REVIEW

MOLECULAR MARKERS FOR ANALYSIS OF PLANT GENETIC DIVERSITY

Vu Tuan Nam[1], Pham Le Bich Hang[1], Nguyen Nhat Linh[1], Luu Han Ly[1], Huynh Thi Thu Hue[1,2], Nguyen Hai Ha[1,2], Ha Hong Hanh[1], Le Thi Thu Hien[1,2,✉]

[1]*Institute of Genome Research, Vietnam Academy of Science and Technology*
[2]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology*

✉To whom correspondence should be addressed. E-mail: hienlethu@igr.ac.vn

SUMMARY

Genetic diversity plays an important role in diversity conservation at multiple levels and supports to monitor and assess genetic variation. In plants, genetic diversity provides the ability to adapt and respond to environmental conditions that helps plants to survive through changing environments. Genetic diversity analyses based on molecular genetic markers are effective tools for conservation and reintroduction of rare and endangered species. In recent years, the development of various chemical and molecular techniques for studying genetic diversity has received great attention. While biochemical markers are primarily used in the diagnosis of pathogens, DNA markers have been developed and widely applied for identification of species and population based on the genotype of an organism that is more stable and not easily affected by the environmental factors. PCR-based molecular marker tools, such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs) are used for analysing the difference in the targeted DNA sequences. With the rapid and robust development of genomic sequencing technology it is now possible to obtain and analyse DNA sequences of the whole genome of studied organisms. However, each type of DNA markers has different principles, as well as the pros and cons of specificity. In this article, we review methods and point out DNA markers, which are considered as reliable and widely used tools for the detection of genetic variation. In addition, we present the application of DNA marker in analysing genetic diversity of wild, domestic and medicinal plants, as well as some perspectives on the future of DNA marker's application in the analysis of genetic diversity.

Keyworks: DNA sequencing analysis, genetic diversity, molecular markers, SNPs, SSRs.

INTRODUCTION

Genetic diversity is a part of biological diversity, which is referred to as any variation of genetic materials (nucleotides, genes, chromosomes, or genomes) in all organisms in the ecosystems that they occur for a given time (Ramanatha, Hodgkin, 2002; Fu, 2015). Genetic diversity is a crucial characteristic of any population or species, which included the total number of genetic characteristics in the genetic makeup of a species, as genes coding for biological traits adapting to environmental conditions (Nevo, 2001). Genetic diversity plays an important role in ecology because of several reasons. Firstly, genetic diversity within a single species may correlate with species diversity both within and between trophic levels (Wimp *et al.*, 2004; Vellend, Geber, 2005; Crutsinger *et al.*, 2006; Johnson *et al.*, 2006; Vellend, 2006).

Secondly, in communities overwhelmed by a single species, hereditary variety inside that species may be as (or more) imperative as variety between species (Hughes, Stachowicz, 2004). One more reason, diversity may serve as a proxy for a complex multivariate phenotypic space that is often difficult to measure (Hughes *et al.*, 2008). Indeed, for a run of plant species, test controls of the number of genotypes have been appeared to influence population, community, and biological system properties (Bailey *et al.*, 2009).

Genetic diversity provides plants with the ability to adapt and response to environmental conditions, especially adverse factors as well as develop new and improved traits relating to yield and production, pest and disease resistance, and so on (Evenson, Gollin, 2003). This can play a very important role in providing adaptive genes that ultimately leads to long-term increase in food productivity related to environmental harm and also overcomes other types of phenomena such as genetic erosion (loss of genetic diversity), extinction of primitive, and adaptive genes (soil loss) caused by long-term exploitation of changes (Holderegger *et al.*, 2006). New strains can be found to be more stable in adaption over time. Each individual is genetically unique, unlike any individual in the same community or population. Genetic diversity also contributes in diversity conservation at multiple levels and supports to monitor and assess genetic variation.

Serious agents relating to environment, such as climate change, environmental pollution, as well as population growth, along with human activities are the dominant pressures for genetic change. Climate change is one of the most significant factors that immediately affects to biodiversity at all levels including genetic diversity and particularly in agriculture (Schmidhuber, Tubiello, 2007). Food production systems rely on highly selective varieties in a better environment but it may be increasingly susceptible to the effects of climate change, such as pests and disease spread (Gornall *et al.*, 2010). The current genetic composition of a crop affects how well its members will adapt to the physical and biological environments of the future. Rapid

world population growth, along with migration in the world, are also some of the pressures to reduce the genetic diversity of living things, especially plants. Due to the need for living and farming space, humans have made the genome range narrower and more homogeneous leading to the life conditions become less favorable to exist and reproduce. Such populations are more likely to become locally extinct, and in extreme cases, all plant species may be at risk of extinction. Reducing genetic variation is a major step in the extinction process. Migration as well as population growth and unsustainable food production systems have been shown to cause a decrease in the variation of genetic loci (gene alleles) that controls physical and phenotypic responses to climate change (Jump, Peñuelas, 2005). Therefore, hereditary variations are considered as the factors that help plants to survive through changing environments. This is also the fundamental for genetic breeding (Xia *et al.*, 2019).

Genetic diversity analysis is necessary for conservation and reintroduction of rare and endangered species (Zhuravlev *et al.*, 2010). A species with a small hereditary changeability may endure from decreased wellness in its current environment. It may not have the developmental potential essential for a changing environment. Hereditary inconstancy is basic for a species to adjust to natural changes and survive in the long term (Reed, Frankham, 2003). Knowledge of genetic diversity within and among populations is imperative for preservation administration, particularly for distinguishing hereditarily interesting basic units inside a species (Reunova *et al.*, 2014). In recent years, the development of various chemical and molecular techniques for studying genetic diversity has received great attention. In this article, we review some popular methods for the detection of genetic variation and point out DNA marker systems, which are considered as reliable and widely available tools used to distinguish variations.

DNA markers

DNA markers are the foremost broadly

utilized markers basically due to their plenitude (Yang *et al.*, 2015). They emerge from distinctive types of DNA mutations, such as point mutations, rearrangements (insertion or deletion) or repeated DNA replication mistakes (Mammadov *et al.*, 2012). These markers are specifically unbiased since they are found in non-coding regions of the DNA within the chromosomes. Unlike others, DNA markers are boundless in number and unaffected by environmental factors as well as plant growth stage (Govindaraj *et al.*, 2015). They can be classified into three categories based on their detection strategy: hybridization, polymerase chain reaction (PCR), and DNA sequencing (Nadeem *et al.*, 2018).

Hybrid-based DNA marker

Restriction fragment length polymorphisms, or RFLPs, was the first DNA-based marker and also the typical hybrid-based marker widely used in biology including evaluating genetic diversity (Petersen *et al.*, 1994). This is a collection of DNA fragments with different lengths cut by restriction enzymes, also known as polymorphisms, which can be visualized separately on agarose gels based on its size, or transferred to a membrane via Southern blot. It was an important marker for genome mapping, localization of genes for genetic disorders, determination of risk for disease, and paternity testing (Ben-Ari, Lavi, 2012). The advantages of RFLPs include unlimited locus detection and are primarily robust, reliable and results can be transferred across populations. However, it is very expensive, time-consuming, labor-intensive, and it requires a larger amount of DNA, and limited polymorphism, when comparing to PCR-based markers developed later (Tabit, 2016).

PCR-based markers

The discovery of PCR technique in 1984 by Kary Mullis opened a second-generation of PCR-based DNA markers (Singh, Singh, 2015). Comparing to the first-generation hybridization-based markers, PCR-based DNA markers were considered as simpler, less expensive, and more convenient marker systems. They requires much smaller quantity of DNA of relatively lower quality and are much more user friendly and amenable to automation (Srivastava *et al.*, 2012). Simple sequence repeats (SSRs), amplified fragment length polymorphisms (AFLPs), and randomly amplified polymorphic DNA (RAPD) markers are some of the widely used PCR-based markers. Even RFLPs, one of former DNA markers has been renewed and improved its efficiency when combined with PCR to form PCR-RFLPs (Singh, Singh, 2015).

Radom amplified polymorphic DNA (RAPD)

RAPD is a PCR-based marker generated by PCR machines that use genomic DNA and random primers to make multiple copies of DNA chains (Simner *et al.*, 2015). Less stringent conditions allow primers binding to both sample DNA strand of template where it is matched or partially matched, resulting in strain-specific heterogeneous DNA products (Sorof, Cheng, 2015). The advantages of RAPD include being fast, simple, inexpensive, and multiple loci can be amplified from a single pair of primers with a small amount of DNA (Sharaf-Eldin *et al.*, 2018). However, the results from RAPD may not be reproduced in different laboratories and can only detect key features of interest (Ramos *et al.*, 2008).

