International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2018; 6(1): 686-690

Received: 17-11-2017
Accepted: 18-12-2017

Jyoti Singh
Department of Plant Molecular Biology and Biotechnology, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Datta P Kakade
Department of Plant Molecular Biology and Biotechnology, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Shubha Banerjee
Department of Plant Molecular Biology and Biotechnology, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Correspondence
Jyoti Singh
Department of Plant Molecular Biology and Biotechnology, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India

Fingerprint: Visual description of variation in multiple sequence alignments of intra-species of rice genotypes belonging to Chhattisgarh

Jyoti Singh, Datta P Kakade and Shubha Banerjee

Abstract
The usefulness of genetic material for use as DNA barcodes is under stable evolution and advancement as new barcodes offering better resolution and efficiency of amplification for specific species groups are identified. The panel of 24 rice genotypes including diverse germplasm lines, elite lines, landraces and wild rice belonging to Chhattisgarh were amplified using *atpH-atpI*, *petA-psbJ*, *trnK*, *rbcL*, *matK*, *trnlc-trnld*, *psbA-trnH* aligned sets of seven chloroplast specific loci, sequences were analyzed using fingerprint available at *http://evol.mcmaster.ca/fingerprint*. Fingerprint helps to identify similarities, differences, and patterns within a multiple sequence alignment which is biologically valuable because it permits visualization of the distribution of a particular feature and inferences about the structure, function, and evolution of the sequences in analysis also represent identity, variability, charge, hydrophobicity, solvent accessibility, and structure along with new visualizations based on composition, heterogeneity, heterozygosity, *dN*/*dS* and nucleotide diversity. The visualization based on composition, heterogeneity, *dN/dS* and nucleotide diversity on 24 were performed. The results shows even though *matK* and *rbcl* loci shows lesser number of haplotypes but are variable throughout the region. On the other part *psbA-trnH* shows variability in the specific region through showed good discrimination power in molecular phylogenetic analysis.

Keywords: Cytochromec oxidase I, DNA barcoding, fingerprint, multiple sequence alignment, polymorphism, sequence diversity

Introduction
Rice is the most important and widely cultivated crop in the world. Rice has more than forty thousand varieties species all across the globe. Asia is the home of rice as more than two billion people are getting 60-70 per cent of their energy requirement from rice and its derived products (Rekha *et al.,* 2015) [15]. Chhattisgarh is traditionally rich in rice diversity containing the wild progenitors of cultivated rice. The recent development of DNA barcoding provides a novel opportunity in delimitating closely related species. DNA barcoding helps to identify species on the basis of single- DNA region or a combination of a few DNA regions in the absence of taxonomic knowledge (Hebert *et al.,* 2003) [10], for these reason the DNA barcode regions should be of short lengths with high recovery rates (success rate for amplifying and sequencing) and have high species differentiation rate (CBOL plant working Group, 2009, Hollingsworth *et al.,* 2008, and Li *et al*., 2011) [3, 11, 13]. At present chloroplast regions were regarded as ideal candidates compared to the mitochondrial regions for plant barcoding by virtue of their faster mutation rate, and recombination, uniparental inheritance and higher recovery rate (Kress *et al.,* 2005) [12] The Barcode of Life Initiative has employed *COI* as the standard gene because it is able to discriminate between many closely related animal species (Hebert *et al* 2003) [10] and there is evidence to suggest that it also works well in algae (Saunders 2005) [16], arthropods (Smith *et al* 2005) [17], fish (Ward *et al* 2005) [19] and some plants (Kress *et al* 2005) [12]. Identifying sequence changes in homologous sites provides insights about the structure functional genomics, and evolution of a protein.
Unfortunately, numerous programs which require download and installation of software, support complicated documentation, impose a fee or limit the number of sequences allowed in the input file, graphical multiple alignment editors, such as clustal_x (Thompson *et al*. 1997) [18], Seaview (Galtier *et al*. 1996) [9], and JALVIEW (Clamp *et al*. 2004) [4], that display an alignment in its entirety. The problem is that it becomes difficult to summarize the characteristics or diversity of as ite relative to other sites within a multiple sequence alignment.

To qualitatively analyze up to 1000 sequences or more at lengths of over 1000 residues is very tedious, time-consuming and difficult.

