East African Medical Journal Vol. 100 No. 1 January 2023
DNA BARCODING USING ITS2 AND RBCL MARKERS FOR *SOLANACEAE* SPECIES IDENTIFICATION
Kathryn Wanjiku Nderitu, Department of Biochemistry, Faculty of Science and Technology, University of Nairobi, P. O. Box 30197-00100 Kenya, Elvince Ager, Department of Microbiology, Faculty of Science and Technology, University of Nairobi, P. O. Box 30197 - 00100, Kenya, Ezekiel Mecha, Department of Biochemistry, Faculty of Science and Technology, University of Nairobi, P. O. Box 30197 - 00100 Kenya, Atunga Nyachieo, Institute of Primate Research, Karen-Nairobi, P. O. Box 24481 - 00502, Kenya.

Corresponding author: Kathryn Wanjiku Nderitu, Department of Biochemistry, University of Nairobi, P. O. Box 30197 -00100, Nairobi, Kenya. Email: knderitu276@gmail.com

# DNA BARCODING USING ITS2 AND RBCL MARKERS FOR *SOLANACEAE* SPECIES IDENTIFICATION

K. W. Nderitu, E. Ager, E. Mecha and A. Nyachieo

## ABSTRACT

***Background*: The *Solanaceae* taxonomic family has substantial economic and commercial importance and is mostly used as food, medicine and spices. In Africa, there are several polyploid species of *Solanum*: *Solanum scabrum, Solanum villosum* and *Solanum nigrum*. Previous studies have suggested that there are different subspecies based on morphological characteristics of *Solanum nigrum* but it is uncertain if these differences are reflected at the genetic level.**

***Objective*: To identify the species taxonomic rank of a member of the *Solanaceae* that showed therapeutic potential against obesity by altering gut microbiome diversity using morphology and DNA barcodes ITS2 and RBCL.**

***Methodology*: Fresh *Solanaceae* leaves (*isolate-001_kate*) were collected from Limuru Sub-County, Kiambu County and identified morphologically. Molecular identification involved DNA extraction, followed by PCR amplification using the primers ITS2 and RBCL at an annealing temperature of 58° C for both primers and sequencing. Phylogenic inference was based on ITS2 and RBCL using maximum likelihood (ML) algorithms.**

***Results and Conclusion*: Morphological data successfully identified the leaves as belonging to *Solanaceae* but molecular primers ITS2 and RBCL gave further insight that *isolate-001_kate* was *Solanum villosum***

*Abbreviation*

*Ctab- Cetyl trimethylammonium bromide, DNA, dNTP – deoxynucleotide triphosphate, ITS- Internal transcribed spacer, Matk – Maturase k, MEGA – Molecular evolutionary genetics analysis, MSA- Multiple sequence alignment, MUSCLE – Multiple sequence comparison by log expectation, PCR, RBCL- Ribulose biphosphate carboxylases, RNA*

## INTRODUCTION

The family *Solanaceae* comprises of about 2500 species distributed in about 100 genera and thrives in a variety of habitats and ecologies and presents varying morphologies. It is used as an important source of food, medicine and spice [1]. Due to its medicinal value, identification and authentication of the species is very important. A few studies indicated that *Solanum* has several polyploid species which include *Solanum scabrum*, *Solanum nigrum* and *Solanum villosum* [2,3]. Among these polyploid species, *Solanum villosum* has previously been reported to be among the most dominant species with the subspecies *S. villosum* subsp. *miniatum* being predominant [4]. In Africa, *Solanum nigrum*, often known as black nightshade, is used to describe all species within the *Solanum nigrum* species complex, those with orange, purple or black fruits. This is due to ethnic names, which are used by most communities to refer to several species collectively. *Solanum nigrum* has been given numerous vernacular names by diverse populations in Kenya, which include: Kiswahili (mnavu), Kikuyu (managu), Kamba (kitulu), and Maasai (momoi). Although previous research has based taxonomic identification on morphological characteristics [5,6] it has not been established whether morphological differences are reflected in the genomic level. Additionally, morphological identification may not be accurate method of species identification within *Solanaceae* because members of this family are cryptic and have overlapping phenotypic characteristics influenced by environmental and genetic factors. DNA barcoding uses short universal standardized DNA sequences and has proven to be an accurate method in the identification of medicinal plants [7]. In this study, we aimed at examining the species' taxonomic rank to a member of the *Solanaceae* that showed therapeutic potential against obesity by altering gut microbiome diversity, using the DNA barcodes ITS2 and RBCL.