Amplified fragment length polymorphisms (AFLPs)

AFLPs are generated by digesting PCR amplified fragments using specific restriction enzymes to cut DNA at or near a specific nucleotide-specific site. AFLPs are highly reproducible and this enables the creation of fast and high-frequency identifiable AFLPs, making it an attractive technique for polymorphic identification and interconnection identification. AFLPs system detects polymorphism due to the sequence variation in and around the recognition sites of restriction endonucleases and uses PCR for marker assay (Paun, Schönswetter, 2012). AFLPs possessed high degrees of polymorphism and good reproducibility, have been widely used in population genetics studies (Mba, Tohme, 2005).

AFLP markers are dominant, so it cannot distinguish heterozygotes from homozygotes, but they can enhance the resolution achieved in population assignment, especially among weakly differentiated populations (Campbell *et al.*, 2003; Mba, Tohme, 2005). Primers for AFLP markers are universal among species, which can be easy to design for any experiments. In order to detect genetic diversity, AFLPs were obviously good in terms of the success rates they achieved and the polymorphism detection (Song *et al.*, 2015). Thus, AFLP markers are still practical markers to use when evaluating the genetic diversity and population structure of plants (Huang *et al.*, 2019).

Microsatellites

Microsatellites are short (1-8 bp) repeat motifs usually associated with a high level of frequency of length polymorphism, which are considered among the best markers for genetic research (Hosseinzadeh-Colagar *et al.*, 2016; Vieira *et al.*, 2016). Microsatellites are also understood as simple sequence repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphisms (SSLPs) (Vieira *et al.*, 2016). SSRs is very diverse and uniformly distributed in the genome and common in eukaryotes. The number of repetition units varies greatly among plant species. The repetition sequence is usually simple, consisting of two, three or four nucleotides, also known as di-, tri- and tetranucleotide, respectively, in which, the most common for microsatellites is repeat dinucleotide (Fan, Chu, 2007). The inter-space areas tend to be conserved within species, although sometimes they can also be preserved at a higher taxonomic level.

SSRs are usually isolated on polyacrylamide gel in combination with $AgNO_3$ staining, autoradiography, or fluorescence detection systems (Hosseinzadeh-Colagar *et al.*, 2016). Agarose gels with ethidium bromide may also be used when the difference in allele size between samples is greater than 10 bp (Li *et al.*, 2017). Electrophoresis bands can be scored in a denominator either as it is present or absent.

SSRs is easy to automate, highly polymorphic, along with good resolution and reproducibility. This is most widely used for individual genotypes, germ cell evaluation, genetic diversity research, genome mapping, phylogenetic and evolutionary research (Vieira *et al.*, 2016). However, the development of microsatellites required extensive comprehension of the DNA sequence. Thus, they are developed primarily for agricultural species, instead of the wild (Mondini *et al.*, 2009).

Inter-simple sequence repeats (ISSRs)

Inter-simple sequence repeats (ISSRs) are used to assess genetic diversity and identify closely related cultivars in many plant species (Verma *et al.*, 2017). The ISSR marker system detects polymorphisms in inter-microsatellite DNA regions without any prior sequence knowledge (Zietkiewicz *et al.*, 1994). Primers are based on a repeat sequence, often with a degenerate 3′anchor, and amplify plenty number of amplicons, providing high reproducibility at a low cost. ISSRs have been used for a wide range of organisms in DNA fingerprinting, diversity analysis, and genome mapping (Rahimi *et al.*, 2019). ISSR technique has demonstrated as a powerful, rapid, simple, reproducible, and inexpensive way, especially ISSR markers are more reproducible than RAPD for the same purpose (Verma *et al.*, 2017). However, a major disadvantage of ISSR markers is that they are not highly reproducible, and some primers generate poorly reproducible band patterns (Singh, Singh, 2015).

Single nucleotide polymorphisms (SNPs)

SNPs are single base substitutions inside DNA sequence of the individuals in the population that occur when a single nucleotide in the genome sequence is changed. SNPs occur more often than any other type of marker and are very close or even in the gene of interest. SNPs, the most abundant in the genome, are widely distributed across genomes with different distributions among species of plants and other organisms (Govindaraj *et al.*, 2015). They provide valuable markers for fast identification

of plant varieties, the construction of ultra-high-density genetic maps, and developing adaptive traits in plant species (Fischer *et al.*, 2017; De Lorenzis *et al.*, 2019; Xia *et al.*, 2019). SNP markers represented to third-generation molecular markers which possessed many significant benefit as high genetic stability and diversity, but their use requires advanced technology and has high costs (Xu *et al.*, 2015).

Cleaved amplified polymorphic sequence (CAPS)

CAPS is a combination of the RFLPs and PCR and it was originally named PCR-RFLPs, which using amplified DNA fragments digested with a restriction endonuclease to display restriction site polymorphisms (Singh, Hazarika, 2020). The technique involves amplification of a target DNA through PCR, followed by digesting with restriction enzymes (Elangbam, Misra, 2016). Therefore, CAPS markers rely on differences in restriction enzyme digestion patterns of PCR fragments caused by nucleotide polymorphism between samples. Critical steps in the CAPS marker approach included DNA extraction, and the number or distribution of polymorphic sites, and PCR conditions. The restriction fragment length polymorphisms are based on PCR amplification, it is much easier and less time-consuming than analysing alternative types of markers that require southern hybridizations. Next, CAPS primers are more useful as genetic markers for comparative mapping study than those markers derived from non-functional sequences such as genomic microsatellite markers (Matsumoto, Tsumura, 2004). In genetic diversity analysis, their genotypes which are easily scored and interpreted, and only a small quantity of DNA is needed for one assay. Also, the cleaved and un-cleaved amplification products can be adjusted arbitrarily by the appropriate placement of the PCR primers. The procedure is technically simple with robust results because the amplification product is always obtained (Idrees, Irshad, 2014).

Expressed sequence tags (ESTs)

DNA markers could be developed from whole nuclear genome or expressed sequence tags (ESTs) (Singh *et al.,* 2016). ESTs are short cDNA sequences reversely transcribed from mRNA. The identification of ESTs has proceeded rapidly, with over 6 million ESTs now available in computerized databases. In general, ESTs were instrumental in gene discovery, for obtaining data on gene expression and regulation, sequence determination, and for developing highly valuable molecular markers, such as EST-based RFLPs, SSRs, SNPs, and CAPS (Idrees, Irshad, 2014). Using EST-derived primer pairs to amplify nuclear genome, the amplicons may consist of intron sequences that displayed higher variation to develop informative markers for variety identification (Shu *et al.,* 2010).

Cytoplasmic genome-derived-DNA markers

DNA markers could be also derived from the cytoplasmic genome, such as the mitochondrial genome (mtDNA) and chloroplast genome (cpDNA). The cytoplasmic CAPS markers are not only maternal inherited from haploid genome but also have a slower nucleotide substitution rate than the nuclear DNA (Kaundun, Matsumoto, 2011). Because of conservative evolution, they have been widely used in detecting geographical origins of plant species (Kaundun, Matsumoto, 2002; Katoh *et al.,* 2003) and population differentiation (Schaal, Olsen, 2000).

The studies often use genetic information contained in cpDNA and mtDNA, as they contain two necessary sets of genes in plants (Skuza *et al.,* 2019). They encode many key proteins for basic cell bioenergy processes. Additionally, they encode many components necessary for the proper expression of their own genes. Considering the essential importance of these two sets of organelle genes, one can expect them to change very slowly during evolution (Bendich, 2010). In addition, organelle DNA is a useful tool in the search for species-specific molecular markers.

Interspecies varieties can be effectively and quickly recognized utilizing chloroplast genome than by utilizing existing widespread DNA standardized tag, and subsequently, the

chloroplast genome has been utilized as the premise for creating markers for species distinguishing proof and classification (Chen *et al.,* 2010). Chloroplast-based SNP markers have as of now been created for the major restorative crops, counting ginseng, and they have been effectively utilized for species recognizable proof, phylogenetic investigation, and hereditary

differing qualities evaluation (Jo *et al.,* 2016). Additionally, chloroplast-derived CAPS markers have moreover been detailed to identify the intra- and interspecies variety of diverse species (Kim *et al.,* 2018).

The basic characteristics of some common DNA markers are shown in Table 1 and their pros and cons are showed in Table 2.

Table 1. Important feature of different types of DNA markers (Miah *et al.*, 2013).

