


## RESEARCH ARTICLE

# Complete Chloroplast Genomes of *Vachellia nilotica* and *Senegalia senegal*: Comparative Genomics and Phylogenomic Placement in a New Generic System

Sajjad Asaf<sup>1</sup> , Arif Khan<sup>1,2</sup> , Abdul Latif Khan<sup>1\*</sup>, Ahmed Al-Harrasi<sup>1\*</sup>, Ahmed Al-Rawahi<sup>1</sup>

**1** Natural and Medical Sciences Research Center, University of Nizwa, Nizwa, Oman, **2** Genomics Group, Faculty of Biosciences and Aquaculture, Nord University, Bodo, Norway

 These authors contributed equally to this work.

\* [latifepm78@yahoo.co.uk](mailto:latifepm78@yahoo.co.uk) (ALK); [aharrasi@unizwa.edu.om](mailto:aharrasi@unizwa.edu.om) (AAH)



## OPEN ACCESS

**Citation:** Asaf S, Khan A, Khan AL, Al-Harrasi A, Al-Rawahi A (2019) Complete Chloroplast Genomes of *Vachellia nilotica* and *Senegalia senegal*: Comparative Genomics and Phylogenomic Placement in a New Generic System. PLoS ONE 14 (11): e0225469. <https://doi.org/10.1371/journal.pone.0225469>

**Editor:** Kandasamy Ulaganathan, Osmania University, INDIA

**Received:** April 22, 2019

**Accepted:** November 5, 2019

**Published:** November 25, 2019

**Copyright:** © 2019 Asaf et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The assembled and annotated complete chloroplast genome were submitted to NCBI under the accession number MK645904 (*V. nilotica*) and MK645903 (*S. senegal*).

**Funding:** The authors are thankful to The Research Council, Oman for their financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

*Vachellia* and *Senegalia* are the most important genera in the subfamily Mimosoideae (Fabaceae). Recently, species from both genera were separated from the long-characterized *Acacia* due to their macro-morphological characteristics. However, this morpho-taxonomic differentiation struggles to discriminate some species, for example, *Vachellia nilotica* and *Senegalia senegal*. Therefore, sequencing the chloroplast (cp) genomes of these species and determining their phylogenetic placement via conserved genes may help to validate the taxonomy. Hence, we sequenced the cp genomes of *V. nilotica* and *S. senegal*, and the results showed that the sizes of the genomes are 165.3 and 162.7 kb, respectively. The cp genomes of both species comprised large single-copy regions (93,849–91,791 bp) and pairs of inverted repeats (IR; 26,093–26,008 bp). The total numbers of genes found in the *V. nilotica* and *S. senegal* cp genomes were 135 and 132, respectively. Approximately 123:130 repeats and 290:281 simple sequence repeats were found in the *S. senegal* and *V. nilotica* cp genomes, respectively. Genomic characterization was undertaken by comparing these genomes with those of 17 species belonging to related genera in Fabaceae. A phylogenetic analysis of the whole genome dataset and 56 shared genes was undertaken by generating cladograms with the same topologies and placing both species in a new generic system. These results support the likelihood of identifying segregate genera from *Acacia* with phylogenomic disposition of both *V. nilotica* and *S. senegal* in the subfamily Mimosoideae. The current study is the first to obtain complete genomic information on both species and may help to elucidate the genome architecture of these species and evaluate the genetic diversity among species.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

*Senegalia senegal* (L.) Britton and *Vachellia nilotica* (L.) P.J.H. Hurter & Mabb are the most important species of the genera *Senegalia* and *Vachellia*, which belong to the family Fabaceae [1]. *S. senegal* was formerly known as *Acacia senegal* (L.) Wild, and *V. nilotica* was known as *Acacia nilotica* [2]. Both species were placed in different genera due to their morphological and taxonomical differences. *S. senegal* is a deciduous tree native to arid and semi-desert regions of sub-Saharan Africa but can also be found in other parts of the world, such as the Indian sub-continent and the Arabian peninsula [3]. The genera are well-known for their exudate gum arabic, a non-timber forest product in international trade possessing medicinal, ecological and commercial importance [3]. The gum derived from the tree is used in such industries as food, pharmaceutical and cosmetics [4]. Moreover, this gum is also used in lithographic ink due to its unique emulsification, encapsulation and film-forming properties, adding to the commercial importance of these species [5, 6]. Furthermore, *S. senegal* has been noted to increase soil fertility through efficient nutrient fixation, whereas the tree provides shade, fodder, wood fuel [7]. In terms of medicinal uses, gum and tree parts have been known to play bioactive roles in cancer, inflammation, oxidative stress and abdominal complications [7, 8].

In a similar vein, *V. nilotica*, a multipurpose legume tree and drought-resistant species, has been well-regarded as a means of rehabilitating dry land ecosystems [9]. This tree increases soil organic carbon, total and available forms of nitrogen and phosphorus under its canopy and can thus be used in soil amelioration [10]. Nitrogenous fertilizers are highly expensive for large-scale afforestation [11]. Utilizing alternative species, such as *V. nilotica*, can assist in fixing atmospheric nitrogen to increase soil fertility [5]. The nutrients generated by *V. nilotica* trees through biological nitrogen fixation can be exploited within the production system, either simultaneously as an intercropping plant or sequentially, as in rotational fallow systems [9]. *V. nilotica* has also been well-documented to possess essential chemical constituents that have been suggested to play roles in fighting cancer, microbial pathogenesis, inflammation, sexually transmitted diseases, oxidative stress, diabetes and mutagenesis [12, 13]. Despite the strong medicinal and local uses of both *V. nilotica* and *S. senegal*, the taxonomy of these species has not been elucidated. These two species were formerly placed in the genus *Acacia*, despite their major variation from the other species of *Acacia* [14]. The genus *Acacia* comprised 1350 species distributed in most of the continents, except Antarctica [13]. The 2011 IBC (International Botanical Congress) meeting in Melbourne finally ratified the previous decision, despite the long-standing controversy, paving the way for name changes to *Vachellia* for a smaller and pan-tropical group [15]. This meeting suggested the use of the genera *Senegalia* and *Vachellia* in the classification of *S. senegal* and *V. nilotica*, respectively. Morphological, biochemical, and palynological data are highly important for the classification of plants into their respective genera [16]. However, emphasis has been placed exegetically to further understand and create more genomic datasets to elucidate these difficult-to-classify and important species [17].

In this regard, chloroplast, the most important organelle in plant cells, plays an important role in photosynthesis, carbon fixation, fatty and amino acid synthesis [18, 19] and has been a focus of attention in recent decades to understand taxonomy, evolution and biological processes. Ideally, a chloroplast (cp) genome of angiosperms exhibits a quadripartite structure size ranging from 110 kb -160 kb. The quadripartite structure is usually composed of a large single copy (LSC) region, a small single copy region (SSC) region and a pair of inverted repeats (IR), which are mirror images of each other [19]. Angiosperm cp genomes generally contain 80 protein-coding genes, 4 ribosomal RNA (rRNA) genes, and 30 transfer RNA (tRNA) genes [20]. The majority of cp genomes exhibit highly conserved structures, some reveal structural

variations, IR loss, and gene loss as a result of adaptation to their environments [21, 22]. Next-generation technologies have allowed the rapid sequencing of many cp genomes in recent years [23]. These abundant cp genomes have facilitated the verification of evolutionary relationships and allowed detailed phylogenetic classifications to be conducted at the group, family, and even generic level in plants [24, 25]. Furthermore, cp genomes can be used for species identification through DNA barcoding and molecular markers that enable morphologically similar species to be distinguished [26]. Despite the highly economic, biological, ecological and social importance of these genera, very little information is available on the comparative chloroplast genomes of *Senegalia* and *Vachellia*. It is difficult to demarcate monophyletic lineages within these genera, despite morphological differences, and they face classification issues [27, 28]. In this study, we sequenced the chloroplast genomes of *V. nilotica* and *S. senegal*, and complete phylogenomic analysis was performed to validate their placement in the genera *Vachellia* and *Senegalia*, respectively. Our study provides sequence resources for future studies of population diversity and taxonomy.

## Materials and methods

### Chloroplast DNA extraction and sequencing

Young and immature green fresh photosynthetic leaves of *V. nilotica* and *S. senegal* were ground to fine powder in liquid nitrogen, and the contamination-free chloroplast DNA was isolated according to the modified protocol of Shi et al., [29]. The Ion Torrent sequencing platform was used for sequencing intact chloroplast DNA using the Ion torrent S5 sequencer with the Ion Torrent server (Life Technologies, USA). Genomic libraries were prepared according to the manufacturer's instructions (Life Technologies, USA). The total chloroplast DNA of each sample was sheared enzymatically into approximately 400-bp fragments using the Ion Shear Plus Reagents kit, and libraries were prepared using the Ion Xpress Plus gDNA Fragment Library kit. Prepared libraries were quantified and qualified on a Qubit 3.0 fluorimeter and an Agilent 2100 Bioanalyzer system. Library preparations were followed by template amplification (Ion one touch 2 instrument, Life Technologies, USA), and enrichment of the amplified template was performed (Ion OneTouch™ ES enrichment system, Life Technologies, USA) using Ion 520 and 530 OT2 reagents. The prepared libraries were loaded onto the Ion S5 sequencing chip, and sequencing was performed according to the Ion torrent S5 protocol (Life Technologies USA).

