

Characterization of the complete chloroplast genome of *Agave* sp. (Asparagales: Asparagaceae: Agavoideae) and its phylogenetic analysis

Xinya Ma

Dali Universit

Xinghui Xu

Dali Universit

Chaoyang Luo

Dali Universit

Zhenyu Xiong

Dali Universit

Juan Zhang

Yunnan Normal University

Yuan Mu

`muy@eastern-himalaya.cn`

Dali Universit

Research Article

Keywords: Asparagaceae, *Agave* sp., Chloroplast genome, phylogenetic analysis, *Agave sensu lato*

Posted Date: January 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8552740/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Genus *Agave* is one of the most diverse and complex group of the family Asparagaceae, originates from the Americas (with Mexico as its native range) and holds significant economic value as a widely distributed crop. However, the systematic phylogenetic relationship of this group is still controversial. In this study, we successfully sequenced, assembled, and annotated the chloroplast genome of *Agave* sp. The result showed: 1) The chloroplast genome of *Agave* sp. was 157,489bp in length with a GC content of 38.57%. It consisted of a large single-copy region (LSC, 86,018 bp), a pair of inverted repeat regions (IR, 26,612bp), and a small single-copy region (SSC, 18,247bp). In addition, the chloroplast genome of *Agave* sp. contains a total of 113 unique genes, including 79 protein-coding genes (PCGs), 30 tRNAs, and 4 rRNAs. 2) phylogenetic analysis of Asparagaceae indicate that the genera *Yucca* and *Hosta* each form a distinct monophyletic clade, clustering independently on separate branches. *Agave* sp. shows a close evolutionary relationship with *Agave attenuata*, forming a sister group. Furthermore, our phylogenetic tree provides support for the taxonomic classification of *Agave sensu lato*. This study enriches the genomic data of *Agave* and offers crucial reference information for subsequent research on the evolutionary history, taxonomic classification, and resource utilization of the Asparagaceae family.

Introduction

Agave originates from the arid and semi-arid regions of the Americas and has diversified under different environmental conditions. The *Agave* genus has about 210 species, of which roughly 160 are found in Mexico, (Eguiarte et al., 2021; González Trillo et al., 2024). *Agave* is widely cultivated around the world and extensively used for producing food, beverages, fibers, and medicines. In addition, due to its strong drought resistance and resource-use efficiency, it has become a potential group for addressing global climate change (Eguiarte et al., 2021; González Trillo et al., 2024). Generally, *Agave* shows potential to compete with or complement traditional bioenergy crops, providing a direction for the development of semi-arid lands (Torres et al., 2015; González Trillo et al., 2024).

The classification and identification of *Agave* have long faced challenges. Widespread morphological variation caused by arid and semi-arid environments, natural hybridization, and convergent evolution together make it difficult to define species. Genetic and morphological differences within populations directly affect the quality fluctuations of *Agave* products (Torres et al., 2015). In previous studies, the core chloroplast fragments *rbcL* and *ndhF* were used to explore the phylogeny of the *Agave* genus, suggesting that *Agave sensu stricto* is not a monophyletic group. *Manfreda*, *Polianthes*, and *Prochnyanthes* are parallel lineages to *Agave sensu stricto*, but due to the limited variable sites of the molecular markers, the precise relationships among the groups could not be clearly determined (Bogler et al., 2006). Good-Avila et al (2006) integrated nuclear gene *ITS* and chloroplast fragments for a combined analysis, proposing the hypothesis of two divergence events in *Agave sensu lato* (clade *Agave sensu stricto*, *Polianthes*, *Manfreda*, and *Prochnyanthes*, which together represent 30% of the genus). However, due to insufficient resolution of the chloroplast data, the branch support was relatively low. Later, by further expanding the sampling of *matk* and nuclear genes, they not only confirmed the

paraphyly of *Agave sensu stricto* but also found that this classification characteristic was not directly associated with the evolution of pollination syndromes (Flores-Abreu et al., 2019). By using the highly variable nuclear gene *ITS* sequences to construct a phylogenetic tree, the broadly defined *Agave* genus was successfully divided into five well-supported clades, basically clarifying the paraphyletic nature of the narrowly defined groups, while also pointing out the obvious limitations of traditional chloroplast fragment analysis (Jiménez-Barrón et al., 2020). However, relying solely on studies of nuclear *ITS* and traditional chloroplast fragments still makes it difficult to verify whether there is a conflict in the co-evolution of the nuclear and plastid genomes. Single-gene markers cannot completely avoid issues such as multi-copy incongruence and interference from random mutations, so the phylogenetic relationships of this group still need to be further investigated.