Feature	RFLPs	RAPDs	AFLPs	SSRs	SNPs
DNA Require (µg)	10	0.02	0.5–1.0	0.05	0.05
PCR based	No	Yes	Yes	Yes	Yes
DNA quality	High	High	Moderate	Moderate	High
No. of polymorph loci analyzed	1–3	1.5–50	20–100	1–3	1
Type of polymorphism	Single base change, insertion, deletion	Single base change, insertion, deletion	Single base change, insertion, deletion	Change in repeat length	Single nucleotide change, insertion, deletion
Dominance	Co-dominant	Dominant	Dominant/Co-dominant	Co-dominant	Co-dominant
Reproducibility	High	Unreliable	High	High	High
Ease of use and development	Not easy	Easy	Easy	Easy	Easy
Automation	Low	Moderate	Moderate	High	High
Cost per analysis	High	Low	Moderate	Low	Low
Developmental cost	Low	Low	Moderate	High	High
Need for sequence data	Yes	No	No	Yes	Yes
Accuracy	Very high	Very low	Medium	High	Very high
Radioactive detection	Usually yes	No	No	No	Yes
Genomic abundance	High	Very high	Very high	Medium	Medium
Part of genome surveyed	Low copy coding regions	Whole genome	Whole genome	Whole genome	Whole genome
Level of polymorphism	Low	Low to moderate	Low to moderate	High	High
Effective multiplex ratio	Low	Medium	High	Medium	Medium
Marker index	Low	Medium	High	Medium	Medium
Inheritance	Codominant	Dominant	Dominant	Codominant	Codominant
Detection of alleles	Yes	No	No	Yes	Yes
Utility for genetic mapping	Species specific	Cross specific	Cross specific	Species specific	Species specific
Utility in marker assisted selection	Moderate	Low to moderate	Low to moderate	High	Low to moderate

Feature	RFLPs	RAPDs	AFLPs	SSRs	SNPs
Cost and labor involved in generation	High	Low to moderate	Low to moderate	High	High

Table 2. Some potential advantage and weakness of the most commonly used DNA markers (Miah *et al.*, 2013)

Marker type	Benefits	Weakness
RFLPs	Co-dominant Genomic abundance high Highly reproducible Better genome exposure Applicable across the species No need for sequence information Reliably used in plants	Need high-quality DNA Laborious (compared to RAPD) Complex to automate Radioactive labeling essential Characterization of probe is essential
RAPDs	Genomic abundance high Better genome coverage Sequence information unneeded Perfect for automation Requires less DNA No radioactive labeling More rapid	No need of probe information Dominant markers Not reproducible Not suitable for across species Not well tested
SSRs	Easy to automate Genomic abundance high Highly reproducible High polymorphism Multiple alleles Moderately genome coverage No radioactive labeling	Not well-examined Not suitable for across species Sequence information needed
AFLPs	High polymorphism Genomic abundance high Can be used across species No need for sequence information Useful in preparing counting maps Works with smaller RFLPs fragments	Very tricky due to changes in materials use Not reproducible Very good primers needed
Sequence-tagged site (STS)	Helpful in preparing counting maps Highly reproducible No radioactive labeling Can use filters many times Moderate genome coverage	Need sequence information Out of the target sites, mutation detection not possible Laborious Cloning and probe characterization required

DNA markers based on the sequencing

DNA sequencing was first introduced by Sanger and then developed as method allowing rapid identification of DNA sequences based the DNA polymerase activity during DNA synthesis (Sanger, Coulson, 1975). The DNA sequencing technique using dideoxy nucleotide to stop the random DNA synthesis; or by chemical using marked DNA later have widely applied around the world (Metzker, 2005). Success in sequencing the human genome in 2003 has opened a new era of development in life science research with many new, complex and modern techniques, developed and applied (Hood, Rowen, 2013). New generation DNA sequencing did not rely on Sanger chemistry, enables efficient and rapid decoding of the entire genome sequencing -WGS). 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 (Margulies *et al.*, 2005; Rothberg, Rothberg, 2015). Commercial 2nd generation sequencing methods included four main platforms such as Roche 454 GS FLX, Illumina Genome Analyzer IIx, ABI SOLiD 3 Plus System and Polonator

G.007 (Lerner, Fleischer, 2010). They were all amplification-based and could be distinguished by the role of PCR in library preparation.

The single-molecule sequencing method, also known as 3rd generation or next-next generation, was recently developed by Helicos Genetic Analysis System using the technology developed by Braslavsky *et al*. (Braslavsky *et al.*, 2003; Rothberg, Rothberg, 2015). There was high-throughput and low-cost sequencing method which provides the first universal genetic analysis platform that requires no amplification. It pursues a molecular sequencing strategy that

simplifies the process of DNA sample preparation, avoids errors and deviations caused by PCR, simplifies data analysis, and tolerates degraded samples (Ozsolak *et al.*, 2009). With the advantage of fast, high-throughput and accuracy, the new gene sequencing systems are widely used in most modern biology research (Gasperskaja, Kučinskas, 2017). However, the traditional sequencing method (Sanger method) with long readings and high accuracy continues to be used in many cases, especially in small-scale projects with limited samples (Shendure, Ji, 2008). The features of each sequencing technology platforms are summarized in Table 3.

Table 3. Summary of the output per sequencing technology platforms (Buermans, den Dunnen, 2014).

		Sequence by	Detection	Run types	Run time	Read length (bp)	# reads per run	Output per run	Remarks
Roche	GS FLX Titanium XL +	Synthesis	Pyrophosphate detection	Single end	23 h	700	1 million	700 Mb	
	GS Junior System	Synthesis	Pyrophosphate detection	Single end	10 h	400	0.1 million	40 Mb	
LifeTechnologies	Ion torrent	Synthesis	Proton release	Single end	4 h	200–400	4 million	1.5–2 Gb	Ion318 Chip
	Proton	Synthesis	Proton release	Single end	4 h	125	60–80 million	8–10 Gb	IonP1 chip
	Abi/solid	Ligation	Fluorescence detection of di-base probes	Single and paired-end	10 days	75 + 35	2.7 billion	300 Gb	
Illumina/solexa	HiSeq2000/2500	Synthesis	Fluorescence; reversible terminators	Single and paired-end	12 days	2 × 100	3 billion	600 Gb	High output mode
	MiSeq	Synthesis	Fluorescence; reversible terminators	Single and paired-end	65 h	2 × 300	25 million	15 Gb	
Pacific biosciences	RSII	Single molecule synthesis	Fluorescence; terminally phospholinked	Single end	2 days	50% of reads > 10 kb	0.8 million	5 Gb	16 SMRT cells
Helicos	Heliscope	Single molecule synthesis	Fluorescence; virtual terminator	Single end	10 days	~ 30	500 million	15 Gb	Two flow cells in parallel

Target-enrichment strategies for next generation sequencing

In plant investigation, targeted re-sequencing of enriched genomic DNA regions has ended up a versatile and cost-effective strategy for the

disclosure of genome-wide sequence varieties to be misused to address diverse organic questions (Kaur, Gaikwad, 2017). Plants have expansive or polyploid genomes, therefore entire genomes may not be promptly gathered and the examination still exceptionally costly, that

596

requires an elective procedure to WGS and to produce a diminished representation of the genome (Fuller *et al.*, 2009). Genetic variations can be obtained by utilizing target enrichment methodologies, which comprises within the separation of particular genomic loci (e.g., qualities, atomic markers, bigger genomic locales, and organelle genomes) coupled with NGS (Gnirke *et al.*, 2009). Compared to WGS, the decrease in sequencing space involves three primary focal points, test multiplexing that embroils a generally lessening of the sequencing fetched per sample; critical decrease within the complexity of the analysis (Mertes *et al.*, 2011).

Re-sequencing the genomic locales that are held is essentially more time- and cost-effective, and the coming about information are significantly less awkward to examine. Few approaches to target enrichment have been created and there are some parameters by which the execution of each can be measured, which shift from one approach to another such as affectability, specificity, consistency, reproducibility, fetched; ease of utilizing; and sum of DNA required. Currently, transcriptome-based, confinement enzyme-based, PCR-based, and hybridization-based strategies, all consistent with the foremost well-known NGS stages, have been created to improve particular targets (Cronn *et al.*, 2012).

Transcriptome-based enrichment is one of the foremost broadly utilized methodologies to diminish genome complexity, since it focuses only on the transcribed portion of the genome. The key point of transcriptome sequencing, moreover known as RNA-seq, is to determine quality expression profiles of each transcript amid advancement and beneath diverse conditions (Martin *et al.*, 2013). SNP discovery and molecular marker development via RNA-seq are often performed, especially in organisms with large genomes (Egan *et al.*, 2012). Essentially, because RNA-seq is autonomous from any a priori comprehension on the genome sequence of the species beneath analysis, it permits the investigation of ineffectively characterized species.

Restriction enzyme-based enrichment makes use of the discriminatory power of the restriction endonucleases to produce restriction fragments among individuals in a population. Three main techniques have been developed so far: RAD-seq (restriction-site associated DNA sequencing), GR-RSC (genomic reduction based on restriction site conservation), and GBS (genotyping-by-sequencing). All these methods are flexible and quite inexpensive and have been used to identify and score, in a group of individuals, thousands of genetic markers randomly distributed along the genome enabling SNP discovery, genotyping as well as quantitative genetic and phylogeographic studies (Baird *et al.*, 2008; Maughan *et al.*, 2009; Rowe *et al.*, 2011; Cronn *et al.*, 2012; He *et al.*, 2014).

