Comparative analysis of the complete plastid genomes of desert trees Neltuma and Strombocarpa genera

Roberto Contreras-Díaz (roberto.contreras@uda.cl)
Universidad de Atacama
Felipe S. Carevic
Universidad Arturo Prat
Liesbeth van den Brink
ECOBIOSIS, Universidad de Concepción
Wilson Huanca-Mamani
Universidad de Tarapacá
Patrick Jung
University of Applied Sciences Kaiserslautern

Article

Keywords:

Posted Date: October 27th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3422232/v1

Additional Declarations: No competing interests reported.

Abstract

Neltuma alba (Algarrobo blanco), *Neltuma chilensis* (Algarrobo Chileno) and *Strombocarpa strombulifera* (Fortuna) are some of the few trees found in small highly fragmented populations, throughout the Atacama Desert, indicating their drought resistance. We found that the complete chloroplast genomes of *N. alba* and *N. chilensis* are larger in size compared to species of the *Strombocarpa* genus. However, the *Strombocarpa* species presented slightly more GC content than the *Neltuma* species. Therefore, we assume that *Strombocarpa* species have been exposed to stronger evolution than *Neltuma* species. We observed high variation values in the number of cpSSRs (chloroplast simple sequence repeats) and repeated elements among *Neltuma* and *Strombocarpa* species. Very low nucleotide diversity values were found in *Neltuma*, while ten highly variable regions found in *Strombocarpa*, can likely be used to resolve uncertainties in phylogeny, and for DNA barcoding. Although in general our study supports the phylogeny of other studies, the biggest inconsistency was the nesting of *Prosopis cineraria* within the *Neltuma* clade and showed a divergence time of 1.85 Mya. With this study we provide valuable information about isolated populations of tree species that provide important ecosystem services in hostile environments before they disappear, due to an ongoing fragmentation of their populations.

Introduction

Legumes have a cosmopolitan distribution and are ecologically important in almost all biomes of the world as their fruits are a food source or their root biology nourishes soils [1,2], even in ecosystems as extreme as the Atacama Desert. *Leguminosae* (*Fabaceae*) is one of the largest angiosperm family in terms of species numbers and one of the most diverse family and is classified into three subfamilies (*Caesalpinioideae*, *Mimosoideae*, and *Papilionoideae*), which have close to 770 genera and over 19,500 species [1,3]. Burkart (1976)[4] taxonomic monograph of the genus *Prosopis* L. (Mesquite) recognized 44 species, which are distributed across Southwest Asia, Africa, and (predominantly) America [5,6]. *Prosopis* species from the *Strombocarpa* Bentham and *Algarobia* DC. Emend. Burk sections are trees and normally inhabitants of arid and semiarid regions [6,7]. These species inhabit the Atacama Desert in Chile, which extends for over 1000 km between latitudes 19°S and 30°S and is bordered by the Coastal Cordillera to the west and the Andean Cordillera to the east [8]. Surviving in a place as hostile as the Atacama Desert, where radiation is high, and so is water stress [9,10] is already an accomplishment, but these trees also provide local people with important resources such as fruits, juice, futter and wood [7]. Three species from the formerly known *Strombocarpa* section (*Prosopis strombulifera* (Argentine screwbean) and the endemics *Prosopis burkartii* and *Prosopis tamarugo* (Tamarugo)) [7,11] and individuals belonging to different species from the formerly know *Algarobia* section (*Prosopis chilensis*, *Prosopis flexuosa*, and *Prosopis alba*) can be found in the Atacama Desert [6]. The scientific names of these species and the concept of *Prosopis* established by Bentham (1875)[12] and Burkart (1976)[4] has only currently been disintegrated, because *Prosopis* was found to be polyphyletic based on both chloroplast (cpDNA) and nuclear DNA (nDNA) [5,13,14]. As a consequence, the old *Prosopis* cluster was divided in six genera – *Anonychium*, *Prosopis*, *Neltuma*, *Strombocarpa*, *Xerocladia* and *Indopiptadenia*. The species of the above mentioned *Algarobia* section were renamed as *Neltuma alba*, *Neltuma chilensis*, *Neltuma flexuosa*, and the species of the above mentioned *Strombocarpa* section as *Strombocarpa tamarugo*, *Strombocarpa burkartii* and *Strombocarpa strombulifera* [5]. This division was based on short DNA sequences, as is common practice in taxonomy. However, compared to short DNA sequences, a complete chloroplast genome of approximately 160,000 bp can offer more information about the phylogenetic relationships and gives a full overview of the specific genes and the structure of its genome.

Chloroplast and mitochondria are the powerhouse of a plant cell, responsible for the majority of the energy produced, through photosynthesis and photorespiration, respectively; these organelles have originated endosymbiotically from eubacterial cells, where most genes were either lost or transferred to the host nucleus [15]. The chloroplast genome is a valuable taxonomic resource with rich genetic information [16], as it is highly conserved and maternally inherited [17]. In

angiosperms, most chloroplast genomes are composed of circular DNA molecules and have a quadripartite organization consisting of two copies of inverted repeats (IRs), which divide the rest of chloroplast genome into a large single copy (LSC) region and small single copy (SSC) region [17]. There are approximately 110–130 genes included in chloroplast DNA, consisting of rRNA-coding genes, protein-coding genes, and tRNA-coding genes [16]. Because the chloroplast genome can provide valuable information to support the conservation of threatened trees [18], gaining insights in chloroplast DNA of the legume tree populations from Atacama Desert could help their conservation. Chloroplast genome sequences are commonly used in plant phylogeny, phylogeographic and genome evolution studies [16,17]. Lately, the use of complete chloroplast genome as a "super-barcoding" method has become an excellent approach allowing for the increase of the phylogenetic resolution at lower taxonomic levels in plants [19,20]. However, in the Atacama Desert only a few plastomes of the native and endemic herbaceous plants [21,22], shrubs [23,24] and leguminous trees [25,26] have been characterized so far.

Unfortunately, several species of trees of the genera *Neltuma* and *Strombocarpa* are in vulnerable and endangered conservation status in Chile, e.g. *Neltuma chilensis* and *Strombocarpa tamarugo*. *Neltuma chilensis* and *Neltuma alba* are restricted to southern Peru, northern and central Chile, southwestern Bolivia and northwestern, western and central Argentina [27–29]. The not threatened *Strombocarpa strombulifera* is widely distributed from the Arizona desert (U.S.A.) to Patagonia (Argentina) [30]. However, in the Atacama Desert, *Neltuma alba, Neltuma chilensis*, as well as *Strombocarpa strombulifera* populations, are fragmented and restricted to oases or valley (forming populations of only a few individuals), and geographically isolated from each other by large areas of land [6,7,31]. Although *Neltuma alba* and *Strombocarpa strombulifera* are in the conservation status of "Least Concern" it is urgently necessary to identify the plastomes of their genera now, before their more endangered cousins go extinct. Until now, there is no complete chloroplast genome available for any *N. alba*, *N. chilensis* and *S. strombulifera*, but they are needed to confirm phylogenetic relationships between them, and with closely related species. Therefore, in this study we provide and analyze the complete chloroplast genomes of *N. alba*, *N. chilensis* and *S. strombulifera*, in terms of structure, gene composition, divergence time and phylogeny.

Methods

Plant material and DNA isolation

Fresh leaves of *Neltuma alba* (Griseb.) C.E. Hughes & G.P. Lewis, *Neltuma chilensis* (Molina) C.E. Hughes & G.P. Lewis and *Strombocarpa strombulifera* (Lam.) A. Gray were collected in Copiapó (27°21'39.3"S 70°20'33.8"W), Chacabuco (33°05'24.9"S 70°39'07.3"W) and Pampa del Tamarugal (20°27'59.9"S 69°33'23.5"W) in Chile, respectively. Identification of samples was done according to the taxonomic criteria described by Burkart (1976)[4]. Additionally, the samples were verified by the forester Boris Burgos of the Corporación Nacional Forestal (CONAF) from the Atacama Region. The specimens were deposited in the Departamento de Silvicultura y Conservación de la Naturaleza herbarium of Universidad de Chile (under the names that were correct at the time of deposition: *Prosopis alba*, EIF13329; *Prosopis chilensis*, EIF13328; and *Prosopis strombulifera*, EIF13350). DNA was isolated from the leaves using the modified cetyl-trimethylammonium bromide (CTAB) protocol [7]. The DNA was quantified with a Qubit™ 3.0 fluorometer and a Qubit™ dsDNA HS Assay Kit, according to the protocol supplied by the manufacturer. DNA integrity was verified with an Agilent 2100 Bioanalyzer prior to sequencing.

Genome sequencing, assembling and annotation

Sequencing libraries were generated by a TruSeq Nano DNA LT Kit (Illumina, San Diego, CA). The final libraries were run on an Agilent 2100 Bioanalyzer to verify the fragment size distribution and concentration. Sequencing was performed with an Illumina sequencing platform, at Genoma Mayor (Universidad Mayor, Chile). Paired-end sequences of 150 bp were generated for each read (R1 and R2). The filtered reads were assembled using SPAdes 4 software version 3.13.0 [32], using

three k-mers parameters: - k 33, 55 and 77. The plastid was annotated with PGA software [33] and CPGAVAS2 [34], after which it was manually corrected when needed. The graphical map of the plastid was generated by Organellar Genome DRAW (OGDRAW) [35], and the complete nucleotide sequences were deposited in the NCBI GenBank database (OP672364, OP672365 and OP672366, under the names *Prosopis alba*, *Prosopis chilensis* and *Prosopis strombulifera*, respectively).