The chloroplast genome is characterized by structural conservation, high copy number, and a slow evolutionary rate, and it contains regions with different evolutionary rates, such as coding and non-coding regions (Hang et al., 2025; Li et al., 2024). It can provide abundant molecular trait information, compensating for the limitations of a single marker or fragment. The complete sequence of the chloroplast genome has been widely used in plant phylogenetic studies (Wu et al., 2021; Ma et al., 2025; Yang et al., 2022). Previous studies of the complete chloroplast genomes have widely used in genus *Agave*, which supported the monophyly of *Agave sensu stricto* from a maternal inheritance perspective (González-Trillo et al., 2024; Yang et al., 2021; Xu et al., 2022).

So far, nearly 1,278 chloroplast genomes of Asparagaceae are stored in NCBI (National Center for Biotechnology Information), but only 12 of them belong to *Agave sensu lato*. This severely restricts the in-depth study of this group. Therefore, this study aims to obtain the complete chloroplast genome sequence of *Agave* sp. discuss the phylogenetic analysis. This study will provide fundamental data support for future research on chloroplast-related evolution, ecology, and resource utilization.

Material and methods

Collection of plant material

The fresh leaves were collected in Dali University, Dali, Yunnan Province, China (25°40'25"N 100°09'22"E) in March 2025. (specimen number: IEHBR-2025001)

DNA extraction and chloroplast genome sequencing and assembly

Genomic DNA was extracted from leaf samples using the kit according to the instructions provided by Tiangen. The quality and quantity of DNA were evaluated by agarose gel electrophoresis, NanoDrop 2000, and Qubit 3.0 fluorometer. DNA was prepared for sequencing using NEB kits, and PCR fragment amplification was performed. The resulting libraries were purified and quality-assessed using the Agilent 5400 system, and their concentrations were determined. High-throughput sequencing (paired-end 150 bp) was conducted on the Illumina HiSeq X system at BGI Genomics (Shaanxi, China) (Illumina, San

Diego, CA). Gene annotation was performed using Geneious software, with the *Agave durangensis* genome as a reference. Annotations were manually reviewed to ensure accuracy, including the start and stop codons of protein-coding genes and the boundaries of introns and exons. A circular map of the *Seravschanicus* chloroplast genome was obtained using OGDRAW. Codon usage characteristics were analyzed with CodonW software, systematically examining the structure and features of the *Agave* sp. chloroplast genome, and comprehensively assessing structural characteristics, codon usage, and relative synonymous codon usage (RSCU) (Sharp PM et al., 1987). Finally, the chloroplast genome was submitted to GenBank with the accession number.

Phylogenetic analysis

To further explore the position of *Agave* sp. and phylogenetic relationships family Asparagaceae, this study constructed a phylogenetic tree using the complete chloroplast genome. The analysis covered 42 species of the Asparagaceae family, *Asparagus cochinchinensis* and *Zephyranthes mesochloa* used as outgroups (Table 1). Sequence acquisition were performed using PhyloSuite v1.2.2 (Zhang et al., 2020). Preliminary alignments were conducted with MAFFT v7.505 (Kato and Standley, 2013) and the alignment results were optimized using MACSE v2.07 (Ranwez et al., 2018). ModelFinder (Kalyaanamoorthy et al., 2017) was used to determine the optimal model based on the Bayesian Information Criterion (BIC), which was found to be TVM. Maximum likelihood (ML) phylogenetic analysis was carried out in IQ-TREE v1.6.12 (Nguyen et al., 2015) with the branch support tested using 1000 ultrafast bootstrap replicates. Visualization and annotation of the phylogenetic tree were performed with the online tool iTOL v6 (Letunic and Bork, 2021), and images were optimized using Adobe Photoshop 2023 to meet publication requirements.