PCR-based target enrichment includes the direct sequencing of small and long PCR products. NGS of PCR fragments has been preferentially applied to chloroplast genomes in systematic studies and in some cases, to nuclear genomic regions despite their complexity. The main disadvantages associated with this method are the high level of failed target amplifications and/or non-specific amplifications as well as the difficulty in obtaining an accurate pooling of samples for NGS multiplexing. Anyway, PCR-based enrichment remains feasible for targeting small to medium-sized regions of the genome, but for high-throughput sequencing of tens of thousands of PCR amplicons, its efficiency falls off, given the initial cost per sample and challenges in sample multiplexing. Microfluidic-based multiplexing PCR can reduce costs but continues to be more expensive than other enrichment methods (Durstewitz *et al.*, 2010; Mamanova *et al.*, 2010; Cronn *et al.*, 2012; Uribe-Convers *et al.*, 2016).

Hybridization-based enrichment or sequence capture methods exploit the high specificity of DNA or RNA probes (also called baits), which are designed to be complementary to target genomic regions. RNA baits have significant advantages over DNA probes because RNA-DNA hybrids have a higher affinity and melting temperature than DNA-DNA hybrids. Two main

597

technologies have been developed for hybrid-capture applications: on-array- or solid-based hybridization, which implies sample hybridization on a solid support and in-solution- or liquid-based hybridization, where pooled baits are used in reaction tubes. Due to their moderate costs and high specificity, low amounts of required DNA per sample and power to simultaneously target large numbers of markers, several protocols and commercial kits have been developed. The most widespread ones and reliable in studies on plant species were provided by Agilent Technologies (SureSelect), Roche NimbleGen (SeqCap EZ), MYcroarray (MYbaits), and Ion Torrent (TargetSeq) (Okou *et al.*, 2007; Gnirke *et al.*, 2009).

GENETIC DIVERSITY ANALYSIS IN WILD, CROP, AND MEDICINAL PLANTS

Genetic diversity analysis in wild plants

DNA markers were widely applied for genetic diversity analysis of wild plant. For instance, a series of DNA markers as AFLPs, ISSRs, SSRs were used to analyse the diversity of 389 accessions of 18 wild almond species, which were considered as crucial importance in breeding. The result showed highest polymorphic information for SSR markers (Sorkheh *et al.*, 2017). SSRs were also applied to analyse the genetic diversity of 86 almond accessions of diverse geographic origin, ranging from Central Asia to the USA. The results showed that there were slight losses of genetic diversity by geographic isolation, human selection but no indication of a major decrease in genetic variability in almond germplasm from Asia to Europe. The results also suggested the need to avoid the decline in genetic diversity during the almond domesticating even though its risk is present in many subpopulations (Halász *et al.*, 2019). Genetic diversity in Chinese wild apple species along with cultivars was evaluated using SSR markers (Zhang *et al.*, 2012), zombi pea (*Vigna vexillata*) (Dachapak *et al.*, 2017), wild soybean (*Glycine soja*) (Nawaz *et al.*, 2017), wild date palm (*Phoenix sylvestris*) (Huda *et al.*,

2018), wild kiwifruit (*Actinidia eriantha*) (Huda *et al.*, 2018; Liao *et al.*, 2019).

RADP is also a traditional marker commonly used in investigation the genetic diversity of wild plants until now. In the past, RAPD markers were utilized in the analysis genetic variation of *Saxifraga cernua* (Bauert *et al.*, 1998,) *Calamagrostis porteri* (Poaceae) (Esselman *et al.*, 1999) or to reveal genetic diversity within and between populations of cashew (*Anacardium occidentale* L.) (Mneney *et al.*, 2001). Genetic variability and species identification within 22 species of the genus *Encephalartos,* the second largest genus with 65 species and 2 sub-species were analysed using RAPD (Prakash, Van-Staden, 2008). The result showed that most species of *Encephalartos* are morphologically very similar and therefore could benefit from additional tools for their correct identification. RAPD markers were also applied to assess genetic diversity and structure of natural *Calophyllum brasiliense* (Clusiaceae) populations in Riparian forests (Mendonça *et al.*, 2014). The genetic differentiations among *Jacaratia mexicana* populations, a native tree in Mexico, also be discovered at even small geographic scale by RAPD method that considered in conservation actions for this genetic resource (Arias *et al.*, 2012). These studies above suggested the potential of the RAPD for correct genetic identification of individual species (Prakash, Van-Staden, 2008).

Some other wild plants were also estimated the level of genetic diversity, such as Phragmites (Poaceae) by AFLP fingerprinting, Roseroot (*Rhodiola rosea* L.) by SSRs and ISSRs, *Saruma henryi,* an endangered plant endemic to China by SSRs, *Galium cracoviense, G. oelandicum, G. sudeticum* (Rubiaceae) and Caragana species by AFLPs or genetic diversity of *Calamagrostis porteri* (Poaceae), *Amentotaxus argotaenia* (Taxaceae) were determined by ISSRs (Esselman *et al.*, 1999; Ge *et al.*, 2005; Cieslak, Szelag, 2010; György *et al.*, 2012; Zhou *et al.*, 2012). Other markers such as AFLPs, ISSRs have been commonly used in the past but are not

currently used frequently for analysis of genetic diversity.

Genetic diversity analysis in crop plants

Plants are very important to the survival and development of humans. Plant domestication has been done for a long time based on desirable traits. The interest in genetic diversity has been paid attention and especially with the helpful support of DNA markers. DNA markers were used in plant breeding for support the evaluation of propagation materials and cross-breeding (Govindaraj *et al.*, 2015). Some supported effects of DNA markers can be listed here for evaluation of genetic diversity, parental selection, identification and purity evaluation of varieties, research on heterozygosity, and identifying the currently selected genomic regions (Nadeem *et al.*, 2018).

Traditional markers as RFLPs have been widely used to compare genomes in major cereal families such as rye, wheat, maize, sorghum, barley, and rice or other plant as pea (Dijkhuizen *et al.*, 1996; Lu *et al.*, 1996). Another one, ISSR marker was used for cultivar identification, genetic diversity analysis and genotypes validation in many species including banana, sorghum, Arabidopsis, sunflower, tomato (Godwin *et al.*, 1997; Bornet, Branchard, 2001; He *et al.*, 2003), and mango varieties in different mango growing regions including Australia (González *et al.*, 2002), China (He *et al.*, 2005), and India (Damodaran *et al.*, 2012).

SSRs seemed to be a DNA marker which is most widely used in analysis of genetic biodiversity. Twenty-four SSR markers were utilized to evaluate the population structure and genetic diversity of thirty-six guava varieties (*Psidium guajava* L.) including wild species to develop varieties with better fruit yield and nutritional quality which assisted assessment for morpho-physiological traits in guava (Kherwar *et al.*, 2018). *Chrysophyllum albidum* is endangered tree species among the forest tree of which fruit is widely consumed and play a significant role in food security in many parts of tropical Africa. A very low genetic variation in

C. albidum population was discovered by SSR markers suggested the need for preserving the remnant inherent diversity towards capturing the existing local adaptation (Boboye *et al.*, 2018). To assess the genetic diversity of 109 newly introduced accessions of narrow-leafed lupin (*Lupinus angustifolius* L.), a great potential to be a new crop in China (Ji *et al.*, 2020), 76 genomic SSR markers were utilized. SSR markers were also applied for evaluating the genetic structure and gene flow pattern of 285 domestication cotton to provide useful information for understanding the genetic base of upland cotton superior varieties and will also promote future high yield and excellent fiber quality breeding (Zhang *et al.*, 2020).

Rice (*Oryza sativa* L.), is an important food crop for human life demand. Genetic diversity analysis of rice germplasm was investigated using PCR-based SNP markers to assess genetic diversity among indigenous rice varieties to identify the different blast resistance genes in a collection of 74 rice germplasms from Tripura (India) (Anupam *et al.*, 2017). Another study using the 32 SSR markers to investigate the genetic diversity of 50 aromatic rice accessions from three regions (Peninsular Malaysia, Sabah, and Sarawak) to determine the high yielding aromatic rice genotypes. This was needed to identify the potential diverse genotypes for use as a parent in future rice breeding program (Jasim Aljumaili *et al.*, 2018).

Rye (*Secale cereale* L.) is also a cereal grass that is an important food crop in Central and Eastern Europe but the genomic resources in rye are underdeveloped because few population genetic studies using genome-wide markers have been published to date. A collection of 603 individuals from 101 genebank accessions of domesticated rye and its wild progenitor *S. cereale* subsp. *vavilovii* and related species in the genus *Secale* was done. The analysis of SNPs data of rye and its wild relatives obtained through sequencing detected 55,744 SNPs relating to cultivated rye yield with present genotype calls in 90% of samples (Schreiber *et al.*, 2019).

