https://doi.org/10.61308/ZOJA3530

Sequence Analysis of some species in the genus *Amaranthus* L. using Ribulose-1.5 - bisphosphate carboxylase/oxygenase (rbcL) genes

Abiodun Sunday Oyelakin[*], Sarah Ifeoma Ossai, Oluwasegun Olamide Fawibe, Oluwaseun Faith Akinyemi, Elizabeth Tolulope Obiwole, Joy Ogeyingbo, Comfort Obaba
Department of Pure and Applied Botany, Federal University of Agriculture, Abeokuta, P.M.B. 2240, Abeokuta, Ogun State, Nigeria.
*E-mail: *oyelakinas@funaab.edu.ng*

Citation: Oyelakin A. S., Ossai, S. I., Fawibe, O. O., Akinyemi, O. F., Obiwole, E. T., Ogeyingbo, J. & Obaba, C. (2024). Sequence Analysis of some species in the genus Amaranthus L. using Ribulose-1.5 - bisphosphate carboxylase/oxygenase (rbcL) genes. *Bularian Journal of Crop Science, 61*(2) 96-103

Abstract: Bioinformatics is a novel method to investigate evolutionary relationships and genetic diversity among species. This study assessed the evolutionary patterns and genetic diversity of *Amaranthus* species using ribulose-1.5-bisphosphate carboxylase/oxygenase (rbcL gene). A total of 20 full FASTA-format rbcL gene sequences for accessions of *Amaranthus tricolor*, *A. dubius*, *A. cruentus*, *A. viridis*, and *A. spinosus* from different regions were retrieved from the National Center for Biotechnology Information. Multiple sequence alignment was used to determine genetic variations. A phylogenetic tree was constructed using neighbor-joining clustering method while a distant matrix was created using Tamura-Nei genetic distance model. From position 1 to 22 in nucleotide alignment, only accession KX082771.1 from USA, aligns. The exclusive alignment in this region suggests an exceptionally conserved sequence among species, implying a common genetic origin. However, from position 23 to 50, nine different accessions namely MN192576.1, MN204800.1, EF590496.1, GQ248546.1, OQ737000.1, MG246937.1, MN104855.1, MN104858.1, and MN04859.1 align. The phylogenetic tree representing the evolutionary relationships offers insights into the genetic diversity and relatedness of these species within specific regions. The genetic distance matrix shows that sequences within the same region share relatively high similarity scores, indicating a genetic connection likely influenced by their geographic proximity. However, sequences from different regions, exhibit moderate genetic differences, emphasizing the genetic diversity across geographically distinct regions. The observed genetic diversity and evolutionary relationships among accessions of *Amaranthus* revealed by rbcL genes highlight the unique genetic features of these species in different regions, and therefore should be explored for their genetic improvement and conservation purposes.

Keywords: Bioinformatics; *Amaranthus*; rbcL; phylogenetic study; genetic diversity; evolution

INTRODUCTION

Members of the genus *Amaranthus* were traditionally classified by experienced experts based on morphological characteristics. However, some species were reported to be wrongly classified during the early period of taxonomy due to many reasons. These include faulty techniques or misidentification as recent taxonomical evidences revealed that many species of *Amaranthus* are uncultivated, underutilized, and neglected, therefore, facing classification as weeds. As a result, a quick and accurate analytical method to study the evolutionary relationship of some species in the genus *Amaranthus* is required for their correct classification and conservation. Bioinformatics is a novel method for species identification that can overcome the aforementioned challenges. It is an interdisciplinary field that plays a critical role in modern life sciences research by providing tools and techniques to handle the massive amounts of biological data generated from "omics" technologies (Pennisi, 2017). It involves various aspects such as sequence analysis, structural bioinformatics, and phylogenetics (Lesk et al, 2016). Genomics is the study of an organism's entire genome, with the intent to understand the structure, function, evolution, and regulation of the organism's genome (Brown & White, 2021). Furthermore, it is independent of morphological characteristics and expert experiences, thus allowing for faster, more subjective, and more accurate results. As a result, bioinformatics has quickly become the focus of attention for many researchers, hence it has been recognized as a powerful tool for species identification, genetic diversity, and tracing of evolutionary relationship among organisms through the construction of phylogenetic trees (Hajibabaei et al., 2005; Hebert & Gregory, 2005).