### Genome assembly

The sequencing of *V. nilotica* and *S. senegal* resulted in 185,114 and 137,673 reads, respectively. The obtained reads of both *Vachellia* and *Senegalia* species were mapped to the selected reference genome of *Vachellia flava* and *Senegalia laeta* using Bowtie ((v.2.2.3) [30] in Geneious Pro (v.10.2.3) [31] software. The mean coverage of the reads for *V. nilotica* and *S. senegal* were 134X and 168X, respectively. The IR (inverted repeat) junction regions were selected from the reference genomes to adjust the sequence length, and the iteration method was used with MITObim (v.1.8) [32].

### Genome annotation

Chloroplast genomes of the sequenced species were annotated by using Dual Organellar Genome Annotator (DOGMA)[33], and the number and position of ribosomal RNAs, transfer RNAs and coding genes present in chloroplast genomes were identified and analyzed using BLASTX and BLASTN, and tRNAscan-SE version 1.21 [34] software was used to annotate

tRNA genes. Additionally, for manual adjustment, Geneious (v11.0) and tRNAscan-SE [34] were used to compare the genome with previously reported reference genomes. Correspondingly, the start and stop codons and intron boundaries were also manually adjusted compared with the pre-published reference cp genome. In addition, the structural features of the cp genomes of both *V. nilotica* and *S. senegal* species were illustrated using OGDRAW [35]. Correspondingly, the MEGA6 software [36] was used to determine the relative synonymous codon usage and deviations in synonymous codon usage by avoiding the influence of amino acid composition. The divergence of *V. nilotica* and *S. senegal* species taxa genomes from those of other related species (Fig 5) was determined using mVISTA [37] in Shuffle-LAGAN mode and using *V. nilotica* and *S. senegal* as reference genomes.

### Repeat identification

REPuter software [38] was used for the identification of palindromic, forward and tandem repeats present in the genome. The criterion was a minimum >15 base pairs with a sequence identity of 90%. Furthermore, SSRs were determined using Phobos version 3.3.12 [39] with the search parameters set for mononucleotide repeats  $\geq 10$  repeat units, dinucleotide repeats  $\geq 8$  repeat units, tri- and tetranucleotide repeats  $\geq 4$  repeat units, and pentanucleotide and hexanucleotide repeats  $\geq 3$  repeat units. Tandem Repeats Finder version 4.07 b [40] with default settings was used to determine tandem repeats.

### Sequence divergence and phylogenetic analysis

The average pairwise sequence divergence of the complete cp genomes of *Vachellia* and *Senegalia* species with related species was determined. Comparative sequence analysis after comparing gene order and performing multiple sequence alignment was used to identify missing and ambiguous gene annotations. MAFFT version 7.222 [41], with default parameters was used for the alignment of complete genomes, and pairwise sequence divergence was calculated by selected Kimura's two-parameter (K2P) model [42]. To resolve the phylogenetic position of *V. nilotica* and *S. senegal* within the family Fabaceae, cp genomes were downloaded from the NCBI database. Alignment of the complete cp genomes was constructed on the basis of conserved gene order and structure of the cp genome. Four methods were used to infer the phylogenetic trees, including maximum parsimony (MP) implemented with PAUP 4.0100, neighbour-joining (NJ) and maximum likelihood (ML) with MEGA 6 [36] and Bayesian inference (BI) with MrBayes 3.1.299 [43] using setting derived from Asaf et al [44] and Wu et al [45]. ML analysis parameters were adjusted with a BIONJ tree with 1000 bootstrap replicates using the Kimura 2-parameter model with gamma-distributed rate heterogeneity and invariant sites. A heuristic search for MP analysis was run with 1000 random addition sequence replicates with the tree-bisection-reconnection (TBR) branch-swapping tree search criterion. The best substitution model GTR + G model was used according to the Akaike information criterion (AIC) by jModelTest version 2.1.10 for Bayesian posterior probabilities (PP) in the BI analyses. The Markov Chain Monte Carlo (MCMC) was run with 4 incrementally heated chains for 1,000,000 generations, starting from random trees and sampling 1 out of every 100 generations. The first 25% of trees were discarded as burn-in to estimate the value of posterior probabilities. In another phylogenetic study, 65 shared genes from the cp genomes of the 102 Fabaceae members downloaded from NCBI were aligned in MAFFT version 7.222 [38]. The above four phylogenetic-inference methods were used to infer trees from these 65 concatenated genes using the same settings described above. The assembled and annotated complete chloroplast genome was submitted to NCBI under the accession numbers MK645904 (*V. nilotica*) and MK645903 (*S. senegal*).

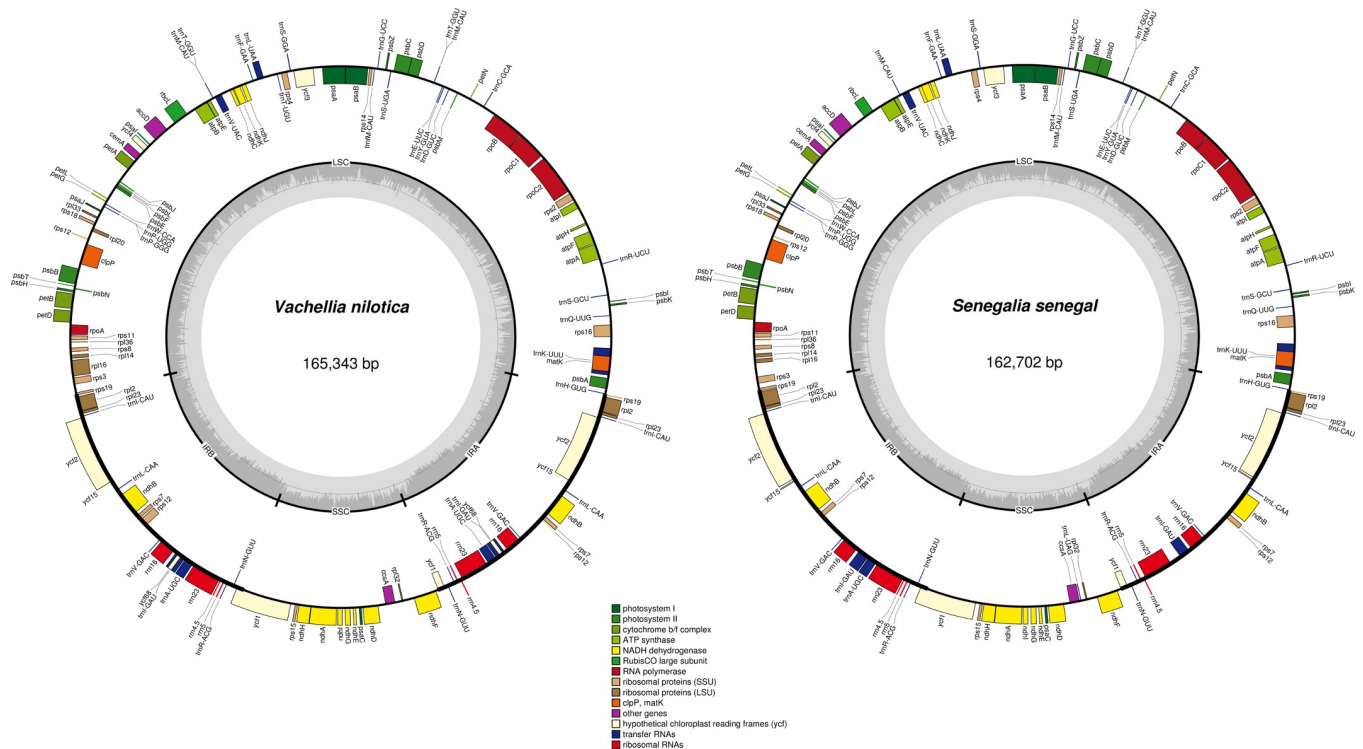
## Results

### General features of *V. nilotica* and *S. senegal* chloroplast genomes

The complete chloroplast genomes of *V. nilotica* and *S. Senegal* exhibited typical sizes of 165,343 bp and 162,702 bp, respectively. These genomes showed a typical quadripartite structure with a large single copy region (LSC) and a small single copy region (SSC) and a pair of inverted repeats (Fig 1). The completely sequenced genomes of *V. nilotica* and *S. Senegal* were compared with seventeen other chloroplast genomes, where the results showed that the sizes of compared genomes ranged from 178,887 bp (*Pithecellobium flexicaula*) to 159,389 bp (*Adeanthera micrsperma*). The overall GC content in *V. nilotica* was found (35.4%) to be less than that in *S. senegal* (35.7%). The LSC regions were 39,849 bp and 91,791 bp, while the SSC regions were 19,308 bp and 18,895 bp, respectively, in *V. nilotica* and *S. senegal*. The IR region in the two cp genomes was found to be similar in *V. nilotica* (26,093) and *S. senegal* (26,008). The number of rRNAs (04) in all the sequenced and compared genomes was the same, while the numbers of tRNAs in the genomes were 37 and 38 in *S. senegal* and *V. nilotica*, respectively (Table 1).

