

A Transitory Review of Molecular Techniques to Assess Plant Diversity

Dr. D. Stripriya

Department of Biotechnology, School of Herbal Studies and Naturo Sciences, Dravidian University, Kuppam – 517 426, A.P., India

Abstract—Immense loss of valuable plant species in the past centuries and its adverse impact on environmental and socioeconomic values has triggered the conservation of plant resources. Appropriate identification and characterization of plant materials is essential for the successful conservation of plant resources and to ensure their sustainable use. Molecular tools developed in the past few years provide easy, less laborious means for assigning known and unknown plant taxa. These techniques answer many new evolutionary and taxonomic questions, which were not previously possible with only phenotypic methods. Molecular techniques such as DNA barcoding, random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites and single nucleotide polymorphisms (SNP) have recently been used for plant diversity studies. Each technique has its own advantages and limitations. These techniques differ in their resolving power to detect genetic differences, type of data they generate and their applicability to particular taxonomic levels. This review presents a basic description of different molecular techniques that can be utilized for DNA fingerprinting and molecular diversity analysis of plant species.

Index Terms—Molecular diversity, DNA fingerprinting, plant species, conservation genetics.

I. INTRODUCTION

The conservation and sustainable use of plant genetic resources require accurate identification of their accession. The emergence of DNA-based markers has changed the practice of species identification techniques. The dramatic advances in molecular genetics over the last few years have provided workers involved in the conservation of plant genetic resources with a range of new techniques for easy and reliable identification of plant species. Many of these techniques have been successfully used to study the

extent and distribution of variation in species gene-pools and to answer typical evolutionary and taxonomic questions. Properties desirable for ideal DNA markers include highly polymorphic nature, codominant inheritance (determination of homozygous and heterozygous states of diploid organisms), frequent occurrence in the genome, selective neutral behavior (the DNA sequences of any organism are neutral to environmental conditions or management practices), easy access (availability), easy and fast assay, high reproducibility, and easy exchange of data between laboratories.

Sequencing based molecular techniques provide better resolution at intra-genus and above level, while frequency data from markers such as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellites provide the means to classify individuals into nominal genotypic categories and are mostly suitable for intra-species genotypic variation study. This distinction is important to grasp for population studies, particularly when the diversity data are used as a basis for making decisions about conservation of plant resources. For instance, a recent study on Napier grass or African grass (*Pennisetum purpureum*) has showed that AFLP is incompatible with RAPD and morphological data; re-registration of all accessions of Napier grass based on DNA barcoding is suggested as a means to resolve the lingering problems regarding the identity of accessions. The main objective of this review is to provide a basic understanding of the recently developed molecular tools and their potential application in the conservation of plant resources.

II. DNA SEQUENCING

DNA sequencing is the determination of the order of the nucleotide bases-A (adenine), G (guanine), C (cytosine) and T (thymine) present in a target molecule of DNA. Early work that was developed for the identification and characterization of clinically important bacterial strains has made it possible to obtain DNA sequences within a few days. We describe the conventional and next generation sequencing techniques separately under the following subheadings.

III. CONVENTIONAL SEQUENCING TECHNIQUE

Presently, dye-terminator sequencing technique is the standard method in automated sequencing analysis. The dye-terminator sequencing method, along with automated high-throughput DNA sequence analyzers, is now being used for the vast majority of sequencing work. The basic technique related with dye terminator sequencing and phylogenetic analysis is illustrated in Figure - 1. Dye-terminator sequencing utilizes labeling of the chain terminator ddNTPs, which allows sequencing in a single reaction, rather than four reactions as in the previously used labeled-primer method. In dye-terminator sequencing, the four dideoxynucleotide chain terminators are labeled with fluorescent dyes, each with a different wavelength of fluorescence emission. The main advantages of this technique are its robustness, automation and high accuracy (>98%). On the other hand, the limitations of this technique include dye effects due to differences in the incorporation of the dye-labeled chain terminators into the DNA fragment. Such incorporation of dye can result in unequal peak heights and shapes in the electronic DNA sequence trace chromatogram after capillary electrophoresis. Another drawback is its inability to handle long sequences; however, it can reliably sequence up to approximately 900 nucleotide long DNA fragments in a single reaction. The advent of new generation sequencers with solid state chemistry has significantly overcome these problems.

Current interest is in the DNA barcoding of plants with the aim to identify an unknown plant in terms of a known classification. DNA barcoding is a technique for characterizing species of organisms using a short

DNA sequence from a standard and agreed-upon position in the genome. DNA barcode sequences are very short relative to the entire genome and they can be obtained reasonably quickly and cheaply. The success of species-level assignment of plants using Basic Local Alignment Search Tool (BLAST) with individual barcodes was obtained with *matK* (99%), followed by *trnH-psbA* (95%) and then *rbcL* (75%). Use of three-locus DNA barcode resulted in >98% correct identifications of 296 species of woody trees, shrubs and palms. Recently, a group of plant DNA barcode researchers proposed two chloroplast genes, *rbcL* and *matK*, taken together, as appropriate for barcoding of plants.

Molecular phylogenies in plants are traditionally based on chloroplast DNA (cpDNA) sequence variation. This approach has proved to be very powerful at the family level through the sequencing of coding regions such as *rbcL*. However, low evolutionary rate of these sequences limits the power of cpDNA for the assignment at the genus or species level. As a consequence, the relationships among closely related taxa have been inferred using non-coding sequences. However, the potential problems due to gene flow of cpDNA among closely related taxa, as well as the lack of phylogenetic resolution, triggered the development of new approaches based on nuclear DNA. The most common alternative corresponds to the sequencing of the ITS (internal transcribed spacer) of 18S–25S nuclear ribosomal DNA. When both cpDNA and ITS sequencing fail to resolve phylogenies, the amplified fragment length polymorphism (AFLP) approach has the potential to solve such difficulties, particularly among closely related species, or at the intra-specific level. Therefore, integration of recently developed barcoding with the following techniques such as RAPD, AFLP, microsatellite and SNP seems to provide better resolution.

