Recent speciation and adaptation to aridity in the ecologically diverse Pilbara region of Australia enabled the native tobaccos (*Nicotiana*; Solanaceae) to colonize all Australian deserts

Running title: Evolutionary history of native Australian tobacco

Luiz A. Cauz-Santos[1*], Rosabelle Samuel[1], Dominik Metschina[1], Maarten J.M. Christenhusz[2,3], Kingsley W. Dixon[3], John G. Conran[4], Ovidiu Paun[1] and Mark W. Chase[1,2,3]

[1]Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, A-1030 Vienna, Austria
[2]Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK
[3]Department of Environment and Agriculture, Curtin University, Perth, Western Australia, Australia
[4]Environment Institute, School of Biological Sciences, The University of Adelaide, SA 5005 Australia

[*]Author for correspondence: Luiz A. Cauz-Santos
Email: luiz.cauz@univie.ac.at

Abstract

For the last six million years, the arid Australian Eremaean Zone (EZ) has been as dry as today. An accepted hypothesis, applied to arid regions worldwide, suggests that flora and fauna were more broadly distributed before aridification began. In Australia, this aridification process started around 20 million years ago (Mya), leading to gradual speciation processes via vicariance as the climate became increasingly arid. Here, we use genomic data to investigate the biogeography and timing of divergence of native allotetraploid tobaccos, *Nicotiana* section *Suaveolentes* (Solanaceae), which putatively entered the EZ 5 Mya. The original allotetraploid migrants from South America were adapted to mesic areas of Australia and putatively radiated recently in the EZ, including sandy dune fields (only 1.2 My old), after developing drought adaptations. Based on coalescent and maximum likelihood analyses designed to corroborate timing of the Australian radiation independently, arrival of *Nicotiana* section *Suaveolentes* on the continent occurred approximately 6 Mya, and ancestors of the Pilbara (Western Australian) lineages radiated there at the onset of extreme aridity 5 Mya by locally adapting to these various ancient, highly stable habitats. The Pilbara thus served as both a mesic refugium and cradle for adaptations to harsher conditions. This dual role is due to its high topographical diversity, providing microhabitats with varying moisture levels, and its proximity to the ocean, which buffers against extreme aridity. Consequently, species like *Nicotiana* have been able to survive in mesic refugia during arid periods and subsequently adapt to more arid conditions. These results demonstrate that initially poorly adapted plant groups can develop novel adaptations in situ, permitting extensive and rapid wide dispersal despite the highly variable and unpredictable extremes of heat and drought in the EZ.

Key words: *Nicotiana* section *Suaveolentes*; phylogenomics; biogeography; drought adaptation; dispersal; diversification.

Introduction

The evolution of plant groups in the Eremaean Zone of Australia has been explained through three primary models: vicariance, pre-adapted immigration, and in situ adaptation. However, these models are not mutually exclusive and may operate concurrently or sequentially within different taxa. Vicariance, for instance, involves the gradual isolation and divergence of widespread taxa due to increasing aridity, as seen in *Eucalyptus* (Mytaceae) and *Acacia* (Fabaceae) (Martin, 2006; Byrne et al., 2008). In contrast, pre-adapted immigration, potentially (African or Asian) immigrant clades, supported by studies on *Triodia*, *Ptilotus* and other Amaranthaceae (Toon et al., 2015; Shepherd *et al.*, 2004; Kadereit & Freitag, 2011; Hammer *et al*., 2021), suggests that some species arrived already equipped for arid conditions. Lastly, in situ adaptation, although less documented, involves the evolution of arid specialization within the Eremaean Zone itself. Australia has several of the oldest known pieces of the Earth's crust (e.g., the Pilbara Craton, 3.8–2.7 Ga), but thanks to extensive erosion over such long timescales, these have a generally low-relief topography and few major physical barriers to dispersal, so it is possible that there is a third model: a group of organisms lacking xeric specializations could adapt in one region and then disperse over much of the Australian continent. These processes collectively contribute to the region's complex biogeography and high levels of endemism.

Dispersal events, such as those facilitated by dry tornadoes (in Australia called willy-willies), have potentially played a crucial role in the diversification of *Nicotiana* in Australia. These events could enable the initial migration and subsequent wide distribution of *Nicotiana* species across the continent, contributing to the genetic isolation and diversification observed in the genus today. Unlike vicariance, which primarily isolates populations gradually, these dispersal events can rapidly introduce species to novel environments, creating new opportunities for speciation. However, the evolutionary history of *Nicotiana* section *Suaveolentes* has been the subject of different hypotheses. One hypothesis suggests that the origin of *N*. section *Suaveolentes* dates back to the early Miocene (ca. 20 Mya), with speciation occurring through a vicariant model of arid-adapted biota. This model, as described by Cracraft (1991) and Ladiges et al. (2011), proposes that species were progressively isolated as aridification moved from north to south, forming organized "tracks" of species distributions. Following these primary vicariance-driven distributions, dispersal events produced a few widespread species. The alternative hypothesis points that *Nicotiana* only reached Australia around 4–6 Mya, as suggested by Mummenhoff and Franzke (2007) and Schiavinato et al. (2020), with a more recent radiation occurring in the Eremaean Zone (Clarkson et al., 2017; Dodsworth et al., 2021). In this scenario, dispersal to the arid interior would have been the predominant mechanism of diversification, with vicariance playing a more secondary localized role.

The Australian distribution of *Nicotiana* section *Suaveolentes* spans the continent, except for Tasmania, which putatively makes them ideal for revealing general factors contributing to arid zone speciation. Furthermore, elucidation of the phylogeographic history for such widespread groups can be instrumental to understand the major drivers of individual species distributions. The Australian species are primarily found in the tropics, subtropics and warm temperate regions and largely absent from cool temperate regions (Ladiges *et al*., 2011). There was putatively rapid initial speciation producing a few species in more mesic areas in northern and eastern Australia, followed by multiple radiations in the EZ (Ladiges *et al*., 2011, Chase *et al*., 2018, 2022a). The current understanding of the biogeography, phylogenetics and chromosome numbers/genome sizes of *N*. section *Suaveolentes* (Table 1; Chase *et al*., 2002b) are consistent with the ancestral distribution of the genus in Australia being confined to these northern and eastern parts, where all species with higher chromosome numbers (n = 22–24) and mostly larger genomes now occur (Fig. 1). As new species appeared and chromosome numbers and generally genome sizes decreased (Chase *et al*., 2002b), they expanded to the dry interior of central and southern Australia, where they are now highly diverse. Although several EZ species are specialists of more mesic niches, growing in the shade of trees or rock outcrops, many others inhabit exposed, extremely arid sites (Fig. 1), unexpected for such thin-leaved plants with no obvious adaptations for these extreme habitats. However, their rapid lifecycle completion following rainfall could sometimes allow them to avoid the harshest conditions.

Here, we have used multispecies coalescent methods on genome-wide single nuclear polymorphisms (SNPs) combined with biogeographical and novel molecular clock analyses to evaluate these two competing hypotheses that the radiation of *N*. sect *Suaveolentes* in the EZ started either as early as 20 Mya or as late as 6 Mya. The former would coincide with the early stages of aridification in Australia, supporting the vicariance model where species gradually became isolated and diversified as the climate changed (Ladiges *et al*., 2011), whereas if the latter is true then dispersal was the preponderant mechanism and diversification took place long after the EZ became dry. Inadequate, fine-scale sampling in such remote areas across the whole continent has limited previous phylogeographic studies, and existing phylogenetic analyses typically have not sampled species thoroughly either through lack of sufficient collections or due to the existence of many undescribed "cryptic" species. After ten years of fieldwork, we have assembled an unprecedented collection of accessions and nearly trebled the number of species known in this group (Chase *et al*., 2021, 2022, 2023). We have also extended our taxonomic coverage by recovering additional taxa/accessions from viable seeds on herbarium specimens up to 25 years old. We first constructed a species tree for *N*. sect *Suaveolentes* and implemented dating of the speciation events to corroborate previous molecule-based age estimates (Clarkson *et al*., 2017; Schiavinato *et al*., 2020; Dodsworth *et al*., 2021). We then carried out biogeographical inference to test the relative contribution of dispersal and vicariance to the overall distribution of the species. Our primary goal is to identify where and when the initial lineages confined to the wetter parts of the Australian continent became adapted to xeric conditions, providing insights to how they diversified into the myriad, relatively unoccupied niches hypothesized to be typical of arid zones.

Material and Methods

Plant material

Sampling of 273 accessions from 58 species of *N.* section *Suaveolentes* (Table S1), extended the recent studies of Cauz *et al*. (2022), which focused on *N. benthamiana*, and Chase *et al*. (2022), which examined genome size and chromosome number evolution in this section of the genus. Most samples were collected by Chase and Christenhusz in the wild and are vouchered in major Australian Herbaria (AD, BRIS, CANB, CNS, DNA, NSW, NT and PERTH, the standard acronyms for these collections). Other accessions were added by germinating viable seeds removed from herbarium specimens. From more than 600 accessions, including roughly 100 from herbarium-stored seeds, we selected the above subset of 278 samples that it comprised all putative species, including several that are undescribed and designated here as sp. nov. with a location where we collected these taxa (e.g., *N. sp. nov.* Coondiner). The maps were generated in QGIS v. 3.20.3 (QGIS Development Team, 2021); the map layer with the vectors from the drainage and rivers divisions were obtained from the Australian Government Bureau of Meteorology (http://www.bom.gov.au/water/geofabric/). The provenance data (latitude, longitude) of the accessions were obtained from the Australasian virtual herbarium (https://avh.chah.org.au) and our personal data gathered during the field collections in Australia.

Collecting and import permits

The following collecting permits, which cover these accessions, were issued to MWC and MJMC: Western Australia SW017148, CE006044, Northern Territory 58658, and

Queensland PTU-18001061. Removal of seeds from herbarium specimens was approved by the curators/collections managers of the following herbaria: AD, BRI, CANB, NSW, NT and PERTH. All seeds imported into the UK followed published guidelines; plants were grown at the Royal Botanic Gardens, Kew, under DEFRA PHL2149/194627/5NIRU CERT:106-2019; HMRC TARIFF CODE: 0601209090. No material collected by us or in our possession will be exploited for commercial purposes without involvement of the Australian and Aboriginal authorities, as required by the collecting/export permits.

DNA isolation, library preparation and sequencing

From ca. 20 mg of silica-dried leaf tissue, we extracted DNA with the cetyltrimethylammonium bromide (CTAB) procedure (Doyle, 1990), following a 20 min ice-cold sorbitol buffer treatment (100 mM tris-HCl, 5mM EDTA, 0.35 M sorbitol, pH 8.0). Then, we used 2.5 µl of RNase A (Thermo Fischer, USA) for 30 min at 37 °C and purified the extracted DNA using the NucleoSpin gDNA clean-up Kit (Machery-Nagel, Germany), according to the manufacturer's instructions.

