Systematic Approach to Establish DNA Barcode of Medicinally Important Plants in Nepal

❑ Sunil Bhandari, Jay Bhandari, Sanjay Lama
Department of Microbiology, Balkumari College, Chitwan, Nepal
National Ayurveda Research and Training Center, Kathmandu, Nepal
E-mail: sunilbhandariv@gmail.com

Abstract

DNA barcoding is an emerging tool for species identification that uses internationally agreed protocols and regions of DNA to create a global database of living organisms. Initiatives are taking place to generate DNA barcodes for all groups of living organisms make these genomic identity publically available to understand, conserve, and utilize the world's biodiversity. Most of the terrestrial plants are characterized using two section of coding region within chloplast, part of chloroplast gene, a more conserved rbcl and more polymorphic MatK gene. In order to create high quality databases, each plants are characterized not only with the rbcl and MatK DNA sequences, an additional sequence information from internal transcribed spacer (ITS) region is more efficient. The quality of barcode is depends on the various factors such as efficient primers, purity of DNA templates, as well as the quality of PCR amplicon from which the sequence data will derive.The protocol described here led to the generation of high efficient PCR amplicon which will aid in the minimization of erroneous DNA sequence information from which bioinformatics procedure will generates efficient barcodes.The primers used to amplified MatK, rbcl and ITS sequencewere MatK-413f-1 and MatK-1227r-1, rbcl-1F and rbcl-724R, ITS1 and ITS4 showed a strong amplification successes of 80% of each in the tasted medicinal plants of Nepal. This study propose that the used sets of primers and amplification condition will help, in part, the development of DNA barcode for medicinally important plants of Nepal to conserve their identity with its nativeness.

Keywords: **DNA barcoding, MatK, rbcl, ITSspecies identification, DNA extraction, PCR**

Introduction

DNA barcoding, although is in its infancy, is an emerging tool for species identification of any species that uses internationally agreed protocols and regions of DNA to create a global database of living organisms. International initiatives are taking place across hundreds of countries to establish the unique DNA barcode of the species to provide the appropriate standing to their nativeness and to produce an openly accessible database of global diversity to the people.

DNA barcoding and genomics share a common emphasis on the acquisition of large scale genetic data although the later encompasses the information of entire genome instead of a short and particular region that will sufficient to provide identity to the species. DNA barcodes typically consist of a standardized short sequence of DNA (notably specific gene of 400–800 bp in size) and with the advent of molecular biology, these sequences can be isolated and sequence information can be generated to characterize the species with certainty [1].

The importance of DNA barcoding of medicinally important plant of Nepal is underlain by the construction of accurate species identification to both conserve and explore the plant. The demand for medicinal plants and herbal remedies and especially its renaissance in the developed countries has been stepped up for the fearless naturopathic therapy [2] and the unexplored biodiversity of developing countries are the conspicuous adornments. These plants are important in both regards: economical as well as with health prospects, but the problem is people does not have sufficient knowledge about their own biodiversity and there is the simultaneous loss of their naive resources. The solution to conserve their identity is a combined strength of molecular genetics, sequencing technology and bioinformatics; DNA barcoding offers an advanced and accurate means to recognize their previous existence or appoints the newly described species. This tool also has the potential to speed the discovery of the thousands of plant species yet to be named, especially in tropical and subtropical biomes [3]. For instance, barcoding approach have been successfully used for the verification of different medicinal plants, kitchen spices, berries, olive oil and tea plants [4]. The genetic sequences obtained in context of DNA barcoding have also been used to create the evolutionary analysis and are also used in expansion of

phylogenetic community ecology [5].

Over the last decade, the application of plant DNA barcodes has accelerated, especially in the fields of ecology, evolution, and conservation. The process of generating and applying plant DNA barcodes for the purpose of identification entails two basic steps: 1) building the DNA barcode library of known species, and 2) matching the DNA barcode sequence of an unknown sample against the DNA barcode library [6]. The first step requires taxonomists to select one to several individuals per species to serve as reference samples in the DNA barcode library. Once the DNA barcode library is complete for the organisms under study, whether they comprise a geographic region, a taxonomic group, or a target assemblage (e.g., medicinal plants, timber trees, etc.), then the DNA barcodes generated for the unidentified samples are compared to the known DNA barcodes using some type of matching algorithm[7].