This genus *Amaranthus* L. in the family Amaranthaceae consists of 70-75 species that are widely distributed throughout tropical and temperate regions. Several species in the genus are classified as neglected and underutilized, but *Amaranthus tricolor* L., *A. dubius* Mart. Ex Thell., *A. cruentus* L., *A. viridis* L., and *A. spinosus* L. are reported to be popular among the cultivated species. These species are known for their adaptability to different environmental conditions. They offer nutritional benefits, being rich in vitamins (A, C, K), high in minerals (Ca, Fe, Mg), and are good sources of antioxidants. They make substantial contributions to local diets and agriculture in the regions where they are prevalent.

Investigating the genetic diversity and evolutionary relationships among *Amaranthus* species from different regions in the world would play a significant role in the conservation efforts of the members of the genus. This is because protecting and preserving genetic diversity is essential for ensuring the long-term survival of these species, especially in the face of environmental changes and habitat loss. Understanding these genetic adaptations would contribute to conservation efforts and food security by enhancing the agricultural systems.

Currently, rbcL genes are widely used for phylogenetic analysis within the angiosperm families, and among seed plant groups because they provide more phylogenetic information and improved phylogenetic resolution among the studied taxa (Chase et al., 2007; Hardig et al., 2010).

Therefore, this present study was initiated to assess the evolutionary patterns and genetic diversity among *Amaranthus tricolor*, *A. dubius*, *A. cruentus*, *A. viridis*, and *A. spinosus* using gene sequences of Ribulose-1.5-bisphosphate carboxylase/oxygenase (*rbcL*) obtained from different regions of the world.

MATERIALS AND METHODS

Gene Sequence Retrieval and Data Preparation for *Amaranthus* species

Sequences of rbcL genes for accessions of different *Amaranthus* species were obtained from the National Center for Biotechnology Information (NCBI). This consists of different specimens from multiple regions, including China, Pakistan, South Africa, USA, Canada, India, Bahrain, and Sri Lanka. These sequences were retrieved from various geographic regions to ensure a broad representation.

Data Organization by Gene and Regions

A total of 20 full FASTA-format rbcL gene sequences were downloaded from the NCBI database to conduct an evolutionary analysis and assess the genetic diversity of the accessions of the species. The sequence information was organized into tables to categorize the genetic material according to the geographic regions from which they were collected.

Multiple Sequence Alignment for *Amaranthus* Species

The retrieved rbcL gene sequences were uploaded to the Geneious software. Due to its efficiency and precision throughout the alignment process, the muscular Alignment tool on Geneious software v9.0.5 was used. The alignment parameters were configured to execute a local alignment with eight iterations. Gaps and mismatches were added throughout the whole alignment procedure to guarantee appropriate alignment of the sequences. The output file, which contained the aligned sequences and the appropriate gap positions, was produced when the alignment was completed.

Phylogenetic Tree Construction and Analysis for *Amaranthus* species

The neighbor-joining clustering method was used to construct a phylogenetic tree to access the genetic distance and evolutionary link as well as identify divergence among the investigated accessions. This method allowed for the creation of clusters that reflected the genetic similarities and differences among the sequences, thereby facilitating the identification of evolutionary trends.

Distant matrix Construction for *Amaranthus* species

A distant matrix was created using Tamura-Nei genetic distance model to analyze the genetic differences over time. To ensure the reliability of the results, the Neighbor-Joining algorithm was run multiple times, using a bootstrap value of 1000. This helped capture the complex and ever-changing nature of the evolutionary process.

Statistical Analysis

In combination, the distance matrix, Neighbor-Joining clustering, and Excel models provided a robust framework for statistical analysis within Geneious platform (https://www.geneious.com). This approach not only verified the divergence among the analyzed accessions but also offered insights into the genetic distances and evolutionary relationships that underline the sequences, ultimately determining their crucial importance within the context of the study.

