

ORIGINAL ARTICLE

Utility of DNA Barcoding Tool for Conservation and Molecular Identification of Intraspecies of Rice Genotypes Belonging to Chhattisgarh Using RBCL and MATK Gene Sequences

Jyoti Singh and Shubha Banerjee

Department of Plant Molecular Biology and Biotechnology, College of Agriculture,

Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh-492012, India

*Email: jyotisurya777@gmail.com

ABSTRACT

India is endowed with a great diversity of rice germplasm, among different state Chhattisgarh is well known for rice cultivation and called "rice bowl" of India. This genetic diversity among and between landraces, exists a wide scope for future crop improvement. In recent years, DNA barcoding has been suggested as a useful molecular technique to complement traditional taxonomic expertise for fast species identification and biodiversity inventories. In this study, in situ application of DNA barcodes was tested on selected diverse rice genotypes with the aim of contributing to the identification, conservation and protection of Intellectual Property Rights (IPR) of state. DNA Barcoding technique was successfully pioneered in animals using a portion of the cytochrome oxidase 1 (CO1) mitochondrial gene. In plants, establishing a standardized DNA barcoding system has been more challenging. Thus potential of the *rbcl* and *matK* markers for the selection of barcoding loci of rice genotypes. The panel of 231 diverse rice genotypes including germplasm lines, elite, varieties and wild rice were used. Our finding showed that amplification efficiency observed in panel of intraspecies of rice was in *rbcl* (89.6%) and *matK* (62.33%). On the basis of amplification efficiency panel of 24 rice genotypes selected for sequencing. The parsimony informative sites was estimated with maximum 305 sites recorded in *Matk*, followed by 264 sites *Rbcl* and number of variable sites reported highest in *rbcl* 672 followed by *MatK*-f 246. While nucleotide diversity per site π reported maximum in *rbcl* 0.21613 (MEGA 7.025). This scientific information data submitted to Barcoding of life database (BOLD) for generation of illustrative barcode.

Key words: DNA barcoding; biodiversity conservation, Cytochrome c oxidase I; Wild plants; *Rbcl*; *MatK*, IPR, BOLD.

Received 11.05.2018

Revised 21.06.2018

Accepted 23.08.2018

How to cite this article:

J Singh and S Banerjee. Utility of DNA Barcoding Tool for Conservation and Molecular Identification of Intraspecies Of Rice Genotypes Belonging To Chhattisgarh Using RBCL and MATK Gene Sequences. Adv. Biores. Vol 9 [5] September 2018: 85-90.

INTRODUCTION

Chhattisgarh is a state in Central India. It is the 10th largest state in India with a geographical area of 137, 90 thousand ha. Chhattisgarh stretches across the latitudinal expanse of 17°46' to 23°15' North on one hand to the longitudinal meridian of 80°30' to 84°23' East. Chhattisgarh known for rice cultivation and called "rice bowl" of India and considered as one of the secondary center of diversity. Further explorations in collaboration with NBPGR, New Delhi were organized and new collections were added to the gene pool which currently had 23,250 accessions including 210 accessions of wild species. This germplasm has only partially been characterized for various biotic and abiotic stress tolerances along with morpho-physiological traits. A few genes for gall-midge and BPH resistance have been identified. Molecular technology is considered a reliable alternative tool for the identification of plant species and DNA barcoding is the latest move towards the generation of universal standards. A DNA barcode is a universally accepted short DNA sequence allowing the prompt and unambiguous identification of species [14], promoted for a variety of biological applications [9], and including biodiversity inventories [2].

On the basis of analysis of recoverability, sequence quality, and levels of species discrimination, the Consortium for the Barcode of Life (CBOL) plant working group has recommended a standard barcode comprising ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcl*) and/or maturase K

(*matK*) for the barcoding of all land plants [1]. There are various researches going on plant barcoding suggests that the development of universal DNA barcoding markers for land plants is challenging; even the choice of the correct loci has been debated [3, 10, 7, 4]. The various scientists' shares their view about the selected core loci for plant barcoding are related to the lack of discriminatory power and/or primer universality [10]. The ability to discriminate between species using barcoding loci has proved more difficult in plants than animals, raising possibilities that plant species boundaries are less well defined. We evaluated the barcoding performance of *rbcL* and *MatK* in rice genotypes at intraspecies level discrimination using this marker. Our final objective is to provide a contribution to the future conservation and cataloging of these valuable resources of Chhattisgarh.