DNA samples were first single digested with the PstI restriction enzyme in advance of library preparations. Although PstI activity is not affected by CG methylation, it is sensitive to that at CHG sites (H stands for any nucleotide apart from G), a type of methylation frequently found around plant transposable elements (e.g., Domb *et al.*, 2020), which are known to have paralogy issues and low phylogenetic signal. The effects of any methylation variation have been mitigated by filtering for missing data. Libraries were prepared following Paun *et al.* (2016), as modified by Cauz *et al*. (2022) and Chase *et al*. (2022). Processing in batches was carried out using index barcodes distinct from one another by at least three bases. Sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at/ngs) on an Illumina HiSeq 2500 with 125 bp paired-end reads.

SNP calling and phylogenomic analysis

The BamIndexDecoder v.1.03 (included in Picard Illumina2Bam package, available from http://gq1.github.io/illumina2bam/) was used first to process the RADseq data and demultiplex via the index barcodes in sublibraries. Subsequently, demultiplexing of individuals via their inline barcodes was conducted in process_radtags from Stacks v.1.47 (Catchen *et al.*, 2013), together with removal of reads containing uncalled bases or with low quality-scores.

The reference genome of *N. benthamiana* v.2.6.1 (Bombarely *et al.*, 2012, available from https://solgenomics.net/organism/Nicotiana_benthamiana/genome), a member of *N.* section *Suaveolentes*, was used on individual read mappings in BWA MEM v. 0.7.17 (Li and Durbin, 2009) using and applying the –M option to flag shorter splits hit as secondary. We also checked for biases potentially driven by phylogenetic relatedness to the reference individual. After alignment, the sam file was sorted by reference coordinates, and read groups were added using Picard Toolkit v.2.27 (available from http://broadinstitute.github.io/picard/). We used the Genome Analysis Toolkit (GATK) v.3.8 (McKenna *et al.*, 2010) to improve alignment quality around indels, thinning the data to a maximum of 100,000 reads per interval.

GATK was used to call variants following the best-practice DNAseq recommendations. First, we inferred genotypes via HaplotypeCaller and GVCF mode for individual samples and subsequently processed all individual intermediate GVCF in a joint genotyping analysis in the GenotypeGVCFs module. The raw vcf file was first processed in VCFtools v.0.1.15 (Danecek *et al.*, 2011) and retained only variants presents in at least 50% of individuals. We then used the VariantFiltration GATK module with the following criteria: (1) depth of coverage (DP) < 500; (2) variant confidence (QUAL) < 30.00; (3) variant confidence divided by the unfiltered depth (QD) < 2; (4) Phred-scaled P-value for the Fisher's exact test to detect strand bias (FS) > 60; (5) a root mean square of mapping quality across all samples (MQ) < 40; (6) u-based z-approximation from the rank sum test for mapping qualities (ReadPosRankSum) < -8.0; and (7) u-based z-approximation from the rank sum test for the distance from the end of the reads with the alternate allele (MQRankSum) < -12.5.

The variant calling and initial filtering steps in GATK produced 7,606,626 variable sites, but we retained only SNPs with a minor allele frequency ≥ 0.008 (i.e., present in at least four haplotypes), an average depth above 20 and 20% maximum missing data. We also filtered the data using the populations pipeline in Stacks to retain only variable positions with a maximum observed heterozygosity of 0.65, thus avoiding the use of pooled paralogs in further analyses.

To investigate phylogenetic relationships among the species of *N.* section *Suaveolentes*, we first converted the final filtered vcf to a PHYLIP file using PGDspider v.2.1.1.0 (Lischer & Excoffier, 2012). We removed invariant sites with the script ascbias.py (https://github.com/btmartin721/raxml_ascbias). A RAxML v.8.2.12 (Stamatakis, 2014) used the remaining 170,552 SNPs with the recommended ascertainment bias correction (Lewis, 2001). The phylogenetic tree was inferred under the GTRCAT model of nucleotide substitution with a search for the best-scoring ML tree and 1,000 rapid bootstrap replicates. We assigned *N. africana* as the outgroup because it was well-supported sister to the rest of *N.* section *Suaveolentes* in all previous studies (Chase *et al.*, 2003; Clarkson *et al*., 2004, 2010, 2017; Marks *et al.*, 2011; Kelly *et al.*, 2013). All species of *N.* section *Suaveolentes* are allotetraploids, including *N. africana* and *N. benthamiana*, so our analyses do not mix diploid and polyploid taxa. Finally, we visualized and annotated the best tree in R, using ape v.5.3

(Paradis & Schliep, 2018), biostrings (Pagès *et al.*, 2020), ggplot2 (Wickham, 2016), ggtree (Yu *et al.*, 2017) and treeio (Wang *et al.*, 2020).

Coalescent-based species tree and divergence time estimation

To construct a species tree for a reduced and representative matrix of *N.* section *Suaveolentes*, we first inferred the relatedness between accessions looking for evidence of introgression, which would interfere with species-tree inference. This exercise was conducted on a reduced matrix of 22 species representing the major clades in *N.* section *Suaveolentes*. After calculating genotype likelihoods in ANGSD v.0.930 (Korneliussen *et al.*, 2014) and retaining only variants with a minor allele shared by at least two individuals, we applied a minimum base mapping quality of 20, SNP calling confidence of $p<1e^{-6}$, and presence in at least 70% of individuals. The major and minor allele frequencies were estimated using the GATK-based genotype likelihood model, and our final dataset resulted in 3,201,820 variable positions. For the inference of coancestry, we used our estimated genotype likelihoods to obtain a covariance matrix using PCangsd (Meisner & Albrechtsen, 2018), and plotted our data using the heatmaps.2 function from GPLOTS v.3.0.1.1 (Warnes *et al*., 2020).

To construct the coalescent species tree, we first used the filtered vcf file to prepare a smaller dataset in VCFtools v.0.1.15 selecting only unlinked biallelic SNPs (> 10,000 bp apart on a contig) and removing missing data at each locus. This procedure produced a matrix of 2,400 unlinked single nucleotide polymorphisms (SNPs) for 36 accessions, representing 18 species (two per species) that broadly cover the phylogenetic diversity of *N.* section *Suaveolentes*. We converted the vcf file containing unlinked SNPs to PHYLIP and then NEXUS format using PGDSpider v.2.1.1.0, and finally we created the input XML files in BEAUti v.2.4.8 (Bouckaert *et al.*, 2014).

We constructed the coalescent species tree using SNAPP v.1.2.5 with a chain length of 10 million and saving a tree every 1,000th generation. We monitored the convergence of the run based on the ESS values from the log-file with Tracer v.1.6 (Rambaut *et al.*, 2018). We removed the initial 10% of trees as burn in, visualized the SNAPP trees as a cloudogram using Densitree v.2.2.6 (Bouckaert & Heled, 2018) and produced the posterior probabilities for each clade with Treeannotator v.1.8.3 (Drummond *et al.*, 2012). To calibrate the species tree, we used $5e^{-09}$ as the rate of substitution per site per generation (Schiavinato *et al.*, 2020) and one year as generation time (i.e., these plants only rarely live more than one season in nature). We estimated divergence times by rescaling the results using the total length of investigated sites for the loci included and total number of polymorphic sites across their length.

Ancestral range estimation

For biogeographic inference, we calibrated a RAxML tree using TreePL (Smith & O'Meara, 2012), which produces a dated tree using a penalized likelihood approach, and minimum and maximum ages to constrain the tree. The minimum and maximum ages used for TreePL dating were based on the divergence times estimated in our SNAPP species tree. The dated tree with node ages was visualized using FigTree v1.3.1 (http://tree.bio.ed.ac.uk). We obtained the confidence intervals for the node ages and the maximum clade credibility (MCC) tree summarizing the RAxML bootstrap replicates with Treeannotator v.1.8.3 (Drummond *et al.*, 2012) (Fig. S1).

We explored the biogeographic history of *N.* section *Suaveolentes* first by comparing models in BioGeoBEARS, including DEC, BAYAREALIKE and DIVA plus using the additional free parameter “j” in each model, which accounts for jump dispersal/founder-event speciation. According to the Akaike information criterion (AIC), the best-fit model was DIVA+J (Table S2). We then performed the ancestral range estimation using BioGeoBEARS testing two models of distribution. The first one considered drainage divisions of Australia as operational areas: A, Africa; B, Pacific, C, Carpentaria Coast; D, Tanami-Timor Sea Coast; E, north-western Plateau; F, Pilbara-Gascoyne; G, Southwest Coast; H, Southwest Plateau; I, Lake Eyre Basin; J, Murray-Darling Basin; K, Northeast Coast (Queensland); L, Southeast Coast (NSW); M, Southeast Coast (Victoria; N, South Australian Gulf. We chose this model because we had noticed in a previous study (Cauz-Santos *et al*., 2022) that species distributions seemed to conform to river drainage basins. Rather than acting as barriers to gene flow, it is more likely that the drainages serve as conduits, facilitating gene flow through the dry tornadoes that typically travel across relatively flat landscapes and dissipate upon encountering uneven terrain. Consequently, dispersal mostly occurring within a drainage system might make sense in this setting, an aspect that we are planning to investigate further.

The second model split *N.* section *Suaveolentes* into the operational areas from Ladiges *et al*. (2011): A, Africa, Namibia; B, Pacific; C, eastern Australia; D, south-eastern Australia; E, south-eastern Interzone; F, north-eastern Interzone 3; G, Adelaide/Eyre; H, south-western Interzone; I, Pilbara; J, north-western Australia; K, western Desert; L, northern Desert; M, eastern Desert; N, Nullarbor; O, central Australia. These geographical areas were defined by Ladiges *et al*. (2011) based on the distribution of narrow-range endemic species, Australian bioregions recognized by other authors (Burbidge, 1960; Cracraft, 1991; Crisp *et al*., 1995) and the Interim Biogeographic Regionalisation of Australia (IBRA) v.6.1 (http://www.environment.gov.au). The BioGeoBEARS model testing for this distribution also resulted in the DIVA+J as best fit AIC model (Table S3). In both models, the species distributions (presence/absence) included a maximum of three areas per species except for one widespread species, *N. velutina*, although we now believe that this species is much more restricted than previously thought due to discovery that this species concept as previously defined includes two previously unrecognized species, considerably reducing its distribution. However, all three form an exclusively related species complex, so in terms of these analyses, this is an acceptable assumption. The accessions included in our analyses are from only one of the revised concepts in the *N. velutina* complex (Cauz-Santos, Metschina & Chase, unpubl.).

Results

Phylogenetic analysis of *Nicotiana* section *Suaveolentes*

We obtained an average of 3,053,843 paired-end reads for the 273 accessions used in this study. The filtered reads mapped onto the *N. benthamiana* reference genome at a high rate (an average of 95.75%), with a final average coverage across samples of 9.4. The mapped reads were then used for variant calling, which after filtering resulted in a total of 240,871 SNPs for a minimum of 80% of individuals. The results shown here expanded the matrix published in Chase *et al*. (2022) to include many more of the new species. These results were also used to illustrate the positions of species being described as new (Chase *et al*., 2023, in press), but the methods and results of this analysis are here published for the first time. Phylogenetic trees of *N.* section *Suaveolentes* (Fig. 2) provide a framework for the other studies conducted for this paper, but they are not a primary focus here and have not been used previously for these purposes.

