JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Single Nucleotide Polymorphism for Assessing Genetic Diversity in *Vitex trifolia* L.

R. Mary Sujin

Research Department of Botany, Pasumpon Thiru Muthuramalinga Thevar Memorial College,

Kamuthi, Ramnad District - 623 604, Tamilnadu, India; e-mail: sujiphd@gmail.com

Abstract

Single Nucleotide Polymorphism (SNPs) has become the most widely used markers in many current genetic applications. We analyzed the population genetic of *Vitex trifolia* L. collected from five different locations of Western Ghats of India. To assess genetic diversity we conducted sequencing of the genomes of five populations of *V. trifolia* and compared the data to a reference genome assembly. Three SNPs were discovered when compared the sequence of five different populations by direct partial sequencing of genes (rbcL). UPGMA clustering indicated two main groups. Our genome-wide assessment of SNP variation in *V. trifolia* revealed low levels of genetic variation. This approach determining population genetic structure using SNPs from genome-wide comparison constitute a framework for high-throughput analyses genetic diversity in plants, particularly in species with limited genetic diversity.

Keywords: Genetic diversity, *Vitex trifolia* L. SNP markers

1. Introduction

Determining the extent and distribution of genetic diversity is an essential component for conservation strategies. Assessment of genetic diversity in medicinal plants has involved increasingly sophisticated approaches, from early allozyme work to amplified frequent length polymorphisms (AFLPs) and microsatellite (Estoup et al. 1998).

DNA makers are valid tools for the evaluation of biological materials, both for genetic diversity studies and for the discrimination of samples (Angiocillo et al. 1999; Bandely et al. 2002; Garant and Kruuk 2005; Morgante and Salamini 2003). It is therefore, a natural extension for this molecules approach to be applied for medicinal plants for identifying the superior genotypes (Dal Marso et al. 2004; Hernandez et al. 2005).

Recently Single Nucleotide polymorphism (SNP) for devised markers, identified in coding sequences of different genes, has been developed to discriminate among populations (Consolandi et al. 2007; Reale et al. 2006). Technological improvements make the one of SNP attractive for high throughput use in markers assist breeding for population studies and to develop high density linkage maps for map-based gene discovery (Palmieri et al. 2004).

SNPs are a viable alternative for assessing population genetic structure for several reasons. SNPs are binary, codominate markers, heterozygosity can be directly measured. Second, unlike micro satellites their power-comes not from the number of alleles, but from the large number of loci that can be answered. Thus, even in a low diversity species the genetic population discrimination power can be equivalent to the same number of loci in a genetically diversity species; once the rare SNPs are discovered. Third, the more evolutionary concerned nature of SNPs makers them less subject to the probe of homoplasy (Brumfield et al. 2003). Finally, SNPs are amenable to high throughput automation, allowing rapid and efficient genotyping of large numbers of sample (Tsuchihashi and Dracopoli 2002). Thus far, the major obstacle has been to discover rare polymorphic sites, but novel sequencing approaches are now mitigating this issue. (Rabinowicz et al. 1999). In *V. trifolia* the assessment of genetic variability has been carried out with SNP markers.

Vitex trifolia L. (Verbenaceae) is a stout, aromatic shrub or a small tree found wild in several parts of India, which is traditionally used by the tribes and native medical practitioners for the treatment of various ailments including liver disorders, tumors, rheumatic pains, inflammation, sprains, fever and used in the treatment of tuberculosis (Anonymous 2003). *V. trifolia* possess larvicidal, wound healing, anti HIV, anticancer, trypanocidal, antibacterial and antipyretic activities (Kannathasan et al. 2007; Manjunatha et al. 2007).

The aim of this work was to discover new SNPs and genetic diversity among five population of *V. trifolia* from Western Ghats of India among SNPs markers.

2. Meterials and Methods

2.1. Plant Material

Five accessions of *V. trifolia* originating from different growing areas of Western Ghats of India were used in this study. Young leaves of each sample uses take in plastic bags for transport from the field to the laboratory. The samples were maintained in deep freezer at.70%.

