Gene tree discordance, rapid diversification and convergence impact phylogenetic inference in Zygophylloideae

Pieter de Wet van der Merwe[a], Michael D. Pirie[b], & Dirk U. Bellstedt[a]*

[a] Department of Biochemistry, University of Stellenbosch, Stellenbosch, South Africa

[b] University Museum, University of Bergen, Postboks 7800, NO-5020 Bergen, Norway

Dirk U Belllstedt, Orcid ID: https://orcid.org/0000-0002-6376-4855

Pieter De Wet van der Merwe, Orcid ID: https://orcid.org/0000-0001-9315-599X

Michael D. Pirie, Orcid ID: https://orcid.org/0000-0003-0403-4470

* Author for correspondence: Dirk Bellstedt, dub@sun.ac.za

Abstract

The Zygophyllaceae subfamily Zygophylloideae includes seven genera and around 180 species distributed worldwide particularly in arid areas where they show various adaptations to drought including C_4 and CAM photosynthesis. Phylogenetic relationships of Zygophylloideae have been analysed in several studies using Sanger sequenced markers, but deeper relationships between clades have remained recalcitrant. Here, we use ion semi-conductor sequencing of chloroplast and nuclear ribosomal regions to analyse these relationships and assess why they have proved so challenging to resolve. We analyse new data for 39 chloroplast genes and nuclear ribosomal 18S, 5.8S, and 26S gene and internal transcribed spacers for key species of *Zygophyllum sensu stricto, Roepera, Tetraena, Fagonia*, *Zygophyllum stapffii* and monotypic *Augea capensis*; and *Tribulus* of the subfamily Tribuloideae as outgroup, compared this to published chloroplast genomes, summarising phylogenetic results separately from non-coding introns and intergenic spacers, photosynthetic genes and non-photosynthetic genes. We combined this data with sequences of three highly variable noncoding chloroplast regions and ITS used in previous studies of greater numbers of taxa and assessed the effect on phylogenetic inference. We used a chloroplast sequence matrix to infer a dated phylogeny. Adding data did not always improve phylogenetic resolution which was impacted by non-informative variation in chloroplast photosynthesis related genes as well as by conflicting chloroplast and nuclear signals. This provides evidence for both selection and introgressive hybridization and/or incomplete lineage sorting during an ancient rapid divergence that will require multiple independent, informative, and neutral gene trees to resolve.

Keywords: Zygophylloideae, *Fagonia*, *Tetraena*, *Augea*, *Stapffia*, *Roepera*, *Zygophyllum*, ion semi-conductor sequencing, ITS region, ITS cassette, C_3 photosynthesis, C_4 photosynthesis, CAM photosynthesis, Photosynthetic genes, Ancient rapid radiation, Phylogenetic reconstruction.

Introduction

The flowering plant family Zygophyllaceae has a world-wide distribution. They can be trees, such as species of the genus *Balanites*; shrubs such as the well-known creosote bush, *Larrea tridentata,* from the USA; or annual herbs, such as *Tribulus terrestris*. The family includes examples of tropical rainforest trees in South America, such as *Bulnesia* and *Guaiacum* which are known for their particularly hard wood. However, Zygophyllaceae are more typically characterised by adaptations to drought tolerance that have allowed the family to radiate into all arid areas of the world.

Sheahan and Chase (1996) presented the earliest molecular systematic studies of the Zygophyllaceae based on sequence data from the chloroplast *rbc*L gene. They reviewed previous morphology-based classifications of the family and re-circumscribed subfamilies Zygophylloideae, Tribuloideae, Seetzenioideae, Larreoideae and Morkillioideae (Sheahan and Chase, 1996). According to their redefinition, the subfamily Zygophylloideae consisted of the genera *Zygophyllum* L. (including ca. 150 species), *Fagonia* L. (30 species) and the monotypic genera *Augea* Thunb. and *Tetraena* Maxim. Sheahan and Chase (2000) published a more detailed hypothesis for relationships within the subfamily Zygophylloideae based on chloroplast *trn*LF and *rbc*L sequences of species of *Zygophyllum sensu lato,* of *Fagonia*, and one each of *Augea* and *Tetraena*. Beier *et al.* (2003) expanded the taxon sampling considerably and, based on non-coding chloroplast *trn*L intron sequences and a morphological character dataset, retrieved a polyphyletic genus *Zygophyllum*. They proposed a new subgeneric classification for Zygophylloideae, recognizing a more narrowly circumscribed *Zygophyllum* (referred to here as *Zygophyllum sensu stricto*), resurrecting the genus *Roepera* A.Juss.to include most members of *Zygophyllum* subgenus *Zygophyllotypus* (Van Huysteen, 1937), expanding *Tetraena* to include most members of *Zygophyllum* subgenus *Agrophyllum* (Van Huysteen, 1937), retaining *Fagonia* and monotypic genus *Augea*, and describing a new genus, *Melocarpum* Beier & Thulin. Beier *et al.* (2004) sequenced the nuclear ribosomal ITS region of all species of *Fagonia* plus Zygophylloideae outgroups to study the relationships and biogeography of *Fagonia*, the only genus in the subfamily to range into the New World. Bellstedt *et al*. (2008) expanded the taxon sampling of southern African members of the subfamily Zygophylloideae, and, based on a phylogeny derived from *trn*LF and *rbc*L sequences, mostly confirmed the results of Beier *et al.* (2003). However, they also included *Zygophyllum stapffii* from the coastal areas of the Namib desert in Namibia and Angola, retrieving this species as sister to *Augea capensis* and thereby implying paraphyly of *Zygophyllum s.s.*.

Subsequent studies also based on Sanger sequenced plastid and nuclear ribosomal sequences have expanded representation of species: Wu *et al.* (2015) sequenced the *trn*LF and ITS of *Zygophyllum sensu stricto* and *Tetraena* from Asia and combined these with published sequences of other groups to produce a combined nuclear and chloroplast phylogeny of the subfamily. Wu *et al.* (2018) sequenced *rbc*L of sixteen additional Asian species of *Zygophyllum sensu stricto* and combined them with *trn*LF and ITS. Lauterbach *et al*. (2016) focussed on relationships and adaptations to aridity within *Roepera* and *Tetraena* in southern Africa, sequencing the chloroplast *atp*I-*atp*H spacer, *trn*G intron and *trn*LF of an even greater number of taxa. Regardless of molecular and taxon sampling, all of these analyses (Bellstedt *et al.*, 2008, Wu *et al.*, 2015; Lauterbach *et al.*, 2016; Wu *et al.*, 2018) retrieved with high support clades corresponding to *Tetraena*; *Roepera*; *Fagonia* plus paraphyletic *Melocarpum* (with two *Melocarpum* species sister to *Fagonia*); *Augea capensis* plus *Zygophyllum stapffii* / *Z. orbiculatum* (the latter treated as conspecific by Bellstedt *et al.,* 2008); and the rest of *Zygophyllum sensu stricto*. Relationships between these clades remained unsupported and/or inconsistent.

There are several possible explanations for lack of support amongst the basal relationships in the subfamily Zygophylloideae. First, it could be a lack of sufficient sequence data and/or rapid lineage divergence. Molecular dating analyses have indicated an ancient rapid radiation event (Bellstedt *et al.*, 2012; Wu *et al.,* 2015; Wang *et al.,* 2018) which, depending on relaxed clock calibration, occurred c. 38-20 mya, the younger bound corresponding to the onset of global aridification during the start of the Miocene (Bellstedt *et al.*, 2012). Resolving rapid successions of speciation events requires more variable sequence data, and the modest numbers of Sanger sequenced markers used may simply have been insufficient. This deficit could be addressed using high throughput sequencing of larger numbers of chloroplast genes or whole genomes (Barret *et al.*, 2013, 2014), particularly to establish relationships between groups in ancient rapid radiations (Zhang *et al.*, 2016). Zhang *et al.* (2021) analysed complete chloroplast genomes of 24 samples of *Zygophyllum sensu stricto* occurring in China plus *Tetraena mongolica* as an outgroup. The results were promising, delivering a generally

robustly supported topology. However, due to the restricted taxon sampling the unresolved basal nodes in Zygophylloideae were not represented.

Beyond a simple lack of data, other explanations for a failure to infer relationships could involve differences in phylogenetic signal between nuclear, chloroplast and mitochondrial genomes or even within genomes. Such conflict could be the result of reticulate processes such as hybridization or demographic processes resulting in incomplete lineage sorting (De Queiroz et al., 1995; Maddison, 1997). In both cases differences between gene trees can be meaningful for interpretation of evolutionary history, but they make inferring a species tree or network more challenging (Nakhleh, 2013) and require more sequence data for effective comparison of individually resolved, independent gene trees, e.g. from targeted high throughput sequencing (Johnson et al, 2019; Larridon et al., 2020; Morales Briones et al., 2021; Yardeni et al., 2022).

Whether derived from Sanger or high throughput sequencing, functional constraints or even convergence could limit or even undermine the usefulness of particular genes (Shen et al., 2017). In the case of Zygophylloideae, those important for different modes of photosynthesis could be such an example. C_4 photosynthesis has been inferred to have evolved from C_3 ancestors numerous times independently in flowering plants in response to the reduction in atmospheric carbon dioxide (CO_2) from the Oligocene onwards (Pagani et al., 2005; Christin et al., 2011). This photosynthesis type has the advantage that atmospheric CO_2 can be captured in chemical intermediates at night, enabling the plant to close its stomata during the day avoiding moisture loss. C_4 plants can therefore survive better in arid environments. C_4 photosynthesis evolved in the subfamily Zygophylloideae in *Tetraena simplex* and independently in clades of *Tribulus*, *Tribulopsis* and *Kallstroemia* species in the sister subfamily Tribuloideae (Lauterbach *et al.*, 2019). Lauterbach *et al.* (2016) analysed adaptations in leaf morphology in the genera *Tetraena* and *Roepera* showing that the vascular bundles which typically facilitate C_4 photosynthesis only evolved in *Tetraena simplex* and not in other species in the genus or in *Roepera*. However, diverse other leaf anatomical types in *Tetraena* do suggest similar adaptations in other species. C_4 photosynthesis involves many genes both within the chloroplast genome such as the *rbc*L gene (Christin *et al.*, 2008), but also in genes that control leaf structure which therefore must be encoded in the nuclear genome. The *rbc*L gene has been used extensively in phylogenetic reconstruction at all hierarchical levels in plant phylogenetics (Chase *et al.*, 1993; APG I., 1998; APG II, 2003, APG III, 2009; APG IV, 2016). However, it is known to exhibit specific convergent mutations in C_4 lineages (Christin et al., 2008). It is important to test whether it and potentially other genes in the chloroplast genome might show convergent signals that impact phylogenetic inference in plant groups in which both C_3 and C_4 photosynthesis occurs.

Another photosynthetic mechanism that serves as an adaptation for aridity, Crassulacean acid metabolism (CAM), evolved in *Roepera cordifolia* (Matimati *et al.* 2012). It is not known whether CAM is present in other species of *Roepera*. CAM photosynthesis is also present in the New-World species *Bulnesia retama* in the subfamily Larreoideae (Mok *et al.* 2023). As CAM metabolism involves a third, very different photosynthetic mechanism for photosynthetic processes to occur in plants exposed to aridity, and is not facilitated by a Kranz type morphology as in C_4 plants but rather an intracellular adaptation, evolution towards this type of photosynthesis apparently requires mutation and expression of a different repertoire of genes, including *rbc*L, to those in plants that possess C_3 or C_4 photosynthetic mechanisms (Berry et al., 2013).

Here, we analyse numerous gene loci from both the chloroplast and nuclear genomes to try to resolve the basal relationships in the subfamily Zygophylloideae. We focus on key species of *Zygophyllum sensu stricto, Roepera, Tetraena, Fagonia*, and the monotypic lineages of *Zygophyllum stapffii/orbiculatum* and *Augea capensis*, with a representative member of the subfamily *Tribuloideae* as outgroup. We compare phylogenetic signals both between genomes and between functional groups within them (photosynthetic genes versus non-photosynthetic genes), with

comparisons to a supermatrix with denser taxon sampling, and assess the impact on results with the aim of inferring a robust and resolved phylogenetic hypothesis for the subfamily.

Materials and Methods

Chloroplast isolation, DNA purification and ion semi-conductor sequencing

Fresh leaf material was obtained from selected species in six genera in the subfamily Zygophylloideae: *Roepera foetida*, *Zygophyllum fabago*, *Zygophyllum stapffii, Tetraena turbinata* ined., *Augea capensis*, *Fagonia rangei* and of the outgroup *Tribulus terrestris* (see Supplementary Table 1). Intact chloroplasts were isolated using a Chloroplast Isolation Kit from Sigma-Aldrich. Approximately 30 g of fresh plant leaf material was macerated in a blender using the detailed instructions of the kit. A 40%/80% Percoll gradient was used to collect intact chloroplasts at the interphase of the two gradients using a Pasteur pipette. After subsequent steps the final isolated chloroplast pellet was resuspended in approximately 500 µl Chloroplast Isolation Buffer (CIB) supplied with isolation kit. This suspension was used in an adapted protocol of Doyle & Doyle, 1987 where upon 500 µl of 2XCTAB was added to lyse the intact chloroplasts without a grinding step, releasing the DNA into solution. The protocol was then followed as described. Contaminating RNA was removed using RNA digestion. The concentration of the purified DNA was determined using a NanoDrop 1000 Spectrophotometer and ion semi-conductor sequencing was performed at Central Analytical Facility (CAF), University of Stellenbosch following the manufacturer's instructions.

Chloroplast gene contiguous sequence assembly and alignments

The short sequence reads obtained from the ion semi-conductor sequencing were subjected to *de novo* assembly to generate contiguous sequences (contigs) using either Newbler (http://454.com/products/analysis-software/index.asp) or Mira (http://sourceforge.net/projects/mira-assembler/) using default settings and were encoded in fasta format. Contiguous sequences were generated from the chloroplast, the nuclear and the mitochondrial genomes.