Genetic diversity analysis in medicinal plants

Medicinal plants are defined as plants that have medicinal properties due to complex chemical substances of different composition such as alkaloid, glycoside, corticosteroid, and essential oils, which are basic ingredients for traditional and modern medicines and therapies for health care and also for a number of other purposes (pharmaceuticals, herbs, herbal foods) (Paramanik, Chikkaswamy, 2014). The World Health Organization (WHO) considers wild plants and plants used for medicinal purposes to be medicinal plants and they are highly significant for their role in the traditional therapies used by 70-80% of citizens of developing countries (Aziz *et al.*, 2018). The number of medicinal plants being used globally varies from 70,000 to 125,000 (Schippmann *et al.*, 2006). In India, from 7,000 to 8,000 plants are considered as medicinal plants (Rajeswara *et al.*, 2012). In Vietnam, the number of medicinal plants used is 4,700 while the resources of medicinal plants can be much larger, estimated to reach 6,000 species (Chi, 2012). However, the breakthrough of modern humans' activities in forests for exploitation, farms, timber, transportation, industrial housing etc., leads to loss/fragmentation/degradation of habitats of plants species including to medical plant, leading to loss of their genetic diversity that cannot be recovered (Haddad *et al.*, 2015).

The efforts to assess genetic diversity and erosion through molecular marker techniques results in positive genetic diversity that exists in existing genetic resources. A typical example, RAPD markers were used to investigate the genetic diversity relationship of 18 medicinal plant species (Paramanik, Chikkaswamy, 2014). In another study, 479 samples of *A. austrokoreense* from 7 populations collected in South Korea were analysed the genetic diversity and structure by microsatellite amplification with the size of the amplified products ranged from 300-5000 bp depended on the sequence of random primers and medicinal plants (Lee *et al.*, 2018). ISSR, PCR-RFLP and SSR markers were applied for analysing the genetic diversity of *Peganum*

harmala L. (Peganaceae), a perennial herbaceous plant for medicinal purposes to treat hypertension, cardiac disease, some nervous system disorders such as Parkinson's disease, *Lumbago asthma*, colic, jaundice and as a stimulant emmenagogue (El-Bakatoushi, Ahmed, 2018). The genetic diversity of the endemic and medicinally important plant *Rheum officinale* also revealed by ISSR markers (Wang *et al.*, 2012).

Genetic diversity analysis of *Panax*

Panax (*Panax* spp.) is one of the most widely used medicinal herbs in the world that contains about 11 species of very slow-growing perennial plants with tubers belonging to the Araliaceae (Shahrajabian *et al.*, 2019). They distributed mostly in Korea, China, Eastern Siberia, and North American distribution and 35 other countries around the world, where characterized by cold countries (Li, Mazza, 1999). The two most important species grown commercially are *Panax quinquefolius* L. (North American ginseng), which is native to eastern North America, and *Panax ginseng* C. A. Meyer (Korean ginseng), which is native to north-eastern China and the *Korean peninsula* (Proctor, Bailey, 2011). In North America, the harvesting of *P. quinquefolius* began over 100 years ago from wild populations in Canada and the new cultivating primarily in British Columbia and Ontario in recent 30 years (Bai *et al.*, 1997; Li, Mazza, 1999). Recently, a new species of *Panax* was named *Panax vietnamensis* Ha et Grushv., discovered in Vietnam, is a true ginseng found in the south pole and to be considered as a newest subject to be studied. *P. vietnamensis* is the most recently identified *Panax* species. The plants are rarely identified at 1700-2000 meters high-elevation forests with cool weather all year round in the tropical Central Vietnam (Dung, Grushvisky, 1985; Court, 2006). The plants were known for their medicinal value in Vietnam like other *Panax* species. The diversity of secondary metabolites and pharmacological effects of *P. vietnamensis* were revealed in recent years (Court, 2006; Le *et al.*, 2015) beside the studies of phylogenetic analysis and transcriptomics in *P. vietnamensis* and related natural variant were

performed (Komatsu *et al.*, 2001; Zhang *et al.*, 2015; Manzanilla *et al.*, 2018). However, the comprehension of current genomic resource of *P. vietnamensis* is very limited.

Ginseng has a relatively high level of genetic diversity, similar to cultivated and wild groups. Ninety-two polymorphic microsatellite markers were developed in 147 individual plants, including cultivars, breeding lines, and wild populations in Korea and neighbouring countries to elucidate genetic diversity. There was no statistically significant difference in genetic diversity between cultivated and wild ginseng groups, but high level of genetic diversity in current breeding populations for breeding of elite ginseng cultivars (Jang *et al.*, 2019). It is limited research effort to analyse the genetic diversity and population structure of ginseng because of its growth habits. Even though, recent studies tried to analyse diversity level by regional distribution. Using SSR markers approach, 94.0%, 5.5% and 0.5% of 1109 accessions of Ginseng (*Panax ginseng* C.A. Meyer) in South Korea were landraces, breeding and cultivars lines, respectively, that were revealed from 56 different alleles from 12 clusters and indicated that average gene diversity was 0.49 (Lee *et al.*, 2020). RAPD-PCR was utilized for investigating the genetic diversity of North American ginseng (*P. quinquefolius* L.) population in Ontario (Canada). The similarity coefficients among the DNA of ginseng plants analysed were low, indicating that a high degree of genetic diversity existed in the ginseng population. However, lower levels of genetic diversity were detected among 3-year-old ginseng plants selected on the basis of greater plant height than among the plants randomly selected from the same subpopulation or over the whole population. This research result suggested that genetic factors at least partly contribute to morphological variation within the ginseng population and that visual selection can be effective in identifying the genetic differences. The significance of a high degree of genetic variation in the ginseng population on its potential for improvement by breeding (Bai *et al.*, 1997). Genetic diversity of *P. stipuleanatus* Tsai in North

Vietnam detected by ISSR markers showed the higher population genetic diversity in the Lao Cai than in Lai Chau region and a highly gene flow within *P. stipuleanatus* populations in Vietnam (Trieu *et al.*, 2016).

CONCLUSION

A variety of plant genetic sources are essential possessions for human. Natural genetic variant in vegetation is extensive, and stays to be completely described and harnessed in plant species. This may additionally be accomplished through the phenotypic and molecular markers of plant genetic assets. In the field of molecular genetics studies, molecular markers have been developed and utilized extensively for the exploitation and identification of plant genetic diversity. Various techniques are used to estimate genetic diversity, in which, the most popular DNA markers for genetic diversity in plants include RAPDs, ISSRs and AFLPs, or RFLPs, SSRs, CAPS, ESTs, SNPs.

Today, new techniques are being developed frequently. No such techniques are ideal yet, each technique has its own advantages and limitations. However, low cost, inexpensive equipment, throughput, alleviation, and ease of check advancement and computerization are indispensable components when deciding on innovation. For example, AFLPs and RFLPs were common techniques for analyzing genetic diversity, however, the large requirement of time and amount of DNA required for analysis makes them no longer effective. Instead, newer markers such as SNPs or CAPS with more advantages become more popular. Some markers that still show efficacy in such analyzes, such as RAPDs or SSRs, continue to be widely used in genetic diversity studies. New methods are frequently being developed to improve their effectiveness and relevance to use although no method is ideal for all applications so scientist should look at both pros and cons of methods when starting a new project. As science advances, it will be feasible to gather near-complete descriptions of genetic variant in natural populations.

REFERENCES

Anupam A, Imam J, Quatadah SM, Siddaiah A, Prasad Das S, Variar M, Prasad Mandal N (2017) Genetic diversity analysis of rice germplasm in tripura state of northeast india using drought and blast linked markers. *Rice Science* 24: 10-20.

Arias DM, Albarrán-Lara AL, González-Rodríguez A, Peñaloza-Ramírez J, Dorado O, Leyva E (2012) Genetic diversity and structure of wild populations of the tropical dry forest tree *Jacaratia mexicana* (Brassicales: Caricaceae) at a local scale in Mexico. *Revista de Biologia Tropical* 60: 1-10.

Aziz MA, Adnan M, Khan AH, Shahat AA, Al-Said MS, Ullah R (2018) Traditional uses of medicinal plants practiced by the indigenous communities at Mohmand Agency, FATA, Pakistan. *J Ethnobiol Ethnomed* 14: 2.

Bai D, Brandle J, Reeleder R (1997) Genetic diversity in North American ginseng (*Panax quinquefolius* L.) grown in Ontario detected by RAPD analysis. *Genome* 40: 111-115.