2.2. DNA extraction from leaves

Genomic DNA was extracted from fresh leaves of the *V. trifolia* by using a CTAB method Uta Pich and Ingo Schubert (1993). After extraction, the samples were treated with RNase A for 3o min at 37 °C and run in 1 % agarose gel in TAE 1X buffer in presence of ethidium bromide (1µg/ml).

2.3. Primer desing

PCR primers were designed using primer 3 program (Rozen and Skaletsky 2000) and the sequence of partial cDNA of rbcL genes available from GenBank at NCBI website (http://www.ncbi.nlm.nih.gov). The sequences were amplified by polymerase chain reaction (PCR). Sequence of the rbcL genes were PCR amplified for the five different accessions of *V. trifolia* using the primer: rbcL F: 5´ ATGTCACCACAAACAGAGACTAAAGC 3´ and rbcL R: 5´ GTAAAATCAAGTCCACCRCG 3´.

2.4. PCR and Sequencing

PCR reaction mixes were prepared for each sample by mixing 5µl of 10X PCR buffer, 2 µl 10 mM dNTPs, 5 µl of 10 mM $MgCl_2$, 1 µl of primer (at 10 µM), 0.5 U of Taq DNA polymerase, 10 µl of diluted genomic DNA (25 ng/ µl) and 25.5 µl of water. Cycling parameters were as follows: 94°C for 5 min followed by 40 cycles of 94 °C for 30 second, hybridizing step for 1 min. Products were separated by agarose gel electrophoresis to check for efficiency of amplification and to ensure that only a single product of the expected size was present. PCR products were then purified by passage through Wizard^R 'SV gel and PCR clean-up system purification columns and sequenced tree times from either end using the same forward primers are used in initial terminator cycle sequencing kit version 3.1 (Applied parameters) was used according to manufacturer's instructions.

2.5. Marker discover strategies

The most direct approach to the discovery of DNA polymorphisms is direct sequencing of PCR products from five populations of *V. trifolia*. Polymorphisms between the sequences were identified by sequence alignment. Furthermore, the trace outputs from the sequences were evaluated by eye to identify possible areas of heterozygous sequence.

2.6. Genetic diversity and Data analyzes

For each SNP marker two alleles are genetically present yielding five populations of V. trifolia. Allelic and genotype frequencies for each marker were estimated by simple counting. Power of discrimination (PD) was calculated using following formula (Kloosterman et al. 1993).

$$PD=1-\sum_{i=1}^{g} f_i^2$$

Where f_i is the frequency of the *i* th genotype and the sum is overall genotypes. The SNP genotypes were then recorded in a 0/1 data matrix, where 1 indicates the presence of a given allele and o its absence. So each SNP was transformed into two columns with three possible states: 1/1 for heterozygote and 0/1 or 1/0 for the two homozygous genotypes.

Jaccard's similarity coefficient was calculated to measure the genetic variability between five populations (Jaccard 1908). Based on the genetic similarity matrix, the populations were clustered by the unweighed pair group method with arithmetic averaging (UPGMA) using the program NTSYS –pc version 2.1 (Rohlf 1999).

3. Results and Discussion

3.1. Targeted sampling

The total length of high quality sequences analyzed obtained for the rbcL genes was 593 bp. Comparison of these sequences to the cDNA original sequences from which the primers were designed by BLAST2 seq (Tatusova and Madden 1999). SNP discovery was validated using direct sequencing of PCR products and comparative analysis of sequences of five different populations of *V. trifolia* (Fig 1). Three SNPs were finally identified within the sequences of the genes (Table 1; Fig 3).

