica* L. suggests numerous shifts in floral macromorphology. *Molecular Phylogenetics and Evolution* **61**: 593–601.

Polihronakis M. 2009. The interface between phylogenetics and population genetics: investigating gene trees, species trees, and population dynamics in the *Phyllophaga fraterna* species group. *Evolution* **64**: 1048–1062.

Rambaut A, Drummond AJ. 2003. *Tracer v. 1.5*. [WWW document] URL http://beast.bio.ed.ac.uk/Tracer [accessed 27 January 2013].

Roalson EH, Skog LE, Zimmer EA. 2003. Phylogenetic relationships and the diversification of floral form in *Achimenes* (Gesneriaceae). *Systematic Botany* **28**: 593–608.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Swofford DL. 2002. *PAUP*: Phylogenetic analysis using Parsimony (*and other methods), version 4*. Sunderland, MA, USA: Sinauer Associates.

Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105–1109.

Tripp EA, Manos PS. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). *Evolution* **62**: 1712–1737.

Weigend M, Edwards TJ. 1994. Notes on *Streptocarpus cyaneus* and *S. parviflorus*. *Sendtnera* **2**: 365–376.

Weisrock DW, Rasoloarison RM, Fiorentino I, Ralison JM, Goodman SM, Kappeler PM, Yoder AD. 2010. Delimiting species without nuclear monophyly in Madagascar's mouse lemurs. *PLoS ONE* **5**: e9883.

Wilgenbusch JC, Warren DL, Swofford DL. 2004. *AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference*. [WWW document] URL http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php [accessed 17 November 2011].

Willyard A, Cronn R, Liston A. 2009. Reticulate evolution and incomplete lineage sorting among the ponderosa pines. *Molecular Phylogenetics and Evolution* **52**: 498–511.

Wright JJ. 2009. Diversity, phylogenetic distribution, and origins of venomous catfishes. *BMC Evolutionary Biology* **9**: 282.

Yu Y, Than C, Degnan JH, Nakhleh L. 2011. Coalescent histories on phylogenetic networks and detection of hybridization despite incomplete lineage sorting. *Systematic Biology* **60**: 138–149.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Strict consensus tree showing the relationships within *Streptocarpus* reconstructed from internal transcribed spacer (ITS) sequences obtained from Möller & Cronk (2001a,b), Hughes *et al.* (2005) and the current study.

Fig. S2 One of the shortest trees reconstructed from the Fitch parsimony analysis of the South Africa (SA) internal transcribed spacer (ITS) matrix and SA plastid matrix, respectively.

Fig. S3 One of the shortest trees reconstructed from the Fitch parsimony analysis of the reduced ITS matrix and the reduced plastid matrix, respectively.

Fig. S4 One of the shortest trees reconstructed from the Fitch parsimony analysis of the reduced combined matrix.

Fig. S5 50% majority-rule consensus tree reconstructed from the Bayesian analysis of the reduced combined matrix.

Fig. S6 One of the shortest trees reconstructed from the Fitch parsimony analysis of the combined+ITS matrix and the combined+plastid matrix, respectively.

Table S1 Locality data for the South African samples included in the study

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.