## METHODOLOGY

*Genomic DNA extraction and amplification*

Fresh young plants of *Solanaceae* that had previously showed anti-obesity effects on rats were collected in November 2019 from Limuru Sub-County, Kiambu County. The whole plants were transported to the Biological Science Department of the University of Nairobi for morphological identification and authentication by a taxonomist and a voucher specimen number was given KWNUON2019/001. The research protocol was approved by the Institutional Review Committee (IRC) of Institute of Primate Research (ISERC/06/19).

Among the morphological traits observed included; leaf shape, margin, surface and fruit color (Figure 1 and 2).

*Figure 1*: isolate-001_kate specie used in the study species

*Figure 2:* Unripe isolate-001_kate

The leaves were later wrapped in a foil paper, shipped to the Molecular Biology Laboratory, University of Nairobi and stored at -20 ◦ C in preparation for molecular identification. DNA extraction of 100mg of the leaf plant was done using the Cetyl trimethylammonium bromide (CTAB) method [8] and thereafter, amplification done using ITS2 and RBCL DNA markers (Table 1). Cleaning and sequencing of the amplicons was later conducted at Institute of Tropical and Infectious Diseases, University of Nairobi.

**Table 1**

*Primers used for PCR amplification* [9]

| Primer Name | Annealing temperature | Direction | Sequence (5' - 3') |
|---|---|---|---|
| RBCL | 58°C | Forward | CTGTATGGACCGATGGACTTAC |
| | | Reverse | CGGTGGATGTGAAGAAGTAGAC |
| ITS2 | 58°C | Forward | GAAGGAGAAGTCGTAACAAGG |
| | | Reverse | TCCTCCGCTTATTGATATGC |

*Sequence analysis*

The sequences were visually inspected and manually edited within BioEdit version 7.2.5 [10] and contiguous (contig) sequence data generated for each of the ITS2 and RBCL primers. For each of the contig sequences, a BLASTN search was conducted in the NCBI-BLAST data base and related sequences (those that had similarity percentage above 95%) were obtained for both set of contig sequences. The selected sequences together with their respective contig sequences were transferred to the phylogentic analysis software NGPhylogeny.fr [11] where multiple sequence alignment (MSA) was done using the inbuilt MUSCLE algorithm. The MSA was curated using BMGE within the NGPhylogeny environment. The resultant cleaned MSA was subjected to Maximum Likelihood phylogenetic inference (ML) using the PhyML [12] and the best substitution model was determined using the algorithm Smart

Model Selection within PhyML (SMS) [13]. The resultant ML trees were transferred to the software Fig Tree for visualization and annotation.

## RESULTS

*Morphological analysis*

The morphological traits of *Solanum* leaves were lanceolate, the margins were sinuate, the fruits were green in color which later turned to orange, leaf surface was hairy while the flowers were white in color (Figure 1 and 2). On the other hand, other *Solanum nigrum* species that were not used in this study had green fruits which later turned to black (Figure 3) and purple (Figure 4) respectively.

***Figure 3 : Solanacea specie sample 2***

***Figure 4: Solanaceae specie sample 3***

*Phylogenetic analysis*

In this study universal primers ITS2 and RBCL were used as the DNA barcodes for *Solanaceae*. Phylogeny was inferred using maximum likelihood (ML) methods for the two data sets generated from Multiple sequence alignment (MSA). The sample, *Isolate-001_Kate,* clustered with known members of *Solanum villosum* for both markers RBCL (Figure 5) and ITS2 (Figure 6) into monophyletic clades. The clustering of *Isolate-001_Kate* with *S. villosum* was supported with strong bootstrap values: 0.92 on ITS2 and 0.90 on RBCL respectively. Generally, the topology of all the phylogenetic trees inferred from ITS2 and RBCL were concordant with each other. In addition to the monophyletic clade that supported *S. villosum* in both the ITS and RBCL inferred trees, the monophyletic clade that housed *S. nigrum, S. americanum and S. photeinocarpun* was also present in both phylogenetic trees and it appeared as a paraphyletic clade to the *S. villosum* clade.