Genome comparison, repeat and phylogenetic analysis

The plastid structures (LSC/IR, IR/SSC) of *N. alba*, *N. chilensis* and *S. strombulifera* and of five closely related species, i.e. *Neltuma juliflora* (Sw.) Raf., *Strombocarpa tamarugo* (Phil.) C.E. Hughes & G.P. Lewis, *Prosopis farcta* (Banks & Sol.) J.F. Macbr. and *Prosopis cineraria* (L.) Druce, of the *Mimoseae* tribe were visualized and compared using IRScope [36]. We used sequence data of whole plastome (obained from GenBank) of species from the genera *Neltuma*, *Strombocarpa* and *Prosopis* for the identification of the simple sequence repeats (SSRs). These SSRs were identified using MISA software [37] with the following search parameters: ten for mononucleotide, eight for dinucleotide, four for trinucleotide and tetranucleotide, and three for pentanucleotide and hexanucleotide. To identify the tandem repeats (forward, palindromic, reverse, and complement) of these species we used REPuter [38] with the following parameters: hamming distance equal to 3, minimal repeat size set to 30 bp, and maximum computed repeats set to 300 bp. A sliding window analysis (window length: 600 pb, step size: 200 bp) was performed to assess the variability (Pi) between plastid of *N. alba* and *N. chilensis*, *S. tamarugo* and *S. strombulifera*, the genera *Strombocarpa* and *Neltuma*, and the tribe *Mimoseae* with DnaSP v. 5 software [39]. The complete plastid genome sequence of *N. alba* (OP672364), *N. chilensis* (OP672365), *S. strombulifera* (OP672366), *S. tamarugo* (MW582314), *N. glandulosa* (NC_026683), *N. juliflora* (MN104889), *P. farcta* (MZ073639), *P. cineraria* (MN104890) and *Leucaena trichandra* (NC_028733) as outgroup species were used in the phylogenetic analysis. Eighty protein-coding genes (PCG) sequences were aligned separately using MAFFT v7 [40] and any gaps in the alignment were trimmed using TrimAL v1.4 [41]. Afterwards, the sequences were concatenated with Mesquite 3.81 software [42]. The analyses of the plastid genomes´ 80 PCG sequences were conducted using the maximum likelihood (ML) method. The complete plastid genome sequence of the nine species were analyzed using the Bayesian inference (BI) methods. The best-fitting nucleotide substitution model of sequence evolution, model TVM + I + G, was determined based on the Akaike Information Criterion (AIC) using the MrModeltest v2.3 [43]. The ML analyses were performed using RAxML-HPC BlackBox v.8.1.24 [44] with 1,000 bootstrap replicates. The BI analysis was conducted using MrBayes v.3.2 [45] with the CIPRES Science Gateway v3.3 [46]. The Markov Chain Monte Carlo (MCMC) algorithm was calculated for 5,000,000 generations, and the sampling tree for every 1,000 generations. The first 25% of generations were discarded as burn-in. In the analysis, bootstrap support (BS) values were estimated in the ML, and the reliability of clades in the Bayesian analysis was evaluated by means of posterior probability (PP). The trees were visualized with FigTree [47].

Divergence time estimate

To estimate the divergence time of the species we used BEAST v2.6.0 [48] based on the complete plastid genome sequence of eight species of the *Prosopis*, *Neltuma* and *Strombocarpa* genera, and *Leucaena trichandra* as an outgroup species. All genome sequences were aligned with MAFFT, and then the file was imported to BEAUTi interface to generate a file for BEAST, after applying HKY + Γ substitution model, “Empirical” frequency, strict molecular clock model and “Yule” model speciation. Divergence times were estimated combining two calibration points. The TimeTree tool (http://www.timetree.org [49]) was used to fix the node age of *Leucaena trichandra* and the *Prosopis* genus, which is known to have diverged 35 Mya (33.2–40.3 Mya) [13,50–52]. We considered a second calibration as well, fixing the node age of *Mezquite* clade (*N. alba*, *N. chilensis*, *N. juliflora*, among others) at 3.65 Mya (3.31–3.99 Mya) after the study of [13]. The Markov Chain Monte Carlo (MCMC) was run for 6 million generations, sampling every 1,000 generations. We ran the program using the input file XML generated by BEAUTi in BEAST. Final log files were checked in Tracer 1.7.1 [53]. We used the TreeAnnotator program [48] with a 10% burn-in. Figtree software was used to visualize phylogenetic trees, using the extent of the 95% highest posterior density (HPD) intervals for each divergence time.

Results

The chloroplast genome lengths of *N. alba*, *N. chilensis* and *S. strombulifera* comprise 162,980 bp, 163,047 bp and 160,569 bp respectively and its structure contains a typical quadripartite structure with two inverted repeat regions (IR; 25,919 bp, 25,919 bp and 26,026 bp respectively) separated by a large single copy region (LSC; 92,300 bp, 92,356 bp and 89,569 bp respectively) and a small single copy region (SSC; 18,842 bp, 18,853 bp and 18,623 bp respectively) (Figure 1, Table 1). Its length and structure are similar to those of the other species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera, which vary in the IRs between 25,919 bp and 25,935 bp, in the LSC between 91,062 bp and 92,937 bp and in the SSC between 18,643 bp and 18,880 bp (Table 1). The cp genomes of the *Strombocarpa* species (*S. strombulifera* 160,569 bp; and *S. tamarugo* 161,575) are smaller (~2,000 bp) than those of the *Neltuma* species (*N. alba* 162,980; and *N. chilensis* 163,047) (Table 1). The GC content in the chloroplast of *N. alba* and *N. chilensis* (35.9%) is slightly less than in *S. strombulifera* (36.2%), but the overall GC content was similar to other species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera (Table 1).

In total, 128 genes were found in cp genomes of *N. alba*, *N. chilensis* and *S. strombulifera*, which included 83 coding genes, 8 rRNA genes, and 37 tRNA genes (Table 1). Of these genes, six coding genes (ndhB, *rpl2*, *rpl23*, *rps12*, *rps7* and *ycf2*), four rRNA genes and seven tRNA genes located in the IR regions contained duplicated genes (Table 2). Eighteen genes had introns, including nine PCGs (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1* and *rps16*), except for *S. strombulifera*, which had ten PCGs, additionally including *ycf2* to the ones mentioned before. Six tRNAs ($trnA^{UGC}$, $trnG^{GCC}$, $trnI^{GAU}$, $trnK^{UUU}$, $trnL^{UAA}$ and $trnV^{UAC}$) contained one intron, while three PCGs (*ycf3*, *rps12* and *clpP*) had two introns (Table 2).

We compared the simple sequence repeats (SSRs) from the cp genomes of the species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera. The maximum number was 100 SSRs in *N. juliflora*, 95 SSRs in *P. cineraria*, 92 SSRs in *N. glandulosa* and *S. strombulifera*, 90 SSRs in *N. chilensis*, 88 SSRs in *N. alba* and, whereas S. *tamarugo* and *P. farcta* only had 70 SSRs (Figure 2A). Mononucleotide repeats (A/T) were the most common repeats and the highest in number, ranging from 61 (*S. tamarugo*) to 80 (*S. strombulifera*); mononucleotide repeats C/G were found in all species except in *S. tamarugo* and *S. strombulifera* (Figure 2B). The number of dinucleotide SSRs (AT/AT) was similar in all species except in *S. tamarugo* and *S. strombulifera* in which no dinucleotide SSRs were found. In all species, 10-12 trinucleotide repeats AAT/ATT were found, but in *S. tamarugo* and *S. strombulifera* only 4 and 5 respectively, were present. Moreover, only *S. tamarugo* had trinucleotide repeats AAG/CTT (Figure 2B). In general, four pentanucleotide SSRs type (AAAAT/ATTTT, AAATT/AATTT, AATAT/ATATT and AATGG/ATTCC) were found in all species, but only one pentanucleotide SSR was present in *P. farcta* (AACTT/AAGTT), *S. strombulifera* and *S. tamarugo* (AATAG/ATTCT), and *S. strombulifera* (AATTC/AATTG) (Figure 2B). Only two hexanucleotide SSRs, AATATT/AATATT and AAATAG/ATTTCT, were observed in *S. strombulifera* and *P. farcta*, respectively (Figure 2B).

The amounts of repeats varied between 57 and 80 in the cp genomes of *S. tamarugo* (88), *P. farcta* (74), *N. juliflora* (73), *P. cineraria* (72), *N. glandulosa* (67), *N. alba* (60), *N. chilensis* (60) and *S. strombulifera* (57) (Figure 3A). The number of complement repeat in *S. tamarugo* was higher (5 repeats) than in the rest of the species (Figure 3A). The total number of palindromic repeats was less in *Strombocarpa* species (*S. strombulifera*, 22; and *S. tamarugo*, 24) than in species of the *Neltuma* (*N. alba*, 28; *N. chilensis*, 27; *N. glandulosa*, 29; and *N. juliflora*, 30) and *Prosopis* (*P. cineraria*, 33; *P. farcta*, 27) genera (Figure 3B). On the other hand, the total number of forward repeats was less in the cp genome of *N. alba* and *N. chilensis* (25 each), than in the rest of the species, where they varied between 30 to 40 (Figure 3C). Palindromic and forward repeats with lengths of 30-39 bp were the most common and abundant repeats in the species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera (Figure 3B-C). The number of reverse repeat (range of 30-39 bp) in *S. tamarugo* was higher (with 25 repeats) than in the rest of the species (Figure 3D).

The expansion and contraction of the IR and SC regions contributes to the differentiation in chloroplast genome size in some genera and families. For that reason, we compared of SSC, LSC, IRA, and IRB border regions of the species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera. In all species, *rps19* genes were located in the junction between LSC and IRb region (JLB), of which 176 to 188 bp were located at the LSC region and 91 to 103 bp located at the IRb region (Figure 4). In all species, the *rpl2* gene was entirely located in the IR regions (Figure 4). The *ndhF* gene of *Neltuma* and *Prosopis* species were located in the SSC region, 137-156 bp away from the IRb-SSC border, while in the two *Strombocarpa* species this gene was located approx. 67 bp away from the IRb-SSC border. At the SSC-IRA border, the *ycf1* gene extended into the SSC region, at varying lengths ranging from 4,760 bp in *S. tamarugo* to 4,794 bp in *N. juliflora* and *P. cineraria*, however, in *N. alba* and *S. strombulifera* the gene was entirely located in the SSC region, 963-973 bp away from the IRA-SSC border. In general, the truncated copy of *ycf1* was located in the IRB region (except in *N. alba* and *S. strombulifera*), while one end extended into the SSC region for 17 bp only in *N. juliflora* and *P. cineraria*. The distance between *rps19* and the LSC/IRA border was only 2 bp in *N. glandulosa* and *N. juliflora*. In most species, the *trnH* gene was located in the LSC region, 2-16 bp away from the IRA-LSC border, but in *N. juliflora* and *P. cineraria* it was much more distant. In general, the structure of the cp genomes of *Neltuma*, *Strombocarpa* and *Prosopis* were similar in arrangement (Figure 4).

In the genus *Neltuma*, the mean estimated value for nucleotide diversity (Pi) between *N. alba* and *N. chilensis* was Pi= 0,00037 (ranging from 0 to 0.0416); and the regions with the highest values of Pi were *rps16-trnQ*^{UUG} (Pi = 0.005), *accD-psaI* (Pi = 0.0416) and *ycf2- trnI*^{CAU} (Pi = 0.005) (Figure 5A). In the genus *Strombocarpa*, the mean estimated Pi value between *S. strombulifera* and *S. tamarugo* was 0,00522 (ranging from 0 to 0.0433) (Figure 5A); and the regions with the highest values of Pi were *matK-rps16* (Pi = 0.0433), *trnK-psbI* (Pi = 0.0366), *trnS*^{GCU}*-trnG*^{GCC} (Pi = 0.0233), *petN-psbM* (Pi = 0.0216), *psaB-psaA* (Pi = 0.0216), *rbcL-accD* (Pi = 0.0350), *psbE-petG* (Pi = 0.0383), *rpoA-rpl36* (Pi = 0.0250), *rps7-ndhB* (Pi = 0.0200) and *ycf2* (Pi = 0.0416) (Figure 5B). Among *Neltuma* and *Strombocarpa* species, the mean estimated Pi value was 0,00759 (ranging from 0 to 0.0627); and the regions with the highest values of Pi were *matK-rps16* (Pi = 0.0563), *trnQ*^{UUG}*-trnS*^{GCU} (Pi = 0.0472), *trnS*^{GCU}*-atpA* (Pi = 0.0627), *petN-trnD*^{GUC} (Pi = 0.0288), *psaB-psaA* (Pi = 0.0444), *atpB-rbcL* (Pi = 0.0541), *psbL-petL* (Pi = 0.0361), *rps18-rps12* (Pi = 0.0341), *ndhE-rrn23S* (Pi = 0.0277), *ndhB* (Pi = 0.0255) and *ycf2* (Pi = 0.0347) (Figure 5C). For all *Mimoseae* species used in this study, the mean estimated Pi value was 0,01236 (ranging from 0 to 0.1171); and ~31 regions with Pi values greater than 0.02 were observed (Figure 5D).