To aid in such a task, there are a variety of multiple alignment shading programs available: alscript (Barton 1993) [1], espript (Gouet *et al*. 1999), boxshade, amas, weblogo (http://weblogo.berkeley.edu/) (Crooks *et al*. 2004) [5], sequence similarity presenter (Frohlich 1994) and TEXSHADE (Beitz 2000) [2]. Furthermore, most of these programs focus on providing sequence-by sequence representations and not alignment overviews. With the continued advancement in technology, increasing amounts of sequence data demands for a program which can be applied to a wide variety of data sets from any sequence data and answer to which is Fingerprint. Overall, fingerprint is an effective tool to quickly and intuitively view the similarities, differences, and patterns in a multiple sequence alignment. The human eye can quickly assimilate these patterns, making data exploration much easier. The concurrence composition fingerprint there is possibility of loss of information since the tool represents the residues with the highest frequency of occurrence. To prevent this loss of information, an alternate presentation is encoded. Each possible residue at a given site corresponds to a distinct coloured percentage of the vertical line drawn to represent a site. The heterogeneous composition of an alignment is viewed using a *heterogeneity* fingerprint. The invariable sites (represented by only one residue are represented by one colour that extends for the entire length of the vertical line representing that particular site; the colour is determined by the residue. If, at a particular site, one residue occurs with a frequency of 0.25 and the second occurs with a frequency of 0.75, then the former colour will represent 25% of the height of the drawn line, and the latter will represent the remaining .75% (Meleanie Lou and G.Brian Golding, 2007) [14].

Material and methods

The experimental materials consisted diverse rice genotype including of 24 germplasm lines, elite lines, landraces and wild rice collection from undivided Madhya Pradesh and Chhattisgarh, popular rice varieties and advanced breeding lines for drought tolerance, yield improvement and nutritional quality traits, Selected 24 genotypes Elayachi, Denteshwari, Poornima, Sahabhagi dhan, MTU 1010, Dubraj, Djogolon – Djogolon, *O.nivara*, *O.nivara*, Shyamala, Chandrahasini, Indira Sugandhit dhan-1, Tarungbhog, Badshahbhog, Cheptigurmatia (3011), Basmati 370, Azucena, IR 64, Swarna, Moroberekan, nagina-22, Kalokuchi, Shenongs, IBD 1.Genomic DNA extraction and PCR amplification: The total genomic DNA was extracted from single tagged plant, by MiniPrep method (Doyle and Doyle, 1987) [7]. The seven chloroplast specific primer *atp H-atpI*, *petA-psbJ*, *trnK*, *rbcL*, *matK*, *trnlc-trnld*, *psbA-trnH* were used for PCR amplification (Dong *et al*. 2012 and Holligsworth *et al*. 2011) [6].

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 PCR was done Veriti follows: 94 ^{0}C for 4 min, followed by 35 cycles of 94 ^{0}C for 30 sec, 55 ^{0}C for 30s, and 72 ^{0}C for 1 min, followed by an elongation step at 72 ^{0}C for 7 min. Resolving PCR product on Agarose gel:PCR products were verified by electrophoresis in 1.5% agarose gel using stained with ethidium. One specific band was recorded at 1000bp on the gel. The PCR product were sent to the Thermo Fisher Scientific Services, New Delhi, India for sequencing based on Sanger method after purification using (Thermo Scientific Gene JET Gel Extraction Kit). The purpose of PCR cleanup is to remove salts, extra nucleotides and primers before sequencing.

Results and discussion

The Visualizations based on composition, heterogeneity, dN/dS and nucleotide diversity were performed on the basis of FINGERPRINT. Figure one to seven shows the results of nucleotide fingerprints based on the seven *atpH-atpI*, *petA-psbJ*, *trnK*, *rbcL*, *matK*, *trnlc-trnld*, *psbA-trnH* gene in 24 selected rice genotypes. Melanie lou and Brain golding., 2007 [14] also reported nucleotide fingerprints based on the cytochrome c oxidase I (COI) gene from the order strepsiptera (twisted-wings parasite), suggested that Fingerprint is effective for identifying sequence variation and for preparing high resolution, intuitive graphics presentation.