Table 1
Basic characteristics of chloroplast genomes in 42 species for phylogenetic analysis.

species	GenBank NO	species	GenBank NO
<i>Agave attenuata</i>	NC_032696.1	<i>Hosta minor</i>	NC_035999.1
<i>Agave amaniensis</i>	NC_058311.1	<i>Hosta plantaginea</i>	NC_053555.1
<i>Agave americana</i>	NC_032053.1	<i>Hosta sieboldiana</i>	NC_060662.1
<i>Agave angustifolia</i>	NC_059876.1	<i>Hosta tsushimensis</i>	NC_060664.1
<i>Agave cantula</i>	NC_070107.1	<i>Hosta tsushimensis var.</i>	NC_060663.1
<i>Agave durangensis</i>	NC_072313.1	<i>Hosta ventricosa</i>	NC_032706.1
<i>Agave fourcroydes</i>	NC_059874.1	<i>Hosta venusta</i>	NC_046895.1
<i>Agave hybrid cultivar</i>	NC_045534.1	<i>Hosta yingeri</i>	MZ919315.1
<i>Agave sisalana</i>	NC_059875.1	<i>Manfreda virginica</i>	NC_032707.1
<i>Beschorneria septentrionalis</i>	NC_032699.1	<i>Polianthes sp.</i>	KX931464.1
<i>Camassia scilloides</i>	NC_032700.1	<i>Schoenolirion croceum</i>	NC_032710.1
<i>Chlorogalum pomeridianum</i>	NC_032701.1	<i>Yucca brevifolia</i>	MW281833.1
<i>Hesperaloe campanulata</i>	NC_032702.1	<i>Yucca brevifolia x Yucca jaegeriana</i>	MW281856.1
<i>Hesperaloe parviflora</i>	NC_032703.1	<i>Yucca filamentosa</i>	NC_032712.1
<i>Hesperocallis undulata</i>	NC_032704.1	<i>Yucca gloriosa</i>	NC_086555.1
<i>Hesperoyucca whipplei</i>	NC_032705.1	<i>Yucca jaegeriana</i>	MW281819.1
<i>Hosta capitata</i>	MZ919305.1	<i>Yucca queretaroensis</i>	NC_032713.1
<i>Hosta clausa</i>	NC_046896.1	<i>Yucca schidigera</i>	NC_032714.1
<i>Hosta clausa var.</i>	OL628765.1	<i>Yucca treculeana</i>	NC_061319.1
<i>Hosta jonesii</i>	MZ919311.1	<i>Zephyranthes mesochloa(outgroup)</i>	NC_057554.1
<i>Hosta longipes</i>	NC_070392.1	<i>Asparagus cochinchinensis(outgroup)</i>	NC_060472.1

Table 2
Functional classification of genes in the chloroplast genome of *Agave* sp.

Category for gene	Group of genes	Name of genes
Self-replication	Large subunit of ribosome	<i>rpl20, rpl22, rpl32, rpl23(X2), rpl14, rpl33, rpl16, rpl36, rpl2(X2)</i>
	Small subunit of ribosome	<i>rps11, rps12(X3), rps14, rps15, rps16, rps3, rps18, rps4, rps19(X2), rps7(X2), rps8</i>
	DNA dependent RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	rRNA gene	<i>rrn5(X2), rrn4.5(X2), rrn16(X2), rrn23(X2)</i>
	tRNA gene	<i>trnR-UCU, trnE-UUC, trnT-GGU, trnS-GGA, trnI-CAU(X2), trnV-GAC(X2), trnR-ACG(X2), trnL-UAA, trnG-GCC, trnD-GUC, trnY-GUA, trnP-UGG, trnM-CAU, trnL-CAA(X2), trnS-GCU, trnW-CCA, trnF-GAA, trnT-UGU, trnS-UGA, trnV-UAC, trnG-UCC, trnL-UAG, trnI-GAU(X2), trnH-GUG(X2), trnM-CAU, trnQ-UUG, trnN-GUU(X2), trnK-UUU, trnA-UGC(X2), trnC-GCA</i>
Gene for photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>
	Subunits of photosystem II	<i>psbL, psbZ, psbM, psbN, psbA, psbB, psbC, psbD, psbE, psbF, psbT, psbH, psbI, psbJ, psbK</i>
	Subunits of NADH-dehydrogenase	<i>ndhG, ndhH, ndhI, ndhJ, ndhK, ndhA, ndhB(X2), ndhC, ndhD, ndhE, ndhF</i>
	Subunits of cytochrome b/f complex	<i>petL, petN, petA, petB, petD, petG</i>
	Subunit for ATP synthase	<i>atpI, atpA, atpB, atpE, atpF, atpH</i>
	Large subunit of rubisco	<i>rbcL</i>
Other genes	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
	Protease	<i>clpP</i>
	Envelope membrane protein	<i>cemA</i>
	Subunit of Acetyl-carboxylase	<i>accD</i>
	C-type cytochrome	<i>ccsA</i>

