Review Article

Importance of Genetic Diversity Assessment in Crop Plants and Its Recent Advances: An Overview of Its Analytical Perspectives

M. Govindaraj,[1,2] M. Vetriventhan,[1,2] and M. Srinivasan[3]

[1]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore 641 003, India
[2]International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Telangana 502324, India
[3]School of Life Science, Bharathidasan University, Tiruchirappalli 620 024, India

Correspondence should be addressed to M. Srinivasan; mahasiva2@gmail.com

The importance of plant genetic diversity (PGD) is now being recognized as a specific area since exploding population with urbanization and decreasing cultivable lands are the critical factors contributing to food insecurity in developing world. Agricultural scientists realized that PGD can be captured and stored in the form of plant genetic resources (PGR) such as gene bank, DNA library, and so forth, in the biorepository which preserve genetic material for long period. However, conserved PGR must be utilized for crop improvement in order to meet future global challenges in relation to food and nutritional security. This paper comprehensively reviews four important areas; (i) the significance of plant genetic diversity (PGD) and PGR especially on agriculturally important crops (mostly field crops); (ii) risk associated with narrowing the genetic base of current commercial cultivars and climate change; (iii) analysis of existing PGD analytical methods in pregenomic and genomic era; and (iv) modern tools available for PGD analysis in postgenomic era. This discussion benefits the plant scientist community in order to use the new methods and technology for better and rapid assessment, for utilization of germplasm from gene banks to their applied breeding programs. With the advent of new biotechnological techniques, this process of genetic manipulation is now being accelerated and carried out with more precision (neglecting environmental effects) and fast-track manner than the classical breeding techniques. It is also to note that gene banks look into several issues in order to improve levels of germplasm distribution and its utilization, duplication of plant identity, and access to database, for prebreeding activities. Since plant breeding research and cultivar development are integral components of improving food production, therefore, availability of and access to diverse genetic sources will ensure that the global food production network becomes more sustainable. The pros and cons of the basic and advanced statistical tools available for measuring genetic diversity are briefly discussed and their source links (mostly) were provided to get easy access; thus, it improves the understanding of tools and its practical applicability to the researchers.

1. Introduction

Diversity in plant genetic resources (PGR) provides opportunity for plant breeders to develop new and improved cultivars with desirable characteristics, which include both farmer-preferred traits (yield potential and large seed, etc.) and breeders preferred traits (pest and disease resistance and photosensitivity, etc.). From the very beginning of agriculture, natural genetic variability has been exploited within crop species to meet subsistence food requirement, and now it is being focused to surplus food for growing populations. In the middle of 1960s developing countries like India experienced the green revolution by meeting food demand with help of high-yielding and fertilizer responsive dwarf hybrids/varieties especially in wheat and rice (Figure 1). These prolonged activities that lead to the huge coverage of single genetic cultivars (boom) made situation again worse in other forms such as genetic erosion (loss of genetic diversity) and extinction of primitive and adaptive genes (loss of landraces). Today with an advancement of agricultural and allied science and technology, we still ask ourselves whether we can feed the world in 2050; this
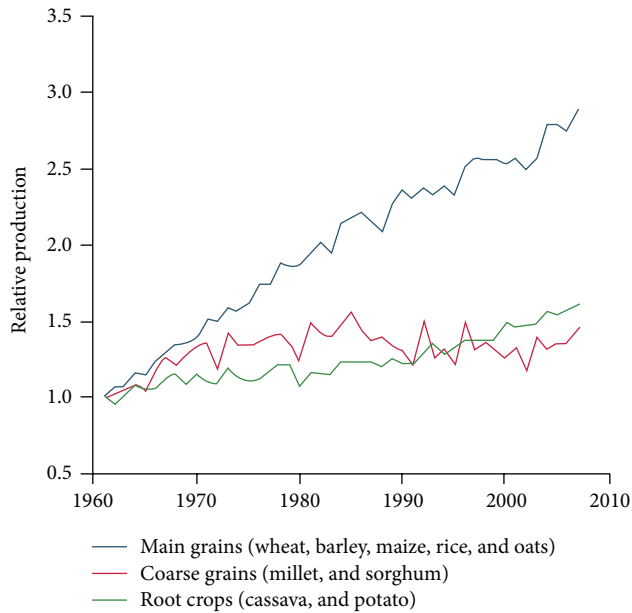
FIGURE 1: Changes in the relative global production of crops since 1961 (when relative production scaled to 1 (m.t) in 1961) (source: http://faostat.fao.org/default.aspx (2010)).