IV. NEXT GENERATION SEQUENCING TECHNIQUES

A new generation of non-Sanger based sequencing technologies has been evolving on its promise of sequencing DNA at unprecedented speed, thereby also having enabled impressive scientific achievements and novel biological applications. These techniques have made it possible to conduct robust population-

genetic studies based on complete genomes rather than just short sequences of a single gene. Rapid progress in genome sequences of various plant species through next generation sequencing will further extend our understanding how genotypic variation translates into phenotypic characteristics. A comparative genomic approach is extraordinarily useful for identifying functional loci related to morphological, geographical and physiological variation, and thus next generation sequencing technology will enable us to better understand the process of plant evolution. Next generation platforms do not rely on Sanger chemistry as did the first-generation machines used for the last 30 years. The first of this kind of 2nd generation of sequencing technique appeared in 2005 with the landmark publication of the sequencing-by-synthesis technology developed by 454 Life Sciences based on pyrosequencing. Commercial 2nd generation sequencing methods can be distinguished by the role of PCR in library preparation. There are four main platforms; all being amplification-based:

- (i) Roche 454 GS FLX,
- (ii) Illumina Genome Analyzer IIX,
- (iii) ABI SOLiD 3 Plus System and
- (iv) Polonator G.007. Common principles of these 2nd generation sequencing techniques are illustrated in Figure 2.

The single-molecule sequencing method (also known as 3rd generation or next-next generation) is independent of PCR. This mode of sequencing protocol was recently developed by Helicos Genetic Analysis System using the technology developed by Braslavsky et al. Other 3rd generation sequencing systems are being developed by Life Technologies and Pacific Biosciences SMRT technology and may appear within one to two years. Oxford Nanopore Technology (www.nanoporetech.com) has been developing a label-free, electrical, single-molecule genuinely revolutionary DNA sequencing method. This technique is aimed at obviating the need for amplification or labeling by instead detecting a direct electrical signal. However, this technique is still in a developing stage. The recently developed Helicos 3rd generation high-throughput and low-cost direct single molecule RNA sequencing method - without requiring prior conversion of RNA to cDNA - opened the door for a comprehensive and bias-free understanding of transcriptomes.

Through directly sequencing single molecules of DNA or RNA, Helicos True Single Molecule Sequencing (tSMS) technology significantly increased the speed of sequencing, while also decreasing the cost. Briefly, the procedure works by: first capturing billions of single molecules of sample DNA on an application-specific proprietary surface within two flow cells. These captured strands serve as templates for the sequencing-by-synthesis. Polymerase and one fluorescently labeled nucleotide (C/G/A/T) are added. The polymerase catalyzes the sequence-specific incorporation of fluorescent nucleotides into nascent complementary strands on all the templates. After a wash step, which removes all free nucleotides, the incorporated nucleotides are imaged and their positions are recorded. The fluorescent group is removed in a highly efficient cleavage process, leaving behind the incorporated nucleotide. The process continues through each of the other three bases (Figure - 3 modified). Using the Helicos DNA Barcoding protocol, scientists at Helicos were able to multiply the system's sample throughput five-fold (from 50 samples to 250 samples per run), without compromising accuracy or representational bias. DNA sequencing data from next generation platforms typically present shorter read lengths, higher coverage and different error profiles compared with Sanger sequencing data. Several software packages have been created specially to cope with the next generation sequencing data. A good review on these recent software tools has been published by Miller et al.

Since the advent of next generations sequencing, these techniques have been helping to uncover secondary metabolic pathways, to analyze cDNA-array based gene expression, for genetic manipulation to improve yield of desirable secondary products and molecular marker identification. For example, an Expressed Sequence Tag (EST) library from whole plantlets of medicinal plant (*Salvia miltiorrhiza*) was generated with the expression patterns of 14 secondary metabolic enzyme genes in different organs. Additionally, a total of 122 microsatellites were identified from the ESTs, with 89 having sufficient flanking sequences for primer design. This set of ESTs represents a significant proportion of the *S. miltiorrhiza* transcriptome and gives preliminary insights into the gene complement of *S. miltiorrhiza*, which was a very laborious task a few years back. Using 454 and

Illumina EST sequencing of the parental diploid species of *Tragopogon miscellus* (Moscow salsify, Asteraceae), 7,782 single nucleotide polymorphisms were identified that differ between the two progenitor's genomes present in this all tetraploid. Next generation high through-put Solexa sequencing technology led to the discovery of 14 novel and 22 conserved miRNA families from peanut. Recently, a new variety of chickpea (*Cicer arietinum*), resistant to *Helicoverpa armigera* (Pod borer), has been developed with the help of valuable information retrieved from next generation sequencing.

V. RANDOM AMPLIFIED POLYMORPHIC DNA (RAPD)

RAPD is based on the amplification of genomic DNA with single primers of arbitrary nucleotide sequence. These primers detect polymorphisms in the absence of specific nucleotide sequence information and the polymorphisms function as genetic markers and can be used to construct genetic maps. Since most of the RAPD markers are dominant, it is not possible to distinguish whether the amplified DNA segment is heterozygous (two different copies) or homozygous (two identical copies) at a particular locus. In rare cases, co-dominant RAPD markers, observed as different-sized DNA segments amplified from the same locus, may be detected.

The basic technique of RAPD involves

- (i) extraction of highly pure DNA,
- (ii) addition of single arbitrary primer,
- (iii) polymerase chain reaction (PCR),
- (iv) separation of fragments by gel electrophoresis,
- (v) visualization of RAPD-PCR fragments after ethidium bromide staining under UV light and
- (vi) determination of fragment size comparing with known molecular marker with the help of gel analysis software.

A diagrammatic presentation of these steps is given in Figure - 4. It is important to note that RAPD technique requires maintaining strictly consistent reaction conditions in order to achieve reproducible profiles. In practice, band profiles can be difficult to reproduce between (and even within) laboratories, if personnel, equipment or conditions are changed. Despite these limitations, the enormous attraction of this technique

is that there is no requirement for DNA probes or sequence information for primer design. The procedure involves no blotting or hybridizing steps. The technique is quick, simple and efficient and requires only the purchase of a thermos cycling machine and agarose gel apparatus and relevant chemicals, which are available as commercial kits (e.g., Ready-To-Go RAPD analysis beads; GE Healthcare, Buckinghamshire, UK). Another advantage is the requirement for only small amounts of DNA (10–100 ng per reaction).