To be practical as a DNA barcode, agene region must satisfy three criteria:(i) contain significant species-level geneticvariability and divergence, (ii) possessconserved flanking sites for developinguniversal PCR primers for wide taxonomicapplication, and (iii) have a shortsequence length so as to facilitate current capabilities of DNA extraction andamplification[8].If the barcode marker is conservative such as rbcL, in a multilocus barcode, it will enable the construction of phylogenetic trees for all of the species in a forest, facilitating investigations of community structure [9] and functional trait evolution [10]. However polymorphic barcode marker such as MatK and ITS will enable to identify the species with high precision[11]. The combined outcomes of both conservative and variable region will enable to establish unique barcode for the specific plant species.

Methodology

The overall aim of the study was to create an optimum condition to establish the foundations for the systematic characterization of medicinally important plants in Nepal. Genomic DNA was extracted usingThermo Scientific GeneJET Plant Genomic DNA Purification Mini Kit K0791, at National Ayurveda Research and Training Center (NARTC), Kirtipur, Kathmandu, Nepal. 10 different medicinally important plants namely, *Artemisia vulgaris* (TitePati), *Asparagus racemosus* (Kurilo), *Acoruscalamus* (Bojho), *Centellaasiatica* (Brahmi), *Schimawallichii* (Chilaune), *Tropaeolummajus* (Nasturtium), *Ocimumtenuiflorum* (Tulsi), *Azadirachtaindica*(Neem), *Catharanthusroseus* white, *Senegalia catechu* (Khair) were selected form Makawanpur and Kathmandu district for their barcode analysis. The plant samples (young leaves), sealed in zip bag, were transferred to the molecular biology laboratory at NARTC and processed for DNA extraction by freezing under liquid nitrogen within the 24 hours of sampling.

DNA samples were maintained in 100µl elution buffer, and agarose gel electrophoresis (2% agarose E-gel, safe green, Invirtogen e-gel power snap) was conducted to ensure the isolation procedures. Purity of sample was quantified by spectrophotometric absorbance (ratio of A_{260}nm and A_{280}nm)from small aliquot of DNA sample (5µl sample in 245µl molecular grade water).Approximative value of absorbance (A_{260}/A_{280}) of DNA samples were then used as the template for the polymerase chain reaction (PCR) to amplify the both conserved and variable region in DNA. Three different sets of primers MatK (MatK-413f-1 and MatK-1227r-1), rbcl (rbcl-1F and rbcl-724R), and ITS (ITS1 and ITS4) were used to amplify the both conserved rbcl and polymorphic MatK gene of chloroplast (See table 1 for details) and 5.8S ITS sequence. PCR reaction was consist of 0.25mMdNTP's(10mM each), 0.5mM primer (10µmol each), 1.25units Taq DNA polymerase (eurofins genomic, 5Units/µl), 1X Taq buffer A (10X), 5µl DNA template for the final volume of 25µl by adding PCR grade water. (See table 2 for the details of the PCR programme for individual sets of primers.)

Results

Genomic DNA was extracted from all 10 plant samples implies 100% DNA extraction rate of Thermo Scientific GeneJET Plant Genomic DNA Purification Mini Kit K0791.The purity of DNA revealed the average absorbance of 1.78 (A_{260}/A_{280}) and more intense band of genomic DNA was appeared from the samples containing very young leaves. Slightly matured leaf samples from *Senegalia catechu* and*Schimawallichii*(Fig 1: 2nd and 9th band),produced a very faint bands on 1.5% agarose gel.

Fig 1: Genomic DNA extracted from plants

One universal (ITS) and two specific sets of primer (MatK and rbcl) with even numbers of bases for both forward and reverse primers were used to amplify the entire gene sequence maturase K encoding gene and ribulose-bisphosphate carboxylase geneand of 5.8S ITS region (ITS1 and ITS4)(Table 1). Out of 10 DNA samples, 80% of them were successfully amplified for the MatK gene of chloroplast. DNA samples of plants *Senegalia catechu* and*Schimawallichii*were not amplified by MatK-413f-1and MatK-1227r-1(fig 2, sample 5 and 6). Similar result was observed for both rbcland ITSgenewith primer sets rbcl-1F and rbcl-724R and ITS1and ITS4with success rate of 80%, the amplification of corresponding genes and was not amplified for *Senegalia catechu* and*Schimawallichii*(fig 3, fig 4: sample 5 and 6). The average length of PCR amplicon for MatK gene were ranging around 1500 base pairs and the average size of amplicon for ITS and rcbl gene were ranges between 500bp to 700bp.Although the similar intense band of genomic DNA was employed for PCR, a variably intense amplicons band were observed.