question was recently sensitized at the world food prize event in 2014 and remains that unanswered in every one hands since global population will exceed 9 billion in 2050. The per capita availability of food and water will become worse year after year coping with the undesirable climate change. Therefore, it becomes more important to look at the agriculture not only as a food-producing machine, but also as an important source of livelihood generation both in the farm and nonfarm sectors. Keeping the reservoir for cultivated and cultivable crops species is a principle for future agriculture, just like keeping a museum of cultural and spiritual specialty of diverse civilized humans in various geography for their historical evidence for future. The former can play a very important role in providing adaptive and productive genes, thus leading to long-term increases in food productivity which is further associated with environmental detriment. This paper will indicate the significance of genetic conservation and its analytical tools and techniques that are made widely available for utilization in postgenomic era. Plant and animal breeders introduced desirable genes and eliminated undesirable ones slowly, altering in the process of underlying heredity principle for several decades [1]. With the advent of new biotechnological tools and techniques, this process of genetic manipulation is being accelerated and it shortened the breeding cycles, and it can be carried out with more precision (neglecting environmental effects) and fast-track manner than the classical breeding techniques.

2. Significance of Genetic Conservation of Crop Plants

The growing population pressure and urbanization of agricultural lands and rapid modernization in every field of our

day-to-day activities that create biodiversity are getting too eroded in direct and indirect way. For instance, land degradation, deforestation, urbanization, coastal development, and environmental stress are collectively leading to large-scale extinction of plant species especially agriculturally important food crops. On the other hand, system driven famine such as, Irish potato famine and Southern corn leaf blight epidemic in USA are the two instances of food crises caused by large-scale cultivation of genetically homogenous varieties of potato and corn, respectively. Even after these historical events, the importance of PGR had only got popular recognition when the spread of green revolution across cultivated crops threatened the conservation of land races [2]. Green revolution technologies introduced improved crop varieties that have higher yields, and it was hoped that they would increase farmers' income. Consequently, the Consultative Group of International Agricultural Researches (CIGAR) initiated gene banks and research centers of domestication for conserving PGR in most of the stable food crops around the world. Center for domestication: maize (Mexico), wheat and barley (middle/near East and North Africa), rice (North China), and potatoes (Peru); for further information see http://www.cigar.org/center/index.html.) The Food and Agriculture Organization (FAO) supported the International Treaty on Plant Genetic Resources (ITPGR) and UN supported the Convention on Biological Diversity (CBD) which are the international agreements that recognize the important role of genetic diversity conservation. Such treaty still plays in current and future food production as one of the major supremo [3].

Genetic diversity is the key pillar of biodiversity and diversity within species, between species, and of ecosystems (CBD, Article 2), which was defined at the Rio de Janeiro Earth Summit. However, the problem is that modern crop varieties, especially, have been developed primarily for high yielding potential under well endowed production conditions. Such varieties are often not suitable for low income farmers in marginal production environments as they are facing highly variable stress conditions [4]. Land races or traditional varieties have been found to have higher stability (adaptation over time) in low-input agriculture under marginal environments, thus, their cultivation may contribute farm level resilience in face of food production shocks [5, 6]. This is especially true in some part of Ethiopia where agroclimatic conditions are challenging, technological progress is slow, and market institutions are poorly developed and have no appropriate infrastructure [7, 8].

Why is genetic diversity important? The goal of conservation genetics is to maintain genetic diversity at many levels and to provide tools for population monitoring and assessment that can be used for conservation planning. Every individual is genetically unique by nature. Conservation efforts and related research are rarely directed towards individuals but genetic variation is always measured in individuals and this can only be estimated for collections of individuals in a population/species. It is possible to identify the genetic variation from phenotypic variation either by quantitative traits (traits that vary continuous and are governed by many genes, e.g., plant height) or discrete traits traits that fall into

discrete categories and are governed by one or few major genes (e.g., white, pink, or red petal color in certain flowers) which are referred to as qualitative traits. Genetic variation can also be identified by examining variation at the level of enzymes using the process of protein electrophoresis. Further, genetic variations can also be examined by the order of nucleotides in the DNA sequence.