RESULTS

Comparative Sequence Analysis of Ribulose-1.5-bisphosphate carboxylase/oxygenase (*rbcL*) genes for *Amaranthus* species

Sequences of *rbcL* genes from various *Amaranthus* species obtained from the NCBI exhibited variations in sequence lengths and GC content as shown in Table 1. This points to the genetic diversity within the genus *Amaranthus*.

The result of multiple sequence alignments using *rbcL* gene sequences to investigate the evolutionary relationships across diverse *Amaranthus* species is shown in Figure 1. The multiple sequence alignment for 20 *rbcL* gene sequences displayed an overall length of 1,741 base pairs. Among these sequences, 251 sites were identical, representing 21.9% identity among the aligned sequences, with a pairwise percentage identity of 53.1%. An analysis of the ungapped lengths of these 20 sequences showed a mean length of 695.0 base pairs, with a standard deviation of 178.9 base pairs. Sequence lengths ranged from a minimum of 544 base pairs to a maximum of 1,240 base pairs. Base pair frequency analysis indicated Adenine (A) at 29.9% (4,157 occurrences), Cytosine (C) at 19.4% (2,699 occurrences), Guanine (G) at 20.5% (2,848 occurrences), and Thymine (T) at 30.2% (4,195 occurrences). The overall GC content of the alignment was computed at 39.9%, totaling 5,547 base pairs. Furthermore, the molecular weight (mean) of the alignment process was calculated, revealing a single-stranded DNA (ssDNA) with an average weight of 214.811 kDa and a double-stranded DNA (dsDNA) with an average weight of 429.341 kDa.

In the nucleotide alignment of various *Amaranthus* species, distinct patterns emerge at different positions. From position 1 to 22, only one species, represented by accession KX082771.1 from the Pakistan, aligns. The exclusive alignment in this region suggests an exceptionally conserved sequence among these species, implying a common genetic origin. Moving to position 23 to 50, nine different accessions/species align, including MN192576.1, MN204800.1, EF590496.1, GQ248546.1, OQ737000.1, MG246937.1, MN104855.1, MN104858.1, and MN04859.1

Evolutionary Analysis of *Amaranthus* species using Phylogenetic Tree based on *rbcL* genes

The phylogenetic tree representing the evolutionary relationships among different *Amaran-*

Table 1. *rbcL* Gene Sequences for *Amaranthus* Species

S/N	Sample ID	Taxon ID	Sequence Length (bp)	GC Content (%)	Region	Specimen Voucher
1	MF135386.1	117272	1240	43.5	China	SZ 201509205
2	JF940804.1	124765	603	43.3	China	Z22
3	JF940803.1	124765	603	43.3	China	Z23
4	JF940802.1	124765	603	43.3	China	Z24
5	OQ737000.1	1193840	573	43.1	India	
6	EF590496.1	124765	544	42.6	USA	NMNH Wells 4469
7	KX079612.1	1193840	608	42.6	South Africa	
8	KX079607.1	1193840	608	42.6	South Africa	
9	KX079606.1	117272	608	42.6	South Africa	
10	KX079600.1	1193840	608	42.6	South Africa	
11	KX079588.1	1193840	608	42.6	South Africa	
12	MG246937.1	117272	552	42.6	Canada	CCDB-18297-C11
13	MK097156.1	56196	553	42.5	Bahrain	BAHMP032040117
14	GQ248546.1	124765	553	42.5	Canada	NMNH Wells 4469
15	MN204800.1	56196	720	42.4	China	TangXS0138
16	MN192576.1	56196	723	42.3	China	NJWLS501
17	KX082771.1	56196	1005	31.7	Pakistan	
18	MN104858.1	117272	862	31.3	Sri Lanka	DMB195
19	MN104859.1	117272	862	31.2	Sri Lanka	DMB196
20	MN104855.1	56196	863	30.8	Sri Lanka	DMB192