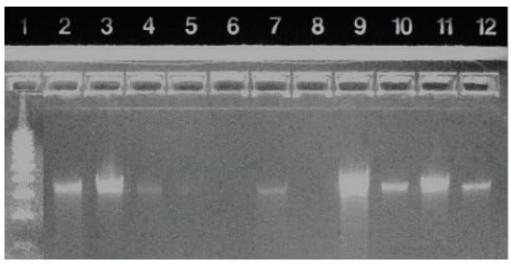

Fig 2: MatK gene amplification using sets of primer ITS1 and ITS4

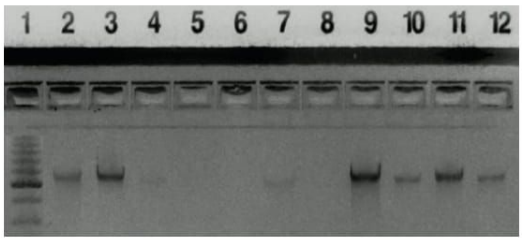

Fig 3: ITS gene amplification using sets of primer MatK-413f-1 and MatK-1227r-1

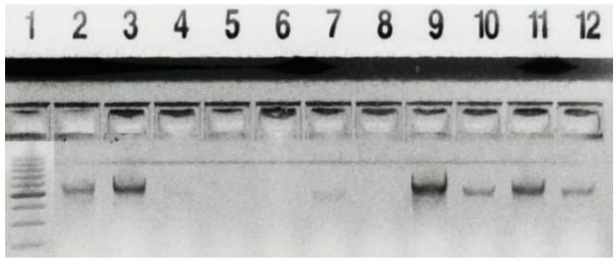

Fig 4: rbcl gene amplification using sets of primer rbcl-1F and rbcl-724R

Table 1: Primers used to amplify rbcl.MatK and ITS regions

Primer	F/R	Sequence	References
MatK-413f-1	F	TAATTTACRATCAATTCATTCAATATTTCC	Heckenhauer et al. 2016
MatK-1227r-1	R	GARGAYCCRCTRTRATAATGAGAAAGATTT	Heckenhauer et al. 2016
rbcl-1F	F	ATGTCACCACAAACAGAAAC	Khan et al. 2012
rbcl-724R	R	TCGCATGATCCTGCAGTAGC	Khan et al. 2012
ITS1	F	TCCGTAGGTGAACCTGCG	White et al, 1990
ITS4	R	TCCTCCGCTTATTGATATGC	White et al, 1990

Table 2:
Thermocyclerprogram condition for each sets of PCR primer. (The values are temperatures and time and were expressed in degree centrigrade (^{0}C) per second (/Sec).)

Target gene	Denaturation	Annealing	Extension	Final Extension
MatK	98/10	55/30	72/30	72/300
Rbcl	94/30	54/45	72/45	72/300
ITS	94/45	54/30	72/60	72/300

Discussion

An ideal DNA barcode should be routinely retrievable with a single primer pair, be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces, and provide maximal discrimination among species. But practically it is harder to define a species certainly with the power of single loci and single set of primer and if followed, there will be the chances of erroneous inclusion of global diversity into the database[14]. Thus, adaptation of cumulative sequence information retrieved from multilocus as well as conserved and variable region will be adequate to minimize the threat of false identification.

Large genome size and evolutionary trendline of autotrophs makes difficulty in choosing a plant efficient barcode for researcher to establish the plants identity. Although MatK, rbcl and ITS sequences possess potential attributes to the barcoding approach, their individual information will not sufficient to perfectly characterize the plant. . Multilocus identity is the powerful means to construct a barcode for plant [15]. Additive sequence information of all these three loci (conserved and polymorphic loci) is desirable tool to characterize plants.

Successful amplification of all Matk, rbcl and ITS genes (80% success rate) provides positive insight to the chances of covering wide groups of medicinal plants in Nepal. The average length of ITS (ITS1-ITS4) and rbcl around 400bp -700bp in the dataset exactly fits here, giving the errorless amplification of these genes. Similarly, MatK with average length 1500bp in the database and in this study was well-suited for DNA barcoding, which will give a free range of gene of their upper length upto 1500bp and lower length encompassing averages of ITS and MatK range for most of the medicinal plants belonging to angiosperm without using taxon-specific internal sequencing primers.