Annotation of the generated contigs was performed in Geneious Prime 2023.1.2 using the complete *Corynocarpus laevigatus* chloroplast genome as a reference genome (NC_014807) (Atherton *et al*., 2010). This allowed for the rapid identification of chloroplast contigs as opposed to the generated mitochondrial and/or nuclear contigs. Only the chloroplast genes (41 in total) that were identified in all seven sequenced taxa were retained to exclude missing data. Nine additional whole chloroplast genomes of key species in the Zygophyllales were downloaded from GenBank. These include two specimens from the Krameriaceae; *Krameria lanceolata* (NC_043801) and *Krameria bicolor* (NC_043800), (Gonçalves *et al.*, 2019), and seven species from the family Zygophyllaceae; *Tribulus terrestris* (MK341055) (Xi, （2018) in the family subfamily Tribuloideae; *Bulnesia arborea* (see Supplementary Table 2) (Moore *et al.*, 2010), *Guaiacum angustifolium* (NC_043796) (Gonçalves *et al.*, 2019) and *Larrea tridentata* (KT272174) (Countway *et al.*, 2015) in the subfamily subfamily Larreoideae; and *Fagonia cretica* (OM952137) (Sonchhatra *et al.*, 2022), *Tetraena mongolica* (ON408223)(Yang, -2022) and *Zygophyllum xanthoxylum* (NC_052769) (Xu *et al.* 2020). From these nine complete chloroplast genomes the 41 abovementioned genes were identified and added to the analyzed alignments. Species in the genera *Fagonia, Tetraena* and *Zygophyllum* all represent deep divergences (i.e. examples of the most distantly related species) within these genera: *Fagonia cretica* and *Fagonia rangei* (Beier *et al.*, 2004); *Tetraena turbinatum* ined. and *Tetraena mongolica* (Lauterbach *et al.*, 2016) and *Zygophyllum fabago* and *Zygophyllum xanthoxylum* (representing both sections of genus *Zygophyllum*, Lauterbach *et al.*, 2016).

Further inspection of the 41 gene sequence alignments revealed that two were non-coding gene regions (*trn*A and *trn*I contain large introns) and that three of the remaining 39 coding genes contained introns (*pet*B, *pet*D, *rpo*C1). These introns were spliced out of the coding genes to not disrupt the reading frame of the protein-coding genes, but the three introns were retained and added to the other two noncoding regions to give an alignment matrix of 5 noncoding regions (referred to as 5NC). The alignment matrix of 39 coding genes (referred to as 39C) was then further subdivided into those genes directly involved in the photosynthetic process, totaling 21 (referred to as 21C), and those genes that were not directly involved in the photosynthetic process (referred to as 18C) (see Supplementary Table 2). All of these alignment matrices contained the sequences of 16 taxa (9 ingroup and 7 outgroup) and were complete, i.e. they contained no missing data.

To augment the non-coding gene sequence information in the phylogenetic analyses, three highly variable noncoding sequences (referred to as 3HVNC) in the chloroplast genome, i.e. *trn*LF *(trn*L *intron + trn*LF *spacer)*, *trn*G and *atp*I-*atp*H, one each per monophyletic species but including the northern and southern populations of *Tetraena simplex* and *Tetraena decumbens* respectively, were taken from previous studies (Sheahan and Chase, 1996 and 2000, Bellstedt *et al.*, 2008, Lauterbach *et al.*, 2016), as well as from the contiguous sequences generated by ion semi-conductor sequencing generated in this study and from the whole genome sequences downloaded from GenBank. They were combined in Geneious Prime 2023.1.2, aligned using MUSCLE v3.8.31 (Katoh & Standley, 2013) followed by manual refinement, and included in a large chloroplast gene alignment matrix consisting of 121 taxa. The sequences of the 3HVNC regions could not be utilised from the whole genome sequence of *Fagonia cretica* (OM952137) because of the poor quality of the genome assembly deposited on GenBank. In previous studies it was found that the *trn*L-*trn*F spacer region of the *trn*LF sequence region of *Augea capensis* could not be sequenced with standard primers (Taberlet *et al.*, 1991). Ion semi-conductor sequencing revealed that the *trn*F gene was inverted in *Augea capensis*. This gene sequence was then reverse-complemented and incorporated into the *trn*L-*trn*F alignment matrix. To study the effect of the addition of the 3HVNC sequences on the phylogenetic analyses, an alignment matrix consisting of these 3HVNC sequences of 15 of the 16 taxa (as those of *Fagonia cretica* were missing) of which the 39 coding genes and 6 noncoding genes were sequenced or obtained from GenBank, was generated. Alignments of increasing numbers of the 3HVNC sequences of an evenly distributed number of species in the different genera were then generated (34, 96 and 121 taxa) to study the effect of gene sequence addition to the 16-taxon coding and noncoding gene matrices on the phylogenies resulting from their combined phylogenetic analyses.

Taxon sampling and DNA extraction for nuclear gene sequencing

DNA was extracted following Doyle and Doyle (1987) from silica gel dried leaf material of *Bulnesia arborea* provided by the Royal Botanical Garden Edinburgh, UK. DNA samples of some taxa were obtained from the Royal Botanic Gardens, Kew, UK. A list of the taxa of which the ITS region (ITS1, 5.8S rRNA, ITS2) gene sequences were generated by PCR and cycle sequencing is shown in Supplementary Table 3.

Nuclear ITS cassette gene sequencing and alignments

Due to nuclear DNA contamination of the chloroplast enriched fraction, the complete nuclear ITS cassette region consisting of the 18S rRNA, ITS1, 5.8S rRNA, ITS2 and 26S rDNA gene cassette region of the seven taxa from which chloroplast sequences were generated could be retrieved from the Ion semi-conductor sequencing data. The complete ITS cassette sequences were, however, not available for *Bulnesia arborea* and several gaps were identified in the ITS cassette region sequences of *Tetraena turbinata* ined. and in *Tribulus terrestris*. The complete ITS cassette region was amplified using

specifically designed primers (primers are listed in Supplementary Table 4) from the DNA isolated from these species. PCR products were purified and sequenced using standard Sanger sequencing (CAF, Stellenbosch University). Some of the primers generated for the amplification of the whole ITS region of the *Bulnesia arborea* could be used to fill in the missing regions in the assembled contigs of *Tribulus terrestris* and *Tetraena turbinata* ined. via Sanger sequencing. The generated contiguous sequences were then used to construct a complete nuclear 18S rDNA-ITS1-5.8S rDNA-ITS2-26S rDNA cassette sequence matrix of a total of 8 taxa.

ITS gene sequencing and alignment

Amplification of the ITS region (ITS1, 5.8s RNA and the ITS2) of those taxa listed in Supplementary Table 3 (59 taxa) was achieved using primers 17SE and 26SE of Sun *et al.* (1994), also referred to as AB101 and AB102 by Douzery *et al.* (1999) (primers are listed in Supplementary Table 4). Alternatively, the primers P16 (forward, TCA CTG AAC CTT ATC ATT TAG AGG A), P17 (forward, CTA CCG ATT GAA TGG TCC GGT GAA), P25 (reverse, GGG TAG TCC CGC CTG ACC TG) 26S-82R (reverse, TCC CGG TTC GCT CGC CGT TAC TA) and 2G (reverse, GTG ACG CCC AGG CAG ACG T) were used to successfully amplify the gene region. PCR reactions were performed in an Applied Biosystems Veriti 96 well Thermal Cycler in 25 µl reactions. Each tube contained 2.5 mM $MgCl_2$, 1x JMR-455 buffer (Southern Cross Biotechnology, Cape Town, RSA), 1 U of Super-Therm Taq polymerase (Southern Cross Biotechnology, Cape Town, RSA), 200 µM of each dNTP and 0.5 µM of the forward and reverse primer respectively. Amplification: 35 cycles, 1 min at 94 °C, 1 min at 55 °C, 90 s at 72 °C, final extension, 6 min at 72 °C.

PCR products were sequenced with the primers as specified above at Stellenbosch University's Central Analytical Facility. Forward and Reverse sequencing chromatograms were superimposed and base calling corrected in ChromasPro v2.1.8 to generated consensus sequences.

Consensus sequences were imported and aligned in Geneious Prime 2023.1.2 using MUSCLE v3.8.31 (Katoh & Standley, 2013). Manual adjustments were made to ensure correct alignment. Indels were removed and treated as missing data. The sequence matrix therefore consisted of the ITS1, 5.8s RNA and the ITS2 region of 59 taxa. The ITS1, 5.8s RNA and the ITS2 region of the 8 taxa from which the full ITS cassette region was sequenced was then added to the alignment matrix of 59 taxa to give a total ITS alignment matrix of 67 taxa.

Strategies for phylogenetic analyses of sequence alignments

The following aligned matrices were analysed using the methods of phylogenetic analysis described hereafter:

1. a. The coding gene sequences (39C, 16 taxa) and subdivisions thereof into a non-photosynthesis coding gene matrix (18C, 16 taxa) and a photosynthesis coding gene matrix (21C, 16 taxa).
   b. A 5NC non-coding gene matrix of 16 taxa combined with a 3HVNC matrix in combinations of increasing taxon numbers (16, 34, 96 and 122).
   c. Combinations of these alignment matrices: Coding non-photosynthetic and non-coding sequences (18C + 5NC +3HVNC, 16 taxa), all coding and non-coding sequences (39C + 5NC +3HVNC, 16 taxa)
   d. Coding and non-coding sequences of 16 taxa with increasing numbers of the 3HVNC sequences: ((39C + 5NC, 16 taxa) +(3HVNC of 34 taxa)); ((39C + 5NC, 16 taxa) +(3HVNC of 96 taxa)); ((39C + 5NC) + (3HVNC of 121 taxa)).
2. a. The complete ITS cassette dataset (18S rRNA, ITS1, 5.8S rRNA, ITS2 and 26S rRNA) consisting of the sequences of the 6 Zygophylloideae ingroup taxa and *Bulnesia arborea* and

Tribulus terrestris as outgroups.
b. The ITS data only (ITS1, 5.8SrRNA and ITS2) of 67 taxa listed in Supplementary Table 3.
c. The combination of 2a and 2b representing the same 67 taxa.

A summarizing diagram of the aligned matrices and their phylogenetic analyses, together with the figure numbers in which their results will be presented is shown in Figure 1.

Phylogenetic analyses

All individual gene or nucleotide alignments were analysed to infer the best-fitting evolutionary model using jmodeltest v2.1.6 as implemented on the CIPRES Science Gateway (Miller *et al.* 2010). In all cases the best-fitting model was found to be GTR+G. We did not partition the alignment matrices for the phylogenetic analyses.

Concatenated alignments of the chloroplast genes and of the concatenated nuclear ITS region genes as described above were used for phylogenetic analyses using maximum likelihood (ML, RAxML 8.0, Stamatakis, 2014). The alignments used in the maximum likelihood analyses in which the number of taxa was 16 were also used for phylogenetic analysis using Bayesian Inference (BI, MrBayes 3.2.6, Ronquist and Huelsenbeck, 2003) on the CIPRES Science Gateway (Miller et al. 2010). Not all alignments were analysed under Bayesian inference because it would have been considerably more time consuming due to the size of the datasets. In all Bayesian analyses, three heated chains and one cold chain were employed, with runs initiated from random trees. Two independent runs were conducted with 10 000 000 generations per run and trees and parameters were sampled every 1 000 generations. Convergence was checked using Tracer 1.5 (Rambaut and Drummond, 2007). For each run, the first 10% of sampled trees were discarded as burn-in. Majority rule consensus trees were generated from the individual trees retrieved from the Bayesian analyses. Tree files were imported into Figtree (Rambaut, 2009) for tree drawing (http://tree.bio.ed.ac.uk/software/figtree/).

In all figures showing phylogenetic trees the genus names of taxa are abbreviated to a single letter; *Z*. for *Zygophyllum*; *T*. for *Tetraena*, and *R*. for *Roepera*, but the monotypic genus *Augea* was not abbreviated. *Zygophyllum stapffii/orbiculatum* is indicated as 'Stapffii'. To indicate the taxa of which sequences were generated by ion semi-conductor sequencing in this study, the names were followed by the suffix NGS (next generation sequencing), and those taxa of which the sequences had been obtained from whole chloroplast sequences deposited on GenBank are indicated with suffix WC or GB in the case of *Bulnesia arborea*.

Molecular clock analysis

A 15 (rather than 16) taxon 18C alignment matrix was used to date the Zygophylloideae phylogeny using BEASTv1.8.4. One *Krameria* outgroup was removed leaving a single outgroup as required using BEAST. The average age of the node between the Krameriaceae and the Zygophyllaceae of 72.2 Mya (Harris & Davies, 2016) as indicated on TimeTree (Kumar et al., 2017) was used to calibrate the phylogeny using an uncorrelated lognormal clock setting and the birth-death model with a standard deviation of 10%. The GTR+G model was also implemented. Two independent analyses were performed for 10 000 000 generations each, sampling every 1 000 generations. A burnin of 10% was applied and the trees of the two independent analyses were combined, and a maximum clade credibility tree was subsequently generated in TreeAnnotator v1.8.4. The dated phylogenetic tree was generated in FigTree v1.4.2.

Results

Chloroplast isolation and chloroplast DNA enrichment

The chloroplast isolation protocol was successful in isolating and enriching DNA from the selected taxa, as was apparent from the subsequently generated ion semiconductor sequence data. However, although the Percoll gradient chloroplast isolation procedure was aimed at the separation of intact chloroplasts from the rest of the cellular content of the plant cells with a view to generating chloroplast DNA sequences only, the generation of some mitochondrial and nuclear sequences from the isolated DNA did indicate contamination with mitochondria and nuclei. However, this apparently had no negative effect on the isolation of chloroplast DNA and additionally allowed the sequencing of the complete ITS cassette region.

NGS data

All *de novo* assembled contiguous chloroplast, nuclear and mitochondrial sequences were deposited in the Stellenbosch University data repository (Currently being submitted). Coverage of chloroplast intergenic and spacer regions and of mitochondrial sequences in general was too poor to allow complete chloroplast or mitochondrial genome construction with a high degree of accuracy. However, the data did allow the consistent retrieval of 39 chloroplast coding genes (39C) and five chloroplast noncoding gene regions (5NC) from all seven taxa (six ingroup, one outgroup) which were combined with sequences of the same genes obtained from the whole chloroplast genome of three additional ingroup and six outgroup taxa from GenBank to give gene sequences of nine ingroup and seven outgroup taxa. Alignment of these genes were then used for subsequent phylogenetic analyses.

Chloroplast sequence alignment and phylogenetic analysis data

We aligned the hypervariable chloroplast gene sequences of 121 taxa from our previous studies (Lauterbach *et al.*, 2016) i.e. *atp*I-*atp*H, *trn*LF and *trn*G spacer regions (3HVNC) with the newly obtained data for phylogenetic analyses.

The characteristics of all alignments including lengths and variability are shown in Table 1 with the topologies of the phylogenies resulting from their phylogenetic analyses.