Fig 1. PCR products of five different accession of *V. trifolia* using SNP primer

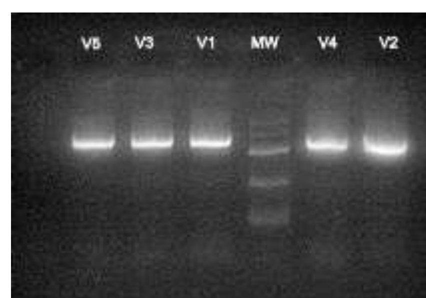
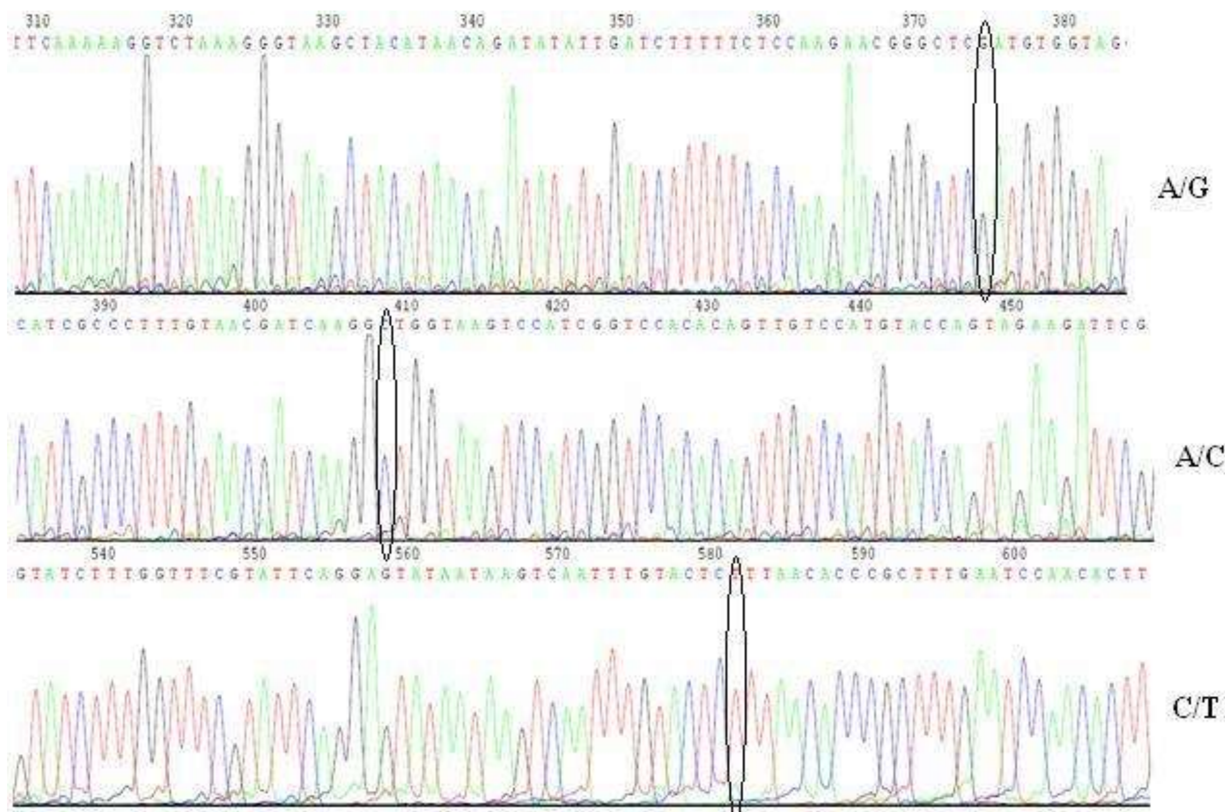


Table 1. Position and types of the three SNPs discovered

Gene	SNP position	Type
rbcL 1	409	A/C
rbcL 2	375	A/G
rbcL 3	582	C/T

Fig 2. SNP position in the sequence comparison graph using the program



3.2. Marker characteristics

The genotype and allelic frequencies of the three SNPs and their power of discrimination (PD) varies from 0.63 to 0.72 for the marker rbcL with an average value of 0.66. The relatively 'poor' value is expected since SNP markers are biallelic. The observed heterozygoisty (*Ho*) ranged from 0.16 to 0.50. The genotypic frequencies and allelic frequencies of each SNPs were varied (Table 2).