3. Erosion of Genetic Diversity due to Population Size: A Bottleneck Concept

It is well known that inbreeding is the most common phenomena in cross-pollinated crops, and in small outcross populations it has resulted in deleterious effects and loss of fitness of the population due to recombination between undesirable genes (recessive identical alleles). In natural population too, severe reductions in population size, the so-called *genetic bottleneck*, leads to loss of genetic diversity and increased susceptibility to infectious pests and diseases that supervene increased chances of extinction of an individual crop in question. Genetic models that predict the proportion of initial heterozygosity retained per generation is $[1 - (1/2N_e)]$ where N_e is the effective population size, usually less than N, the actual population size. Thus a population of $N_e = 10$ individuals loses 5% of its heterozygosity per generation. This indicates that severe bottlenecks degrade heterozygosity and genetic diversity [9]. Therefore, plant breeders have been advised to maintain the optimum population size for any trait conservation for specific purpose and its utilization for crop improvement. Thus, before quantifying the genetic diversity, it is essential to know the optimum population size and its representatives to ensure no biasness in diversity assessment that leads to wrong prediction of its value.

4. Climate Change and Its Impact on Plant Genetic Resources

The most profound and direct impacts of climate change over previous decade and the next few decades will surely be on agriculture and food security. The effects of climate change will also depend on current production conditions. The area where already being obstructed by other stresses, such as pollution and will likely to have more adverse impact by changing climate. Food production systems rely on highly selected cultivars under better endowed environments but it might be increasingly vulnerable to climate change impacts such as pest and disease spread. If food production levels decreases over the year, there will be huge pressure to cultivate the crops under marginal lands or implement unsustainable practices that, over the long-term, degrade lands and resources and adversely impact biodiversity on and near agricultural areas. In fact, such situations have already been experienced by most of the developing countries. These changes have been seen to cause a decrease in the variability of those genetic loci (alleles of a gene) controlling physical and phenotypic responses to changing climate [10]. Therefore, genetic variation holds the key to the ability of populations and species to persist over evolutionary period of time through changing environments

[11]. If this persists, neither any organism can predict its future (and evolutionary theory does not require them to) nor can any of those organisms be optimally adapted for all environmental conditions. Nonetheless, the current genetic composition of a crop species influences how well its members will adapt to future physical and biotic environments.

The population can also migrate across the landscape over generations. By contrast, populations that have a narrow range of genotypes and are more phenotypically uniform may merely fail to survive and reproduce at all as the conditions become less locally favorable. Such populations are more likely to become extirpated (locally extinct), and in extreme cases the entire plant species may end up at risk of extinction. For example, the Florida Yew (*Torreya taxifolia*) is currently one of the rarest conifer species in North America. But in the early Holocene (10,000 years ago), when conditions in southeastern North America were cooler and wetter than today, the species was probably widespread. The reasons for that are not completely understood, but *T. taxifolia* failed to migrate towards the northward as climate changed during the Holocene. Today, it is restricted to a few locations in the Apalachicola River Basin in southern Georgia and the Florida panhandle. As the *T. taxifolia* story illustrates, once plant species are pushed into marginal habitat at the limitations of their physiological tolerance, they may enter an extinction vortex, a downward cycle of small populations, and so on [12, 13]. Reduced genetic variability is a key step in the extinction vortex. Gene banks must be better to respond to novel and increased demands on germplasm for adapting agriculture to climate change. Gene banks need to include different characteristics in their screening processes and their collections need to be comprehensive, including what are now considered minor crops, and that may come with huge impact on food baskets.

5. Assessment of Genetic Diversity in Crop Plants

The assessment of genetic diversity within and between plant populations is routinely performed using various techniques such as (i) morphological, (ii) biochemical characterization/evaluation (allozyme), in the pregenomic era, and (iii) DNA (or molecular) marker analysis especially single nucleotide polymorphism (SNPs) in postgenomic era. Markers can exhibit similar modes of inheritance, as we observe for any other traits, that is, dominant/recessive or codominant. If the genetic pattern of homozygotes can be distinguished from that of heterozygotes, then a marker is said to be codominant. Generally codominant markers are more informative than the dominant markers.

Morphological markers are based on visually accessible traits such as flower color, seed shape, growth habits, and pigmentation, and it does not require expensive technology but large tracts of land area are often required for these field experiments, making it possibly more expensive than molecular assessment in western (developed) countries and equally expensive in Asian and Middle East (developing) countries considering the labour cost and availability. These marker

traits are often susceptible to phenotypic plasticity; conversely, this allows assessment of diversity in the presence of environmental variation which cannot be neglected from the genotypic variation. These types of markers are still having advantage and they are mandatory for distinguishing the adult plants from their genetic contamination in the field, for example, spiny seeds, bristled panicle, and flower/leaf color variants.

