PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.3317

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

In-silico approaches for discrimination of *Curcuma* species and their closely related family using the novel technique of DNA Barcoding

Vinod Kumar Sahu[1] *, Keerti Tantwai[1], Sharad Tiwari[1], Swapnil Sapre[1], Nishi Mishra[1] & Shobha Sondhia[2]

[1] Biotechnology Centre, Jawaharlal Nehru Krishi Vishwa Vidyalaya, Jabalpur 482004, MP, India

[2] Directorate of Weed Research, Jabalpur 482004, MP, India

*Email: vinodsahu084@gmail.com

OPEN ACCESS

ARTICLE HISTORY

Received: 27 January 2024
Accepted: 16 June 2024

Available online
Version 1.0 : 17 July 2024

Check for updates

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE

Sahu VK , Tantwai K , Tiwari S, Sapre S, Mishra N, Sondhia S. In-silico approaches for discrimination of Curcuma species and their closely related family using the novel technique of DNA Barcoding. Plant Science Today (Early Access). https:/doi.org/10.14719/pst.3317

Abstract

In this study, we have discriminated and identified the Genus of *Curcuma* and related species Zingiberaceae using *rbcL* and *trnL* DNA barcode primers. *Curcuma* genus related to the family Zingiberaceae comprises a significant number of medicinal plants renowned for their use in ethnomedicine, playing a pivotal role in the medical, health, and pharmaceutical sectors. Traditionally, morphological methods alone have proven insufficient for accurately identifying species within this family. However, DNA barcoding technology provides a contemporary solution by utilizing plant DNA sequences for species identification, thus enabling effective conservation efforts. We used DNA barcoding techniques and for analysis used the Maximum Parsimony tree in MEGA 11 with the Kimura 2-parameter (KP model) to analyse the genetic relationships between species. Out of the 13 accessions that were studied, 12 accessions belonged to *Curcuma caesia* and 1 accession belonged to *Curcuma aeruginosa*. The genetic relationships observed were correlated with the geographical distributions of these species. It was determined that *C. aeruginosa* is a mutated species of *C. caesia*. Additionally, 1 specimen of *Alpinia galanga*, a plant species related to the Zingiberaceae. Barcode primer *trnL* primer demonstrated a 92% efficiency during the investigation. The *rbcL* and *trnL* loci are recommended as potential barcode markers for discriminating between different plant species. This study developed a comprehensive DNA barcoding database that can confidently differentiate between species by combining morphological and molecular data. This database has the potential to identify adulteration in herbal products, combat illegal trade and adulteration of plant species, and assist in germplasm conservation efforts.

Keywords

Curcuma; Kali Haldi; DNA Barcoding; PCR; Primers; *Zingiberaceae*; *rbcL*; *trnL*

Introduction

The genus *Curcuma*, belonging to the Zingiberaceae family, is widely distributed across tropical and subtropical regions of Asia, including India, China, Myanmar, Bangladesh, and Thailand. It comprises approximately 120 species, with more than 50 species found in India. *Curcuma* species exhibit diverse morphology in their rhizomes, flowers, and fruits. Local and tribal communities have recognized the medicinal properties of these plants for centuries (1).In India, *Curcuma* holds significant cultural and traditional medicinal importance, with turmeric being an integral part of religious rituals and traditional healing practices (2). Various species of *Curcuma* have

been utilized in traditional methods to treat illnesses such as migraines, body pain, liver disorders, skin diseases, and diabetes. A bioactive compound such as curcumin has been identified as a key component in *Curcuma* species, and the essential oils derived from these plants contain compounds for instance turmerone, zingiberene, and sesquiterpenes (3,4,5,6). Identifying Zingiberaceae plants based on morphology is challenging due to their complex classification and closely related species. (7,8,9). Hybridization and polyploidization have also played a role in creating taxonomic uncertainties in the *Curcuma* genus (10,11). Traditional identification methods heavily rely on the expertise of taxonomists, and the developmental stage of the plants and can be influenced by environmental factors. DNA barcoding has emerged as a powerful tool in biotechnology for species identification, utilizing the specificity of DNA sequences located in the plastid, mitochondria, and nucleus. To address the need for a reliable method to identify *Curcuma* and related plant species, therefore this study focuses on DNA barcoding techniques using the *rbcL* and *trnL* barcode regions. These barcode regions have been widely utilized in DNA barcoding studies for plant identification. This investigation aims to assess the efficacy of barcode regions in identifying closely related *Curcuma* species, as well as the degree of variation present.

Materials and Methods

Collection of Plant Materials

This study included a total of 13 specimens. These samples were obtained from various locations in India, as listed in Table 1. The plants were collected in their natural state and carefully preserved to minimize DNA damage. The plant species that were collected, including individuals, branches, or tissues, were wrapped meticulously in aluminum foil, sealed in plastic bags, and safely stored at a temperature of -80°C in a deep freezer. This method of storage was chosen to reduce the degradation of DNA and ensure that the samples remained well-preserved until DNA extraction.

Amplification of DNA Barcoding Markers

The genomic DNA from the plant samples was extracted using a modified Cetyl-tri-methyl ammonium bromide (CTAB) method (12). To ensure high-quality DNA with enough quantity, the concentration was determined using both gel electrophoresis and a Nanodrop 2000c spectrophotometer (Thermo Scientific, USA). DNA was quantified by measuring the absorbance at 260nm and 280nm on a UV-spectrophotometer. 50µg/ml concentration of double-stranded DNA showed an absorbance of 1 at 260nm. The concentration of DNA samples was calculated using the following formula.

$$\text{Concentration of DNA} = \frac{\text{O.D.260 nm X 50 µg DNA/ml X Dilution factor}}{1000}$$

Table 1. List of plant material collected from different locations in India.