### Important genes and base composition in sequenced cp genome

Furthermore, the gene content, gene size and gene order of *V. nilotica* and *S. senegal* were largely similar, comprising 135 and 132 genes, respectively. Among all the compared genomes,



**Fig 1. Genome map of the *Vachellia nilotica* and *Senegalia senegal* chloroplast genome.** Thick lines indicate the extent of the inverted repeat regions (IRa and IRb), which separate the genome into small (SSC) and large (LSC) single copy regions. Genes drawn inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The dark grey in the inner circle corresponds to the GC content, and the light grey corresponds to the AT content.

<https://doi.org/10.1371/journal.pone.0225469.g001>



*P. flexicaula* possessed the highest number of genes (139), and *S. laeta* showed the lowest number of genes (127). The numbers of protein coding genes (PCGs) were 89 and 87 in *V. nilotica* and *S. senegalia*, respectively. However, these numbers were found to be highest (94) in the *P. flexicaula* cp genome. The PCGs in the chloroplast genome include some important genes responsible for photosynthesis, i.e., Photosystem I (*psaA, B, C, I, J*) and Photosystem II (*psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z*). The genes responsible for tRNA, rRNA, large subunit and small subunit of ribosomal proteins were also present in the chloroplast genome. Other important genes annotated in the chloroplast genome were *matK, clpP, cemA, accD, ccsA, ycf1, 3, 4, 15*, which were also present in the chloroplast genome (Table 2). Approximately 22 intron-containing genes were observed in both sequenced genomes (Table 1).

In the complete genome, the composition of (T) is higher than other base nucleotides present in the genome, which is 32.9% and 32.7% in *V. nilotica* and *S. senegal*, respectively. Adenine (A), which comprises the first position in both *V. nilotica* and *S. senegal*, accounts for 34.47 and 30.8, respectively. The (T/U) base at the 2<sup>nd</sup> position was found to be higher than the other genomes, accounting for 33.59 and 32.2 in *V. nilotica* and *S. senegal*, respectively. Similarly, the (T/U) base was also found to be abundant at the 3<sup>rd</sup> position (Table 3).

### Comparison of sequenced genomes with other genomes

Comparison of the currently two sequenced and seventeen other genomes from the database (NCBI) revealed that the *P. flexicaula* (178,887 bp) cp genome was the largest, and that of *A. microsperma* (159,389) was the smallest (Table 1). *V. nilotica* contains the highest number of

Table 1. Composition of *Vachellia nilotica* and *Senegalia senegal* cp genomes with related species.

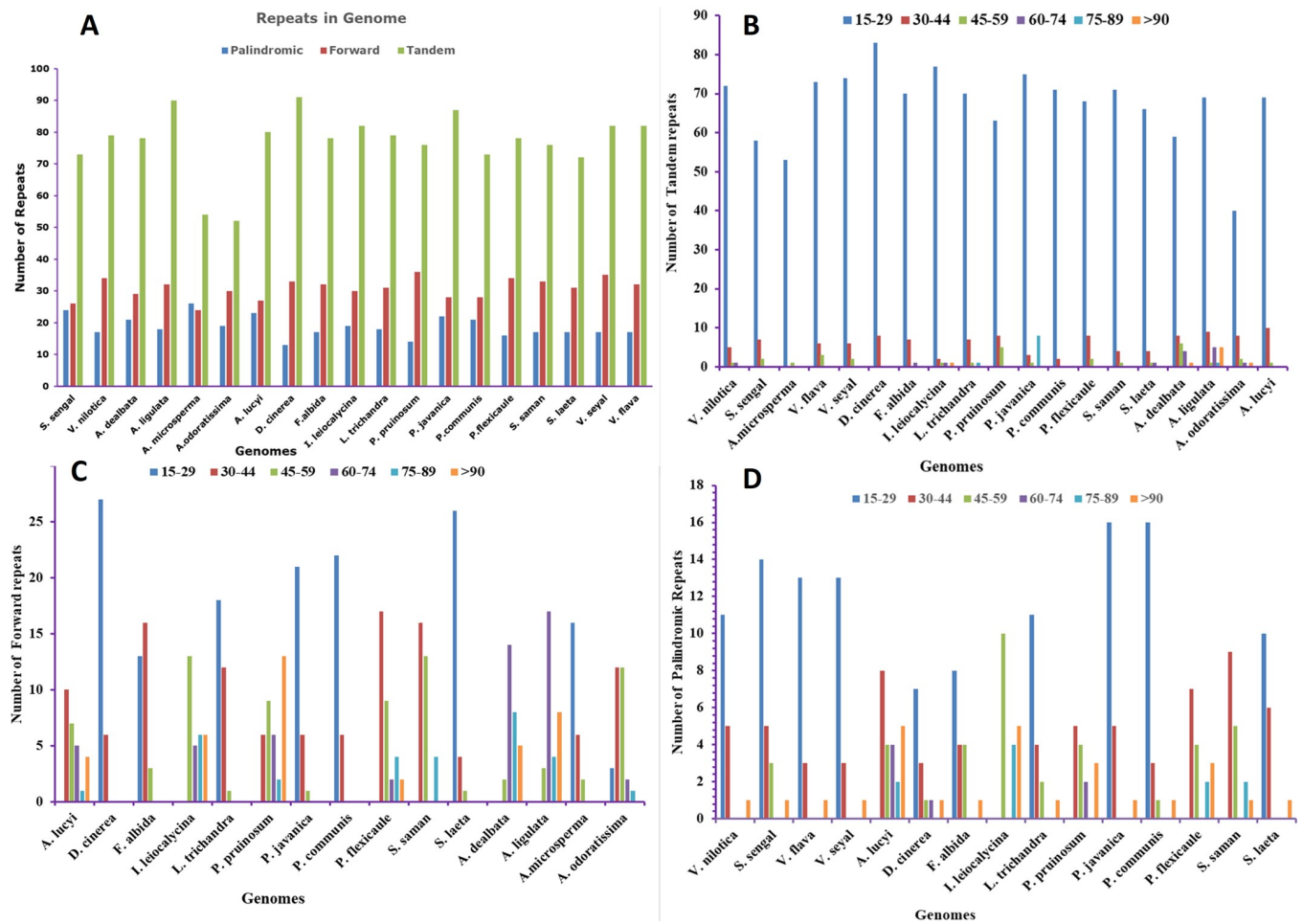
	Size (bp)	Overall GC contents	LSC size in bp	SSC size in bp	IR size in bp	Protein coding regions size in bp	tRNA size in bp	rRNA size in bp	Number of genes	Number of protein coding genes	Number of rRNA	Number of tRNA	Genes with introns
<i>V. nilotica</i>	165,343	35.4	93849	19308	26093	79520	2847	9052	135	89	8	38	22
<i>V. seyal</i>	165,383	35.3	93901	19347	26068	78009	2828	9052	127	82	8	37	21
<i>V. flava</i>	165,829	35.3	94220	19474	26068	78033	2828	9052	127	82	8	37	21
<i>S. Senegal</i>	162702	35.7	91791	18895	26008	75571	2794	9052	132	87	8	37	22
<i>S. Laeta</i>	162754	35.8	91750	18911	26047	77997	2829	9052	127	82	8	37	20
<i>S. saman</i>	176717	35.3	92036	5053	39814	89202	2793	9052	138	92	8	37	21
<i>p. flexicaula</i>	178887	35.1	91076	4805	41503	89784	2793	9078	139	94	8	37	23
<i>P. communis</i>	162,552	35.9	91517	18941	26047	77949	2793	9052	130	83	8	37	23
<i>P. javanica</i>	161,681	35.9	91093	18574	26007	78075	2794	9052	130	83	8	37	23
<i>P. pruinsum</i>	176,692	35.3	92320	5036	39668	89271	2793	9052	138	92	8	37	24
<i>L. trichandra</i>	164692	35.6	93690	18890	26056	78759	2815	9049	129	84	8	37	22
<i>I. leiocalycina</i>	175489	35.5	90987	4948	39777	89244	2820	9056	137	92	8	37	24
<i>F. albida</i>	175646	35.3	91045	4761	39920	88638	2793	9068	138	90	8	37	23
<i>D. cinerea</i>	161240	35.9	90430	18526	26142	77958	2793	9068	130	83	8	37	23
<i>A. lucyi</i>	176870	35.2	92319	4573	39989	89133	2793	9052	138	92	8	37	24
<i>A. odoratissima</i>	174861	35.6	90169	4928	39882	89130	2793	9066	138	92	8	37	24
<i>A. microsperma</i>	159389	36.5	88577	18756	26028	78030	2793	9052	130	83	8	37	23
<i>A. ligulata</i>	174233	35.4	92770	4986	38225	88107	2802	9062	134	89	8	37	22
<i>A. daelbata</i>	174217	35.4	92753	4956	38254	88551	2793	9060	137	91	8	37	22

<https://doi.org/10.1371/journal.pone.0225469.t001>

tRNAs (38) among all the compared genomes. The highest number of genes was found in the *P. flexicaula* (139), and the lowest number was found in 127 genes and was similar in the *V. seyal*, *V. flava* and *S. laeta* chloroplast genomes. The highest number of PCGs (protein coding genes) was observed in *P. flexicaula* (94 genes), and the lowest number (82) was found to be similar in *V. seyal*, *V. flava* and *S. laeta*. The number of rRNAs was similar in all of the compared and sequenced chloroplast genomes, while the number of tRNA- and intron-containing genes varied in all of the chloroplast genomes (Table 1). The largest LSC region was found in *V. flava* (94,220 bp), and the smallest LSC was observed in *A. microsperma* (88,577 bp), which is also the smallest genome. The largest SSC region was found in *V. flava* (19,474 bp), while the smallest was found in *A. lucyi* (4,573 bp).

