Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Assembly, annotation and analysis of the chloroplast genome of the Algarrobo tree Neltuma pallida (Subfamily: Caesalpinioideae)

Esteban Caycho
Universidad Nacional Mayor de San Marcos
Renato La Torre
Universidad Nacional Mayor de San Marcos
Gisella Orjeda (morjedaf@unmsm.edu.pe)
Universidad Nacional Mayor de San Marcos

Research Article

Keywords: Neltuma pallida, chloroplast genome, gene annotation, repetitive sequences, sequence comparison, boundary shift

Posted Date: July 20th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3173243/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on November 16th, 2023. See the published version at https://doi.org/10.1186/s12870-023-04581-5.

Abstract

Background

Neltuma pallida is a tree that grows in arid soils in northwestern Peru. Being a predominant species of the Equatorial Dry Forest ecoregion, it has a high economic and ecological importance for people and environment. Despite this, the species is severely threatened and there are few genetic and genomic studies about it, making it difficult to propose evidence-based conservation strategies.

Results

In this work, the chloroplast genome of a *N. pallida* specimen was assembled, annotated, analyzed and compared with those of related species. The length of the assembled chloroplast genome was 162381 bp with a typical quadripartite structure (LSC-IRA-SSC-IRB). The calculated GC% for the genome was 35.97%, although this is variable between regions, with a higher GC% found in the IRs. A total of 132 genes were annotated, of which 19 were duplicates and 18 contained at least one intron in their sequence. A large number of repetitive sequences of different types were identified in the assembled genome, being the most frequent the tandem repeats (> 300), especially microsatellites (SSR). The phylogenetic reconstruction of *Prosopis* s.l. showed that this group is monophyletic. *N. pallida* was closely related to *P. cineraria*, *N. juliflora* and *N. glandulosa*, forming a subclade with these species. When comparing the chloroplast genome sequence of *N. pallida* with *N. juliflora*, *P. farcta* and *Strombocarpa tamarugo*, it was found that the sequences were highly similar.

Conclusions

The chloroplast genome of *N. pallida* is composed of 162381 bp with the classical quadripartite structure and GC% of 35.97%, a value similar to that of other closely related species. Also, the genetic structure of the genome was found to be similar to other members of *Prosopis* s.l. In the chloroplast genome of *N. pallida*, more than 800 repetitive sequences were identified, being the most common the tandem repeats. *N. pallida* showed closeness to the other *Neltuma* spp., forming a subclade within the *Prosopis* s.l. group. The divergence analysis indicated that there were little differences when comparing the *Prosopis* s.l. chloroplast genomes, evidencing high levels of conservation between these species.

Background

The Algarrobo tree *Neltuma pallida* (Humb. & Bonpl. ex Willd.) Hughes & Lewis is a tree of 8 to 20 meters tall, with small grayish-green leaves, uninodales axillary spines and yellow pod-shaped fruits [1, 2]. This species is classified within the genus *Neltuma* (Subfamily: Caesalpinioideae), consisting of 43 species that are characterized by being arboreal or shrubby, have uninodales axillary spines and are distributed in dry tropical and arid regions of America. *Neltuma pallida*, specifically, is native to arid regions of Peru, Colombia and Ecuador, and is present in the desert areas of western South America in Peru, Chile and the arid regions of the Gran Chaco. In Peru, Algarrobo trees grow in the Equatorial Dry Forest (3.45% of the country's total area), an ecoregion presents in the northern coastal regions such as La Libertad, Lambayeque, Piura and Tumbes [3–6], being one of the predominant species there.

Algarrobo trees have high economic and ecological importance in South American countries, where they grow [7, 8]. The species is considered "multipurpose" due to the different ways in which it benefits local people [9] The most common uses are as fuel, medicine, cattle feed or construction material [10]. Similarly, the Algarrobo trees contribute in many ways to the maintenance of their environment [11]. They form a physical barrier to prevent wind erosion, contribute to soil fertility, maintain a microclimate and bioremediation [12, 13].

Despite the economic and ecological importance of *Neltuma* spp., their populations in Peru, Argentina, Chile, Venezuela, and the Chaco region (which includes some departments of Bolivia, Brazil, Argentina, and Paraguay) are in serious decline [14–18]. In the specific case of Peru, according to a report from the Peruvian National Forest and Wildlife Inventory [19], up to

40% of the *N. pallida* trees are standing dead and 9% are found as stumps. Of the 51% that are alive, 27% show mild to severe damage. While the real cause of the high mortality of the Algarrobo tree population is still uncertain and likely complex; some of the hypotheses contemplate abiotic factors such as climate change [20] and drought [21] in the northern coast of Peru. Others hypotheses include biotic factors such as the increasing presence of phytophagous and sucking plague insects [17, 22] thought today to be an effect of the decrease of natural biological controllers due to an ecological imbalance after recent ENSO events, or viruses from the Closteroviridae family [23, 24].

The literature shows few genetic studies on *N. pallida*, and genomic studies have not yet been carried out. In 2022, the National Institute of Agrarian Innovation (INIA) of Peru, together with the National University of Frontera (Piura - Peru), started a plant breeding project to improve the Algarrobo in Piura, a department in northern Peru [Arbizu, pers comm]. Also, a project on molecular delimitation of Neltuma species has recently begun in our laboratory [Rivas M. pers comm].

Some other works that have been published about the species focus on knowing its phenotypic variability, particularly its characteristics of agronomic importance [25–27]. On the other hand, there are some studies from more than a decade ago that have sought to characterize the species at the genetic level [6]. These use classic molecular markers [28, 29], study the ploidy of the species [30–32], or analyze one nucleotide sequence [33]. The lack of information about a species of such high importance, which is currently endangered, is detrimental to the development of comprehensive conservation and improvement strategies. Thus, the decline of Algarrobo tree populations continues to impact its highly fragile ecosystems such as the deserts and dry forests of many South American countries. With this, a decrease in endemism and biodiversity, and harsher lifestyles is expected to occur.

A first step in the development of genomic studies in plants is the assembly and annotation of the chloroplast genome. This represents a faster and simpler task than sequencing and analyzing the nuclear genome due to its size and level of complexity [34, 35]. Chloroplast genome sizes are in the range of 120 to 160 Kb, most often with 100 to 130 genes [36, 37]. In addition, genomic data is valuable for diversity studies [38, 39], phylogenetic analyses [40, 41], genetic improvement [42, 43] and genetic engineering of the species and closely related species [44, 45].

In this work, we discover the chloroplast genome content of the Algarrobo tree *N. pallida* (Subfamily: Caesalpinioideae) through its assembly, annotation and structural analysis. We also perform a comparative analysis using the sequenced chloroplast genomes of other species *Prosopis sensu lato* (s.l.), now the genera *Neltuma*, *Strombocarpa* and *Prosopis sensu stricto* (s.s.), and make a phylogenetic reconstruction in order to identify the relationships of *N. pallida* within the clade.

Results

Assembly and annotation of the chloroplast genome of Neltuma pallida:

We assembled the chloroplast genome and obtained a graphical file of the assembly and the genomic sequence. The graphical file was used to examine the structure and sequencing depth of the assembled genome. The assembled *N. pallida* chloroplast genome (Genbank: OR178743) had a length of 162381 bp and the classical quadripartite structure (Figure S1): a long single copy sequence (LSC) of 91805 bp (~ 56.54% of the genome), a short single copy sequence (SSC) of 18748 bp (~ 11.55% o) and two inverted repeat (IR) regions of 25914 bp (~ 31.91%) (Table 1).

Table 1
Summary of chloroplast genome features of *N. pallida* and related species.

Species	*Neltuma pallida*	*Neltuma juliflora*	*Prosopis farcta*	*Strombocarpa tamarugo*	*Acacia ligulata*
Size (bp)	162381	163237	162900	161575	174233
Total GC content (%)	35.97	35.9	35.88	36	35.4
LSC (bp)	91805	92495	92156	91062	88576
SSC (bp)	18748	18880	18880	18643	18298
IRs (bp)	25914	25931	25932	25935	25925
Protein-coding regions (bp)	78834	78421	-	-	-
tRNA coding regions (bp)	2938	2927	-	-	-
rRNA coding regions (bp)	9052	9052	-	-	-
Number of genes	132	132	127	127	133
Number of protein-coding genes (PCG)	85	85	82	82	88
Number of rRNA	8	8	8	8	8
Number of tRNA	39	39	37	37	37
Number of genes with introns	22	21	25	18	18

An analysis of the assembled genome structure was performed by examining the nucleotide composition of each region (Table S1). The GC content (GC%) of the total genome was 35.97%. LSC and SSC had lower GC content: 33.26% and 30.46%, respectively. The IRs presented a higher GC% than the rest of the regions and the total genome, with 42.77%. Among the coding regions, rRNA coding regions (which are located in the IRs) had the highest GC% with 55.41%, followed by tRNA coding regions with 53.10%. The lowest GC content was found in the protein coding regions with 37.45%.

The annotation of the assembled genome was done taking as reference the chloroplast genome of *N. juliflora*. A total of 132 genes were found within the genome (19 duplicated genes), being 85 protein-coding genes, 39 tRNA-coding genes and 8 rRNA-coding genes. The 85 protein-coding genes correspond to 78834 bp, the 39 tRNA-coding genes to 2938 bp and the 8 rRNA-coding genes to 9052 bp (Table 2).

Table 2
Genes annotated in the *N. pallida* chloroplast genome grouped by category and functional group.

Category	Functional group	Genes
Transcription and translation	Large subunit of ribosomal proteins	*rpl*2”*, 14, 16*, 20, 23”, 32, 33, 36
	Small subunit of ribosomal proteins	*rps*2, 3, 4, 7”, 8, 11, 12”*, 14, 15, 16*, 18, 19
	DNA dependent RNA polymerase	*rpo*A, B, C1*, C2
	Translation	*inf*A
	rRNA	*rrn*4.5”, *rrn*5”, *rrn*16”, *rrn*23”
	tRNA	*trn*A-UGC”*, *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC*, *trn*G-UCC, *trn*H-GUG, *trn*I-CAU”, *trn*I-GAU”*, *trn*K-UUU*, *trn*L-CAA”, *trn*L-UAA*, *trn*L-UAG, *trn*fM-CAU, *trn*M-CAU”, *trn*N-GUU”, *trn*P-GGG, *trn*P-UGG, *trn*Q-UUG, *trn*R-ACG”, *trn*R-UCU, *trn*S-GCU, *trn*S-GGA, *trn*S-UGA, *trn*T-GGU, *trn*T-UGU, *trn*V-GAC”, *trn*V-UAC*, *trn*W-CCA, *trn*Y-GUA
Photosynthesis	Photosystem I	*psa*A, B, C, I, J
	Photosystem II	*psb*A, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
	NADH dehydrogenase	*ndh*A*, B”*, C, D, E, F, G, H, I, J, K
	Cytochrome b6/f complex	*pet*A, B*, D*, G, L, N
	ATP synthase	*atp*A, B, E, F*, H, I
	Rubisco	*rbc*L
Other genes	Maturase	*mat*K
	Protease	*clp*P*
	Envelop membrane protein	*cem*A
	Subunit Acetyl-CoA-Carboxylate	*acc*D
	c-type cytochrome synthesis gene	*ccs*A
Unknown	Conserved Open reading frames	*ycf*1”, 2”, 3*, 4
” Duplicated genes, * Genes with introns		

The annotated genes included the following (Fig. 1): 12 small ribosomal proteins (*rps*), 8 large ribosomal proteins (*rpl*), 4 DNA-dependent RNA polymerases (*rpo*), 4 different rRNA-coding genes (*rrn*), 31 different tRNA-coding genes (*trn*), 5

photosystem I proteins (*psa*), 15 photosystem II proteins (*psb*), 11 NADH dehydrogenase proteins (*ndh*), 6 cytochrome b6/f complex proteins (*pet*), 6 ATP synthase complex proteins (*atp*), the major subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L), maturase K (*mat*K), the proteolytic subunit of ATP-dependent Clp protease (*clp*P), membrane envelope protein (*cem*A), the beta subunit of acetyl-CoA carboxylase (*acc*D), the cytochrome C biogenesis protein (*ccs*A), 4 hypothetical proteins with unknown function (*ycf*) and translation initiation factor 1 (*inf*A). This makes a total of 113 different genes, among which, 18 present introns (16 one intron and 2 two introns) (Table 3).

Table 3
Genes with introns in the *N. pallida* chloroplast genome and the length of their exons.

Gene	Region	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
*atp*F	LSC	146	726	407		
*clp*P	LSC	69	789	291	641	228
*ndh*A	SSC	553	1450	539		
*ndh*B	IR	777	685	756		
*pet*B	LSC	6	814	642		
*pet*D	LSC	8	720	475		
*rpl*2	IR	393	662	435		
*rpl*16	LSC	9	1129	399		
*rpo*C1	LSC	432	801	1617		
*rps*12	LSC-IR	114		232	536	26
*rps*16	LSC	40	883	245		
*trn*A-UGC	IR	38	802	35		
*trn*G-GCC	LSC	23	702	49		
*trn*I-GAU	IR	42	948	35		
*trn*K-UUU	LSC	37	2611	29		
*trn*L-UAA	LSC	37	536	50		
*trn*V-UAC	LSC	39	621	35		
*ycf*3	LSC	126	728	228	739	153

With the sequences of the protein-coding genes, a functional annotation was performed looking for the metabolic pathways and processes in which they are involved. According to the information in the KEGG database, these genes are distributed in 4 primary classes (metabolism, genetic information processing, cellular processes and organ systems) (Fig. 2; Table S2). The class with the highest number of genes is metabolism, with energy metabolism standing out. In the case of the UniProt database, it showed that most of the genes are involved in biological processes, photosynthesis and generation of metabolite precursors and energy (Fig. 3). Also noteworthy is the large number of genes pertaining to cellular components and the chloroplast.