Serial Number	Collection Numbers	Region/Location	Lat-Lon
1	Ca5	Forest area Mandla MP	22.57°N, 80.45°E
2	Ca24	Jabalpur, MP	23.21°°N, 79.95°E
3	Ca25	Katni, MP	24.00°N, 80.61°E
4	Ca28	Mandla, MP	22.59°N, 80.36°E
5	Ca30	Raipur, CG	21.23°N, 81.70°E
6	Ca31	Sagar Forest MP	23.81°N, 78.76°E
7	Ca32	Shahdol, MP	23.29°N, 81.40°E
8	Ca39	Shillong, Meghalaya	25.57°N, 91.88°E
9	Ca45	Sidhi, MP	24.10°N, 81.74°E
10	Ca35	State Forest Research Institute, Jabalpur MP	23.12°N, 79.93°E
11	Ca49	Tropical Forest Research Institute, Jabalpur MP	23.09°N, 79.98°E
12	Ca17	Indian Institute of Spice Research, Kozhikode, Kerala	11.29°N, 75.84°E
13	Ca27	Botanical survey of India, Shillong, Meghalaya	25.67°N, 91.90°E

The isolated genomic DNA was then preserved at -20°C for further use. PCR amplification of the three barcode primers was conducted in an Agilent Sure Cycler 8800 thermal cycler (USA). The PCR reactions began with a heated lid at 94°C for 5 seconds, followed by 35 cycles consisting of denaturation at 94°C for 30 seconds, annealing at 55°C (according to the primer's melting temperature) for 45 seconds, extension at 72°C for 30 seconds, and a final extension at 72°C for 10 minutes. The PCR reactions were carried out using an Agilent Technologies thermal cycler, and the resulting bands were visualized under UV light using a gel documentation system. The amplified PCR product was confirmed through 1.5% agarose gel electrophoresis to check the efficiency of barcode primers. For the sequencing process, the reaction volumes were increased to 50 µl after the subsequent PCR amplification. The final PCR products were sequenced via outsourcing to an ABI 3730XL sequencer (Applied Biosystems Inc.) using the amplification primers mentioned in Table 2.

Data Analysis

Sequence alignment

Sequence alignments were conducted using ClustalW (13), and manual adjustments were made using BioEdit (14).

Sequence similarity searches

Each generated barcode sequence from the respective barcode primer was searched for similarity using BLASTn and the online Plant barcode identification tools available at BOLD System v3 (www.boldsystems.org/) to initially identify the plant species. Reference sequences for each barcode locus, as well as the outgroup sequence, were obtained from GenBank (https://www.ncbi.nlm.nih.gov/).

Table 2. DNA barcoding primers and their sequences.

Primer	Plastidial gene	Sequence
rbcL	ribulose-1,5-bisphosphate carboxylase/ oxygenase large	F 5' GACAACTGTGTGGACCGATG 3'
		R 5' CCACCGCGAAGACATTCATA 3'
trnL	Chloroplast, Intron of *trnL*	F 5' CGAAATCGGTAGACGCTACG 3'

To determine the most effective barcoding loci for identifying species in the *Zingiberaceae* family and infer the taxonomy of the sampled plants, phylogenetic analysis based on the maximum parsimony (MP) method was performed on the barcode data and outgroup species using the MEGA 11 program. The tree was constructed using the Tree-Bisection-Regrafting (TBR) algorithm with a search level of 1, and the initial trees were generated by randomly adding sequences in 10 replications. The resulting tree was drawn to scale, with branch lengths calculated using the average pathway method and expressed in terms of the number of changes across the entire sequence. Codon positions were considered, including 1st, 2nd, 3rd, and noncoding positions. Positions with less than 95% site coverage were excluded. If the species were identified as monophyletic for each barcode, they were considered resolved and identified as the same species.

Database enrichment

Specimen data for each region were deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank) and can be accessed publicly using the assigned accession numbers.

Plant DNA Barcode Generation

To visualize and obtain an illustrative barcode depicting similarities, differences, and nucleotide compositions of each sequence, barcodes for each primer were generated using the online tool http://biorad-ads.com/DNABarcodeWeb/.

Results

PCR amplification and sequencing

PCR amplification was performed on selected plant species using gene loci-specific primers from the chloroplast genome (*rbcL* and *trnL*) based on sequence analysis. The extraction of DNA from the plant samples was successful, resulting in high-quality DNA with satisfactory yields. The *trnL* region showed the highest PCR amplification success rate at 92%, while the *rbcL* region had a PCR amplification success rate of 70%. The list of successful amplifications and the corresponding accession numbers assigned to the specimens in the NCBI database are provided in Table 3.

Sequence length and GC percentage

The sequence length and GC content were calculated for each sample obtained after PCR amplification through MEGA 11 and sequencing from all DNA barcode loci. The average sequence length of the *rbcL* region was 339 bp, with GC content ranging from 40% to 41.2%. For the *trnL* primer, the average sequence length was 594 bp, and the GC percentage varied from 34% to 37%. The following values are summarized in Table 3.

Table 3. List of submitted and rediscovered species through comparison BOLD database, BLAST, and NCBI Gene bank Database.