The RAxML maximum likelihood (ML) tree was in general highly supported (most nodes with bootstrap percentage, BP, >90; Fig. 2). For several species, we included multiple accessions that formed unique and well-supported groups. In total, the tree comprised 18 major clades (numbered as Roman numerals I–XVIII), with *N. africana* (I) outgroup to the rest of *N.* section *Suaveolentes* (Clarkson *et al*., 2011). The basal node comprises *N.*

fatuhivensis (II) sister to *N. gibbosa*+*N. forsteri* (III), but with low support (BP 76; the only BP less than 90 along the spine of the ML tree) relative to the position of *N. monoschizocarpa* (IV). *Nicotiana heterantha* (V) and *N. umbratica* (VI) are then successively sister to the rest, which split into six geographically widespread clades (VII, VIII–XI, X/XI, XII/XIII, XIV, XV–XVIII).

Species tree and divergence times

The coancestry heatmap (Fig. 3) shows the clear relatedness among accessions of *N. monoschizocarpa*, which is one of the species at the basal node of the *N.* section *Suaveolentes*, but this pattern is not evident in the other two species near the basal nodes, *N. africana* and *N. forsteri*. *Nicotiana africana* grouped within the *N. forsteri* accessions with high coancestry, and additionally, one of the *N. africana* accessions exhibited high coancestry with the species of the *N. benthamiana* complex. Considering these patterns of introgression and as only two accessions of *N. africana* have been available to us (a minimum of two accessions per species is required to produce the species tree), we decided to remove *N. forsteri* and species of the *N. benthamiana* complex from the species tree inference. The highest coancestry is exhibited by the most recent species groups, *N. gossei*/*N. velutina*, *N. truncata*/*N. excelsior* and *N. maritima/N. suaveolens*. Introgression is not a general phenomenon in *N.* section *Suaveolentes*, but it does seem to be a factor for a few species pairs.

The coalescent species tree for a reduced dataset comprised representative species from all major clades in *N.* section *Suaveolentes*. We obtained one topology in the SNAPP analysis with strong support representing 99% of the posterior density distribution (Fig. 4). The ESS values were used to monitor the convergence of the analysis, from which we obtained values higher than 200 in each parameter. The species tree places *N. africana* as sister to the rest, consistent with its outgroup position in the ML results (above) followed by *N. monoschizocarpa*. Subsequently we observed three clades, one comprising (*N. occidentalis*+*N. murchisonica*)+*N. hesperis* (clade XII/XIII in the ML tree), a second with (*N. gascoynica*+*N. simulans*)+*N. stenocarpa* (clade VII–XI in the ML tree) and a major clade (*N. rotundifolia* to *N. gossei*) comprising species from the two remaining major clades in the ML tree (XIV, XIV–XVIII).

In the divergence time estimation, *N. africana* diverged from the remainder of the section around 6.5 Mya, in the late Miocene. Among the Australian species, the first split, *N. monoschizocarpa* from the rest, was estimated at 6 Mya, followed by clades XI/XII and VI–XI clades, at 5 and 4 Mya, respectively. Finally, the largest clade, XIV–XVIII, diversified in the last 1 My.

Biogeographic history of *Nicotiana* section *Suaveolentes*

For the biogeographic analysis, we first evaluated the best model for our dataset including the possibility of founder-event speciation or jump dispersal (adding the j parameter in BioGeoBEARS). According to AIC values, the DIVALIKE+J model was best, accounting for 79% of the predictive power found in all tested models. This model allows for the possibility of anagenetic (dispersal and extinction) and cladogenetic events (vicariance), and the J parameter added to the model the possibility of founder events. Considering our favored distribution model (Australian river drainages) and the fact that most species do not occur in more than one drainage basin, selection of DIVALIKE+J has biogeographical support.

The ancestral area reconstruction (DIVALIKE+J and Australian river drainages as divisions for the distribution species in *N.* section *Suaveolentes*) resulted in a total of 63 biogeographic events (Table 1, Fig. 5). The results show a combination of anagenetic and cladogenetic events playing a role in the distribution of this group, with founder events

(60.6%) being the main source of speciation, followed by within-area speciation events (36.4%) and a small proportion of vicariance (3.6%). In the biogeographic tree, the arrival of the ancestor of the *N.* section *Suaveolentes* in Australia occurred around 5.2 Mya. Even if the ancestral range of the species at the basal nodes is unclear (*N. monoschizocarpa*, *N. gibbosa* and *N. forsteri*), the common ancestor of the rest expanded its distribution around 5 Mya to the Pilbara region (F), in which several within-area speciation events occurred, resulting in this region becoming a *Nicotiana* biodiversity hotspot. The Pilbara region (F) as referred in this study encompasses the Pilbara Craton and adjacent basins of the Gascoyne, Wooramel, and Murchison Rivers in north-western Western Australia (north of latitude 25°00′ S and west of longitude 121°30′ E; Pepper *et al*., 2013) (Fig. 5). This area is larger than the Pilbara Bioregion of the IBRA classification, which only covers a portion of region (F), and is characterized by its unique geological formations, including the ancient Pilbara Craton, and features a variety of landscapes such as mountain ranges, coastal plains, and arid desert areas. From the Pilbara, a series of dispersals to the other parts of the EZ occurred, the last of which was to central and southern Australia. The second model using the areas as defined in Ladiges *et al*. (2011) also resulted in the Pilbara (region I) being colonized by the species of *N.* section *Suaveolentes* around 5 Mya, with a series of dispersal events from there, leading to the current distribution of this section in central and southern Australia (Table S4, Fig. S2).

Discussion

Timing of diversification in the Pilbara region and dispersal to the rest of the Eremaean Zone

The evolutionary processes shaping arid-adapted biota are complex and multifactorial, involving in many cases processes as vicariance, isolating populations due to climatic or geological changes, long-distance dispersal, enabling colonization of new habitats, and in situ speciation, driven by microhabitat differentiation and local environmental pressures. These mechanisms interact dynamically, creating genetic diversity and endemism. This complexity, influenced by the harsh and fluctuating conditions of arid environments, results in unique evolutionary trajectories. Other studies on arid-adapted biota have already demonstrated the complexity of evolutionary models in these environments (Cracraft, 1991; Byrne et al., 2008; Ladiges et al., 2011).

Our study contributes to this understanding by showing that *N.* section *Suaveolentes* first occupied mesic regions of Australia around 6 Mya. Subsequently, these species adapted to the arid conditions of the Pilbara region around 5 Mya, leading to their widespread distribution across the Eremaean Zone. This pattern of adaptation and dispersal is significant within the context of the Australian flora, and while highlights a specific example within *N.* section *Suaveolentes*, it is important to recognize that similar processes may have occurred in other plant groups, and further studies may reveal additional examples of such evolutionary dynamics. Molecular clock methods using multiple calibrations based on species pairs on oceanic islands/mainland (Clarkson *et al*., 2005), secondary ages from other analyses (Clarkson *et al*., 2017), and various methods of estimating rates of molecular divergence (Mummenhoff & Frankze 2007; Schiavinato *et al*., 2021; this paper) produce relatively consistent age estimates for formation of *N.* section *Suaveolentes,* ca. 5–7 Mya. Ladiges *et al.* (2011) disparaged the estimate of Mummenhoff & Frankze (2007) by citing other molecular clock studies (more generally on Australian angiosperms) that have provided highly divergent timings, but those for *Nicotiana* using these several approaches thus far have been relatively consistent for *N.* section *Suaveolentes*.

We found dispersal to be the predominant mode of evolution for *N.* section *Suaveolentes* throughout the EZ with some local examples of *in situ* speciation and vicariance being secondary features. Diversification in each major clade appears to be recent and ongoing, including adaptation to sandy dune fields (e. g., *N. pila*, *N. inglba*, *N. latzii*, *N. velutina*), which only appeared c. 1.2 Mya (Fujioka *et al*., 2009), concurrent with lower frequency/higher amplitude glacial cycles worldwide c. 0.8–1.2 Mya (Clark *et al.*, 1999). The recentness of speciation accounts for the cryptic nature of many recently described *Nicotiana* species (Chase *et al.*, 2023).

However, diversification has not been limited to recent times, and our phylogeny (Figure 5) reveals significant speciation events that occurred over a prolonged period and across multiple regions (for example in areas D and H). This pattern highlights a complex evolutionary history where both recent diversification (approximately 1 Mya) and older events have played crucial roles in shaping the current diversity of *N.* section *Suaveolentes*. The presence of these older diversification events reinforces the importance of considering the entire temporal framework of evolution within this group, as it reveals the contributions of different historical processes to their adaptation and speciation in the Eremaean Zone.

The Pilbara as a biodiversity hotspot and cradle of novel adaptations

Our results clearly demonstrate that *N.* section *Suaveolentes* first occupied the more mesic portions of the Australian continent 6–5 Mya and then diversified (Fig. 5) in the Pilbara region. Aridity increases from the forested areas of the south-western and eastern coasts (mean annual rainfall of 1000–2500+ mm) to the hummock grasslands in the desert inland regions of the continent (< 200 mm; Groves, 1999). Mountains in the Pilbara are located near the ocean, providing a thermal buffer that maintains relatively stable temperatures. This proximity to the ocean and varied topography contributes to the region's unique biodiversity and ecological stability. The plants and animals there are exceptionally diverse for an arid region (Booth *et al.*, 2022), undoubtedly due to the complexity of the landscape with diverse soil types and topography, long-term geological stability, and presence of abundant refugia (e.g., multiple gorges with permanent water) providing mesic havens during even the most arid years (e.g., El Niño years). The high mountains of the Archaean Pilbara Craton (Hammersley Ranges at 1,500 m) differ from those of the Proterozoic Gascoyne Complex to the south, which are lower and the product of the collision of the Pilbara and Proterozoic Yilgarn Cratons (Myers, 1993), and the overall effect of this geological history on the Pilbara region has been the production of a mosaic of arid and mesic habitats in close proximity, making them unique on a planetary scale and a laboratory for diversification and evolutionary novelty.

The once tropical, forested center of Australia was long ago replaced by extensive arid lands beginning in the Miocene (23.03–5.33 Mya) reaching into the Pliocene (5.33 to 2.58 Mya) and Pleistocene (2.58 mya to 11,700 kya; Flower & Kennett, 1994). The arid center of the Australian continent has received much less attention (Byrne *et al.*, 2008) than other regions, and most detailed studies of evolutionary diversity in Australia have focused primarily on the tropical rainforests (Bell *et al*., 2010) and temperate forests (Chapple *et al.*, 2011; Kay & Keogh, 2012).

Some plant phylogeographic studies have focused on genetic diversity in the Pilbara bioregion, but most have examined single or pairs of species endemic to the Pilbara (Levy *et al*., 2016; Nistelberger *et al.*, 2020; Millar *et al.*, 2022), in which they found high levels of variation that was not geographically structured due to gene flow across the region. The Pilbara bioregion has often been found to harbor high levels of diversity also at the species level, making it a biodiversity hotspot for both plants (Anderson *et al.*, 2016) and animals (see below), but how much this species diversity has influenced broader biotic patterns in the EZ has been understudied. Comprehensive environmental surveys in advance of mining projects, 2002–2007 (McKenzie *et al*., 2009), resulted in the discovery of hundreds of new plant and animal species. For beetles and scorpions, 68% and 83%, respectively, could not be assigned to described species (Guthrie *et al.*, 2010; Volschenk *et al.*, 2010). Fine-scale genetic studies have suggested substantial cryptic diversity and complex genetic patterns across the Pilbara (Pepper *et al*., 2008, 2011a; Shoo *et a*l., 2008; Doughty *et al.*, 2010, 2011b; Catullo *et al*., 2011; Anderson *et al*., 2016). In contrast to our results, Ladiges *et al.* (2011) found species diversity in *N*. section *Suaveolentes* to be greatest in central Australia, where several of their "tracks" overlap, but our findings with highly revised species circumscriptions identify the Pilbara region as the most diverse in both species numbers and lineages.