The RAPD markers have been used for detecting genomic variations within and between varieties of sweet potato. A total of 160 primers were tested and eight showed consistent amplified band patterns among the plants with variations within and between varieties of sweet potato. Genetic diversity was evaluated by RAPD markers and morpho-agronomic characters for a total of 42 accessions of Barberton daisy (*Gerbera jamesonii*) employing a set of 12 primer pairs. Germplasm accession of 80 *Plantago* spp. was studied by using RAPD with the help of 20 random primers. Recently, RAPD has been used for estimation of genetic diversity in various endangered plant species.

VI. AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP)

The AFLP technique is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. The technique involves:

- (i) extraction of highly purified DNA,
- (ii) restriction endonuclease digestion of DNA (enzyme mixture, usually EcoRI + MseI),
- (iii) ligation of adapters (enzyme adapters),
- (iv) pre-PCR (amplification of the restriction fragments; pre-selective amplification with EcoRI primer + A and MseI primer + C),
- (v) selective-PCR with labeled primer pair (Primer + 3 base pairs; forward labeled, reverse unlabeled), and
- (vi) gel electrophoresis and fragment analysis by automated sequencing machine (Figure - 5). The electrophoretograms can be analyzed using programs like Gene Mapper.

AFLP is applicable to all species, and unlike RAPD, this technique is highly reproducible as it combines restriction digestion and PCR. However, AFLP requires more DNA (300–1000 ng per reaction) and is more technically demanding than RAPD, however the automation and recent availability of kits means that this technology can be brought to a higher level. Nuclear and chloroplast sequences sometimes fail to reveal variability when plant species are closely related. However, AFLP distributed throughout the whole genome provides a robust solution to overcome the hurdles in plants fingerprinting.

A good review on AFLP markers in surveys of plant diversity is published by Mba and Tohme. Recently, several plants in germplasm collections such as *Jatropha curcas* and *Rhodiola rosea* have been characterized by AFLP. Teyer et al. have studied the wild populations of *Agave angustifolia* in the desert using AFLP to measure the genetic variability within and between natural populations. AFLP markers have been extensively used for phylogenetic analysis and determining the genetic diversity for conservation of endangered plant species.

VII. MICROSATELLITES

Microsatellites, or simple sequence repeats (SSRs), are polymorphic loci present in DNA that consist of repeating units of one to six base pairs in length. One common example of a microsatellite is a (CA)_n repeat, where n is variable among different alleles. These markers often present high levels of inter- and intra-specific polymorphism, particularly when the tandem repeats number is 10 or greater. The repeated sequence is often simple, consisting of two, three or four nucleotides (di-, tri- and tetra- nucleotide repeats) and can be repeated many times. The basic principle of microsatellite is illustrated in Figure - 6.

Microsatellites can be amplified for identification by PCR using the unique sequences of flanking regions as primers. The most common way to detect microsatellites is to design PCR primers that are unique to one locus in the genome and that base pair on either side of the repeated portion. Therefore, a single pair of PCR primers will work for every individual in the species and produce different sized products for each of the different length of microsatellites. The PCR products are separated either

by slab gel electrophoresis or capillary gel electrophoresis in an automated sequencer.

Microsatellites have proved to be versatile molecular markers, particularly for population analysis, but they are not without limitations. With the abundance of PCR technology, primers that flank microsatellite loci are simple and quick to use, but the development of correctly functioning primers is often a tedious and costly process. Microsatellites developed for particular species can often be applied to closely related species, but the percentage of loci that successfully amplify may decrease with increasing genetic distance. Microsatellite technique has recently been used to establish conservation strategy of endangered plants like *Calystegia soldanella*, *Tricyrtis ishiiana* and *Galium catalinense* subspecies *acrispum*.

VIII. SINGLE NUCLEOTIDE POLYMORPHISM (SNP)

Single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide (A, T, G or C) differs among members of a species. SNP is the most abundant marker system both in animal and plant genomes and has recently emerged as the new generation molecular markers for various applications. Being binary or co-dominant status, they are able to efficiently discriminate between homozygous and heterozygous alleles. Moreover, unlike microsatellites their power comes not from the number of alleles but from the large number of loci that can be assessed. Once the rare SNPs are discovered in a low diversity species, the genetic population discrimination power can be equivalent to the same number of loci in a genetically diverse species. The more evolutionary conserved nature of SNPs makes them less subject to the problem of homoplasy. Most importantly, SNPs are amenable to high throughput automation, allowing rapid and efficient genotyping of large numbers of samples.

In plants, SNP can be designed from ESTs and single-stranded pyrosequencing. A high-throughput genome analysis method called diversity array technology (DArT), based on microarray platform, has been developed for the analysis of plant DNA polymorphism. Eijk et al. described a novel SNP genotyping technique, SNPWave. Chip-based SNP

arrays use thousands of oligonucleotide probes attached to a solid surface (e.g., glass, silicon wafer) allowing for a large number of SNPs to be interrogated simultaneously. The ABI PRISM SNaPshot Multiplex Kit is designed to interrogate up to 10 single nucleotide polymorphisms (SNPs) at known locations on one to 10 DNA templates in a single tube. The basic principle of SNP and detection method is illustrated in Figure - 7. In brief, the protocol includes preparation of sample reactions using template and primer, performing SNaPshot reactions by thermal cycling and conduction of post-extension treatment of the products. Then automated electrophoresis of the samples and finally, analyzing the data.

SNP is able to determine genetic diversity in plants, particularly in species with limited genetic diversity. Determination of population genetic structure of Castor oil plant (*Ricinus communis*) using SNPs from genome-wide comparisons showed low levels of genetic diversity and mixing of genotypes, leading to minimal geographic structuring of castor bean populations worldwide. Clustering of Castor bean indicated five main groups worldwide and a repeated pattern of mixed genotypes in most countries; most molecular variance occurred within-populations (74%) followed by 22% among-populations and 4% among-continent.

Recently, a single nucleotide primer extension (SNUPE) assay targeting *gyrB* gene has been developed to identify bacteria belonging to the *Burkholderia cepacia* complex, which are very difficult to identify using commonly used phenotypic and molecular techniques. Novel SNP based technique allowed the successful detection and distinction of specific genetic variations and is effectively applied in routine medical diagnosis since it permits to analyze routinely many samples in a short time. Similar approaches need to be utilized in plants that are difficult to discriminate for low level of genetic diversity.