Sl No.	Collection number	*rbcL* Accession Number NCBI	GC %	Sequence Length	*trnL/ Accession* Number NCBI	GC %	Sequence Length	Similarity % in Bold system database	E-Value	Query cover % Gene bank	Revealed Species through the NCBI Database and Bold database system
1	Ca17	ON652451	41	347	ON615011	34.7	605	100	0	99.99	*Curcuma aeruginosa*
2	Ca24	ON652454	41.4	353	-	-	-	99.9	0	100	*Curcuma caesia*
3	Ca25	-	-	-	ON615017	34.6	596	99.9	0.01	99.99	*Curcuma caesia*
4	Ca27	ON652456	41.7	299	ON615019	33.8	597	100	0	100	*Alpinia galanga*
5	Ca28	ON652457	40.7	320	ON615020	36.9	601	100	0	100	*Curcuma caesia*
6	Ca30	-			ON615022	34.7	598	100	0	99.98	*Curcuma caesia*
7	Ca31	ON652459	41.5	352	ON615023	36.1	579	100	0	99.99	*Curcuma caesia*
8	Ca32	ON652460	41.3	347	ON615024	35.4	604	100	0	99.99	*Curcuma caesia*
9	Ca35	-	-	-	ON615027	34.6	594	100	0	100	*Curcuma caesia*
10	Ca39	-	-	-	ON615031	36.2	598	99.99	0	100	*Curcuma caesia*
11	Ca45	ON652468	41.3	347	ON615030	35.5	587	100	0	100	*Curcuma caesia*
12	Ca49	ON652471	41	347	ON615041	34	597	100	0	100	*Curcuma caesia*
13	Ca5	ON652444	40.9	345	ON615001	36.1	578	100	0	99.99	*Curcuma caesia*

Species Identification

To identify the species of each plant sample, an online BOLD database, BLASTn, and gene bank NCBI search analysis was performed for each sequence. The search results provided species resolution up to the genus level. However, due to multiple hits with similar parameters, such as maximum score, total score, query cover, check (ranging from 99.9% to 100%), E value (0%), and similarity percentage identities (ranging from 99.9% to 100%), it was challenging to distinguish among the species included in this study. BOLD and NCBI gene bank database search engines provide a list of up to 100 closest specimens. We selected the top suggestion from this list, based on the highest percentage match and base pair overlap. When the top suggestion matched a sequence from our project, we included an alternative match to avoid circular analysis. Statistical tests were conducted separately to ensure the integrity of the results. Matches were verified at the species, genus, and family levels. The barcoding loci were analyzed to differentiate between the different species. The Maximum Parsimony (MP) was carried out with the following options: Parsimony informative characters were unordered and equally weighted, full bootstrap option. The MP tree (Fig. 1) was obtained by using the Close-Neighbour-Interchange algorithm in which the initial trees were obtained with the random addition of sequences (100 replicates). Both barcode markers, *rbcL*, and *trnL*, demonstrated excellent species resolution in the phylogenetic tree. From the *rbcL* dendrogram, the specimens were clustered into two groups. All the specimens of *Curcuma longa* and *Curcuma aeruginosa* formed Group 1, Group 2 was composed of *Alpinia galanga* which shows major diversity among the species. The relationship between *Curcuma longa* and *Curcuma aeruginosa* was closer, compared with Alpinia galanga. In Group 1, there are two adjacent sister subclades (Clade I and Clade II) Fig.1. The phylogram generated through maximum parsimony analysis displayed a well-resolved tree with moderate to high support for most branches. The utilization of these DNA barcode primers facilitated clear discrimination of species at the genus level and their relationships with other species. The outgroup species, *Alpinia galanga*, was separated from the studied genera of the Zingiberaceae family (Fig.1 and Fig. 2). Additionally, *Curcuma aeruginosa* exhibited a close relationship with *Curcuma caesia*, indicating genus-based clustering. In this study, all species formed distinct and well-defined monophyletic clades with their respective species, enabling reliable species identification.

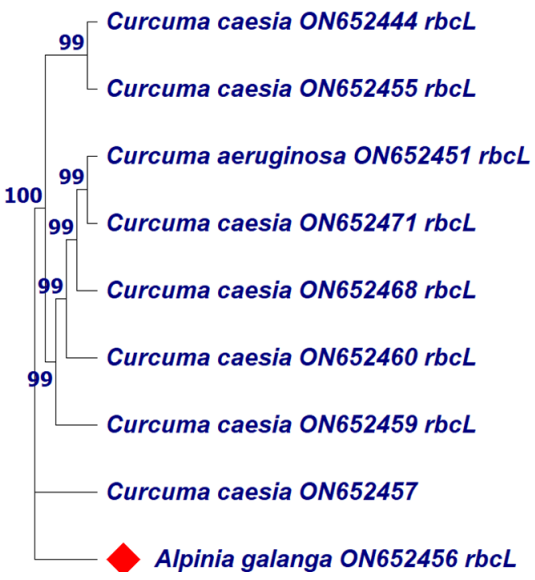

Fig. 1. Using maximum parsimony analysis for *Curcuma* and closely related genera using the *rbcL* DNA barcode. Accessions were ON652456 showing outgroup due to different genera and accession ON652451 clustering near *Curcuma caesia* due to Neighbour relationships within species.

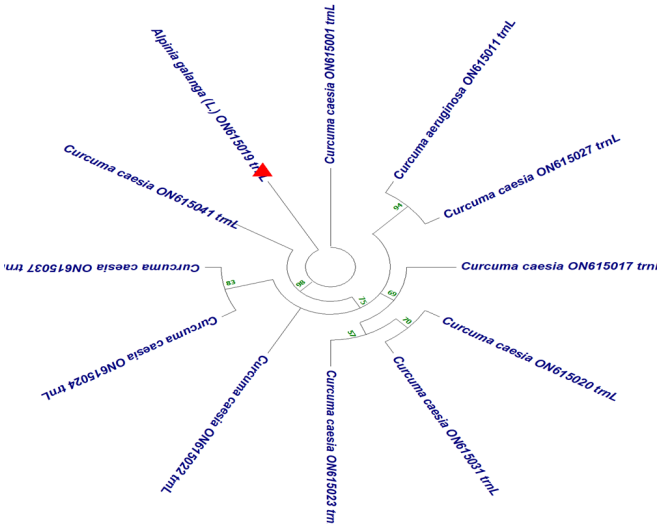

Fig. 2. Using maximum parsimony analysis for *Curcuma* and closely related genera using the *trnL* DNA barcode. Accessions were ON615019 showing outgroup due to different genera and accession ON615011 clustering near *Curcuma caesia* due to Neighbour relationships within species.