Bailey JK, Schweitzer JA, Úbeda F, Koricheva J, LeRoy CJ, Madritch MD, Rehill BJ, Bangert RK, Fischer DG, Allan GJ, Whitham TG (2009) From genes to ecosystems: A synthesis of the effects of plant genetic factors across levels of organization. *Philos Trans R Soc B* 364: 1607-1616.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* 3: e3376.

Bauert, M.R., M. KÄLin, M. Baltisberger, and P.J. Edwards (1998). No genetic variation detected within isolated relict populations of *Saxifraga cernua* in the Alps using RAPD markers. *Mol Ecol*, 7(11): p. 1519-1527.

Ben-Ari G, Lavi U (2012) Marker-assisted selection in plant breeding. In: Altman A and Hasegawa PM (Eds) *Plant Biotechnology and Agriculture*. San Diego, Academic Press: 163-184.

Bendich AJ (2010). Mitochondrial DNA, chloroplast DNA and the origins of development in eukaryotic organisms. *Biol Direct* 5: 42-42.

Boboye OM, Lawal A, Oyerinde OV (2018) Assessment of the genetic variation in *Chrysophyllum albidum* G. Don using CP SSR marker. *Appl Trop Agricul* 23: 35-40.

Bornet B, Branchard M (2001) Nonanchored inter simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting. *Plant Mol Biol Rep* 19: 209-215.

Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci USA* 100: 3960-3964.

Buermans HPJ, den Dunnen JT (2014) Next generation sequencing technology: Advances and applications. *Biochim Biophys Acta Mol Basis Dis* 1842: 1932-1941.

Campbell D, Duchesne P, Bernatchez L (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Mol Ecol* 12: 1979-1991.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5(1): e8613.

Chi VV (2012) Dictionary of Vietnamese medicinal plants, Medical Publishing House.

Cieslak, Szelag (2010) Genetic diversity of *Galium cracoviense, G. oelandicum and G. sudeticum* (Rubiaceae)-Narrow endemic species of *Galium sect.* Leptogalium in northeastern Europe. *Acta Societatis Botanicorum Poloniae* 74: 269-275.

Court WE (2006) Ginseng: the genus *Panax*. In: Hardman R (Ed) *Medicinal and aromatic plants: Industrial profiles*, Taylor & Francis 15: 13-21.

Cronn R, Knaus BJ, Liston A, Maughan PJ, Parks M, Syring JV, Udall J (2012) Targeted enrichment strategies for next-generation plant biology. *Amer J Bot* 99: 291-311.

Crutsinger GM, Collins MD, Fordyce JA, Gompert Z, Nice CC, Sanders NJ (2006) Plant genotypic diversity predicts community structure and governs an ecosystem process. *Science* 313: 966.

Dachapak S, Somta P, Poonchaivilaisak S, Yimram T, Srinives P (2017) Genetic diversity and structure of the zombi pea (*Vigna vexillata* (L.) A. Rich) gene pool based on SSR marker analysis. *Genetica* 145: 189-200.

Damodaran T, Kannan R, Ahmed I, Srivastava RC, Rai RB, Umamaheshwari S (2012) Assessing genetic relationships among mango (*Mangifera indica* L.) accessions of Andaman islands using inter simple sequence repeat markers. *New Zealand J Crop Horticul Sci* 40: 229-240.

De Lorenzis G, Mercati F, Bergamini C, Cardone MF, Lupini A, Mauceri A, Caputo AR, Abbate L, Barbagallo MG, Antonacci D, Sunseri F, Brancadoro L (2019) SNP genotyping elucidates the genetic diversity of *Magna Graecia* grapevine germplasm and its historical origin and dissemination. *BMC Plant Biol* 19: 7.

Dijkhuizen A, Kennard WC, Havey MJ, Staub JE (1996) RFLP variation and genetic relationships in cultivated cucumber. *Euphytica* 90: 79-87.

Dung H, Grushvisky I (1985) A new species of the genus *Panax* L., Araliaceae in Vietnam: *Panax vietnamensis* Ha et Grushv. *Bot J Vietnam* 70: 518-522.

Durstewitz G, Polley A, Plieske J, Luerssen H, Graner EM, Wieseke R, Ganal MW (2010) SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome* 53: 948-956.

Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. *Amer J Bot* 99: 175-185.

El-Bakatoushi R, Ahmed DGA (2018) Evaluation of genetic diversity in wild populations of *Peganum harmala* L., a medicinal plant. *J Genet Engin Biotechnol* 16: 143-151.

Elangbam M, Misra A (2016). Development of CAPS markers to identify Indian tea (*Camellia sinensis*) clones with high catechin content. *Genet Mol Res* 15(2): gmr.15027860.

Esselman EJ, Jianqiang L, Crawford DJ, Windus JL, Wolfe AD (1999) Clonal diversity in the rare *Calamagrostis porteri* ssp. *insperata* (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and intersimple sequence repeat (ISSR) markers. *Mol Ecol* 8: 443-451.

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 to 2000. *Science* 300: 758.

Fan H, Chu JY (2007) A brief review of short tandem repeat mutation. *Genomics Proteomics Bioinformatics* 5: 7-14.

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* 18: 69.

Fu YB (2015) Understanding crop genetic diversity under modern plant breeding. *Theor Appl Genet* 128: 2131-2142.

Fuller CW, Middendorf LR, Benner SA, Church GM, Harris T, Huang X, Jovanovich SB, Nelson JR, Schloss JA, Schwartz DC, Vezenov DV (2009) The challenges of sequencing by synthesis. *Nat Biotechnol* 27: 1013-1023.

Gasperskaja E, Kučinskas V (2017) The most common technologies and tools for functional genome analysis. *Acta Med Litu* 24: 1-11.

Ge X-J, Zhou X-L, Li Z-C, Hsu T-W, Schaal BA, Chiang T-Y (2005) Low genetic diversity and significant population structuring in the relict *Amentotaxus argotaenia* complex (Taxaceae) based on ISSR fingerprinting. *J Plant Res* 118: 415-422.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27: 182-189.

Godwin ID, Aitken EAB, Smith LW (1997) Application of inter simple sequence repeat (ISSR) markers to plant genetics. *Electrophoresis* 18: 1524-1528.

González A, Coulson M, Brettell R (2002). Development of DNA markers (ISSRs) in mango, International Society for Horticultural Science (ISHS), Leuven, Belgium.

Gornall J, Betts R, Burke E, Clark R, Camp J, Willett K, Wiltshire A (2010) Implications of climate change for agricultural productivity in the early twenty-first century. *Philos Trans R Soc B* 365: 2973-2989.

Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop

plants and its recent advances: an overview of its analytical perspectives. *Genet Res* 2015: 431487-431487.

György Z, Szabó M, Bacharov D, Pedryc A (2012) Genetic diversity within and among populations of roseroot (*Rhodiola rosea* L.) based on molecular markers. *Not Bot Horti Agrobo* 40: 266-273.

Haddad NM, Brudvig LA, Clobert J, Davies KF, Gonzalez A, Holt RD, Lovejoy TE, Sexton JO, Austin MP, Collins CD, Cook WM, Damschen EI, Ewers RM, Foster BL, Jenkins CN, King AJ, Laurance WF, Levey DJ, Margules CR, Melbourne BA, Nicholls AO, Orrock JL, Song D-X, Townshend JR (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci Adv* 1: e1500052-e1500052.

Halász J, Kodad O, Galiba GM, Skola I, Ercisli S, Ledbetter CA, Hegedűs A (2019) Genetic variability is preserved among strongly differentiated and geographically diverse almond germplasm: an assessment by simple sequence repeat markers. *Tree Genet Genomes* 15: 12.

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. *Theor Appl Genet* 106: 363-373.

He J, Zhao X, Laroche A, Lu ZX, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Fron Plant Sci* 5.

He XH, Li YR, Guo YZ, Tang ZP, Li RB (2005) Genetic analysis of 23 mango cultivars collection in Guangxi province revealed by ISSR. *Mol Plant Breed* 3: 829-834.

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Lands Ecol* 21: 797-807.

Hood L, Rowen L (2013) The human genome project: big science transforms biology and medicine. *Gen Med* 5: 79-79.

Hosseinzadeh-Colagar A, Haghighatnia MJ, Amiri Z, Mohadjerani M, Tafrihi M (2016) Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: evidence which shows replication slippage occurs in extend or nascent DNA strands. *Mol Biol Res Commun* 5: 167-174.

Huang Y, Zhou N, Yang M, Shen Y, Zhang D (2019) A comparative study of the population genetics of wild and cultivated populations of *Paris polyphylla* var. *yunnanensis* based on amplified fragment length polymorphism markers. *Ecol Evol* 9: 10707-10722.

Huda MN, Hasan M, Abdullah HM, Sarker U (2018) Spatial distribution and genetic diversity of wild date palm (*Phoenix sylvestris*) growing in coastal Bangladesh. *Tree Genet Gen* 15: 3.

Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M (2008) Ecological consequences of genetic diversity. *Ecol Lett* 11: 609-623.

Hughes AR, Stachowicz JJ (2004) Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proc Natl Acad Sci USA* 101: 8998.

Idrees M, Irshad M (2014). Molecular markers in plants for analysis of genetic diversity: a review. *Eur Acad Res* 2(2): 1513-1540.

Jang W, Jang Y, Kim NH, Waminal NE, Kim YC, Lee JW, Yang TJ (2019) Genetic diversity among cultivated and wild *Panax ginseng* populations revealed by high resolution microsatellite markers. *J Ginseng Res.*

Jasim Aljumaili S, Rafii MY, Latif MA, Sakimin SZ, Arolu IW, Miah G (2018) Genetic diversity of aromatic rice germplasm revealed by SSR markers. *BioMed Res Int* 2018: 7658032.

Ji Y, Liu R, Hu J, Huang Y, Wang D, Li G, Rahman MM, Zhang H, Wang C, Li M, Yang T, Zong X (2020) Genetic diversity analysis for narrow-leafed lupin (*Lupinus angustifolius* L.) by SSR markers. *Mol Biol Rep:* 10.1007/s11033-020-05596-z.

Jo IH, Bang KH, Hong CE, Kim JU, Lee JW, Hyun DY, Ryu H, Kim DH (2016) Analysis of the chloroplast genome and SNP detection in a salt tolerant breeding line in Korean ginseng. *J Plant Biotechnol* 43:417-421.

Johnson MTJ, Lajeunesse MJ, Agrawal AA (2006) Additive and interactive effects of plant genotypic diversity on arthropod communities and plant fitness. *Ecol Lett* 9: 24-34.

Jump AS, Peñuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecol Lett* 8: 1010-1020.

Katoh Y, Katoh M, Takeda Y, Omori M (2003) Genetic diversity within cultivated teas based on nucleotide sequence comparison of ribosomal RNA

maturase in chloroplast DNA. *Euphytica* 134(3): 287-295.

Kaundun SS, Matsumoto S (2002) Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, *Camellia sinensis*. *Genome* 45(6): 1041-1048.

Kaundun S, Matsumoto S (2011) Molecular evidence for maternal inheritance of the chloroplast genome in tea, *Camellia sinensis* (L.) O. Kuntze. *J Sci Food Agric* 91(14): 2660–2663.

Kaur P, Gaikwad K (2017) From genomes to gene-omes: exome sequencing concept and applications in crop improvement. *Fron Plant Sci* 8: 2164-2164.

Kherwar D, Usha K, Mithra SVA, Singh B (2018) Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). *J Plant Biochem Biotechnol* 27: 284-292.

Kim S, Lee J, Baek S, Ahn J, Hong K (2018) Identification of DNA molecular markers by comparison of *Pinus densiflora* and *Pinus sylvestris* chloroplast genomes. *PeerJ* 6: e26506v1.

Komatsu K, Zhu S, Fushimi H, Qui TK, Cai S, Kadota S (2001) Phylogenetic analysis based on 18S rRNA gene and matK gene sequences of P*anax vietnamensis* and five related species. *Planta medica* 67: 461-465.

Le TH, Lee GJ, Vu HK, Kwon SW, Nguyen NK, Park JH, Nguyen MD (2015) Ginseng saponins in different parts of *Panax vietnamensis*. *Chem Pharm Bull* 63: 950-954.

Lee KJ, Lee JR, Sebastin R, Cho GT, Hyun DY (2020) Molecular genetic diversity and population structure of ginseng germplasm in RDA-genebank: Implications for breeding and conservation. *Agronomy* 10: 68.

Lee SR, Choi JE, Lee BY, Yu JN, Lim CE (2018) Genetic diversity and structure of an endangered medicinal herb: implications for conservation. *AoB Plants* 10.

Lerner HRL, Fleischer RC (2010) Prospects for the use of next-generation sequencing methods in ornithology. *The Auk* 127: 4-15.

Li L, Fang Z, Zhou J, Chen H, Hu Z, Gao L, Chen L, Ren S, Ma H, Lu L, Zhang W, Peng H (2017) An accurate and efficient method for large-scale SSR genotyping and applications. *Nucleic Acids Res.* 45: e88-e88.

Li TSC, Mazza G (1999) Correlations between leaf and soil mineral concentrations and ginsenoside contents in American ginseng. *HortSci* 34: 85-87.

Liao G, Li Z, Huang C, Zhong M, Tao J, Qu X, Chen L, Xu X (2019) Genetic diversity of inner quality and SSR association analysis of wild kiwifruit (*Actinidia eriantha*). *Scientia Horticulturae* 248: 241-247.

Lu J, Knox MR, Ambrose MJ, Brown JKM, Ellis THN (1996) Comparative analysis of genetic diversity in pea assessed by RFLP and PCR-based methods. *Theor Appl Genet* 93: 1103-1111.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7: 111-118.

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. *Int J Plant Genomics* 2012: 728398-728398.

Manzanilla V, Kool A, Nguyen Nhat L, Nong Van H, Le Thi Thu H, de Boer HJ (2018) Phylogenomics and barcoding of *Panax*: toward the identification of ginseng species. *BMC Evol Biol* 18: 44.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.

Martin L, Fei Z, Giovannoni J, Rose J (2013) Catalyzing plant science research with RNA-seq. *Fron Plant Sci* 4.

Matsumoto A, Tsumura Y (2004). Evaluation of cleaved amplified polymorphic sequence markers for *Chamaecyparis obtusa* based on expressed sequence tag information from *Cryptomeria japonica*. *Theor Appl Genet* 110(1): 80-91.

Maughan PJ, Yourstone SM, Jellen EN, Udall JA (2009) SNP discovery via genomic reduction,

barcoding, and 454-Pyrosequencing in Amaranth. *Plant Genome* 2: 260-270.

Mba C, Tohme J (2005) Use of AFLP markers in surveys of plant diversity. *Methods Enzymol* 395: 177-201.

Mendonça EG, de Souza AM, Vieira FdA, Estopa RA, Reis CAF, Carvalho Dd (2014) Using Random amplified polymorphic DNA to assess genetic diversity and structure of natural *Calophyllum brasiliense* (Clusiaceae) populations in Riparian forests. *Int J Forest Res* 2014: 305286.

Mertes F, Elsharawy A, Sauer S, van Helvoort JMLM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ (2011) Targeted enrichment of genomic DNA regions for next-generation sequencing. *Brief Funct Genomics* 10: 374-386.

Metzker ML (2005) Emerging technologies in DNA sequencing. *Genome Res.* 15.

Miah G, Rafii MY, Ismail MR, Puteh AB, Rahim HA, Islam KN, Latif MA (2013) A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *Int J Mol Sci* 14: 22499-22528.

Mneney E, Mantell S, Bennett M (2001) Use of random amplified polymorphic DNA (RAPD) markers to reveal genetic diversity within and between populations of cashew (*Anacardium occidentale* L.). *J Hort Sci Biotechnol* 76: 375-383.

Mondini L, Noorani A, Pagnotta MA (2009) Assessing plant genetic diversity by molecular tools. *Divers Distrib* 1: 19-35.

Nadeem MA, Nawaz MA, Shahid MQ, Doğan Y, Comertpay G, Yıldız M, Hatipoğlu R, Ahmad F, Alsaleh A, Labhane N, Özkan H, Chung G, Baloch FS (2018) DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol Biotechnol Equip* 32: 261-285.

Nawaz MA, Yang SH, Rehman HM, Baloch FS, Lee JD, Park JH, Chung G (2017) Genetic diversity and population structure of Korean wild soybean (*Glycine soja* Sieb. and Zucc.) inferred from microsatellite markers. *Biochem Syst Ecol* 71: 87-96.

Nevo E (2001) Evolution of genome-phenome diversity under environmental stress. *Proc Nat Acad Sci USA* 98: 6233-6240.

Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME (2007) Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* 4: 907-909.

Ozsolak F, Platt AR, Jones DR, Reifenberger JG, Sass LE, McInerney P, Thompson JF, Bowers J, Jarosz M, Milos PM (2009) Direct RNA sequencing. *Nature* 461: 814-818.

Paramanik RC, Chikkaswamy BK (2014) Study on genetic diversity relationship some medicinal plants using RAPD molecular marker. *Int J Curr Microbiol Appl Sci* 3: 1004-1015.

Paun O, Schönswetter P (2012) Amplified fragment length polymorphism: an invaluable fingerprinting technique for genomic, transcriptomic, and epigenetic studies. *Methods Mol Biol* 862: 75-87.

Petersen L, Østergård H, Giese H (1994) Genetic diversity among wild and cultivated barley as revealed by RFLP. *Theor Appl Genet* 89: 676-681.

Prakash S, Van-Staden J (2008) Genetic variability and species identification within encephalartos using random amplified polymorphic DNA (RAPD) markers. *S Afr J Bot* 74: 735-739.