Second type of genetic marker is called biochemical markers, allelic variants of enzymes called isozymes that are detected by electrophoresis and specific staining. Isozyme markers are codominant in nature. They detect diversity at functional gene level and have simple inheritance. It requires only small amounts of plant material for its detection. However, only a limited number of enzymes markers are available and these enzymes are not alone but it has complex structural and special problems; thus, the resolution of genetic diversity is limited to explore.

The third and most widely used genetic marker type is molecular markers, comprising a large variety of DNA molecular markers, which can be employed for analysis of genetic and molecular variation. These markers can detect the variation that arises from deletion, duplication, inversion, and/or insertion in the chromosomes. Such markers themselves do not affect the phenotype of the traits of interest because they are located only near or linked to genes controlling the traits. These markers are inherited both in dominant and codominant patterns. Different markers have different genetic qualities (they can be dominant or codominant, can amplify anonymous or characterized loci, can contain expressed or nonexpressed sequences, etc.). A molecular marker can be defined as a genomic locus, detected through probe or specific starter (primer) which, in virtue of its presence, distinguishes unequivocally the chromosomic trait which it represents as well as the flanking regions at the $3'$ and $5'$ extremity [14]. Molecular markers may or may not correlate with phenotypic expression of a genomic trait. They offer numerous advantages over conventional, phenotype-based alternatives as they are stable and detectable in all tissues regardless of growth, differentiation, development, or defense status of the cell. Additionally, they are not confounded by environmental, pleiotropic, and epistatic effects. We are not describing much about the pregenomic era tools, since our paper deals with genomic advances and its assistance in crop genetic diversity assessment.

6. Analyses of Genetic Diversity in Genomic Era

A comprehensive study of the molecular genetic variation present in germplasm would be useful for determining whether morphologically based taxonomic classifications reveal patterns of genomic differentiation. This can also provide information on the population structure, allelic richness, and diversity parameters of germplasm to help breeders to use genetic resources with less prebreeding activities for cultivar development more effectively. Now germplasm characterization based on molecular markers has gained importance due

to the speedy and quality of data generated. For the readers benefit, the availability of different DNA markers acronyms is given in Abbreviations section.

6.1. Molecular Markers. DNA (or molecular) markers are the most widely used type of marker predominantly due to their abundance. They arise from different classes of DNA mutations such as substitution mutations (point mutations), rearrangements (insertions or deletions), or errors in replication of tandemly repeated DNA [15]. These markers are selectively neutral because they are usually located in noncoding regions of DNA in a chromosome. Unlike other markers, DNA markers are unlimited in number and are not affected by environmental factors and/or the developmental stage of the plant [16]. DNA markers have numerous applications in plant breeding such as (i) marker assisted evaluation of breeding materials like assessing the level of genetic diversity, parental selection, cultivar identity and assessment of cultivar purity [16–26], study of heterosis, and identification of genomic regions under selection, (ii) marker assisted backcrossing, and (iii) marker assisted pyramiding [27].

Molecular markers may be broadly divided into three classes based on the method of their detection: hybridization-based, polymerase chain reaction- (PCR-) based, and DNA sequence-based. Restriction fragment length polymorphisms (RFLPs) are hybridization-based markers developed first in human-based genetic study during 1980s [28, 29] and later they were used in plant research [30]. RFLP is based on the variation(s) in the length of DNA fragments produced by a digestion of genomic DNAs and hybridization to specific markers of two or more individuals of a species is compared. RFLPs have been used extensively to compare genomes in the major cereal families such as rye, wheat, maize, sorghum, barley, and rice [31–33]. The advantages of RFLPs include detecting unlimited number of loci and being codominant, robust, and reliable and results are transferable across populations. However, RFLPs are highly expensive, time consuming, labour intensive, larger amounts of DNA required, limited polymorphism especially in closely related lines [34]. At present polymerase chain reaction- (PCR-) based marker systems are more rapid and require less plant material for DNA extraction. Rapid amplified polymorphic DNAs (RAPDs) were the first of PCR-based markers and are produced by PCR machines using genomic DNA and arbitrary (random) primers which act as both forward and backward primers in creation of multiple copies of DNA strands [35, 36]. The advantages of RAPDs include being quick and simple and inexpensive and the facts that multiple loci from a single primer are possible and a small amount of DNA is required. However, the results from RAPDs may not be reproduced in different laboratories and only can detect the dominant traits of interest [34]. Amplified fragment length polymorphisms (AFLPs) combine both PCR and RFLP [37]. AFLP is generated by digestion of PCR amplified fragments using specific restriction enzymes that cut DNA at or near specific recognition site in nucleotide sequence. AFLPs are highly reproducible and this enables rapid generation and high frequency of identifiable AFLPs, making it an attractive technique for identifying polymorphisms and for determining linkages by

analyzing individuals from a segregating population [37]. Another class of molecular markers which depends on the availability of short oligonucleotide repeats sequences in the genome of plants such as SSR, STS, SCAR, EST-SSR, and SNP. Many authors reviewed in detail different markers techniques [38, 39]. In this paper we are presenting the most widely used molecular markers and next generation sequencing technologies in detail in the following section.