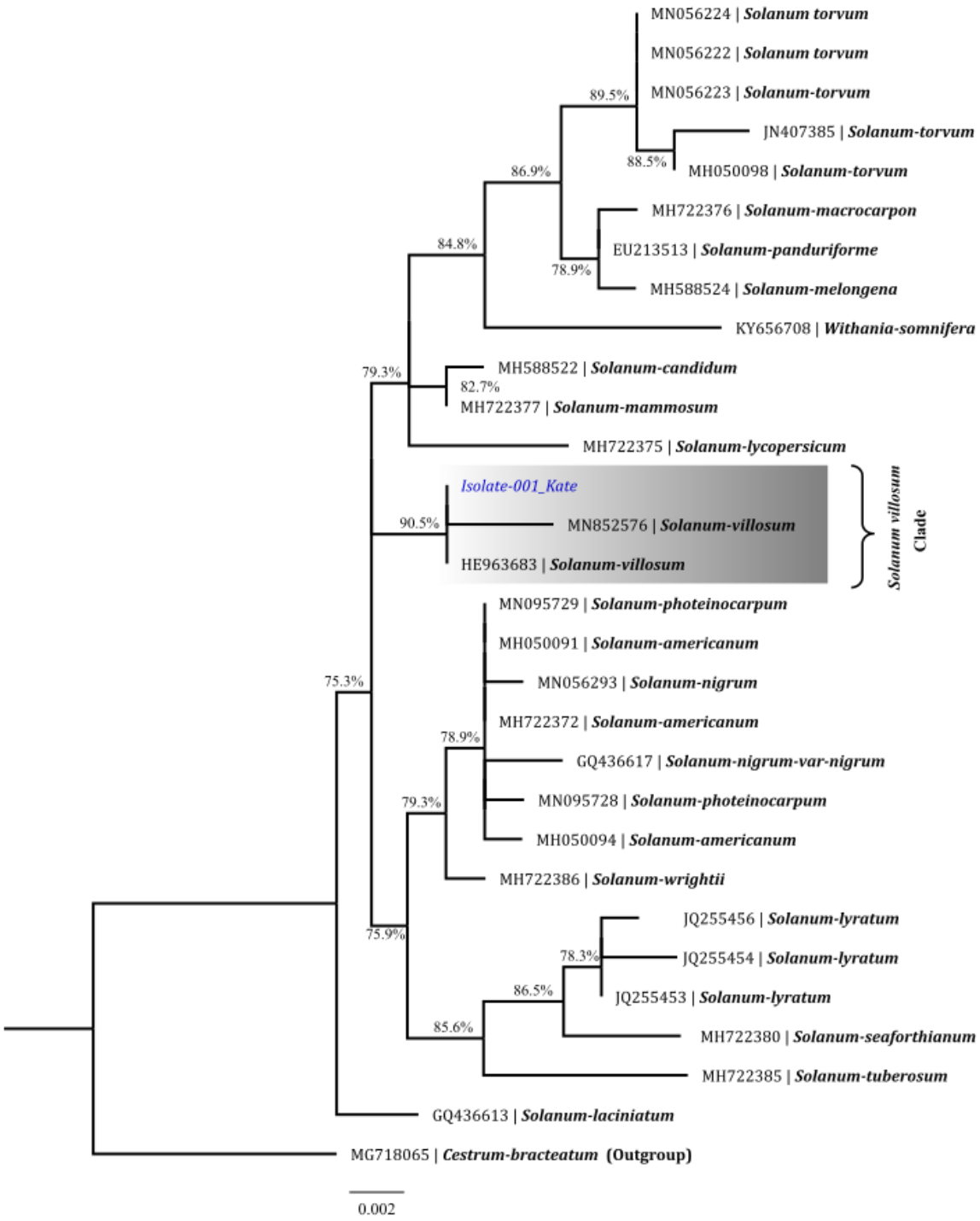


*Figure 5: A maximum likelihood phylogenetic tree constructed from RBCL sequences, our sequence nests with other known sequences of* **S. villosum** *from GenBank, the tree is drawn to scale with the scale bar showing number of nucleotide substitutions, the values next to the nodes depict SH-aLRT replications, the ID on the left of the tips are the GenBank accession numbers.*