DNA Extraction reveals the 100% extraction rate, gives the applicability of protocols can be adapted to the broad range of plants. However, appropriate amplification of all three gene was not attained on two DNA samples from *Senegalia catechu* and*Schimawallichii*respectively. Reason for thiscontrasting results may not be an absolute factors and this may be due to one of the following: taxon nonspecific primer, inadequate thermocycler condition or the poor quality of extracted DNA in comparison to others (fig 1). It is insufficient claim the identity of species without the molecular sequence information. The yield of successful amplification of the essential genes for plant barcode lead us to conclude that its sequence could be retrieved and subsequent implementation of bioinformatics tools will lead to the identity of the plant and to explore their barcode on the database.

Conclusion

This study revealed the successful amplification of complete sets of both conserved and polymorphic gene necessary to characterize plants at molecular level.DNA barcoading is an emerging tool and to date the exact method or algorithm to be used in searching the barcode database has not been thoroughly investigated, particularly as regards a multilocus DNA barcode. Collective and efficient genomic data generated from Matk, rbcl and ITS sequencing will tremendously useful not only to identify the plant, it would be enough to provide them with unique barcode to trace them in the global digital database. DNA barcoding has great potential for enhancing ecological and evolutionary investigations if the right genetic markers are selected. The unexplored and unidentified medicinal plants which counts greatly on both economy and global concerns due to endangeredspecies, their nativeness in Nepal can be protected with their habitual identity if they are systematized with DNA barcoding approach.

References

1. Savolainen V, Cowan R, Vogler A, Roderick G, & Lane R (2005). Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions Of The Royal Society B: Biological Sciences, 360(1462), 1805-1811. Doi: 10.1098/rstb.2005.1730.

2. SchipmannU, Leaman D, Cunningham A, & WalterS, (2005). Impact of cultivation and collection on

the conservation of medicinal plants: global trends and issues. Actahorticulturae, (676), 31-44. Doi: 10.17660/actahortic.2005.676.3

3. Cowan RS, Chase MW, Kress WJ, Savolainen V. 2006. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55: 611–616.

4. Vere N, Rich T.C.G, Trinder S.A, Long C (2015). DNA Barcoding for plants. In: Batley J. (eds) plant genotyping. methods in molecular biology (methods and protocols), vol 1245.

5. Kress W, Erickson D, Swenson N, Thompson J, Uriarte M, & Zimmerman J (2010). Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. plos one, 5(11), e15409. Doi: 10.1371/journal.pone.0015409

6. KressW, Lopez I, & Erickson D (2012). Generating plant DNA barcodes for trees in long-term forest dynamics plots. DNA barcodes, 441-458. Doi: 10.1007/978-1-61779-591-6_22441–458.

7. Kress W (2017). Plant DNA barcodes: Applications today and in the future. Journal of Systematics And Evolution, 55(4), 291-307. Doi: 10.1111/jse.12254

8. Kembel S, & Hubbell S (2006). The phylogenetic structure of a neotropical forest tree community. Ecology,87(sp7), S86-S99. Doi: 10.1890/0012-9658(2006)87[86:tpsoan]2.0.co;2.

9. Westoby M, & Wright I (2006). Land-plant ecology on the basis of functional traits. Trends in ecology & evolution, 21(5), 261-268. Doi: 10.1016/j.tree.2006.02.004

10. Heckenhauer J, Barfuss M, & Samuel R (2016). Universal multiplexablematK primers for DNA barcoding of angiosperms. Applications in plant sciences, 4(6), 1500137. Doi: 10.3732/apps.1500137

11. Bafeel S, Arif I, Bakir M, Al Homaidan A, Al Farhan A, & Khan H (2012). DNA barcoding of arid wild plants using rbcL gene sequences. Genetics and molecular research, 11(3), 1934-1941. Doi: 10.4238/2012.july.19.12

12. White T, Gardes M,Fortin J, Bruns T, & Taylor J (1990). Identification of indigenous and introduced symbiotic fungi in ectomycorrhizae by amplification of nuclear and mitochondrial ribosomal DNA. Canadian journal of botany, 69(1), 180-190. doi: 10.1139/b91-026

13. Innnis MA, Gelfand DH, Sninsky JJ, White TJ (1990). PCR protocols: a guide to methods and applications. Academic Press, San Diego, CA.

14. Hollingsworth P, Forrest L, Spouge J, Hajibabaei M, Ratnasingham S, & Van der Bank M (2009). A DNA barcode for land plants. Proceedings of the national academy of sciences, 106(31), 12794-12797. Doi: 10.1073/pnas.0905845106

15. Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, & Gigot G. et al. (2008). DNA barcoding the floras of biodiversity hotspots. Proceedings of the national academy of sciences, 105(8), 2923-2928. Doi: 10.1073/pnas.0709936105