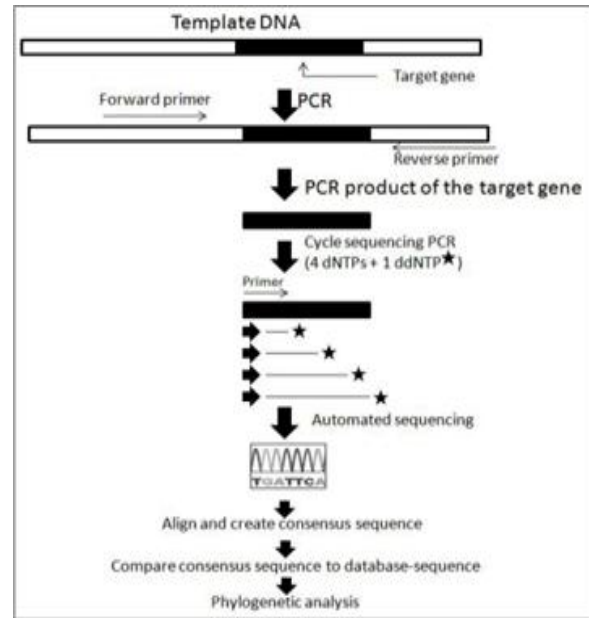


FIGURE – 1

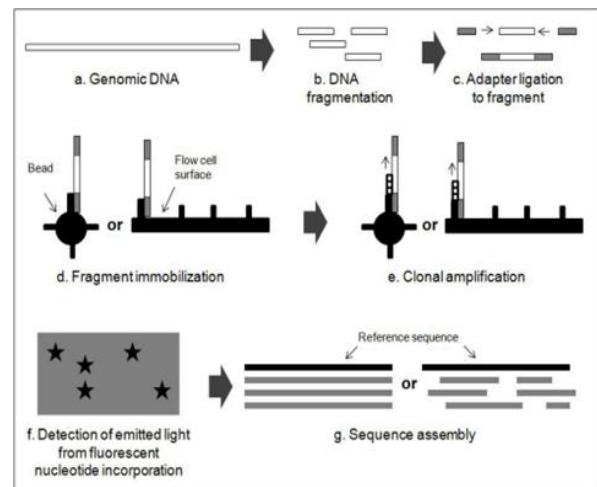


FIGURE – 2

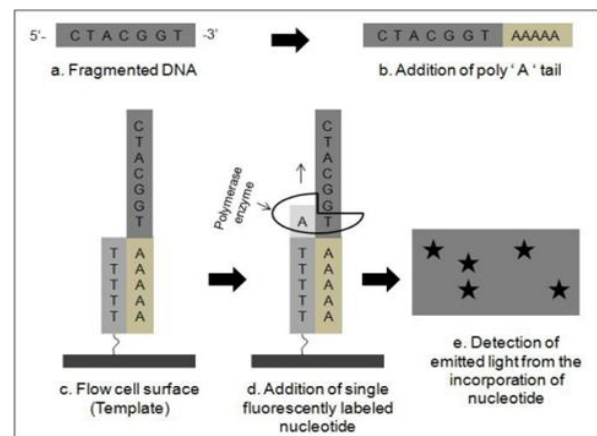


FIGURE – 3

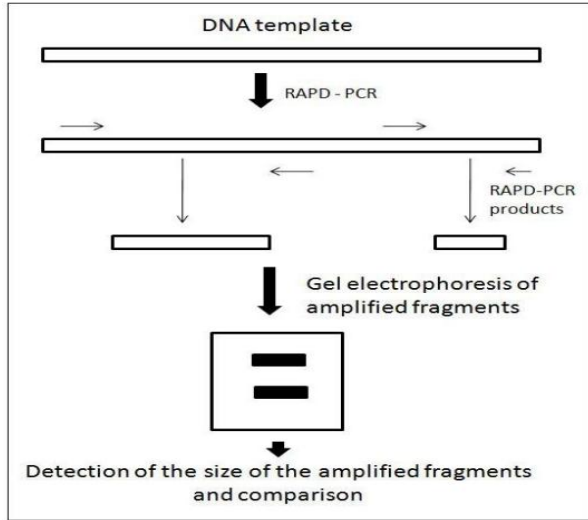


FIGURE – 4

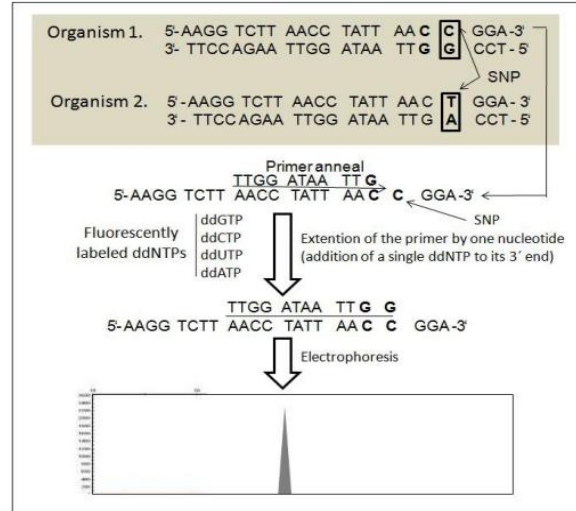


FIGURE – 7

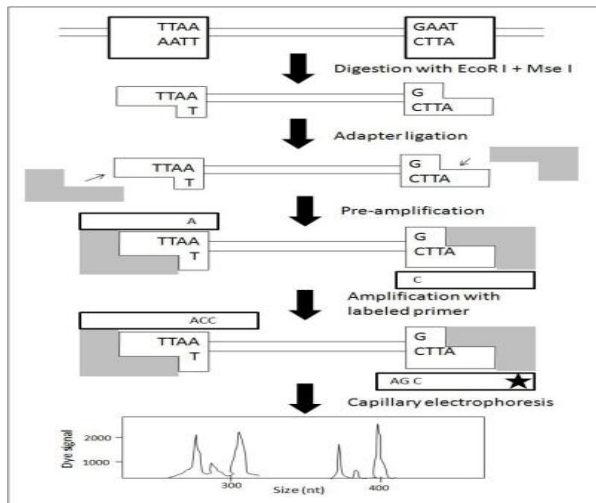


FIGURE – 5

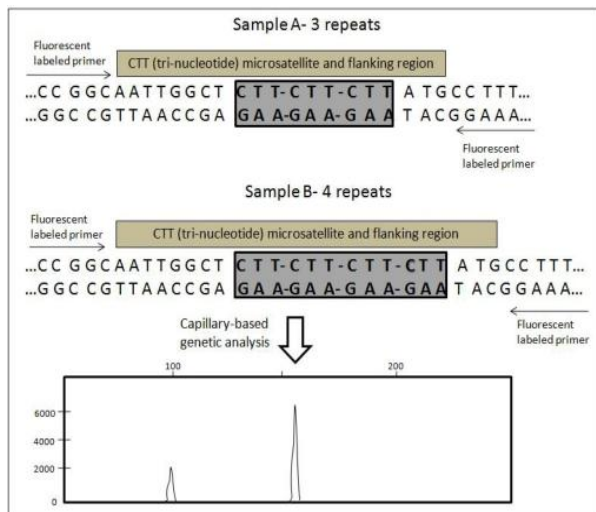


FIGURE – 6

IX. CONCLUDING OBSERVATIONS

Molecular markers are indispensable tools for measuring the diversity of plant species. Low assay cost, affordable hardware, throughput, convenience and ease of assay development and automation are important factors when choosing a technology. Databases based on a large number of potential characters are readily available for inferring relationships using sequence data. Further advantage of sequencing includes substitutions within structural genes that produce differentiation from changes in morphology. Information from the sequences themselves can be useful for specifying parameters of the model of sequence evolution, which in turn, influences the topology of the inferred tree. To date, next generation sequencing technologies have been applied in a variety of contexts, including whole-genome sequencing, targeted resequencing, discovery of transcription factor binding sites and noncoding RNA expression profiling. A disadvantage of sequencing includes inferences of positional homology that are frequently more problematic for non-coding nucleotide sequences because penalties for insertion-deletion events determine the extent of sequence similarity during pairwise and multiple-alignment. Another potential shortcoming of sequencing is that the evolutionary history of species can be inconsistent with the genealogy of a single gene and it is not possible to assess if a tree topology based on a single gene sequence is likely to represent the original genealogy of the species. The advantages of

RAPD include its simplicity, low cost, rapid, use of arbitrary primers, no need of initial genetic or genomic information, and the requirement of only tiny quantities of target DNA. Disadvantages of this technique are dominant type and the lack of a prior knowledge on the identity of the amplification products which in turn creates problems with reproducibility and co-migration. The major advantage of the AFLP technique is the large number of polymorphisms that the method generates compared with other markers. The ability of AFLP to differentiate individuals in a population makes the technique useful for paternity analyses, gene-flow experiments and also for plant variety registration. However, the methodology of AFLP experiment and post-run data analysis are complex and time consuming compared with other markers like RAPD. The great advantage of microsatellite analysis is the large number of polymorphisms that the method reveals. The ability of the method to differentiate individuals when a combination of loci is examined makes the technique very useful for gene-flow experiments, cultivar identification and paternity analyses. Major problem with the microsatellite relates with the initial