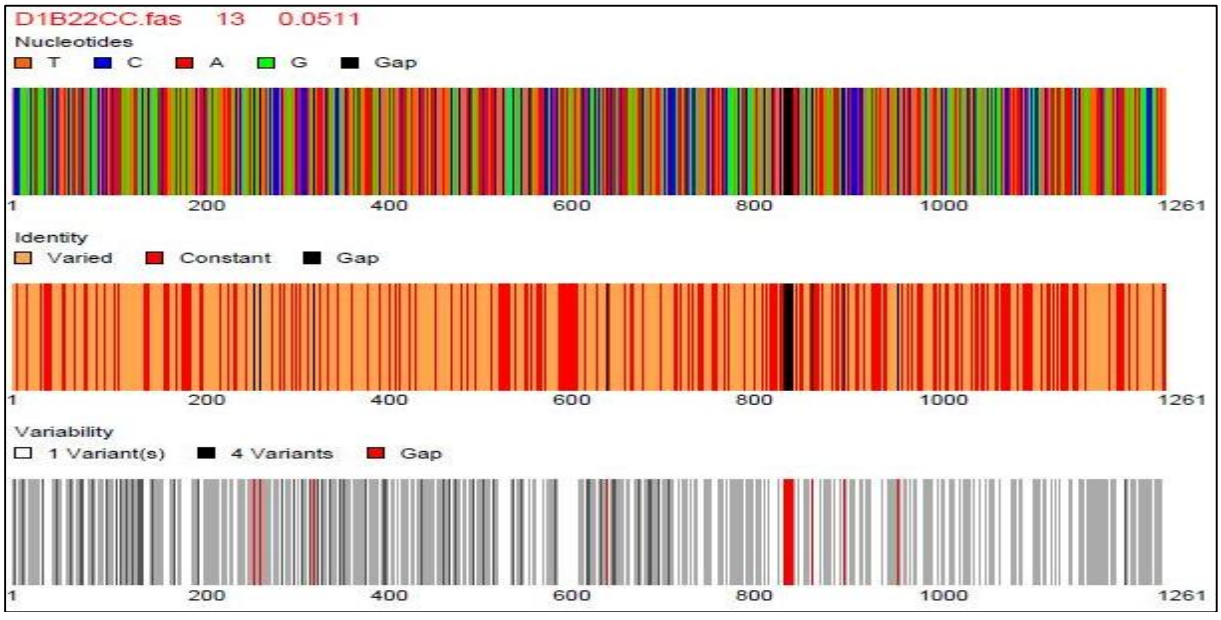

Fig 1: Fingerprint: visual delineation of variation in multiple sequence alignments of *rbcl* loci