Category for gene	Group of genes	Name of genes
	synthesis gene	
	Open reading frames(ORF, <i>ycf</i>)	<i>ycf1</i> , <i>ycf2</i> (X2), <i>ycf3</i> , <i>ycf4</i>

Results and discussion

Chloroplast genome features

The *Agave* sp. cp genome is 157,489 bp in size and contains two IR regions (IRa and IRb; each 26,612 bp), separating the LSC (86,018 bp) and SSC (18,247 bp) regions, forming the quadripartite structure characteristic of angiosperms. It contains 113 unique genes, including 4 ribosomal rRNAs, 79 CDS, and 30 tRNA genes. We found 17 duplicated genes in the IR regions, including 4 rRNAs, 5 CDS, and 8 tRNA genes. The overall guanine-cytosine (GC) content of the chloroplast genome is 38.57%, with the GC content of the LSC, IR, and SSC regions being 35.9%, 43.1%, and 31.8%, respectively. Analysis of GC content by different gene types showed that rRNAs have a GC content of 55.4%, tRNAs 53.1%, and coding sequences 38.2%. The total genome size, the size of each region, and gene content are consistent with those reported for *Agave durangensis*, *Agave sisalana*, and *Agave amaniensis* (González-Trillo et al., 2024; Yang et al., 2021; Xu et al., 2022).

Relative synonymous codon usage

The codon preference of the chloroplast genome is directly related to the genome's AT/GC content and can reflect the AT-rich characteristics of the *Agave* chloroplast genome. The chloroplast genome of *Agave* sp. contains 79 protein-coding genes (PCGs). Relative synonymous codon usage (RSCU) was calculated, and the results (Fig. 2) are shown. These genes are encoded by 19,582 codons and exhibit highly conserved codon usage bias (CUB). Leucine (Leu) is the most frequently used amino acid, appearing 1,963 times, while cysteine (Cys) is the least common, occurring 215 times. RSCU analysis shows that AGU has the highest RSCU value at 3.43, while UCG has the lowest at 0.10. Among these codons, 29 have RSCU values greater than 1, and 27 have values less than 1. Additionally, the codon for methionine (Met), AUG, and the codon for tryptophan (Trp), UGG, both have an RSCU value of 1 (Wang et al., 2023).

Phylogenetic reconstruction

The analysis of the complete chloroplast genome sequence constructed a maximum likelihood phylogenetic tree with strong support. Phylogenetic analysis indicates that this tree has four major clades, each with a bootstrap value of 100%. The genera *Yucca* and *Hosta* cluster together on a separate branch, representing a relatively independent evolutionary line within the Asparagaceae, forming a typical monophyletic group, which was consistent with previous morphological and molecular studies (Bogler et

al., 2006; Xu et al., 2022). *Camassia scilloides* and *Chlorogalum pomeridianum* are sister species, forming a paraphyletic relationship with *Hesperocallis undulata*, *Hesperoyucca whipplei*, *Schoenolirion croceum*, and the *Hesperaloe* genus, which aligns with previous research findings (Yang et al., 2021). *Agave* sp. is closely related to *Agave attenuata*, forming a sister group. The traditionally defined *Agave sensu stricto* is not monophyletic and diverged over a short period. *Agave sensu lato* is paraphyletic, including *Polianthes* and *Manfreda virginica*. This is consistent with phylogeny from nuclear *ITS* gene (Jiménez-Barron et al., 2020). The phylogenetic tree clearly illustrates the evolutionary branching relationships of the genera *Agave*, *Yucca*, *Hosta*, and other closely related groups within the Asparagaceae.

Conclusion

This study presents the complete chloroplast genome sequence of *Agave* sp., with a cp genome size of 157,489 bp, featuring a typical quadripartite structure and 113 unique genes, similar to other agave species. Our chloroplast phylogenetic analysis of 42 species supports the paraphyly within the broad genus *Agave* while *Yucca*, *Agave sensu lato*, and *Hosta* are well-supported monophyletic groups. Our results provide a valuable insight into chloroplast evolution within the family Asparagaceae.

Declarations

Competing interests

The authors declare no competing interests.