The results of the ML and IB trees had similar topologies when we compared 80 protein-coding genes of the plastid genomes of the nine *Mimosacea* species (Figure 6). The ML phylogenetic analysis revealed four clades, one joined *S. tamarugo* and *S. strombulifera* (BP=100), the second clade contained *N. glandulosa*, *N. juliflora*, *P. cineraria*, *N. chilensis* and *N. alba* (BP=100), the third clade was formed by *P. farcta*, and the fourth clade was formed by the outgroup *L. trichandra* (Figure 6A). The BI phylogenetic analysis with the nine *Mimosaceae* species showed high support (PP=1.00) and their topology was identical to ML analysis (Figure 6B). The second clade showed three subclades: one containing *N. glandulosa* which was separated from the rest of species with high support (BP=100; PP=1.00), while the second contained *N. juliflora* and *P. cineraria* (ML, BP=100; BI, PP=1.00), and the third contained *N. chilensis* and *N. alba* with high support (BP=100; PP=1.00) (Figure 6).

Divergence time for the *Neltuma*, *Strombocarpa* and *Prosopis* species based on the sequence of the chloroplast genomes is shown in Figure 7, and suggests that *Netuma*, *Strombocarpa* and *Prosopis* species shared a common ancestor around 43.11 Mya (95% highest posterior density (HPD): 37.72-48.07 Mya) in the Eocene. The age estimate for the split between *Prosopis* in the Old World and the new world species was 33.52 Mya (95% HPD: 29.48-37.70 Mya) in the early Oligocene. *Strombocarpa* and *Neltuma* genera diverged in the New World around 22.32 Mya (95% HPD: 19.55-25.08 Mya) in the early Miocene (Figure 7). Within of the genus *Strombocarpa*, *S. strombulifera* and *S. tamarugo* diverged around 8.70 Mya (95% HPD: 7.52-9.89 Mya) in the late Miocene (Figure 7). While the species of the genus *Neltuma* diverged into two clades much later, around 2.96 Mya (95% HPD: 2.62-3.33 Mya) in the Pliocene (Figure 7). One clade containing *N. alba*, *N. chilensis* and *N. glandulosa* diverged 1.12 Mya (95% HPD: 0.85-1.39 Mya) when a subclade formed between *N. alba* and

N. chilensis that diverged 0.66 Mya (95% HPD: 0.49-0.84 Mya). The other clade containing *P. cineraria* and *N. juliflora* diverged 1.85 Mya (95% HPD: 1.56-2.12 Mya) (Figure 7).

Table 1. General features of the *Neltuma*, *Strombocarpa* and *Prosopis* plastid genomes.

Species	Accession	Size (bp)	GC (%)	LSC (bp)	SSC (bp)	IR (bp)	N° Genes	Protein-coding genes	tRNA genes	rRNA genes
N. alba	OP672364	162,980	35.9	92,300	18,842	25,919	128	83	37	8
N. chilensis	OP672365	163,047	35.9	92,356	18,853	25,919	128	83	37	8
N. glandulosa	NC_026683	163,040	35.9	92,322	18,880	25,919	128	83	37	8
N. juliflora *	MN104889	163,237	35.9	92,495	18,880	25,931	132	85	39	8
S. strombulifera	OP672366	160,569	36.2	89,569	18,623	26,026	128	83	37	8
S. tamarugo	MW582314	161,575	36.0	91,062	18,643	25,935	127	82	37	8
*P. cineraria**	MN104890	163,677	35.9	92,937	18,878	25,931	131	85	38	8
P. farcta	MZ073639	162,900	35.9	92,156	18,880	25,932	127	82	37	8

(*) values of features of the species are described by Asaf et al (2020)

Table 2. Gene composition of the plastid genome of *N. alba, N. chilensis and S. strombulifera*.

Category of genes	Group of genes	Names of genes	N°
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*	5
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*	15
	ATP synthase	*atpA, atpB, atpE, atpF*[b], *atpH, atpI*	6
	NADH-dehydrogenase	*ndhA*[b], *ndhB*[ab], *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12
	cytochrome b/f complex	*petA, petB*[b], *petD*[b], *petG, petL, petN*	6
	Large subunit RUBISCO	*rbcL*	1
Protein synthesis and DNA replication	Transfer RNAs	*trnA-UGC*[ab], *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-UCC, trnG-GCC*[b], *trnH-GUG, trnI-GAU*[ab], *trnI-CAU*[a], *trnK-UUU*[b], *trnL-UAA*[b], *trnL-CAA*[a], *trnL-UAG, trnM-CAU, trnN-GUU*[a], *trnP-UGG, trnQ-UUG, trnR-ACG*[a], *trnR-UCU, trnS-GGA, trnS-UGA, trnS-GCU, trnT-GGU, trnT-UGU, trnV-UAC*[b], *trnV-GAC*[a], *trnW-CCA, trnY-GUA*	37
	Ribosomal RNAs	*rrn16S*[a], *rrn23S*[a], *rrn4.5S*[a], *rrn5S*[a]	8
	Ribosomal Protein large-subunit	*rpl14, rpl2*[ab], *rpl16*[b], *rpl20, rpl23*[a], *rpl32, rpl33, rpl36*	10
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*[b], *rpoC2*	4
	Ribosomal Protein Small-subunit	*rps11, rps12*[ac], *rps14, rps15, rps16*[b], *rps18, rps19, rps2, rps3, rps4, rps7*[a], *rps8*	14
Other functions	Subunit of Acetyl-CoA-carboxylase	*accD*	1
	c-type cytochrome synthesis gene	*ccsA*	1
	Envelop membrane protein	*cemA*	1
	Protease	*clpP*[c]	1
	Maturase	*matK*	1
	Initiation Factor	*infA*	0
Unknown function	Conserved open reading frames	*ycf1, ycf2*[a], *ycf3*[c], *ycf4*	5

Total	128

[a] Duplicated genes; [b] Genes containing introns; [c] Genes containing two introns

Discussion

Genomic research with NGS technology has developed rapidly, allowing efficient sequencing of complete plastid genomes [54]. Molecular differences in the complete chloroplast genome between species and individuals provide a good means of comparison [55]. The cp genome offers several advantages over the nuclear genome, such as unique haploid structure, structural conservation, maternal inheritance, and moderate rate of evolution [55,56]. In our comparative study of the plastid genomes of *N. alba*, *N. chilensis*, and *S. strombulifera* analyzing gene content, structure, divergence time, and phylogeny we found that the complete chloroplast genomes of *N. alba* and *N. chilensis* are conserved in size compared to species of the *Strombocarpa* genus. The chloroplast of *N. alba* and *N. chilensis* showed similar values for genome size and the number of genes compared to *Neltuma juliflora* and *Neltuma glandulosa* described by Asaf[57], ~ 163.000 bp for both, while the number of genes varied between 131 and 128. The number of genes was similar between the *Neltuma* and *Strombocarpa* genera, although *S. tamarugo* lost the gene *psbL* (remaining with 127 genes only) [25]. The absence of the *psbL* gene has been observed in some other eudicots, magnoliids, and monocots as well [58]. The genome sizes of *S. strombulifera* (160,569 bp) and *S. tamarugo* (161,575 bp) were smaller compared to the *Neltuma* species (~ 163.000 bp) [25]. However, the *Strombocarpa* species presented slightly more GC content (36.0%-36.2%) compared to the *Neltuma* species (35.9%). These GC values fall within the limit of variation registered in others studies [25,57]. Furthermore, a study about several orchid species, showed that the species with a smallest chloroplast size (*Pholidota cantonensis*, 158,786 bp), had a highest GC content (37.47%) [59], similar to our observations. The chloroplast genome tends to reduce its size during evolution [60], and gene length might be affected by selection during the evolution of spermatophytes [61]. The variations in chloroplast genome size among closely related species can be attributed to IRs, LSC, SSC, intergenic regions, and gene numbers [61]. In this study, very little variation in IRs and intergenic regions was observed between *N. alba* and *N. chilensis*, resulting in very few differences in genome size, while there is a large variation in these regions in the genomes of *S. strombulifera* and *S. tamarugo*. Therefore, we assume that *Strombocarpa* species have been exposed to stronger evolution than *Neltuma* species.

A total of 70 to 100 chloroplast simple sequence repeats (cpSSRs) were founded in the cp genomes of the species of the *Neltuma*, *Strombocarpa* and *Prosopis* genera. Our results showed high variation values in the number of cpSSRs among *Neltuma* and *Strombocarpa* species, being the highest for *N. juliflora* (100) and the lowest for *S. tamarugo* (70). The most abundant cpSSR motif types in *Neltuma*, *Strombocarpa* and *Prosopis* were mono-nucleotides, which is the most abundant repeat type in angiosperms cp genomes [62]. Only *Strombocarpa* species did not show mononucleotide C/G motifs, nor dinucleotide motifs and additionally, they had a lower number of trinucleotide AAT/ATT motifs. However, the *Strombocarpa* species were the only species that presented the pentanucleotide AATAG/ATTCT motifs. It has been shown in *Cyatheaceae*, that the characteristics of cpSSRs can provide useful phylogenetic information at the genus level, such as phylogenetic relationships, but also about the number, relative abundance, motif type and relative density of cpSSRs [63]. In a similar way, our results demonstrate that the cpSSRs among *Neltuma* and *Strombocarpa*, both in number and cpSSR motifs, are likely genus specific.

Repeat sequences are considered to play an important role in genome recombination, rearrangements and contain fundamental phylogenetic information [64,65]. We found differences in the repeated elements of the cp genome between *Neltuma* and *Strombocarpa* species. The highest total number of repeat elements (palindrome, forward, reverse and complement) was found in *S. tamarugo* (88) and the lowest in *S. strombulifera* (57). In general, the total number of palindromic repeats was less in *Strombocarpa* species than in *Neltuma* species. However, the total number of forward

repeats was less in *N. alba* and *N. chilensis* than in the *Strombocarpa* species. On the other hand, the number of complement and reverse (range of 30–39 bp) repeats in *S. tamarugo* was higher than in the *Neltuma* species. In the majority of the species in this study, the most abundant repeat elements detected were, in order, forward, palindromic and reverse. This corresponds to other studies about cp genomes of mimosoid species [66,67], although *S. tamarugo* is an exception in terms of reverse and complements repeats numbers.