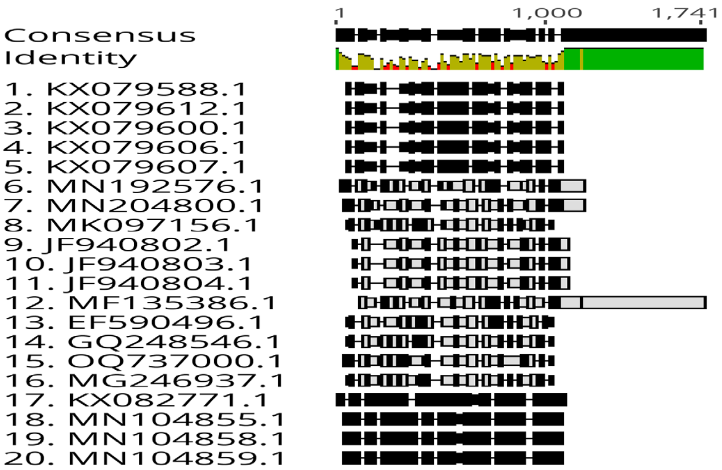


Figure 1. *rbcL* Nucleotide Alignment of *Amaranthus* Species

thus species based on *rbcL* gene sequences offers valuable insights into their genetic diversity and relatedness within specific country regions as depicted in Figure 2. Individual branches within the tree, such as MN104858.1 and MN104859.1 indicated distinct *Amaranthus* species, with underline species-level genetic variation. The presence of a prominent red branch, led by MN104855.1

with a bootstrap value of 53, signifies an ancient divergence point in the evolutionary history of these species, specifically within Sri Lanka. This ancient divergence suggests that MN104855.1's lineage has undergone substantial genetic differentiation over time in Sri Lankan context. There green branch with a higher bootstrap value of 65 which connect to the red branch signifies a more recent common ancestor among species within the South African region. This suggests a closer genetic connection among South African *Amaranthus* species, and indicates recent evolutionary events within this geographical area.

Furthermore, the striking blue branch characterized by a strong bootstrap value of 100 represents a distinctive subgroup within the genus, possibly comprising a diverse set of species from various regions. The high bootstrap value indicates the high confidence in the genetic distinction of this subgroup. The phylogenetic tree also