Author Contribution

Xinya Ma: Conceptualization, Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing. Xionghui Xu: Methodology, Investigation. Chaoyang Luo: Conceptualization. Zhenyu Xiong: Methodology, Visualization. Juan Zhang: Visualization. Yuan Mu: Conceptualization, Methodology, Investigation, Project administration, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

Data Availability

The dataset generated and analyzed in this study is available in the NCBI-Nucleotide database. For the biological samples included in this study (with the study-generated identifier: PX665995), the relevant accessions are: NC_032696.1, NC_058311.1, NC_032053.1, NC_059876.1, NC_070107.1, NC_072313.1, NC_059874.1, NC_045534.1, NC_059875.1, NC_032699.1, NC_032700.1, NC_032701.1, NC_032702.1, NC_032703.1, NC_032704.1, NC_032705.1, MZ919305.1, NC_046896.1, OL628765.1, MZ919311.1,

NC_070392.1, NC_035999.1, NC_053555.1, NC_060662.1, NC_060664.1, NC_060663.1, NC_032706.1, NC_046895.1, MZ919315.1, NC_032707.1, KX931464.1, NC_032710.1, MW281833.1, MW281856.1, NC_032712.1, NC_086555.1, MW281819.1, NC_032713.1, NC_032714.1, NC_061319.1, NC_057554.1, NC_060472.1

References

1. Eguiarte LE, Jiménez Barrón OA, Aguirre-Planter E et al (2021) Evolutionary ecology of Agave: distribution patterns, phylogeny, and coevolution (an homage to Howard S. Gentry). *Am J Bot* 108(2):216–235
2. González-Trillo AC, Reyes López MÁ, Almaraz-Abarca N et al 2024.Characterization of the complete chloroplast genome sequence of *Agave durangensis* (Asparagales: Asparagaceae: Agavoideae). *Mitochondrial DNA Part B* 9 (4): 536–540
3. Torres I, Casas A, Vega E et al 2015.Population dynamics and sustainable management of mescal agaves in central Mexico: *Agave potatorum* in the Tehuacán-Cuicatlán Valley[J]. *Econ Bot* 69 (1): 26–41
4. Bogler DJ, Pires JC, Francisco-Ortega J (2006) Phylogeny of Agavaceae Based on *ndhF*, *rbcL*, and its Sequences. *Aliso: A Journal of Systematic and Floristic Botany* 22(1) : 313–328
5. Good-Avila SV, Souza V, Gaut BS, Eguiarte LE (2006) Timing and rate of speciation in *Agave* (Agavaceae). *Proc Natl Acad Sci USA* 103(24):9124–9129
6. Flores-Abreu IN, Trejo-Salazar RE, Sánchez-Reyes LL et al (2019) Tempo and mode in coevolution of *Agave sensu lato* (Agavoideae, Asparagaceae) and its bat pollinators, Glossophaginae (Phyllostomidae). *Mol Phylogenet Evol* 133:176–188
7. Jiménez-Barron O, García-Sandoval R, Magallón S et al (2020) Phylogeny, Diversification Rate, and Divergence Time of *Agave sensu lato* (Asparagaceae), a Group of Recent Origin in the Process of Diversification. *Front Plant Sci* 11:536135
8. Yang X, Huang X, Tan S et al (2021) The complete chloroplast genome of *Agave sisalana*. *Mitochondrial DNA Part B* 6(7):1855–1856
9. Xu B, Tan S, Qin X et al (2022) The complete chloroplast genome of *Agave amaniensis* (Asparagales: Asparagaceae: Agavoideae). *Mitochondrial DNA Part B* 7(8):1519–1521
10. Selvaraj D, Sarma RK, Sathishkumar R (2008) Phylogenetic analysis of chloroplast *matK* gene from Zingiberaceae for plant DNA barcoding. *Bioinformation* 3(1):24
11. Hang LM, Zhang ZP, Zhang XH et al (2025) Comparative chloroplast genome analyses of *potentilleae*: insights into genome characteristics, mutational hotspots, and adaptive evolution. *Genetica* 153(1):22
12. Li QQ, Zhang ZP, Aogan et al (2024) Comparative chloroplast genomes of *Argentina* species: genome evolution and phylogenomic implications. *Front Plant Sci* 15:1349358