Throughout of the evolution of plastid genomes, structural rearrangements occur, for example in the IRs, which are frequently subject to expansion, contraction or even complete loss [68]. An increased length of IR-SSC boundaries plays an important role in mimosoid plastome size variation [69]. For example, eight mimosoid plastomes of the tribe *Acacia* and *Inga* exhibited an unusual 13 kb IR-SSC boundary shift into the SSC region [67,69], and the size of these plastomes was found significantly affected by a IR-SC boundary shift, as well as by repeat content [67]. We observed a slight IR expansion into SSC in *S. strombulifera* (26.026 bp) and *S. tamarugo* (25.935 bp) in comparison to the *Netuma* species. Therefore, the SSC regions of the *Strombocarpa* species showed contraction, and were the shortest SSC regions compared with those of the *Neltuma* and *Prosopis* genera. Asaf et al [57] did not detect IR expansion in *Neltuma* and *Prosopis* species, however, they detected a slight expansion in the outgroup species of the genus *Adenanthera* (with a length of 26,028 bp), similar to what we found the in *Strombocarpa* species. The study of Asaf Asaf et al [57] did not, however, include *Strombocarpa* species to compare to the *Neltuma* and *Prosopis* species. Similar to Asaf [57], we found a partially duplicated *rps19* gene at the beginnings and ends of the IR regions in *N. alba*, *N. chilensis*, *S. strombulifera* and *S. tamarugo* (including 91 bp in IR). In of most *Mimosoideae* species, the *rps19* is located in the LSC/IRB junction (JLB), with 98–109 bp of the 5' end of this gene into the IR region [67]. The *ndhF* gene was located closer to the IRB-SSC border (JSB) in *Strombocarpa* species (up to 67 bp) than in *Neltuma* and *Prosopis* species (137 to 156 bp). Likewise, the *ndhF* gene in the species of the genera *Adenanthera*, *Parkia*, *Piptadenia*, *Leucaena* and *Dichrostachys* (*Mimosoideae*) was found entirely within the SSC region (ranging 11 to 150 away from the JSB junction), however, in species of the tribe *Acacia* and *Inga* (*Mimosoideae*) it was found within the JSB junction, resulting in the duplication of this gene [67]. Several models concerning the expansion and contraction of IR regions have been proposed to explain the possible mechanisms that result in shifts in the IR-LSC junctions [70]. In our case, we detected that *Strombocarpa* species had a larger contraction of the LSC region then *Neltuma* and *Prosopis* species. The structural differences presented among the plastomes of the *Neltuma* and *Strombocarpa* species reinforce the idea and necessity to disintegrate the *Prosopis* cluster, as proposed by Hughes et al [5]. However, for the new genera it would have been recommendable to have kept the names of the sections (*Algarobia* and *Strombocarpa*, as proposed by Burkart [4] for the new genera.

The nucleotide diversity (Pi) analysis of *Neltuma* and *Strombocarpa* plastomes showed more variations in the LSC and SSC regions than the IR regions. In addition, strong differences of nucleotide diversity value were found between *Neltuma* and *Strombocarpa* species. The Pi values between *Neltuma* species were so low that we found only three variable regions (*rps16-trnQ*^{UUG}, *accD-psaI* and *ycf2- trnI*^{CAU}), whereas in *Strombocarpa* species we found ten regions with high Pi values (*matK-rps16*, *trnK-psbI*, *trnS*^{GCU}*-trnG*^{GCC}, *petN-psbM*, *psaB-psaA*, *rbcL-accD*, *psbE-petG*, *rpoA-rpl36*, *rps7-ndhB* and *ycf2*). We believe that these ten highly variable regions found in *Strombocarpa* species, can be of use to resolve uncertainties in phylogenetic analysis of the genus, as well as for DNA barcoding. However, as a very low number of variable regions was found in the species of the *Neltuma* genus, it will be necessary for further studies to include a sufficient number of samples in order to identify the best regions for identification within the genus *Neltuma*.

The phylogenetic results (ML and BI) based on 80 protein-coding genes of the plastid genome of nine *Mimosoideae* species showed that *S. strombulifera* formed a strongly supported group with *S. tamarugo* (BP = 100; PP = 1.00), and the *Neltuma* group appeared as paraphyletic because *P. cineraria* was part of a well-supported clade (BP = 62; PP = 1.00) with *N. juliflora*, *N. alba* and *N. chilensis*. *P. farcta*, however appeared as sister group of *Neltuma* and *Strombocarpa* clade, as expected. Within the *Neltuma* clade, *N. alba* formed a highly supported clade with *N. chilensis* (BP = 100; PP = 1.00), and

so did *N. juliflora* with *P. cineraria* (BP = 100; PP = 1.00), whereas *N. glandulosa* appeared as a strongly supported sister group to both (BP = 100; PP = 1.00). With the exception of *P. cineraria* (further discussed in the next paragraph), the *Neltuma* group was monophyletic with *Strombocarpa* group as its sister clade. Although *S. strombulifera* and *S. tamarugo* formed a well-supported group, these two species showed important differences in genome size, number of genes and nucleotide diversity with high degree of variation. These genetic differences in the chloroplast correspond to the findings of Burkart [4] who separated *S. tamarugo* and *S. strombulifera* into the *Cavernicarpae* and *Strombocarpa*e series, respectively. The same was observed by Catalano et al [13] through a three-marker analysis (*trnS-psbC*, *G3pdh*, *NlA*), who found two well supported groups, one of them corresponding to the *Cavenicarpae* series (including *Prosopis ferox* and *P. tamarugo*) and the other formed by North American species of the *Strombocarpa*e series (including *Prosopis pubescens* and *Prosopis palmeri*).

Undoubtedly, the biggest inconsistency observed in our phylogenetic analysis was the nesting of *P. cineraria* within the *Neltuma* clade. According to the results of Asaf et al [57], *P. cineraria* forms a group with high support with *N. juliflora*. It is interesting and unexpected that *P. cineraria* did not form a group with *P. farcta*, both of them being Old World species, but nested with *N. juliflora*, *N. glandulosa*, *N. alba* and *N. chilensis*, which are New World species. However, according to the phylogenetic analysis performed by Catalano et al[13], there are more distant relationships among species from the Old World sections and closer relationships among species of the American sections (*Strombocarpa*, *Algarobia*, and *Monilicarpa* sections). *Prosopis cineraria* is one of the most common trees of the Indian desert, Arabian Peninsula and, in general, is abundant throughout the middle east [57,71], whereas *N. juliflora* is native to the Caribbean, Central and northern South America [72]. However, *Neltuma juliflora* was introduced to Ethiopia and the Middle East around 1970 and over the years this species has spread outside the plantation areas, adversely affecting natural habitats and rangelands [73]. This invasive plant is characterized by vigorous growth which helps it to outcompete indigenous plant species [74]. *Neltuma juliflora* seeds survive in livestock and warthogs' droppings, which serve as a vehicle for the plant to reach distant areas and to expand their distribution throughout the region [74,75]. We hypothesize that *N. juliflora* might have crossed with some individuals of *P. cineraria* in a natural way, giving offspring to a hybrid with a phenotype resembling *P. cineraria* but, when *N. juliflora* acted as the maternal part, with the cp genome of *N. juliflora*. This could be a logic explanation for the nesting of *P. cineraria* within the *Neltuma* clade, if the samples used by Asaf et al. [57] were obtained from a *P. cineraria* resembling hybrid.

Estimate of divergence time in plant groups have been important in order to understand their phylogeographic history and evolutionary biology [76]. Our molecular dating analysis suggests that *Leucaena trichandra* as root species diverged in the Middle Eocene (mean = 43.11 Mya; 95% HPD = 37.72–48.07 Mya). Later, *P. farcta* diverged in the Early Oligocene (mean = 33.52 Mya; 95% HPD = 29.48–37.70 Mya), while *P. cineraria* diverged together with the *Neltuma* species in the Pleistocene. Sudalaimuthuasari et al [71], using whole genome sequencing with 76,554 genes, estimated that *P. cineraria* and *P. alba* diverged ~ 23Mya. Undoubtedly, this divergence time is closer to what would be expected for species of the genus *Prosopis* that belongs to the Old World, but not for *P. cineraria*, whose complete chloroplast genome data show a divergence time of 1.85 Mya (95% HPD: 1.56–2.12 Mya), strengthening our suspicion of this sample being a natural hybrid. Although a previous study indicates that the divergence between *Strombocarpa* and *Neltuma* genera occurred in the Oligocene [13], our results show that these genera diverged in the early Miocene (mean = 22.32 Mya; 95% HPD = 19.55–25.08 Mya). The molecular divergence time found in *Neltuma* and *Strombocarpa* genera is close to the diversification of the major clades in the subfamily *Mimosoideae*, which occurred in the Late Miocene [13,50]. Our results showed that *Strombocarpa* diverged in the Late Miocene (mean = 8.70 Mya; 95% HPD = 7.52–9.89 Mya), which is supported by the fossil *Prosopisinoxylon anciborae*, a *Mimosoideae* species with a high similarity to genus *Prosopis* L. (currently re-delimitated), reported to have occurred during the Late Miocene in the Catamarca Province, Argentina [77]. Additionally, a similar divergence time, around 9.21 Mya (8.35–10.07), for the genus *Strombocarpa* was found Catalano et al [13]. Our results also showed that the *Neltuma* genus started diverging in the Pliocene (mean = 2.96 Mya; 95% HPD = 2.62–3.33

Mya) and continued in the Pleistocene. This corresponds to the Mesquite species (e.g. *N. alba*, *N. juliflora*, *N. glandulosa*, *N. chilensis*, *N. alpataco* and *N. nigra*) whose divergence time started in the Pliocene and continued in the Pleistocene, (mean = 3.65 Mya; 95% HPD = 3.31–3.99 Mya) [13].

Tree species such as *Neltuma* and *Strombocarpa* are subject to a number of ecological selective pressures due to the hostile conditions of the Atacama Desert. Chloroplast genes are involved in regulatory responses to various abiotic stresses, including heat, chilling, salinity, drought and radiation [78,79]. Therefore, the here presented chloroplast genomes of the *Neltuma* and *Strombocarpa* species can play an important role in understanding the plants adaptations to these hostile environments.

The chloroplast genome structure of legumes is particularly interesting, because it contains multiple rearrangements, expansions, contractions, and loss of genetic content, which are all very useful for phylogenetic studies [79]. Phylogenetic analysis can aid conservation of species through the confirmation of taxonomic status, clarification of evolutionary relationships and consequently the determination fo conservation priorities [80]. Additionally, phylogeographic studies offer valuable information for conservation purposes as they describe the geographical distribution of genetic variability among species populations [81]. With this study, we discovered differences in chloroplast genomes of *Neltuma* and *Strombocarpa*, species improving our understanding of its phylogeny and evolution, in hope to aid the conservation of these valuable species before it is too late and they disappear.