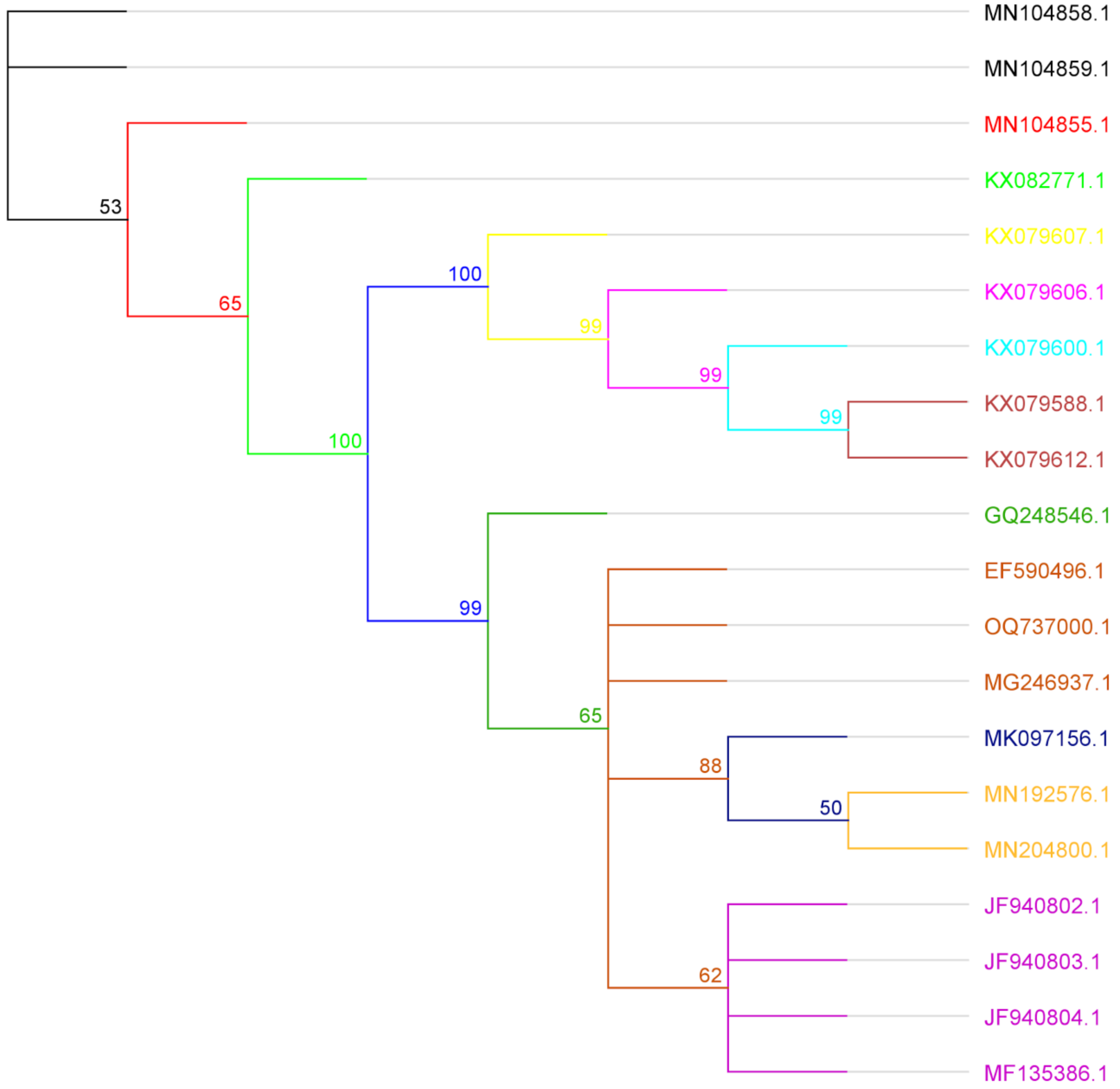


Figure 2. Phylogenetic tree of *Amaranthus* species based on *rbcL* genes

classifies species into clusters groups, within which species are denoted by colors like brown, sky blue, pink, and yellow, possibly indicative of their geographical origins. The brown branch, represented by KX079588.1 exhibits a close genetic relationship with the sky blue branch, indicated by KX079600.1. The high bootstrap value of 99 connecting these two branches suggests a very close genetic connection between species originating from South Africa and suggests a recent shared ancestry.

Furthermore, the sky blue branch connects to the pink branch with accession KX079606.1 and a bootstrap value of 99, suggests a shared genetic lineage among accessions of species within the South African region. Also, the pink branch connects to the yellow branch by KX079607.1 with a high bootstrap value of 100 signifies an exceptionally strong genetic relationship within this cluster, and possibly indicates a recent common ancestor among South African species. The high bootstrap values across the branches highlight the confidence in the genetic relatedness among species from this specific region and emphasize a recent evolutionary divergence within Cluster 1, suggesting the possibility of dynamic evolutionary history within the South African context. Also, Cluster 2 demonstrates similarly strong bootstrap values, suggesting distinct genetic lineage and relationships among species from other geographic regions such as Canada, China, and Sri Lanka. These strong bootstrap values highlight the genetic diversity and evolutionary dynamics among *Amaranthus* species in various country regions, and it offers valuable insights into their evolutionary history and relationships.