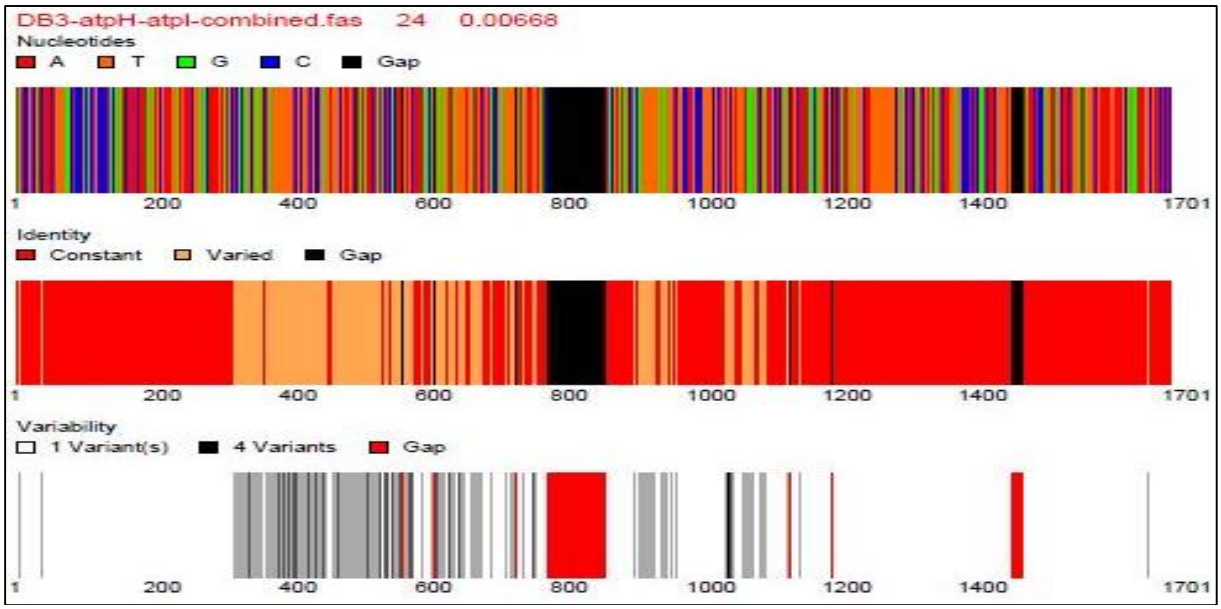

Fig 2: Fingerprint: visual delineation of variation in multiple sequence alignments of *atpH-atpI* loci

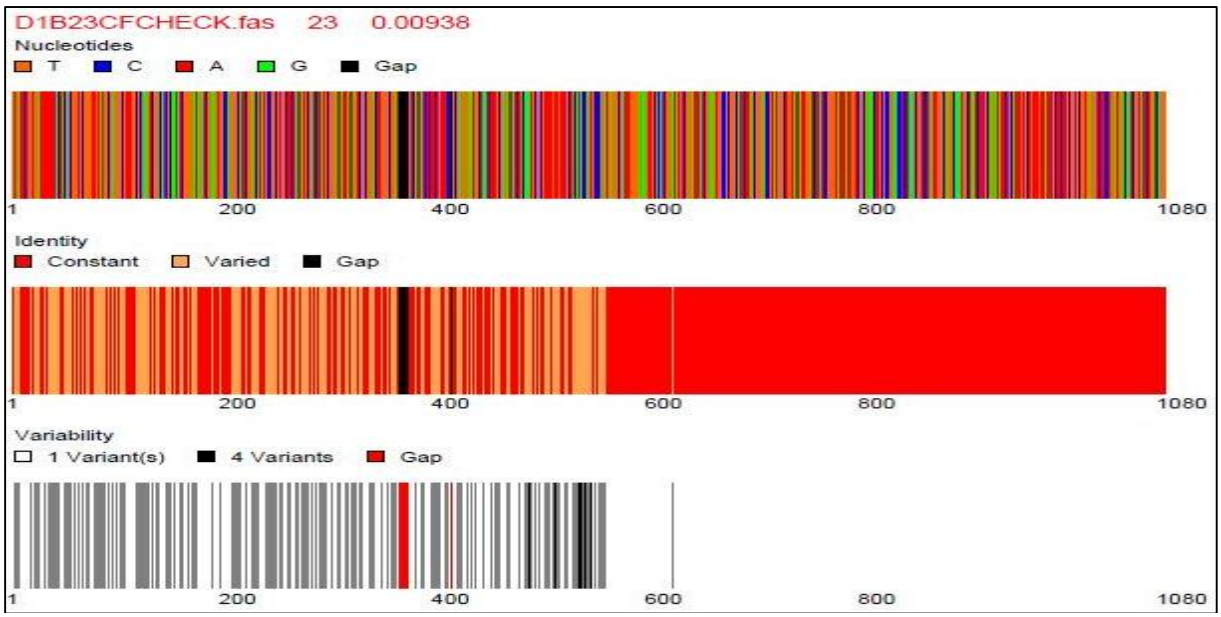

Fig 3: Fingerprint: visual delineation of variation in multiple sequence alignments of *matK* loci