Codon usage analysis

The codon frequency (Table S3) and the Relative Synonymous Codon Usage (RSCU) were calculated for the whole exome (protein-coding regions). The most frequently used codon in the genome was AAU, which codes for Isoleucine (n = 1144),

followed by AAA, which codes for Lysine (n = 1069). At the other extreme, the least used codons were the three types of STOP codons UGA (n = 17), UAG (n = 18) and UAA (n = 50). But the less frequent amino acid-coding codons were UGC, coding for Cysteine (n = 85), and CGC, coding for Arginine (n = 103). Analyzing the RSCU values it could be seen that, for each amino acid, half of the codons were used with a higher relative frequency than the other half. Also, the preferred codons were A or T/U-ending codons (Fig. 4).

Identification of repetitive sequences:

We looked for repetitive sequences in the genome. The tandem repeats were identified as microsatellites (SSRs) and tandem repeats in general. A total of 142 SSRs were identified (Fig. 5A), with the most abundant repeats being mononucleotide repeats (n = 78), mostly A/T mononucleotides (Fig. 5B). For the remaining SSR (di- to hexanucleotide repeats) between 4 and 18 repeats were found. Tandem repeats larger than hexanucleotides were also identified, finding 164 repeats (Fig. 5C). Thus, a total of 306 tandem repeats were found.

The rest of the repeats were palindromes, direct and inverted repeats. In the genome, 33 palindromes, 250 direct repeats and 240 inverted repeats were found, most of them having between 15 and 29 nucleotides (Fig. 5C).

Phylogenomic relationships of Neltuma pallida:

A phylogeny of the subfamily Caesalpinioideae was made with the chloroplast genomes of 30 species previously published to determine the position of *N. pallida* in the clade. We found that *N. pallida* is grouped together with the other species of *Neltuma* spp., supporting the monophyly of the genus (Fig. 6). The *Neltuma* genus is grouped together with the rest of species that belonged to the group *Prosopis* s.l.: *Prosopis* s.s. and *Strombocarpa*. *Prosopis* s.l. also shows closeness to *Cylicodiscus gabunensis* and the Dichrostachys clade. On the other hand, *P. cineraria* was found within the genus *Neltuma*, instead of within the genus *Prosopis* s.s.

Sequence divergence analysis:

Divergence between the sequences of *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata* with *N. pallida* was calculated (Fig. 7). The least divergence was seen with *N. juliflora*, while the greatest was with *A. ligulata*. When looking at genomic regions, it was found that the most conserved regions are the IRs, regardless of the species with which they are compared. The opposite was determined for the single copy regions (LSC and SSC), where the greatest divergence between sequences is found.

As expected, non-coding regions show the greatest divergence, highlighting some intergenic regions such as *trn*K-*rps*16, *trn*S-*trn*G, *trn*C-*pet*N, *psb*Z-*trn*G, *trn*T-*trn*L, *rbc*L-*acc*D, *acc*D-*ycf*4, *rps*8-*rpl*14 and *rps*3-*rps*19 in LSC, or *ccs*A-*ndh*D and *rps*15-*ycf*1 in SSC. Also, some intronic regions show a higher level of divergence such as in *trn*K, *rps*16, *atp*F, *rpo*C1, *rpl*2, *rpl*16 and *clp*P in LSC or *pet*B, *pet*D and *ndh*A in SSC.

In the coding regions the divergence was smaller, moreover when comparing to other *Prosopis* s.l. species. The genes with the greatest difference in their coding sequences are *rpo*C2, *rpo*B, *acc*D, *cem*A, *pet*A, *rpl*20, *ycf*2, *ndh*F and *ycf*1. There are also genes that only show divergence in one of the species of this group, this is the case for *atp*A, *psb*C, *psa*A and *psb*B in *P. farcta*.

Genetic distance of coding sequences analysis:

Genetic distance was determined using the sequences of 75 chloroplast genes taken pairwise. For this purpose, the genome of *N. pallida* was compared as a reference with those of *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata* (Fig. 8). As expected, the highest average distance was found between *N. pallida* and *A. ligulata* (0.0200), while the lowest was between *N. pallida* and *N. juliflora* (0.0007).

The 10 genes showing on average the greatest genetic distance to *N. pallida* sequences were *clp*P (0.0754), *ycf*1 (0.0231), *rpl*20 (0.0188), *acc*D (0.0187), *psb*T (0.0139), *pet*N (0.0139), *psb*K (0.0134), *rps*3 (0.0133), *rbc*L (0.0131) and *rps*18 (0.0122).

Boundary between regions:

The expansion and contraction of the LSC, IRB, SSC, and IRA of the *N. pallida* chloroplast genome was analyzed by examining the distance between their boundaries with their nearest genes (Fig. 9), and comparing these distances with related species.

In all *Prosopis* s.l. species, the genes closest to the boundary between the LSC and IRB regions (called JLB) were *rps*19 and *rpl*2. In all cases, the *rps*19 gene overlaps with both regions, and *rpl*2 falls in the IRB region. When comparing the distances of *rpl*2 to JLB between *Neltuma* spp. and *Prosopis* s.l. species, a contraction (12 bp) in IRB and an expansion in LSC are found.

Moving on to the next boundary, between IRB and SSC (called JSB), in all *Prosopis* s.l. species, the closest genes are *ycf*1 and *ndh*F. When comparing *N. pallida* with the rest of the species of *Prosopis* s.l., we found a 17 bp contraction in IRB and a contraction in SSC of 20–112 bp.

At the boundary between SSC and IRA (called JSA), the closest genes for *Prosopis* s.l. species were the other copy of *ycf*1 and *trn*N. When comparing the distance between this boundary with the closest genes, it was found in the *N. pallida* SSC, a contraction of 6 bp with respect to the distance found in *N. juliflora* and an expansion of 17 bp with respect to the distance found in *P. farcta* and *S. tamarugo*. On the other hand, in *N. pallida* IRA, a contraction of 16 bp was observed with respect to the rest of *Prosopis* s.l. species.

Finally, near the boundary between IRA and LSC (called JLA) the *rps*19 pseudogene and the trnH gene were found in *Prosopis* s.l. species. In the *Prosopis* s.l. IRAs, no differences were found between species, starting the *rps*19 pseudogene in IRA and reaching the boundary with LSC. The opposite case was observed in LSC, when comparing *N. pallida* with the rest of the *Prosopis* s.l. species, a contraction of 12 and 182 bp was found with respect to *P. farcta* and *N. juliflora*, respectively. With respect to *S. tamarugo* an expansion of only 1 bp difference was found.

Different distances were observed between boundaries and their closest genes, these differences were small among species within *Prosopis* s.l. We also analyzed the case of *A. ligulata* and found two differences. First, the boundaries between regions in the chloroplast genome of *A. ligulata* were at completely different positions than those of the *Prosopis* s.l. species used in this analysis. Second, the closest genes to the boundaries were different.

Discussion

The total genomic DNA of *N. pallida* was sequenced with Illumina technology and its chloroplast genome sequences were extracted and assembled with GetOrganelle. This genome was annotated and compared with chloroplast genomes of other species of *Prosopis* s.l. The assembled chloroplast genome of *N. pallida* was 162381 bp (162.4 Kb) long, with a classical circular quadripartite structure: two inverted repeats (IRA and IRB), a short single-copy region (SSC) and a long single-copy region (LSC) (Fig. 1; Table 1). The same quadripartite structure and similar chloroplast genome sizes have been reported in species of the genera *Prosopis*, *Neltuma* and *Strombocarpa* (161.5–163.7 kb) [46–48]. The same has been found in other mimosoid species [49, 50] and legumes [46]. In general, chloroplast genomic regions conserve similar sizes among closely related species, as in *Prosopis* s.l. However, some mimosoid species present regions with different lengths. This is the case with species of Inga clade and *Albizia* spp. [51, 52], which exhibit expansions in their IRs and reductions in their SSCs, generating longer genomes. Our results agree with some studies mentioning that differences in chloroplast genome sizes are explained by variations in the length of single copy regions, LSC and SSC [47, 53]. Comparing size differences between

IR and single copy regions in *Prosopis* s.l., it was found that LSC and SSC differences were always larger than those between IRs.

In the chloroplast genome of *N. pallida*, 132 genes were identified and the 19 genes located in the IRs were duplicated (Fig. 1, Table 2). The genes in the IRs were also duplicated in the chloroplast genomes of *P. cineraria* and *N. juliflora* [47]. Duplication of genes in the IRs is usual as it has been observed in other mimosoid species, including those with longer IRs. The difference with these species is that their IRs comprise a larger number of genes [51, 52]. In other legumes out of mimosoids, the same 19 duplicated genes were also identified as duplicates [54]. The duplication of the IRs genes in the chloroplast is usual irrespective of the length of the region and the phylogeny

Gene prediction annotation identified 15 genes with one intron and 3 with 2 introns (*clp*P, *rps*12 and *ycf*3) (Table 3). As in other chloroplast genomes, *rps*12 was unevenly distributed in LSC (first exon) and IRB (second and third exons) [55, 56]. On the other hand, in the chloroplast genome of *N. pallida* we found genes that were not present in related species. One case is *trn*G-GCC, that has not been found in *P. cineraria* or in *N. juliflora* chloroplast genomes, but only in *N. glandulosa* [47]. In *N. pallida*, *inf*A, the gene coding for the transcription initiator factor, was also annotated, as it presents an open reading frame (ORF). This ORF is also present in the chloroplast genomes of *Prosopis*, *Strombocarpa* or *Neltuma*, but has not been annotated in these genomes. This gene is absent or present only as a pseudogene in many legume species [54, 57, 58]. The ORF of this gene has been identified in different chloroplast genomes of other legumes as *Albizia julibrissin* and *Lespedeza* spp. [52, 59]. One study has proposed that *inf*A is a gene that has been lost numerous times from chloroplast genomes in the evolutionary history of angiosperms, having been mobilized to the nuclear genome of plants [60]. Another interesting case is that of a *rps*19 segment (103 pb) which is located in IRA as a pseudogene. In all the other species of *Prosopis* s.l., this *rps*19 pseudogene has been found, with incomplete ORFs [47]. The same has been found in species that are not legumes such as *Cerasus humilis* (family Rosaceae) [61] or in *Garcinia paucinervis* (family Clusiaceae) [62].

The GC content (GC%) of the total chloroplast genome of *N. pallida* was determined to be 35.97% (Table 1). Similar values have been found in other chloroplast genomes of *P. cineraria* and *N. juliflora* [47]. This has been seen in most of sequenced chloroplast genomes, with values around 36.2%, except for a few taxonomic groups such as *Selaginella* spp. (family Selaginellaceae) with 54.8% [63, 64]. Relatively lower GC% values were found in single copy regions (33.26% and 30.46% for LSC and SSC, respectively), but this was higher in IRs (42.77%). Other work has assumed that this higher GC% value in IRs is due to the presence of rRNA and tRNA genes [47, 65, 66]. In this work, the GC% of rRNA and tRNA genes was calculated to be greater than 53%, supporting this assumption. It is important to highlight the fact that this high GC% in IRs helps their conservation, as it has been shown to slow down the rate of nucleotide substitution [67].

Regarding codon usage in protein-coding genes, those ending in A or T/U were preferred, having RSCU values higher than 1. Similar results were reported in the chloroplast genome of *Albizia julibrissin* [52] and *Stryphnodendron adstringens* [49], both species from the subfamily Caesalpinioideae. This has also been found in more distant species such as *Salix floderusii* [68] or *Chrysosplenium* spp. [69]. The work of Duan et al. [70], also found a preference for the use of codons ending in A or T/U. They determined that there is a selective pressure favoring the use of these codons. This evolutionary pressure has only been found in the third codon position, but not in the rest of the chloroplast genome. Positive selective pressure favors the expression efficiency and conservation of highly important genes [71]; this can be an explanation for of the observed conservation in the third 3rd codon position in the protein-coding genes.

We identified 142 microsatellites (SSRs) in the *N. pallida* chloroplast genome, with the most abundant being mononucleotide repeats. This is similar to what was found by Asaf et al. [47] in the chloroplast genomes of *P. cineraria* and *N. juliflora*. Nonetheless, they found a lower amount of SSRs and found neither penta- nor hexanucleotides in the aforementioned genomes. We also identified more repeats of all types compared to other chloroplast genomes of *Prosopis*, *Strombocarpa* or *Neltuma*. This could be explained by the use of different tools to identify repetitive sequences. Das & Ghosh [72] mention that specialized programs for the identification of repetitive sequences often present different results depending on the algorithm used. Despite that differences between algorithms make accurate comparisons difficult,

general comparisons, such as the higher number of repetitive mononucleotides among SSRs in the chloroplast genomes of distant species, are still possible [40, 73, 74]. There is a need to establish a standardized methodology for the identification of repetitive sequences in the chloroplast genomes. This will allow for more reliable comparisons. The analysis of repetitive sequences within chloroplast genomes is of high importance as they serve to study genetic diversity and phylogeny through the development of molecular markers [73, 75]. Similarly, in the chloroplast genomes, the repetitive sequences have a role in the presence or absence of polymorphisms [76].