Conclusion

In this work, we present for the first time the assembly and characterization of the chloroplast genomes of *Neltuma alba*, *Neltuma chilensis* and *Strombocarpa strombulifera*. We found enough variation in genome size, GC content, repetitive elements and nucleotide diversity to support the disintegration of the former genus *Prosopis* L. The chloroplasts presented in this study provide a better understanding of the diversification of *Neltuma*, *Strombocarpa* and *Prosopis* as well as important information for evolutionary, phylogenomic and biogeographic studies for other species of the *Fabaceae* family.

Declarations

Acknowledgements

RC thanks projects ANID-FONDECYT initiation 11230668. RC and FC thanks AFOREST, a Millenium Nucleus supported by ANID – MILENIO – NCS2022_024. WH thanks "Convenio ESR UTA2295". RC thanks Neri Contreras-Ascencio and Ana Díaz-Sanchez for their sampling support.

Funding

This work was supported by the **ANID-FONDECYT Initiation** into Research grant 11230668 and by the Universidad de Atacama (DIUDA 22423 project). PJ was funded by the German Research Foundation (DFG) with the grant number JU 3228/1-1. LvdB was funded by ANID PIA/ACT 210038.

Authors' contributions.

R.C., L.vdB., F.C.: data analyses, writing manuscript. R.C., L.vdB., F.C., W.H., P.J.: data interpretation, writing and editing manuscript. R.C., F.C., L.vdB: experimental analysis. R.C., L.vdB., W.H.: Analysis and interpretation of result. R.C., L.vdB., P.J.: editing-review original draft. All the authors have approved the final manuscript.

Conflicts of interest/Competing interests.

There is no conflict of interest between the authors of this manuscript.

Ethics approval.

This article does not contain any studies with human participants or animals performed by any of the authors.

Research Permit.

This research complies with the corresponding research permits according to national and international standards, for the collection of material from Neltuma alba, Neltuma chilensis and Strombocarpa strombulifera, and the care of flora and fauna. The research permit was granted by CONAF (National Forestry Corporation) N° N00024/08-11-2019 (JBH/FAP/JVO) and N° N00003-2023/27-01-2023 (NOO/FAP/JVO).

Consent to participate.

All the authors of this manuscript declare that we participated in the design and preparation of this manuscript.

Consent for publication.

All authors authorize the publication of this manuscript.

Availability of data and material.

The datasets generated and analyzed during the current study are available in the Genome Database on National Center for Biotechnology Information (NCBI) repository under the accession number OP672364 for Neltuma alba, OP672365 for Neltuma chilensis and OP672366 for Strombocarpa strombulifera. The BioProject and BioSample accession numbers on NCBI for Neltuma alba are PRJNA1026123 and SAMN37734720, for Neltuma chilensis are PRJNA1026131 and SAMN37735133, and for Strombocarpa strombulifera are PRJNA1026137 and SAMN37735326. The identification of the plant material was carried out by Roberto Contreras-Díaz, according to the keys described by Burkart (1976). It was also confirmed by CONAF professionals and recognized by the researcher Nicolás García of the Universidad de Chile.

Code availability.

The sequence data is available in GenBank with the codes OP672364, OP672365, OP672366

References

1. LPWG (The Legume Phylogeny Working Group). A new subfamily classification of the leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66, 44–77 (2017).
2. Bhatia, H., Srivastava, G. & Mehrotra, R. C. Legumes from the Paleocene sediments of India and their ecological significance. Plant Divers. 45, 199–210 (2023).
3. Lewis, G., Schrire, B., Mackinder, B. & Lock, M. *Legumes of the World*. (Royal Botanic Gardens, Kew, 2005).
4. Burkart, A. A monograph of the genus Prosopis (Leguminosae subfam. Mimosoideae). J. Arnold Arbor. 450–525 (1976).
5. Hughes, C. E., Ringelberg, J. J., Lewis, G. P. & Catalano, S. A. Disintegration of the genus Prosopis L. (Leguminosae, Caesalpinioideae, mimosoid clade). *PhytoKeys* 205, 147–189 (2022).
6. Bessega, C. *et al.* Genetic studies of various Prosopis species (Leguminosae, Section Algarobia) co-occurring in oases of the Atacama Desert (northern Chile). Ecol. Evol. 11, 2375–2390 (2021).

7. Contreras, R., Carevic, F. S., Porcile, V. & Arias, M. Development of ssr loci in prosopis tamarugo phillipi and assessment of their transferability to species of the strombocarpa section. For. Syst. 29, 1–10 (2020).
8. Bull, A. T., Andrews, B. A., Dorador, C. & Goodfellow, M. Introducing the Atacama Desert. Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol. 111, 1269–1272 (2018).
9. Lehner, G., Delatorre, J., Lütz, C. & Cardemil, L. Field studies on the photosynthesis of two desert Chilean plants: Prosopis chilensis and Prosopis tamarugo. J. Photochem. Photobiol. B Biol. 64, 36–44 (2001).
10. Garrido, M. *et al.* The adjustment of Prosopis tamarugo hydraulic architecture traits has a homeostatic effect over its performance under descent of phreatic level in the Atacama Desert. Trees - Struct. Funct. 34, 89–99 (2020).
11. Contreras, R., van den Brink, L., Burgos, B., González, M. & Gacitúa, S. Genetic characterization of an endangered chilean endemic species, prosopis burkartii muñoz, reveals its hybrids parentage. Plants 9, 1–19 (2020).
12. Bentham, G. VII. Revision of the Suborder Mimoseae. Trans. Linn. Soc. London 30, 335–664 (1875).
13. Catalano, S. A., Vilardi, J. C., Tosto, D. & Saidman, B. O. Molecular phylogeny and diversification history of Prosopis (Fabaceae: Mimosoideae). Biol. J. Linn. Soc. 93, 621–640 (2008).
14. Ringelberg, J. J. *et al.* Phylogenomic analysis of 997 nuclear genes reveals the need for extensive generic re-delimitation in Caesalpinioideae (Leguminosae). *PhytoKeys* 205, 3–58 (2022).
15. Green, B. R. Chloroplast genomes of photosynthetic eukaryotes. Plant J. 66, 34–44 (2011).
16. Guo, S. *et al.* A Comparative Analysis of the Chloroplast Genomes of Four Polygonum Medicinal Plants. Front. Genet. 13, (2022).
17. Li, B. & Zheng, Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. Sci. Rep. 8, (2018).
18. Shang, C. *et al.* Chloroplast Genomic Resources and Genetic Divergence of Endangered Species Bretschneidera sinensis (Bretschneideraceae). Front. Ecol. Evol. 10, (2022).
19. Li, Q. *et al.* Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of Pulsatilla (Ranunculaceae). Sci. Rep. 10, (2020).
20. Li, X. *et al.* Plant DNA barcoding: from gene to genome. Biol. Rev. Camb. Philos. Soc. 90, 157–166 (2015).
21. Contreras-Díaz, R., Arias-Aburto, M. & van den Brink, L. Characterization of the complete chloroplast genome of Zephyranthes phycelloides (Amaryllidaceae, tribe Hippeastreae) from Atacama region of Chile. Saudi J. Biol. Sci. 29, 650–659 (2022).
22. Díaz, R. C., Mamani, W. H., Brink, L. van den, Fuentes, M. N. & Aburto, M. A. The complete chloroplast genome of the endangered species garra de león [Bomarea ovallei (Phil.) Ravenna] from Chile. Indian J. Genet. Plant Breed. 82, 365–368 (2022).
23. Chávez-Galarza, J. C., Cardenas-Ninasivincha, S., Contreras, R., Ferro-Mauricio, R. & Huanca-Mamani, W. Chloroplast genome of Tillandsia landbeckii Phil. (Bromeliaceae) a species adapted to the hyper-arid conditions of the Atacama and Peruvian desert. Mitochondrial DNA Part B Resour. 6, 3375–3377 (2021).
24. Chávez-Galarza, J. C., Cardenas-Ninasivincha, S., Contreras, R., Ferro-Mauricio, R. & Huanca-Mamani, W. Chloroplast genome of Tillandsia marconae till & Vitek (Bromeliaceae), a hyperarid desert endangered species. Mitochondrial DNA Part B Resour. 6, 2562–2564 (2021).
25. Contreras-Díaz, R., van den Brink, L., Navarrete-Fuentes, M. J. & Arias-Aburto, M. Characterization of the complete chloroplast genome of Prosopis tamarugo (Prosopis, Leguminosae), an endangered endemic tree species from the Atacama Desert. Bosque 42, 365–370 (2021).
26. Contreras-Díaz, R. *et al.* Chloroplast genome structure and phylogeny of Geoffroea decorticans, a native tree from Atacama Desert. Electron. J. Biotechnol. 60, 19–25 (2022).

27. Ferreyra, L. I., Vilardi, J. C., Tosto, D. S., Julio, N. B. & Saidman, B. O. Adaptive genetic diversity and population structure of the 'algarrobo' [Prosopis chilensis (Molina) Stuntz] analysed by RAPD and isozyme markers. Eur. J. For. Res. 129, 1011–1025 (2010).
28. Bessega, C. *et al.* Genetic diversity and differentiation among Prosopis alba (Leguminosae) populations from dry valleys of Bolivia with different levels of human disturbance and altitude. Ecol. Evol. 8, 11309–11321 (2018).
29. Díaz, R. C., González, M. G., Cruces, F. A., Arias, S. G. & Saavedra, V. P. Analysis of genetic diversity and differentiation of Prosopis chilensis populations between Chacabuco and San Felipe de Aconcagua provinces using microsatellite molecular markers. Bosque 42, 371–382 (2021).
30. Reginato, M., Llanes, A., Devinar, G., Garello, F. & Luna, M. V. Morphophysiology and Biochemistry of Prosopis strombulifera Under Salinity. Are Halophytes Tolerant to All Salts? 57–71 (2016) doi:10.1007/978-3-319-27093-7_4.
31. Carevic, F. S., Contreras, R., Carmona, E. & Barrientos, E. Identifying Plant Physiological and Climatic Drivers in the Woody Shrub Prosopis strombulifera: Effects of Spring Traits on Flower Sprouting and Fruit Production. Forests 14, 1167 (2023).
32. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477 (2012).
33. Qu, X. J., Moore, M. J., Li, D. Z. & Yi, T. S. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods 15, (2019).
34. Shi, L. *et al.* CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47, W65–W73 (2019).
35. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47, W59–W64 (2019).
36. Amiryousefi, A., Hyvönen, J. & Poczai, P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34, 3030–3031 (2018).
37. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. Bioinformatics 33, 2583–2585 (2017).
38. Kurtz, S. & Schleiermacher, C. REPuter: Fast computation of maximal repeats in complete genomes. Bioinformatics 15, 426–427 (1999).
39. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452 (2009).
40. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).
41. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).
42. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis. http://www.mesquiteproject.org (2023).
43. Nylander, J. A. A. *MrModeltest Version 2*. (Program distributed by the author. Evolutionary Biology Centre, Uppsala University, 2004).
44. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).
45. Ronquist, F. *et al.* Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542 (2012).
46. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew. Comput. Environ. Work. GCE *2010* (2010) doi:10.1109/GCE.2010.5676129.