screening of an organism for microsatellite library creation. The distribution and frequencies of SNPs are the key factors to understand molecular diversity between closely related populations and species. SNP is potentially useful for the analysis of degraded samples by use of short amplicons and low mutation rates. However, this technique has bi-allelic nature and less resolution power compared with multi-allelic microsatellites; though this shortcoming is overcome by its inherent capacity of scanning a large number of loci. SNP markers are best for characterizing and conserving the gene bank materials and the AFLP and microsatellite markers are more suitable for diversity analysis and fingerprinting.

Unfortunately, an ideal marker does not exist for use in all studies; rather a technique or techniques will be suited to a range of investigations. In this context, we agree with Robinson and Harris and conclude that RAPD, AFLP and microsatellites should not be considered appropriate for phylogenetic analyses above the species level. These markers are undoubtedly valuable tools for addressing population genetics and plant breeding issues, but for phylogeny

reconstruction and taxonomy they could be problematic and sometimes even misleading, so they must be used with caution. Molecular genetics is a fast-moving field and new techniques are likely to be developed in the near future which will have their own strengths and limitations. Development of molecular technique based on error free database is another essential demand for easy assignment of unknown plant samples into appropriate taxa.

REFERENCES

- [1] Botstein D, White RL, Skolnick MH, Davis RW. Construction of a genetic map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 1980; 32:314–331. [PMC free article] [PubMed] [Google Scholar]
- [2] Karp A, Seberg O, Buiatti M. Molecular Techniques in the Assessment of Botanical Diversity. *Ann. Bot.* 1996; 78:143–149. [Google Scholar]
- [3] Karp A, Kresovich S, Bhat KV, Ayad WG, Hodgkin T. *Molecular Tools in Plant Genetic Resources Conservation: A Guide to the Technologies* IPGRI Technical Bulletin No. 2; International Plant Genetic Resources Institute; Rome, Italy: 1997 [Google Scholar]
- [4] Joshi SP, Ranjanekar PK, Gupta VS. Molecular markers in plant genome analysis. *Curr. Sci.* 1999; 77:230–240. [Google Scholar]
- [5] Robinson JP, Harris SA. In: Which DNA Marker for Which Purpose. Gillet EM, editor. *Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen; Göttingen, Germany: 1999.* pp. 1–27. [Google Scholar]
- [6] Struwig M, Mienie CMS, van den Berg J, Mucina L, Buys MH. AFLPs are incompatible with RAPD and morphological data in *Pennisetum purpureum* (Napier grass) *Biochem. Sys. Ecol.* 2009; 37:645–652. [Google Scholar]
- [7] Hultman T, Stahl S, Hornes E, Uhlen M. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucl. Acids Res.* 1989; 17:4937–4946. [PMC free article] [PubMed] [Google Scholar]
- [8] Brytting M, Wahlberg J, Lundeberg J, Wahren B, Uhlen M, Sundqvist VA. Variations in the cytomegalovirus major immediate-early gene found by direct genomic sequencing. *J. Clin.*