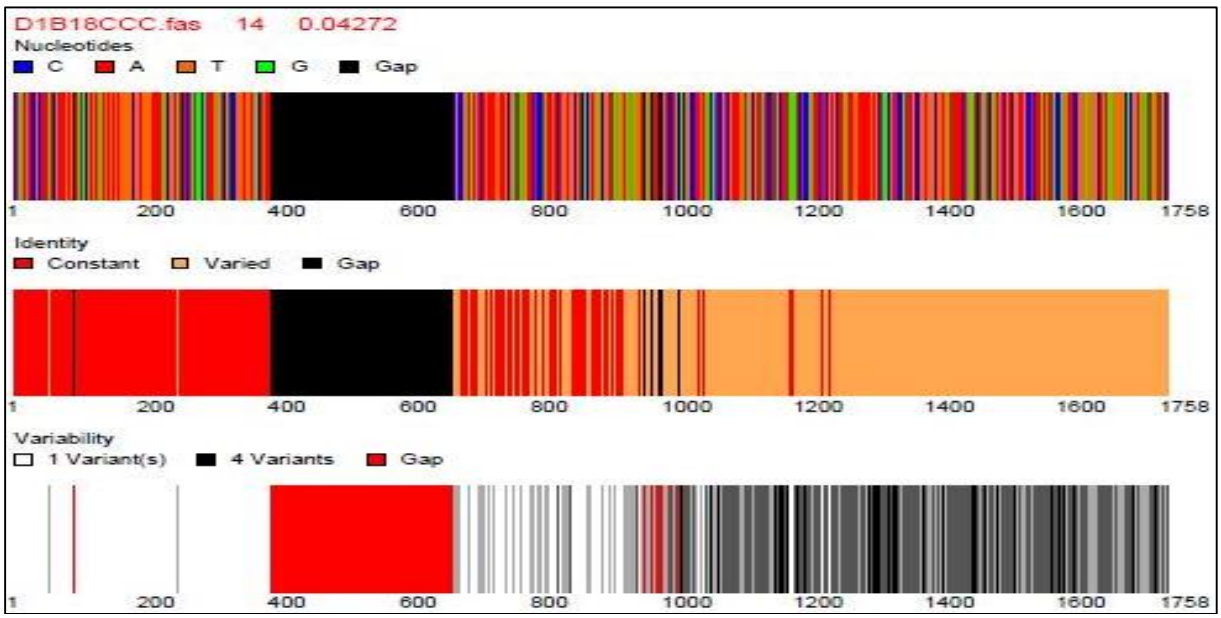

Fig 4: Fingerprint: visual delineation of variation in multiple sequence alignments of *trnlc-trnld* loci

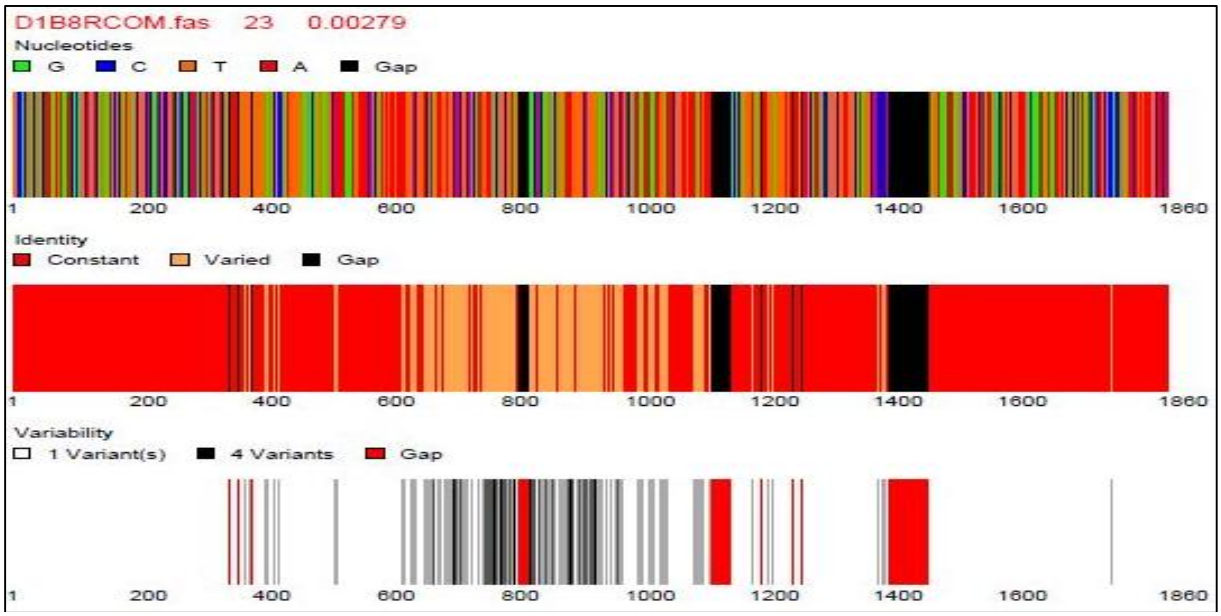

Fig 5: Fingerprint: visual delineation of variation in multiple sequence alignments *petA-psbJ* of loci