The only other studies comparable in terms of nearly complete species-level sampling to ours are those of *Ptilotus* (Amaranthaceae; Hammer *et al*., 2021) and *Triodia* (Toon *et al.*, 2015; Anderson *et al*. 2016), which both found that the detected high levels of species diversity in the Pilbara could act as a source of pre-adapted xeric diversity for the rest of the EZ. However, it is important to note that the species-level sampling in Triodia was under-estimated and included many incorrect identifications, making some interpretations problematic. Unlike *Ptilotus* and *Triodia*, which both arrived from Africa and Asia pre-adapted to aridity, the species of *N*. section *Suaveolentes* developed novel adaptation(s) to aridity *de novo* in the Pilbara before dispersing throughout the EZ. We suspect that a "key" adaptation for *N*. section *Suaveolentes* is strict inhibition of germination until the precise conditions for growth and seed production occur. They exhibit no succulence or other obvious physical attributes typically associated with aridity, and they disappear into the soil seed bank before the annual onset of summer drought and extreme heat, a key adaptation to surviving aridity. The species of *N*. section *Suaveolentes* thus experience the arid zone when it is neither too dry nor too hot, indicating that their distinct ancestral habitats (like those of tomato and potato), while different from current arid conditions, did not hinder their ability to find suitable environments for survival and subsequent conquest of the EZ.

Because of our improved species circumscriptions (not the different geographical areas underpinning their analysis), the main distinction between our conclusions and those of Ladiges *et al.* (2011) is that dispersal rather than vicariance is more explanatory in the evolution of *N.* section *Suaveolentes*. In Ladiges *et al*. (2011), dispersal only came into the picture as the explanation for species with broad distributions, which occurred after vicariance laid down the general patterns, whereas we envisage lineage diversification to have taken place in the Pilbara region, after which dispersal took place twice to other parts of the tropical/subtropical zones (i.e., the *N. benthamiana* and *N. occidentalis* clades; the former inhabiting only mesic sites), followed by the central and southern districts (the *N. simulans* clade) and finally the southern central and far southern areas (the *N. suaveolens* clade). In each of these clades, further diversification outside the Pilbara Region produced species inhabiting both mesic and xeric habitats, and to adequately address in which direction preferences changed more study of species limits/phylogenetics is needed. Most widespread species, e. g., the *N. benthamiana* complex (Cauz-Santos *et al*., 2022; Chase *et al.*, 2022), appear to be species complexes, and we are adding more accessions to our analyses to try to address this issue. Dispersal-mediated founder events and genetic drift could have limited their evolutionary potential as the individual clades left the Pilbara, reducing their evolutionary potential and constraining further change, much as these new species were perhaps spatially constrained by their specializations. The data produced for this study cannot address these issues, but we have collected additional data for more accessions that will permit these topics to be investigated in future studies.

The importance of evaluating species delimitation

For robust understanding of evolutionary history, species limits must be well understood. If species limits have not been assessed properly, then the distributions of such taxa are meaningless biologically. The previous phylogeographic analysis of *N.* section *Suaveolentes*

(Ladiges *et al*., 2011) relied on the species as delimited by Horton (1981), which was the basis for the treatment in the *Flora of Australia* (Purdie *et al*., 1982), plus four others described after the *Flora* treatment. Here, we updated the analyses using a set of species with highly modified circumscriptions, including many newly recognized species (Chase *et al*., 2018, 2021, 2023). For example, in the wider Pilbara region (north-western Western Australia), only three of 15 species were recognized by Horton (1981); the remaining ten species have been described in the last five years, mostly discovered during recent fieldwork (Chase et al., 2021, 2023).

Implications for conservation planning

Our results highlight the evolutionary significance of the Pilbara region, providing a foundation for prioritizing areas for conservation and developing management plans. The phylogenetic and biogeographical results presented in this study can inform conservation strategies by pointing areas with high species diversity and endemism, suggesting these as high-priority zones for conservation efforts. Protecting these areas can help preserve the evolutionary potential and ecological functions of the Pilbara's unique flora.

Here, we show that the Pilbara region represents a major cradle of diversification in *Nicotiana*, which is echoed in other plant (Anderson *et al*., 2016) and animal groups (Ashman *et al*., 2018). It is an extensive, geologically complex, culturally important region for which only 6% sits in formally protected reserves (Government of Western Australia, 2017). However, the explosion of mining activity throughout the region in the past 40 years, with major mine expansions underway and planned, combined with an emerging knowledge of the biodiversity of the region highlight the risks of continued development in the absence of robust, detailed biological surveys by specialists in each group.

An alarming discovery from our study concerns the number of evolutionarily distinct *Nicotiana* species that appear to have extremely restricted distributions, particularly in the Hamersley Basin, the region comprising the unique iron-rich rocks at the core of the Australian mining industry. Extensive biotic surveys prior to mining notwithstanding (McKenzie *et al*., 2009), none of the new species of *Nicotiana* section *Suaveolentes* in the Pilbara region was identified, highlighting the need for detailed genetic and taxonomic studies as part of specialist treatments. This parallels the findings in Anderson *et al.* (2016) in *Triodia*, in which multiple new species were identified in the Pilbara (Barrett & Trudgeon, 2018; Barrett, 2019; Barrett *et al*., 2023). One of our new, narrowly distributed, cryptic-species discoveries from the Pilbara Craton is *Nicotiana karijini* (Chase & Christenhusz, 2018; previously identified as the more widespread *N. umbractica*), for which six of seven known collections were made during mine-site surveys. The possible fate of such species is obvious and leads us to speculate that many new species have gone extinct before they were described.

Conclusions

Our results here have demonstrated that the large arid portions of Australia can act as a catalyst for rapid adaptation and diversification. While this phenomenon has been documented in some reptile groups, particularly arid-zone geckos (Pepper *et al*., 2011ab; Pepper *et al*., 2013; Ashman *et al*., 2018), our study contributes providing detailed insights into plant diversification. Without first examining species limits, our biogeographic conclusions could never have been reached, and thus the first period of our studies focused on their taxonomy. The Australian arid zone is a relatively geologically stable and still largely undisturbed set of environments (despite numerous mines in areas like the Pilbara), making it an ideal setting for studies of speciation and diversification. Topographical heterogeneity combined with nearby marine influences have created localized regions within the Pilbara region with more buffered environments compared to the much larger and more homogeneous surrounding arid zone (Macphail & Stone, 2004; Byrne et al., 2008), permitting ancestrally mesic-adapted taxa like *Nicotiana* to experiment repeatedly with adaptations that have then allowed them to exploit available niches in the rest of the arid zone. Detailed molecular studies across the EZ are at present highly limited, and the high species diversity of the Pilbara Region, especially of cryptic, undescribed species attributable to its status as a mesic refuge, means that it should be an important focus of attention in the future.

Plants and perhaps also terrestrial invertebrates typically have more direct ties to the physical environment than vertebrates because they are inherently less vagile and thus more likely to provide important models to investigate genetic patterns/barriers across the EZ. The Pilbara is an ancient, topographically complex landscape of plateaus, gorges, valleys, and ranges with meteorological extremes and seasonal monsoons/cyclones. We hypothesize that the ancestors of the Pilbara lineages in *N.* section *Suaveolentes* entered the mesic refuges of the Pilbara Craton from the more coastally focused monsoonal region roughly 5 Mya. They became locally adapted to these various ancient and highly stable terrain types and subsequently were exposed repeatedly and became adapted to the interdigitated arid micro-habitats, which then permitted dispersals in several waves to other parts of the arid zone, including most recently (in the last million years) to sandy dune fields and the most homogenous, flat and extremely arid, southern parts of continent (e.g., the Nullarbor). The biotic history of this most ancient landscape, the Pilbara Craton, remains largely unknown and speculative. Our results should provide impetus to develop further an understanding of it and its contribution to the EZ flora and fauna.

Acknowledgements

We thank Juliane Baar and Daniela Paun for constructing RADseq libraries. Financial support to prepare the RAD library and analyze data were funded by research grants P26548-B22, P33028-B of the Austrian Science Fund (FWF) and an award (Dr. Anton Oelzelt-Newin Foundation) from the Austrian Academy of Sciences (ÖAW) to Rosabelle Samuel. The work was also supported by the Marie Curie individual fellowship for CondensDrought project (Grant agreement: 101029312) awarded to Luiz Cauz-Santos. Leaf samples of *N. fatuhivensis* were provided by Kenneth Wood, National Tropical Botanical Garden, Hawai'i. We thank Aureliano Bombarely for giving us earlier access to the *N. benthamiana* v.2.6.1 reference genome.

Data availability statement

The data have been deposited in the NCBI Sequence Read Archive (BioProject ID PRJNA681916), and the files are available under the SRA Study SRP295424.

Benefit sharing statement

Benefits from this research accrue from sharing our datasets on public databases as described above. Research collaborations were developed with scientists from the countries providing genetic samples, and all collaborators are included as co-authors.

Author contributions

MWC and RS conceived this study. The fieldwork and collected/grew of accessions were conducted by MWC and MJMC. LACS, DM and OP analyzed the data. LACS and MWC wrote the manuscript with input from RS, KWD, JGC and OP.

References

Anderson, B. M., Barrett, M. D., Krauss, S. L., & Thiele, K. (2016). Untangling a species complex of arid zone grasses (*Triodia*) reveals patterns congruent with co-occurring animals. *Molecular Phylogenetics and Evolution*, **101**, 142–162.

Ashman, L. G., Bragg, J. G., Doughty, P., Hutchinson, M. N., Bank, S., Matzke, N. J., Oliver, P., & Moritz, C. (2018). Diversification across biomes in a continental lizard radiation. *Evolution*, 72, 1553–1569. https://doi.org/10.1111/evo.13541

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376. https://doi.org/10.1371/journal.pone.0003376

Barrett, M. D. (2019). *Triodia veniciae* (Poaceae), a new species from the Pilbara region, Western Australia. *Nuytsia*, 30, 221–228.

Barrett, M. D., & Trudgeon, M. E. (2018). *Triodia pisoliticola* (Poaceae), a new species from the Pilbara region, Western Australia, and a description for T. sp. Mt Ella (ME Trudgen MET 12739). *Nuytsia*, 29, 271–281.