Table 2. Features and frequencies of new SNPs

Genes	SNP	*Ho*	PD	Genotypic Frequencies			Allelic Frequencies	
rbcL	rbcL 375	0.33	63	28.33A/A	22.59G/G	8.58A/G	56.32:A	38.61:G
	rbcL 409	0.16	72	28.16A/A	19.86C/C	8.06A/C	56.99:A	40.64:C
	rbcL 582	0.50	64	19.73C/C	28.16T/T	8.07C/T	40.64:C	56.49:T

3.3.

Genetic diversity level

The three SNP markers in *V. trifolia* from five populations and assess their potential in studying genetic diversity. The lower coefficient of similarity (GS=0.998) was observed between the most of the populations. The highest similarity (GS=1) occurred between V_4 & V_5 and V_1&V_2.

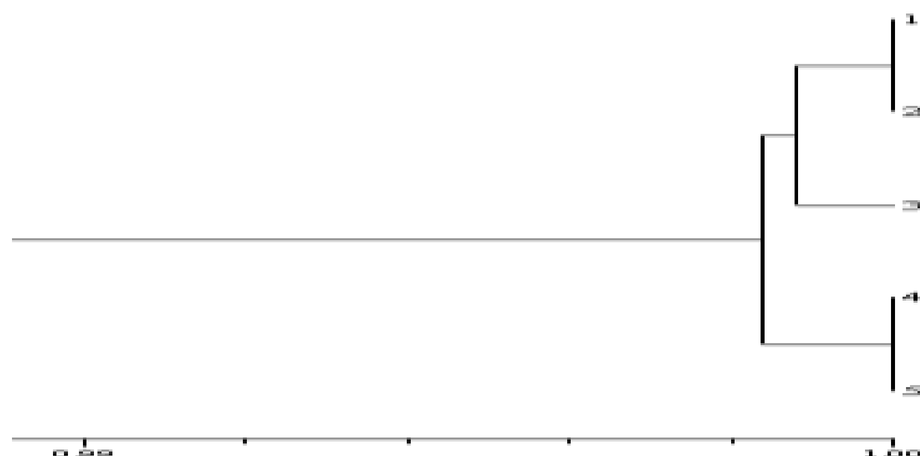
3.4. Patterns of genetic diversity

The present study revealed a relatively low level of genetic diversity in *V. trifolia* based on SNP markers. Using UPGMA algorithm a dendrogram was constructed to infer phylogenetic relationships between the five populations (Fig. 3). A Jaccard's matrix was used to produce a dendrogram based on SI, which showed distinct separation of the five accessions into two major cluster having 99.8 % similarity. Among the two major clusters, the accessions belonging to the lowest cluster (LC) were collected from V_4 and V_5 while accessions V_1, V_2 and V_3 belongs to the upper cluster (UC) and the sub clusters had 99.9 % similarity. Further, the accessions of the UC were grouped in to two major sub clusters USC1 from V_1 and V_2 having 100 % similarity and USC2 from V_3. The lower cluster (LC) was divided into two sub cluster LSC1 and LSC2. The lower sub cluster LSC1 was further sub divided into two (LSC1A) from V_4 and (LSC1B) from V_5 with a similarity of 100 % (Table 3; Fig 3).

Table 3. Similarity indices of five population of *V. trifolia* using SNP markers

Pop	1	2	3	4	5
1	1.00				
2	1.00	1.00			
3	0.999	0.998	1.00		
4	0.998	0.998	0.998	1.00	
5	0.998	0.998	0.998	1.00	1.00

Fig 3. Dendrogram of *V. trifolia* using Jaccard similarity coefficients from SNP markers and UPGMA method



Coefficient of similarity of Jaccard

Our assessment of genome wide diversity of *V. trifolia* suggests that it has low genetic diversity and structure for all populations. Even our upwardly biased estimate of nucleotide diversity is far less than the average number of SNPs found in plants *Ricinus Communis* (Imen et al. 2010). Low genetic diversity is likely a consequence of a genetic bottleneck due to domestication, as seen in a range of other plants. Alternative, the fragmentation of populations, subsequent loss of gene flow and the effects of genetic drift could also account for loss of heterozygosity, but more research on the timing of introductions is needed to verify these alternative explanations (Jeffrey et al. 2010). Furthermore, our data suggest that SNPs may be necessary for better resolution of relationships of samples among populations.