Phylogenetic analyses of combined chloroplast gene sequences

The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 39 coding genes is shown in Figure 2. The majority-rule consensus tree from the Bayesian inference analysis generated a tree with an identical topology. The posterior probabilities from the Bayesian analysis were therefore also indicated on the nodes of the maximum likelihood tree in Figure 2. The monophyly of the subfamily Zygophylloideae and the relationships within the Zygophylloideae are strongly supported except for two nodes which were poorly supported in the maximum likelihood analysis (bootstrap support [BS] of 6 and of 41) but more strongly in the Bayesian analysis (posterior probabilities [PP] of 100 and 68 for the corresponding nodes). The relationships between *Augea*, *Zygophyllum orbiculatum/stapffii* and *Roepera* were strongly supported. Internal branches between the genera were short, whilst the terminal branches were much longer.

To test potential conflict within the dataset the 16 taxa 39C gene matrix was subdivided into the 18C gene matrix consisting of genes not directly involved in the photosynthetic process and the 21C gene matrix consisting of genes involved in the photosynthetic process. Phylogenies inferred from these matrices are shown in Figures 3 and 4 respectively. The 18C gene matrix gives a better-supported topology compared to the 39C matrix (maximum likelihood BS ≥70 and Bayesian PP ≥0.95, except for one node in the maximum likelihood phylogeny, as shown in Figure 3. By contrast, the 21C gene matrix

consisting of genes involved in the photosynthetic process shows poor support of several nodes (See Figure 4).

The phylogenetic tree inferred from the 16 taxa eight non-coding sequences (5NC and 3HVNC) is shown in Figure S1 and that of the 3HVNC sequences only in supplementary Figure S2. Both failed to resolve the focal relationships (see Table 1). The addition of increasing numbers of 3HVNC sequences (representing 34, 96 and 121 taxa, Supplementary Figures S3, S4 and S5) to the 16 taxa 5NC gene sequence matrix, resulted in weakened support values. None resulted in a fully resolved phylogeny.

Phylogenetic results of combined 16 taxa 18C gene matrix and 16 taxa 5NC and 3 HVNC gene matrices are presented in Figure S6 and that of the 16 taxa 39C coding gene matrix with the 16 taxa 5NC and 3 HVNC alignments in Figure S7. Both the addition of the 16 taxa noncoding genes (5NC and 3HVNC) as well as the 16 taxa 21C gene matrix consisting of genes directly involved in the photosynthetic process, weakened the support compared to that of the 16 taxa 18C gene matrix consisting of genes not directly involved in the photosynthetic process.

Additionally, as more sequences of the 3HVNC were added to the 16 taxa 39C and 5NC gene matrix, the BS values for focal nodes became weaker as shown in supplementary Figures S8, S9 and S10.

Nuclear ITS cassette gene sequencing and alignment

From the ion semiconductor generated data, the ITS cassette gene sequences of all six ingroup taxa and one outgroup could be obtained. The ITS cassette region genes of *Bulnesia arborea* and gaps in the generated sequences of *Tribulus terrestris* and *Tetraena turbinata* ined. were successfully sequenced using Sanger sequencing. From these sequences an alignment matrix of six ingroup and two outgroup taxa was generated.

ITS gene sequencing and alignment

The amplification of the ITS region (ITS1-5SRNA-ITS2), using Sanger sequencing, sequence assembly and alignment was successful for the 59 taxa included in the study.

Phylogenetic analyses of the nuclear rDNA gene sequences

The topologies of the phylogenetic trees of the nuclear rDNA ITS cassette region retrieved using maximum likelihood and Bayesian inference were identical but there were differences in clade support in the different analyses as shown in Figure S11. There is strong support for the monophyly of subfamily Zygophylloideae. The sister relationship of *Augea* and *Zygophyllum orbiculatum*/*stapffii* was strongly supported and this group was sister to *Roepera* with strong support. Relationships between *Zygophyllum, Fagonia* and *Tetraena* clades were not supported.

The phylogenetic tree of the nuclear rDNA ITS region is shown in Figure S12. Relationships within the genera are generally well supported. The basal resolution is also much better than the nuclear ITS cassette-based phylogeny.

Combining the ITS cassette and the ITS region data results in an overall well supported phylogeny (Figure 5) with the exception of the node between genus *Tetraena* and the clade consisting of the genera *Zygophyllum* and *Fagonia*.

Conflicts between phylogenies generated from chloroplast and nuclear gene matrices

Chloroplast and the nuclear genes tree topologies within the subfamily Zygophylloideae (Figures 3 and 5) showed conflict with regard the position of both the Asian *Zygophyllum* clade and *Tetraena*.

Molecular clock analysis

The dated phylogenetic tree generated from the 15 taxa 18C alignment matrix is shown in Figure 6. All of the nodes within the Zygophylloideae are strongly supported. The age of the family Zygophylloideae is dated at 24.75 [15.06-36.71] Mya. The genera *Fagonia*, *Zygophyllum*, *Tetraena* and the *Roepera*, *Zygophyllum orbiculatum*/*stapffii, Augea* group, radiated by 20.76 [12.06-29.01] (basal node between *Tetraena* and the *Roepera*, *Zygophyllum orbiculatum*/*stapffii, Augea* group) within the relatively short period of time of just under 4 million years indicating an ancient rapid radiation in the subfamily.

Discussion

The plant order Zygophyllales currently consists of two morphologically dissimilar plant families: Zygophyllaceae and Krameriaceae (APG IV, 2016), the latter comprising a single genus, *Krameria*. The hemi-parasitic, arid adapted, North and South American *Krameria* species show little similarity to any other flowering plant groups. It was only in 1993 that their relationship to Zygophyllaceae was discovered using chloroplastic *rbcL* sequences (Chase *et al.*, 1993). Despite being based on just a single protein coding gene, this result has stood the test of time. However, our results suggest that within Zygophyllaceae neither *rbc*L nor chloroplast-encoded photosynthesis genes in general are always reliable for phylogenetic inference.

The chloroplast genome is generally inherited uniparentally, and in the absence of recombination we might expect phylogenetic analysis of increasing numbers of chloroplast markers to give increasingly robust phylogenies (Givnish et al., 2018). However, in the subfamily Zygophylloideae, where we analysed an alignment matrix of 39 coding genes of 16 taxa representing all of the groups in the subfamily (C39, Figure 2), this was not the case. Why did the addition of many more sequences not result in a better supported phylogeny in the Zygophylloideae? With the knowledge that there are three documented photosynthesis types in the subfamily (C_3, C_4 and CAM) we removed those sequences that encode genes directly involved in the photosynthetic process (*rbc*L included) from the alignment matrix. The resultant topology (C18, Figure 3) is almost completely supported. Phylogenetic analysis of the 21 coding genes (C21, Figure 4) that were removed from the alignment matrix produces a phylogeny without discrete topological differences, but with very little support, particularly given the amount of data it represents. We did not identify specific examples of convergence associated with shifts in the mode of photosynthesis (from C_3 to C_4 and to CAM) that might cause misleading phylogenetic signal. However, our results suggest that some combination of functional constraints and selection acting on genes directly involved in the photosynthetic process hinders rather than helps us to infer the chloroplast phylogeny.

Lauterbach *et al.* (2016) showed that just three variable chloroplast gene alignments give a phylogeny with excellent resolution within genera of Zygophylloideae, but little between them (Figure S2 here generated from a re-aligned sequence matrix). Addition of the sequences of nine noncoding gene regions (16 taxa 5NC and 3HVNC) to the 16 taxa 18C gene matrix followed by phylogenetic analysis also gives a largely resolved phylogeny (Figure S1), but with inconsistent support for two apparently conflicting but very short internal branches (Figure S1 vs Figure 3). Addition of even greater numbers of 3HVNC genes to the 5NC genes of 16 taxa does not change the topology or increase support values of the spine of the phylogeny (Supplementary figure S3, S4 and S5), nor does adding more such data from additional taxa (34 taxa 3HVNC) to the 16 taxa 39C + 5NC + 3HVNC (Figure S6) phylogeny (64 taxa, Figure S7, and 122 taxa, Figure S8). Despite the potential with denser taxon sampling to improve alignments and reveal informative sites, these regions also appear uninformative at this level.

Perhaps nuclear gene sequences may be better suited for phylogenetic inference in the subfamily? Our ITS cassette region data on its own delivers a phylogeny with less support than the 18C chloroplast phylogeny but is based on far less sequence data (compare alignment lengths in Table 1) and, perhaps more importantly, fewer taxa (8 taxa as opposed to the 16 used to establish the chloroplast phylogenies). Addition of noncoding ITS sequences to coding (rRNA) gene sequences fails to resolve all the basal nodes. However, the addition of large numbers of taxa with noncoding genes to a smaller number of taxa with coding genes resulted in a large, combined phylogeny that is otherwise strongly supported (Figure 5). Thus, in contrast to what we have found when we combine similar sources of sequence data from the chloroplast genome, the addition of large numbers of ITS sequences to the smaller number of ITS cassette sequences from the nuclear genome strengthens the phylogenetic signal.

Better resolved chloroplast (Figure 3) and nuclear (Figure 5) phylogenies of Zygophylloideae reveal supported topological differences. This could reflect either ancient introgressive hybridization or incomplete lineage sorting. We have not attempted to test which of these mechanisms could be responsible: the data represent just two independent gene trees (precluding e.g. ABBA BABA frequency tests; e.g. Yardeni et al., 2025), and coalescent simulation approaches would depend on assumptions about population sizes through a long time period and would be unlikely to reject coalescence (De Villiers et al., 2013). With datasets representing few independent gene trees, causes of gene tree conflict may remain uncertain (e.g., Barret *et al.*, 2013; Bell *et al.*, 2010; Sun *et al.*, 2014, Xi *et al.*, 2014).

Both possibilities are plausible: The short internal branch lengths between different clades in the Zygophylloideae subtending long stem lineages is indicative of a rapid ancient radiation. Our dated phylogeny shows that divergence between the groups in the subfamily Zygophylloideae occurred in quick succession after 24.75 Mya (Figure 6). This coincides with the start of the worldwide Miocene aridification (Zachos *et al.*, 2001). Some of these lineages radiated into larger genera (*Fagonia, Tetraena*, *Zygophyllum* and *Roepera*) occupying a variety of habitats and habitat niches (Lauterbach *et al.*, 2016). The clade of monotypic *Augea*, *Zygophyllum orbiculatum/stapffii* and *Roepera* supported in both the nuclear and chloroplast phylogenies (when excluding photosynthesis genes) represent a southern African radiation. Other lineages are now represented by single species with widespread distributions (e.g., *Augea capensis* and *Zygophyllum orbiculatum/stapffii*). In other Zygophyllaceae subfamilies, Tribuloideae and Seetzeniodieae, there is a similar pattern of lineages that have radiated across arid biomes, as in *Tribulus* (Lauterbach et al., 2019), and individual species with wide distributions in these arid biomes, such as those of monotypic genera *Sisyndite spartea*, *Neoluederitzia sericeocarpa* and *Seetzenia lanata*.

The species with the widest geographical distribution in the Zygophylloideae (from southern- to northern Africa including the Azores) is *Tetraena simplex*. It is an annual prostrate herb with a tumbleweed seed distribution mechanism and C_4 photosynthesis. *Roepera cordifolia* that exhibits CAM photosynthesis also has a large range and has dispersed into the hyper-arid southern Namib desert in Namibia. *Tribulus terrestris*, another example of C_4, in the Tribuloideae, is distributed almost worldwide. In ancestral lineages, preadaptation to aridity potentially including arid adapted photosynthetic mechanisms may have allowed Zygophylloideae to disperse from the arid areas in southern Africa into similar arid areas in northern Africa and Asia via the African Arid Corridor, to Madagascar, and more recently into Australia by long distance dispersal (Bellstedt *et al.*, 2012). Sequence variation that reflects shifts in photosynthetic mode may therefore be directly linked to important phenomena such as drought tolerance and potential for dispersal.

Conclusion

Phylogenetic relationships within the subfamily Zygophylloideae have been a point of contention for many years, as neither multiple linked chloroplast genes nor single nuclear marker molecular studies proved conclusive. The evidence we present here shows the impact of ancient coalescent and/or reticulate processes in Zygophylloideae, and following Hahn and Nakhleh (2016), we may hence need to temper our 'exuberance' for inferring a meaningful bifurcating species tree. However, with enough data, we ought to be able to infer meaningfully resolved individual gene trees. By representing the entire geographic distribution of Zygophylloideae in analyses of multiple independent gene trees we will be able to further improve our understanding of relationships, whether they are tree-like, or not. Chloroplast genomes are an excellent source of data in this context, but we must be aware of genes under selection for different photosynthesis modes (C_3, C_4 and CAM): our results suggest that some kinds of data may impede rather than improve phylogenetic inference. They may underlie phenomena that warrant further investigation with a phylogenetic hypothesis based on other, more appropriate data.

Acknowledgements:

Dr Felix Forest, the Royal Botanic Gardens, Kew and Dr Michael Moeller, Royal Botanic Garden Edinburgh provided samples.

References

Atherton, R.A., McComish, B.J., Shepherd, L.D., Berry, L.A., Albert, N.W. and Lockhart P.J. 2010. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. *Plant Methods* **6**, 22 . https://doi.org/10.1186/1746-4811-6-22

Barrett, C.F., Davis, J.I., Leebens-Mack, J., Conran, J.G., Stevenson, D.W. 2013. Plastid genomes and deep relationships among the Commelinid monocot angiosperms. *Cladistics*. 29(1):65–87

Barrett, C.F., Specht, C.D., Leebens-Mack, J., Stevenson, D.W., Zomlefer, W.B., Davis, J.I. 2014. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales). *Ann. Bot.* 113(1):119–133

Beier, B-A., Chase, M.W., Thulin, M. 2003. Phylogenetic relationships and taxonomy of subfamily Zygophylloideae (Zygophyllaceae) based on molecular and morphological data. *Plant Syst. Evol.* 240(1-4):11–39

Beier, B-A., Nylander, J.A.A., Chase, M.W., Thulin, M. 2004. Phylogenetic relationships and biogeography of the desert plant genus *Fagonia* (Zygophyllaceae), inferred by parsimony and Bayesian model averaging. *Mol. Phylogenet. Evol.* 33(1):91–108

Bell, C.D., Soltis, D.E., Soltis, P.S. 2010. The age and diversification of the Angiosperms rerevisited. *Am. J. Bot.* 97(8):1296–1303

Bellstedt, D.U., van Zyl, L., Marais, E.M., Bytebier, B., de Villiers, C.A., Makwarela, A. M., Dreyer L. L. 2008. Phylogenetic relationships, character evolution and biogeography of southern African members of *Zygophyllum* (Zygophyllaceae) based on three plastid regions. *Mol. Phylogenet. Evol.* 47(3):932–949

Bellstedt, D.U., Galley, C., Pirie, M.D., Linder, H.P. 2012. The Migration of the Palaeotropical Arid Flora: Zygophylloideae as an Example. *Syst. Bot.* 37(4):951–959

Berry, J.O., Yerramsetty, P., Zielinski, A.M. and Mure, C.M., 2013. Photosynthetic gene expression in higher plants. *Photosynthesis Research*, *117*(1), pp.91-120.