Barrett, M. D., de Kock, P. L., & Dillon, S. J. (2023). Four new species of *Triodia* (Poaceae) from the Pilbara and adjacent bioregions, Western Australia. *Nuytsia*, 34, 261-294. https://doi.org/10.58828/nuy01060

Bell, R., Parra, J. L., Tonione, M., Hoskin, C., Mackenzie, J. B., Williams, S. E. & Moritz, C. (2010). Patterns of persistence and isolation indicate resilience to climate change in montane rainforest lizards. *Molecular Ecology*, **19**, 2531–2544. https://doi.org/10.1111/j.1365-294X.2010.04676.x

Bombarely, A., Rosli, H. G., Vrebalov, J., Moffett, P., Mueller, L. A. & Martin, G. B. (2012). A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research. *Molecular Plant–Microbe Interactions*, **25**, 1523–1530**.** https://doi.org/10.1094/MPMI-06-12-0148-TA

Booth, C., Adams, V., Kruse, B, & Douglass L. (2022). *The enduring Pilbara: a conservation vision for a land rich in nature, culture and resources.* Malaga, Western Australia: The Environmental Printing Company, Centre for Conservation Geography and University of Tasmania.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bouckaert, R. R., & Heled, J. (2018). DensiTree 2: seeing trees through the forest. *bioRχiv*, 012401. https://doi.org/10.1101/012401

Burbidge, N. (1960). The Australian species of *Nicotiana* L. (Solanaceae). *Australian Journal of Botany*, **8**, 342–380. https://doi.org/10.1071/BT9600342

Byrne, M., Yeates, D. K., Joseph, L., Kearney, M., Bowler, J., Williams, A. J., Cooper, S., Donnellan, S. C., Keogh, J. S., Leys, R., Melville, J., Murphy, D. J., Porch, N., & Wyrwoll, K. H. (2008). Birth of a biome: insights into the assembly and maintenance of the Australian arid zone biota. *Molecular Ecology*, **17**, 4398–4417. https://doi.org/10.1111/j.1365-294X.2008.03899.x

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140. https://doi.org/10.1111/mec.12354

Catullo, R. A., Doughty, P., Roberts, J. D., & Keogh, J. S. (2011). Multi-locus phylogeny and taxonomic revision of *Uperoleia* toadlets (Anura: Myobatrachidae) from the western arid zone of Australia, with a description of a new species. *Zootaxa*, **2902**, 1–43. https://doi.org/10.11646/zootaxa.2902.1.1

Cauz-Santos, L. A., Dodsworth, S., Samuel, R., Christenhusz, M. J. M., Patel, D., Shittu, T., Jakob, A., Paun, O., & Chase, M. W. (2022). Genomic insights into recent species divergence in *Nicotiana benthamiana* and natural variation in *Rdr1* gene controlling viral susceptibility. *The Plant Journal*, **111**, 7–18. https://doi.org/10.1111/tpj.15869.

Chapple, D. G., Chapple, S. N. & Thompson, M. B. (2011). Biogeographic barriers in south□eastern Australia drive phylogeographic divergence in the garden skink, *Lampropholis guichenoti*. *Journal of Biogeography*, **38**, 1761–1775. https://doi.org/10.1111/j.1365-2699.2011.02531.x

Chase, M. W. & Christenhusz, M. J. M. (2018). *Nicotiana karijini*. *Curtis's Botanical Magazine*, **35**, 228–236. http://dx.doi.org/10.1111/curt.12242

Chase, M. W., Christenhusz, M. J. M., Cauz-Santos, L. A., Nollet, F., Bruhl, J. J., Andrew, D. D., Palsson, R., Jobson, R. W., Taseski, G. M. & Samuel, R. (2023). Nine new species of Australian *Nicotiana* (Solanaceae). *Australian Systematic Botany*, 36: 167–205.

Chase, M. W., Christenhusz, M. J. M., Conran, J. G., Dodsworth, S., Medeiros de Assis, F. N., Felix, L. P. & Fay, M. F. (2018). Unexpected diversity of Australian tobacco species (*Nicotiana* section *Suaveolentes*, Solanaceae). *Curtis's Botanical Magazine*, **35**, 212–227. http://dx.doi.org/10.1111/curt.12241

Chase, M. W., Christenhusz, M. J. M., Palsson, R. L., Fay, M. F., Dodsworth, S., Conran, J. G., Cauz□Santos, L. A., Nollet, F., Samuel, R. & Paun, O. (2021). Species delimitation in *Nicotiana* sect. *Suaveolentes* (Solanaceae): reciprocal illumination leads to recognition of many new species. *Curtis's Botanical Magazine*, **38**, 266–286. https://doi.org/10.1111/curt.12410

Chase, M. W., Samuel, R., Leitch, A. R., Conran, J. G., Nollet, F., Fletcher, P., Jakob, A., Cauz-Santos, L. A., Vignolle, G., Dodsworth, S., Christenhusz, M. J. M., Buril, M. T. & Paun, O. (2022). Down, then up: non-parallel genome size changes and a descending chromosome series in a recent radiation of Australian allotetraploid plant species in *Nicotiana* section *Suaveolentes* (Solanaceae). *Annals of Botany*, 131, 123–142. https://doi.org/10.1093/aob/mcac006

Clark, P., Alley, R., & Pollard, D. 1999. Northern Hemisphere ice-sheet influences on global climate change. *Science*, 286, 1104–1111.

Clarkson, J. J., Knapp, S., Garcia, V. F., Olmstead, R. G., Leitch, A. R., & Chase, M. W. (2004). Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions. *Molecular Phylogenetics and Evolution*, **33**, 75–90. https://doi.org/10.1016/j.ympev.2004.05.002

Clarkson, J. J., Lim, K. Y., Kovarik, A., Chase, M. W., Knapp, S., & Leitch, A. R. (2005). Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). *New Phytologist*, **168**, 241–252. https://doi.org/10.1111/j.1469-8137.2005.01480.x

Clarkson, J. J., Kelly, L. J., Leitch, A. R., Knapp, S., & Chase, M. W. (2010). Nuclear glutamine synthetase evolution in *Nicotiana*: phylogenetics and the origins of allotetraploid and homoploid (diploid) hybrids. *Molecular Phylogenetics and Evolution*, **55**, 99–112. https://doi.org/10.1016/j.ympev.2009.10.003

Clarkson, J. J., Dodsworth, S., & Chase, M. W. (2017**).** Time-calibrated phylogenetic trees establish a lag between polyploidisation and diversification in *Nicotiana* (Solanaceae). *Plant Systematics and Evolution*, **303**, 1001–1012. https://doi.org/10.1007/s00606-017-1416-9

Cracraft, J. (1991). Patterns of diversification within continental biotas: hierarchical congruence among the areas of endemism of Australian vertebrates. *Australian Systematic Botany*, **4**, 211–227. https://doi.org/10.1071/SB9910211

Crisp, M. D., Linder, P. H. & Weston, P. H. (1995). Cladistic biogeography of plants in Australia and New Guinea: congruent pattern reveals two endemic tropical tracks. *Systematic Biology*, **44**, 457–473. https://doi.org/10.2307/2413654

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Domb, K., Katz, A., Harris, K. D., Yaari, R., Kaisler, E., Nguyen, V. H., Hong, U. V. T., Griess, O., Heskiau, K. G., Ohad, N. & Zemach, A. (2020). DNA methylation mutants in *Physcomitrella patens* elucidate individual roles of CG and non-CG methylation in genome regulation. *Proceedings of the National Academy of Sciences*, **117**, 33700–33710. https://doi.org/10.1073/pnas.2011361117

Donoghue, M. J. (2008). A phylogenetic perspective on the distributio*n of plant diversity*. *Proceedings of the National Academy of Sciences USA*, **105**, 11549–51155. https://doi.org/10.1073/pnas.0801962105

Doughty, P. A., Pepper, M. & Keogh, J. S. (2010). Morphological and molecular assessment of the *Diplodactylus savagei* species complex in the Pilbara region, Western Australia, with a description of a new species. *Zootaxa*, 2393, 33–45. https://doi.org/10.11646/zootaxa.2393.1.3

Doughty, P., Kealley, L., & Donnellan, S. C. (2011). Revision of the pygmy spiny-tailed skinks (*Egernia depressa* species-group) from Western Australia, with descriptions of three new species. *Records of the Western Australian Museum*, **26**, 115–137. https://doi.org/10.18195/issn.0312-3162.26(2).2011.115-137

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973. https://doi.org/10.1093/molbev/mss075

Edwards, E. J., & M. J. Donoghue. (2013). Is it easy to move and easy to evolve? Evolutionary accessibility and adaptation. *Journal of Experimental Botany*, **64**, 4047–4052. https://doi.org/10.1093/jxb/ert220

Flower, B. P., & Kennett, J. P. (1994). The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **108**, 537– 555. https://doi.org/10.1016/0031-0182(94)90251-8

Fujioka, T., Chappell, J., Fifield, L.K., & Rhodes, E.J., 2009. Australian desert dune fields initiated with Pliocene-Pleistocene global climatic shift. *Geology* **37**, 51–54.

Guthrie, N. A., Weir, T. & Will, K. (2010). Localised and regional patterns in ground-dwelling beetle assemblages in a semi-tropical arid zone environment. *Records of the Western Australian Museum*, **78**, 169–184. https://doi.org/10.18195/issn.0313-122x.78(1).2010.169-184.

Government of Western Australia. (2017). Pilbara conservation strategy. https://www.dpaw.wa.gov.au/images/documents/conservation-management/pilbara/pilbara_conservation_strategy.pdf

Hammer, T. A., Renton, M., Mucina, L. & Thiele, K. R. (2021). Arid Australia as a source of plant diversity: the origin and climatic evolution of *Ptilotus* (Amaranthaceae). *Australian Systematic Botany*, **34**, 570–586. https://doi.org/10.1071/SB21012

Horton, P. (1981). A taxonomic revision of *Nicotiana* (Solanaceae) in Australia. *Journal of the Adelaide Botanic Garden*, **3**, 1–56. https://www.jstor.org/stable/23872354

Kadereit, G. & Freitag, H. (2011). Molecular phylogeny of Camphorosmeae (Camphorosmoideae, Chenopodiaceae): implications for biogeography, evolution of C4-photosynthesis and taxonomy. *Taxon*, **60**, 51–78. https://.doi.org/10.1002/tax.601006

Kay, G. & Keogh, J. S. (2012). Molecular phylogeny and morphological revision of the *Ctenotus labillardieri* (Reptilia: Squamata: Scincidae) species group and a new species of immediate conservation concern. *Zootaxa*, **3390**, 1–18. https://doi.org/10.11646/zootaxa.3390.1.1

Kelly, L. J., Leitch, A. R., Clarkson, J. J., Knapp, S,. & Chase, M. W. (2013). Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section *Suaveolentes*). *Evolution*, **76**, 80–94. https://doi.org/10.1111/j.1558-5646.2012.01748.x

Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, **15**, 356. https://doi.org/10.1186/s12859-014-0356-4

Ladiges, P., Parra□O, C., Gibbs, A., Udovicic, F., Nelson, G., & Bayly, M., 2011. Historical biogeographical patterns in continental Australia: congruence among areas of endemism of two major clades of eucalypts. *Cladistics*, **27**, 29–41.