- Microbiol. 1992; 30:955–960. [PMC free article] [PubMed] [Google Scholar]
- [9] Olsvik O, Wahlberg J, Petterson B, Uhlén M, Popovic T, Wachsmuth IK, Fields PI. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J. Clin. Microbiol.* 1993; 31:22–25. [PMC free article] [PubMed] [Google Scholar]
- [10] Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science.* 1987; 238:336–341. [PubMed] [Google Scholar]
- [11] Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci.* 2005; 102:8369–8374. [PMC free article] [PubMed] [Google Scholar]
- [12] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol.* 1990; 215:403–410. [PubMed] [Google Scholar]
- [13] Kress WJ, Erickson DL, Jones FA, Swenson NG, Perez R, Sanjur O, Bermingham E. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. USA.* 2009; 106:18621–18626. [PMC free article] [PubMed] [Google Scholar]
- [14] CBOL Plant Working Group A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA.* 2009; 106:12794–12797. [PMC free article] [PubMed] [Google Scholar]
- [15] Despres L, Gielly L, Redoutet B, Taberlet P. Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Mol. Phylogenet. Evol.* 2003; 27:185–196. [PubMed] [Google Scholar]
- [16] Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL, et al. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* 1993; 80:528–580. [Google Scholar]
- [17] Soltis DE, Morgan DR, Grable A, Soltis PS, Kuzoff R. Molecular systematics of Saxifragaceae sensu stricto. *Am. J. Bot.* 1993; 80:1056–1081. [Google Scholar]
- [18] Gielly L, Taberlet P. A phylogeny of the European gentians inferred from chloroplast *trnL* UAA intron sequences. *Bot. J. Linn. Soc.* 1996; 120:57–75. [Google Scholar]
- [19] Baldwin BG. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the Compositae. *Mol. Phylogenet. Evol.* 1992; 1:3–16. [PubMed] [Google Scholar]
- [20] Yuan YM, Küpfer P, Doyle J. Infrageneric phylogeny of the genus *Gentiana* (Gentianaceae) inferred from nucleotide sequences of the internal transcribed spacers ITSs of nuclear ribosomal DNA. *Am. J. Bot.* 1996; 83:641–652. [Google Scholar]
- [21] Koopman WJM, Zevenbergen MJ, Van-den-Berg RG. Species relationships in *Lactuca* S.L. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. *Am. J. Bot.* 2001; 88:1881–1887. [PubMed] [Google Scholar]
- [22] Sun M. Comparative analysis of phylogenetic relationships of grain amaranths and their wild relatives (*Amaranthus*; amaranthaceae) using internal transcribed spacer, amplified length polymorphism, and double-primer fluorescent intersimple sequence repeat markers. *Mol. Phylogenet. Evol.* 2001; 21:372–387. [PubMed] [Google Scholar]
- [23] Zhang LB, Comes HP, Kadereit JW. Phylogeny and quaternary history of the European montane/alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. *Am. J. Bot.* 2001; 88:2331–2345. [PubMed] [Google Scholar]
- [24] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 1977; 74:5463–5467. [PMC free article] [PubMed] [Google Scholar]
- [25] Schuster SC. Next-generation sequencing transforms today's biology. *Nat. Methods.* 2008; 5:16–18. [PubMed] [Google Scholar]
- [26] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. *Nature.* 2005; 437:376–380. [PMC free article] [PubMed] [Google Scholar]

- [27] Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 2006; 242:84–89. [PubMed] [Google Scholar]
- [28] Nyrén P. The history of pyrosequencing. *Methods Mol. Biol.* 2007; 373:1–14. [PubMed] [Google Scholar]
- [29] Lerner HRL, Fleischer RC. Prospects for the use of next-generation sequencing methods in Ornithology. *The Auk.* 2010; 127:4–15. [Google Scholar]
- [30] Blow N. DNA sequencing: Generation next-next. *Nat. Methods.* 2008; 5:267–274. [Google Scholar]
- [31] Braslavsky I, Hebert B, Kartalov E, Quake SR. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA.* 2003; 100:3960–3964. [PMC free article] [PubMed] [Google Scholar]
- [32] Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotechnol.* 2009; 4:265–270. [PubMed] [Google Scholar]
- [33] Ozsolak F, Platt AR, Jones DR, Reifengerger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM. Direct RNA sequencing. *Nature.* 2009; 461:814–818. [PubMed] [Google Scholar]
- [34] Miller JR, Koren S, Sutton G. Assembly algorithms for next generation sequencing data *Genomics* 2010. in press, doi: 10.1016/j.ygeno.2010.03.001. [PMC free article] [PubMed] [Google Scholar]
- [35] Yan Y, Wang Z, Tian W, Dong Z, Spencer D. Generation and analysis of expressed sequence tags from the medicinal plant *Salvia miltiorrhiza*. *Sci. Chin. Life Sci.* 2010; 53:273–285. [PubMed] [Google Scholar]
- [36] Buggs RJA, Chamala S, Wu W, Gao L, May GD, Schnable PS, Soltis DE, Soltis PS, Barbazuk WB. Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX Mass ARRAY genotyping. *Mol. Ecol.* 2010; 19:132–146. [PubMed] [Google Scholar]
- [37] Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, Li AQ, Li MJ, Li CS, Zhang BH, Wang XJ. Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.) *BMC Plant Biol.* 2010; 10:1–12. [PMC free article] [PubMed] [Google Scholar]
- [38] Acharjee S, Sarmah BK, Kumar PA, Olsen K, Mahon R, Moar WJ, Moore A, Higgins TJV. Transgenic chickpeas (*Cicer arietinum* L.) expressing a sequence-modified cry2Aa gene. *Plant Sci.* 2010; 178:333–339. [Google Scholar]
- [39] Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 1990; 18:6531–6535. [PMC free article] [PubMed] [Google Scholar]
- [40] Charney J, Vershon D, Bill Sofer B. RAPD PCR, Genes, Genomes and Human Genetics; Department of Genetics, Department of Molecular Biology and Biochemistry, Division of Biological Sciences and the Waksman Institute: New York, NY, USA, 2007. Available at <http://avery.rutgers.edu/WSSP/StudentScholars/project/archives/onions/rapd.html> (accessed on 29 April 2010).
- [41] Lin KH, Lai YC, Li HC, Lo SF, Chen LFO, Lo HF. Genetic variation and its relationship to root weight in the sweet potato as revealed by RAPD analysis. *Sci. Hort.* 2009; 120:2–7. [Google Scholar]
- [42] da-Mata TL, Segeren MI, Fonseca AS, Colombo CA. Genetic divergence among gerbera accessions evaluated by RAPD. *Sci. Hort.* 2009; 121:92–96. [Google Scholar]
- [43] Singh N, Lal RK, Shasany AK. Phenotypic and RAPD diversity among 80 germplasm accessions of the medicinal plant isabgol (*Plantago ovata*), Plantaginaceae. *Genet. Mol. Res.* 2009; 8:1273–1284. [PubMed] [Google Scholar]
- [44] Zheng W, Wang L, Meng L, Liu J. Genetic variation in the endangered *Anisodus tanguticus* (Solanaceae), an alpine perennial endemic to the Qinghai-Tibetan Plateau. *Genetica.* 2008; 132:123–129. [PubMed] [Google Scholar]
- [45] Liu P, Yang YS, Hao CY, Guo WD. Ecological risk assessment using RAPD and distribution pattern of a rare and endangered species. *Chemosphere.* 2007; 68:1497–1505. [PubMed] [Google Scholar]