MATERIAL AND METHODS

The leaf samples from 231 diverse rice genotypes were used. DNA was extracted using MiniPrep method [5]. The concentration and quality of the extracted DNA were determined using gel electrophoresis and a Nano Drop spectrophotometer (Thermo scientific 30304-Ace-600). The isolated genomic DNA was stored at -20 C until used.

PCR and gene sequencing

A total volume of 20 µl of PCR reaction mixture contained the following: 2 µl (50 ng /µl) DNA, 2µl 10mM dNTPs mix (Invitrogen), 2µl of 10X PCR buffer with 15mM MgCl₂ (Invitrogen), 2µl of 10 pMo primer (1µl of each forward and reverse primer), 0.1µl of Taq DNA poly 5U/µl (Invitrogen) and rest was adjusted with nuclease free water (Sigma Aldrich). The primer *rbcL* (forward) atgtcaccacaaacagaaac, *rbcL* (reverse) tcgcatgtactcgtagc and *matK* (forward) cgatctattcattcaatattc *matK*(reverse) tctagcacacgaaagtcaagt pairs were used for the PCR (Imperial Life Sciences). The PCR was done Veriti 96-Well Thermal Cycler (Applied Biosystems) as follows: 94°C for 4 min, followed by 35 cycles of 94°C for 30 s, 50°C-65°C for 30s, and 72°C for 1 min, followed by an elongation step at 72°C for 7 min. A long (Horizontal electrophoresis unit Max Fill) 1.5% horizontal agarose gel using 1X TAE buffer containing 0.5µl/mL ethidium bromide was used for resolving PCR. Gel images were documented using a (Gel Doc XR+ BIORAD ET9970616AA) UV transilluminator opticom imaging system. The PCR product sizes were determined using a 100-bp ladder. PCR products were purified using (Thermo Scientific Gene JET Gel Extraction Kit) as per manufacturing instruction.

Data analysis

The DNA sequences of the all twenty-four genotypes were minimally edited using BioEdit software and aligned using ClustalW performed in MEGA 7.0.25 software. The analysis of DNA sequences was conducted by Neighbour-Joining to assess topology with MEGA 7.0.25. All the regions containing gaps and missing data were eliminated. The species identification and homology between the sequences was identified using BLAST method. The phylogenetic tree was developed using Neighbour-Joining (NJ) method which was tested with Kimura 2- parameter for evolutionary distances in MEGA 7.0 and node support was assessed on 1000 bootstrap replicated. Pairwise distance, transitional/transversional substitutions, and the maximum likelihood substitution matrix were estimated using MEGA 7.0.25 software. Genetic variation among the rice genotypes were estimated by calculating the number of haplotypes, haplotype diversity (HD), and parsimony informative sites using the DNAsp ver. 5.10. To test population expansion neutrality tests were performed to examine the null hypothesis that sequences are evolving according to the neutral expectations also submission of sequence data to Barcode of life data base for generation of illustrative barcode (BOLD).

RESULTS AND DISCUSSION

Chloroplast genome sequence specific loci were used in the study to assess their potential and identified candidate DNA Barcode loci for intraspecies discrimination in rice. Barcode regions must be relatively short in length to facilitate easy PCR amplification and DNA sequencing. The size of the amplification products obtained from other plant barcode loci ranged between 700-1200bp. The *rbcL* and *matK* loci were not amplified in all the 231 genotypes amplification efficiency observed in panel of intraspecies of rice was in *rbcL* (89.6%) and *matK* (62.33%). The *rbcL* gene, encoding "RUBISCO", *ribulose 1, 5-biophosphate-carboxylase/oxygenase* a free enzyme, present as single copy region of chloroplast genome and with intergenic spacer (600-800) nucleotide [14] is as an integral component for plant species discrimination [10].