Ladiges, P. Y., Marks, C. E., & Nelson, G. (2011). Biogeography of *Nicotiana* section *Suaveolentes* (Solanaceae) reveals geographical tracks in arid Australia. *Journal of Biogeography*, **38**, 2066–2077. https://doi.org/10.1111/j.1365-2699.2011.02554.x

Levy, E., Byrne, M., Coates, D. J., Macdonald, B. M., McArthur, S., & Van Leeuwen, S. (2016). Contrasting influences of geographic range and distribution of populations on patterns of genetic diversity in two sympatric Pilbara acacias. *PLoS One*, 11, e0163995.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50**, 913–925. https://doi.org/10.1080/106351501753462876

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299. https://doi.org/10.1093/bioinformatics/btr642

Marks, C. E., Newbigin, E., & Ladiges, P. Y. (2011). Comparative morphology and phylogeny of *Nicotiana* section *Suaveolentes* (Solanaceae) in Australia and the South Pacific. *Australian Systematic Botany*, **24**, 61–86. https://doi.org/10.1071/SB11006

Martin, H. A. (2006.) Cenozoic climatic change and the development of the arid vegetation in Australia. *Journal of Arid Environments*, **66,** 533–563.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303. https://doi.org/10.1101/gr.107524.110.

McKenzie, N. L., van Leeuwen, S. & Pinder, A. M. (2009). Introduction to the Pilbara biodiversity survey, 2002–2007. *Records of the Western Australian Museum*, **78**, 3–89. https://doi.org/10.18195/issn.0313-122x.78(1).2009.003-089

Macphail, M. K. & Stone, M. S. (2004). Age and palaeoenvironmental constraints on the genesis of the Yandi channel iron deposits, Marillana Formation, Pilbara, northwestern Australia. *Australian Journal of Earth Sciences*, **51**, 497–520. https://doi.org/10.1111/j.1400-0952.2004.01071.x

Meisner, J. & Albrechtsen, A. (2018**)**. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, **210**, 719–731. https://doi.org/10.1534/genetics.118.301336

Millar, M. A., Binks, R. M., Tapper, S. L., Macdonald, B. M., McArthur, S. L., Hankinson, M., Coates, D. J., van Leeuwen, S., & Byrne, M. (2022). Limited phylogeographic and genetic connectivity in *Acacia* species of low stature in an arid landscape. Ecology and Evolution, **12**, e9052.

Mummenhoff, K. & Franzke, A. (2007). Gone with the bird: Late Tertiary and Quaternary intercontinental long-distance dispersal and allopolyploidization in plants. *Systematics and Biodiversity*, **5**, 255–260. https://doi.org/10.1017/S1477200007002393

Myers, J. S. (1993). Precambrian history of the west Australian craton and adjacent orogens. *Annual Review of Earth and Planetary Sciences*, **21**, 453–485. https://doi.org/10.1146/annurev.ea.21.050193.002321

Nistelberger, H. M., Binks, R. M., van Leeuwen, S., Coates, D. J., McArthur, S. L., Macdonald, B. M., Hankinson, M., & Byrne, M. (2020). Extensive genetic connectivity and historical persistence are features of two widespread tree species in the ancient Pilbara region of Western Australia. *Genes*, **11**, 863.

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M. W. & Samuel, R. (2016). Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. *Systematic Biology*, **65**, 212–227. https://doi.org/10.1093/sysbio/syv076

Pepper, M., Doughty, P., Arculus, R. & Keogh, J. S. (2008**)**. Landforms predict phylogenetic structure on one of the world’s most ancient surfaces. *BMC Evolutionary Biology*, **8**, 152. https://doi.org/10.1186/1471-2148-8-152

Pepper, M., Ho, S., Fujita, M. K. & Keogh, J. S. (2011a**).** The genetic legacy of aridification: Miocene refugia fostered diversification while Pleistocene climate cycles erased diversity in desert lizards. *Molecular Phylogenetics and Evolution*, **61**, 750–759. https://doi.org/10.1016/j.ympev.2011.08.009

Pepper, M., Doughty, P., Hutchinson, M. N., & Keogh, J. S. (2011b). Ancient drainages divide cryptic species in Australia’s arid zone: Morphological and multi-gene evidence for four new species of Beaked Geckos (Rhynchoedura). Molecular Phylogenetics and Evolution, 61(3), 810-822. https://doi.org/10.1016/j.ympev.2011.08.012

Pepper, M., Doughty, P. & Keogh, J. S. (2013). Geodiversity and endemism in the iconic Australian Pilbara region: a review of landscape evolution and biotic response in an ancient refugium. *Journal of Biogeography*, **40**, 1225–1239. https://doi.org/10.1111/jbi.12080

Pickrell, J. K. & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, **8**, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Purdie, R. W., Symon, D. E. & Haegi, L. (1982). *Nicotiana*. In: George AS, ed. *Flora of Australia*, vol. 29. Canberra, Australia: Australian Government Publishing Service, 38–57.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, **67**, 901–904. https://doi.org/10.1093/sysbio/syy032

Schiavinato, M., Marcet□Houben, M., Dohm, J. C., Gabaldón, T. & Himmelbauer, H. (2020). Parental origin of the allotetraploid tobacco *Nicotiana benthamiana*. *The Plant Journal*, **102**, 541–554. https://doi.org/10.1111/tpj.14648

Shepherd, K. A., Waycott, M. & Calladine, A. (2004). Radiation of the Australian Salicornioideae (Chenopodiaceae) – based on evidence from nuclear and chloroplast

DNA sequence. *American Journal of Botany*, **91**, 1387–1397. https://doi.org/10.3732/ajb.91.9.1387

Shoo, L. P., Rose, R., Doughty, P., Austin, J. J. & Melville, J. (2008**).** Diversification patterns of pebble-mimic dragons are consistent with historical disruption of important habitat corridors in arid Australia. *Molecular Phylogenetics and Evolution*, 48, 528–542. https://doi.org/10.1016/j.ympev.2008.03.022

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Toon, A., Crisp, M. D., Gamage, H., Mant, J., Morris, D. C., Schmidt, S., Cook, L. G. (2015). Key innovation or adaptive change? A test of leaf traits using Triodiinae in Australia. *Scientific Reports*, **5**, 12398. https://doi.org/10.1038/srep12398

Warnes, G. R., Bolker, B., Bonebakker L, et al. (2020). gplots: Various R programming tools for plotting data. R package gplots version 3.1. 0.

Weeks, J. P. (2010). plink: An *R* package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, **35**, 1–33. https://doi.org/10.18637/jss.v035.i12

Table 1. Biogeographic stochastic mapping (BSM) counts for *N*. section *Suaveolentes* using the Australian drainage regions as operational areas and the DIVALIKE+J model selected in BioGeoBEARS. Mean values (mean) and standard deviations (stdev) are event counts of 100 BSMs.

Mode	**Type**	**mean**	**stdev**	**%**
Within area speciation	Sympatry - narrow	22.98	1.59	36.47
Dispersal	Jump-dispersal	38.23	1.69	60.68
Vicariance		1.79	0.46	2.84
Total events		63		

Figure captions

Fig. 1. Habitat diversity, mesic and arid, among species of *Nicotiana* section *Suaveolentes* in the Australian Eremaean Zone. (a) Mesic site for *N. cavicola* near the entrance to a cave near Meekathara (Western Australia). (b) Mesic site for *N. maritima* on the sea cliffs near St. Vincent, York Peninsula (South Australia). (c) Arid site for *N. simulans* on the gibber plains south of Ooodnadatta (South Australia). (d) Mesic site for *N. insecticida* in a mulga woodland (*Acacia aneura* species complex) near Carbla (Western Australia). (e) Arid site for *N. velutina* in a fossilised sand dune near Coober Peddy (South Australia). (f) Mesic site for *N. insecticida* in the Anthwerrke Gap, East MacDonnell Ranges (Northern Territory).

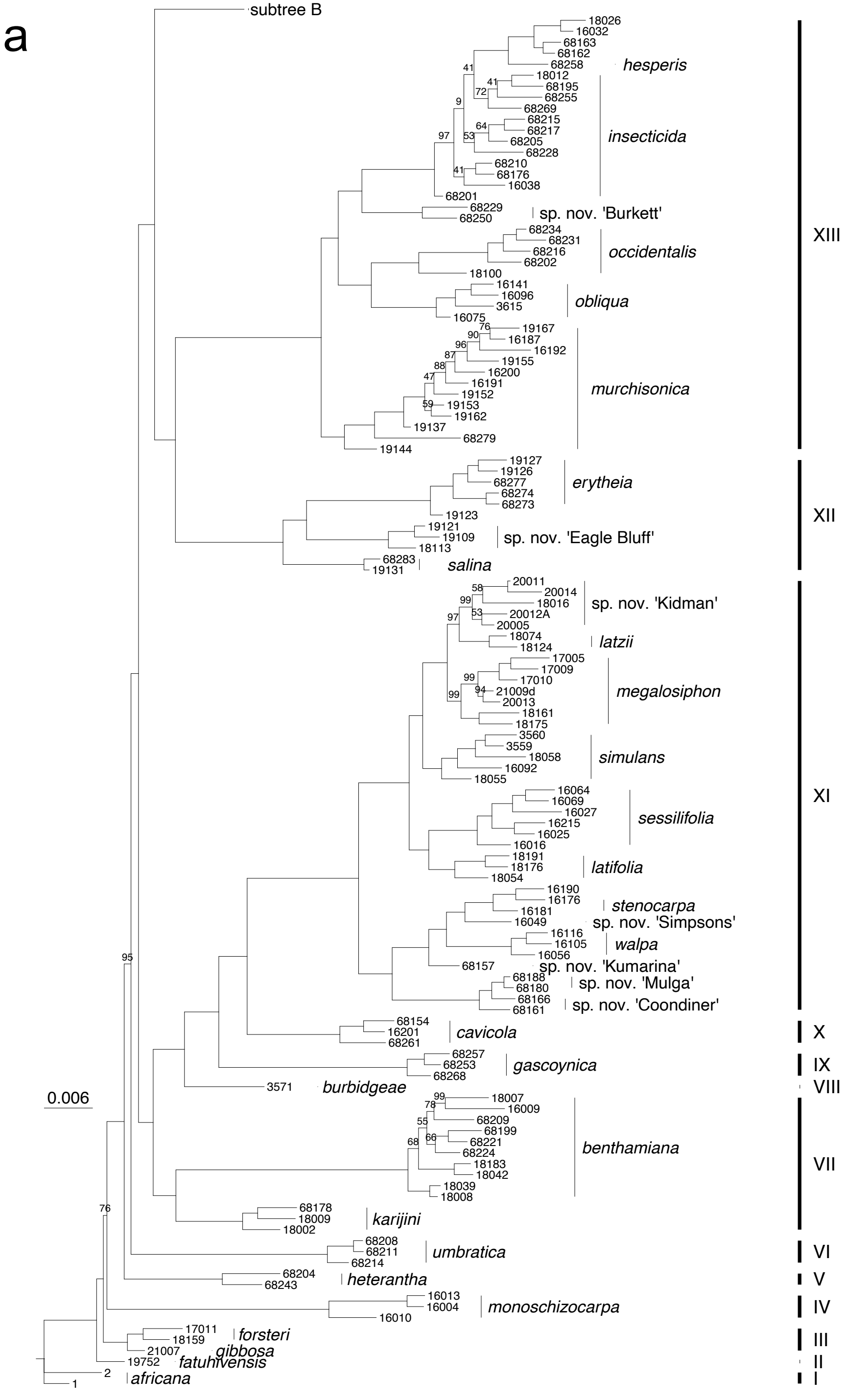
Fig. 2. RAxML phylogenetic tree from *N.* section *Suaveolentes*, subtrees (a) and (b), based on 240,871 SNPs. The numbers are bootstrap percentages; branches without numbers received 100.