The chloroplast genome sequence is a powerful tool often used to elucidate phylogenetic relationships [77, 78]. Although cases of discordance between chloroplast and nuclear phylogenies can also be found [79, 80]. This is why information from both sources should be used concurrently when possible. In this work, *N. pallida* formed a subclade with *P. cineraria*, *N. juliflora*, and *N. glandulosa*. The subclade with these four species, formerly included in the *Prosopis* genus, was reported by Asaf et al. [47], as a monophyletic group. This differs completely with Hughes et al. [2] whose results recently supported that the *Prosopis* s.l. is polyphyletic. In their work, they used about 1000 nuclear genes sequenced by Ringelberg et al. [81], in combination with morphological characters. They proposed the separation of *Prosopis* s.l. into 4 different genera, placing *N. pallida*, *N. juliflora* and *N. glandulosa* within the genus *Neltuma*. On the other hand, *P. cineraria* and *P. farcta*, remained as species of the genus *Prosopis* s.s. This is one of the most discordant points with what was found in our work, since it can be seen in Fig. 6 that *P. cineraria* and *P. farcta* are relatively distant. According to their geographic distribution, *P. cineraria* and *P. farcta* originated and can be found in western Asia [82–84]. While *N. pallida*, *N. juliflora* and *N. glandulosa* are native to the Americas [2, 7]. Other cases of incongruent phylogeny between nuclear and chloroplast information have been reported in the literature [85–90]. In all of these cases, incomplete lineage delimitation or introgression/hybridization has been determined to have influenced the discordances. In some Caesalpinioideae species both options are plausible, this is the case of *Prosopis cineraria*. There is also evidence of interspecific and intergeneric hybridization in *Prosopis* s.l. [91, 92]. However, due to the lack of reference chloroplast genome sequences for the rest of Caesalpinioideae species, they could not be included in our analysis.

When analyzing the divergence between chloroplast genome sequences among species of *Prosopis* s.l., a general high level of conservation was observed. But it was also found: first, that non-coding regions, such as intergenic sequences or introns, had the lowest similarity; second, that LSC and SSC are the regions with the lowest identity, as has been seen in different taxonomic groups [93–95]; and finally, that IRs are highly conserved, concordant with the importance of the biological functions they serve, such as genomic stabilization and gene conservation [96]. Thus, a mutation in IRs can lead to numerous structural and sequence changes at many other sites in chloroplast genomes [97–99]. Another interesting point when comparing the chloroplast sequences of *Prosopis* s.l. is that variations were only detected in 9 genes. Something similar was found by Asaf et al. [47] when comparing *N. juliflora* and *P. cineraria*, with *N. glandulosa*, finding divergence in 5 of these 9 genes. This may indicate that these are the most prone to mutations in the chloroplast genomes of this group. Despite the variation found, the genetic distances are low. The *clp*P gene showed the largest variation (0.0754). This gene, which codes for a subunit of the ATP-dependent chloroplast protease, has been reported to show a high rate of amino acid substitution, which is related to protein structure [100]. The second gene with the highest genetic distance was *ycf*1 (0.0231). Some work highlights the ability of *ycf*1, together with the intergenic region between it and the *ndh*F gene, as a barcode for land plants [101, 102], indicating its high diversity level.

Boundary shift analysis between chloroplast genome regions showed that all boundaries were at different positions in *N. pallida* when compared to the other *Prosopis* s.l. species. Yet, these differences were small, with the maximum being a few hundred bases at the beginning of LSC. The same was found by Asaf et al. [47] when evaluating other *Prosopis* s.l. species. The shifting of boundaries between regions is one of the main explanations for variation in chloroplast genome size. This has been found when comparing close species in different groups, being explained by expansion and contraction of chloroplast regions [103–105]. But these variations are usually small among related species, so the lengths of their chloroplast genomes and regions are similar as well.

In this work, we succeeded in sequencing, assembling and annotating the chloroplast genome of *N. pallida*. The assembled and annotated genome was useful to make comparisons with other species of the genus *Prosopis* s.l., finding that this genome is highly similar to those of other species, but showing some differences at the structural and genetic level. We also used the sequence produced and other published chloroplast sequences to perform a phylogenomic analysis that showed that *Prosopis* s.l. is a monophyletic group when using chloroplast information. Finally, divergence comparisons with other chloroplast genomes of *Neltuma* and *Prosopis* s.l. showed that within the group there is a high level of sequence identity. However, there are certain divergent sequences and genes that may be of interest for the development of molecular markers. The information generated by this research can be used in the development of new lines of research that contribute to understanding the diversity and conservation of this species in a more efficient way.

Conclusions

High quality assembly and annotation of the chloroplast genome of the Algarrobo tree *N. pallida* was constructed from paired end reads sequenced by Illumina. This 162381 bp genome exhibits a classical quadripartite structure with a LSC of 91805 bp, an SSC of 18748 bp and two IRs of 25914 bp, each. The average GC% of the genome was 35.97%, while this value was lower in the single copy regions (30.46% – 33.26%) and higher in the IRs (42.77%). Within the genome, 132 genes were annotated, with 19 genes being duplicated and 18 having introns. Of the total number of genes, 85 were protein coding (78 unique genes), 8 rRNA coding genes (4 unique genes) and 39 tRNA coding genes (31 unique genes).

Codon usage of the 85 protein-coding genes showed that the most abundant amino acids in these genes are Isoleucine and Lysine. The RSCU value reveals a clear bias for codons having A or T/U in their third position, repeating this pattern for all amino acids. A total of 142 SSRs were identified, mainly mononucleotides (n = 78). More tandem repeats (n = 164), palindromes (n = 33) and 490 interspaced repeats between direct and inverted repeats could also be identified.

At the phylogenetic level, it was found that *N. pallida* forms a clade with the other species of the genus *Neltuma* and *P. cineraria*. In turn, the whole *Prosopis* group is shown as a monophyletic clade, being able to delimit the genera *Neltuma* and *Strombocarpa*, while *Prosopis* s.s. shows itself as polyphyletic. When comparing the genomic sequences of several *Prosopis* s.l. species with the chloroplast genome of *N. pallida*, a high level of conservation was evident. The most divergent regions were the non-coding sequences, while divergence was found in only 9 genes coding sequences. Also, when analyzing the displacement of the boundaries between genomic regions, it was determined that the differences between *Prosopis* s.l. species were small, showing a high degree of conservation among them.

Methods

Plant material:

Young leaves of *N. pallida* were collected from an adult Algarrobo tree located in the Bosque de Pómac Historic Sanctuary, Lambayeque Department, Peru (6°26'39.4" S 79°48'16.6.6" W). The collection was carried out under the authorization granted by “Resolución Jefatural de Santuario Histórico Bosque de Pómac N° 003-2020-SERNANP-JEF”. The collected samples were placed in paper envelopes, in which they were transported for DNA extraction. The species was identified according to the descriptions of Burkart [1] and Hughes et al. [2] and the voucher can be found in the Museo de Historia Natural - UNMSM herbarium (USM N° 335439) in Lima, Peru.

DNA extraction and sequencing:

For DNA extraction, we used 0.2 g of ground sample in 2% CTAB buffer and followed the protocol published by Doyle [106] with minor modifications. The extracted DNA was treated purified with a DNAse-free RNAse A treatment at 37°C for 1 hr. DNA quality and concentration were assessed with Nanodrop™ One C (Thermo Scientific, Massachusetts, USA) and Qubit™ 4 (Invitrogen, Massachusetts, USA), respectively. Sample integrity was also verified by 1% agarose gel electrophoresis.

DNA sequencing was performed using Illumina methodology by contracting the services of an external laboratory. The TruSeq DNA PCR-Free kit (Illumina Inc., California, USA) was used to construct the sequencing library.

The quality of the generated reads was evaluated with FastQC v0.11.9 (https://github.com/s-andrews/FastQC) and filtered with Trimmomatic v0.39 (https://github.com/usadellab/Trimmomatic) [107].

De novo assembly of Neltuma pallida chloroplast genome:

To obtain the sequence of the *N. pallida* chloroplast genome (Genbank: OR178743), a de novo assembly was performed with GetOrganelle v1.7.6.1 (https://github.com/Kinggerm/GetOrganelle) [108]. This is a toolkit that combines Bowtie2 [109], BLAST [110], SPAdes [111], and Python libraries to identify sequences belonging to the chloroplast DNA and assemble the chloroplast genome. GetOrganelle options were tuned to use the data filtered in the previous step (forward and reverse reads), to do 1000 rounds of extension iterations. We also set up GetOrganelle to search sequences corresponding to plant plastids. The other options were left as default. The obtained GFA file was visualized with Bandage v0.9.0 (https://rrwick.github.io/Bandage/) [112] to explore the chloroplast genome structure. The average coverage of the final genome assembly was 130X.

Genome annotation of Neltuma pallida chloroplast genome:

Genome annotation by prediction was performed using GeSeq v2.03 [113] in the Chlorobox web server (https://chlorobox.mpimp-golm.mpg.de/). The chloroplast genome of *N. juliflora* (Genbank: NC049132) was taken as reference. Chloroplast Inverted Repeats (IR), *rps*12 interspersed gene, protein-coding sequences, transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) were annotated. For proteins and RNAs, 25% and 85% identity were used as thresholds for annotation, respectively. Furthermore, tRNAscan-SE v2.0.7 [114], found on the same server, was used as a second tRNA annotator. Additionally, the external annotator Chloë v0.1.0 (https://github.com/ian-small/chloe), which is also found on the web server, was used. A manual curation was performed to finish the annotation for each gene, comparing the genes with their homologues in other chloroplast genomes of species of the genera *Neltuma*, *Prosopis* and *Strombocarpa*.

The sequences of the annotated protein-coding genes were blasted in the KEGG (https://www.kegg.jp/) and UniProt (https://www.uniprot.org/) databases. In this way we sought to perform functional annotation.

Codon usage of protein-coding sequences:

This analysis was only performed on protein coding sequences. Codon usage was analyzed by calculating codon frequency and the Relative Synonym Codon Usage (RSCU) values. If the RSCU was greater than 1, this codon was considered to be used more frequently, whereas if the RSCU was less than 1, the opposite was considered to be true. DAMBE5 v7.3.2 (http://dambe.bio.uottawa.ca/index.aspx) [115] was used to perform this analysis.

Identification of repetitive sequences in the chloroplast genome:

Several programs were used for the identification of repetitive sequences. The identification of tandem repeats, both microsatellite repeats (SSR, Short Sequence Repeat) and long sequences was performed with the Phobos v3.3.12 program (https://www.ruhr-uni-bochum.de/ecoevo/cm/cm_phobos.htm) [116]. The threshold for determining repeats was a minimum of 10 repeats for mononucleotides, 8 repeats for dinucleotides, 4 repeats for trinucleotides and tetranucleotides, and 3 repeats for pentanucleotides and hexanucleotides.

For the remaining repetitive sequences: palindromes (P), forward (F), and reverse (R), the RepEX web server (http://bioserver2.physics.iisc.ac.in/RepEx/index.html) [117] was used. Additionally, IUPACpal was used (https://sourceforge.net/projects/iupacpal/) [118] for palindromic sequences, and Vmatch v2.3.1 (http://www.vmatch.de/) [119] for forward repeats. For all of these repeats, a minimum length of 15 bp was considered, and 90% identity with respect to their template as threshold.

Phylogenomic relationships of Neltuma pallida:

The *N. pallida* chloroplast genome was aligned with previously published chloroplast genomes of 30 species of the Caesalpinioideae subfamily (Table S7). Sequences were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) and aligned using the MAFFT web server (https://mafft.cbrc.jp/alignment/software/) [120] with default options.

Phylogenetic inference was done with two types of algorithms: Bayesian Inference (BI), in Beast2 v2.7.3 (https://www.beast2.org/) [121], and Maximum Likelihood (ML), in RAxML-HPC2 v8.2.12 [122]. For BI, the GTR + I + G substitution model (Yang 1994) with a Markov Chain Monte Carlo of 1500000 steps was used, removing the initial 10% of these. For ML, the GTR + I + G model with 1000 Bootstrap replicates was used. We used jModelTest 2.1.10 v20160303 (https://github.com/ddarriba/jmodeltest2) [123] to determine the substitution models.

Sequence divergence of chloroplast genomes:

The Vista web server (https://genome.lbl.gov/vista/index.shtml) [124], with the mVista function in Shuffle-LAGAN mode was used for sequence divergence analysis. This mode is used for detecting rearrangements within genomes. The chloroplast genome sequence of *N. pallida* was used as a reference along with its annotation. The comparison was made with the chloroplast genomes of *N. juliflora*, *P. farcta* (Genbank: NC061227), *S. tamarugo* (Genbank: NC060845) and *A. ligulata* (Genbank: NC026134).