47. Rambaut, A. FigTree: Tree Figure Drawing Tool Version 1.4. *University of Edinburgh, Edinburgh, Scotland.* http://tree.bio.ed.ac.uk/software/figtree (2012).
48. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, (2007).
49. Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. Mol. Biol. Evol. 39, (2022).
50. Lavin, M., Herendeen, P. S. & Wojciechowski, M. F. Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. Syst. Biol. 54, 575–594 (2005).
51. Zhao, K. *et al.* Comparative Analyses of Chloroplast Genomes From 14 Zanthoxylum Species: Identification of Variable DNA Markers and Phylogenetic Relationships Within the Genus. Front. Plant Sci. 11, (2021).
52. Särkinen, T., Pennington, R. T., Lavin, M., Simon, M. F. & Hughes, C. E. Evolutionary islands in the Andes: Persistence and isolation explain high endemism in Andean dry tropical forests. J. Biogeogr. 39, 884–900 (2012).
53. Rambaut, A., Suchard, M., Xie, D. & Drummond, A. Tracer V1.7.1. (2018).
54. Moore, M. J. *et al.* Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. 6, (2006).
55. Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R. & Li, D. Z. Complete chloroplast genome of the genus Cymbidium: Lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol. Biol. 13, (2013).
56. Goremykin, V. V., Hirsch-Ernst, K. I., Wölfl, S. & Hellwig, F. H. Analysis of the Amborella trichopoda chloroplast genome sequence suggests that Amborella is not a basal angiosperm. Mol. Biol. Evol. 20, 1499–1505 (2003).
57. Asaf, S., Khan, A. L., Khan, A. & Al-Harrasi, A. Unraveling the chloroplast genomes of two prosopis species to identify its genomic information, comparative analyses and phylogenetic relationship. Int. J. Mol. Sci. 21, (2020).
58. Mohanta, T. K. *et al.* Gene loss and evolution of the plastome. Genes (Basel). 11, 1–23 (2020).
59. Li, L. *et al.* Comparative analyses and phylogenetic relationships of thirteen Pholidota species (Orchidaceae) inferred from complete chloroplast genomes. BMC Plant Biol. 23, (2023).
60. Ren, Q. *et al.* The extremely large chloroplast genome of the green alga Haematococcus pluvialis: Genome structure, and comparative analysis. Algal Res. 56, (2021).
61. Xiao-Ming, Z. *et al.* Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. Sci. Rep. 7, (2017).
62. Castro, A. A. *et al.* Chloroplast genome characterization of Uncaria guianensis and Uncaria tomentosa and evolutive dynamics of the Cinchonoideae subfamily. Sci. Rep. 13, (2023).
63. Zhu, M. *et al.* Phylogenetic significance of the characteristics of simple sequence repeats at the genus level based on the complete chloroplast genome sequences of Cyatheaceae. Ecol. Evol. 11, 14327–14340 (2021).
64. Milligan, B. G., Hampton, J. N. & Palmer, J. D. Dispersed repeats and structural reorganization in subclover chloroplast DNA. Mol. Biol. Evol. 6, 355–368 (1989).
65. Sun, J. *et al.* Evolutionary and phylogenetic aspects of the chloroplast genome of Chaenomeles species. Sci. Rep. 10, (2020).
66. Souza, U. J. B. de, Nunes, R., Targueta, C. P., Diniz-Filho, J. A. F. & Telles, M. P. de C. The complete chloroplast genome of Stryphnodendron adstringens (Leguminosae - Caesalpinioideae): comparative analysis with related Mimosoid species. *Sci. Rep.* 9, (2019).
67. Wang, Y. H., Qu, X. J., Chen, S. Y., Li, D. Z. & Yi, T. S. Plastomes of Mimosoideae: structural and size variation, sequence divergence, and phylogenetic implication. Tree Genet. Genomes 13, (2017).
68. Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F. & Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297 (2011).
69. Dugas, D. V. *et al.* Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. Sci. Rep. 5, (2015).

70. Wang, R. J. *et al.* Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8, (2008).
71. Sudalaimuthuasari, N. *et al.* The Genome of the Mimosoid Legume Prosopis cineraria, a Desert Tree. Int. J. Mol. Sci. 23, (2022).
72. Pasiecznik, N. M. Prosopis: Pest or Providence, Weed or Wonder Tree? Eur. Trop. For. Res. Netw. Newsl. 28, 12–14 (1999).
73. Ali, A. S., Tudsri, S., Rungmekarat, S. & Kaewtrakulpong, K. Effect of feeding Prosopis juliflora pods and leaves on performance and carcass characteristics of Afar sheep. Kasetsart J. - Nat. Sci. 46, 871–881 (2012).
74. Mehari, Z. H. The invasion of Prosopis juliflora and Afar pastoral livelihoods in the Middle Awash area of Ethiopia. Ecol. Process. 4, 1–9 (2015).
75. Shiferaw, H., Teketay, D., Nemomissa, S. & Assefa, F. Some biological characteristics that foster the invasion of Prosopis juliflora (Sw.) DC. at Middle Awash Rift Valley Area, north-eastern Ethiopia. J. Arid Environ. 58, 135–154 (2004).
76. Premoli, A. C., Mathiasen, P., Cristina Acosta, M. & Ramos, V. A. Phylogeographically concordant chloroplast DNA divergence in sympatric Nothofagus s.s. How deep can it be? New Phytol. 193, 261–275 (2012).
77. Martínez, L. C. A. Prosopisinoxylon anciborae nov. gen. et sp. (Leguminosae, Mimosoideae) from the Late Miocene Chiquimil Formation (Santa María Group), Catamarca, Argentina. *Rev. Palaeobot. Palynol.* 158, 262–271 (2010).
78. Song, Y., Feng, L., Alyafei, M. A. M., Jaleel, A. & Ren, M. Function of chloroplasts in plant stress responses. Int. J. Mol. Sci. 22, (2021).
79. Daniell, H., Lin, C. S., Yu, M. & Chang, W. J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol. 17, (2016).
80. Byrne, M. Phylogenetics and the conservation of a diverse and ancient flora. Comptes Rendus - Biol. 326, (2003).
81. Bobo-Pinilla, J., Salmerón-Sánchez, E., Mendoza-Fernández, A. J., Mota, J. F. & Peñas, J. Conservation and Phylogeography of Plants: From the Mediterranean to the Rest of the World. *Diversity* 14, (2022).

Figures

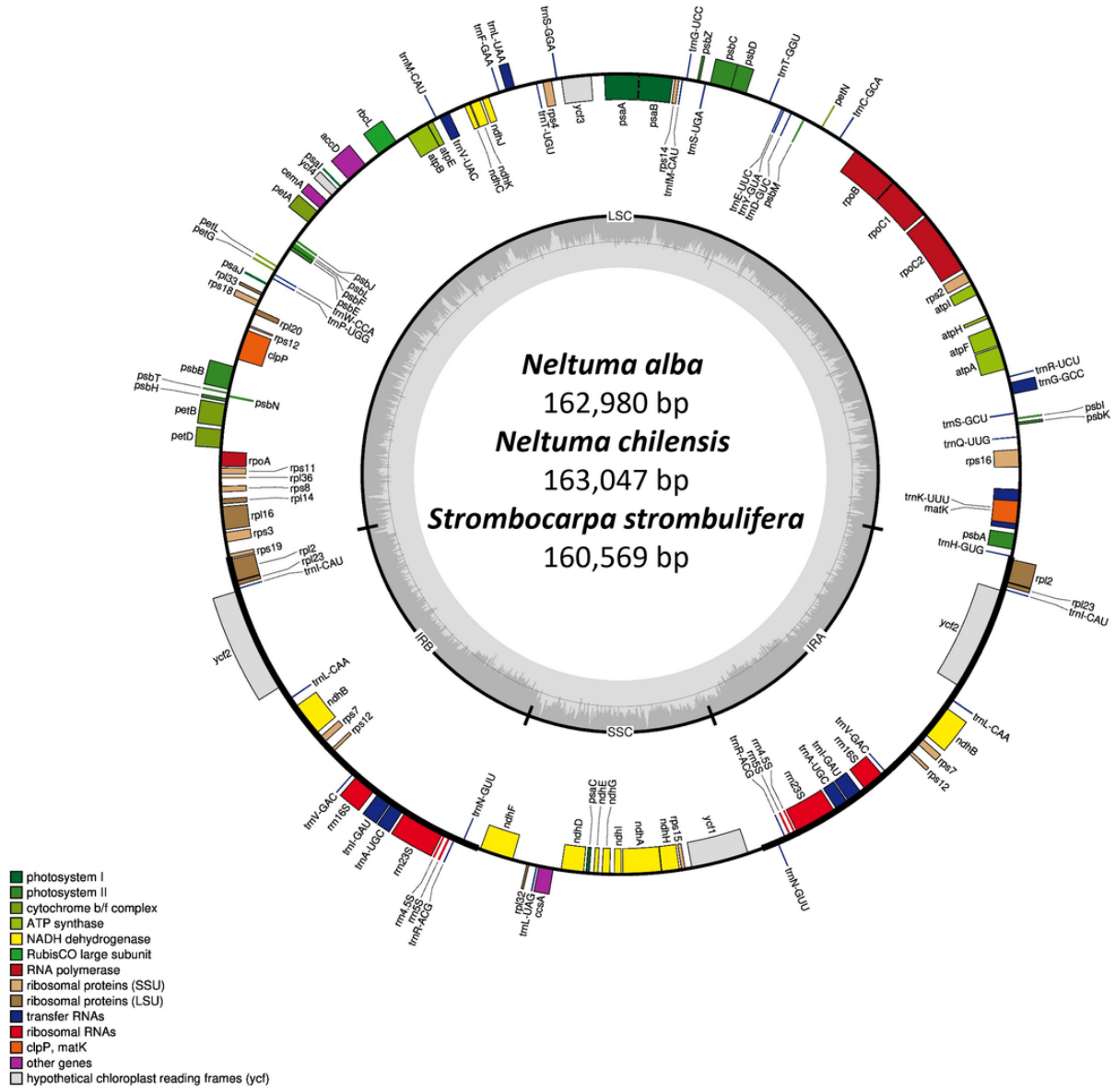

Figure 1

Circular gene map of the chloroplast genomes of *Neltuma alba*, *Neltuma chilensis* and *Strombocarpa strombulifera*. Genes were colored according to their functional group. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRA, IRB) were indicated.