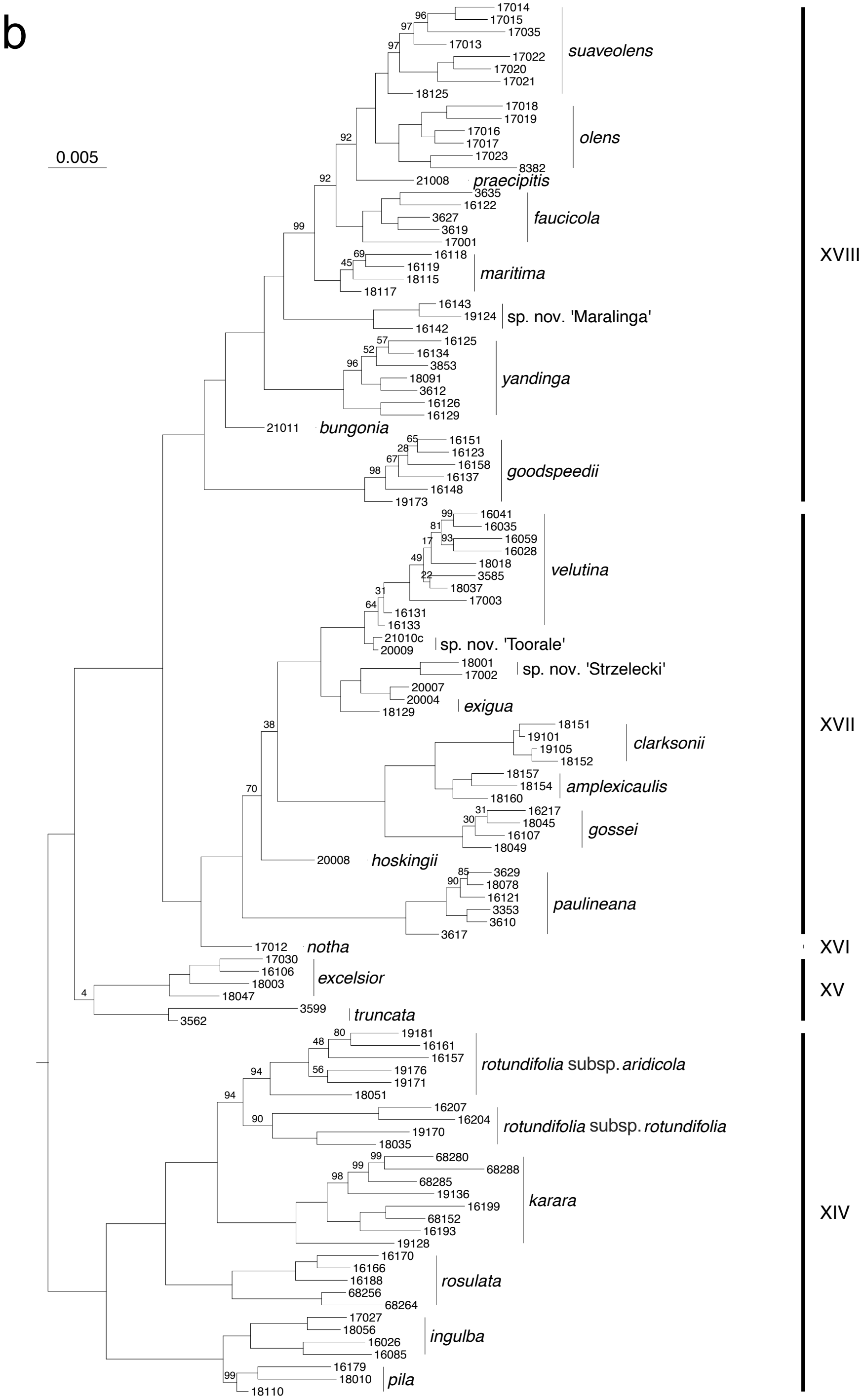
Fig. 3. Coancestry heatmap from the pairwise relatedness between accessions from *N.* section *Suaveolentes*. The darker colours indicate higher relatedness; the estimates of relationship of individuals to themselves are not shown. The blue lines specify the accessions of *N. africana*. High levels of coancestry among recently diverged species (e.g., *N. goodspeedii/N. maritima/N. suaveolens*) are to be expected.

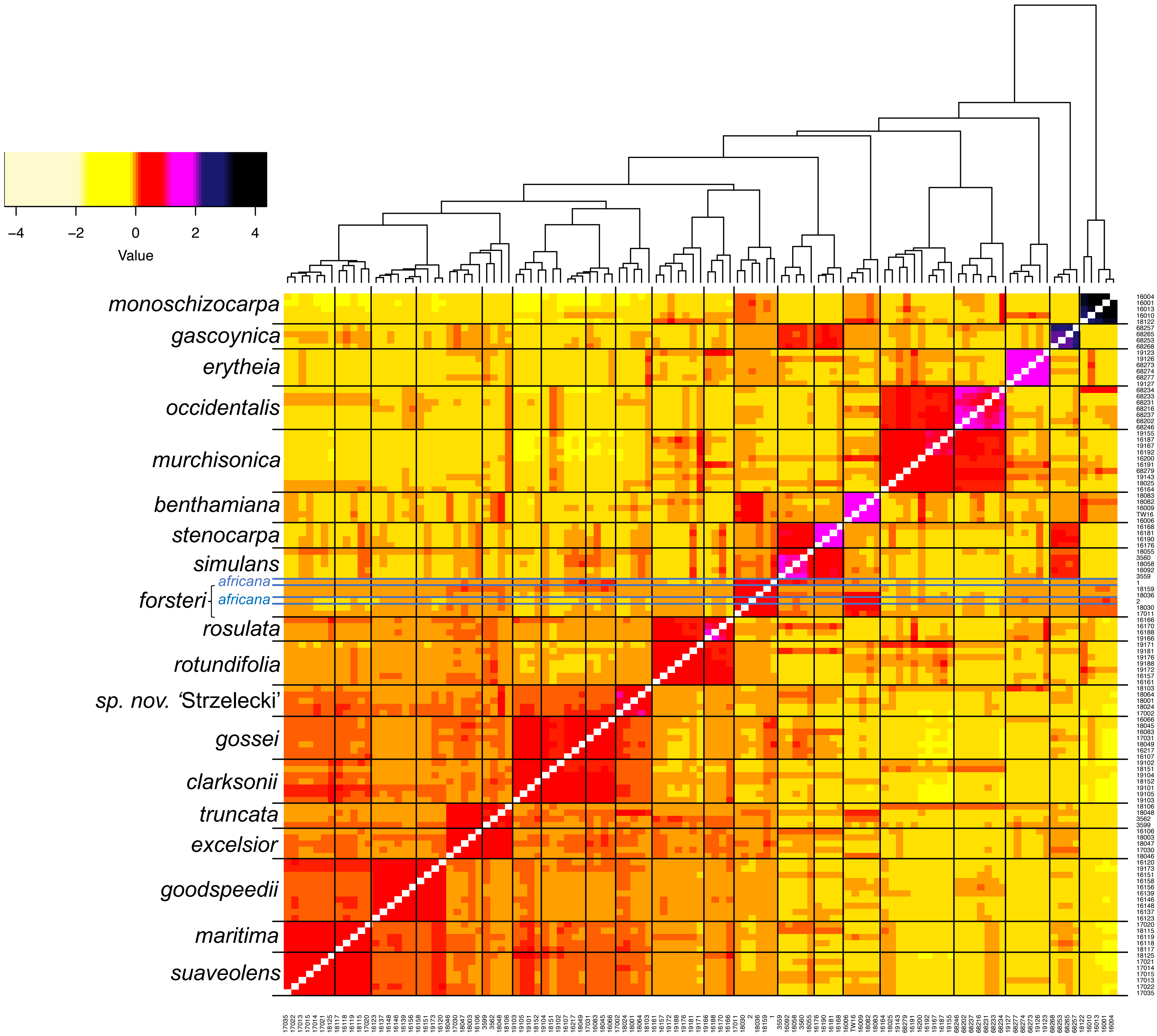
Fig. 4. Cloudogram from species trees generated for *Nicotiana section Suaveolentes*, scaled to divergence times. The species tree reveals three of the main radiations in the section, the *N. occidentalis*, *N. gascoynica* and *N. suaveolens* clades, the last being the most recent radiation (ca. 1 Mya) comprising the species occurring in central and southern Australia.

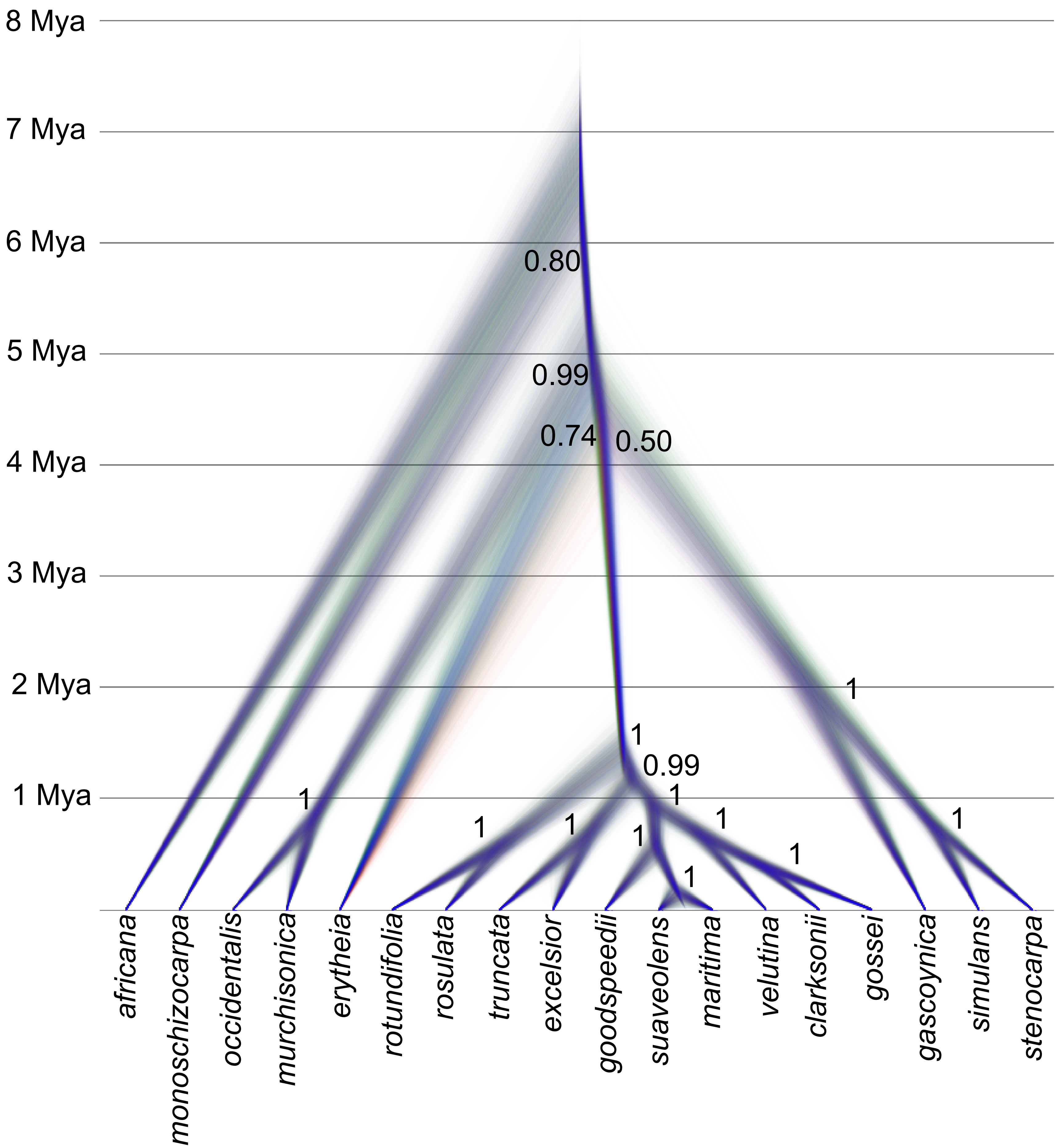
Fig. 5. Biogeographic history of *N.* section *Suaveolentes*. Ancestral area reconstruction using the drainage areas as a model and showing greater Pilbara region (region F/light green) as the ancestral range for the species adapted to the Eremaean Zone (ca. 5 Mya). The map in the upper part of the figure represents the biogeography model used for distributions of the species in *N.* section *Suaveolentes*. The map shows the Australian drainages employed here to evaluate dispersal versus vicariance during diversification of *N.* section *Suaveolentes*.