6.2. Simple Sequence Repeat or Microsatellite. Microsatellites [40] are also known as simple sequence repeats (SSRs), short tandem repeats (STRs), or simple sequence length polymorphisms (SSLPs) which are short tandem repeats, their length being 1 to 10 bp. Some of the literatures define microsatellites as 2–8 bp [41], 1–6 bp [42], or even 1–5 pb repeats [43]. SSRs are highly variable and evenly distributed throughout the genome and common in eukaryotes, their number of repeated units varying widely among crop species. The repeated sequence is often simple, consisting of two, three, or four nucleotides (di-, tri-, and tetranucleotide repeats, resp.). One common example of a microsatellite is a dinucleotide repeat $(CA)n$, where n refers to the total number of repeats that ranges between 10 and 100. These markers often present high levels of inter- and intraspecific polymorphism, particularly when tandem repeats number is 10 or greater [44]. PCR reactions for SSRs are performed in the presence of forward and reverse primers that anneal at the $5'$ and $3'$ ends of the template DNA, respectively. These polymorphisms are identified by constructing PCR primers for the DNA flanking the microsatellite region. The flanking regions tend to be conserved within the species, although sometimes they may also be conserved in higher taxonomic levels.

PCR fragments are usually separated on polyacrylamide gels in combination with $AgNO_3$ staining, autoradiography, or fluorescent detection systems. Agarose gels (usually 3%) with ethidium bromide (EBr) can also be used when differences in allele size among samples are larger than 10 bp. However, the establishment of microsatellite primers from scratch for a new species presents a considerable technical challenge. Several protocols have been developed [43, 45–47] and details of the methodologies are reviewed by many authors [48–50]. The loci identified are usually multiallelic and codominant. Bands can be scored either in a codominant or as present or absent. The microsatellite-derived primers can often be used with many varieties and even other species because the flanking DNA is more likely to be conserved. These required markers are evenly distributed throughout the genome, easily automated, and highly polymorphic and have good analytic resolution and high reproducibility making them a preferred choice of markers [51], most widely used for individual genotyping, germplasm evaluation, genetic diversity studies, genome mapping, and phylogenetic and evolutionary studies. However, the development of microsatellites requires extensive knowledge of DNA sequences, and sometimes they underestimate genetic structure measurements; hence they have been developed primarily for agricultural species, rather than wild species [39].

6.3. EST-SSRs. An alternative source of SSRs development is development of expressed sequence tag- (EST-) based SSRs using EST databases has been utilized [52–58]. With the availability of large numbers of ESTs and other DNA sequence data, development of EST-based SSR markers through data mining has become fast, efficient, and relatively inexpensive compared with the development of genomic SSRs [59]. This is due to the fact that the time-consuming and expensive processes of generating genomic libraries and sequencing of large numbers of clones for finding the SSR containing DNA regions are not needed in this approach [60]. However, the development of EST-SSRs is limited to species for which this type of database exists as well as being reported to have lower rate of polymorphism compared to the SSR markers derived from genomic libraries [61–64].

6.4. Single Nucleotide Polymorphisms (SNPs). Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is changed, that is, single nucleotide variations in genome sequence of individuals of a population. These polymorphisms are single-base substitutions between sequences. SNPs occur more frequently than any other type of markers and are very near to or even within the gene of interest. SNPs are the most abundant in the genomes of the majority of organisms, including plants, and are widely dispersed throughout genomes with a variable distribution among species. SNPs can be identified by using either microarrays or DHPLC (denaturing high-performance liquid chromatography) machines. They are used for a wide range of purposes, including rapid identification of crop cultivars and construction of ultrahigh-density genetic maps. They provide valuable markers for the study of agronomic or adaptive traits in plant species, using strategies based on genetic mapping or association genetics studies.

6.5. Diversity Arrays Technology (DArT). A DArT marker is a segment of genomic DNA, the presence of which is polymorphic in a defined genomic representation. A DArT was developed to provide a practical and cost-effective whole genome fingerprinting tool. This method provides high throughput and low cost data production. It is independent from DNA sequence; that is, the discovery of polymorphic DArT markers and their scoring in subsequent analysis does not require any DNA sequence data. The detail of methodology for DArT is described by Jaccoud et al. [65] and Semagn et al. [38] as well as in website http://www.diversityarrays.com/.