Genetic distance analysis of coding sequences:

Sequences of 74 chloroplast genes of *N. pallida*, *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata* (Genbank: NC026134) were aligned using ClustalW [125] in Bioedit v7.2.6 [126]. The generated alignments were used to calculate the genetic distance of the sequences of all species from *N. pallida*. The distance was calculated using the p-distance model with Mega X v10.1.8 (https://www.megasoftware.net/) [127].

Boundary shift analysis between chloroplast genome regions:

A comparison was made of the position of boundaries between single copy regions (LSC and SSC) and inverted regions (IRA and IRB) in 5 species of the subfamily Caesalpinioideae: *N. pallida*, *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata*. These boundaries were called JLB (LSC-IRB boundary), JSB (IRB-SSC), JSA (SSC-IRA) and JLA (IRA-LSC). For the positions of the boundaries, the distance between the boundaries and the genes closest to or over the boundaries was calculated.

Abbreviations

BI	Bayesian Inference
F	Forward repeat
IR	Inverted Repeat
JLA	Boundary between LSC and IRA
JLB	Boundary between LSC and IRB
JSA	Boundary between SSC and IRA
JSB	Boundary between SSC and IRB
LSC	Large Single Copy
ML	Maximum Likelihood
ORF	Open Reading Frame
P	Palindrome repeat
R	Reverse repeat
rRNA	Ribosomal RNA
RSCU	Relative Synonymous Codon Usage
SSC	Small Single Copy
SSR	Simple Sequence Repeat or Microsatellite
tRNA	Transfer RNA

Declarations

Acknowledgements:

We would like to thank the chief of the Santuario Histórico Bosque de Pómac (SHBP), Eng. Sirley Bernabé and the SHBP rangers, in particular William Zeña for helping us with transportation, sample collection and his extensive knowledge of the forest. We would also like to thank Edgar Neyra for his help.

Author Contribution:

GO conceived the study and acquired the funding. EC and RL performed the DNA extraction and the bioinformatic analysis. All authors participated in the manuscript writing and approved its final version.

Funding:

The work was funded financed by the Research and Graduate Studies Office of the UNMSM within the framework of the research project: "Sequencing of the genome of the Algarrobo tree (*Prosopis pallida*) and identification of genes for resistance to biotic factors using bioinformatics tools" through the Research Projects Program for Research Groups [RECTORAL RESOLUTION No. 01686-R-20, Code: B20100311].

Data availability

The data supporting the findings of this study is freely available in GenBank on the NCBI website (https://www.ncbi.nlm.nih.gov/) using the accession number OR178743, which corresponds to the *Neltuma pallida*

chloroplast genome sequence. The reads that were used to assemble the chloroplast genome in this study were deposited at the NCBI Sequence Read Archive (SRA) under accession SRR25007997.

Ethics approval and consent to participate

The authors confirm that all methods comply with local and national regulations. The authors have collected plant material under the research permit N° 003-2020-SERNANP-JEF and the certificate of origin for biological samples N° 001-2021-SERNANP-SHBP from the “Servicio Nacional de Áreas Naturales Protegidas por el Estado” (SERNANP).

Consent for publication:

Not applicable

Competing interests:

The authors declare no competing interests

References

1. Burkart A. A Monograph of the Genus Prosopis (leguminosae Subfam. Mimosoideae). J Arnold Arbor. 1976;57(4):450–525.
2. Hughes CE, Ringelberg JJ, Lewis GP, Catalano SA. Disintegration of the genus Prosopis L. (Leguminosae, Caesalpinioideae, mimosoid clade). PhytoKeys. 2022;205:147–89.
3. Padrón E, Navarro-Cerrillo RM. Aboveground biomass in Prosopis pallida (Humb. and Bonpl. ex Willd.) H. B. K. ecosystems using Landsat 7 ETM + images. Rev Chil Hist Nat. 2007;80(1).
4. Organismo de Supervisión de los Recursos Forestales y de Fauna Silvestre. Aprovechamiento forestal maderable en bosques secos en el norte del Perú. 2018. https://www.osinfor.gob.pe/publicaciones/aprovechamiento-forestal-maderable-en-bosques-secos-en-el-norte-del-peru/. Accessed 2 Jan 2023.
5. Cruzado Jacinto L, Chávez-Villavicencio CL, Charcape-Ravelo JM. Uso y selección de las partes aéreas del Algarrobo Prosopis pallida (Fabaceae) por reptiles, aves y mamíferos en Sechura (Piura – Perú). Rev Peru Biol. 2019;26(1):081–6.
6. Servicio Nacional Forestal y de Fauna Silvestre. Avances Sobre la Investigación de. “Algarrobo” Prosopis (Fabaceae) en la Costa Norte del Perú. Lima, Peru: Ministerio de Desarrollo Agrario y Riego; 2021. https://www.gob.pe/institucion/serfor/informes-publicaciones/1977079-avances-sobre-la-investigacion-de-algarrobo-en-la-costa-norte-del-peru. Accessed 2 Jan 2023.
7. Barros Asenjo S. El género Prosopis, valioso recurso forestal de las zonas áridas y semiáridas de América, Asia y Africa. Cienc Investig For. 2010;16(1):91–127.
8. Moreno MC, Torres LM, del Campos R. CM. Nuevos aportes al uso de Prosopis flexuosa en el centro oeste de Argentina y su interpretación en el marco general de la ecorregión del monte. 2018;16(3),18–35.
9. Mostacero León J, editor. Plantas medicinales del Perú: taxonomía, ecogeografía, fenología y etnobotánica. 1st ed. Lima, Peru: Asamblea Nacional de Rectores; 2011.
10. Depenthal J, Yoder LSM. Community Use and Knowledge of Algarrobo (Prosopis pallida) and Implications for Peruvian Dry Forest Conservation. Rev Cienc Ambient. 2017;52(1):49.
11. Salazar Zarzosa P, Mendieta-Leiva G, Navarro-Cerrillo RM, Cruz G, Grados N, Villar R. An ecological overview of Prosopis pallida, one of the most adapted dryland species to extreme climate events. J Arid Environ. 2021;193:104576.
12. Mokgalaka-Matlala NS, Flores-Tavizón E, Castillo-Michel H, Peralta-Videa JR, Gardea-Torresdey JL. Arsenic tolerance in mesquite (Prosopis sp.): Low molecular weight thiols synthesis and glutathione activity in response to arsenic. Plant

Physiol Biochem. 2009;47(9):822–6.

13. Santos-Jallath J, Castro-Rodríguez A, Huezo-Casillas J, Torres-Bustillos L. Arsenic and heavy metals in native plants at tailings impoundments in Queretaro, Mexico. Phys Chem Earth Parts ABC. 2012;37–39:10–7.
14. D´Aubeterre R, Principal J, Barrios C, Graterol Z. Insectos plaga en vainas de Acacia tortuosa y Prosopis juliflora para consumo animal en las zonas semiáridas del estado Lara, Venezuela. Zootec Trop. 2012;30(2):147–53.
15. Calderón Jiles GB. Prosopis tamarugo Phil. un árbol endémico del desierto de Atacama. Propuesta de niveles umbrales de profundidad freática para su conservación [Tesis para optar al título de Ingeniero en Recursos Naturales Renovables]. Santiago, Chile: Universidad de Chile; 2016. https://repositorio.uchile.cl/handle/2250/151714.
16. Iturre MC, Araujo PA, Rueda MP, Rueda CV, Pece MG. Reclutamiento y mortalidad de las principales especies arbóreas del Chaco Semiárido, Argentina. Quebracho - Rev Cienc For. 2020;28(2):134–44.
17. Whaley OQ, Borda C, Moat J, Wilkinson T, Bravo Sánchez A, Gagné RJ, et al. Ecología y diagnóstico de Enallodiplosis discordis (Diptera: Cecidomyiidae): un nuevo defoliador ferozcon repercusiones directas en la pérdida del bosqueseco de Prosopis y los medios de vida en Perú. Rev Peru Biol. 2020;27(4):451–82.
18. García J, Alvárez P, Izzo M, Zurita-Bianchini C, Sarmiento M. Establecimiento y reposición de pérdidas en plantaciones de Prosopis alba en Santiago del Estero, Argentina. Revista Forestal del Perú. 2022;37(1):41–53.
19. Servicio Nacional Forestal y de Fauna Silvestre. Informe del Inventario Nacional Forestal y de Fauna Silvestre - Primer Panel. Lima, Peru: Ministerio de Desarrollo Agrario y Riego. ; 2019. https://www.serfor.gob.pe/portal/wp-content/uploads/2020/03/INFORME-DEL-INFFS-PANEL-1.pdf. Accessed 2 Jan 2023.
20. Ramos Jauregui LYR. Estimación del Efecto del Cambio Climático en la Precipitación en la Costa Norte del Perú Usando Simulaciones de Modelos Climáticos Globales [Tesis para optar al título de Ingeniero Meteorólogo]. Lima, Peru: Universidad Nacional Agraria La Molina; 2014. http://met.igp.gob.pe/publicaciones/2014/tesis_final_yakelyn_ramos.pdf.
21. Autoridad Nacional del Agua. Evaluación de la veda en el acuífero de Motupe y margen derecha del río La Leche. Ministerio de Desarrollo Agrario y Riego. ; 2016. https://www.ana.gob.pe/sites/default/files/normatividad/files/estudio_evaluacion_de_la_veda_acuifero_motupe.pdf. Accessed 16 Jan 2023.
22. Juárez NG, Grados QN, Cruz AG. Insectos asociados a Prosopis pallida (Humb.& Bonpl.ex.Wild.) en el campus de la Universidad de Piura, Perú. Zonas Áridas. 2016;16(1):28.
23. Inca' Biotec. Servicio de Análisis Molecular de Agentes Patógenos del Algarrobo para Determinar Causas de Mortalidad. Tumbes, Peru; 2019.
24. Servicio Nacional Forestal y de Fauna Silvestre. Fichas Resumenes de las Autorizaciones de Investigación Científica de Flora Silvestre Fuera de Áreas Naturales Protegidas Otorgadas por el SERFOR en el Año 2020. Lima, Peru: Ministerio de Desarrollo Agrario y Riego. ; 2021. https://www.gob.pe/institucion/serfor/informes-publicaciones/2174548-fichas-resumenes-de-las-autorizaciones-de-investigacion-cientifica-de-flora-silvestre-fuera-de-areas-naturales-protegidas-otorgadas-por-serfor-en-el-ano-2020. Accessed 2 Jan 2023.
25. Alban L, Matorel M, Romero J, Grados N, Cruz G, Felker P. Cloning of elite, multipurpose trees of the Prosopis juliflora/pallida complex in Piura, Peru. Agrofor Syst. 2002;54(3):173–82.
26. Balcazar Huamán MA. Variabilidad de rasgos funcionales de 7 poblaciones de Algarrobo (Prosopis pallida H.B.K) bajo condiciones controladas [Tesis para optar al título de Ingeniero Agrónomo]. Piura, Peru: Universidad Nacional de Piura; 2017. http://repositorio.unp.edu.pe/handle/UNP/1253.
27. Chinguel-Laban DO, Charcape-Ravelo JM. Biomasa y secuestro de carbono radicular, en bosque natural de "Algarrobo" en el campus de la Universidad de Piura (UDEP). Rev Investig Agroproducción Sustentable. 2019;3(3):53–66.
28. Landeras G, Alfonso M, Pasiecznik NM, Harris PJC, Ramírez L. Identification of Prosopis juliflora and Prosopis pallida Accessions Using Molecular Markers. Biodivers Conserv. 2006;15(5):1829–44.