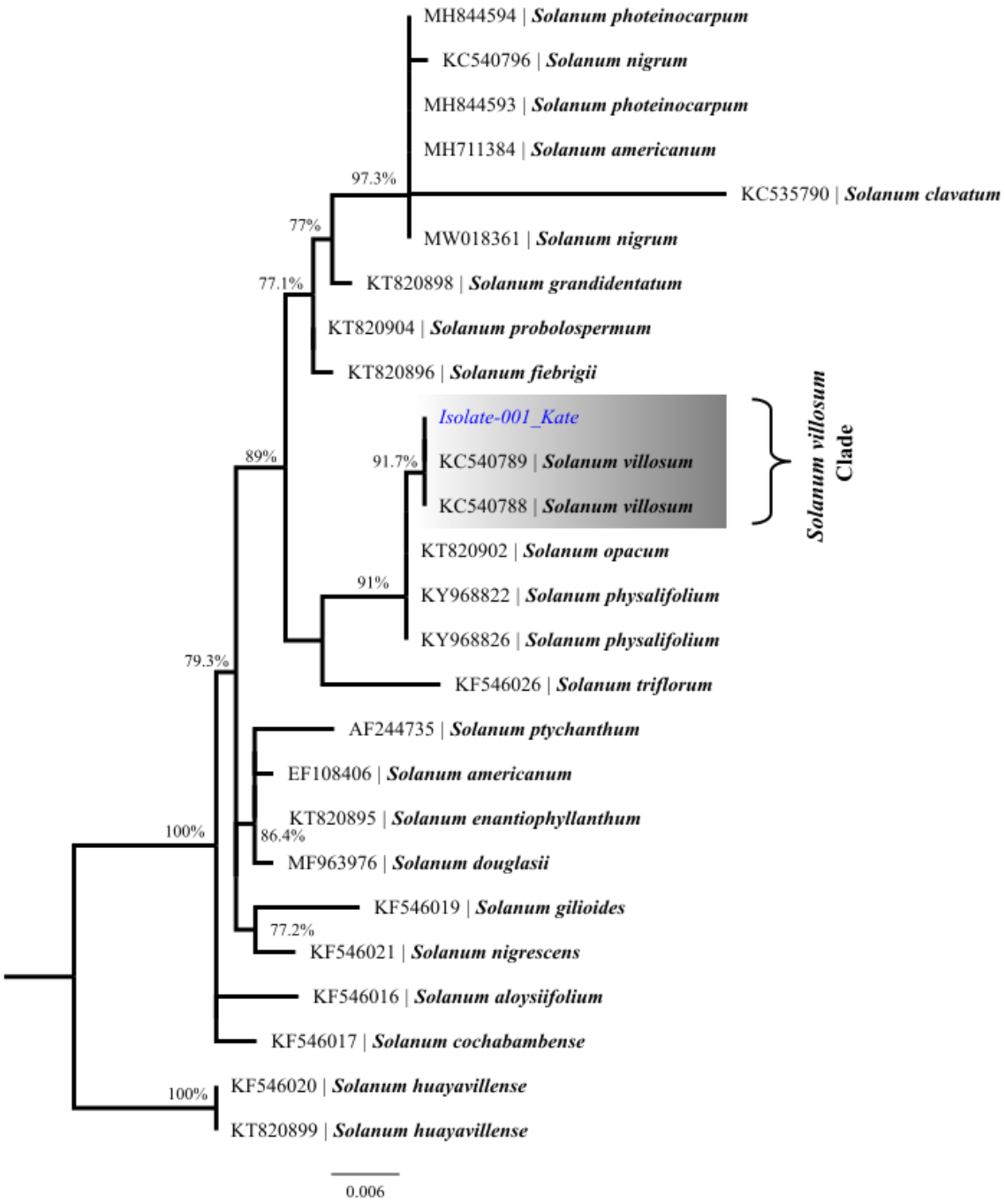


***Figure 6: A maximum likelihood phylogenetic tree constructed from ITS2 sequences, our sequence nests with other known sequences of* S. villosum *from GenBank, the tree is drawn to scale with the scale bar showing number of nucleotide substitutions, the values next to the nodes depict SH-aLRT replications, the ID on the left of the tips are the GenBank accession numbers.***