Distance Matrix for *Amaranthus* Species using *rbcL* genes

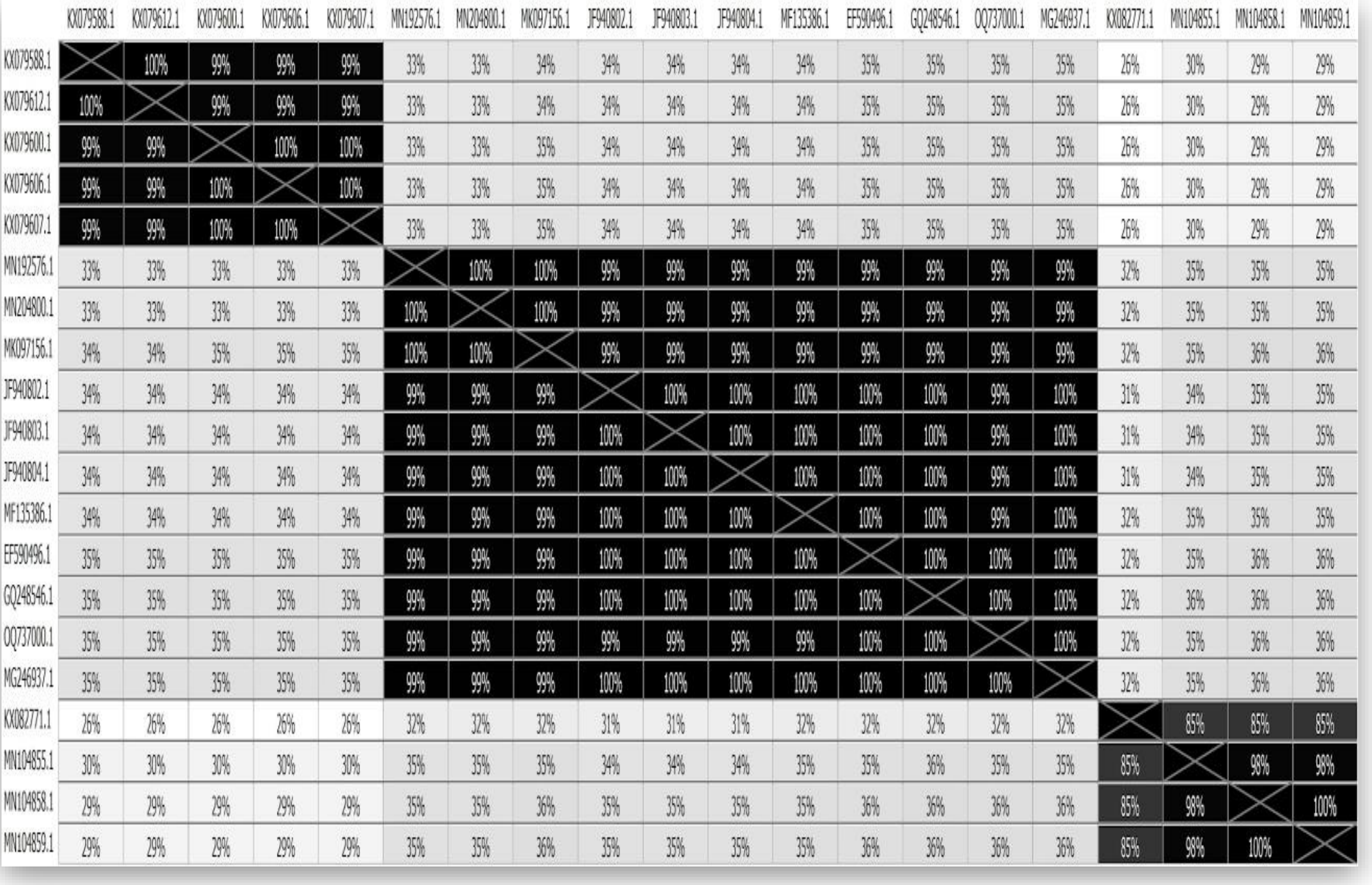
The genetic distance matrix provides insights into the genetic relationships among various species of *Amaranthus*, considering their genetic diversity across regions as presented in Figure 3. Sequences of accessions within the same region, such as KX079588.1, KX079612.1, KX079600.1, KX079606.1, and KX079607.1 share relatively high similarity scores, which suggest that a genetic connection is likely influenced by their geographic proximity. These sequences exhibit similarity scores ranging from 99 to 100 and this suggests the close genetic relationship among them. However, moderate genetic differences with similarity scores of 85 and 100 exhibited by sequences of accessions from the same region, like accession MN104858.1 and MN104859.1 from Sri Lanka suggests the genetic diversity across geographically distinct regions. This matrix, when considering the regions, underscores the diverse genetic relationships among various *Amaranthus* species, with variations likely attributed to both evolutionary processes and regional influences.