The power of SNP discovery should not be misconstrued as an indication of diversity in a species that shows low overall genetic diversity; our SNP discovery found relatively few SNPs despite extensive survey of several other genome of *V. trifolia*. Although chloroplast markers have been effectively used for studying plant distributions, low effective population size in chloroplast DNA and reduced genetic diversity, compared with nuclear DNA. Chlorplast SNPs are more variable and amenable to high throughput genotyping and will likely be the marker of choice for population-level analyses of species with sequences genomes (Estoup et al. 2002).

From this study, we have identified 3 SNPs that can be used as genetic markers in *V. trifolia*. Cluster analysis showed that diversity among the different accessions of *V. trifolia* corresponds well with the geographic origins of each population. Similar work was done in *Eurycoma longifolia* by Asiah *et al.* (2003); in *Ricinus communis* by Jeffrey et al. (2010); in Oat by Shiaomon et al. (2011); *Panaz ginseng* by

Hongtao et al. (2010). These markers should prove useful in preserving genetic diversity among populations. The distinction between the populations could be explained by its geographical isolation, which is expected to limit gene exchange between populations (Asiah et al. 2003).

Conclusions

Our study demonstrates the utility of a SNP-based approach for assessing the population genetics of medicinal plants. As new sequencing technologies emerge and more genome become more available, our approach promised to be particularly useful for plant population studies due to the resolving power of SNPs and the ability to rapidly assess diversity in a large number of samples. This study represents one of the most extensive genomic studies of worldwide SNP variation in this medicinal plant. With rapidly increasing capabilities in genome sequencing, this work provides a template for assessing population structure. These results indicate that SNP markers would provide a rapid and easy way to establish a fingerprint of each population for genetic variability.

Acknowledgement

The authors are grateful to the Indian Council of Medical Research, Government of India, New Delhi (Ref: 59/12/2006/BMS/TRM dt. 26.03.2009) for financial support.

References

1. Angiolillo A, Mencuccini M, Baldoni L (1999). Olive genetic diversity assessed using amplified fragment length polymorphisms. Theoret Appl Genet 10:411-421.
2. Anonymous (2003). The wealth of India – Raw Materials, Vol. X, Council for Scientific and Industrial Research. New Delhi, pp 525.
3. Asiah Osman, Barbara Jordan, Philip A, Lessard, Norwati Muhammad M, Rosli Haron, Norifiza Mat Riffin, Anthony J, Sinskey, ChoKyun Rha, David Housman E (2003). Genetic Diversity of *Eurycoma longifolia* Inferred from Single Nucleotide Polymorphisms. Plant Physiol 131(3):1294–1301.
4. Bandelj D, Jakse J, Javornik B (2002). DNA fingerprinting of olive varieties by microsatellite markers. Food Technol Biotechnol 5:185-190.
5. Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003). The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18:249-256.
6. Consolandi C, Palmieri L, Doveri S, Maestri E, Marmiroli N, Reale S, Lee D, Baldoni L., Tosti N, Severgnini M, De Bellis G, Castiglioi B (2007). Olive variety identification by ligation detection reaction in a universal array format. Afr J Biotechnol 9:565-574.
7. Dalmasso A, Fontanella E, Piatti P, Civera T, Rosati S, Bottero MT (2004). A multiplex PCR assay for the identification of animal species in feedstuffs. Mol Cell Probes 6:81-87.