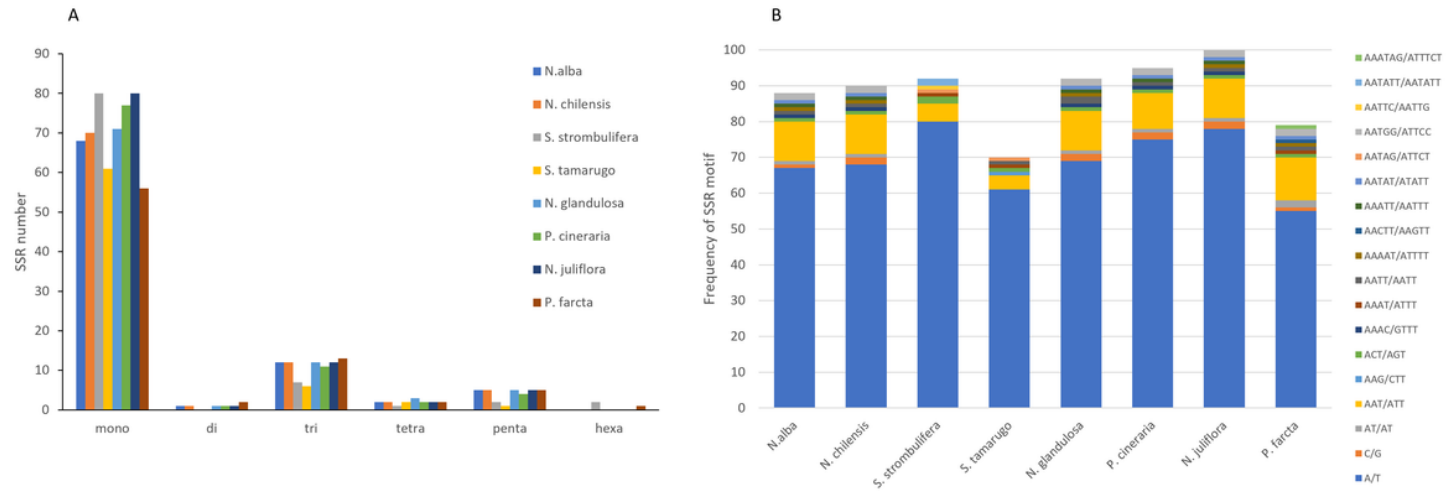

Figure 2

Analysis of simple sequence repeats (SSRs) of the *N. alba*, *N. chilensis*, *S. strombulifera*, *S. tamarugo*, *N. glandulosa*, *N. juliflora*, *P. cineraria*and *P. farcta* chloroplast genomes. Total numbers of SSRs of each motif unit (A) and frequency of SSR motifs in different repeat class types (B).

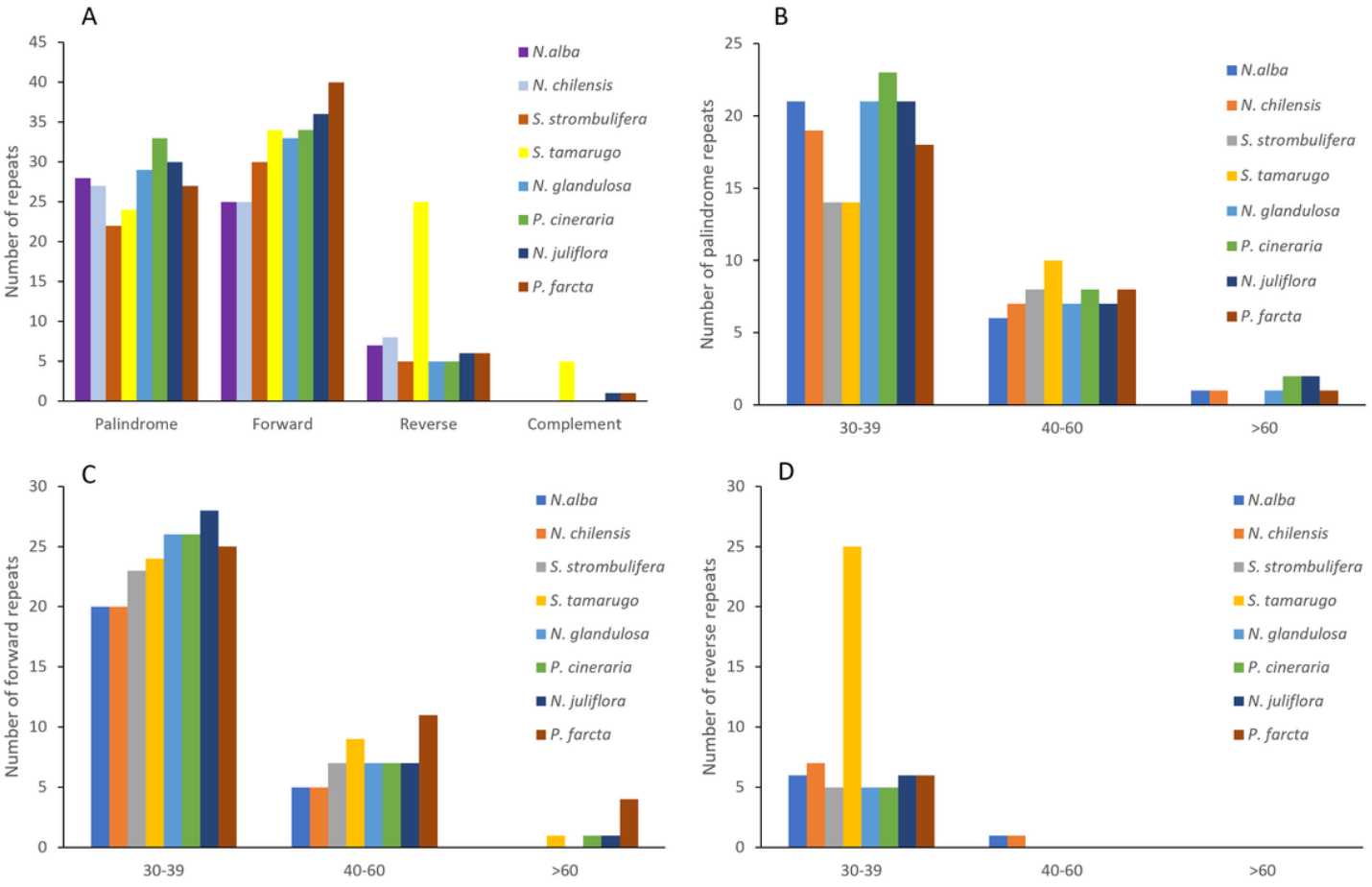

Figure 3

Repeat structure analysis of the *N. alba*, *N. chilensis*, *S. strombulifera*, *S. tamarugo*, *N. glandulosa*, *N. juliflora*, *P. cineraria* and *P. farcta*chloroplast genomes. Total numbers long repeat types: Palindrome, Forward, Reverse and Complement (A),

number of palindrome repeats (B), number of forward repeats (C) and number of reverse repeats (D) by length.

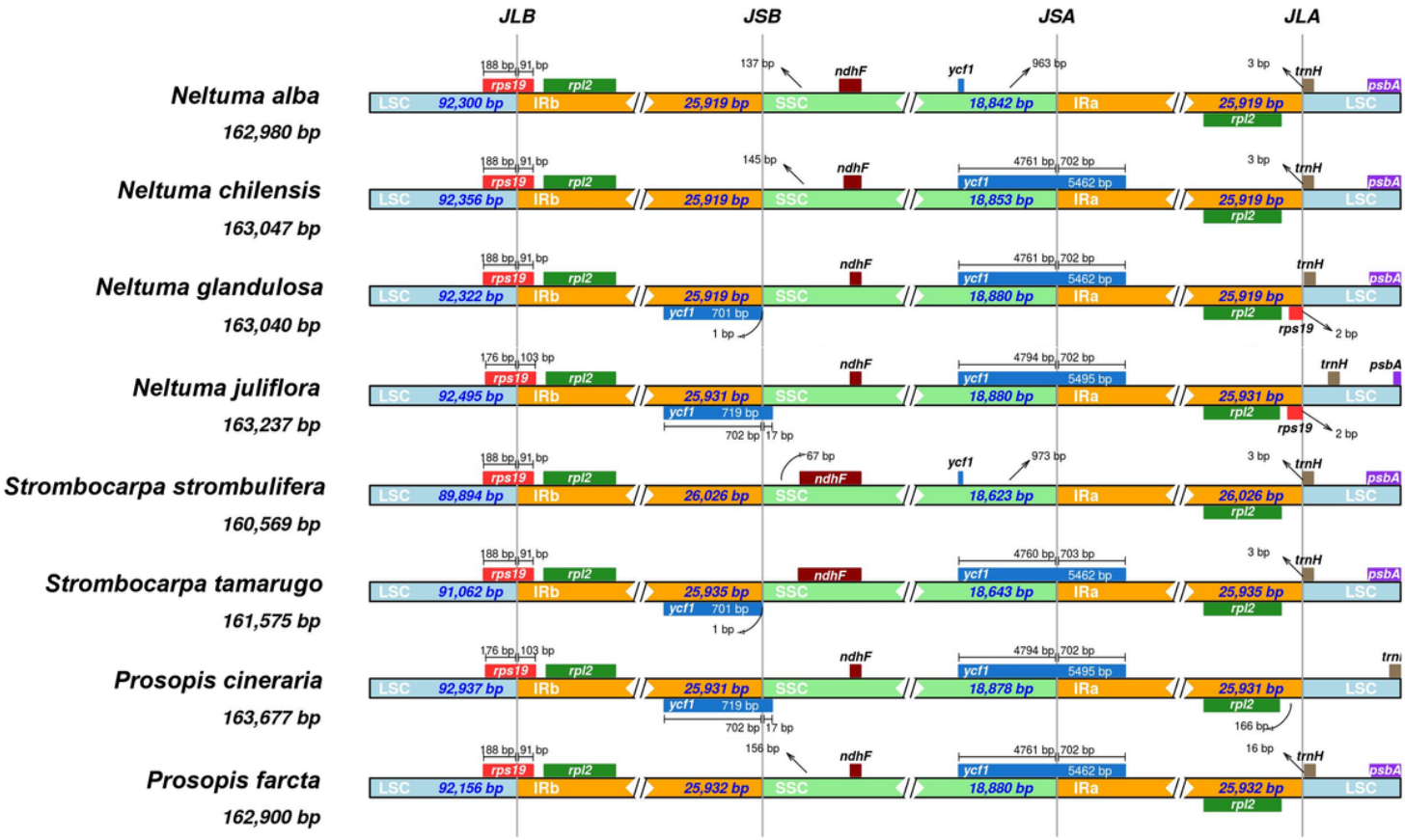

Figure 4

Comparison of chloroplast genomes between the Long Single Copy (LSC), Short Single Copy (SSC) and Inverted Repeat (IRa and IRb) junction regions among *Neltuma*, *Strombocarpa* and *Prosopis*species.