Bult, C., Kallersjo, M., Suh, Y. 1992. Amplification and Sequencing of 16/18S rDNA from Gel-Purified Total Plant DNA. *Plant Mol. Biol. Report.* 10(3):273–284

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y.-L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., … Albert, V. A. (1993). Phylogenetics of Seed Plants: An Analysis of Nucleotide Sequences from the Plastid Gene *rbc*L. *Annals of the Missouri Botanical Garden*, *80*(3), 528–580. https://doi.org/10.2307/2399846

Christin, P.A., Salamin, N., Muasya, A.M., Roalson, E.H., Russier, F. and Besnard, G., 2008. Evolutionary switch and genetic convergence on rbcL following the evolution of C4 photosynthesis. *Molecular biology and evolution*, *25*(11), pp.2361-2368.

Christin, P.-A., Osborne, C.P., Sage, R.F., Arakaki, M., & Edwards, E.J. 2011. C4 eudicots are not younger than C4 monocots. *Journal of Experimental Botany* 62: 3171–3181. https://doi.org/10.1093/jxb/err041

Countway, A., Kane, N., Keepers, K. and LaPort, R. 2015. Direct Submission, Ecology & Evolutionary Biology, University of Colorado at Boulder, Ramaley N122 Campus Box 334, Boulder, CO 80309, USA (07-JUL-2015)

Cronquist, A. 1981. *An Integrated System of Classification of Flowering Plants*. Columbia University Press. pp. 707–708

de Jussieu, A.L. (1789). *Genera Plantarum*. Paris: Herissant & Barrois. pp. 234–235

De Queiroz, A., Donoghue, M.J., Kim, J. 1995. Separate Versus Combined Analysis of. Annual Review of Ecology and Systematics 26: 657–681. https://doi.org/10.1146/annurev.es.26.110195.003301

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of leaf tissue. Phytochem. Bull. 19, 11–15.)

Douzery, E.J.P., Pridgeon, A.M., Kores, P., Linder, H.P., Kurzweil, H.,Chase, M.W. 1999. Molecular phylogenetics of Diseae (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. Am. J. Bot. 86,887–899.

De Villiers, M.J., Pirie, M.D., Hughes, M., Möller, M., Edwards, T.J. and 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 Phytologist 198, 284-300.

Givnish, T. J., A. Zuluaga, D. Spalink, M. Soto Gomez, V. K. Y. Lam, J. M. Saarela, C. Sass, et al. 2018. Monocot plastid phylogenomics, timeline, net rates of species diversification, the power of multi-gene analyses, and a functional model for the origin of monocots. *American Journal of Botany* 105(11): 1888–1910.

Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. D., Shimizu, G. H., and Jansen, R. K. (2019). Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. Mol. Phylogenet. Evol. 138, 219–232. doi: 10.1016/j.ympev.2019.05.022

Hahn, M.W., Nakhleh, L. (2016) Irrational exuberance for resolved species trees. Evolution 70: 7–17. https://doi.org/10.1111/evo.12832

Harris, L.W., Davies, T.J. A Complete Fossil-Calibrated Phylogeny of Seed Plant Families as a Tool for Comparative Analyses: Testing the 'Time for Speciation' Hypothesis. PLoS One. 2016 Oct 5;11(10):e0162907. doi: 10.1371/journal.pone.0162907. Erratum in: PLoS One. 2017 Feb 21;12(2):e0172816. doi: 10.1371/journal.pone.0172816. PMID: 27706173; PMCID: PMC5051821.

Huelsenbeck, J.P., Ronquist, F. 2001. MRBAYES_: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17(8):754–755

Ismael, A.M.A., 1983. Some factors controlling the water economy of *Zygophyllum qatarense* growing in Qatar. J. Arid Environ. 6, 239–246.

Johnson, M.G., Pokorny, L., Dodsworth, S., Botigué, L.R., Cowan, R.S., Devault, A., Eiserhardt, W.L., Epitawalage, N., Forest, F., Kim, J.T., Leebens-Mack, J.H., Leitch, I.J., Maurin, O., Soltis, D.E., Soltis, P.S., Wong, G.K., Baker, W.J., Wickett, N.J. (2019) A Universal Probe Set for Targeted Sequencing of 353 Nuclear Genes from Any Flowering Plant Designed Using k-Medoids Clustering. Systematic Biology 68: 594–606. https://doi.org/10.1093/sysbio/syy086

Kuzoff, R.K., Sweere, J.A., Soltis, D.E., Soltis, P.S., Zimmer, E.A. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Mol. Biol. Evol.* 15(3):251–263

Larridon, I., Villaverde, T., Zuntini, A. R., Pokorny, L., Brewer, G. E., Epitawalage, N., Fairlie, I., Hahn, M., Kim, J., Maguilla, E., Maurin, O., Xanthos, M., Hipp, A.L., Forest, F., Baker, W. J. 2020. Tackling Rapid Radiations With Targeted Sequencing, Frontiers in Plant Science, 10 – 2019 DOI=10.3389/fpls.2019.01655

Lauterbach, M., van der Merwe, P.D.W., Keßler, L., Pirie, M.D., Bellstedt, D.U., Kadereit, G. (2016). "Evolution of leaf anatomy in arid environments – A case study in southern African *Tetraena* and *Roepera* (Zygophyllaceae)". Molecular Phylogenetics and Evolution. 97: 129–144. doi:10.1016/j.ympev.2016.01.002

Lauterbach, M., Zimmer, R., Alexa, A.C., Adachi, S., Sage, R., Sage, T. et al. (2019) Variation in leaf anatomical traits relates to the evolution of C4 photosynthesis in Tribuloideae (Zygophyllaceae). Perspectives in Plant Ecology, Evolution and Systematics, 39, 125463. https://doi.org/10.1016/j.ppees.2019.125463.

LogCombiner v1.8.0 of the BEAST v1.8.0 software package (45)

Maddison, W.P. (1997) Gene trees in species trees. Systematic Biology 46: 523–536. https://doi.org/10.1093/sysbio/46.3.523

Matimati, I., Musil, C.F., Raitt, L., February, E.C. 2012. Diurnal stem diameter variations show CAM and C3 photosynthetic modes and CAM–C3 switches in arid South African succulent shrubs. Agric. For. Meteorol. 161:72–79

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8. DOI: 10.1109/GCE.2010.5676129 PDF.

Mok, D., Leung, A., Searles, P., Sage, T.L., Sage, R.F. CAM photosynthesis in *Bulnesia retama* (Zygophyllaceae), a non-succulent desert shrub from South America. Ann Bot. 2023 Nov 25;132(4):655-670. doi: 10.1093/aob/mcad114. PMID: 37625031; PMCID: PMC10799978.

Moore, M. J. , Soltis, P. S., Bell, C. D., Burleigh, J. G. and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots, Proc. Natl. Acad. Sci. U.S.A. 107 (10) 4623-4628, https://doi.org/10.1073/pnas.0907801107

Morales-Briones, D.F., Kadereit, G., Tefarikis, D.T., Moore, M.J., Smith, S.A., Brockington, S.F., Timoneda, A., Yim, W.C., Cushman, J.C. and Yang, Y., 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae sl. *Systematic Biology*, *70*(2), pp.219-235.

Nakhleh, L. (2013) Computational approaches to species phylogeny inference and gene tree reconciliation. Trends in ecology & evolution 28: 719–28. https://doi.org/10.1016/j.tree.2013.09.004

Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. 2005. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. Science 309, 600– 603. (doi:10.1126/science.1110063)

SequenceMatrix v1.7.8 (204)

Sheahan, M.C., Chase, M.W. 1996. A phylogenetic analysis of Zygophyllaceae R. Br. based on morphological, anatomical and *rbc*L DNA sequence data. *Bot. J. Linn. Soc.* 122:279–300

Sheahan, M.C., Chase, M.W. 2000. Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions , with special emphasis on Zygophylloideae phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions , with special Emp. *Syst. Bot.* 25(2):371–384Beier et al (2003)

Shen, X., Hittinger, C.T., Rokas, A. (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nature Ecology & Evolution 1: 0126. https://doi.org/10.1038/s41559-017-0126

Sonchhatra, N. P., Monpara, J. K., Bhatt, P. P. and Thaker, V. S. (2022). The complete chloroplast genome of *Fagonia cretica*, Direct Submission, Department of Biosciences, Saurashtra University, University Road, Rajkot, Gujarat 360311, India (09-MAR-2022)

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313

Sun, Y., D. Z. Skinner, G. H. Liang SHH. 1994. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. *Theor. Appl. Genet.* 89:26–32

Sun, M., Soltis, D.E., Soltis, P.S., Zhu, X., Burleigh, J.G. and Chen, Z., 2015. Deep phylogenetic incongruence in the angiosperm clade Rosidae. *Molecular phylogenetics and evolution*, *83*, pp.156-166.

Swofford, D.L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). *Sinauer Assoc. Sunderland, Massachusetts*

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

The Angiosperm Phylogeny Group. An Ordinal Classification for the Families of Flowering Plants. *Annals of the Missouri Botanical Garden* 85, no. 4 (1998): 531–53. https://doi.org/10.2307/2992015.

Takhtajan, A. 1997. *Diversity and Classification of Flowering Plants*. Columbia University Press. pp. 319–320

Taubert, P. 1892. In: *Engler & Prantl, Die Natürlichen Pflanzenfamilien*, Vol. III, Leipzig: Engelmann. pp. 347–348

The Angiosperm Phylogeny Group, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II, *Botanical Journal of the Linnean Society*, Volume 141, Issue 4, April 2003, Pages 399–436, https://doi.org/10.1046/j.1095-8339.2003.t01-1-00158.x

The Angiosperm Phylogeny Group, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III, *Botanical Journal of the Linnean Society*, Volume 161, Issue 2, October 2009, Pages 105–121, https://doi.org/10.1111/j.1095-8339.2009.00996.x

The Angiosperm Phylogeny Group, M. W. Chase, M. J. M. Christenhusz, M. F. Fay, J. W. Byng, W. S. Judd, D. E. Soltis, D. J. Mabberley, A. N. Sennikov, P. S. Soltis, P. F. Stevens, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV, *Botanical Journal of the Linnean Society*, Volume 181, Issue 1, May 2016, Pages 1–20, https://doi.org/10.1111/boj.12385

Van Zyl, L. 2000. *A Systematic Revision of Zygophyllum (Zygophyllaceae) in the Southern African Region.* Ph D thesis, University of Stellenbosch

Wang, Q., Wu, S., Su, X., Zhang, L., Xu, X., Lyu, L., Cai, H., Shrestha, N., Liu, Y., Wang, W., and Wang, Z. (2018) Niche conservatism and elevated diversification shape species diversity in drylands:

evidence from Zygophyllaceae. Proc. R. Soc. B. 285, 20181742. http://doi.org/10.1098/rspb.2018.1742

Wu, S-D., Lin, L., Li, H-L., Yu, S-X., Zhang, L-J.,Wang, W. (2015) Evolution of Asian Interior Arid-Zone Biota: Evidence from the Diversification of Asian Zygophyllum (Zygophyllaceae). PLoS ONE 10(9):e0138697. doi:10.1371/journal.pone.0138697

Wu, S.D., Zhang, L.J., Lin, L. Yu, S-X., Chen, Z-D.,Wang, W. 2018.Insights into the historical assembly of global dryland floras: the diversification of Zygophyllaceae. BMC *Evol Biol* 18, 166. https://doi.org/10.1186/s12862-018-1277-z

Xi, Z., Liu, L., Rest, J.S., Davis, C.C. 2014. Coalescent versus Concatenation Methods and the Placement of *Amborella* as Sister to Water Lilies. *Syst. Biol.* 63(6):919– 932

Xu, H., Fu, W. and Xie, W. 2021. The complete chloroplast genome sequences of Sarcozygium xanthoxylum and Zygophyllum fabago from Zygophyllaceae, Direct Submission, National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA, School of Life Sciences Fudan University, Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering, 2005 Songhu Rd., Shanghai, Shanghai 200438, P. R. China (09-FEB-2021)

Yang, Y. 2022. The complete chloroplast genome of *Tetraena mongolica*, Direct Submission, Baotou Teachers'College School of Biological Science and Technology, KEXUE street 3, Baotou, Inner Mongolia, China

Yardeni, G., Viruel, J., Paris, M., Hess, J., Groot Crego, C., de La Harpe, M., Rivera, N., Barfuss, M.H.J., Till, W., Guzmán-Jacob, V., Krömer, T., Lexer, C., Paun, O., Leroy, T. 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. Molecular Ecology Resources 22: 927–945. https://doi.org/10.1111/1755-0998.13523

Yardeni G, Barfuss MHJ, Till W, Thornton MR, Groot Crego C, Lexer C, Leroy T, Paun O 2025. The Explosive Radiation of the Neotropical Tillandsia Subgenus Tillandsia (Bromeliaceae) Has Been Accompanied by Pervasive Hybridization. Systematic Biology: syaf039. https://doi.org/10.1093/sysbio/syaf039

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K. 2001. Trends , Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science (80-.).* 292(5517):686–693

Zha, X. 2018. Complete Chloroplast Genome of *Tribulus terrestris*, Direct Submission, College of Life and Environmental Sciences, Minzu University of China, #27, Zhongguancun South Street, Haidian District, Beijing, Beijing 100081, China (26-DEC-2018)

Zhang, N., Wen, J. and Zimmer, E.A., 2016. Another look at the phylogenetic position of the grape order Vitales: chloroplast phylogenomics with an expanded sampling of key lineages. *Molecular Phylogenetics and Evolution*, *101*, pp.216-223.

Zhang, L., Wang, S., Su, C., Harris, A.J., Zhao, L., Su, N., Wang, J-R., Duan, L., Chang, Z-Y. 2021. Comparative Chloroplast Genomics and Phylogenetic Analysis of *Zygophyllum* (Zygophyllaceae) of China. Front. *Plant Sci*. 12:723622. doi: 10.3389/fpls.2021.723622

Table 1: A summary of sequence variation of the different gene regions that were sequenced and the topology of the phylogenies generated from them.