To identify the polymorphic markers, a complexity reduction method is applied on the metagenome, a pool of genomes representing the germplasm of interest. The genomic representation obtained from this pool is then cloned and individual inserts are arrayed on a microarray resulting in a "discovery array." Labelled genomic representations prepared from the individual genomes included in the pool are hybridized to the discovery array. Polymorphic clones (DArT markers) show variable hybridization signal intensities for different individuals. These clones are subsequently assembled into a "genotyping array" for routine genotyping. DArT is one of

the recently developed molecular techniques and it has been used in rice [66], wheat [38, 67, 68], barley [69], eucalyptus [70], *Arabidopsis* [71], cassava [72], pigeon-pea [73], and so forth.

DArT markers can be used as any other genetic marker. With DArT, comprehensive genome profiles are becoming affordable regardless of the molecular information available for the crop. DArT genome profiles are very useful for characterization of germplasm collections, QTL mapping, reliable and precise phenotyping, and so forth. However, DArT technique involves several steps, including preparation of genomic representation for the target species, cloning, data management, and analysis, requiring dedicated software such as DArTsoft and DArTdb. DArT markers are primarily dominant (present or absent) or differences in intensity, which limits its value in some application [38].

7. Next Generation 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. DNA sequencing technology has played a pivotal role in the advancement of molecular biology [74]. Next generation sequencing (NGS) or second generation sequencing technologies are revolutionizing the study of variation among individuals in a population. Most NGS technologies reduce the cost and time required for sequencing than Sanger-style sequencing machines (first generation sequencing). The following is the list of NGS technologies available at present, namely, the Roche/454 FLX, the Illumina/Solexa Genome Analyzer, the Applied Biosystems SOLiD System, the Helicos single-molecule sequencing, and pacific Biosciences SMRT instruments. 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.

The Roche/454 FLX, based on sequencing-by-synthesis with pyrophosphate chemistry, was developed by 454 Life Sciences and was the first next generation sequencing platform available on the market [75]. The Solexa sequencing platform was commercialized in 2006. The working principle is sequencing-by-synthesis chemistry. The Life Technologies SOLiD system is based on a sequencing-by-ligation technology. This platform has its origins in the system described by Shendure et al. [76] and in work by McKernan et al. [77] at Agencourt Personal Genomics (acquired by Applied Biosystems in 2006). Helicos true single molecule sequencing (tSMS) technology is an entirely novel approach to DNA sequencing and genetic analysis and offers significant advantages over both traditional and "next generation" sequencing technologies. Helicos offers the first universal genetic analysis platform that does not require amplification. Pursuing a single molecule sequencing strategy simplifies the DNA sample preparation process, avoids PCR-induced bias and errors, simplifies data analysis, and tolerates degraded samples. Helicos single-molecule sequencing is often referred to as third generation sequencing. The detailed methodology, advantages, and disadvantages of each NGS technology were reviewed by many authors [78–81].

8. Analysis of Genetic Diversity from Molecular Data

It is essential to know the different ways that the data generated by molecular techniques can be analyzed before their application to diversity studies. Two main types of analysis are generally followed: (i) analysis of genetic relationships among samples and (ii) calculation of population genetics parameters (in particular diversity and its partitioning at different levels). The analysis of genetic relationships among samples starts with the construction of a matrix, sample × sample pair-wise genetic distance (or similarities).

The advent and explorations of molecular genetics led to a better definition of Euclidean distance to mean a quantitative measure of genetic difference calculated between individuals, populations, or species at DNA sequence level or allele frequency level. Genetic distance and/or similarity between two genotypes, populations, or individuals may be calculated by various statistical measures depending on the data set. The commonly used measures of genetic distance (GD) or genetic similarity (GS) are (i) Nei and Li's [82] coefficient (GD_{NL}), (ii) Jaccard's [83] coefficient (GD_J), (iii) simple matching coefficient (GD_{SM}) [84], and (iv) modified Rogers' distance (GD_{MR}). Genetic distance determined by the above measures can be estimated as follows:

$$
\begin{aligned}
GD_{NL} &= 1 = \left[\frac{2N_{11}}{(2N_{11} + N_{10} + N_{01})} \right], \\
GD_J &= 1 = \left[\frac{N_{11}}{(N_{11} + N_{10} + N_{01})} \right], \\
GD_{SM} &= 1 = \left[\frac{(N_{11} + N_{00})}{(N_{11} + N_{10} + N_{01} + N_{00})} \right], \\
GD_{MR} &= \left[\frac{(N_{10} + N_{01})}{2N} \right]^{0.5},
\end{aligned}
\tag{1}
$$

where N_{11} is the number of bands/alleles present in both individuals; N_{00} is number of bands/alleles absent in both individuals; N_{10} is the number of bands/alleles present only in the individual i; N_{01} is the number of bands/alleles present only in the individual j; and N represents the total number of bands/alleles. Readers are requested to read Mohammadi and Prasanna [85] review paper for more details about different GD measures.

There are two main ways of analyzing the resulting distance (or similarity) matrix, namely, principal coordinate analysis (PCA) and dendrogram (or clustering, tree diagram). PCA is used to produce a 2 or 3 dimensional scatter plot of the samples such that the distances among the samples in the plot reflect the genetic distances among them with a minimum of distortion. Another approach is to produce a dendrogram (or tree diagram), that is, grouping of samples together in clusters that are more genetically similar to each other than to samples in other clusters. Different algorithms were used for clustering, but some of the more widely used ones include unweighted pair group method with arithmetic

TABLE 1: Some basic statistical concept on genomic data for genetic diversity assessment.

Concept terms	Description/features	Formulae/pros/cons
Band-based approaches	Easiest way to analyze and measure diversity by focusing on presence or absence of banding pattern.	Routinely use individual level. Totally relay on marker type and polymorphism
(1) Measuring polymorphism	Observing the total number of polymorphic bands (PB) and then calculating the percentage of polymorphic bands.	This "band informativeness" (Ib) can be represented on a scale ranging from 0 to 1 according to the formula Ib $= 1 - (2 \times \|0.5 - p\|)$, where p is the portion of genotypes containing the band.
(2) Shannon's information index (I)	It is called the Shannon index of phenotypic diversity and is widely applied.	$I = - \sum p_i \log_2 p_i$. These methods depend on the extraction of allelic frequencies.
(3) Similarity coefficients	Utilize similarity or dissimilarity (the inverse of the previous one) coefficients. The Jaccard coefficient (J) only takes into account the bands present in at least one of the two individuals. It is therefore unaffected by homoplasic absent bands (where the absence of the same band is due to different mutations). The simple-matching index (SM) maximizes the amount of information provided by the banding patterns considering all scored loci. The Neil and Li index (SD) doubles the weight for bands present in both individuals, thus giving more attention to similarity than dissimilarity.	(i) **Jaccard similarity coefficient** or Jaccard index $J = a/(a + b + c)$. (ii) **Simple matching coefficient** or index SM $= (n - b - c)/n$. (iii) **Sørensen-Dice index** or Nei and Li index SD $= 2a/2a + b + c$ where a is the number of bands (1 s) shared by both individuals; b is the number of positions where individual i has a band, but j does not; c is the number of positions where individual j has a band, but i does not; and n is the total number of bands (0 s and 1 s).
(4) Allele frequency based approaches	Measure variability by describing changes in allele frequencies for a particular trait over time, more population oriented than band-based approaches.	These methods depend on the extraction of allelic frequencies from the data. The accurate estimates of frequencies essentially influence the results of different indices calculated for further measurements of genetic diversity.
(5) Allelic diversity (A)	Easiest ways to measure genetic diversity is to quantify the number of alleles present. Allelic diversity (A) is the average number of alleles per locus and is used to describe genetic diversity.	$A = n_i/n_l$ where n_i is the total number of alleles over all loci; n_l is the number of loci. It is less sensitive to sample size and rare alleles and is calculated as $n_e = 1/ \sum p_i^2$ p_i^2 ability; it provides information about the dispersal ability of the organism and the degree of isolation among populations.
(6) Effective population size (N_e)	It provides a measure of the rate of genetic drift, the rate of genetic diversity loss, and increase of inbreeding within a population.	Effective size of a population is an idealized number, since many calculations depend on the genetic parameters used and on the reference generation. Thus, a single population may have many different effective sizes which are biologically meaningful but distinct from each other.
(7) Heterozygosity (H)	There are two types of heterozygosity observed (H_O) and expected (H_E). The H_O is the portion of genes that are heterozygous in a population and H_E is estimated fraction of all individuals that would be heterozygous for any randomly chosen locus. Typically values for H_E and H_O range from 0 (no heterozygosity) to nearly 1 (a large number of equally frequent alleles). If H_O and H_E are similar (they do not differ significantly), mating in the populations is random. If $H_O < H_E$, the population is inbreeding; if $H_O > H_E$, the population has a mating system avoiding inbreeding.	Expected H_E is calculated based on the square root of the frequency of the null (recessive) allele as follows: $H_E = 1 - \sum_i^n p_i^2$ where p_i is the frequency of the ith allele. H_O is calculated for each locus as the total number of heterozygotes divided by sample size.