8. Estoup A, Angers B (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. *Adv Mol Ecol.* Amsterdam:IOS Press Carvalho GR, pp 55-86.
9. Estoup A, Jame P, Comuet JM (2002). Homoplasy and muatation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol* 11:1591-1604.
10. Garant D Kruuk LE (2005). How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol Ecol* 15:1843-1859.
11. Hernandez M, Esteve T, Pla M (2005). Real-time polymerase chain reaction based assays for quantitative detection of barley, rice, sunflower and wheat. *J Agric Food* Chem 6:7003-7009.
12. Hongtao Wang, Hua Sun, Woo-Saeng Kwon, Haizhu Jin Deok-Chun Yang (201). A PCR based SNP marker for specific authentication of Korean gingeng (Panax gingeng) cultivar “Chunpoong”. Mol Biol Rep 37(2):1053-1057.
13. Imen Rekik Hakim, Nazhiha Grati Kammoun, Emna Makhloufi, Ahmed Rebai (2010). Discovery and potential of SNP markers in characterization of Tunisian olive germplasm. Diversity. 2:17-27.
14. Jaccard P (1908). Nouvelles recherché sur la distribution florale. Bull Soc Vaud Sci Nat 44:223-270.
15. Jeffrey Foster T, Gerard Allan J, Agnes Chan P, Pablo Rabinowicz D, Jacques Ravel, Pai Jackson J, Paul Keim (2010). Single nucleotide polymorphisms for assessing genetic diversity in castor bean (Ricinus communis). BMC Plant Biol 10 (13):1-11.
16. Kannathasan K, Senthilkumar A, Chandrasekaran M, Venkatesalu V (2007). Differential larvicidal efficacy of four species of *Vitex* against *Culex quinquefasciatus* larvae. Parasitol Res 101:1721-1723.
17. Kloosterman AD, Budowle B, Daselaar P (1993). PCR- amplification and detection of the human DIS80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. Int J Legal Med 7:257-264.
18. Manjunatha BK, Vidya SM, Krishna V, Mankani KL, Singh SD, Manohara YN (2007). Comparative evaluation of wound healing potency of *Vitex trifolia* L. and *Vitex altissima* L. *Phytother Res* 21:457-461.
19. Morgante M, Salmini F (2003). From plant genomics to breeding practice. Curr Opin Biotechnol 5:214-219.
20. Palmieri L, Vietina M, Bottura C, Donini P, Nelson M (2004). SNP development in olive and their application on varietal determination in olive oil. In Proceedings of the XLVIII It Joint Meeting, Lecce, Italy, pp 15-18.
21. Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie, WR Martienssen RA (1999). Differential methylation of genes and retrotransposons facilities shotgun sequencing of the maize genome. *Nat Genet* 23:305-308.
22. Reale S, Doveri S, Diaz A, Angiolillo A, Lucentini L, Pilla F, Martin A, Donini P, Lee D (2006). SNP-based markers for discriminating olive (*Olea europapaea* L.) cultivars. Genome 15:1193-1209.

23. Rohlf FJ (1999). NTSYS-PC: Numerical Taxonomy and Multivariate Analysis System (Version 2.1), Exeter Software, Setauket, NY, USA.
24. Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 21:365-386.
25. Shiaoman Chao, Rebekah Oliver, Gerard LAzo, Nick Tinker, Jean-Luc Jannink, Rachel Redman, Eric Jackson (2011). Analysis of genetic diversity using SNP markers in Oat. Plant and Animal XIX conference. Jan.15-19. Town & country convention center. San Diego, CA.
26. Tatusova TA, Madden LT (1999). BLAST 2 sequences. A new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 3:247-250.
27. Tsuchihashi Z, Dracopoli NC (2002). Progress in high-throughput SNP genotyping methods. Pharmacogenomics J. 2(2):103-110.
28. Uta Pich, Ingo Schubert (1993). Midiprep method for isolation of DNA from plants with a high content of polyphenolics. Nucl Acids Res 21(14):3328.