Sequence Divergence for DNA Barcode Loci

In this study, the measurement of sequence divergence for DNA barcode loci was conducted using DnaSP v4.5 software. Several parameters, including polymorphic sites, the total number of sites, variable sites, and nucleotide diversity, were analyzed for each barcode locus (Table 4). On an individual barcode basis, the highest number of polymorphic sites and nucleotide diversity was observed in the *trnL* locus, with 619 polymorphic sites and a nucleotide diversity of 0.07831. Conversely, the *rbcL* locus exhibited the lowest number of polymorphic sites (343) and the lowest nucleotide diversity (0.002). The variable sites were predominantly found in the *trnL* locus (168 sites), while the *rbcL* locus had only 2 variable sites. The sequence divergence results for each barcode locus are summarized in Table 4.

Barcode generation

To visually represent the nucleotide composition of the generated sequences from each DNA barcode locus in the studied plant species of the Zingiberaceae family, a color-coded barcode system was employed. Each nucleotide (Adenine, Cytosine, Guanine, and Thymine) was assigned a specific color, with red representing adenine, blue representing cytosine, green representing guanine, and orange representing thymine. This color-coded fingerprint accurately depicted the distinct nucleotide composition encoded by each sequence and was presented in the form of barcode primers (Fig. 3).

Table 4. Measurement Sequence divergence for DNA barcode loci.

Sl. No	Variable/Trait	*rbcL*	*trnL*
1	The number of polymorphic sites	343	619
2	The number of variable sites	2	168
3	Nucleotide diversity	0.00214	0.07831

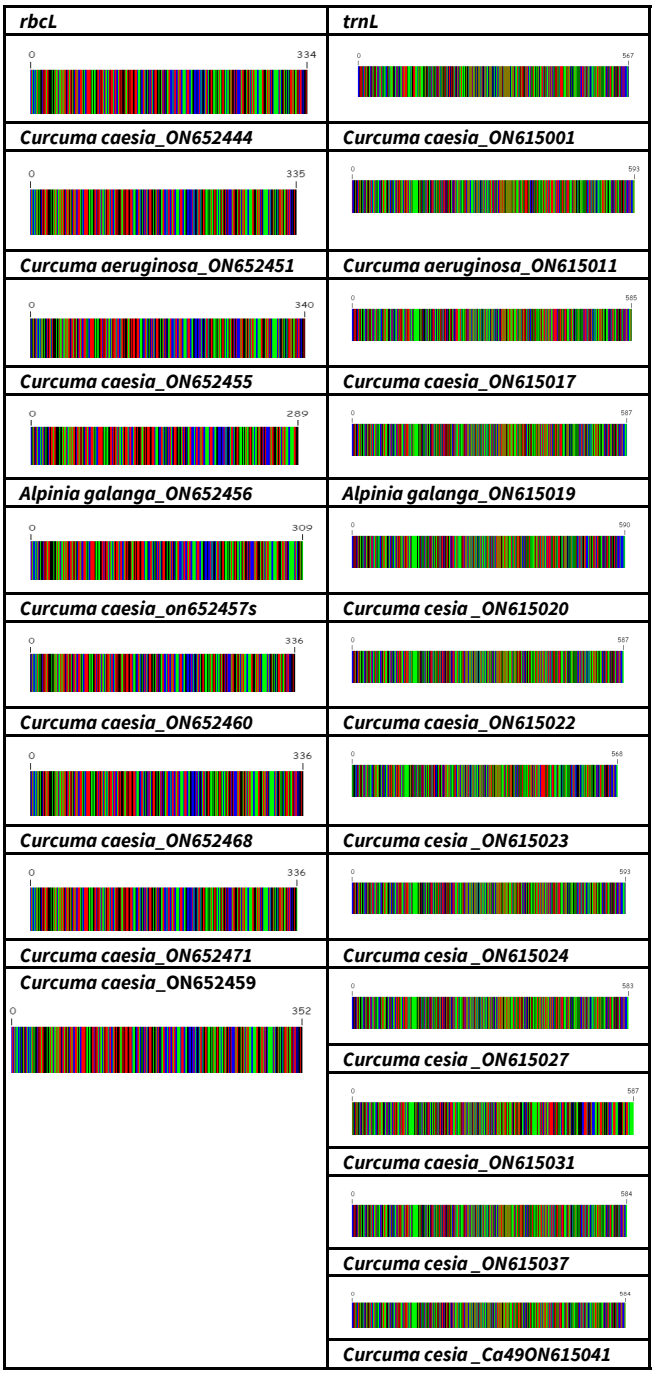

Fig. 3. Species-specific DNA Barcodes [Nucleotide bar colors: G-Black, C-Blue, A-Green, and T- Red

Discussion and Conclusion

The accurate identification and safe usage of medicinal herbs derived from plants in the Zingiberaceae family required proper species authentication. DNA barcoding is a reliable tool to identify species, and it has been successfully used in different organisms such as insects, fishes, and birds (15). At first, the concept was not widely accepted in the plant taxonomy field, and it was viewed with suspicion. DNA barcoding, however, has proven its effectiveness in distinguishing plant species using specific regions of DNA (16). In this study, DNA barcoding was performed using barcode loci derived from plant plastid DNA (*rbcL* and *trnL*). These regions have been widely recognized and recommended by the Consortium for the Barcode of Life (CBOL) for plant DNA barcoding due to their ability to discriminate between species. The process involved the generation of molecular signatures specific to each species, which were then compared to a reference library of known DNA barcodes (17,18,19,20,21).