The *rbcL* based DNA barcoding has been exhibit at efficiency of inter/intrageneric levels lived in cupressaceae, Cornaceae, Ericaceae, Graniaceae [8]. Several studies have reported that *rbcL* proved to be the most promising locus in terms of amplification and sequencing success in plants followed by *rpoC1*, *rpoB*, *matK*, *ITS2*, *trnH-psbA*, *trnL-F* as well as *psbK-I* and lastly *atpF-H*. The panel representative of 24

representative genotypes consisted of landraces, wild rice, variety, elite lines along with aromatic genotypes. The loci were amplified for validation based on sequencing to obtain informative data for analyzing intra species variation in rice. The PCR amplified products of expected size were extracted from the Gel then purified and then subjected to DNA sequencing. The DNA sequences were analyzed using *FinchTV* program. (<http://www.geospiza/products/finchtv.stml>) allows to view DNA sequence files and assess their quality from the chromatograms. Frequency of nucleotide substitution for *rbcl* and *matK* loci among 24 genotypes is mentioned in (Table 1) the overall average frequency of *MatK* loci was as follows: A=29.3, T/U = 36.2, C=18.4 and G= 16.0 and for *rbcl* A=28.1, T/U = 28.5, C=20.9 and G= 22.5 (Table 1). Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics (Table2). The Maximum likelihood values of maximum *transitional* rate was 53.56 and maximum *transversional* rate was 10.22 substitution of nucleotides based on the seven markers for 24 genotypes of rice calculated by MEGA 7.0.25. The estimated evolutionary divergence between Sequences recorded as 3.061 (average-1.236). The number of base substitutions per site from between sequences was shown in (Table 3).

Table 1: Average Nucleotide composition

Sr.no.	Marker name	T(U)	C	A	G	Total
1	MatK	36.2	18.4	29.3	16.0	1624.8
2	MatK-f	35.9	18.3	29.6	16.1	797.7
3	MatK-r	28.9	15.8	36.7	18.7	811.1
4	Rbcl-f	28.5	21.2	28.0	22.4	658.7
5	Rbcl-r	28.0	22.8	28.7	20.5	632.5
6	Rbcl	28.5	20.9	28.1	22.5	1276.6

Note:-All frequencies are given in percent.

Table 2: Maximum likelihood values of transitional (bold) and transversional (italics) substitution of nucleotides based on the seven markers for 24 genotypes of rice.

MatK-f	A	T	C	G	Rbcl-f	A	T	C	G
A	0	9.75	4.46	8.92	A	0	2.53	2.07	2.6
T	7.89	0	6.88	4.31	T	2.42	0	34.33	1.99
C	9.89	15.94	0	4.31	C	2.42	41.9	0	1.99
G	16.34	9.75	4.46	0	G	3.15	2.53	2.07	0
MatK-r	A	T	C	G	Rbcl-r	A	T	C	G
A	0	6.56	4.06	0	A	0	9.52	8.22	12.29
T	7.93	0	20.65	4.43	T	10.22	0	0	7.42
C	7.93	33.41	0	4.43	C	10.22	0	0	7.42
G	0	6.56	4.06	0	G	16.94	9.52	8.22	0
MatK	A	T	C	G	Rbcl	A	T	C	G
A	0	2.9	1.45	30.27	A	0	7.93	5.75	9.88
T	2.33	0	0.06	1.32	T	7.6	0	9.68	6.24
C	2.33	0.12	0	1.32	C	7.6	13.35	0	6.24
G	53.56	2.9	1.45	0	G	12.05	7.93	5.75	0

Table 3: Estimates of Evolutionary Divergence between Sequences

SR.NO.	MARKER	Minimum	Maximum	Average	No. of genotypes	No. of positions
1	MatK	0	1.686	0.151	23	1088
2	MatK-f	0	1.772	0.155	23	432
3	MatK-r	0	0.002	0	24	498
4	Rbcl-f	0	3.061	1.236	23	440
5	Rbcl-r	0	0.014	0.002	13	491
6	Rbcl	0	0	0.28	13	964

Table 4: Sequence polymorphism among the 24 rice genotypes based on *matK* and *rbcl* loci

SR.NO.	MARKER	VARIABLE SITES	PARSIMONY INFORMATIVE SITES	NUCLOTIDE DIVERSITY (per site)Pi
1	MatK	7	305	0.02611
2	MatK-f	246	0	0.01357
3	MatK-r	5	1	0.00099
4	Rbcl-f	236	235	0.21613
5	Rbcl-r	7	0	0.00219
6	Rbcl	672	264	0.18639