## DISCUSSION

The morphological data placed the sample used in the current study in the genus *Solanum* sp. but it was not possible to assign a lower taxonomic rank to the sample. Based on the suggestion of [14] the primer ITS2 and RBCL was used for DNA barcoding to supplement this morphological data. There were several ITS2 DNA sequence data of *Solanum nigrum* available from Genbank, which demonstrated high PCR amplification efficiency and high sequence quality. ITS2 is recognized as a universal barcode that has distinguished more than 6600 plant samples [15]. While RBCL did offer great universality, it also showed that there was not enough sequence variation to distinguish closely related species. There was a significance regarding the grouping of the *isolate-001 kate* and the phylogenetic trees created using ITS2 sequences. ITS2 phylogenetic tree agreed that *isolate-001_kate* was *S. villosum* and the tree showed sequence KC540788 and KC540789 were more related to *isolate-001_kate* than any other. The use of ITS2 for identification is also supported by previous reports indicating that ITS2 shows more accurate results on interspecific variation other than intraspecific variation compared to RBCL [16]. ITS2 indicates significant variability at the species level and is also reported as the preferred DNA barcode for identifying medicinal plants [17] . In addition, the RBCL dataset, that was much larger than the ITS2 dataset agreed that *isolate-001_kate* is a member of *S. villosum*. RBCL is a universal chloroplast genome used for DNA barcoding and is involved in discrimination of large groups of angiosperm. In comparison to all other sequences, RBCL tree isolate-001 kate showed the greatest similarity to MN852576 and HE963683. However, nuclear genes are said to provide more information than the use of organellar markers such as RBCL which is inherited from one parent [18].

In conclusion, morphological data and molecular identification using both markers strongly suggest *isolate-001_kate* species is *Solanum villosum*. This study is in agreement with previous research which showed that ITS2 and RBCL can be used as a standard DNA barcode which identify closely related medicinal plant species [19].

**REFERENCES**

1. Afroz M, Akter S, Ahmed A, Rouf R, Shilpi JA, Tiralongo E, et al. Ethnobotany and Antimicrobial Peptides From Plants of the Solanaceae Family: An Update and Future Prospects. Front Pharmacol. 2020;11(May).

2. Poczai P, Hyvönen J. On the origin of Solanum nigrum: Can networks help? Mol Biol Rep. 2011;38(2):1171–85.

3. Manoko MLK, Van Den Berg RG, Feron RMC, Van Der Weerden GM, Mariani C. Genetic diversity of the African hexaploid species Solanum scabrum Mill. and Solanum nigrum L. (Solanaceae). Genet Resour Crop Evol. 2008;55(3):409–18.

4. Olet EA, Lye KA, Heun M. Amplified fragment length polymorphisms (AFLPs) analysis of species of solanum section Solanum (Solanaceae) from Uganda. Afr J Biotechnol. 2011;10(34):6387–95.

5. Cakir Z, Balkaya A, Saribas S, Kandemir D. The Morphological Diversity and Fruit Characterization of Turkish Eggplant (Solanum melongena L.) Populations. Ekin J Crop Breed Genet. 2017;3(2):34–44.

6. Arslanoglu F, Aytac S, Oner EK. Morphological characterization of the local Potato (Solanum tuberosum L.) Genotypes collected from the Eastern Black Sea region of Turkey. Afr J Biotechnol. 2011;10(6):922–32.

7. Zhu RW, Li YC, Zhong DL, Zhang JQ. Establishment of the most comprehensive ITS2 barcode database to date of the traditional medicinal plant Rhodiola (Crassulaceae). Sci Rep. 2017;7(1):1–9.

8. Turaki AA, Ahmad B, Magaji UF, Abdulrazak UK, Yusuf BA, Hamza AB. Optimised cetyltrimethylammonium bromide (CTAB) DNA extraction method of plant leaf with high polysaccharide and polyphenolic compounds for downstream reliable molecular analyses. Afr J Biotechnol. 2017;16(24):1354–65.
9. Zahra NB, Shinwari ZK, Qaiser M. DNA barcoding: A tool for standardization of herbal medicinal products (HMPs) of Lamiaceae from Pakistan. Pak J Bot. 2016;48(5):2167–74.
10. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41:95–8.
11. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, et al. NGPhylogeny.fr: New generation phylogenetic services for non-specialists. Nucleic Acids Res. 2019;47(W1):W260–5.
12. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.
13. Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017;34(9):2422–4.
14. Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. J Ethnopharmacol. 2010;130(1):116–21.
15. Liu Z, Zeng X, Yang D, Chu G, Yuan Z, Chen S. Applying DNA barcodes for identification of plant species in the family Araliaceae. Gene. 2012;499(1):76–80.
16. Duan H, Wang W, Zeng Y, Guo M, Zhou Y. The screening and identification of DNA barcode sequences for Rehmannia. Sci Rep. 2019;9(1):1–12.
17. Dereeper A, Audic S, Claverie JM, Blanc G. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol. 2010;10(1):1–6.
18. Ralte L, Singh YT. Use of rbcL and ITS2 for DNA barcoding and identification of Solanaceae plants in hilly state of Mizoram, India. Res Crops. 2021;22(3).
19. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS One. 2010 Jan 7;5(1):e8613.