To successfully implement DNA barcoding, several techniques and steps were involved, including DNA extraction, PCR amplification of barcode loci, sequencing, and bioinformatic analysis. The effective amplification of PCR requires DNA of high purity. In this study, high-quality DNA was extracted from plant samples. The efficiency of PCR amplification varied among the barcode loci, with the *trnL* primer showing the highest efficiency of 92%. This finding is consistent with previous studies reporting successful amplification of *trnL* loci in various plant species, including ferns, *Scolymus* multiseriate plants, and *Phalaenopsis* species.

The high amplification efficiency of *trnL* loci makes it particularly suitable for projects involving DNA amplification from degraded tissues This is crucial for DNA barcoding studies in cases where the DNA sample quality is poor or degraded (22,23,24). The current study demonstrated successful DNA barcoding in identifying Zingiberaceae species, with *trnL* and *rbcL* loci providing the best species resolution in the phylogenetic tree. In conclusion, the use of specific barcode loci, such as *trnL* and *rbcL*, along with proper laboratory techniques and bioinformatic analysis, allowed for accurate species discrimination. Further research and expansion of DNA barcoding databases will enhance our understanding of Zingiberaceae species diversity and aid in the conservation and management of these valuable plant resources.

Sequence length and average GC contents

The length of DNA sequences and their GC contents have important implications for gene regulation, DNA stability, primer binding, and genomic studies. In this study, the sequence length and average GC contents of the DNA barcode loci were calculated using MEGA 11.

The mean sequence length of the *trnL* loci across all studied species was 594 bp, which was the highest. This finding is consistent with previous reports in *Thunbergia* species, which showed similar sequence lengths. In the *Uncaria* species, the length of the *trnL* sequence was found to be slightly longer (25). The GC content of the *trnL* sequences obtained in this study was found to be 34.65%. These GC contents were lower compared to *Thunbergia* and *Uncaria* species, which had higher GC contents. The variation in sequence lengths and GC contents reflect the genetic diversity within the Zingiberaceae family.

On the other hand, *rbcL* is a well-characterized gene sequence and is widely used for DNA barcoding in plant species identification. During the present study, the mean

sequence length obtained from *rbcL* primers was 324.09 bp, which is relatively short compared to other published reports. Previous studies in the Lemnaceae family and tropical trees of India reported longer *rbcL* sequence lengths (26,27). This discrepancy in sequence lengths may be attributed to variations in the species studied or differences in the primers used for PCR amplification and sequencing. There were variations in the sequence length and GC contents of DNA barcode loci among different species and primer sets of the Zingiberaceae family. The differences observed among the members of the family were a result of their genetic diversity and evolutionary relationships. Further studies involving a larger number of species and standardized primer sets will contribute to a more comprehensive understanding of the sequence characteristics in the Zingiberaceae family and enhance the accuracy of DNA barcoding techniques. Among the plastid genes, *rbcL* stands out as the most extensively studied and characterized gene sequence. Its suitability for barcoding purposes has been widely tested by various research groups. The *rbcL* gene encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO), a crucial enzyme involved in photosynthesis. The gene *rbcL* was the first plant gene sequenced due to its importance in photosynthesis (28).

While selecting a suitable barcoding marker, it is essential to consider the unique requirements of each study and the genetic makeup of the organisms being studied. In future studies, it may be beneficial to investigate the development of better primers or alternative barcoding markers to address the limitations of shorter sequence lengths found in the *rbcL* marker. Additionally, combining multiple barcode regions or integrating other genomic data could enhance the accuracy and reliability of species identification in DNA barcoding studies. To facilitate PCR amplification and sequencing of the relatively short sequences within the *rbcL* gene, primers have been developed for a wide range of taxa. However, in the present study, the mean sequence length obtained from *rbcL* primers was relatively short at 324.09 bp compared to other published reports. For instance, in the Lemnaceae family, the sequence length of *rbcL* was reported to be 522 bp, and in a tropical tree from India, it was documented as 618 bp, which contradicts the results obtained in this study (29,30). Although the present study observed a shorter sequence length, *rbcL* continues to be a valuable marker for DNA barcoding. This is due to its extensive characterization and the availability of specific primers for amplification. Even though sequence lengths may vary among different taxa, the conserved regions targeted by the *rbcL* primers still provide enough information for species identification and discrimination