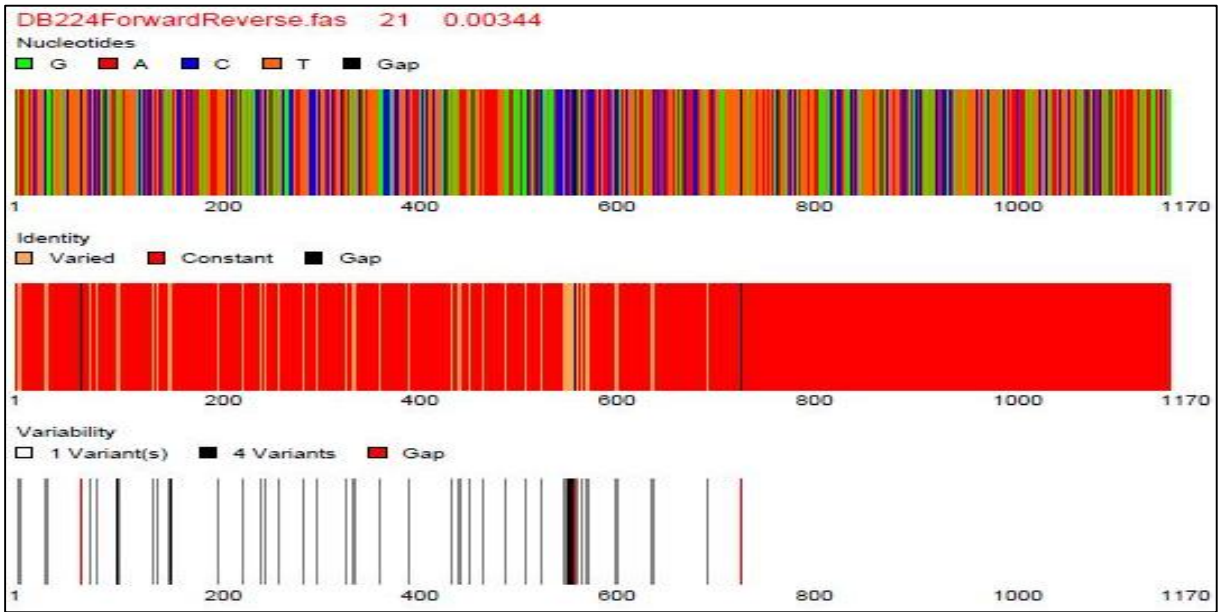

Fig 6: Fingerprint: visual delineation of variation in multiple sequence alignments *psbA-trnH* of loci