Informative character and phylogenetic analysis summary								
Alignment matrix	Total characters	Constant	Uninforma-tive	Informative	Vari-able	% vari-ability	Figure	Topology
Chloroplast coding								
16 taxa Combined 39C	37764	30789	1729	5246	6975	18.47	1	F(Z(T(RSA)))
16 taxa Non-Photosynthesis 18C genes	20253	16194	1033	3026	4059	20.04	2	F(Z(T(RSA)))
16 taxa Photosynthesis 21C genes	17511	14595	696	2220	2916	16.65	3	T(RF)((SA)Z)
Chloroplast noncoding								
16 taxa noncoding 6NC & 16 taxa 3HVNC	9422	6875	743	1804	2547	27.03	S1	T(Z(F(RSA)))
121 taxa, noncoding 3HVNC	4162	2689	423	1050	1473	35.39	S2	T(Z(F(RSA)))
16 taxa noncoding 6NC & 34 taxa 3HVNC	9422	6794	802	1826	2628	27.89	S3	T(Z(F(RSA)))
16 taxa noncoding 6NC & 96 taxa 3HVNC	9422	6645	843	1934	2777	29.47	S4	T(Z(F(RSA)))
16 taxa noncoding 6NC & 121 taxa 3HVNC	9422	6606	858	1958	2816	29.89	S5	T(Z(F(RSA)))
Chloroplast coding & noncoding combinations								
16 taxa coding Non-Photo 18C & 16 taxa noncoding 6 NC & 3HVNC	29675	23133	1829	4713	6542	22.05	S6	F(Z(T(RSA)))
16 taxa coding Combined 39C & 16 taxa noncoding 6NC & 3HVNC	47186	37728	2525	6933	9458	20.04	S7	F(Z(T(RSA)))
16 taxa coding Combined 39C & 16 taxa noncoding 6NC & 34 taxa noncoding 3HVNC	47186	37583	2531	7072	9603	20.35	S8	T((FZ)(RSA))
16 taxa coding combined 39C & 16 taxa noncoding 6NC & 96 taxa noncoding 3HVNC	47186	37434	2572	7180	9752	20.67	S9	F((ZT)(RSA))
16 taxa coding combined 39C & 16 taxa noncoding 6NC & 122 taxa noncoding 3HVNC	47186	37395	2587	7204	9791	20.75	S10	F((ZT)(RSA))
Nuclear coding								
8 taxa ITS cassette (18S & ITS1 & 5.8S & ITS2 & 26S)	5930	5204	422	284	706	11.9	S11	(Z(FT))(RSA)
Nuclear noncoding								
67 taxa ITS (ITS1 & 5.8S & ITS2)	819	365	96	358	454	55.4	S12	(T(FZ))(RSA)
Nuclear coding & noncoding combined								
8 taxa ITS cassette & 59 taxa ITS	5930	5073	351	506	857	14.4	4	(T(FZ))(RSA)

Figure 1: A diagram summarizing the aligned matrices generated in this study and their phylogenetic analyses. The matrices of which the phylogenetic analyses are presented as figures and supplementary figures are highlighted in light blue and yellow respectively. The figures and supplementary figure numbers as presented in the results are shown in brackets.

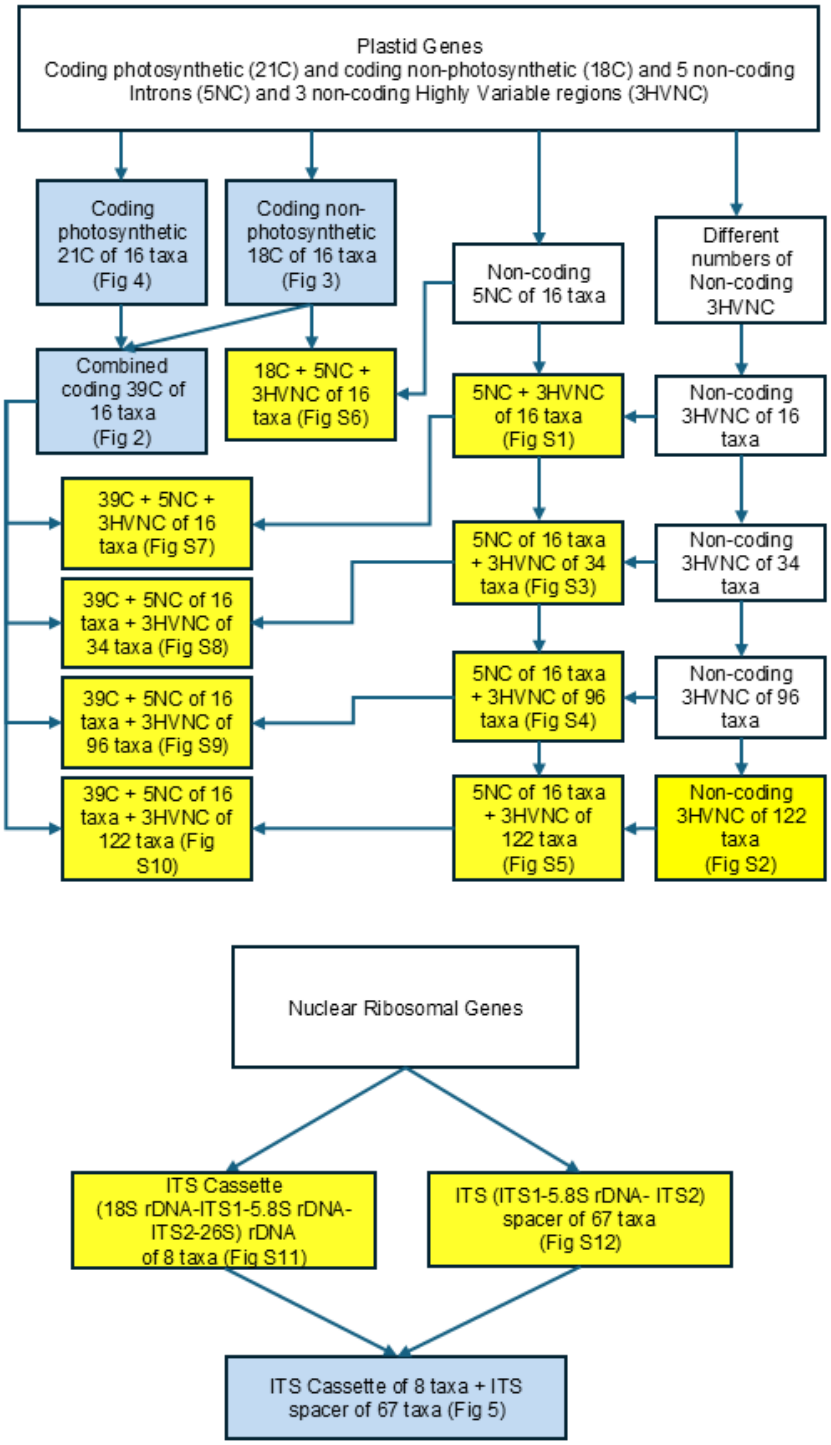


Figure 2: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 39 coding genes alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

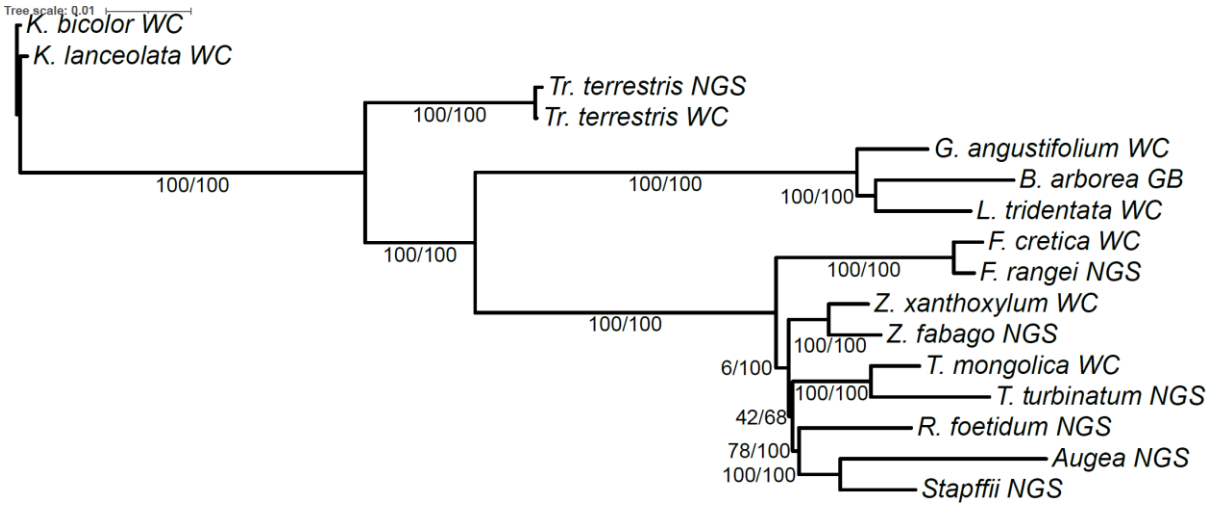


Figure 3: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 18 coding genes alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

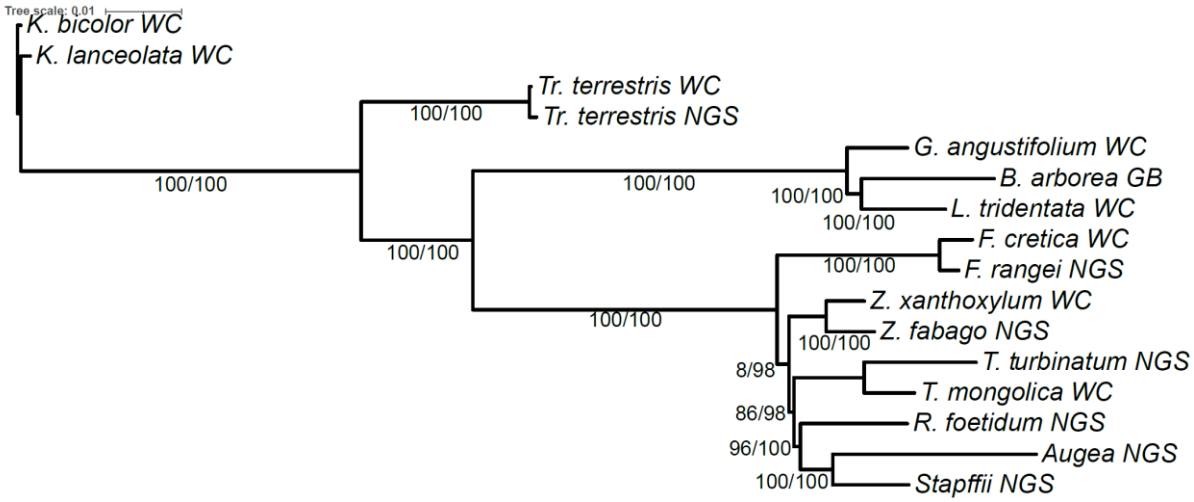


Figure 4: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 21 coding genes alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

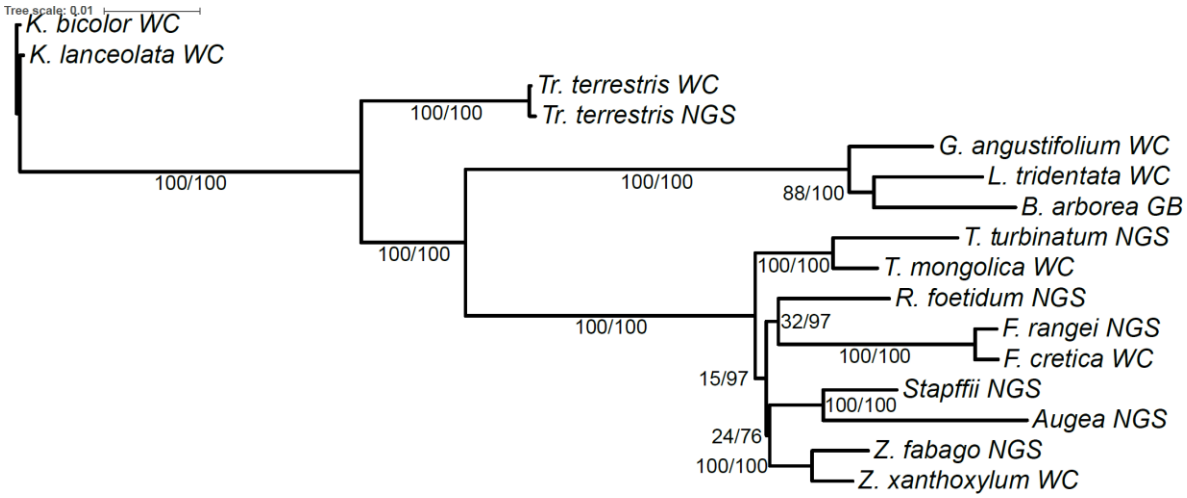


Figure 5: The most likely phylogenetic tree from the maximum likelihood analysis of the 67 taxa ITS cassette and ITS region combined alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes. Combined cassette and ITS of 67 taxa.