Species identification ability of all barcode loci

In this study, three methods were used for species identification: similarity-based (BLASTn, BOLD database), distance-based (Species Identifier/TaxonDNA), and phylogenetic analyses (UPGMA, Maximum parsimony). The barcode sequences were initially subjected to similarity analysis using the BLASTn tool of Gene Bank NCBI, which compares the sequences with those available in the GenBank repository. While the searches resolved the plant sequences up to the genus level, most species could not be identified unambiguously due to multiple hits at identical parameters. This led to the recovery of identical sequences with different species names under the genus *Curcuma* and related species. The Barcode of Life Data Systems (BOLD) was also employed, and it successfully resolved the plant sequences up to the genus level using the barcode sequences from this study. For identifications, species used BOLD engine and BLAST gene bank searches, as well as the subsequent repository of GenBank accession numbers, are presented in Table 3. The comparative performance analysis of BOLD versus BLAST Gene bank for specimen identification revealed a high degree of similarity, ranging from 99.99% to 100%, for pooled comparisons of all species. Among the 13 accessions analyzed using BOLD and NCBI GenBank submissions, 11 species were identified as *Curcuma caesia*, one species as *Curcuma aeruginosa,* and one as *Alpinia galanga*, which belongs to the Zingiberaceae family. These results underscore the effectiveness of both BOLD and BLAST in accurately identifying species within the Zingiberaceae family (Fig.4,5). During the investigation, another aspect that was explored was the divergence of sequences. Various studies have noted that there are different sites and levels of variability in various barcode regions. In the *rbcL* region, only two variable sites were found which indicates the lowest variability. The *ITS2* region exhibits the highest variability, followed by the *trnH-psbA*, *matK*, and *trnL*-F regions. This information is crucial for any researcher or geneticist looking to understand and analyze the diversity of different species (31,32). By employing multiple analysis methods to assess sequence variation, previous research has paved the way for findings that align with the approach used in this study. These results highlight the importance of employing diverse and comprehensive methods to evaluate inter- and intraspecific sequence variation. Visual representation of the nucleotide composition of each sequence obtained from the DNA barcode loci is essential for accurate analysis. To ensure precision, unique color codes have been assigned to each nucleotide. This method guarantees precise identification and differentiation of nucleotides, providing reliable results every time. By using the colours red, blue, green, and orange, we can easily distinguish between the four nitrogenous bases of DNA. Adenine is represented by red, cytosine by blue, guanine by green, and thymine by orange. This color-coded system simplifies the process of identifying and studying DNA, making it an essential tool for researchers in various fields. This representation, which resembles a barcode, unequivocally displays the unique nucleotide composition encoded by each sequence.

In conclusion, this study demonstrated the effectiveness of plastid DNA barcodes in species identification within the Zingiberaceae family. The *rbcL* and *trnL* loci were identified as potential barcode regions for different species. The results obtained in this study lay the foundation for constructing a robust DNA-barcoding

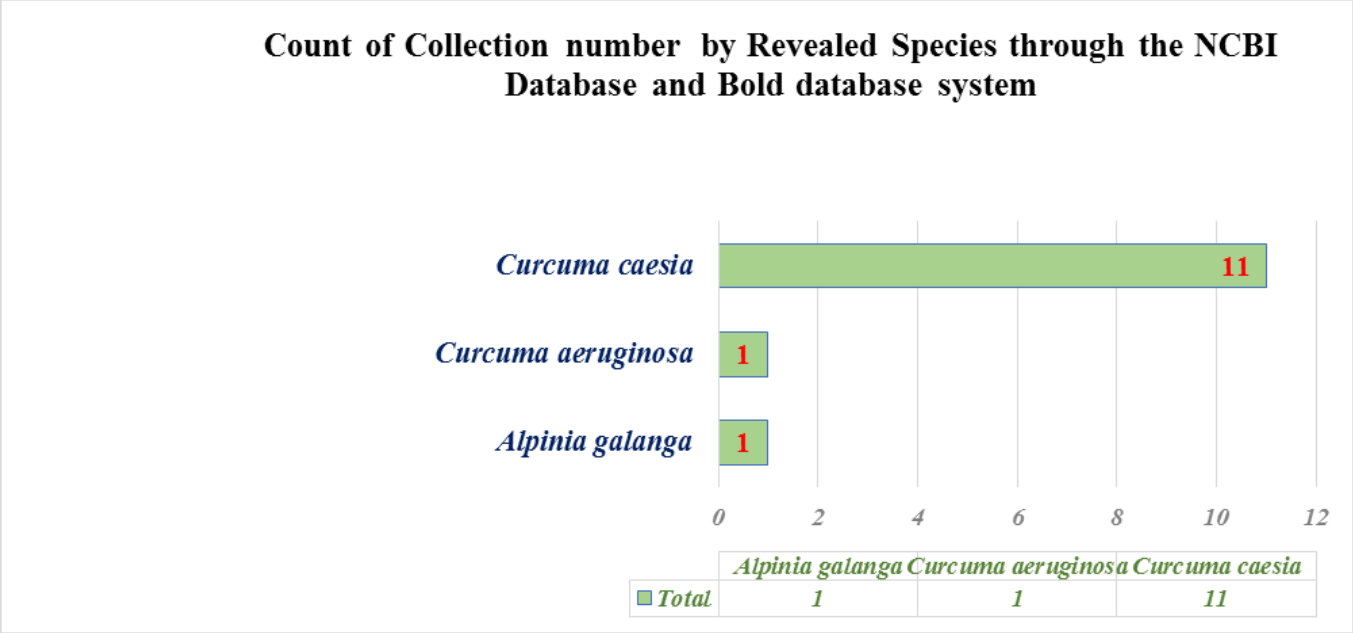

Fig. 4. Count of Collection number by Revealed Species through the NCBI Database and Bold database system Out of 13 11 species of *Curcuma Caesia*, 1 *Curcuma aeruginosa,* and 1 *Alpina galanga*