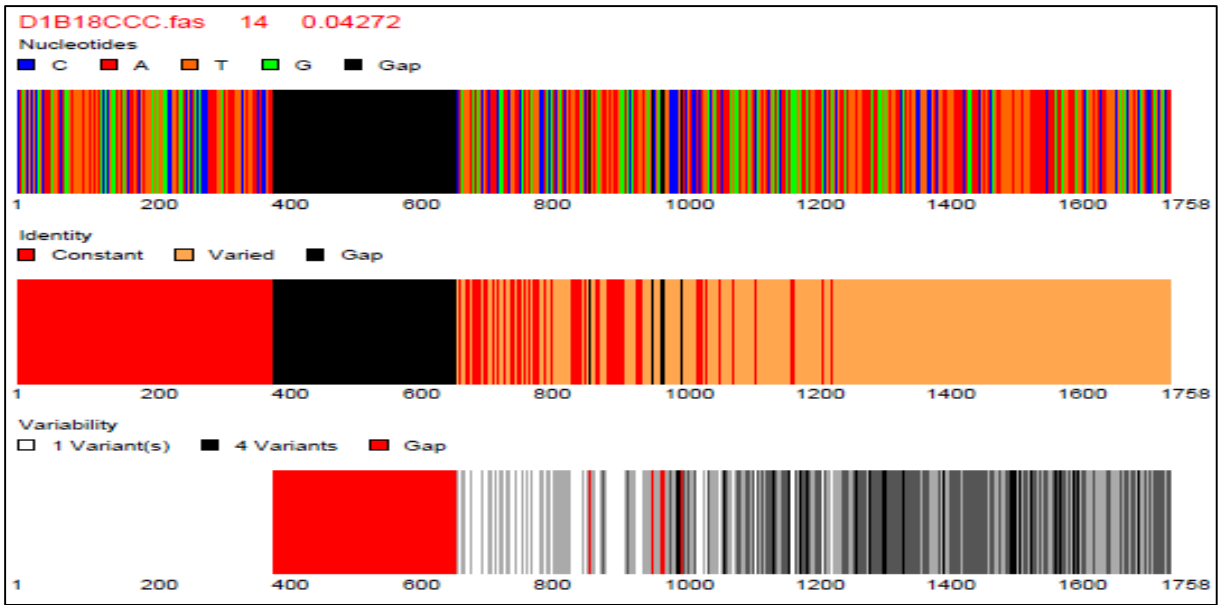

Fig 7: Fingerprint: visual delineation of variation in multiple sequence alignments *trnk* of loci

The outcomes of the study based on the use of Fingerprint, a web server application produces diagrams called fingerprints, show that even though *matK and rbcl* loci show lesser number of haplotypes but are variable throughout the region. On the other part *psbA-trnH* shows variability in the specific region though showed good discrimination power in molecular phylogenetic analysis. In general Fingerprint is an efficient tool to fast and intuitively view the similarities, differences and patterns in multiple sequence alignment. The human eye can quickly learn these patterns, making data exploration easier.

References

1. Barton GJ. Alscript: A tool to format multiple sequence alignments. Protein Engineering Design Protein and Selection, 1993, 637-40.
2. Beitz E. texshade: shading and labeling of multiple sequence alignments using latex 2 epsilon. Bioinformatics. 2000; 16:135-139.
3. CBOL Plant Working Group. A DNA barcode for land plants. Proceeding of the National Academy of Sciences. 2009; 106(31):12794-12797.
4. Clamp M, Cuff J, Searle SM, Barton GJ. The jalview Java alignment editor. Bioinformatics, 2004; 20:426-427.
5. Crooks GE, Hon G, Chandonia JM, Brenner SE. weblogo: a sequence logo generator. Genome Research, 2004; 14:1188-1190.
6. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PloS one, 2012; 7:e35071.
7. Doyle J, Doyle J. Genomic plant DNA preparation from fresh tissue-CTAB method. Phytochem Bull, 1987; 19(11):11-15.
8. Felsenstein J. phylip-Phylogeny Inference Package (Version 3.2). Cladistics, 1989; 5:164-166.
9. Galtier N, Gouy M, Gautier C. seaview and phylo_win: two graphic tools for sequence alignment and molecular phylogeny. Computer Applications in the Biosciences. 1996; 12:543-548.
10. Hebert PD, Cywinska A, Ball SL. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences, 2003; 270:313-321.
11. Hollingsworth PM. DNA barcoding plants in biodiversity hot spots: progress and outstanding questions. Heredity, 2008; 101:1-2.
12. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences, USA, 2005; 102:8369-8374.
13. LI DZ, Liu JQ, Chen ZD, Wang H, GE XJ, Zhou SL *et al*. Plant DNA barcoding in China. Journal of Systematics and Evolution. 2011; 49:165-168.
14. Melanie Lou, Brian Golding G. Fingerprint: visual depiction of variation in multiple sequence alignment, 2007.
15. Rekha M, Jayadeva H, Kombali G. Economic evaluation of drip fertigation in aerobic rice. In Compendium of abstracts of the 2nd international conference on bio-resource and stress management, ANGRAU & PJTSAU, Hyderabad, India, 2015, 7-10.
16. Saunders GW. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 2005; 360:1879-1888.
17. Smith MA, Fisher BL, Hebert PD. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2005; 360:1825-1834.
18. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The clustal_x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 1997; 25:4876-4882.
19. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 2005; 360:1847-1857.