### Comparative sequence divergent regions in genome

The complete chloroplast genomes of *V. nilotica* and *S. senegal* were compared with seventeen species for sequence divergent regions from the NCBI database using mVISTA [37]. The comparative analyses of the chloroplast genome showed a high level of similarity. Overall, the comparison of these chloroplast genomes observed similarity in coding regions, while non-coding



**Fig 2. Analysis of repeated sequences in *V. nilotica* and *S. senegal*.** (A) Totals of three repeat types, (B) Frequency of palindromic repeats by length, (C) Frequency of forward repeats by length and (D) Frequency of tandem repeats by length.

<https://doi.org/10.1371/journal.pone.0225469.g002>

**Table 2. Genes in the sequenced *V. nilotica* and *S. senegal* species chloroplast genomes.**

Category	Group of genes	Name of genes
Self-replication	Large subunit of ribosomal proteins	<i>rpl2, 14, 16, 20, 22, 23, 32, 33, 36</i>
	Small subunit of ribosomal proteins	<i>rps2, 3, 4, 7, 8, 11, 12, 14,15, 16, 18, 19</i>
	DNA dependent RNA polymerase	<i>rpoA, B, C1, C2</i>
	rRNA genes	<i>rrn 4.5, rrn 5, rrn 16, rrn23</i>
	tRNA genes	<i>trnC-GCA, trnD-GUC, trnfM-CAU, trnG-UCC, trnH-GUG, trnI-CAU, trnI-GAU, trnK-UUU, trnL-CAA, trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU, trnP-GGG, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC, trnW-CCA, trnY-GUA</i>
Photosynthesis	Photosystem I	<i>psaA, B, C, I, J,</i>
	Photosystem II	<i>psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z</i>
	Cytochrome b6/f complex	<i>petA, B, D, G, L, N</i>
	ATP synthase	<i>atpA, B, E, F, H, I</i>
	Rubisco	<i>rbcl</i>
Other genes	Maturase	<i>matK</i>
	Protease	<i>clpP</i>
	Envelop membrane protein	<i>cemA</i>
	Subunit Acetyl-CoA-Carboxylate	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
Unknown	Conserved open reading frames	<i>ycf1, 3,4, 15</i>

<https://doi.org/10.1371/journal.pone.0225469.t002>

regions had more variation, which is almost two times that of coding regions (S1 Fig). The *V. nilotica* chloroplast genome was used as a reference genome. The comparative analyses of *V. nilotica* with related species revealed high sequence similarity with no obvious difference from *V. flava*. The most variable coding regions found in these genomes are *tran K, rps16, rpoC1, petB, petD, ycf2, rrn23,* and *ndhA*. In particular, the *ycf1* gene displayed more variation among all variable regions (S1 Fig).

### Analysis of repetitive sequences in genomes

Repeat analysis of the sequenced cp genomes showed that there were 123 repeats in the *S. senegal* cp genome, which comprised 24 palindromic repeats, 26 forward repeats and 73 tandem repeats. Similarly, in *V. nilotica*, 130 repeats were present, containing 17 palindromic, 34 forward and 79 tandem repeats (Fig 2). In *V. nilotica*, the highest number of repeats was observed, and the sizes ranged from 15–29 in all palindromic, forward and tandem repeats containing 11, 20 and 72 repeats, respectively. A similar trend was observed in *S. senegal* containing 15–29 repeat sizes with 18, 23 and 70 repeats, respectively. Analysis of total repeats showed that *V. nilotica* had similarity with *Archidendron lucyi* in repeat number, with each containing 130 repeats. Similarly, *V. flava* and *V. seyal* also had 131 and 134 repeats, which showed similarity in repeat number to *V. nilotica*. The other species that was similar to *S. senegal* regarding repeats was *S. laeta*, containing 120 repeats, suggesting that *S. senegal* shows similarity in terms of repeats. Overall, in the compared genomes, *Acacia ligulata* comprised the highest number of repeats (140), and *Albizia odoratissima* contained the lowest number of repeats (101) (Fig 2).



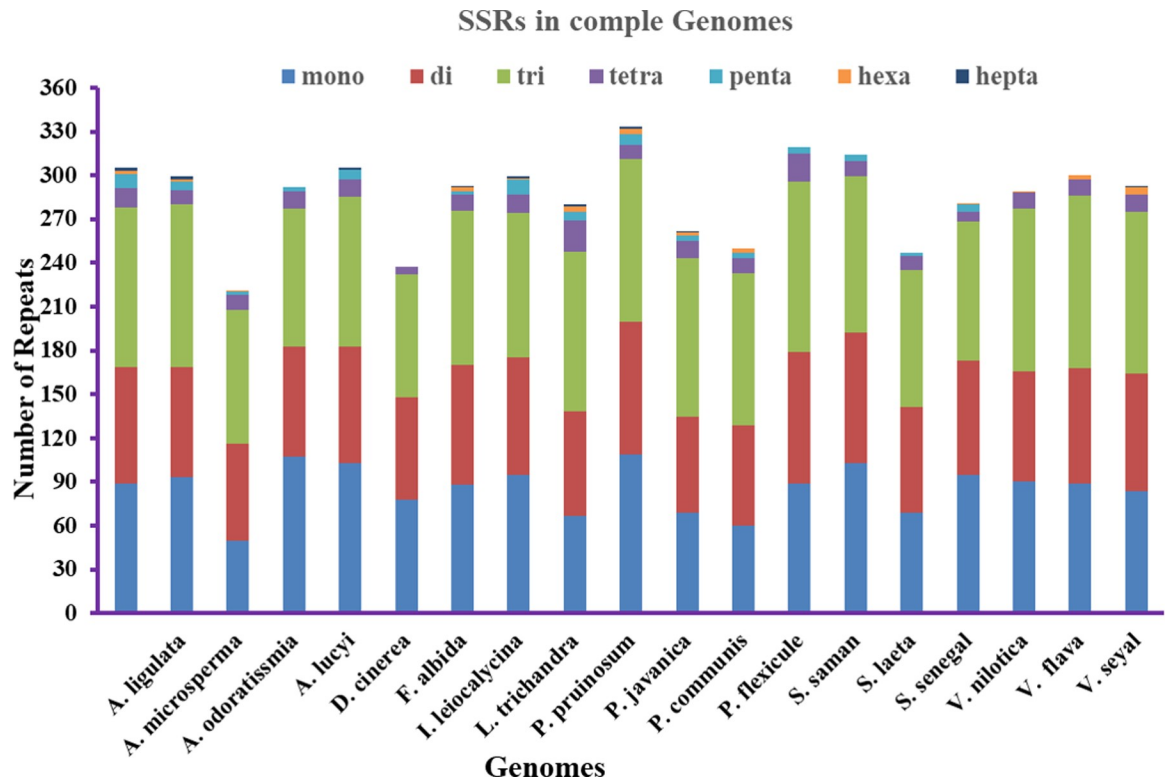


Fig 3. Analysis of simple sequence repeat (SSR) in *V. nilotica* and *S. senegal* genomes with related species cp genomes. Number of different SSR types detected in these genomes.

<https://doi.org/10.1371/journal.pone.0225469.g003>

### SSRs in the genomes

The SSRs (1–7) present in the *V. nilotica* genome were analyzed, and a total of 290 and 281 SSRs were present in *V. nilotica* and *S. senegal*, respectively. In *V. nilotica*, the most numerous SSRs were trinucleotide repeats (111) followed by mononucleotide (90) and dinucleotide (76) SSRs (Fig 3). The highest number of SSR nucleotides present in the *V. nilotica* genome was an octanucleotide (1). Similarly, in *S. senegal*, the total number of SSRs found was 281, where the

Table 3. Base composition of the *Vachellia nilotica* and *Senegalia senegal* chloroplast genome.