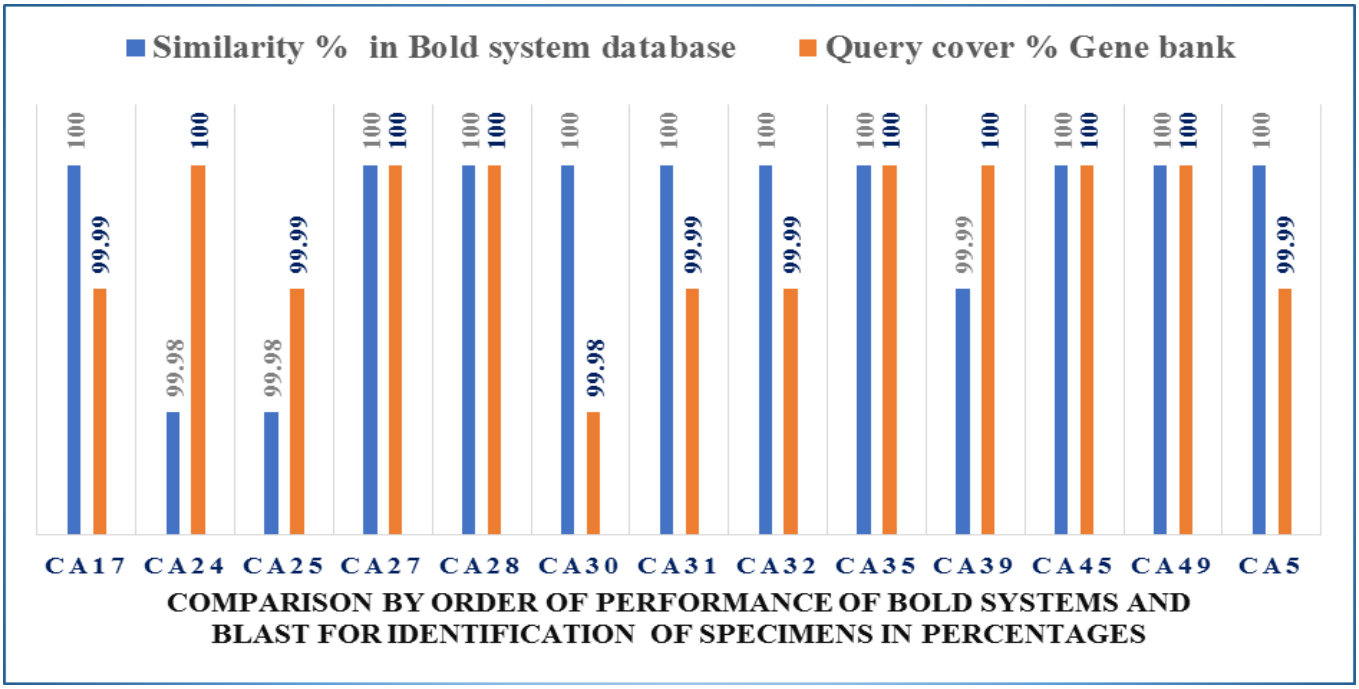

Fig. 5. Comparison by order of performance of BOLD Systems and BLAST for identification of specimens in percentages.

database, where species can be verified by combining morphological and molecular data. A database can aid species identification, leading to the development of species-specific markers and bioinformatics tools for endangered species detection. The integration of DNA barcoding techniques with traditional taxonomic approaches can enhance our understanding of species diversity and provide valuable tools for species identification and conservation efforts. This study highlights the potential of DNA barcoding in the Zingiberaceae family. It sets the stage for further research and the development of comprehensive DNA-barcoding databases for accurate species identification.

Acknowledgements

We would like to thank all the authors for their support in this research. Special thanks to the Botanical Survey of India Meghalaya, JNKVV Jabalpur herbal garden, TFRI Jabalpur MP, SFRI Jabalpur MP, and IISC Kozhikode for providing the germplasm of Zingiberaceae. This research was supported by the PhD Research Fund from the Biotechnology Centre, Jawaharlal Nehru Agriculture University, Jabalpur, Madhya Pradesh, India

Authors' contributions

The research was designed by KT and ST. Experiments were conducted by VKS, who also performed data analysis alongside SS. The manuscript was drafted by VKS and KT with support in scientific writing from SS, SwS, and NM. All authors reviewed and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: The authors declare no conflicts of interest.

Ethical issues: None.

Funding: This research received no external funding.

Data Availability Statement

In this research, the *Curcuma* genus was discriminated using barcoding techniques, and the species-specific DNA barcode gene sequences were deposited in the public database domain of the National Center for Biotechnology Information (NCBI). The submitted accession numbers list in Table 2 *rbcL* and *trnL* gene accession are published in public NCBI GenBank and this deposited accession can be found at NCBI Gene bank nucleotide dashboard using the following link: https://www.ncbi.nlm.nih.gov/nuccore/?term=vinod+sahu