DISCUSSION

The extensive collection of rbcL gene sequences revealed significant genetic diversity and evolutionary relationships among the various *Amaranthus* species. The presence of variations in sequence lengths and GC content among these sequences suggests the diverse genetic makeup in the genus *Amaranthus*. This diversity is crucial in understanding how different species within the genus have evolved and adapted to various environmental condition.

The multiple sequence alignments reveal both similarities and differences among these species. The presence of multiple species aligning signifies a mix of conserved genetic elements and variations. These observed variations of the rbcL gene in conserved regions among *Amaranthus* species indicates shared genetic heritage while variations at specific positions indicate genetic diversity.

The application of phylogenetic trees as a powerful tool to unravel the genetic diversity and evolutionary relationships in the genus *Amaranthus* cannot be overemphasized because it unveils the diversification of the species. The distinct color-coded branches in the phylogenetic tree indicate regional genetic lineages, ancient divergences, and shared ancestry. The presence of ancestral and recent genetic variants, as exemplified by sequences with high bootstrap values



	KX079588.1	KX079612.1	KX079600.1	KX079606.1	KX079607.1	MN192576.1	MN204800.1	MK097156.1	JF940802.1	JF940803.1	JF940804.1	MF135386.1	EF590496.1	GQ248546.1	OQ737000.1	MG246937.1	KX082771.1	MN104855.1	MN104858.1	MN104859.1
KX079588.1		100%	99%	99%	99%	33%	33%	34%	34%	34%	34%	34%	35%	35%	35%	35%	26%	30%	29%	29%
KX079612.1	100%		99%	99%	99%	33%	33%	34%	34%	34%	34%	34%	35%	35%	35%	35%	26%	30%	29%	29%
KX079600.1	99%	99%		100%	100%	33%	33%	35%	34%	34%	34%	34%	35%	35%	35%	35%	26%	30%	29%	29%
KX079606.1	99%	99%	100%		100%	33%	33%	35%	34%	34%	34%	34%	35%	35%	35%	35%	26%	30%	29%	29%
KX079607.1	99%	99%	100%	100%		33%	33%	35%	34%	34%	34%	34%	35%	35%	35%	35%	26%	30%	29%	29%
MN192576.1	33%	33%	33%	33%	33%		100%	100%	99%	99%	99%	99%	99%	99%	99%	99%	32%	35%	35%	35%
MN204800.1	33%	33%	33%	33%	33%	100%		100%	99%	99%	99%	99%	99%	99%	99%	99%	32%	35%	35%	35%
MK097156.1	34%	34%	35%	35%	35%	100%	100%		99%	99%	99%	99%	99%	99%	99%	99%	32%	35%	36%	36%
JF940802.1	34%	34%	34%	34%	34%	99%	99%	99%		100%	100%	100%	100%	100%	99%	100%	31%	34%	35%	35%
JF940803.1	34%	34%	34%	34%	34%	99%	99%	99%	100%		100%	100%	100%	100%	99%	100%	31%	34%	35%	35%
JF940804.1	34%	34%	34%	34%	34%	99%	99%	99%	100%	100%		100%	100%	100%	99%	100%	31%	34%	35%	35%
MF135386.1	34%	34%	34%	34%	34%	99%	99%	99%	100%	100%	100%		100%	100%	99%	100%	32%	35%	35%	35%
EF590496.1	35%	35%	35%	35%	35%	99%	99%	99%	100%	100%	100%	100%		100%	100%	100%	32%	35%	36%	36%
GQ248546.1	35%	35%	35%	35%	35%	99%	99%	99%	100%	100%	100%	100%	100%		100%	100%	32%	36%	36%	36%
OQ737000.1	35%	35%	35%	35%	35%	99%	99%	99%	99%	99%	99%	99%	100%	100%		100%	32%	35%	36%	36%
MG246937.1	35%	35%	35%	35%	35%	99%	99%	99%	100%	100%	100%	100%	100%	100%	100%		32%	35%	36%	36%
KX082771.1	26%	26%	26%	26%	26%	32%	32%	32%	31%	31%	31%	32%	32%	32%	32%	32%		85%	85%	85%
MN104855.1	30%	30%	30%	30%	30%	35%	35%	35%	34%	34%	34%	35%	35%	36%	35%	35%	85%		98%	98%
MN104858.1	29%	29%	29%	29%	29%	35%	35%	36%	35%	35%	35%	35%	36%	36%	36%	36%	85%	98%		100%
MN104859.1	29%	29%	29%	29%	29%	35%	35%	36%	35%	35%	35%	35%	36%	36%	36%	36%	85%	98%	100%	