	T/U		C		A		G		Length (bp)	
	V.N	S.S	V.N	S.S	V.N	S.S	V.N	S.S	V.N	S.S
Genome	32.9	32.7	17.9	18	31.7	31.6	17.5	17.7	165343	162702
LSC	34.6	34.3	16.7	16.9	32.9	32.7	15.8	16	93849	91791
SSC	35.8	35.8	14.2	14.4	34.5	34.2	15.5	15.5	193.8	18895
IR	28.4	28.3	22.1	22.1	28.9	29.1	20.6	20.6	26093	26008
tRNA	25.3	24.9	23.3	23.4	22.1	22.2	29.3	29.5	2847	2794
rRNA	19	18.9	23.7	23.7	25.6	25.7	31.7	31.7	9052	9052
Protein Coding genes	31.9	32.1	17.4	17.3	30.6	30.7	20.1	19.9	79520	75571
1st position	23.36	25.2	17.43	18.20	34.47	30.8	25.57	25.16	26506	25189
2nd position	33.59	32.2	19.09	19.78	29.26	29.62	19.44	17.49	26506	25189
3rd position	40.12	38.7	15.24	13.8	27.82	31.6	15.15	14.4	26506	25189

V.N = *Vachellia nilotica*, S. S = *Senegalia Senegal*.

<https://doi.org/10.1371/journal.pone.0225469.t003>

highest number of nucleotides were trinucleotide (95) followed by mononucleotide (94) and dinucleotide repeats (78). Furthermore, *V. nilotica* contains the least number of SSRs when compared to other *Vachellia* species, i.e., *V. flava* and *V. seyal* with 302 and 295 SSR repeats, respectively. *S. senegal* had the highest number of SSRs compared to *S. laeta*. The number of SSRs was abundant in coding regions of all the sequenced and compared cp genomes (Fig 4). *V. nilotica* and *S. senegal* had 211 and 195 SSRs in the coding regions, respectively. Furthermore, *V. flava* contained the highest number (218) of SSRs in the coding region among all the compared genomes (Fig 4).

### Contraction and expansion of IR regions

Comprehensive comparative analysis of the junction region was performed among the 19 species for the contraction and expansion in  $J_{Lb}$  (LSC-IRb),  $J_{SB}$  (IRb-SSC),  $J_{SA}$  (SSC-IRa), and  $J_{LA}$  (IRa-LSC) and also for the position of genes present on these junctions. The largest inverted repeat region was found in the largest chloroplast genome of *P. flexicaula*, which was 41,503 bp in size, and the smallest IR region was found in the *P. javanica* (26,007 bp) chloroplast genome.

Although genomic structure and gene composition are highly conserved among these genomes, there are some differences in the IR region. Comparison of the  $J_{SB}$  junction of *Vachellia* species (*V. nilotica*, *V. seyal*, *V. flava*) and *Senegalia* species (*S. senegal*, *S. laeta*) revealed small differences, and the genes at the junction regions are also conserved. In the junction regions of all the compared genomes, the *ycf1* gene is conserved and present at the

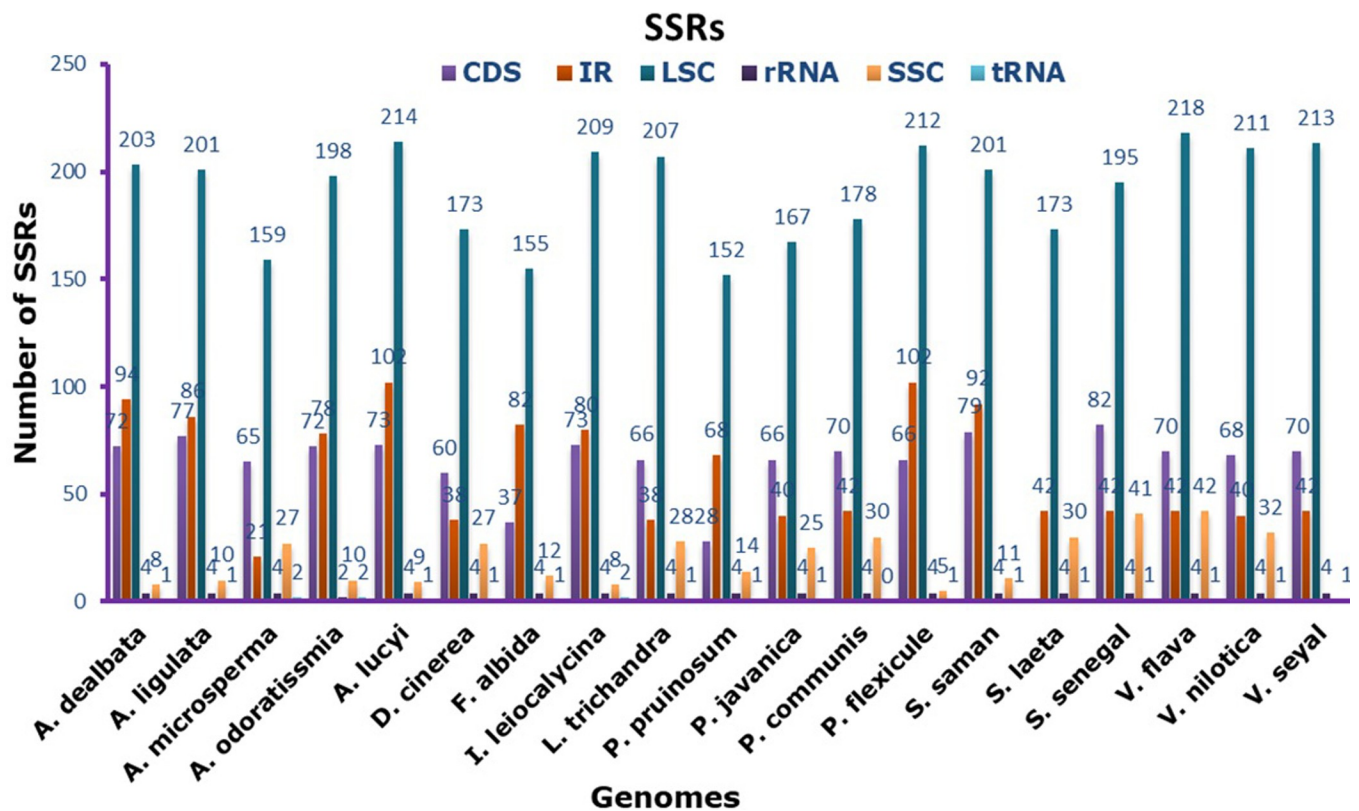
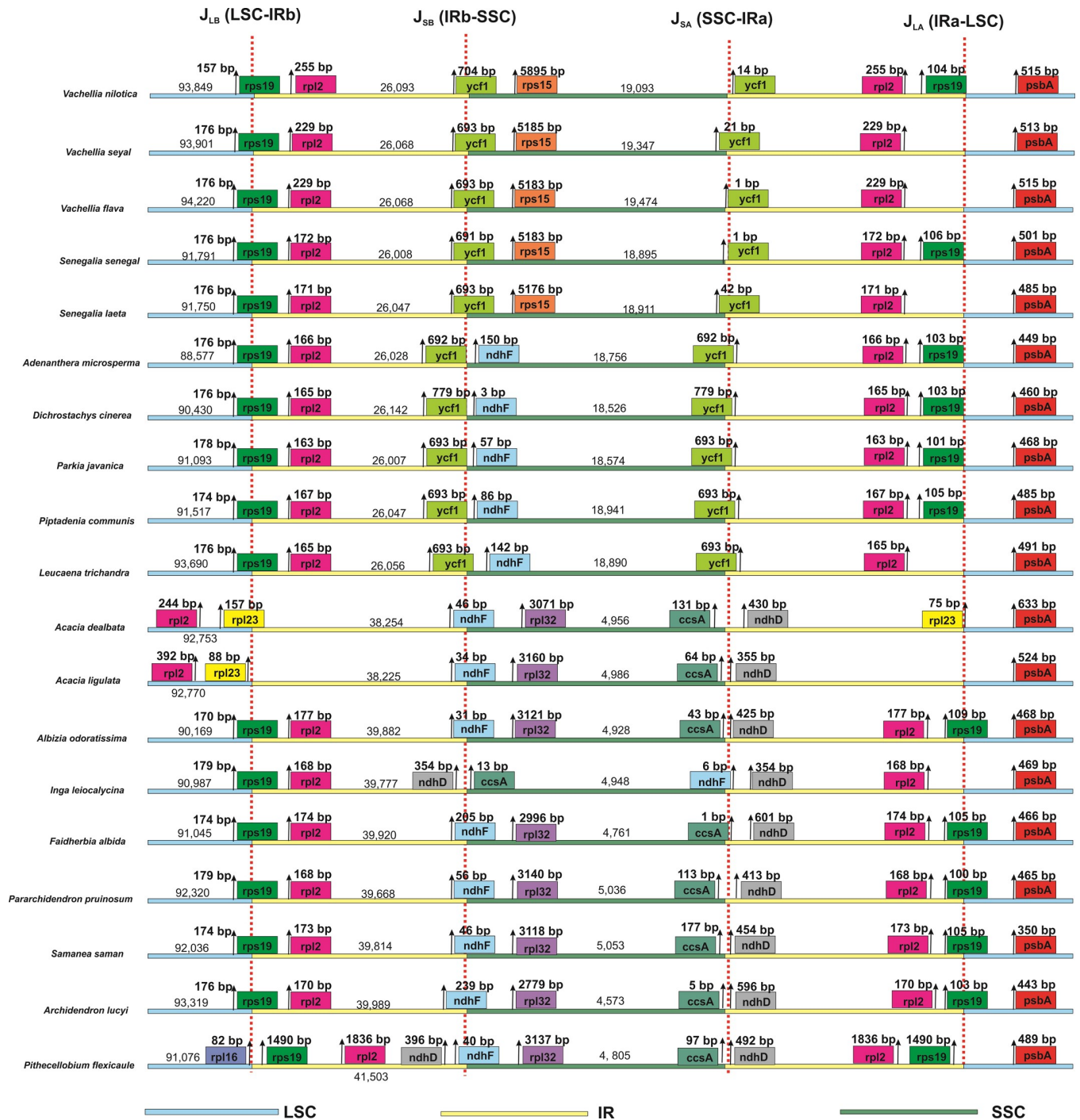


Fig 4. Analysis of simple sequence repeat (SSR) in the *V. nilotica* and *S. senegal* genomes. Frequency of identified SSRs in the Small Single-Copy (SSC), Large Simple-Copy (LSC), Inverted Repeat (IR), transfer RNA (tRNA), ribosomal RNA (rRNA), and coding sequence (CDS) regions.