13. Wu L, Nie L, Wang Q et al (2021) Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae. *Sci Rep* 11(1):14643
14. Ma D, Tian Q, Wang Y et al (2025) Comparative chloroplast genome of six species in Hypoxidaceae from China: insights into phylogenetic relationships and molecular marker development. *BMC Plant Biol* 25(1):1232
15. Yang L, Abduraimov O, Tojibaev K et al (2022) Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of *Ferula*. *BMC Genomics* 23(1):643
16. Sharp PM, Li WH (1987) The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15(3):1281–1295
17. Zhang D, Gao F, Jakovlić I et al (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour* 20(1):348–355
18. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780
19. Ranwez V, Douzery EJP, Cambon C et al (2018) MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol* 35(10):2582–2584
20. Kalyaanamoorthy S, Minh BQ, Wong TKF (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* (6): 587–589
21. Nguyen LT, Schmidt HA, von Haeseler A et al 20145. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* (1): 268–274
22. Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 52(W1):W78–W82
23. Wang Y, Jiang D, Guo K et al (2023) Comparative analysis of codon usage patterns in chloroplast genomes of ten *Epimedium* species. *BMC Genom Data* 24(1):3

Figures

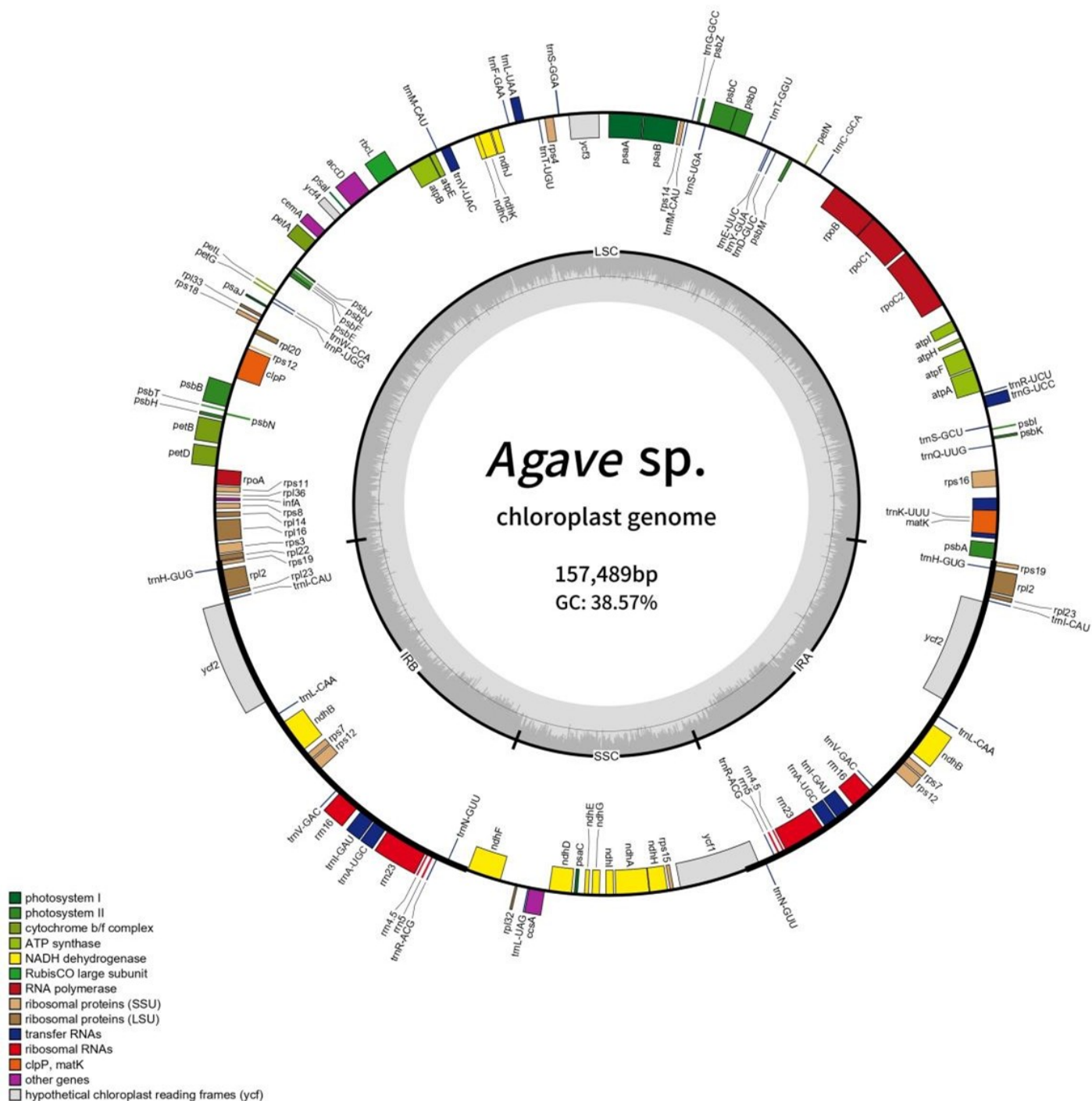


Figure 1

Circular map of the chloroplast genome of *Agave sp.* Genes are color coded according to their functions. Genes located on the outer edge transcribe in an anticlockwise direction, whereas those on the inner side transcribe clockwise.

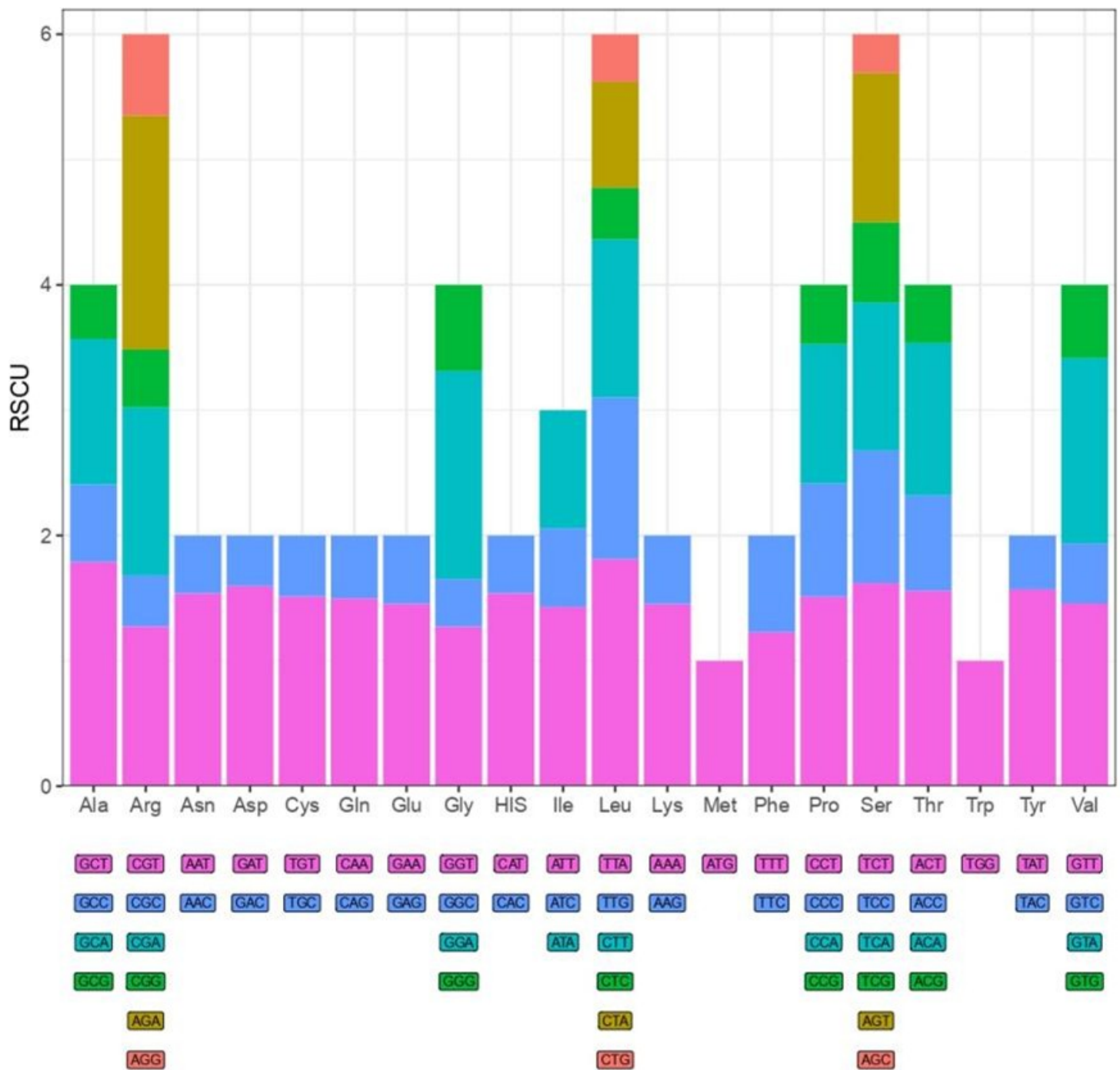


Figure 2

The Bar chart of codon usage for amino acids of 79 protein-coding genes in the chloroplast genome of *Agave* sp.