29. Palacios RA, Burghardt AD, Frías-Hernández JT, Olalde-Portugal V, Grados N, Alban L, et al. Comparative study (AFLP and morphology) of three species of Prosopis of the Section Algarobia: P. juliflora, P. pallida, and P. limensis. Evidence for resolution of the "P. pallida–P. juliflora complex. Plant Syst Evol. 2012;298(1):165–71.
30. Bukhari YM. Nuclear DNA Amounts in Acacia and Prosopis (Mimosaceae) and their Evolutionary Implications. Hereditas. 2004;126(1):45–51.
31. Harris P. Differentiation of Prosopis juliflora (Sw.) DC. and, pallida P. (H. & B. ex. Willd.) H.B.K. using foliar characters and ploidy. For Ecol Manag. 2003;180(1–3):153–64.
32. Trenchard LJ, Harris PJC, Smith SJ, Pasiecznik NM. A review of ploidy in the genus Prosopis (Leguminosae). Bot J Linn Soc. 2008;156(3):425–38.
33. Bessega C, Vilardi JC, Saidman BO. Genetic relationships among American species of the genus Prosopis (Mimosoideae, Leguminosae) inferred from ITS sequences: evidence for long-distance dispersal. J Biogeogr. 2006;33(11):1905–15.
34. Palmer JD. Comparative organization of chloroplast genomes. Annu Rev Genet. 1985;19(1):325–54.
35. De Las Rivas J, Lozano JJ, Ortiz AR. Comparative Analysis of Chloroplast Genomes: Functional Annotation, Genome-Based Phylogeny, and Deduced Evolutionary Patterns. Genome Res. 2002;12(4):567–83.
36. Sugiura M, Hirose T, Sugita M. Evolution and mechanism of translation in chloroplasts. Annu Rev Genet. 1998;32(1):437–59.
37. Turudić A, Liber Z, Grdiša M, Jakše J, Varga F, Šatović Z. Towards the Well-Tempered Chloroplast DNA Sequences. Plants. 2021;10(7):1360.
38. Ahrens CW, Supple MA, Aitken NC, Cantrill DJ, Borevitz JO, James EA. Genomic diversity guides conservation strategies among rare terrestrial orchid species when taxonomy remains uncertain. Ann Bot. 2017;119(8):1267–77.
39. Supple MA, Shapiro B. Conservation of biodiversity in the genomics era. Genome Biol. 2018;19(1):131.
40. Li QJ, Su N, Zhang L, Tong RC, Zhang XH, Wang JR, et al. Chloroplast genomes elucidate diversity, phylogeny, and taxonomy of Pulsatilla (Ranunculaceae). Sci Rep. 2020;10(1):19781.
41. Shang C, Li E, Yu Z, Lian M, Chen Z, Liu K, et al. Chloroplast Genomic Resources and Genetic Divergence of Endangered Species Bretschneidera sinensis (Bretschneideraceae). Front Ecol Evol. 2022;10:873100.
42. Bell DA, Robinson ZL, Funk WC, Fitzpatrick SW, Allendorf FW, Tallmon DA, et al. The Exciting Potential and Remaining Uncertainties of Genetic Rescue. Trends Ecol Evol. 2019;34(12):1070–9.
43. Onley IR, Moseby KE, Austin JJ. Genomic Approaches for Conservation Management in Australia under Climate Change. Life. 2021;11(7):653.
44. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17(1):134.
45. Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. Acta Physiol Plant. 2020;42(6):98.
46. Schwarz EN, Ruhlman TA, Sabir JSM, Hajrah NH, Alharbi NS, Al-Malki AL, et al. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. J Syst Evol. 2015;53(5):458–68.
47. Asaf S, Khan AL, 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. 2020;21(9):3280.
48. van den Contreras Díaz R, Navarrete Fuentes MJ, Arias Aburto M, van den Contreras Díaz R, et al. Caracterización del genoma completo del cloroplasto de Prosopis tamarugo (Prosopis, Leguminosae), una especie arbórea endémica en peligro del Desierto de Atacama. Bosque Valdivia. 2021;42(3):365–70.
49. de Souza UJB, Nunes R, Targueta CP, Diniz-Filho JAF, de Telles MP. The complete chloroplast genome of Stryphnodendron adstringens (Leguminosae - Caesalpinioideae): comparative analysis with related Mimosoid species. Sci Rep. 2019;9(1):14206.

50. Thi NPA, Khang DT, Khoa NT, Le VM, Do HDK. The complete chloroplast genome of Mimosa pigra L. (Fabaceae), a notorious invasive plant. Mitochondrial DNA Part B. 2022;7(6):1063–5.
51. Asaf S, Khan A, Khan AL, Al-Harrasi A, Al-Rawahi A. Complete Chloroplast Genomes of Vachellia nilotica and Senegalia senegal: Comparative Genomics and Phylogenomic Placement in a New Generic System. PLoS ONE. 2019;14(11):e0225469.
52. Zhang J, Huang H, Qu C, Meng X, Meng F, Yao X, et al. Comprehensive analysis of chloroplast genome of Albizia julibrissin Durazz. (Leguminosae sp). Planta. 2021;255(1):26.
53. Su Q, Liu L, Zhao M, Zhang C, Zhang D, Li Y, et al. The complete chloroplast genomes of seventeen Aegilops tauschii: genome comparative analysis and phylogenetic inference. PeerJ. 2020;8:e8678.
54. Liu W, Kong H, Zhou J, Fritsch PW, Hao G, Gong W. Complete Chloroplast Genome of Cercis chuniana (Fabaceae) with Structural and Genetic Comparison to Six Species in Caesalpinioideae. Int J Mol Sci. 2018;19(5):1286.
55. Hildebrand M, Hallick RB, Passavant CW, Bourque DP. Trans-splicing in chloroplasts: the rps 12 loci of Nicotiana tabacum. Proc Natl Acad Sci USA. 1988;85(2):372–6.
56. Liu S, Wang Z, Wang H, Su Y, Wang T. Patterns and Rates of Plastid rps12 Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns. Sci Rep. 2020;10(1):9394.
57. Lei W, Ni D, Wang Y, Shao J, Wang X, Yang D, et al. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Sci Rep. 2016;6(1):21669.
58. Li C, Zhao Y, Xu Z, Yang G, Peng J, Peng X. Initial Characterization of the Chloroplast Genome of Vicia sepium, an Important Wild Resource Plant, and Related Inferences About Its Evolution. Front Genet. 2020;11:73.
59. Somaratne Y, Guan DL, Wang WQ, Zhao L, Xu SQ. The Complete Chloroplast Genomes of Two Lespedeza Species: Insights into Codon Usage Bias, RNA Editing Sites, and Phylogenetic Relationships in Desmodieae (Fabaceae: Papilionoideae). Plants. 2020;9(1):51.
60. Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, et al. Many Parallel Losses of infA from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus. Plant Cell. 2001;13(3):645–58.
61. Mu X, Wang P, Du J, Gao YG, Zhang J. The chloroplast genome of Cerasus humilis: Genomic characterization and phylogenetic analysis. PLoS ONE. 2018;13(4):e0196473.
62. Wang Y, Zhao B, Lu Z, Shi Y, Li J. The complete chloroplast genome provides insight into the polymorphism and adaptive evolution of Garcinia paucinervis. Biotechnol Biotechnol Equip. 2021;35(1):377–91.
63. Smith DR. Unparalleled GC content in the plastid DNA of Selaginella. Plant Mol Biol. 2009;71(6):627–39.
64. Jiao Y, Guo H. Chapter Nine - Prehistory of the Angiosperms: Characterization of the Ancient Genomes. In: Paterson AH, editor. Advances in Botanical Research. London, UK: Academic Press; 2014. pp. 223–45.
65. Talat F, Wang K. Comparative Bioinformatics Analysis of the Chloroplast Genomes of a Wild Diploid Gossypium and Two Cultivated Allotetraploid Species. Iran J Biotechnol. 2015;13(3):47–56.
66. Guo S, Liao X, Chen S, Liao B, Guo Y, Cheng R, et al. A Comparative Analysis of the Chloroplast Genomes of Four Polygonum Medicinal Plants. Front Genet. 2022;13:764534.
67. Li FW, Kuo LY, Pryer KM, Rothfels CJ. Genes Translocated into the Plastid Inverted Repeat Show Decelerated Substitution Rates and Elevated GC Content. Genome Biol Evol. 2016;8(8):2452–8.
68. Ren W, Jiang Z, Zhang M, Kong L, Zhang H, Liu Y, et al. The chloroplast genome of Salix floderusii and characterization of chloroplast regulatory elements. Front Plant Sci. 2022;13:987443.
69. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of Chrysosplenium (Saxifragaceae). BMC Genomics. 2020;21(1):621.
70. Duan H, Zhang Q, Wang C, Li F, Tian F, Lu Y, et al. Analysis of codon usage patterns of the chloroplast genome in Delphinium grandiflorum L. reveals a preference for AT-ending codons as a result of major selection constraints. PeerJ. 2021;9:e10787.

71. Guan DL, Ma LB, Khan MS, Zhang XX, Xu SQ, Xie JY. Analysis of codon usage patterns in Hirudinaria manillensis reveals a preference for GC-ending codons caused by dominant selection constraints. BMC Genomics. 2018;19(1):542.
72. Das G, Ghosh I. Benchmarking tools for DNA repeat identification in diverse genomes. bioRxiv. 2021. https://doi.org/10.1101/2021.09.10.459798.
73. Shukla N, Kuntal H, Shanker A, Sharma SN. Mining and analysis of simple sequence repeats in the chloroplast genomes of genus Vigna. Biotechnol Res Innov. 2018;2(1):9–18.
74. Asaf S, Khan AL, Khan MA, Shahzad R, Lubna, Kang SM, et al. Complete chloroplast genome sequence and comparative analysis of loblolly pine (Pinus taeda L.) with related species. PLoS ONE. 2018;13(3):e0192966.
75. Ping J, Feng P, Li J, Zhang R, Su Y, Wang T. Molecular evolution and SSRs analysis based on the chloroplast genome of Callitropsis funebris. Ecol Evol. 2021;11(9):4786–802.
76. Vieira L do, Faoro N, Rogalski H, Fraga M, de Cardoso HP, de Souza RLA, et al. The Complete Chloroplast Genome Sequence of Podocarpus lambertii: Genome Structure, Evolutionary Aspects, Gene Content and SSR Detection. PLoS ONE. 2014;9(3):e90618.
77. Xu J, Shen X, Liao B, Xu J, Hou D. Comparing and phylogenetic analysis chloroplast genome of three Achyranthes species. Sci Rep. 2020;10(1):10818.
78. Namgung J, Do HDK, Kim C, Choi HJ, Kim JH. Complete chloroplast genomes shed light on phylogenetic relationships, divergence time, and biogeography of Allioideae (Amaryllidaceae). Sci Rep. 2021;11(1):3262.
79. Rose JP, Toledo CAP, Lemmon EM, Lemmon AR, Sytsma KJ. Out of Sight, Out of Mind: Widespread Nuclear and Plastid-Nuclear Discordance in the Flowering Plant Genus Polemonium (Polemoniaceae) Suggests Widespread Historical Gene Flow Despite Limited Nuclear Signal. Syst Biol. 2021;70(1):162–80.
80. Terra V, Ringelberg JJ, Maslin B, Koenen EJM, Ebinger J, Seigler D et al. Dilemmas in generic delimitation of Senegalia and allies (Caesalpinioideae, mimosoid clade): how to reconcile phylogenomic evidence with morphology and taxonomy? PhytoKeys. 2022;205:261–78.
81. Ringelberg JJ, Koenen EJM, Iganci JR, de Queiroz LP, Murphy DJ, Gaudeul M, et al. Phylogenomic analysis of 997 nuclear genes reveals the need for extensive generic re-delimitation in Caesalpinioideae (Leguminosae). PhytoKeys. 2022;205:3–58.
82. Pasiecznik N, Harris P, Smith S. Identifying Tropical Prosopis Species: A Field Guide. Coventry, UK: International Research Department; 2004.
83. El-ahmir S, hoob, Mahklouf M, Azzu Y. The First Record of A medicinal Plant Species Prosopis farcta (, Banks, Sol JF, Macbr. (Fabaceae; Mimosoideae) from Libya. 2021;8:39–43.
84. Baibout M, Corcket E, Kothari SL, Fievet V. Ecosystem services provided by Prosopis cineraria (L.) Druce in the drylands of Southern and Western Asia. Bot Lett. 2022;169(1):30–42.
85. Soltis DE, Kuzoff RK. Discordance Between Nuclear and Chloroplast Phylogenies in the Heuchera Group (saxifragaceae). Evolution. 1995;49(4):727–42.
86. Van Raamsdonk L, Smiech MP, Sandbrink JM. Introgression explains incongruence between nuclear and chloroplast DNA-based phylogenies in Allium section Cepa. Bot J Linn Soc. 1997;123(2):91–108.
87. Roelofs D, Bachmann K. Comparison of chloroplast and nuclear phylogeny in the autogamous annualMicroseris douglasii (Asteraceae: Lactuceae). Plant Syst Evol. 1997;204(1):49–63.
88. Nishimoto Y, Ohnishi O, Hasegawa M. Topological incongruence between nuclear and chloroplast DNA trees suggesting hybridization in the urophyllum group of the genus Fagopyrum (Polygonaceae). Genes Genet Syst. 2003;78(2):139–53.
89. Renoult JP, Kjellberg F, Grout C, Santoni S, Khadari B. Cyto-nuclear discordance in the phylogeny of Ficus section Galoglychia and host shifts in plant-pollinator associations. BMC Evol Biol. 2009;9(1):248.