Proctor JTA, Bailey WG (2011) Ginseng: Industry, botany, and culture. In: Janick J (Ed) *Horticultural Reviews*: 187-236.

Rahimi M, Kordrostami M, Safaeichaeikar S (2019) Genetic variation, population structure and the possibility of association mapping of biochemical and agronomic traits using dominant molecular markers in iranian tea accessions. *Iran J Sci Technol Trans A: Sci* 43: 2769-2780.

Rajeswara RBR, Rajput DK, Nagaraju G, Adinarayana G (2012) Scope and potential of medicinal and aromatic plants products for small and medium enterprises. *J Pharm* 3: 112-114.

Ramanatha RV, Hodgkin T (2002) Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tiss Organ Cult* 68: 1-19.

Ramos JR, Telles MP, Diniz-Filho JA, Soares TN, Melo DB, Oliveira G (2008) Optimizing reproducibility evaluation for random amplified polymorphic DNA markers. *Genet and Mol Res: GMR* 7: 1384-1391.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conserv Biol* 17: 230-237.

Reunova G, Koren O, Muzarok T, Zhuravlev Y (2014) Microsatellite analysis of *Panax ginseng* natural populations in Russia. *Chin Med* 5: 231-243.

Rothberg BEG, Rothberg JM (2015) Massively parallel ("next-generation") DNA sequencing. *Clin Chem* 61: 997-998.

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. *Mol Ecol* 20: 3499-3502.

Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94: 441-448.

Schaal BA, Olsen KM (2000) Gene genealogies and population variation in plants. *Proc Natl Acad Sci USA* 97(13): 7024–7029.

Schippmann U, Leaman D, Cunnigham A (2006) A comparison of cultivation and wildcollection of medicinal and aromatic plants under sustainable aspects. In: Rogers RJ, Craker LE, Lange D (Eds) *Medicinal and aromatic plants. Agricultural, commercial, ecological, legal,pharmacological and social aspects*. Dordrecht Springer. 17: 75-95.

Schmidhuber J, Tubiello FN (2007) Global food security under climate change. *Proc Nat Acad Sci USA* 104: 19703.

Schreiber M, Himmelbach A, Börner A, Mascher M (2019) Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evol Appl* 12: 66-77.

Shahrajabian MH, Sun W, Cheng Q (2019) A review of Ginseng species in different regions as a multipurpose herb in traditional Chinese medicine, modern herbology and pharmacological science. *J Med Plants Res* 13: 213-226.

Sharaf-Eldin MA, Alam P, Elkholy SF (2018) Molecular and chemical characterization of mutant and nonmutant genotypes of saffron grown in Saudi Arabia. *Food Sci Nutr* 7: 247-255.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-1145.

Shu Y, Li Y, Zhu Y, Zhu Z, Lv D, Bai X, Cai H, Ji W, Guo D (2010). Genome-wide identification of intron fragment insertion mutations and their potential use as SCAR molecular markers in the soybean. *Theor Appl Genet*, 121(1): 1-8.

Skuza L, Szućko I, Filip E, Strzała T (2019). Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. *PloS One*, 14(2): e0213023-e0213023.

Simner PJ, Khare R, Wengenack NL (2015) Chapter 95 - Rapidly growing Mycobacteria. In: Tang YW, Sussman M, Liu D, Poxton I and Schwartzman J (Eds) *Molecular Medical Microbiology (Second Edition)*. Boston, Academic Press**:** 1679-1690.

Singh BD, Singh AK (2015) Polymerase chain reaction-based markers. In: Singh BD and Singh AK (Eds) *Marker-assisted plant breeding: principles and practices*. New Delhi, Springer**:** 47-75.

Singh A, Dikshit HK, Singh D, Jain N, Aski M, Sarker A, Sharma TR (2016). Use of expressed sequence tag microsatellite markers for exploring genetic diversity in lentil and related wild species. *J Agricul Sci* 154(7): 1254-1269.

Singh HR, Hazarika P (2020) *4 - Biotechnological approaches for tea improvement*, in Grumezescu AM and Holban AM, (Eds) *Biotechnological progress and beverage consumption*, Academic Press, 111-148.

Song Y, Li MF, Xu J, Zhao Z, Chen NZ (2015) Polymorphic microsatellite markers in the traditional Chinese medicinal plant *Paris polyphylla* var. *yunnanensis. Genet Mol Res: GMR* 14: 9939-9942.

Sorkheh K, Dehkordi MK, Ercisli S, Hegedus A, Halász J (2017) Comparison of traditional and new generation DNA markers declares high genetic diversity and differentiated population structure of wild almond species. *Sci Rep* 7: 5966.

Sorof UM, Cheng Q (2015) 12 - Recent application of biotechniques for the improvement of mango research. In: Poltronieri P and Hong Y (Eds) *Applied Plant Genomics and Biotechnology*. Oxford, Woodhead Publishing**:** 195-212.

Srivastava N, Bajpai A, Chandra R, Rajan S, Muthukumar M, Srivastava MK (2012) Comparison of PCR based marker systems for genetic analysis in different cultivars of mango. *J Environ Biol* 33: 159-166.

Tabit FT (2016) Advantages and limitations of potential methods for the analysis of bacteria in milk: a review. *J Food Sci Technol* 53: 42-49.

Trieu LN, Mien NT, Van Tien T, Van Ket N, Van Duy N (2016) Genetic diversity of *Panax stipuleanatus*

Tsai in North Vietnam detected by inter simple sequence repeat (ISSR) markers. *Biotechnol Biotechnol Equip* 30: 506-511.

Uribe-Convers S, Settles ML, Tank DC (2016) A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of *Bartsia* L. (Orobanchaceae). *PloS One* 11: e0148203.

Vellend M (2006) The consequences of genetic diversity in competitive communities. *Ecol* 87: 304-311.

Vellend M, Geber MA (2005) Connections between species diversity and genetic diversity. *Ecol Lett* 8: 767-781.

Verma KS, Haq SU, Kachhwaha S, Kachhwaha S, Kothari SL (2017) RAPD and ISSR marker assessment of genetic diversity in *Citrullus colocynthis* (L.) Schrad: a unique source of germplasm highly adapted to drought and high-temperature stress. *3 Biotech* 7: 288-288.

Vieira MLC, Santini L, Diniz AL, Munhoz CdF (2016) Microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol* 39: 312-328.

Wang XM, Hou XQ, Zhang YQ, Yang R, Feng SF, Li Y, Ren Y (2012) Genetic diversity of the endemic and medicinally important plant *Rheum officinale* as revealed by inter-simpe sequence repeat (ISSR) markers. *Int J Mol Sci* 13: 3900-3915.

Wimp GM, Young WP, Woolbright SA, Martinsen GD, Keim P, Whitham TG (2004) Conserving plant genetic diversity for dependent animal communities. *Ecol Lett* 7: 776-780.

Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, Tang W, Dou Y, Zhang C, Xiao Y (2019) Development of high-density snp markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Fron Plant Sci* 10.

Xu JL, Wang Y, Hou M, Li Q (2015) Progress on detection methods of SNP. *Mol Plant Breed* 13: 475-482.

Yang HB, Kang WH, Nahm SH, Kang BC (2015) Methods for developing molecular markers. In: Koh H, Kwon S and Thomson M (Eds) *Current Technologies in Plant Molecular Breeding*. Dordrecht, Springer**:** 15-50.

Zhang GH, Ma CH, Zhang JJ, Chen JW, Tang QY, He MH, Xu XZ, Jiang NH, Yang SC (2015) Transcriptome analysis of *Panax vietnamensis* var. *fuscidicus* discovers putative ocotillol-type ginsenosides biosynthesis genes and genetic markers. *BMC Genomics* 16: 159.

Zhang Q, Li J, Zhao Y, Korban SS, Han Y (2012) Evaluation of genetic diversity in chinese wild apple species along with apple cultivars using SSR markers. *Plant Mol Biol Rep* 30: 539-546.

Zhang TT, Zhang NY, Li W, Zhou XJ, Pei XY, Liu YG, Ren ZY, He KL, Zhang WS, Zhou KH, Zhang F, Ma XF, Yang DG, Li ZH (2020) Genetic structure, gene flow pattern, and association analysis of superior germplasm resources in domesticated upland cotton (*Gossypium hirsutum* L.). *Plant Divers* 42: 189-197.

Zhou TH, Dong SS, Li S, Zhao GF (2012) Genetic structure within and among populations of *Saruma henryi,* an endangered plant endemic to China. *Biochem Genet* 50: 146-158.

Zhuravlev YN, Reunova GD, Kats IL, Muzarok TI, Bondar AA (2010) Genetic variability and population structure of endangered *Panax ginseng* in the Russian Primorye. *Chin Med* 5: 21.

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176-183.