References

1. Škorničková J. Taxonomic studies in Indian Curcuma L. Dissertation, Charles University, Prague, Czech Republic 2007.
2. Ratnasingham S, Hebert PDN. bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular ecology notes. 2007;7(3). https://doi.org/10.1111/j.1471-8286.2007.01678.x
3. Angel GR, Menon N, Vimala B. Essential oil composition of eight starchy *Curcuma species*. Industrial Crops and Products. 2014;60:233-8. https://doi.org/10.1016/j.indcrop.2014.06.028
4. Sahu VK, Tantawi K, Sapre S. Improved method of DNA extraction from leaf and rhizome samples of black turmeric (*Curcuma caesia*) for molecular analysis. Journal of Phytopharmacology.2022a; 11(4): 286-88. https://doi.org/10.31254/phyto.2022.11411
5. Xiang Z, Wang XQ, Cai X, Zeng S. Metabolomics study on quality control and discrimination of three Curcuma species based on gas chromatograph-mass spectrometry. Phytochemical Analysis. 2011;22(5):411-18.https://doi.org/10.1002/pca.1296
6. Zhang L, Yang Z, Wei J. Contrastive analysis of chemical composition of essential oil from twelve Curcuma species distributed in China. Industrial Crops and Products. 2017;108:17 -25. https://doi.org/10.1016/j.indcrop.2017.06.005
7. Ravindran PN and Babu KN. Ginger: The Genus Zingiber [1st ed.]. CRC Press.2004.
8. Vanchhawng L, Lalramnghinglova H. Notes on the genus Hedychium J. Koen. (Zingiberaceae) in Mizoram, northeast India. Int J Waste Resour. 2016;6(3):234. https://doi.org/10.4172/2252-5211.1000234
9. Sahu, V, Tantwai K, Mishra N. Morphological characterization of Curcuma caesia germplasm collected from different regions of Madhya Pradesh. The Pharma Innovation Journal. 2022b;11(7S): 4080-4084.
10. Chen J, Xia NH, Zhao JT. Chromosome numbers and ploidy levels of Chinese Curcuma species.Hortscience.2013;48(5):525-30. https://doi.org/10.21273/hortsci.48.5.525
11. Záveská E, Fér T, Šída O, Krak K, Marhold K, Leong-Škorničková J. Phylogeny of Curcuma (Zingiberaceae) based on plastid and nuclear sequences: Proposal of the new subgenus Ecomata.Taxon.2012;61(4):747-63. https://doi.org/10.1002/tax.614004
12. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard R. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proceedings of the National Academy of Sciences.1984;81(24):8014-8018. https://doi.org/10.1073/pnas.81.24.8014.
13. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Research. 1994;22 (22):4673-4680.https://doi.org/10.1093/nar/22.22.4673
14. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series.1999;41:95-98.
15. Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences.2003;270 (1512):313-321. https://doi.org/10.1098/rspb.2002.2218
16. Chase MW, Salamin N, Wilkinson ML. Plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society.2005; 360(1462):1889-95. https://doi.org/10.1098/rstb.2005.1720
17. Amandita FY, Rembold K, Vornam B. DNA barcoding of flowering plants in Sumatra, Indonesia. Ecology and Evolution. 2019;9 (4):1858-68.https://doi.org/10.1002/ece3.4875
18. Cahyaningsih R, Compton LJ, Rahayu S. DNA barcoding medicinal plant species from Indonesia.Plants.2022;11 (10):1375. https://doi.org/10.3390/plants11101375
19. Fazekas AJ, Kuzmina ML, Newmaster. DNA barcoding methods for land plants. Methods Molecular Biology. 2012;858:223-52. https://doi.org/10.1007/978-1-61779-591-6_11
20. Kress WJ, García-Robledo C, Uriarte M.DNA barcodes for ecology, evolution, and conservation. Trends in Ecology and Evolution. 2015;30(1):25-35. https://doi.org/10.1016/j.tree.2014.10.008
21. Sucher NJ, Hennell JR, Carles MC. DNA fingerprinting, DNA barcoding, and next-generation sequencing technology in plants. Methods in Molecular Biology.2012;862:13-22. https://doi.org/10.1007/978-1-61779-609-8_2
22. Gravendeel LA, Kouwenhoven MC, Gevaert O. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Research. 2019;699(23): 9065-9072. https://doi.org/10.1158/0008-5472.CAN-09-2307
23. Chen S, Yao H, Han J. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE. 2010;5(1): e8613. https://doi.org/10.1371/journal.pone.0008613
24. Tsai CC, Chiang YC, Lin YS. Plastid *trnL* intron polymorphisms among *Phalaenopsis species* used for identifying the plastid genome type of Phalaenopsis hybrids. Scientia Horticulturae. 2012;142:84-91. https://doi.org/10.1016/j.scienta.2012.05.004
25. Wongakson P, Rungpragayphan S, Powthongchin B. Evaluation of potential DNA barcodes for identifying *Thunbergia spp*. Journal of Pharmaceutical Sciences. 2015 10(supple:)147-159.
26. Osathanunkul MP, Lithanatudom P, Madesis. Identification of *Uvaria sp* by barcoding coupled with high-resolution melting analysis (Bar-HRM). Genetics and Molecular Research.2016;15: 13. https://doi.org/10.4238/gmr.15017405
27. Pettengill JB, Neel MC. An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus *Agalinis* (Orobanchaceae).American Journal of Botany. 2010;97(8):1391-406. https://doi.org/10.3732/ajb.0900176
28. Zurawski G, Gunsalus RP, Brown KD. Structure and regulation of aroH, the structural gene for the tryptophan-repressible 3-deoxy-D-arabino-heptulosonic acid-7-phosphate synthetase of *Escherichia coli*. Journal of Molecular Biology. 1981;145(1):47-73. https://doi.org/10.1016/0022-2836(81)90334-X
29. Tripathi AM, Tyagi A, Kumar A. The internal transcribed spacer (ITS) region and trnhH-psbA are suitable candidate loci for DNA

barcoding of tropical tree species of India. PloS ONE. 2013;8 (2):e57934. https://doi.org/10.1371/journal.pone.0057934

30. Wang W, Wu Y, Yan Y.DNA barcoding of the Lemnaceae, a family of aquatic monocots. BMC Plant Biology. 2010;10:205. https://doi.org/10.1186/1471-2229-10-205

31. Chen J, Xia NH. Curcuma gulinqingensis sp.nov. (Zingiberaceae) from Yunnan, China. Nordic Journal of Botany. 2013;31(6):711-6. https://doi.org/10.1111/j.1756-1051.2012.01408.x

32. Lahaye R, Van der Bank M, Bogarin D. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences. 2008;105(8):2923-8. https://doi.org/10.1073/pnas.0709936105

Weblink of database

http://www.barcodeoflife.org

http://www.barcoding.si.edu/protocols.html

http://www.barcodinglife.com

http://www.boldsystems.org