90. Yu WB, Huang PH, Li DZ, Wang H. Incongruence between Nuclear and Chloroplast DNA Phylogenies in Pedicularis Section Cyathophora (Orobanchaceae). PLoS ONE. 2013;8(9):e74828.
91. Vega MV, Hernández P. Molecular evidence for natural interspecific hybridization in Prosopis. Agrofor Syst. 2005;64(3):197–202.
92. Fontana ML, Pérez VR, Luna CV. Características evolutivas en Prosopis spp.: citogenética, genética e hibridaciones. Rodriguésia. 2018;69(2):409–21.
93. Nguyen HQ, Nguyen TNL, Doan TN, Nguyen TTN, Phạm MH, Le TL, et al. Complete chloroplast genome of novel Adinandra megaphylla Hu species: molecular structure, comparative and phylogenetic analysis. Sci Rep. 2021;11(1):11731.
94. Yang L, Abduraimov O, Tojibaev K, Shomurodov K, Zhang YM, Li WJ. Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of Ferula L. BMC Genomics. 2022;23(1):643.
95. Fan ZF, Ma CL. Comparative chloroplast genome and phylogenetic analyses of Chinese Polyspora. Sci Rep. 2022;12(1):15984.
96. Turudić A, Liber Z, Grdiša M, Jakše J, Varga F, Šatović Z. Chloroplast Genome Annotation Tools: Prolegomena to the Identification of Inverted Repeats. Int J Mol Sci. 2022;23(18):10804.
97. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982;29(2):537–50.
98. Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. Recent Acceleration of Plastid Sequence and Structural Evolution Coincides with Extreme Mitochondrial Divergence in the Angiosperm Genus Silene. Genome Biol Evol. 2012;4(3):294–306.
99. Turmel M, Otis C, Lemieux C. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. Sci Rep. 2017;7(1):994.
100. Williams AM, Friso G, van Wijk KJ, Sloan DB. Extreme variation in rates of evolution in the plastid Clp protease complex. Plant J. 2019;98(2):243–59.
101. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, et al. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015;5(1):8348.
102. Amar MH. ycf1-ndhF genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in Prunus persica. J Genet Eng Biotechnol. 2020;18(1):42.
103. Asaf S, Waqas M, Khan AL, Khan MA, Kang SM, Imran QM, et al. The Complete Chloroplast Genome of Wild Rice (Oryza minuta) and Its Comparison to Related Species. Front Plant Sci. 2017;8:304.
104. Zhang F, Wang T, Shu X, Wang N, Zhuang W, Wang Z. Complete Chloroplast Genomes and Comparative Analyses of L. chinensis, L. anhuiensis, and L. aurea (Amaryllidaceae). Int J Mol Sci. 2020;21(16):5729.
105. Zhang XF, Landis JB, Wang HX, Zhu ZX, Wang HF. Comparative analysis of chloroplast genome structure and molecular dating in Myrtales. BMC Plant Biol. 2021;21(1):219.
106. Doyle J. DNA Protocols for Plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. Molecular Techniques in Taxonomy. Berlin, Alemania: Springer; 1991. pp. 283–93.
107. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
108. Jin JJ, Yu WB, Yang JB, Song Y, de Pamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
109. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
110. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.

111. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19(5):455–77.
112. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.
113. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
114. Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol Clifton NJ. 2019;1962:1–14.
115. Xia X. DAMBE5: A Comprehensive Software Package for Data Analysis in Molecular Biology and Evolution. Mol Biol Evol. 2013;30(7):1720–8.
116. Mayer C, Phobos. Highly accurate search for perfect and imperfect tandem repeats in complete genomes by Christoph Mayer. Ruhr Universität Bochum. 2007. https://www.ruhr-uni-bochum.de/spezzoo/cm/cm_phobos.htm. Accessed 15 Oct 2022.
117. Michael D, Gurusaran M, Santhosh R, Hussain MdK, Satheesh SN, Suhan S, et al. RepEx: A web server to extract sequence repeats from protein and DNA sequences. Comput Biol Chem. 2019;78:424–30.
118. Alamro H, Alzamel M, Iliopoulos CS, Pissis SP, Watts S. IUPACpal: efficient identification of inverted repeats in IUPAC-encoded DNA sequences. BMC Bioinformatics. 2021;22(1):51.
119. Kurtz S. The Vmatch large scale sequence analysis software A Manual. University of Hamburg. 2010; Available from: http://www.vmatch.de/virtman.pdf.
120. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80.
121. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2014;10(4):e1003537.
122. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
123. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9(8):772.
124. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.
125. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22(22):4673–80.
126. Hall TA, BioEdit:. A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;(41):95–8.
127. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Battistuzzi FU, editor. Mol Biol Evol. 2018;35(6):1547–9.

Figures

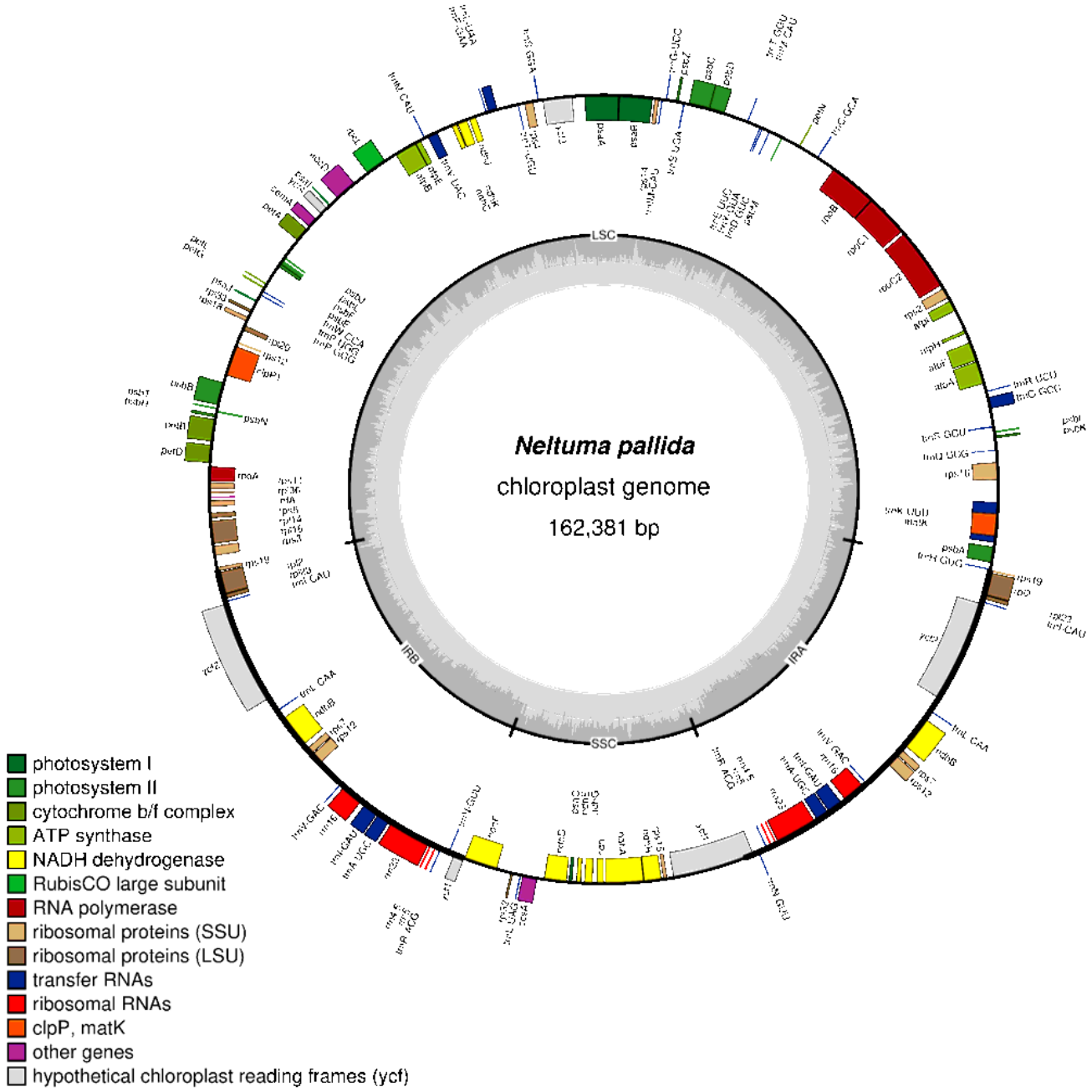


Figure 1

Genome map of *Neltuma pallida* chloroplast genome. It shows its four genomic regions (LSC, SSC, IRA and IRB). Genes located in the inside of the circle are transcribed clockwise, while those located in the outside are transcribed on the opposite direction. Genes are grouped according to their functional group by color codes. The inner circle exhibits de GC content (dark gray) and the AT content (light gray).

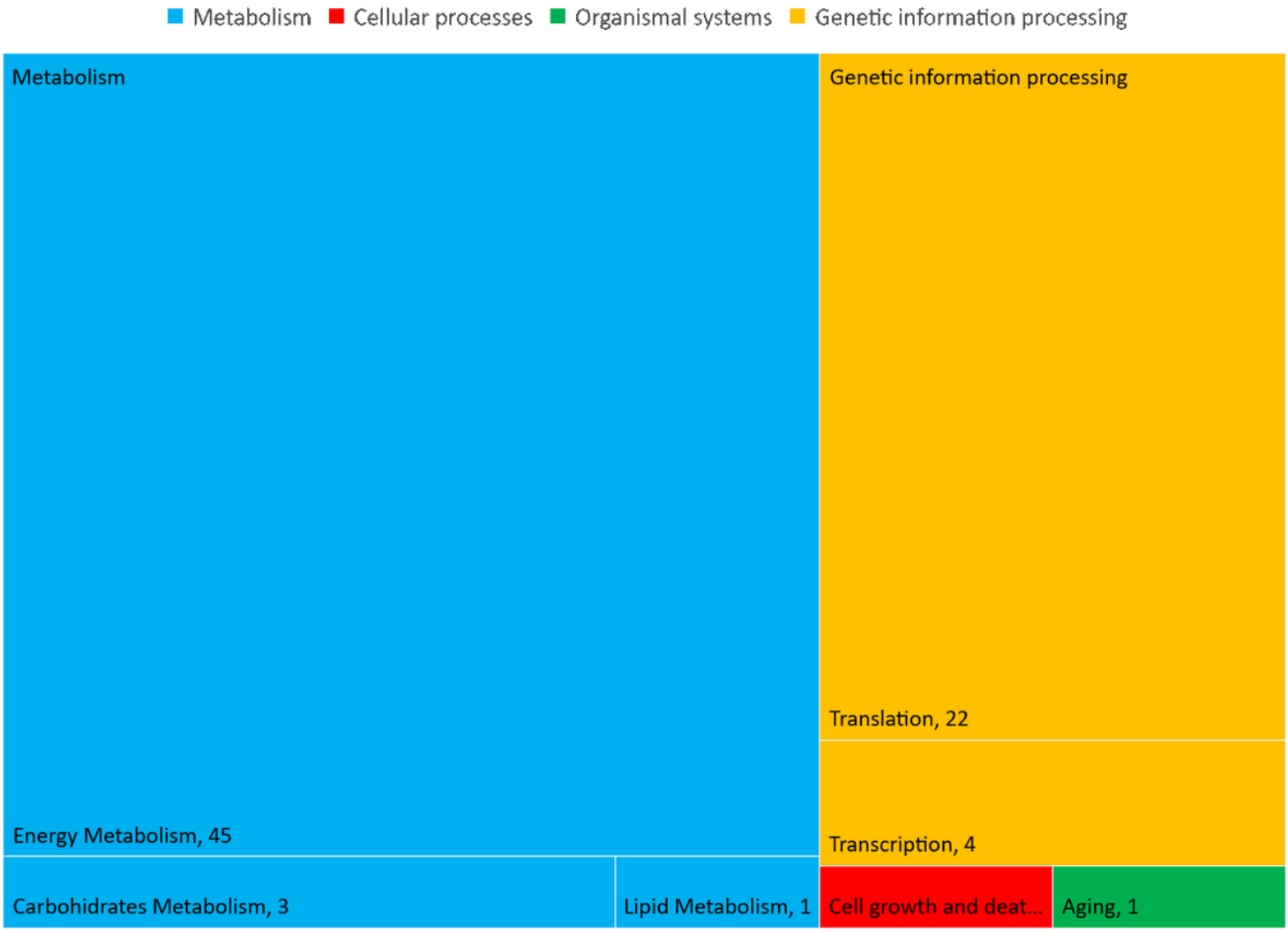


Figure 2

KEGG pathway categories of *N. pallida* chloroplast genome protein-coding genes. The colors are used to assign the primary classification; the secondary classification is named within the diagram. The number of genes belonging to each classification is shown next to the secondary classification.