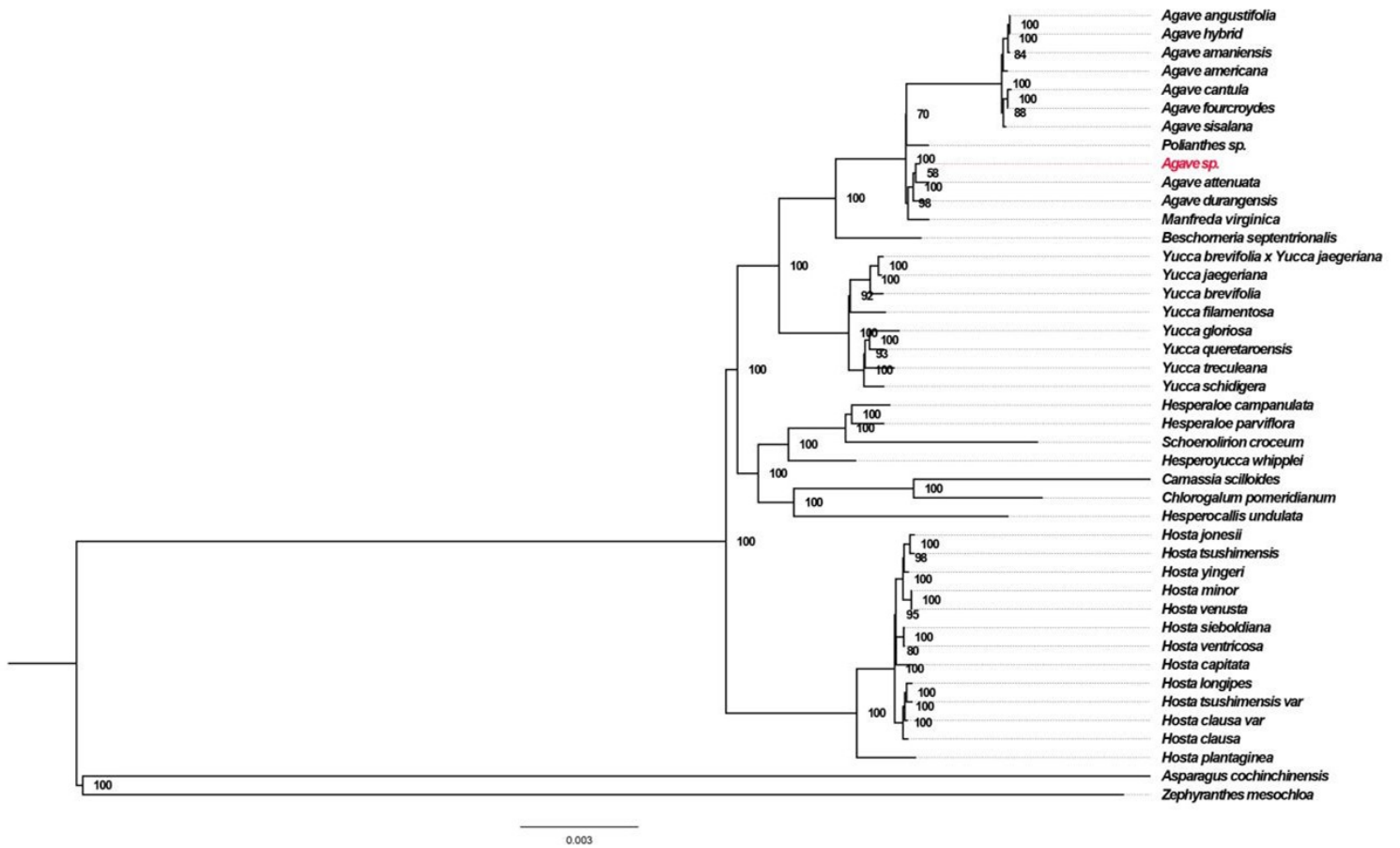


Figure 3

Phylogenetic tree constructed using Maximum Likelihood (ML) method based on the complete chloroplast genome sequences of 42 Asparagaceae species.