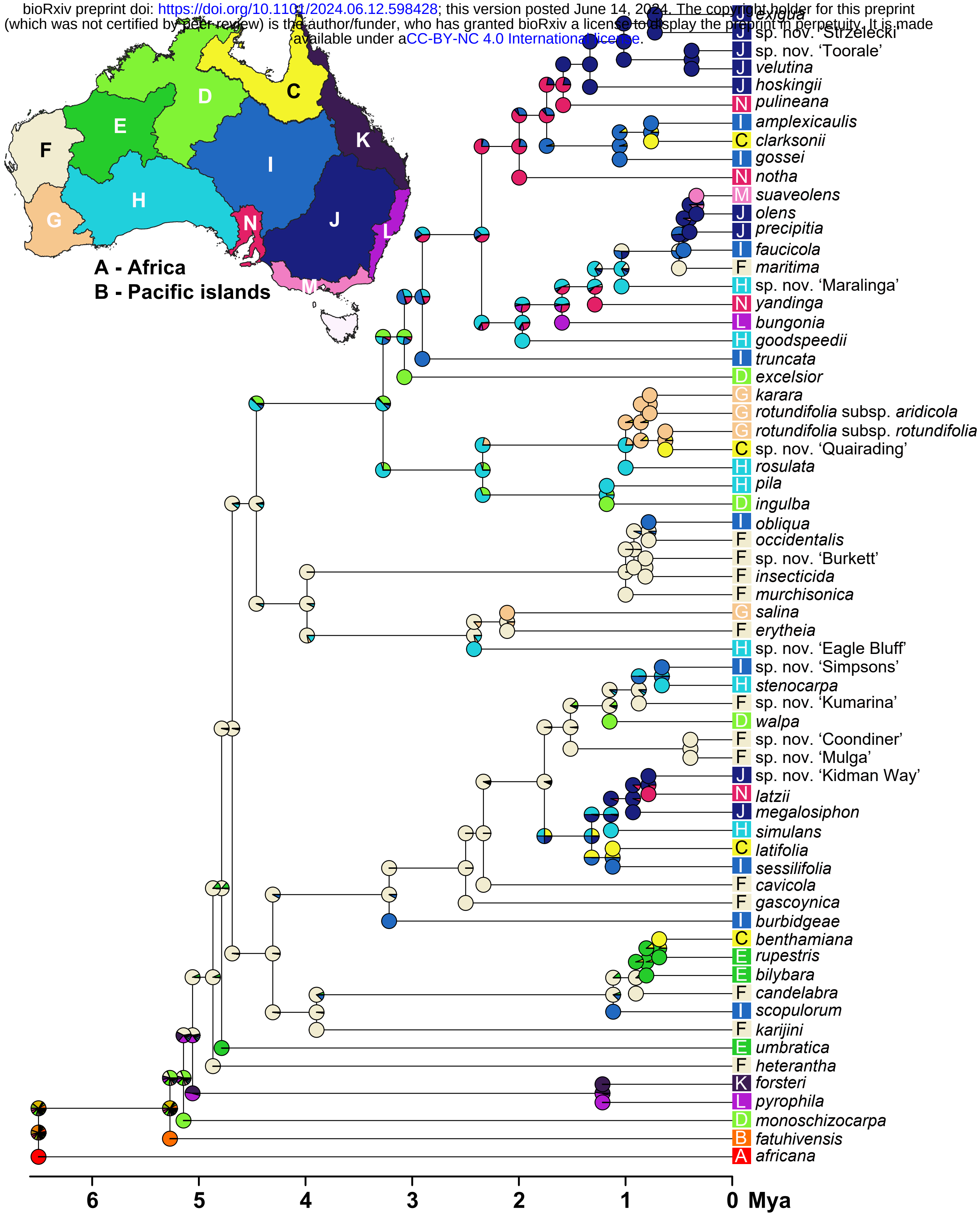

(a)
(b)
(c)
(d)
(e)
(f)


a
subtree B
18026
16032
68163
68162
68258
hesperis
41
18012
41
68195
72
68255
68269
9
insecticida
64
68215
68217
97
53
68205
68228
68210
41
68176
16038
68201
68229
68250
sp. nov. 'Burkett'
XIII
68234
68231
68216
occidentalis
68202
18100
16141
16096
obliqua
3615
16075
76
19167
90
16187
96
16192
87
19155
88
16200
47
16191
murchisonica
19152
59
19153
19162
19137
68279
19144
19127
19126
68277
erytheia
68274
68273
XII
19123
19121
19109
sp. nov. 'Eagle Bluff'
18113
68283
salina
19131
20011
58
20014
99
18016
sp. nov. 'Kidman'
53
20012A
97
20005
18074
latzii
18124
17005
17009
99
17010
megalosiphon
94
21009d
99
20013
18161
18175
3560
3559
simulans
18058
16092
18055
XI
16064
16069
16027
sessilifolia
16215
16025
16016
18191
18176
latifolia
18054
16190
16176
stenocarpa
16181
16049
sp. nov. 'Simpsons'
16116
16105
walpa
16056
95
68157
sp. nov. 'Kumarina'
68188
sp. nov. 'Mulga'
68180
68166
sp. nov. 'Coondiner'
68161
68154
cavicola
X
16201
68261
68257
68253
gascoynica
IX
68268
3571
burbidgeae
VIII
0.006
99
18007
78
16009
55
68209
68199
66
68221
68
68224
benthamiana
VII
18183
18042
18039
18008
68178
18009
karijini
18002
76
68208
68211
umbratica
VI
68214
68204
heterantha
V
68243
16013
16004
monoschizocarpa
IV
16010
17011
forsteri
18159
III
21007
gibbosa
19752
fatuhivensis
II
2
africana
I
1


b
0.005
17014
17015
17035
17013
17022
17020
17021
18125
suaveolens
17018
17019
17016
17017
17023
8382
olens
21008
praecipitis
3635
16122
3627
3619
17001
faucicola
16118
16119
18115
18117
maritima
16143
19124
16142
sp. nov. 'Maralinga'
16125
16134
3853
18091
3612
16126
16129
yandinga
21011
bungonia
16151
16123
16158
16137
16148
19173
goodspeedii
XVIII
16041
16035
16059
16028
18018
3585
18037
17003
16131
16133
velutina
21010c
20009
sp. nov. 'Toorale'
18001
17002
sp. nov. 'Strzelecki'
20007
20004
18129
exigua
18151
19101
19105
18152
clarksonii
18157
18154
18160
amplexicaulis
16217
18045
16107
18049
gossei
20008
hoskingii
3629
18078
16121
3353
3610
3617
paulineana
XVII
17012
notha
XVI
17030
16106
18003
18047
excelsior
3599
3562
truncata
XV
19181
16161
16157
19176
19171
18051
rotundifolia subsp. aridicola
16207
16204
19170
18035
rotundifolia subsp. rotundifolia
68280
68288
68285
19136
16199
68152
16193
19128
karara
16170
16166
16188
68256
68264
rosulata
17027
18056
16026
16085
ingulba
16179
18010
18110
pila
XIV
96
97
97
92
92
99
69
45
57
52
96
65
28
67
98
99
81
93
17
49
22
31
64
38
70
4
80
48
56
94
94
90
99
99
98
99


−4
−2
0
2
4
Value
monoschizocarpa
gascoynica
erytheia
occidentalis
murchisonica
benthamiana
stenocarpa
simulans
africana
forsteri
africana
rosulata
rotundifolia
sp. nov. ‘Strzelecki’
gossei
clarksonii
truncata
excelsior
goodspeedii
maritima
suaveolens


8 Mya
7 Mya
6 Mya
5 Mya
4 Mya
3 Mya
2 Mya
1 Mya
0.80
0.99
0.74
0.50
1
1
0.99
1
1
1
1
1
1
1
1
1
1
africana
monoschizocarpa
occidentalis
murchisonica
erytheia
rotundifolia
rosulata
truncata
excelsior
goodspeedii
suaveolens
maritima
velutina
clarksonii
gossei
gascoynica
simulans
stenocarpa



C
D
E
F
G
H
I
J
K
L
M
N
A - Africa
B - Pacific islands
J exigua
J sp. nov. 'Strzelecki'
J sp. nov. 'Toorale'
J velutina
J hoskingii
N pulineana
I amplexicaulis
C clarksonii
I gossei
N notha
M suaveolens
J olens
J precipitia
I faucicola
F maritima
H sp. nov. 'Maralinga'
N yandinga
L bungonia
H goodspeedii
I truncata
D excelsior
G karara
G rotundifolia subsp. aridicola
G rotundifolia subsp. rotundifolia
C sp. nov. 'Quairading'
H rosulata
H pila
D ingulba
I obliqua
F occidentalis
F sp. nov. 'Burkett'
F insecticida
F murchisonica
G salina
F erytheia
H sp. nov. 'Eagle Bluff'
I sp. nov. 'Simpsons'
H stenocarpa
F sp. nov. 'Kumarina'
D walpa
F sp. nov. 'Coondiner'
F sp. nov. 'Mulga'
J sp. nov. 'Kidman Way'
N latzii
J megalosiphon
H simulans
C latifolia
I sessilifolia
F cavicola
F gascoynica
I burbidgeae
C benthamiana
E rupestris
E bilybara
F candelabra
I scopulorum
F karijini
E umbratica
F heterantha
K forsteri
L pyrophila
D monoschizocarpa
B fatuhivensis
A africana
6
5
4
3
2
1
0 Mya