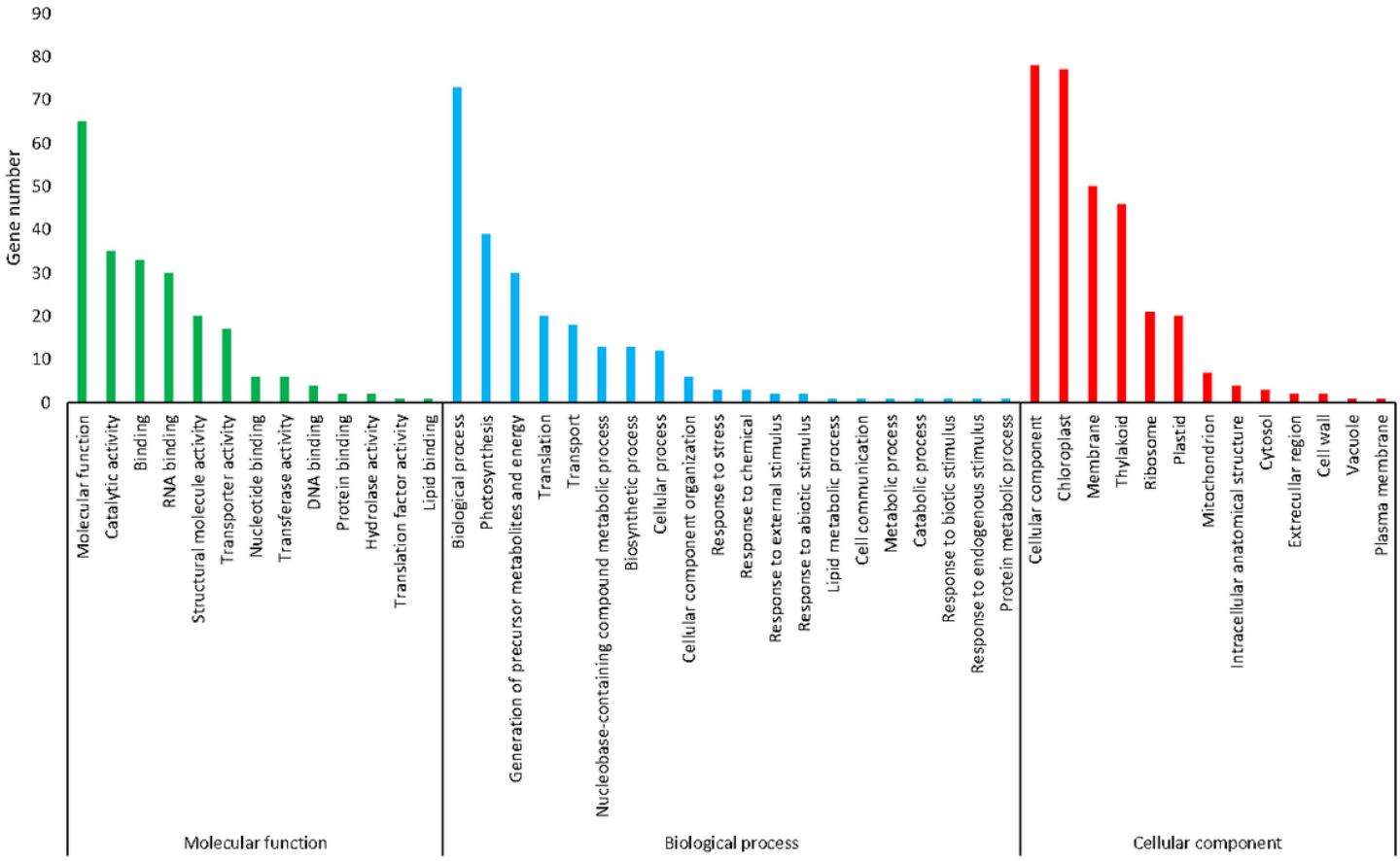


Figure 3

Histogram of *N. pallida* chloroplast genes distribution in Gene Ontology (GO) terms.

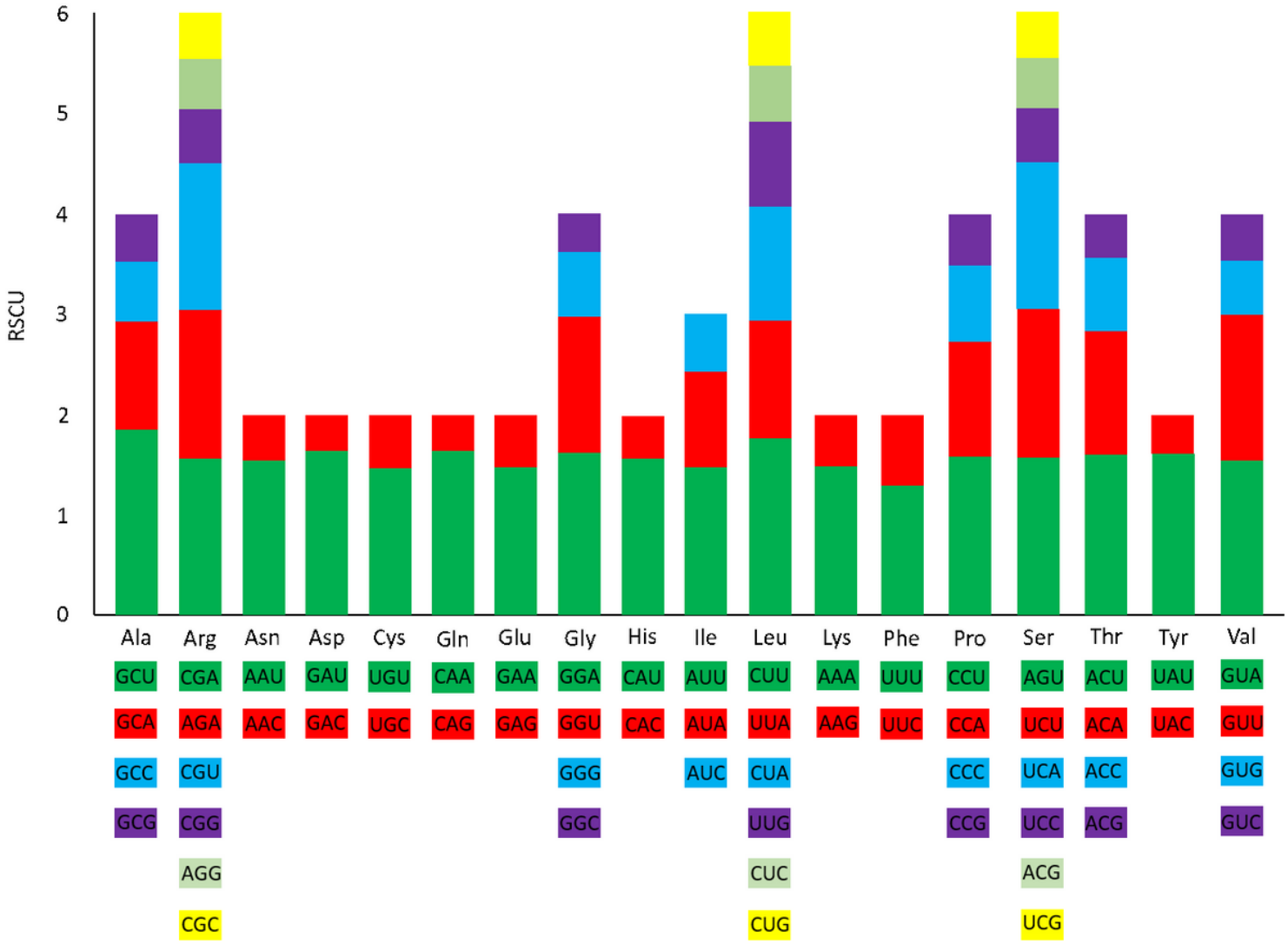


Figure 4

RSCU values of amino acids in 85 protein-coding genes of the *Neltuma pallida* chloroplast genome.

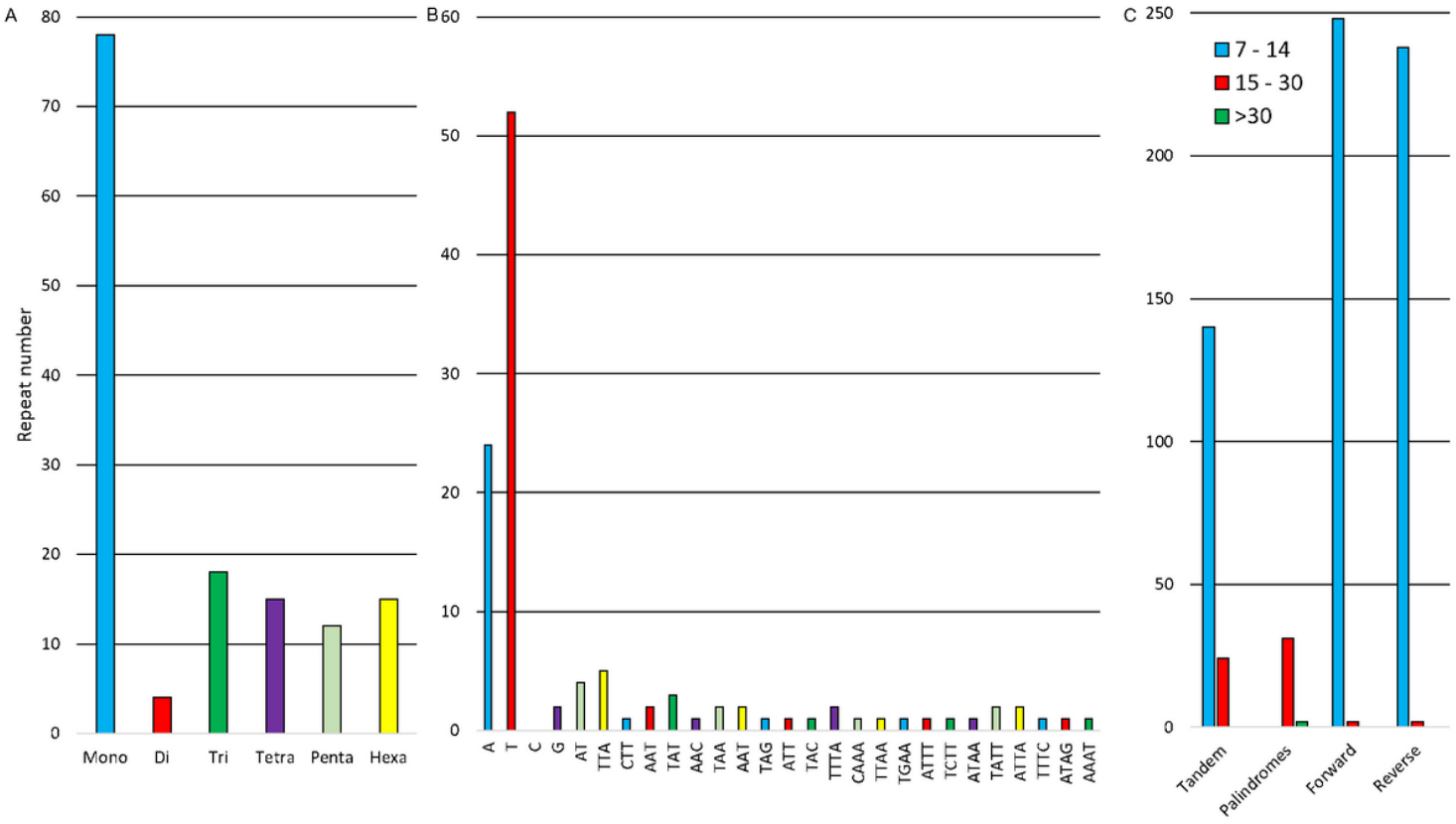


Figure 5

Repetitive sequences frequency in the *Neltuma pallida* chloroplast genome. (A) Total SSR frequency. (B) SSR frequency according to its repeat motif. (C) Tandem, palindrome, forward and reverse repeats frequency

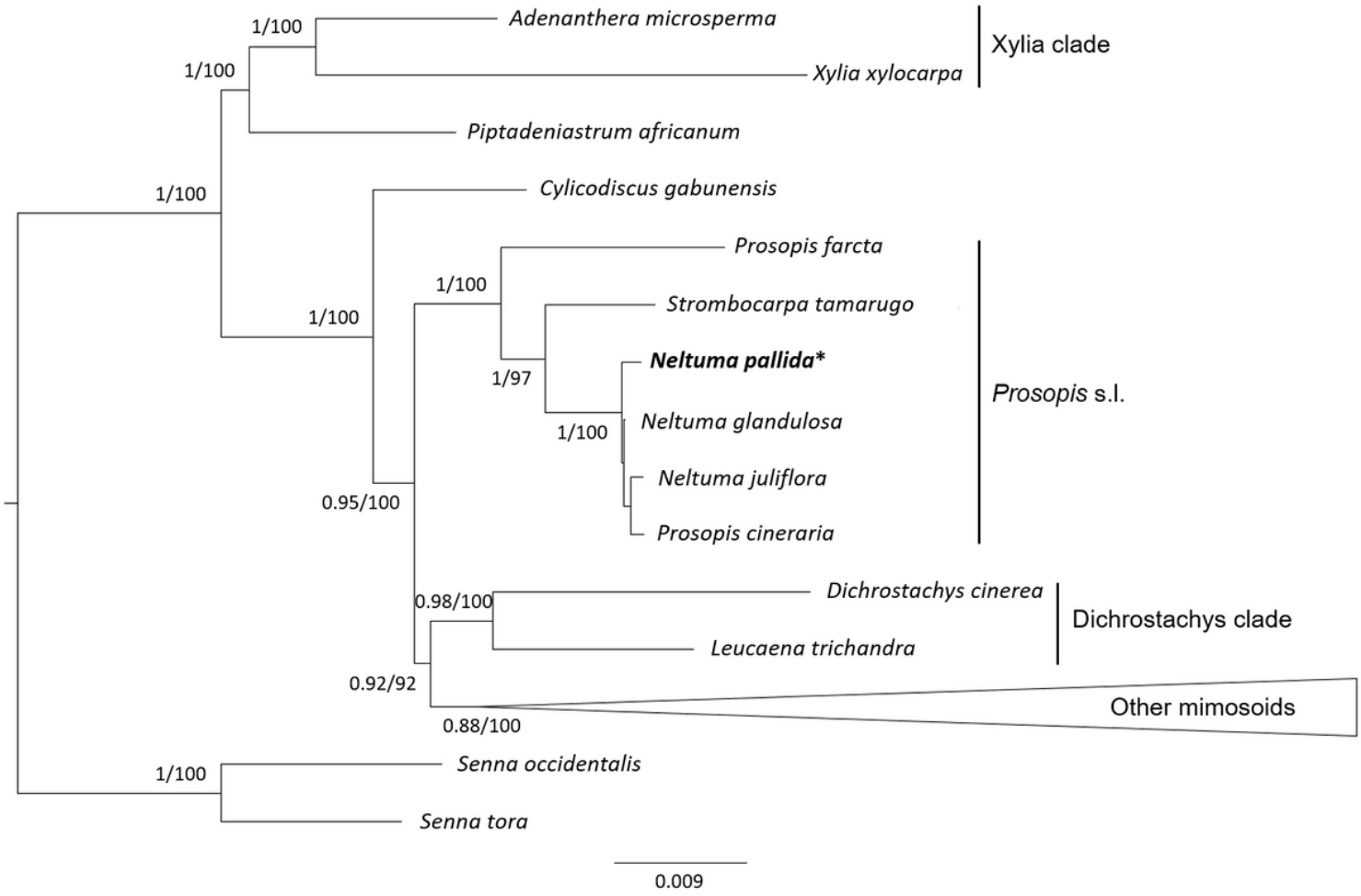


Figure 6

Phylogenomic tree of *N. pallida* within subfamily Caesalpinioideae. The whole chloroplast genomes were used for this reconstruction. The methods employed were Bayesian Inference and Maximum Likelihood, and their statistics values are represented by the numbers on the left of each node. *N. pallida* position in the tree is highlighted with an asterisk (*).