<https://doi.org/10.1371/journal.pone.0225469.g004>



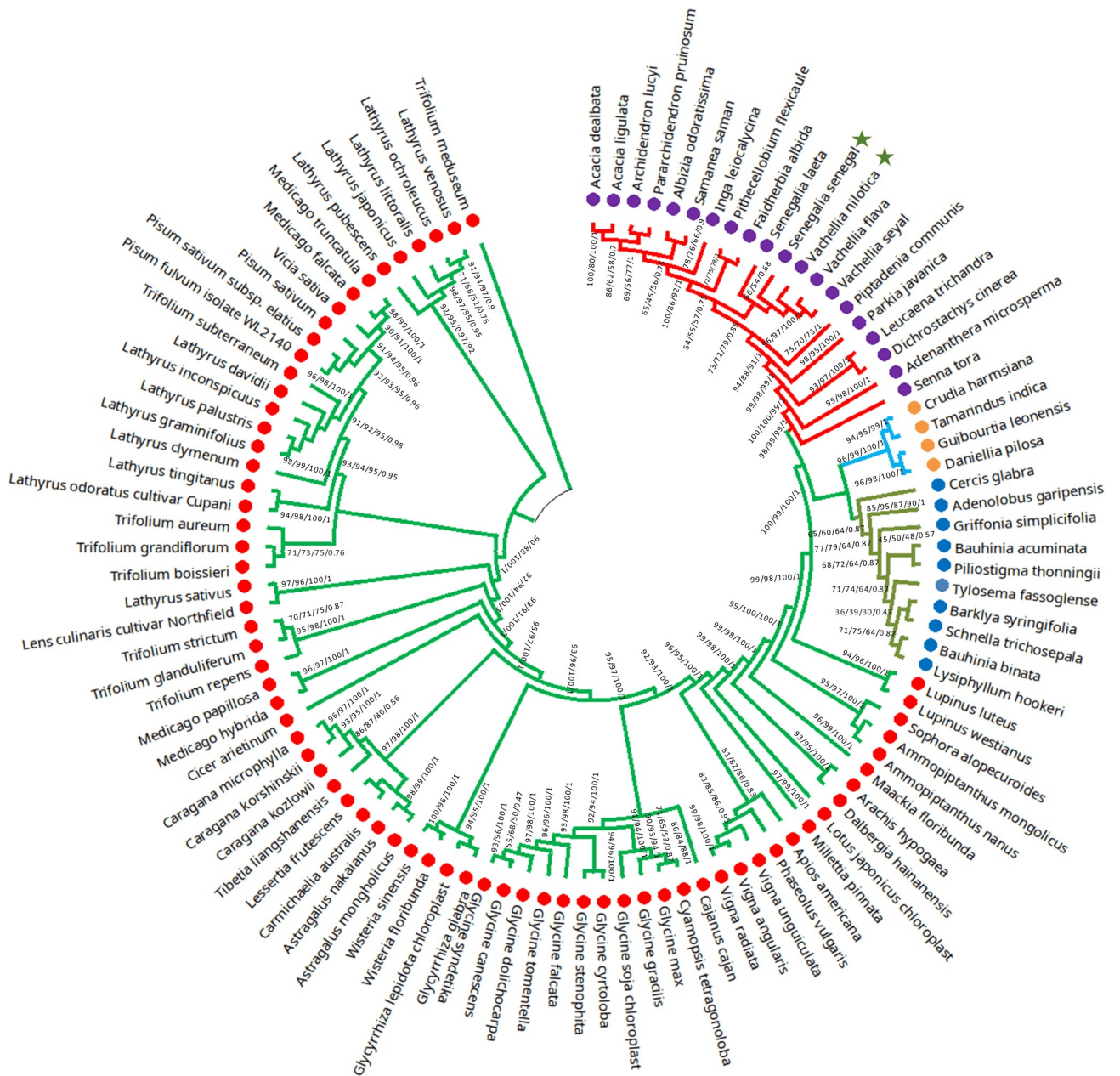
**Fig 5. Comparison of border distance between adjacent genes and junctions of LSC, SSC and two IR regions among the chloroplast genomes of *V. nilotica* and *S. senegalia* with related species.** Boxes above or below the main line indicate the adjacent border genes. The figure is not to scale with respect to sequence length and only shows relative changes at or near the IR/SC borders.

<https://doi.org/10.1371/journal.pone.0225469.g005>

same position (Fig 5). In the J<sub>SB</sub> junction in *Vachellia* and *Senegalia* species, the *ycf1* gene is present, while in the remaining species, it is located in the IRb region. Furthermore, at the J<sub>LB</sub> junction in all the genomes, the *rpl2* gene is located in the IRb region, except for *A. ligulata*



and *A. dealbata*, in which the *rpl2* gene is present in the LSC region. Moreover, at the  $J_{SB}$  junction, the *rps15* gene was found in the SSC region of all *Vachellia* and *Senegalia* species, while other compared genomes were absent. In addition, *rpl23* was only at the  $J_{LB}$  junction in *Accacia dealbata*. Similarly, in *S. senegal*, the *ycf1* gene is present at the  $J_{SA}$  junction, while in *V. nilotica*, it was found 14 bp away from  $J_{SA}$  in the IRa region (Fig 5).



**Fig 6. Phylogenetic tree constructed on the basis of whole genome dataset using four different methods: Bayesian inference (BI), maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ).** Numbers above the branches are the posterior probabilities of BI and bootstrap values for ML, MP and NJ. The star represents the position of *V. nilotica* and *S. senegal*.

<https://doi.org/10.1371/journal.pone.0225469.g006>

## Phylogenetic analyses: Confirmation of recent classification based on complete CP

Previously, numerous studies were conducted to resolve the phylogenetic position of Mimosoideae [46], but no study to date has investigated the basis of the complete chloroplast genome of *Vachellia* and *Senegalia* species. In this study, the phylogenetic position of *V. nilotica* and *S. senegal* within the family Fabaceae was established by analyzing multiple sequence alignments of complete cp genomes and 56 shared genes of 104 Fabaceae members (Fig 6 and S2). The 56 shared genes (from all species) and the complete cp genome sequence generated phylogenetic trees with identical topologies (Fig 6 and S2). In these phylogenetic trees, *S. senegal* formed a sister clade with *S. laeta*, while *V. nilotica* shared a sister clade with *V. flava* and *V. seyal* with high posterior probability and bootstrap support values using four different methods (Fig 6). Our results supported the recent classification of *V. nilotica* and *S. senegal* in the genera *Vachellia* and *Senegal*, respectively, and did not support the former placement of these species in the genus *Acacia*.

## Discussion

This study reports the complete chloroplast genomes of *S. Senegal* and *V. nilotica*, ranging from 162.7~165.3 kb in length. Both cp genomes exhibit a typical quadripartite conserved structure, as reported for other angiosperm genomes [44, 47]. Both *V. nilotica* and *S. Senegal* encode 135 and 132 genes, including 89 and 87 protein-coding genes, respectively. Similar differences in the protein coding genes were also observed, as in previously reported genomes [46]. The important genes present in these genomes were also similar to those of previously reported angiosperm cp genomes [44, 48]. The main reason for size variation among the chloroplast genomes is the contraction and expansion in the IR regions of the genome [49]. The size variation was observed (161,681 bp ~178,887 bp) to be in keeping with the previously reported angiosperm genomes [46]. Genome conservation was observed in both genera with some minor changes in IR/SSC regions, which reveal evidence of variation in the chloroplast genomes and also provides some information in the evolutionary context of chloroplast genomes [50]. Divergence hotspots among the species facilitate comparative genomics, species identification [51] and phylogenetic studies at different levels [52]. Comparative analysis of these genomes through mVISTA revealed that coding regions, such as *rps16*, *rpoC2*, *atpF*, *rpoC1*, *accD*, *clpP*, *petD*, *rpl16*, *ycf1*, *ycf2* and *ndhA*, were more divergent than the non-coding regions, which is similar to the findings obtained with previously reported cp genomes [46, 53]. The significance of these divergent regions can be further used as potential DNA markers for phylogenetic studies, population genetics studies and species identification studies [54]. Some of the protein coding genes present in the plastid genomes were found to have versatile roles in the resolution of phylogenetic relationships of complex plant taxa, such as *rpoA*, *psal*, *petB* and *rps19* in *Notopterygium* species [55] and *ycf1* in *Anemopaegma* species [56]. Moreover, in some other species, such as *Veroniceae*, the *petD-rpoA*, *ycf4-cemA*, and *rpl32-trnL* genes were used for the identification of the species. In our study, the PCG regions were more conserved and showed less sequence divergence than the intergenic spacer region, which had a higher degree of divergence among the compared species. Surprisingly, the IR regions in these compared cp genomes were less divergent compared to the LSC and SSC regions, which were also previously reported [53].