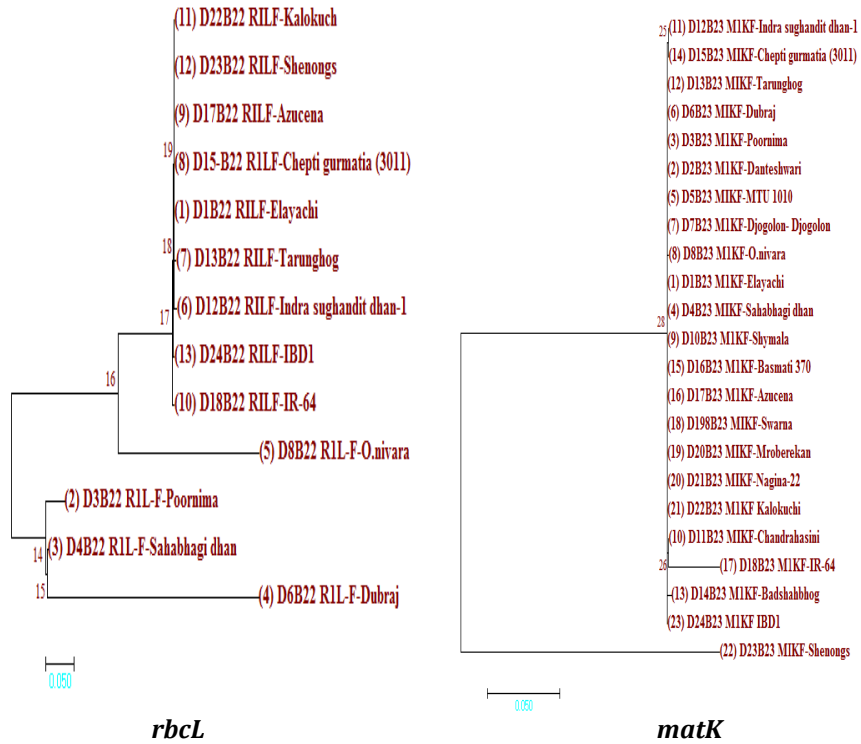


Fig.1: Phylogenetic tree showing Evolutionary relationship on the basis *rbcl* and *matK* loci inferred using the Neighbour joining method

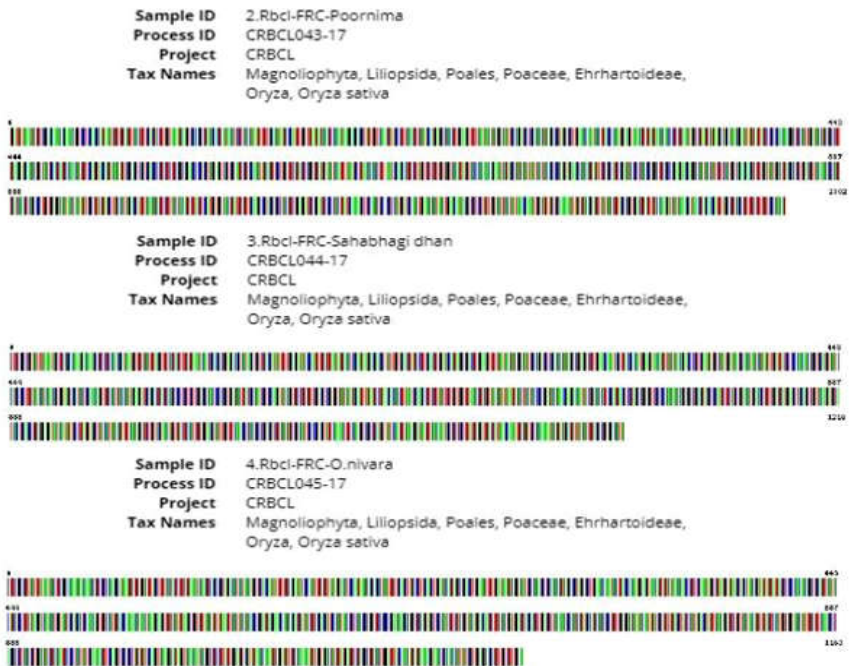


Fig 2. Barcodes of highly informative double stranded sequence of *rbcl* loci in Poornima sahbhagi and *O. nivara*



Fig 3. Barcodes of highly informative reverse stranded sequence of *matK* loci in Elaychi, Danteshwari and Poornima

Analyses were conducted using the Maximum Composite Likelihood model. The parsimony informative sites was estimated with maximum 305 sites recorded in *matK*, followed by 264 sites *RbcL* and 226 sites *trnK* and number of variable sites reported highest in *rbcL* 672 and *matK-f* 246. While nucleotide diversity (per site pi) reported maximum in *rbcL* 0.21613 (Table 4).