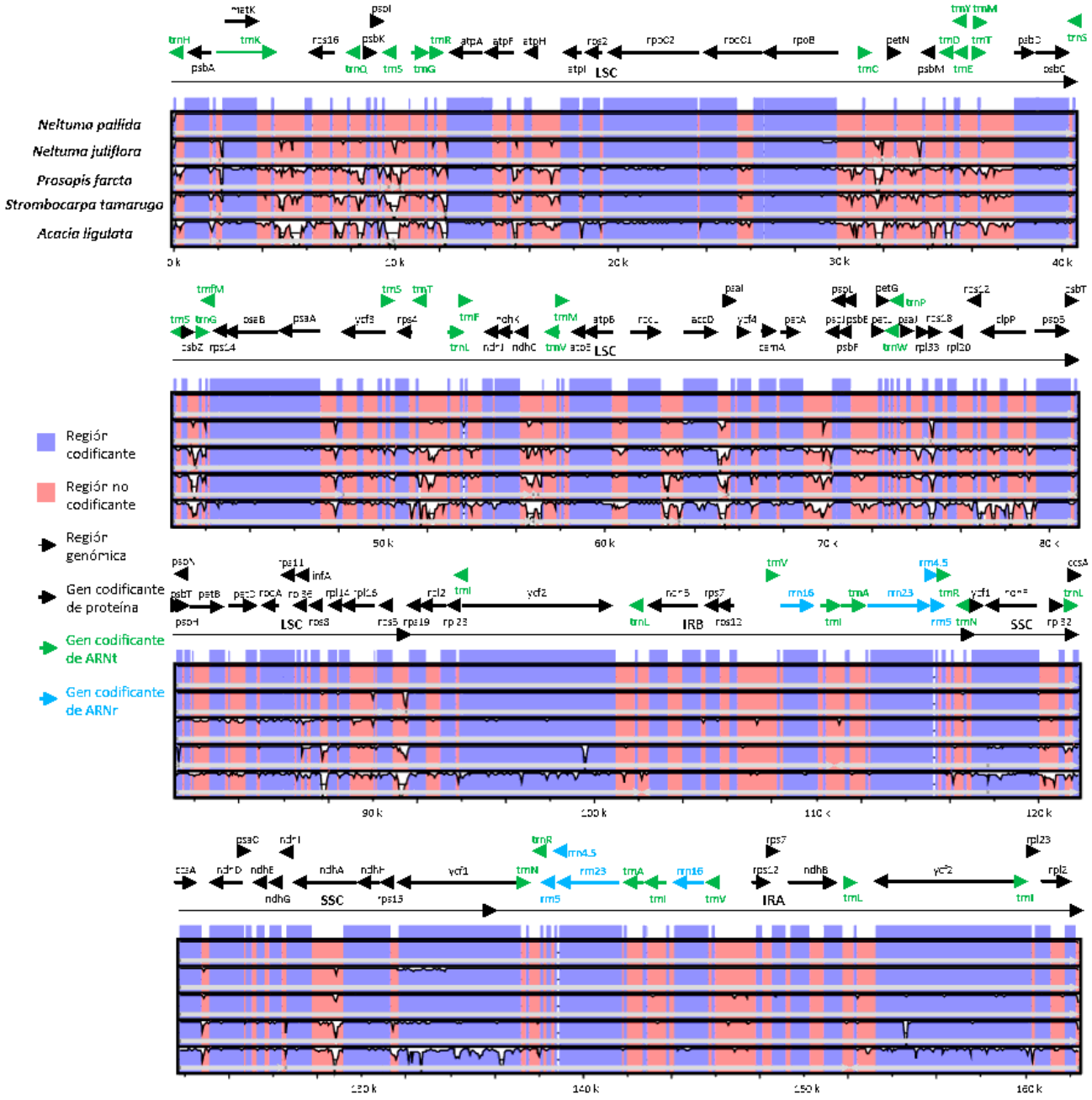


Figure 7

Graphical representation of the alignment of the chloroplast genomes of *N. pallida* and related species. *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata* are used for comparison. The graph shows the level of identity (50% - 100%, Y-axis) along the genomes (X-axis) taking as reference the chloroplast genome of *Neltuma pallida*.

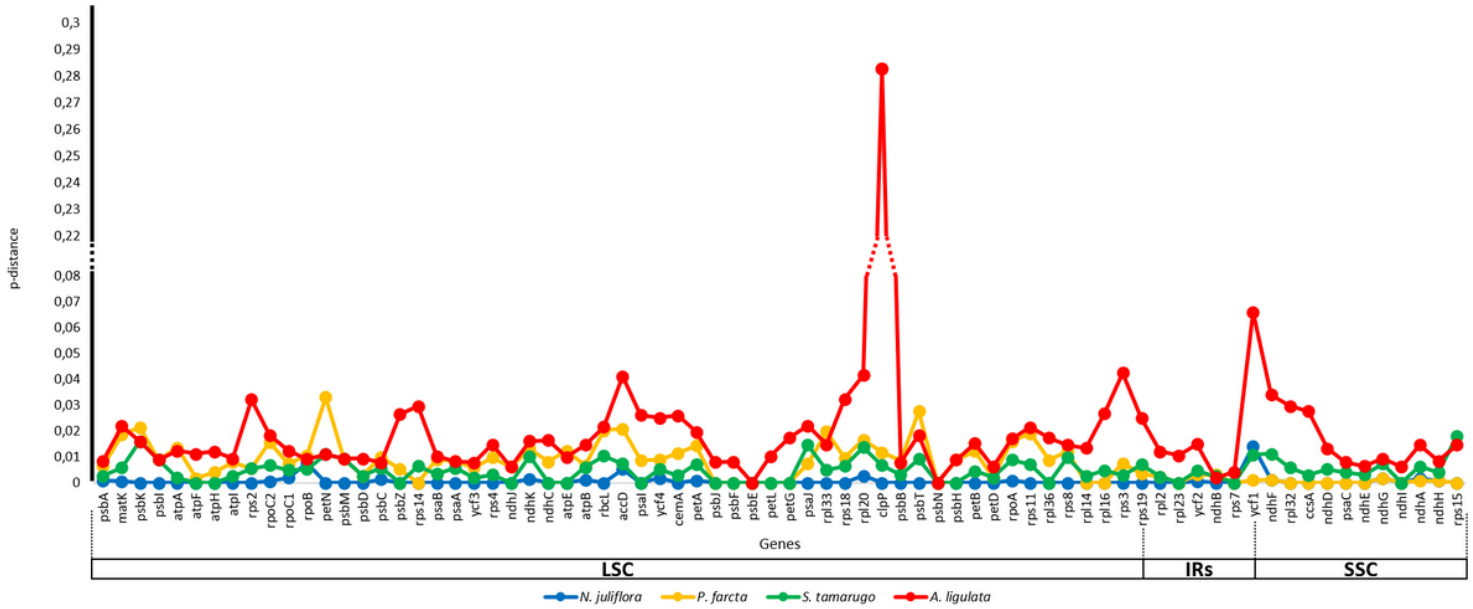


Figure 8

Pairwise distance of 74 protein-coding genes from *N. pallida* chloroplast genome with homologs. The genes homologs were obtained from *N. juliflora*, *P. farcta*, *S. tamarugo* and *A. ligulata* chloroplast genomes.

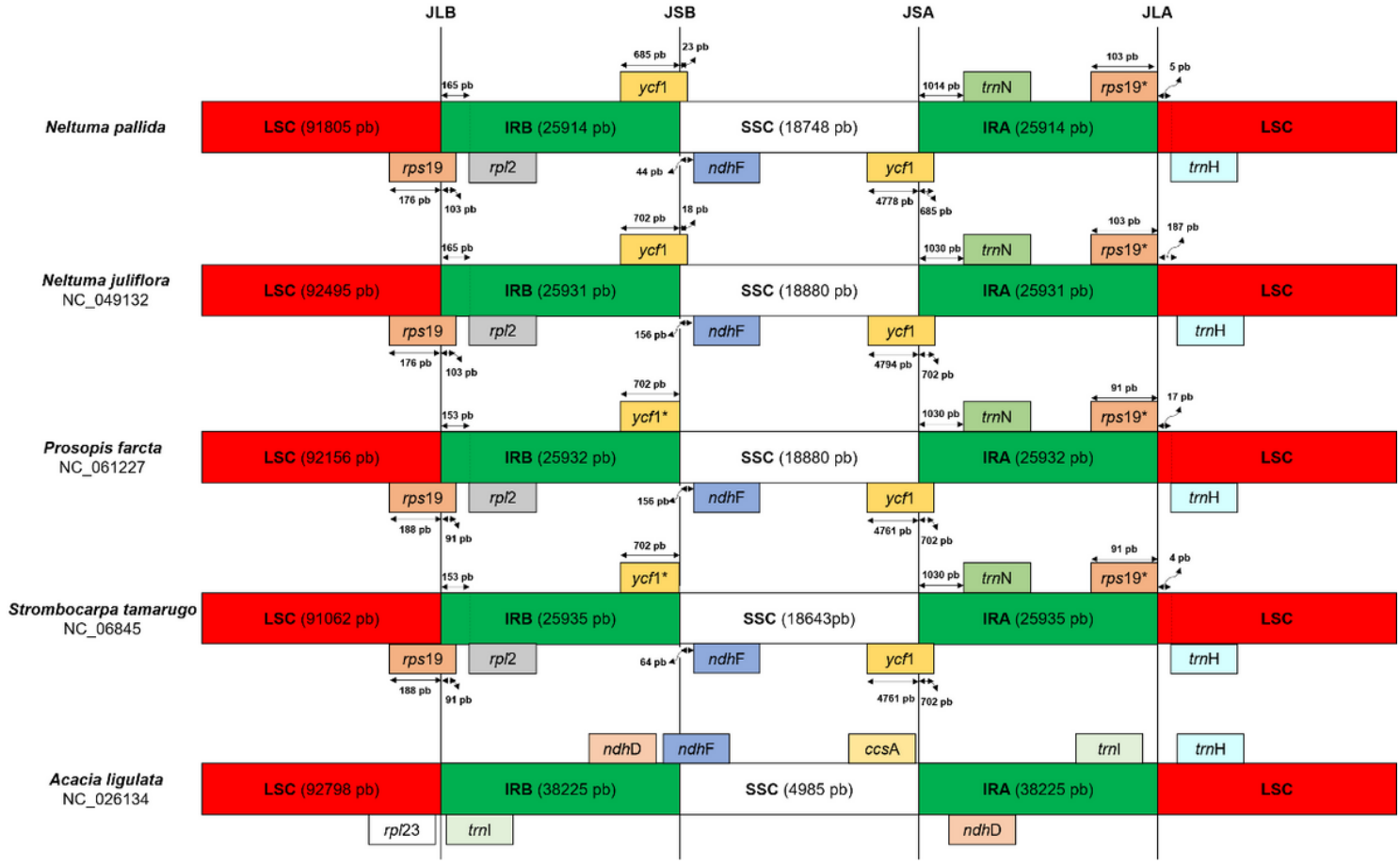


Figure 9

Comparison of boundaries between the regions of the *N. pallida* chloroplast genome with related species. The species used for this comparison were *N. juliflora*, *P. farcta*, *S. tamarugo*and *A. ligulata*. The genomes are represented as split bars in

each region. The boxes above and below the bars are representations of the genes. The arrows indicate the distance in bp between the ends of the genes with the boundaries closest to this one. These representations are not proportional to sequence lengths. rps19* (non-coding) found in IRA is a portion of the complete rps19 gene found in IRB.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTablesAssemblyannotationandanalysisofthechloroplastgenomeoftheAlgarrobotreeNeltumapallida.xlsx
- SupplementaryfiguresAssemblyannotationandanalysisofthechloroplastgenomeoftheAlgarrobotreeNeltumapallida.docx