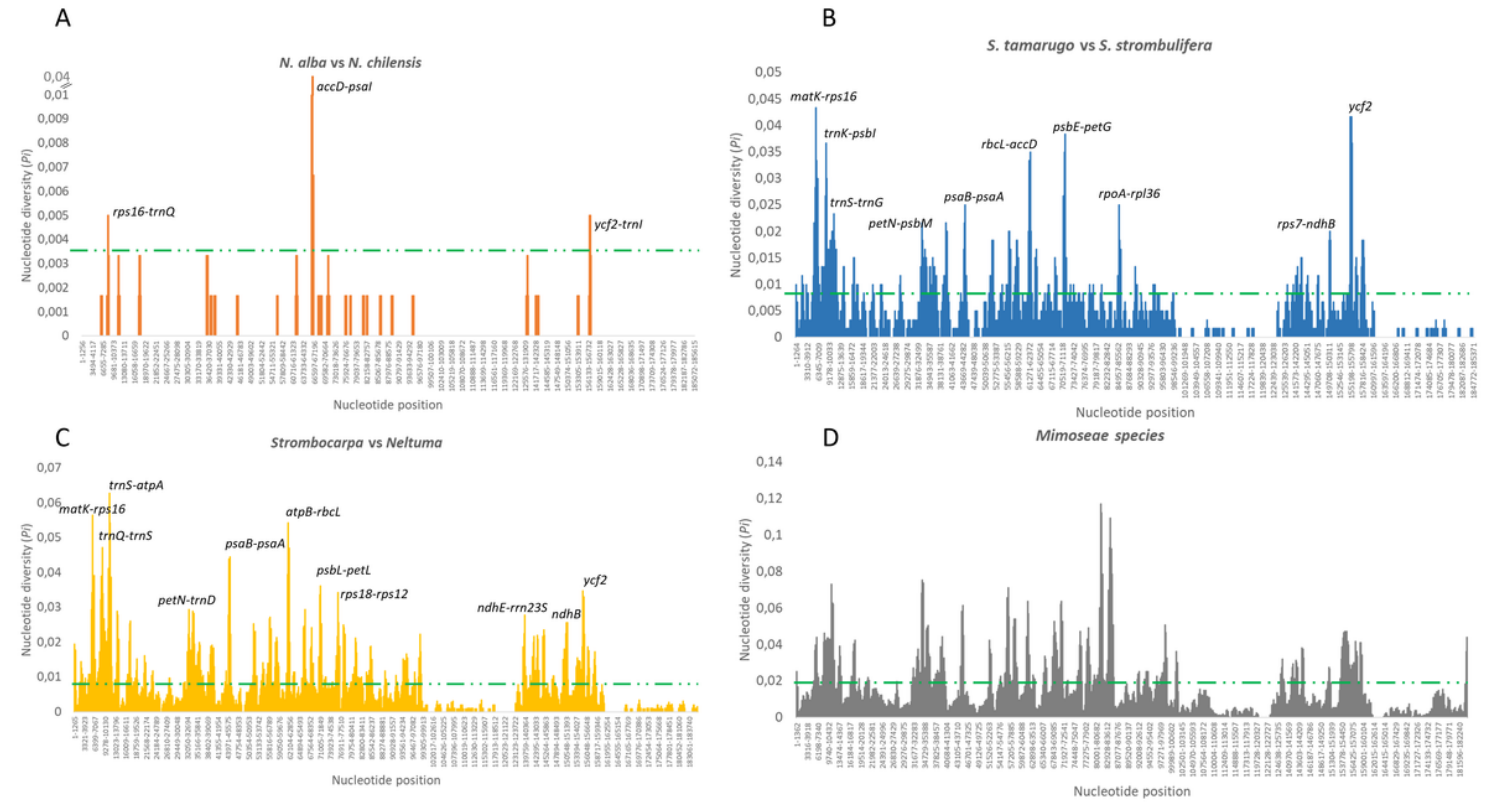

Figure 5

Sliding window analysis of the whole plastid of the two *Neltuma*species (a), of the two *Strombocarpa*species (b), of the genera *Strombocarpa*and *Neltuma* (c), and of the tribe *Mimoseae* (d). X-axis: Nucleotide position, Y-axis: Nucleotide diversity (Pi). The regions with highest Pi values were plotted considering 2× the median value as the cutoff point indicated by the green dashed line.

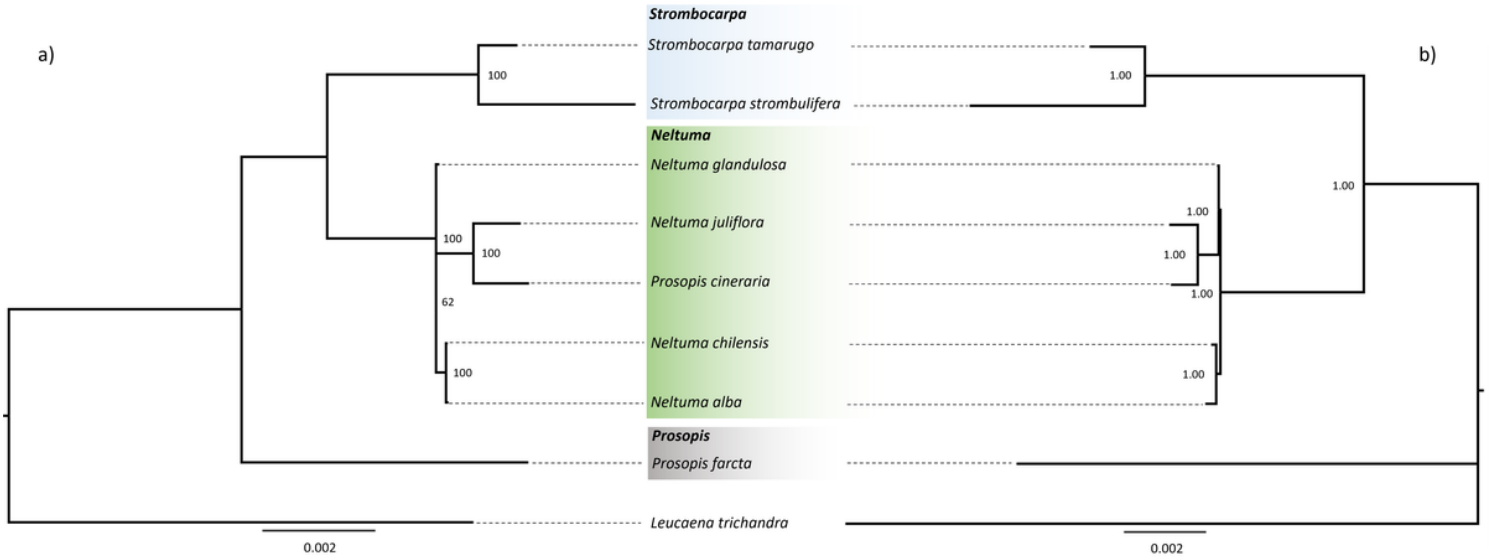

Figure 6

Molecular phylogenetic analysis based on 80 protein-coding genes of the plastid genome of nine *Mimoseae* species inferred by maximum likelihood (a) and Bayesian inference (b) methods. Numbers in the nodes are ML bootstrap values

(a) and Bayesian posterior probabilities values (b).

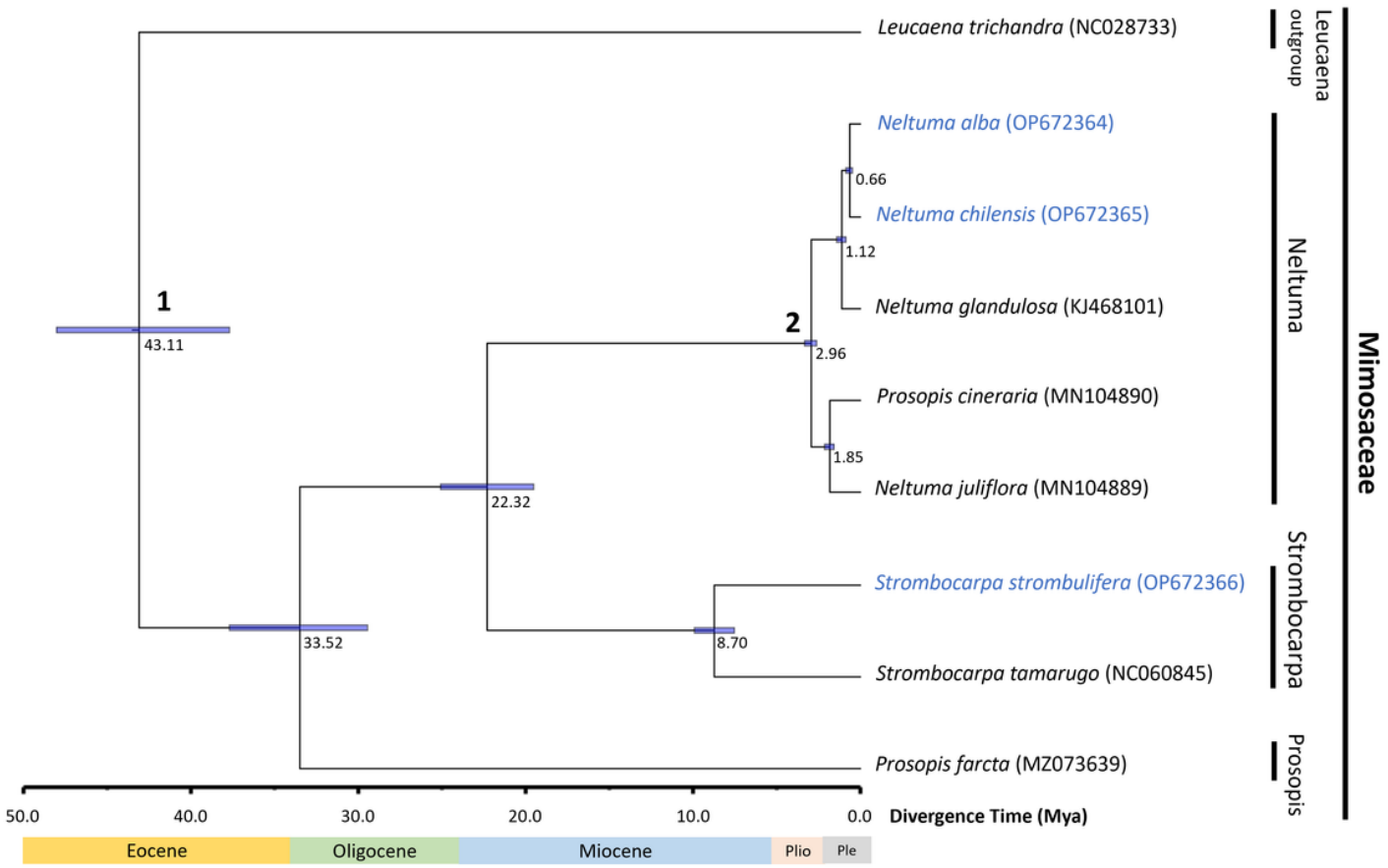

Figure 7

Neltuma, Strombocarpa and Prosopis chronogram showing divergence times estimated using BEAST program based on data from nine whole plastid genomes. The divergence times of each clade are displayed near each node. Blue bars represent 95% highest posterior density values for the estimated mean dates. The nodes 1 and 2 correspond to calibration points.