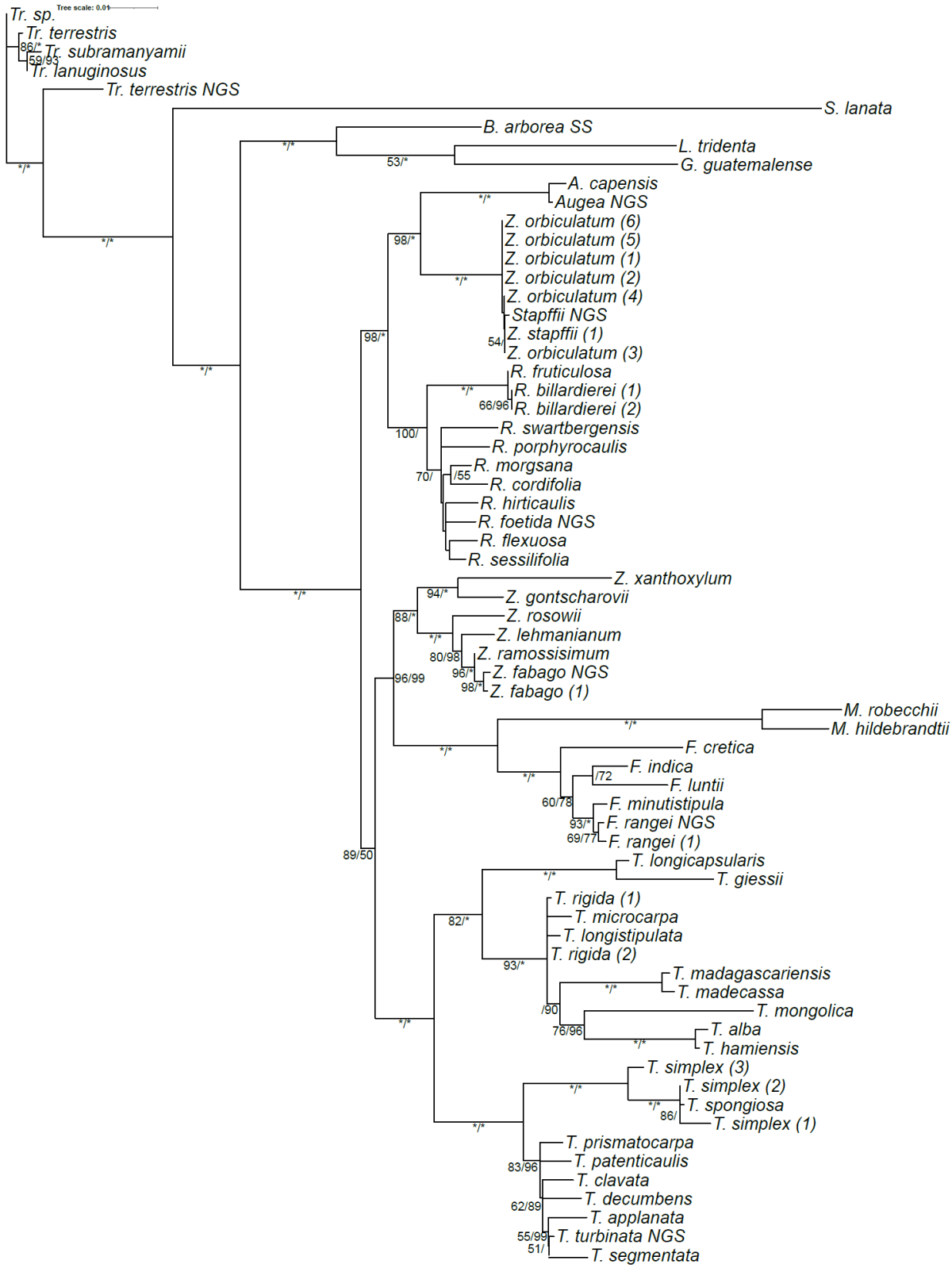
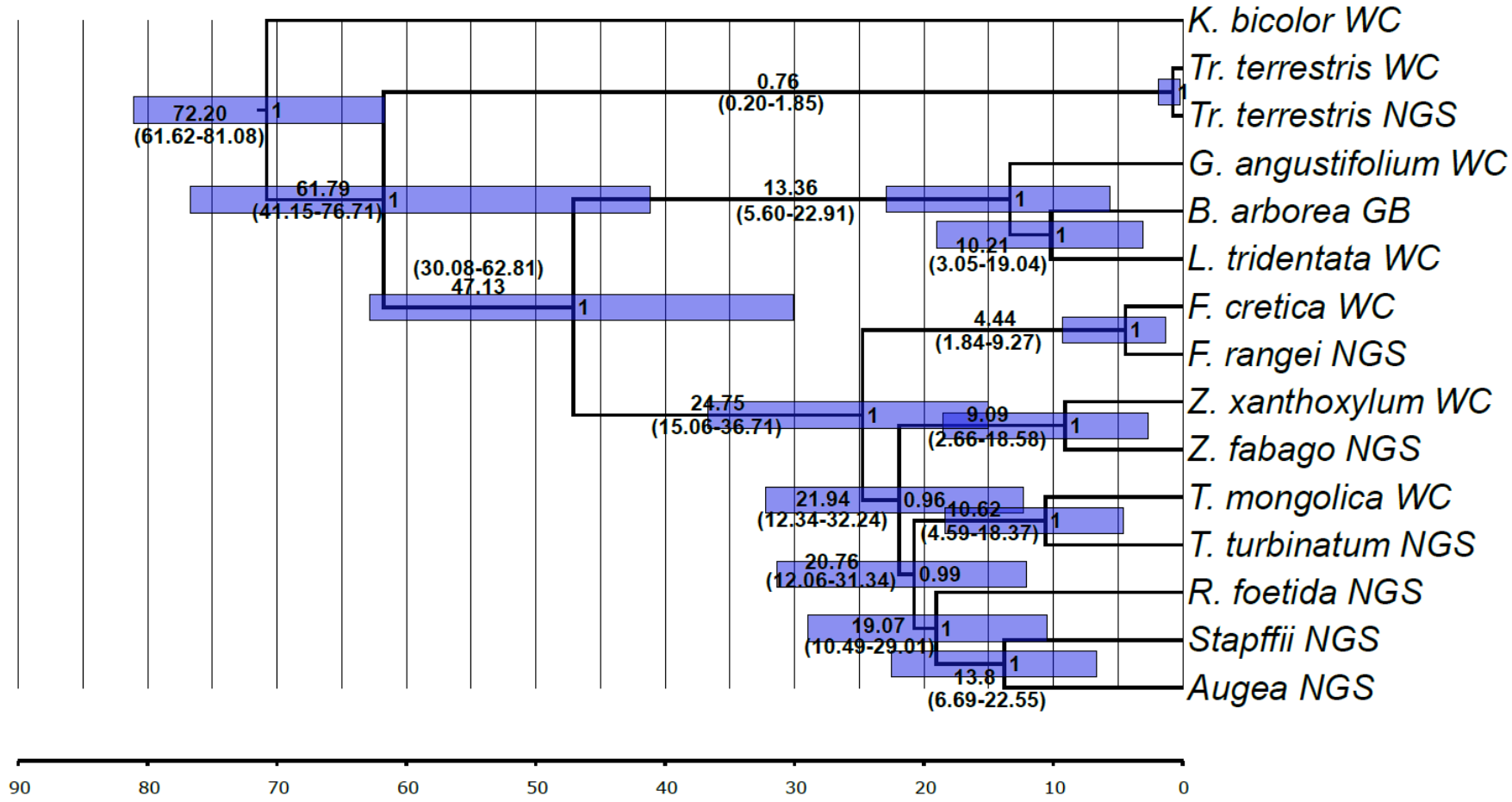


Figure 6: Maximum clade credibility tree inferred from BEAST based on the 15 taxa 18 coding genes alignment matrix to estimate the age of the Zygophylloideae radiation. Node support is indicated with posterior probability values, and the bars around node ages indicate 95% highest posterior density intervals.



Supplementary Tables

Supplementary Table 1: Collected details

Species	**Collection locality**	**Country**	**Collectors**
Augea capensis	Tankwa Karoo near Ceres, Western Cape, South Africa. (Bellstedt, 934)	South Africa	Dirk U. Bellstedt & P. De Wet van der Merwe
Fagonia rangei	Close to embankment of the Orange River near Vioolsdrif border outpost between South Africa and Namibia. (Namibian side) (Bellstedt, 1314)	Namibia	Dirk U. Bellstedt
Roepera foetida	Bridge across the Breede River between Robertson and McGregor Western Cape, South Africa (Robertson side) (Bellstedt 1387)	South Africa	P. De Wet van der Merwe
Tetraena turbinata ined.	Tankwa Karoo near Ceres, Western Cape, South Africa. (Bellstedt, 1386)	South Africa	Dirk U. Bellstedt & P. De Wet van der Merwe
Tribulus terrestris	Stellenbosch Train station, Western Cape, South Africa. (P.D.W. van der Merwe, s.n.)	South Africa	Dirk U. Bellstedt & P. De Wet van der Merwe
Zygophyllum fabago	Germany (Botanical Garden, Mainz University), Native to the Middle East.	Germany	Michael D. Pirie
Zygophyllum stapffii	Near Swakopmund, Namibia. (Bellstedt, 1222)	Namibia	Dirk U. Bellstedt

Supplementary Table 2: Genbank accession codes of gene sequences used, currently being submitted.

	Chloroplast									
	Photosynthetic Genes									
Number	**Gene abbreviation**	**Gene name**	**Genbank Accessions**							
1	*atp*B (protein-coding)	ATPase beta subunit	***Tribulus terre-stris***	***Bulnesia arborea***	***Fagonia rangei***	***Tetraena turbinata* ined.**	***Zygo-phyllum fabago***	***Roepera foetida***	***Augea capensis***	***Zygo-phyllum stapffii***
2	*atp*E (protein-coding)	ATPase epsilon subunit								
3	*atp*I (protein-coding)	ATP synthase CF0 A subunit								
4	*cem*A (protein-coding)	Chloroplast envelope membrane protein (ycf10)								
5	*pet*A (protein-coding)	Cytochrome f								
6	*pet*B (protein-coding)	Cytochrome b6								
7	*pet*D (protein-coding)	Cytochrome b6/f complex subunit IV								
8	*pet*G (protein-coding)	Cytochrome b6/f complex subunit V								
9	*pet*L (protein-coding)	Cytochrome b6/f complex subunit VI								
10	*psa*A (protein-coding)	Photosystem I P700 apoprotein A1								
11	*psa*B (protein-coding)	Photosystem I P700 apoprotein A2								
12	*psa*J (protein-coding)	Photosystem I reaction center subunit IX								
13	*psb*A (protein-coding)	Photosystem II 32 kDa protein								
14	*psb*B (protein-coding)	Photosystem II 47 kDa protein								
15	*psb*C (protein-coding)	Photosystem II 44 kDa protein								
16	*psb*D (protein-coding)	Photosystem II D2 protein								
17	*psb*H (protein-coding)	Photosystem II 10 kDa Phosphoprotein								
18	*psb*N (protein-coding)	Photosystem II N protein								
19	*psb*T (protein-coding)	Photosystem II T protein								

20	*psb*Z (protein-coding)	Photosystem II reaction center Z protein								
21	*rbc*L (protein-coding)	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit								
	Non-Photosynthetic Genes									
Number	**Gene abbreviation**	**Gene name**								
1	*mat*K (protein-coding)	maturase K								
2	*rpl*14 (protein-coding)	Ribosomal protein L14								
3	*rpl*16 (protein-coding)	Ribosomal protein L16								
4	*rpl*20 (protein-coding)	Ribosomal protein L20								
5	*rpo*A (protein-coding)	RNA polymerase alpha subunit								
6	*rpo*B (protein-coding)	RNA polymerase beta subunit								
7	*rpo*C1 (protein-coding)	RNA polymerase beta I subunit								
8	*rpo*C2 (protein-coding)	RNA polymerase beta II subunit								
9	*rps*11 (protein-coding)	Ribosomal protein S11								
10	*rps*12 (protein-coding)	Ribosomal protein S12								
11	*rps*14 (protein-coding)	Ribosomal protein S14								
12	*rps*18 (protein-coding)	Ribosomal protein S18								
13	*rps*2 (protein-coding)	Ribosomal protein S2								
14	*rps*8 (protein-coding)	Ribosomal protein S8								
15	*rrn*16 (RNA-coding)	16S ribosomal RNA								
16	*rrn*23 (RNA-coding)	23S ribosomal RNA								
17	*rrn*4.5 (RNA-coding)	4.5S ribosomal RNA								
18	*rrn*5 (RNA-coding)	5S ribosomal RNA								
Number	**Non-Coding**									
	Gene abbreviation	Gene name								
1	*pet*B (intron)	pet B intron								

2	*pet*D (intron)	pet D intron								
3	*rpo*C1 (intron)	rpoC1 intron								
4	*trn*I (intron)	trnA (Isoleucine)								
5	*trn*G (intron)	trnG (Glycine) intron								
	Nuclear									
Number	**Used name in text**	**Genetic region covered**								
1	ITS Cassette	18S rRNA-ITS1-5.8S rRNA-ITS2-26S rRNA								
2	ITS region	ITS1-5.8S rRNA-ITS2								

Supplementary Table 3: Collections details of all species collected or used to generate sequences in this study.

Number	Species	Name on trees	Geographic distribution	Voucher details	Collection locality
1	*Augea capensis*	*A. capensis*	South Africa	Bellstedt (STE)	Calitzdorp, South Africa
2	*Augea capensis*	*Augea NGS*	South Africa	Bellstedt 934	Tankwa Karoo near Ceres, Western Cape
3	*Bulnesia arborea*	*B. arborea SS*	New World	Michael Moeller	Royal Botanical Gardens, Edinburgh
4	*Fagonia cretica*	*F. cretica*	Canary Islands	Beier 125 (UPS)	
5	*Fagonia indica*	*F. indica*	United Arab Emirates	Thulin *et al.* 10024 (UPS)	
6	*Fagonia luntii*	*F. luntii*	Yemen	Thulin *et al.* 9881 (UPS)	
7	*Fagonia minutistipula*	*F. minutistipula*	Namibia	Giess and Muller 13952	
8	*Fagonia rangei*	*F. rangei (1)*	South Africa	Leistner 3388	
9	*Fagonia rangei*	*F. rangei NGS*	Namibia	Bellstedt 1314	Close to embankment of the Orange River near Vioolsdrif border outpost between South Africa and Namibia, South Africa
10	*Guaiacum guatemalense*	*G. guatemalense*	New World	Chase 640 (K)	
11	*Larrea tridentata*	*L. tridentata*	New World	Chase 636 (K)	
12	*Melocarpum hildebrandtii*	*M. hildebrandtii*	Somalia	Thulin *et al.* 9012 (UPS)	
13	*Melocarpum robecchii*	*M. robecchii*	Yemen	Thulin *et al.* 9537 (UPS)	
14	*Roepera billardierei* (1)	*R. billardierei* (1)	Australia	SR.417 (Adelaide BG.)	Adelaide
15	*Roepera billardierei* (2)	*R. billardierei* (2)	Australia	2287	
16	*Roepera cordifolia*	*R. cordifolia*	South Africa	Marais 446 (STE)	Koekenaap, South Africa
17	*Roepera flexuosa*	*R. flexuosa*	South Africa	Bellstedt 794 (STE)	Langebaan, WC, South Africa