Molecular Phylogenetic analysis by Maximum Likelihood method

The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. For Maximum Likelihood method based (ML) phylogenetic analysis of each loci, the forward, reverse and both strand sequences were used.

The sequences of both the strand *matK* loci are having 7 haplotypes. IR 64 and Shenongs are distinctly different separated out in saperate clusters. While the strand sequences of the *rbcL* loci is having 9 haplotypes. The forward strand of loci *rbcL* separates *O.nivara*, Dubraj and poornima as most distinct genotypes grouped in separate cluster. Poornima and Sahabhagi dhan are distinctly similarly the reverse strand sequences of the *rbcL* loci are having 3 haplotypes. The discrimination ability of reverse strand of loci *rbcL* separates Indira sugandhit dhan 1 and poornima as most distinct genotypes grouped in separate cluster than other 22 rice genotypes (Fig.1).

Based on assessments of recoverability, sequence quality, and levels of species discrimination, CBOL in 2003 recommend the 2-locus combination of *rbcL+matK* as the plant barcode. In the present study the forward, reverse and both the strand sequences *matK* and *rbcL* loci shows significant identification ability among interspecies of rice genotypes also less discrimination ability while on the basis of results obtained variable sites, parsimony informative sites, nucleotide diversity (per site) Pi proves to be useful loci. While it is reported by various researchers that the generation of *matK* sequences for some plants groups seems to be problematic, because the part of the chloroplast genome underwent a large-scale restructuring during evolution [6]. There is no primer set which seems to be suitable for all lineages of land plants [9, 12, 13].

On the basis of analysis sequence data is submitted to Barcoding of Life Database (BOLD) and generated barcode (fig.2 & 3) Thus *rbcL* and *matK* shows good amplification efficiency but are less accurate for discrimination ability overall analysis proves *matK* and *rbcL* are reliable loci also along with combination with other loci will be alternative for identification of the intraspecies variation of rice genotypes but also are useful for explaining phylogenetic relationships.

In conclusion, this study provides preliminary assessment data that will be useful for wider application of DNA barcoding which will be useful to improve discrimination ability of intraspecies level in rice. as more

loci will be identified and validated using sequencing informative data for analyzing intra species variation in rice will be achievable. On the basis of overall analysis we found that *rbcl* and *matK* are useful for barcoding intraspecies of rice genotypes belongs to Chhattisgarh.

REFERENCES

1. CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceeding of the National Academy of Sciences* 106:31 12794-12797.
2. Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A (2011). Plant DNA Barcodes Can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE* 6: e26841. doi: 10.1371/journal.pone.0026841.
3. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi PR, Haidar N & Savolainen V (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philosophical Transactions of the Royal society of London*. 360: 1889-1895.
4. de Groot GA, During HJ, Maas JW, Schneider H, et al. (2011). Use of *rbcl* and *trnL-F* as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. *PLoS One* 6: e16371.
5. Doyle, J. and Doyle, J. (1987). Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem Bull*, 19(11): 11-15.
6. Duffy AM, Kelchner SA and Wolf PG (2009). Conservation of selection on *matK* following an ancient loss of its flanking intron. *Gene* 438: 17-25.
7. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.
8. Gielly, L. and Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcl* sequences. *Molecular Biol. Evol.*, 11: 769-777.
9. Hollingsworth ML, Andra CA, Forrest LL, Richardson J, et al. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* 9: 439-457.
10. Janzen, D.H., Hallwachs, W., Blandin, P., Burns, J.M., CADIQU, J., Chacon, I., Dapkey, T., Deans, A.R., Epstein, M.E., and Espinoza, B. (2009). Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular ecology resources* 9: 1-26.
11. Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., and Janzen, D.H. (2005). Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 102: 8369-8374.
12. Li FW, Tan BC, Buchbender V, Moran RC, et al. (2009). Identifying a mysterious aquatic fern gametophyte. *Plant Syst. Evol.* 281: 77-86.
13. Roy, S., Tyagi, A., Shukla, V., Kumar, A., Singh, U.M., Chaudhary, L.B., Datt, B., Bag, S.K., Singh, P.K., and Nair, N.K. (2010). Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. *PloS one* 5: e13674.
14. Savolainen V., Cowan R.S., Vogler A.P., Roderick G.K., Lane R. (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360:1805-1811.

Copyright: © 2018 Society of Education. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.