Figure 3. Distant matrix for *Amaranthus* species using *rbcL* genes

highlight the genetic diversity and evolutionary dynamics among *Amaranthus* species in various regions, thus, offering valuable insights into their evolutionary history and relationships. The robust clustering of species in groups, such as Cluster 1 and Cluster 2, provides insights into shared evolutionary history among South African and other species from various geographical regions. These phylogenetic patterns support the significance of genetic diversity and its role in evolutionary dynamics. These observations are in line with the studies conducted by Lee et al. (2021) and Davis et al. (2022).

The genetic distance matrix analysis emphasized the impact of regional factors on genetic relationships as evident in the sequences from the same region exhibiting relatively high similarity scores, while sequences from different regions displayed lower similarity scores. This indicates the influence of geographic factors on genetic makeup and evolutionary connections within the genus *Amaranthus*. The observed genetic diversity and evolutionary relationships among accessions of *Amaranthus* species in this study highlights the unique genetic features of the species in different regions, hence should be explored for conservation purposes.

CONCLUSION

The observed genetic diversity and evolutionary relationships among accessions of *Amaranthus tricolor*, *A. dubius*, *A. cruentus*, *A. viridis*, and *A. spinosus* revealed by ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*) genes highlight the unique genetic features of these species in different regions, and therefore should be explored for their genetic improvement and conservation purposes.

CONFLICT OF INTERST

There is no conflict of interest among the authors.

REFERENCES

Brown, E. F. & White, S. D. (2021). Genetic Relationships among South African *Amaranthus* Species. *South African Journal of Botany*, *19*(4), 271-288.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., Van Den Berg, C., Madriñán, S., Petersen, G., ... & Wilkinson, M. (2007). A proposal for a standardized protocol to barcode all land plants. *Taxon*, *56*(2), 295-299.

Davis, P. W., Lee, S. A., & Smith, A. B. (2022). Insights into *Amaranthus* Genetic Diversity from *rbcL* Phylogenetic Tree Analysis. *Plant Biology International*, *50*(6), 782-798.

Hajibabaei, M., DeWaard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., ... & Hebert, P. D. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1959-1967.

Hardig, T. M., Anttila, C. K., & Brunsfeld, S. J. (2010). A phylogenetic analysis of Salix (Salicaceae) based on matK and ribosomal DNA sequence data. *Journal of Botany.*

Hebert, P. D., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic biology*, *54*(5), 852-859.

Lee, S. A., Davis, P. W., & Smith, A. B. (2021). Phylogeographic Analysis of *Amaranthus* Species Using *rbcL* Gene Sequences. *International Journal of Plant Science*, *42*(1), 78-94.

Lesk, C., Rowhani, P. & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, *529*(7584), 84-87.

Pennisi, E. (2017). Sequencing plant genomes. *Science*, *355*(6323), 545-547.

Received: January 17 2024; *Approved:* March 28 2024; *Published:* April 2024