18	*Roepera foetida*	*R. foetida NGS*	South Africa	Bellstedt 1387	Bridge across the Breede River between Robertson and McGregor Western Cape, (Robertson side), South Africa
19	*Roepera fruticulosa*	*R. fruticulosa*	W. Australia	Chase 2203 (K)	Port Gregory, North of Perth
20	*Roepera hirticaulis*	*R. hirticaulis*	Namibia	Craven 2857	
21	*Roepera morgsana*	*R. morgsana*	South Africa	Bellstedt 890 (STE)	Steinkopf, NC, South Africa
22	*Roepera porphyrocaulis*	*R. porphyrocaulis*		Bellstedt 800 (STE)	
23	*Roepera sessilifolia*	*R. sessilifolia*	South Africa	Marais 434 (STE)	Morreesburg, WC, South Africa
24	*Roepera swartbergensis*	*R. swartbergensis*	South Africa	Bellstedt 798 (STE)	Swartberg Pass, WC, South Africa
25	*Seetzenia lanata*	*S. lanata*	South Africa	Bellstedt 938	Koue Bokkeveld, Western Cape, South Africa
26	*Tetraena alba*	*T. alba*	Yemen	Thulin *et al.* 7977 (1697) (K)	
27	*Tetraena applanata*	*T. applanata*	Namibia	Bellstedt 870 (STE)	Rosh Pinah, Namibia
28	*Tetraena clavata*	*T. clavata*	Namibia	Bellstedt 878 (STE)	Lüderitz peninsula, Namibia
29	*Tetraena giessii*	*T. giessii*	Namibia	Bellstedt 1323	Arimas Farm, Witputz, Namibia
30	*Tetraena giessii*	*T. giessii*	Namibia	Bellstedt 1326	Arimas Farm, Witputz, Namibia
31	*Tetraena hamiensis*	*T. hamiensis*	Yemen	Thulin and Berer 9840 (UPS)	
32	*Tetraena longicapsularis*	*T. longicapsularis*	Namibia	Bellstedt 879 (STE)	Lüderitz peninsula. Namibia
33	*Tetraena longistipulata*	*T. longistipulata*		Bellstedt 1322	Arimas Farm, Witputz, Namibia
34	*Tetraena madagascariensis*	*T. madagascariensis*	Madagascar	Bellstedt 1239	Tulear district, Madagascar
35	*Tetraena madecassa*	*T. madecassa*	Madagascar	Bellstedt 1243	Taolagnaro District, Madagascar
36	*Tetraena microcarpa*	*T. microcarpa*	Namibia	Bellstedt 1315	Arimas Farm, Witputz, Namibia
37	*Tetraena mongolica*	*T. mongolica*	Mongolia	Sheahan 1105	Dung Kou, Inner Mongolia

38	*Tetraena patenticaulis*	*T. patenticaulis*	Namibia	Bellstedt 868 (STE)	Rosh Pinah, Namibia
39	*Tetraena prismatocarpa*	*T. prismatocarpa*	Namibia	Bellstedt 863 (STE)	Dumusibriver, near Orangeriver, Namibia
40	*Tetraena rigidum*	*T. rigidum* (1)	Namibia	Bellstedt 852 (STE)	Noordoewer, Namibia
41	*Tetraena rigidum*	*T. rigidum* (2)	Namibia	Bellstedt 855 (STE)	Aussenkehr, Namibia
42	*Tetraena segmentata*	*T. segmentata*	Namibia	Bellstedt 861 (STE)	Rosh Pinah, Namibia
43	*Tetraena simplex*	*T. simplex* (1)	Namibia	Bellstedt 1226 (STE)	Grünau, Namibia
44	*Tetraena simplex*	*T. simplex* (2)	Namibia	Bellstedt 1224 (STE)	Swakopriver mouth, Namibia
45	*Tetraena simplex*	*T. simplex* (3)	Africa and Asia	El Hadidi, Sheahan 806	Northern Hemisphere
46	*Tetraena spongiosa*	*T. spongiosa*	Angola	Bellstedt 1214 (STE)	Praia Azul, Benguela, Angola
47	*Tetraena turbinata* ined.	*T. turbinata NGS*	South Africa	Bellstedt 1386	Tankwa Karoo near Ceres, Western Cape
48	*Tribulus lanuginosus*	*Tr. lanuginosus*		Sathishkumar,R, *et al.*	
49	*Tribulus sp.*	*Tr. sp.*	Namibia	Bellstedt 1333	
50	*Tribulus subramanyamii*	*Tr. subramanyamii*		Sathishkumar,R, *et al.*	
51	*Tribulus terrestris*	*Tr. terrestris (1)*		B. B. Simpson 16-VI-00-1	USA: TX, Burnet Co
52	*Tribulus terrestris*	*Tribulus terrestris NGS*	South Africa		Stellenbosch Train station, Western Cape, South Africa. (P.D.W. van der Merwe, s.n.)
53	*Zygophyllum fabago*	*Z. fabago* (1)	France	Chase 516 (K)	Lyons Bot. Gard.
54	*Zygophyllum fabago*	*Z. fabago NGS*	Germany (Native to the Middle East)	MP	Botanical Garden, Mainz University)
55	*Zygophyllum gontscharovii*	*Z. gontscharovii*	Tadzikistan	Ashnova 10191 (K)	
56	*Zygophyllum lehmanianum*	*Z. lehmanianum*	Turkmenistan	10191 (K)	

57	*Zygophyllum orbiculatum*	*Z. orbiculatum* (1)	S. Angola	Bellstedt 1202 (STE)	Namibe District, Angola
58	*Zygophyllum orbiculatum*	*Z. orbiculatum* (5)	S. Angola	Craven 5096 (WIND)	30 km west of Foz de Cunene, Angola
59	*Zygophyllum orbiculatum*	*Z. orbiculatum* (6)	Angola	Craven 5101 (WIND)	River tributary, South of Tambor, Angola
60	*Zygophyllum orbiculatum*	*Z. orbiculatum* (2)	S. Angola	Bellstedt 1213 (STE)	Lake Acoa, Angola
61	*Zygophyllum orbiculatum*	*Z. orbiculatum* (4)	Namibia	HK 2757 (WIND)	
62	*Zygophyllum orbiculatum*	*Z. orbiculatum* (3)	Namibia	HK 2753 (WIND)	
63	*Zygophyllum ramossisimum*	*Z. ramossisimum*	Kazachstan	Giamolov 10182 (K)	
64	*Zygophyllum rosowii*	*Z. rosowii*	China	Chaney 57	Karakum
65	*Zygophyllum stapffii*	*Z. stapffii*	Namibia	Bellstedt 1222 (STE)	Swakopmund, Namibia
66	*Zygophyllum stapffii*	*Stapffii NGS*	Namibia	Bellstedt 1222	Near Swakopmund, Namibia
67	*Zygophyllum xanthoxylum*	*Z. xanthoxylum*	China	Chase 1700 (K)	

Supplementary Table 4: Primers used for sequencing ITS genes in *Bulnesia arborea*.

	Primer name	**Primer sequence (5'-3')**	**Primer use**
18S rDNA	N-NS1	GTA GTC ATA TGC TTG TCT G	Amplification & sequencing
	C-18H	GCC CTT CCG TCA ATT CCT TTA AGT TTC AGC	Amplification & sequencing
	18Srev	CCT TCC TCT AAA CGA TAA GGT TC	Amplification & sequencing
ITS region	AB101 (Forward)	ACG AAT TCA TGG TCC GGT GAA GTG TTC G	Amplification & sequencing
	AB102 (Reverse)	TAG AAT TCC CCG GTT CGC TCG CCG TTA C	Amplification
	P16 (Forward)	TCA CTG AAC CTT ATC ATT TAG AGG A	Amplification & sequencing
	P17 (Forward)	CTA CCG ATT GAA TGG TCC GGT GAA	Amplification & sequencing
	P25 (Reverse)	GGG TAG TCC CGC CTG ACC TG	Amplification
	26S-82R (Reverse)	TCC CGG TTC GCT CGC CGT TAC TA	Amplification
	2G (Reverse)	GTG ACG CCC AGG CAG ACG T	Sequencing
26S rDNA	641rev	TTG GTC CGT GTT TCA AGA CG	Amplification & sequencing
	950rev	GCT ATC CTG AGG GAA ACT TC	Amplification & sequencing
	1229rev	ACT TCC ATG ACC ACC GTC CT	Amplification & sequencing
	1499rev	ACC CAT GTG CAA GTG CCG TT	Amplification & sequencing
	1839rev	TTC ACC TTG GAG ACC TGA TG	Amplification & sequencing
	2134rev	GGA CCA TCG CAA TGC TTT GT	Amplification & sequencing
	2426rev	CCT ACA CCT CTC AAG TCA T	Amplification & sequencing
	268rev	GCA TTC CCA AAC AAC CCG AC	Amplification & sequencing
	2782rev	GGT AAC TTT TCT GAC ACC TC	Amplification & sequencing
	3058rev	TTC GCG CCA CTG GCT TTT CA	Amplification & sequencing
	3331rev	ATC TCA GTG GAT CGT GGC AG	Amplification & sequencing
	N-nc26S1	CGA CCC CAG GTC AGG CG	Amplification & sequencing
	N-nc26S2	GAG TCG GGT TGT TTG GGA	Amplification & sequencing
	N-nc26S3	AGG GAA GCG GAT GGG GGC	Amplification & sequencing

	N-nc26S4	TTG AAA CAC GGA CCA A	Amplification & sequencing
	N-nc26S5	CGT GCA AAT CGT TCG TCT	Amplification & sequencing
	N-nc26S6	TGG TAA GCA GAA CTG GCG	Amplification & sequencing
	N-nc26S7	GAT GAG TAG GAG GGC GCG	Amplification & sequencing
	N-nc26S8	ACG TTA GGA AGT CCG GAG	Amplification & sequencing
	N-nc26S9	AAT GTA GGC AAG GGA AGT	Amplification & sequencing
	N-nc26S10	TAA AAC AAA GCA TTG CGA	Amplification & sequencing
	N-nc26S11	AAT CAG CGG GGA AAG AAG	Amplification & sequencing
	N-nc26S12	GTC CTA AGA TGA GCT CAA	Amplification & sequencing
	N-nc26S13	CCT ATC ATT GTG AAG CAG	Amplification & sequencing
	N-nc26S14	TTA TGA CTG AAC GCC TCT	Amplification & sequencing
	N-nc26S15	TGC CAC GAT CCA CTG AGA	Amplification & sequencing

Supplementary Figures S1 – S12:

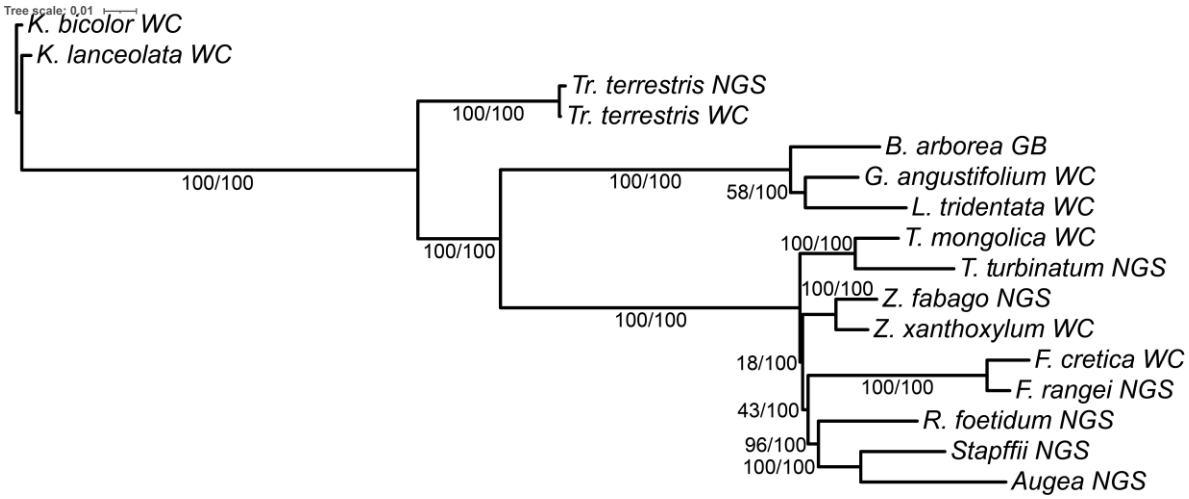


Figure S1: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 5 NC gene and 3 HVNC gene alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

Figure S2: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC gene alignment matrix of 9 ingroup taxa and 8 outgroups plus 105 species = 121 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

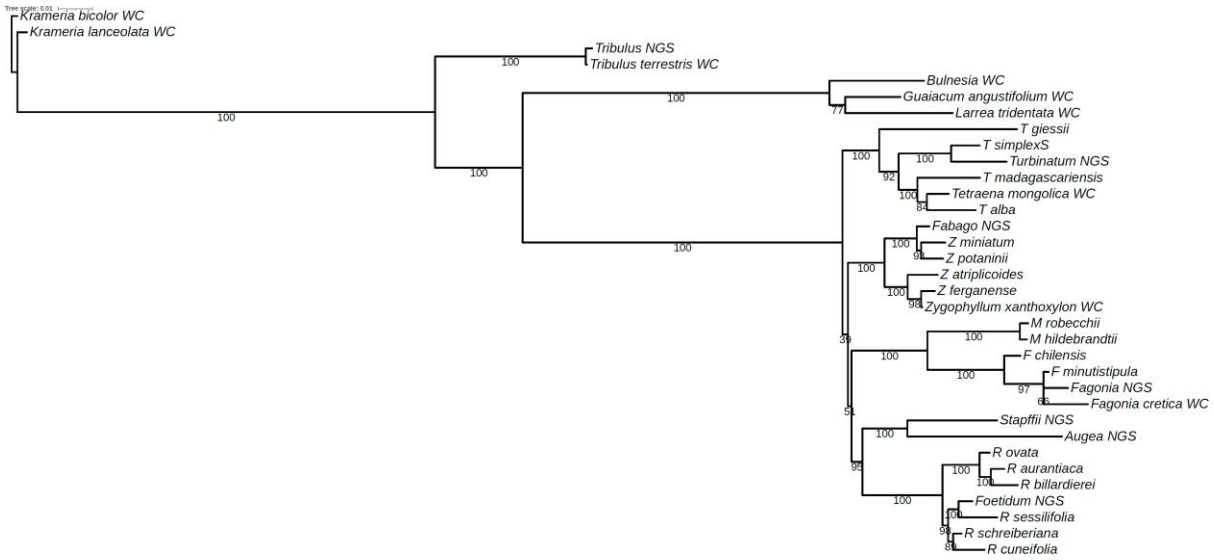


Figure S3: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene alignment matrix of 9 ingroup taxa and 7 outgroups plus 18 species = 34 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

Figure S4: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene alignment matrix of 9 ingroup taxa and 7 outgroups plus 80 species = 96 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

Figure S5: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene alignment matrix of 9 ingroup taxa and 8 outgroups plus 105 species = 122 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

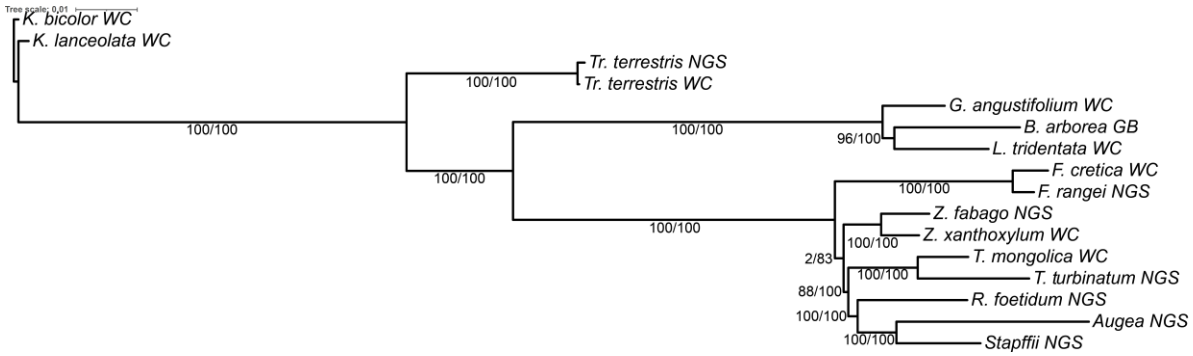


Figure S6: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 18 coding genes plus 5 NC gene and 3 HVNC gene alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