Repetitive sequences within the chloroplast genome play a crucial role in evolution, divergence studies and cp genome rearrangement. Moreover, microsatellite-like SSRs play an important role in molecular-level identification and in population genetics [57, 58]. The identification of repetitive sequences in the IGS provides useful information in various angiosperm



species [58]. Among all the compared genomes, *Albizia odoratissima* had the lowest number of total repeats (104), and *Acacia ligulata* had the highest number of total repeats in cp genomes of subfamily Mimosoideae [46] and among other angiosperms [59]. The *Adenanthera microsperma* genome was found to contain the highest number of palindromic repeats (26), and the lowest number (16) was reported in *Pithecellobium flexicaule*. The highest number of forward repeats (36) was found in *Pararchidendron pruinosum*, while the lowest (24) was found in *A. microsperma*. The tandem repeats were highest (91) in *Dichrostachys cinerea*, and the lowest (52) was in *Albizia odoratissima*. Plastome size variation leads to the variation in tandem repeats [60] and dispersed repeats as previously reported by [57]. Earlier studies also showed that these repeats play an important role in structural variation [61]. The highest number of SSRs among these genomes was 333 in *Parachidendron pruinosum*, while the lowest number of SSRs was observed in *A. microsperma*. This result was consistent with the previously reported chloroplast genome of wild roses [62].

The phylogenetic relationship of the genus *Vachellia* and *Senegalia* belonging to the subfamily Mimosoideae (Fabaceae) was poorly resolved previously using only a few plastid markers [63–65]. Phylogenomic analysis based on the complete chloroplast genome can be widely used to resolve the complex relationship at the family level, as previously reported in *orchidaceae* [66], and *Bambusoideae* [67]. A detailed comprehensive study of the subfamily Mimosoideae was reported by Wang et al. [46], but there was no mention of the phylogenomic placement of *V. nilotica* and *S. senegal* into the genus *Vachellia* and *Senegalia*. The results of our study indicate that phylogenetic trees based on the complete genome dataset and 56 shared genes of *V. nilotica* and *S. senegal* contain the same phylogenetic signals and support the recent classification of *V. nilotica* and *S. senegal* in the genera *Vachellia* and *Senegal*, respectively (Fig 6). A complete phylogeny of Mimosoideae was constructed to resolve the evolutionary relationship of Mimosoideae with Fabaceae. Structural rearrangement in the chloroplast genome is an important phylogenetic signal and is used to define monophyletic lineages in plant groups [68].

## Conclusion

We sequenced the chloroplast genome of *V. nilotica* and *S. senegal*. Both genomes shared the same gene organization and overall genome structure, which were also found in related species. The quadripartite structure (LSC/SSC/IRA/IRB) of the genomes was compared for Mimosoideae species, and no significant variation was noted in these genomes, instead showing the closest similarity to these species. The phylogenetic relationships of these species, which were formerly classified in the genus *Acacia* and later placed in the genera *Vachellia* and *Senegalia*, were validated on the basis of the complete chloroplast genome. Furthermore, the phylogenetic analyses revealed that both *V. nilotica* and *S. senegal* formed monophyletic clades, while *V. nilotica* further shared sub-monophyletic clades with *V. flava* and *V. seyal*, while the *S. senegal* shared the same clade with *S. laeta*. These findings may help to elucidate the complex taxonomy of these genera and the studied species *V. nilotica* and *S. senegal*.

## Availability of data and materials

All data generated or analyzed during this study are included in this published article.

## Supporting information

**S1 Fig. Alignment visualization of the *V. nilotica* and *S. Senegal* chloroplast genome sequences.** VISTA-based identity plot showing sequence identity among nineteen species, using *V. nilotica* as a reference genome. The vertical scale indicates the percentage of identity,

ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Arrows indicate the annotated genes and their transcriptional direction. (PDF)

**S2 Fig. Phylogenetic trees of *V. nilotica* and *S. senegal* based on protein coding genes (PCGs).** A phylogenetic tree was constructed for 104 species from the family Fabaceae based on 56 shared protein coding genes. The following four different methods were used for the 56 shared gene data sets: Bayesian inference (BI), maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ). Numbers above the branches are the posterior probabilities of BI and bootstrap values for ML, MP and NJ. (PDF)

## Author Contributions

**Data curation:** Arif Khan.

**Formal analysis:** Sajjad Asaf.

**Methodology:** Arif Khan.

**Project administration:** Abdul Latif Khan.

**Resources:** Ahmed Al-Harrasi.

**Software:** Ahmed Al-Rawahi.

**Supervision:** Ahmed Al-Harrasi, Ahmed Al-Rawahi.

**Writing – original draft:** Sajjad Asaf.

**Writing – review & editing:** Abdul Latif Khan.

## References

1. Haque A. Investigation of the fungi associated with dieback of prickly acacia (*Vachellia nilotica* subsp. *indica*) in Northern Australia. PhD, The University of Queensland. 2015.
2. Kyalangalilwa B, Boatwright JS, Daru BH, Maurin O, van der Bank M. Phylogenetic position and revised classification of *Acacia* s.l. (Fabaceae: Mimosoideae) in Africa, including new combinations in *Vachellia* and *Senegalia*. *Botanical Journal of the Linnean Society*. 2013; 172(4):500–23.
3. Beshai AA. The economics of a primary commodity: Gum Arabic. *Oxford bulletin of economics and statistics*. 1984; 46(4):371–81.
4. Khan IA, Abourashed EA. *Leung's encyclopedia of common natural ingredients: used in food, drugs and cosmetics*: John Wiley & Sons; 2011.
5. Al-Assaf S, Phillips GO, Aoki H, Sasaki Y. Characterization and properties of *Acacia senegal* (L.) Willd. var. *senegal* with enhanced properties (*Acacia* (sen) SUPER GUM™): Part 1—Controlled maturation of *Acacia senegal* var. *senegal* to increase viscoelasticity, produce a hydrogel form and convert a poor into a good emulsifier. *Food Hydrocolloids*. 2007; 21(3):319–28.
6. Motlagh S, Ravines P, Karamallah K, Ma Q. The analysis of *Acacia* gums using electrophoresis. *Food hydrocolloids*. 2006; 20(6):848–54.
7. Arce LR, Banks H. A preliminary survey of pollen and other morphological characters in neotropical *Acacia* subgenus *Aculeiferum* (Leguminosae: Mimosoideae). *Botanical Journal of the Linnean Society*. 2001; 135(3):263–70.
8. Salih AA, El Fadl MA, Kaarakka V, Luukkanen O. Symbiotic nitrogen fixation in eight *Acacia senegal* provenances in dryland clays of the Blue Nile Sudan estimated by the 15 N natural abundance method. *Plant and Soil*. 2005; 275(1–2):261–9.
9. Bargali K, Bargali S. *Acacia nilotica*: a multipurpose leguminous plant. *Nature and Science*. 2009; 7(4):11–9.
10. Raj A, Chandrawanshi S. *Acacia nilotica*: a multipurpose tree and source of Indian gum Arabic. *South Indian Journal of Biological Sciences*. 2015; 1(2):66–9.