- [46] Wang ZS, An SQ, Liu H, Leng X, Zheng JW, Liu YH. Genetic structure of the endangered plant *Neolitsea sericea* (Lauraceae) from the Zhoushan archipelago using RAPD markers. *Ann. Bot.* 2005; 95:305–313. [PMC free article] [PubMed] [Google Scholar]
- [47] Lu HP, Cai YW, Chen XY, Zhang X, Gu YJ, Zhang GF. High RAPD but no cpDNA sequence variation in the endemic and endangered plant, *Heptacodium miconioides* Rehd. (Caprifoliaceae) *Genetica.* 2006; 128:409–417. [PubMed] [Google Scholar]
- [48] Vos P, Hogers R, Bleeker M, Reijmans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. AFLP: A new technique for DNA fingerprinting. *Nucl. Acids Res.* 1995; 23:4407–4414. [PMC free article] [PubMed] [Google Scholar]
- [49] Romero G, Adeva C, Battad Z. Genetic fingerprinting: Advancing the frontiers of crop biology research. *Philipp. Sci. Lett.* 2009; 2:8–13. [Google Scholar]
- [50] Amplified Fragment Length Polymorphism (AFLP) Analysis on Applied Biosystems Capillary Electrophoresis Systems, Application Note AFLP on the 3130/3730; Applied Biosystems: Foster City, CA, USA, 2005
- [51] Mba C, Tohme J. Use of AFLP markers in surveys of plant diversity. *Methods Enzymol.* 2005; 395:177–201. [PubMed] [Google Scholar]
- [52] Tatikonda L, Wani SP, Kannan S, Beerelli N, Sreedevi TK, Hoisington DA, Devi P, Varshney RK. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L.: A biofuel plant. *Plant Sci.* 2009; 176:505–513. [PubMed] [Google Scholar]
- [53] Elameen A, Klemsdal SS, Dragland S, Fjellheim S, Rognli OA. Genetic diversity in a germplasm collection of roseroots (*Rhodiola rosea*) in Norway studied by AFLP. *Biochem. Syst. Ecol.* 2008; 36:706–715. [Google Scholar]
- [54] Teyer FS, Salazar MS, Esqueda M, Barraza A, Robert ML. Genetic variability of wild *Agave angustifolia* populations based on AFLP: A basic study for conservation. *J. Arid Environ.* 2009; 73:611–616. [Google Scholar]
- [55] van Ee BW, Jelinski N, Berry PE, Hipp AL. Phylogeny and biogeography of *Croton alabamensis* (Euphorbiaceae), a rare shrub from Texas and Alabama, using DNA sequence and AFLP data. *Mol. Ecol.* 2006; 15:2735–2751. [PubMed] [Google Scholar]
- [56] Ronikier M. The use of AFLP markers in conservation genetics—a case study on *Pulsatilla vernalis* in the Polish lowlands. *Cell. Mol. Biol. Lett.* 2002; 7:677–684. [PubMed] [Google Scholar]
- [57] Li X, Ding X, Chu B, Zhou Q, Ding G, Gu S. Genetic diversity analysis and conservation of the endangered Chinese endemic herb *Dendrobium officinale* Kimura et Migo (Orchidaceae) based on AFLP. *Genetica.* 2008; 133:159–166. [PubMed] [Google Scholar]
- [58] Zawko G, Krauss SL, Dixon KW, Sivasithamparam K. Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae) *Mol. Ecol.* 2001;10:2389–2396. [PubMed] [Google Scholar]
- [59] Bidichandani S, Ashizawa T, Patel PI. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. *Am. J. Hum. Genet.* 1998; 62:111–121. [PMC free article] [PubMed] [Google Scholar]
- [60] Queller DC, Strassman JE, Hughes CR. Microsatellites and Kinship. *Trends Ecol. Evol.* 1993; 8:285–288. [PubMed] [Google Scholar]
- [61] Allender C. Microsatellite Information Exchange for Brassica HRI; Wellesbourne, UK: 2004. Available at: <http://www.brassica.info/resource/markers/ssr-exchange.php> (accessed on 3 May 2010). [Google Scholar]
- [62] Jarne P, Lagoda P. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* 1996; 11:424–429. [PubMed] [Google Scholar]
- [63] Noda A, Nomura N, Mitsui Y, Setoguchi H. Isolation and characterisation of microsatellite loci in *Calystegia soldanella* (Convolvulaceae), an endangered coastal plant isolated in Lake Biwa, Japan. *Conserv. Genet.* 2009; 10:1077–1079. [Google Scholar]
- [64] Setoguchi H, Mitsui Y, Ikeda H, Nomura N, Tamura A. Development and characterization of microsatellite loci in the endangered *Tricyrtis ishiana* (Convallariaceae), a local endemic plant