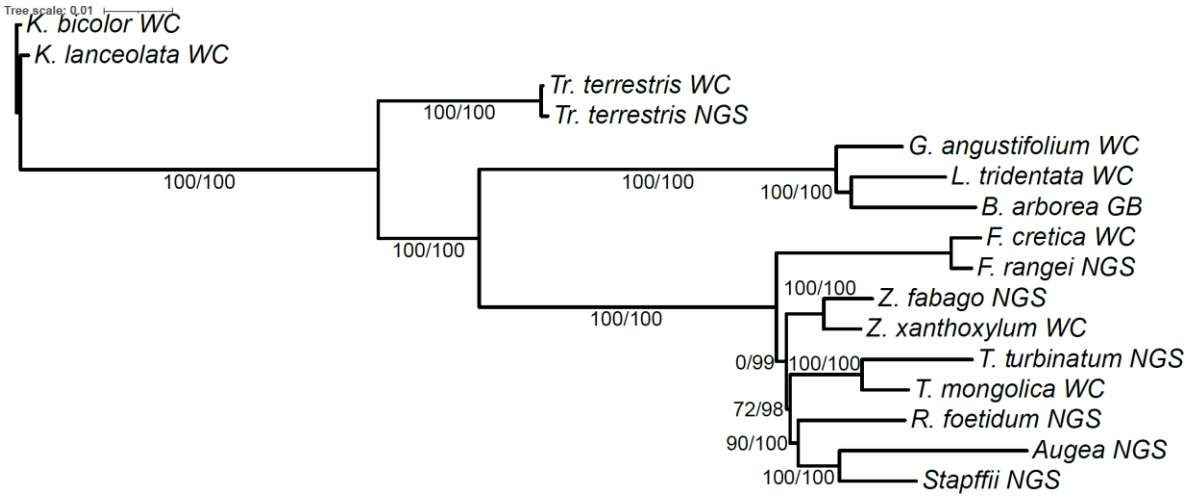


Figure S7: The most likely phylogenetic tree from the maximum likelihood analysis of the 16 taxa 39 coding genes plus 5 NC gene and 3 HVNC gene alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

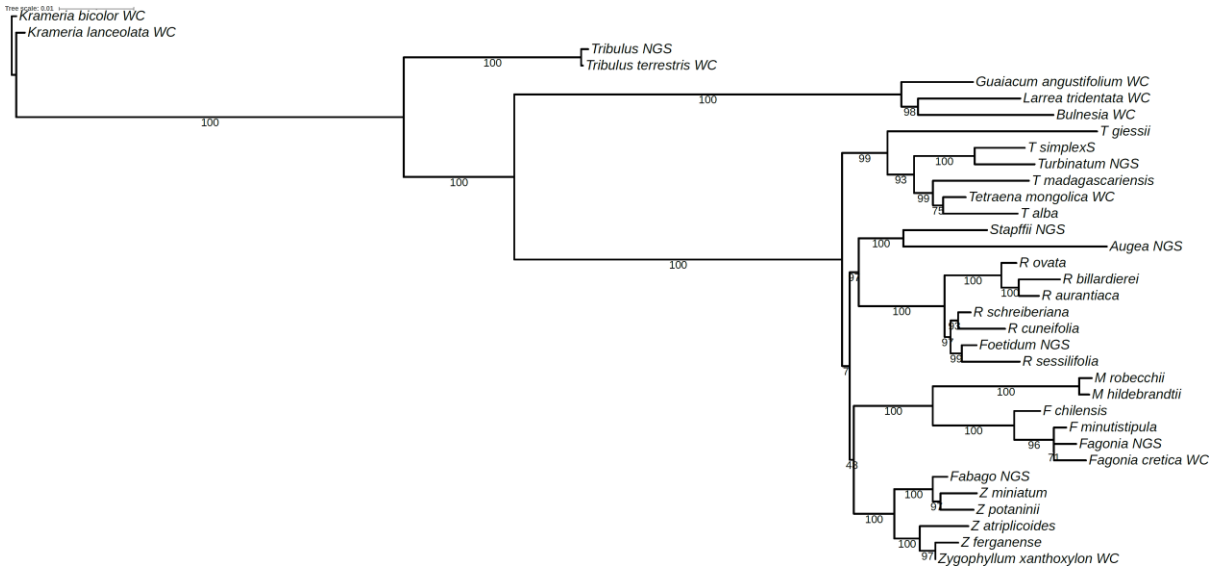


Figure S8: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene plus 39 coding gene alignment matrix of 9 ingroup taxa and 7 outgroups plus 18 species = 34 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

Figure S9: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene plus 39 coding gene alignment matrix of 9 ingroup taxa and 7 outgroups plus 80 species = 96 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

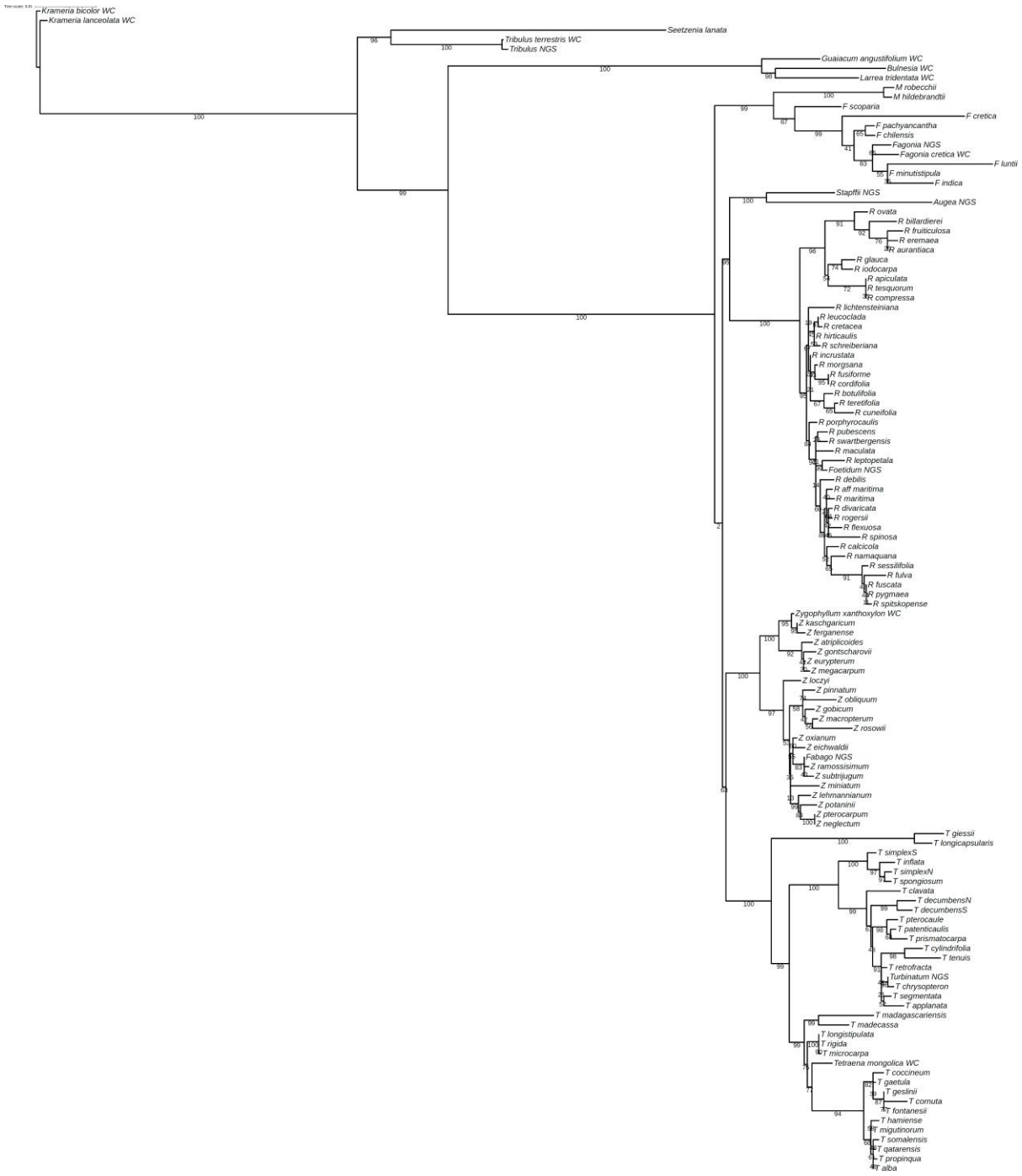


Figure S10: The most likely phylogenetic tree from the maximum likelihood analysis of the 3 HVNC genes plus 5 NC gene plus 39 coding gene alignment matrix of 9 ingroup taxa and 8 outgroups plus 105 species = 122 taxa. Maximum likelihood bootstrap values are indicated on the nodes.

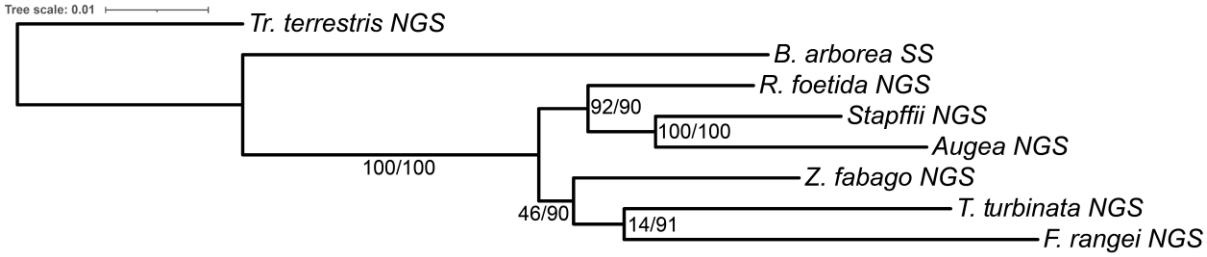


Figure S11: The most likely phylogenetic tree from the maximum likelihood analysis of the 8 taxa ITS cassette gene alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.

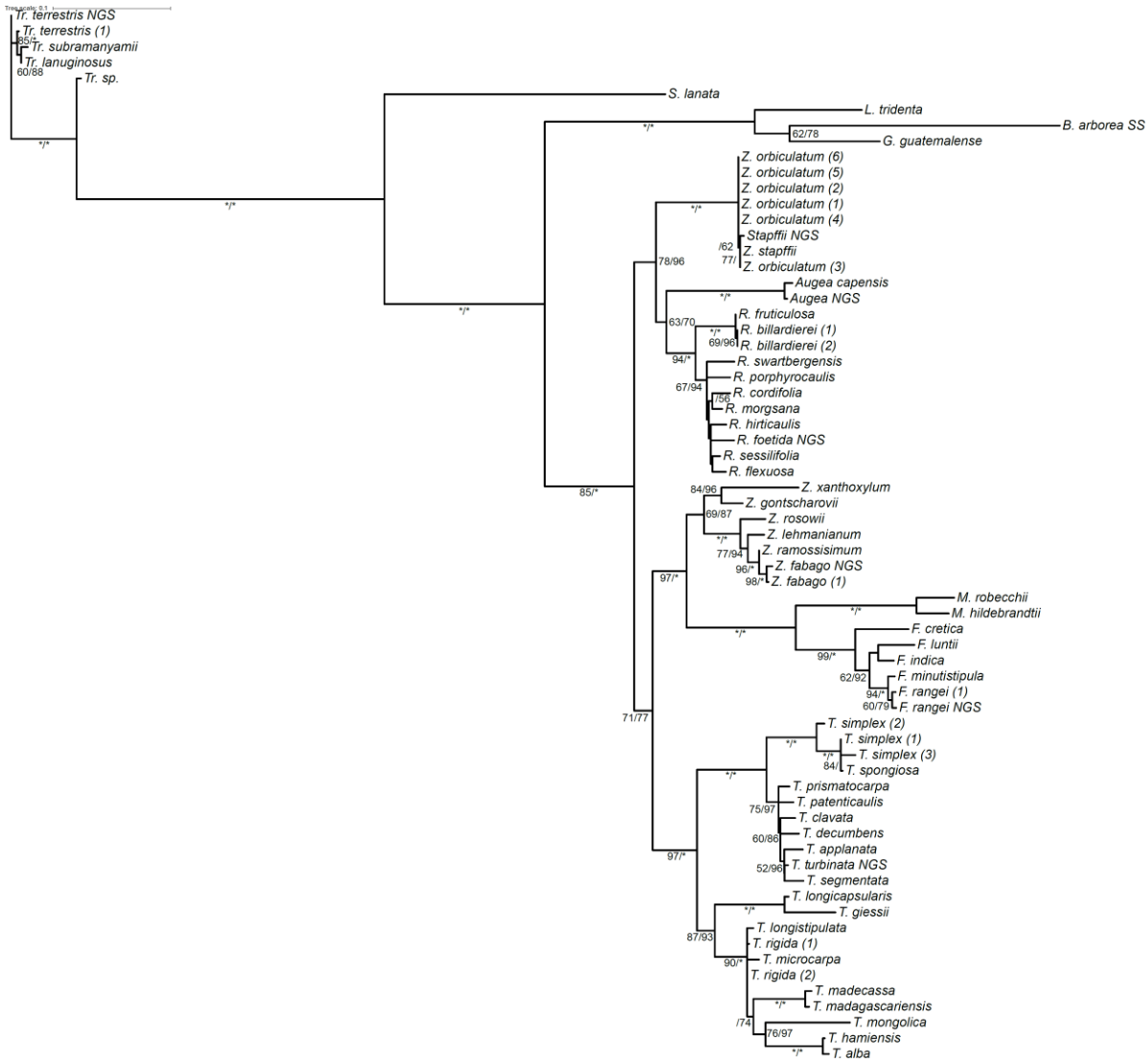


Figure S12: The most likely phylogenetic tree from the maximum likelihood analysis of the 67 taxa ITS region alignment matrix. Maximum likelihood bootstrap values and Bayesian posterior probabilities are indicated on the nodes.