11. Good A. Toward nitrogen-fixing plants. *Science*. 2018; 359(6378):869–70. <https://doi.org/10.1126/science.aas8737> PMID: 29472469
12. Donalisio M, Cagno V, Civra A, Gibellini D, Musumeci G, Rittà M, et al. The traditional use of *Vachellia nilotica* for sexually transmitted diseases is substantiated by the antiviral activity of its bark extract against sexually transmitted viruses. *Journal of ethnopharmacology*. 2018; 213:403–8. <https://doi.org/10.1016/j.jep.2017.11.039> PMID: 29203273
13. Rather LJ, Mohammad F. *Acacia nilotica* (L.): a review of its traditional uses, phytochemistry, and pharmacology. *Sustainable Chemistry and Pharmacy*. 2015; 2:12–30.
14. Daru BH, Kyalangalilwa B, Maurin O, van der Bank M, Boatwright JS. Phylogenetic position and revised classification of *Acacia* s.l. (Fabaceae: Mimosoideae) in Africa, including new combinations in *Vachellia* and *Senegalia*. *Botanical Journal of the Linnean Society*. 2013; 172(4):500–23. <https://doi.org/10.1111/boj.12047>
15. Orchard AE, Maslin BR. The case for conserving *Acacia* with a new type. *Taxon*. 2005; 54(2):509–12.
16. Gardner FP, Pearce RB, Mitchell RL. *Physiology of crop plants*: Scientific Publishers; 2017.
17. Taylor DB, Dhileepan K. Implications of the changing phylogenetic relationships of *Acacia* s.l. on the biological control of *Vachellia nilotica* ssp. *indica* in Australia. *Annals of Applied Biology*. 2019; 174(2):238–47.
18. Jansen RK, Ruhlman TA. *Plastid genomes of seed plants. Genomics of chloroplasts and mitochondria*: Springer; 2012. p. 103–26.
19. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome biology*. 2016; 17(1):134. <https://doi.org/10.1186/s13059-016-1004-2> PMID: 27339192
20. Wicke S, Schneeweiss GM, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant molecular biology*. 2011; 76(3–5):273–97. <https://doi.org/10.1007/s11103-011-9762-4> PMID: 21424877
21. Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution*. 2011; 28(7):2077–86. <https://doi.org/10.1093/molbev/msr028> PMID: 21289370
22. Wicke S, Müller KF, de Pamphilis CW, Quandt D, Wickett NJ, Zhang Y, et al. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *The Plant Cell*. 2013; 25(10):3711–25. <https://doi.org/10.1105/tpc.113.113373> PMID: 24143802
23. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. *Functional & integrative genomics*. 2015; 15(2):141–61.
24. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC biology*. 2009; 7(1):84.
25. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*. 2007; 104(49):19369–74.
26. Kim K, Lee S-C, Lee J, Lee HO, Joh HJ, Kim N-H, et al. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PloS one*. 2015; 10(6):e0117159. <https://doi.org/10.1371/journal.pone.0117159> PMID: 26061692
27. Ross JH. *A conspectus of African acacia species* 1979.
28. Miller JT, Bayer RJ. Molecular phylogenetics of *Acacia* subgenera *Acacia* and *Aculeiferum* (Fabaceae: Mimosoideae), based on the chloroplast *matK* coding sequence and flanking *trnK* intron spacer regions. *Australian Systematic Botany*. 2003; 16(1):27–33.
29. Shi C, Hu N, Huang H, Gao J, Zhao Y-J, Gao L-Z. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. *PloS one*. 2012; 7(2):e31468. <https://doi.org/10.1371/journal.pone.0031468> PMID: 22384027
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9(4):357. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
31. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
32. Hahn C, Bachmann L, Chevreur B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic acids research*. 2013; 41(13):e129–e. <https://doi.org/10.1093/nar/gkt371> PMID: 23661685

33. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 2004; 20(17):3252–5. <https://doi.org/10.1093/bioinformatics/bth352> PMID: 15180927
34. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic acids research*. 2005; 33(suppl\_2):W686–W9.
35. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current genetics*. 2007; 52(5–6):267–74. <https://doi.org/10.1007/s00294-007-0161-y> PMID: 17957369
36. Kumar S, Nei M, Dudley J, Tamura K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*. 2008; 9(4):299–306. <https://doi.org/10.1093/bib/bbn017> PMID: 18417537
37. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic acids research*. 2004; 32(suppl\_2):W273–W9.
38. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic acids research*. 2001; 29(22):4633–42. <https://doi.org/10.1093/nar/29.22.4633> PMID: 11713313
39. Kraemer L, Beszteri B, Gäbler-Schwarz S, Held C, Leese F, Mayer C, et al. S TAMP: Extensions to the S TADEN sequence analysis package for high throughput interactive microsatellite marker design. *BMC bioinformatics*. 2009; 10(1):41.
40. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*. 1999; 27(2):573. <https://doi.org/10.1093/nar/27.2.573> PMID: 9862982
41. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013; 30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
42. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*. 1980; 16(2):111–20. <https://doi.org/10.1007/bf01731581> PMID: 7463489
43. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19(12):1572–4. <https://doi.org/10.1093/bioinformatics/btg180> PMID: 12912839
44. Asaf S, Khan AL, Khan MA, Waqas M, Kang S-M, Yun B-W, et al. Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. *Scientific reports*. 2017; 7(1):7556. <https://doi.org/10.1038/s41598-017-07891-5> PMID: 28790364
45. Wu Z, Tembrock LR, Ge S. Are differences in genomic data sets due to true biological variants or errors in genome assembly: an example from two chloroplast genomes. *PLoS One*. 2015; 10(2):e0118019. <https://doi.org/10.1371/journal.pone.0118019> PMID: 25658309
46. 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 genetics & genomes*. 2017; 13(2):41.
47. Yang JB, Li DZ, Li HT. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Molecular Ecology Resources*. 2014; 14(5):1024–31. <https://doi.org/10.1111/1755-0998.12251> PMID: 24620934
48. Asaf S, Waqas M, Khan AL, Khan MA, Kang S-M, Imran QM, et al. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Frontiers in plant science*. 2017; 8:304. <https://doi.org/10.3389/fpls.2017.00304> PMID: 28326093
49. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics*. 2007; 8(1):174. <https://doi.org/10.1186/1471-2164-8-174> PMID: 17573971
50. Zhang S-D, Jin J-J, Chen S-Y, Chase MW, Soltis DE, Li H-T, et al. Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. *New Phytologist*. 2017; 214(3):1355–67. <https://doi.org/10.1111/nph.14461> PMID: 28186635
51. Ahmed I, Matthews PJ, Biggs PJ, Naeem M, McLenachan PA, Lockhart PJ. Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Molecular Ecology Resources*. 2013; 13(5):929–37. <https://doi.org/10.1111/1755-0998.12128> PMID: 23718317
52. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany*. 2015; 40(1):336–51.
53. Dong W-L, Wang R-N, Zhang N-Y, Fan W-B, Fang M-F, Li Z-H. Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution.

- International Journal of Molecular Sciences. 2018; 19(3):716. <https://doi.org/10.3390/ijms19030716> PMID: 29498674
54. Ye X, Hu D, Guo Y, Sun R. Complete Chloroplast Genome of *Castanopsis sclerophylla* (Lindl.) Schott: Genome Structures, Comparative and Phylogenetic Analysis. *BioRxiv*. 2019:540617.
  55. Yang J, Vázquez L, Chen X, Li H, Zhang H, Liu Z, et al. Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (*Quercus* Subgenus *Quercus*) and Assessment of Their Utility as DNA Barcodes. *Frontiers in Plant Science*. 2017; 8(816). <https://doi.org/10.3389/fpls.2017.00816> PMID: 28579999
  56. Firetti F, Zuntini AR, Gaiarsa JW, Oliveira RS, Lohmann LG, Van Sluys M-A. Complete chloroplast genome sequences contribute to plant species delimitation: A case study of the *Anemopaegma* species complex. *American Journal of Botany*. 2017; 104(10):1493–509. <https://doi.org/10.3732/ajb.1700302> PMID: 29885220
  57. Weng M-L, Blazier JC, Govindu M, Jansen RK. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular biology and evolution*. 2013; 31(3):645–59. <https://doi.org/10.1093/molbev/mst257> PMID: 24336877
  58. Xue J, Wang S, Zhou S-L. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *American Journal of Botany*. 2012; 99(6):e240–e4. <https://doi.org/10.3732/ajb.1100547> PMID: 22615305
  59. Khan AL, Al-Harrasi A, Asaf S, Park CE, Park G-S, Khan AR, et al. The first chloroplast genome sequence of *Boswellia sacra*, a resin-producing plant in Oman. *PloS one*. 2017; 12(1):e0169794. <https://doi.org/10.1371/journal.pone.0169794> PMID: 28085925
  60. Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in clpP. *Scientific reports*. 2015; 5:16958. <https://doi.org/10.1038/srep16958> PMID: 26592928
  61. Greiner S, Wang X, Rauwolf U, Silber MV, Mayer K, Meurer J, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: I. Sequence evaluation and plastome evolution. *Nucleic acids research*. 2008; 36(7):2366–78. <https://doi.org/10.1093/nar/gkn081> PMID: 18299283
  62. Jeon J-H, Kim S-C. Comparative Analysis of the Complete Chloroplast Genome Sequences of Three Closely Related East-Asian Wild Roses (*Rosa* sect. *Synstylae*; Rosaceae). *Genes*. 2019; 10(1):23.
  63. Bouchenak-Khelladi Y, Maurin O, Hurter J, Van der Bank M. The evolutionary history and biogeography of Mimosoideae (Leguminosae): an emphasis on African acacias. *Molecular Phylogenetics and Evolution*. 2010; 57(2):495–508. <https://doi.org/10.1016/j.ympev.2010.07.019> PMID: 20696261
  64. Luckow M, Miller JT, Murphy DJ, Livshultz T. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. *Advances in legume systematics, part*. 2003; 10:197–220.
  65. Miller JT, Seigler D. Evolutionary and taxonomic relationships of *Acacia* s.l. (Leguminosae: Mimosoideae). *Australian Systematic Botany*. 2012; 25(3):217–24.
  66. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences*. 2015; 282(1814):20151553.
  67. Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duvall MR. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC evolutionary biology*. 2015; 15(1):50.
  68. Henry RJ. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*: Cabi Publishing; 2005.