- in Japan. *Conserv. Genet.* 2009; 10:705–707. [Google Scholar]
- [65] McLaughlin ME, Riley L, Helenurm K. Isolation of microsatellite loci from the endangered plant *Galium catalinense* subspecies *acrispum* (Rubiaceae) *Mol. Ecol. Resources.* 2009; 9:984–986. [PubMed] [Google Scholar]
- [66] Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P. Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*) *BMC Plant Biol.* 2010;10:13–23. [PMC free article] [PubMed] [Google Scholar]
- [67] Brumfield RT, Beerli P, Nickerson DA, Edwards SV. The utility of single nucleotide polymorphisms in inferences of population history. *Trends Ecol. Evol. (Amst.)* 2003;18:249–256. [Google Scholar]
- [68] Tsuchihashi Z, Dracopoli NC. Progress in high-throughput SNP genotyping methods. *Pharmacogenomics J.* 2002; 2:103–110. [PubMed] [Google Scholar]
- [69] Coles ND, Coleman CE, Christensen SA, Jellen EN, Stevens MR, Bonifacio RA, Rojas-Beltran JA, Fairbanks DJ, Maughan PJ. Development and use of an expressed sequenced tag library in quinoa (*Chenopodium quinoa* Willd.) for the discovery of single nucleotide polymorphisms. *Plant Soc.* 2005; 168:439–447. [Google Scholar]
- [70] Gómez MJM, Maloo JN. Sequence diversity in three tomato species: SNPs, markers and molecular evolution. *BMC Plant Biol.* 2009; 9:85. [PMC free article] [PubMed] [Google Scholar]
- [71] Miller PT, Henry R. Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. *Anal. Biochem.* 2003; 317:166–170. [PubMed] [Google Scholar]
- [72] Jaccoud D, Peng K, Feinstein D, Kilian A. Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucl. Acids Res.* 2001;29: e25. [PMC free article] [PubMed] [Google Scholar]
- [73] van-Eijk MJT, Broekhof JLN, van-der-Poel HJA, Hogers RCJ, Schneiders H, Verstege KJE, van-Aart JW, Geerlings H, Buntjer JB, van-Oeveren AJ, Vos P. SNPWave: A flexible multiplexed SNP genotyping technology. *Nucl. Acids Res.* 2004;32: e47. [PMC free article] [PubMed] [Google Scholar]
- [74] Rapley R, Harbron S. *Molecular Analysis and Genome Discovery.* John Wiley & Sons Ltd; Chichester, UK: 2001. [Google Scholar]
- [75] Ingelsson M, Shin Y, Irizarry MC, Hyman BT, Lilius L, Forsell C, Graff C. Genotyping of apolipoprotein E: Comparative evaluation of different protocols. *Curr Protoc Hum Genet.* 2003; 9:914:1–914:13. [PubMed] [Google Scholar]
- [76] Coenye T, Vandamme P, Govan JRW, LiPuma JL. Taxonomy and Identification of the *Burkholderia cepacia* Complex. *J. Clin. Microbiol.* 2001; 39:3427–3436. [PMC free article] [PubMed] [Google Scholar]
- [77] Ferri L, Perrin E, Campana S, Tabacchioni S, Taccetti G, Cocchi P, Ravenni N, Dalmastrì C, Chiarini L, Bevivino A, Manno G, Mentasti M, Fani R. Application of multiplex single nucleotide primer extension (mSNuPE) to the identification of bacteria: The *Burkholderia cepacia* complex case. *J. Microbiol. Methods.* 2010; 80:251–256. [PubMed] [Google Scholar]
- [78] Rafalski JA, Tingey SV. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. *Trends Genet.* 1993; 9:275–280. [PubMed] [Google Scholar]
- [79] Rafalski JA. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Soc.* 2002; 162:329–333. [Google Scholar]
- [80] Wilson AC. The molecular basis of evolution. *Sci. Am.* 1985; 253:164–173. [PubMed] [Google Scholar]
- [81] Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. *Genomics.* 2008; 92:255–264. [PubMed] [Google Scholar]
- [82] Nei M. *Molecular Evolutionary Genetics.* Columbia University Press; New York, NY, USA: 1987. [Google Scholar]
- [83] Nadler SA. Advantages and disadvantages of molecular phylogenetics: A case study of ascaridoid nematodes. *J. Nematol.* 1995; 27:423–432. [PMC free article] [PubMed] [Google Scholar]
- [84] Munthali M, Ford-Lloyd BV, Newbury HJ. The random amplification of polymorphic DNA for

- fingerprinting plants. *PCR Methods Appl.* 1992; 1:274–276. [PubMed] [Google Scholar]
- [85] Lowe AJ, Hanotte O, Guarino L. Standardization of molecular genetic techniques for the characterization of germplasm collections: The case of random amplified polymorphic DNA (RAPD) *Plant Genet. Resour. Newsl.* 1996; 107:50–54. [Google Scholar]
- [86] Krauss SL. Complete exclusion of nonsires in an analysis of paternity in a natural plant population using amplified fragment length polymorphism (AFLP) *Mol. Ecol.* 1999; 8:217–226. [Google Scholar]
- [87] Law JR, Donini P, Koebner RMD, Jones CR, Cooke RJ. DNA profiling and plant variety registration III: The statistical assessment of distinctness in wheat using amplified fragment length polymorphisms. *Euphytica.* 1998; 102:335–342. [Google Scholar]
- [88] Hokanson SC, Szewc-McFadden AK, Lamboy WF, McFerson JR. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection. *Theor. Appl. Genet.* 1998; 97:671–683. [Google Scholar]
- [89] Tautz D, Ellegren H, Weigel D. Next generation molecular ecology. *Mol. Ecol.* 2010; 19:1–3. [PubMed] [Google Scholar]
- [90] Sobrino B, Brión M, Carracedo A. SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Sci Int.* 2005;154, 181–194. [PubMed] [Google Scholar]
- [91] Engle LJ, Simpson CL, Landers JE. Using high-throughput SNP technologies to study cancer. *Oncogene.* 2006; 25:1594–1601. [PubMed] [Google Scholar]
- [92] Varshney RK, Chabane K, Hendre PS, Aggarwal RK, Graner A. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Soc.* 2007; 173:638–649. [Google Scholar]