TABLE 1: Continued.

Concept terms	Description/features	Formulae/pros/cons
(8) F-statistics	In population genetics the most widely applied measurements besides heterozygosity are F-statistics, or fixation indices, to measure the amount of allelic fixation by genetic drift. The F-statistics are related to heterozygosity and genetic drift. Since inbreeding increases the frequency of homozygotes, as a consequence, it decreases the frequency of heterozygotes and genetic diversity.	Three indexes can be calculated as follows: $F_{IT} = 1 - (H_I/H_T)$, $F_{IS} = 1 - (H_I/H_S)$, $F_{ST} = 1 - (H_S/H_T)$, where H_I is the average H_O within each population, H_S is the average H_E of subpopulations assuming random mating within each population, and H_T is the H_E of the total population assuming random mating within subpopulations and no divergence of allele frequencies among subpopulations.

averages (UPGMA), neighbour-joining method, and Ward's method [86].

The molecular data can be scored in presence/absence matrices manually or with the aid of specific software. However, because these techniques are based on the incorporation of genomic elements in the primer sets or else target specific regions in the genome, biases affecting the evaluation process can occur. Although many recently developed targeting methods detect large numbers of polymorphisms, not many studies to date have utilized them, largely due to their unfamiliarity. In many cases the drawbacks are unknown. These mainly affect the analysis of the banding patterns produced, largely depending on the nature of the methods and whether they generate dominant or codominant markers. We presented a brief description of common/basic statistical approaches and its principle with the pros and cons of each method for measuring genetic diversity and it is given in Table 1. These are self-explanatory; therefore, the features and method of calculations were not much discussed separately in our text.

9. Assessment of Genetic Diversity in Postgenomic Era

Many software programs are available for assessing genetic diversity; however, most of them are freely available through source link to internet and corresponding institute web links are given in Table 2. In this section, we described some of the programs available which are mostly used in molecular diversity analyses in the postgenomic era (Table 2). Many of these perform similar tasks, with the main differences being in the user interface, type of data input and output, and platform. Thus, choosing which to use depends heavily on individual preferences.

10. Conclusion

Agriculturist has been realized that diverse plant genetic resources are priceless assets for humankind which cannot be lost. Such materials increasingly required to accessible for feeding a burgeoning world population in future (>9 billion in 2050). Presence of genetic variability in crops is essential for its further improvement by providing options for the breeders to develop new varieties and hybrids. This can be achieved through phenotypic and molecular characterization of PGR. Sometimes, large size of germplasm may limit their use in breeding. This may be overcome by developing and using subsets like core and minicore collection representing the diversity of the entire collection of the species. 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. Now with the high throughput molecular marker technologies ensuring speed and quality of data generated, it is possible to characterize the larger number of germplasm with limited time and resources. Next generation sequencing reduced the cost and time required for sequencing the whole genome. Many software packages are available for assessing phenotypic and molecular diversity parameters that increased the efficiency of germplasm curators and, plant breeders to speed up the crop improvement. Therefore, we believe that this paper provides useful and contemporary information at one place; thus, it improves the understanding of tools for graduate students and also practical applicability to the researchers.

Abbreviations

AFLP: Amplified fragment length polymorphism
AP-PCR: Arbitrarily primed PCR
ARMS: Amplification refractory mutation system
ASAP: Arbitrary signatures from amplification
ASH: Allele-specific hybridization
ASLP: Amplified sequence length polymorphism
ASO: Allele specific oligonucleotide
CAPS: Cleaved amplification polymorphic sequence
CAS: Coupled amplification and sequencing
DAF: DNA amplification fingerprint
DGGE: Denaturing gradient gel electrophoresis
GBA: Genetic bit analysis
IRAO: Interretrotransposon amplified polymorphism
ISSR: Intersimple sequence repeats
ISTR: Inverse sequence-tagged repeats
MP-PCR: Microsatellite-primed PCR
OLA: Oligonucleotide ligation assay
RAHM: Randomly amplified hybridizing microsatellites
RAMPs: Randomly amplified microsatellite polymorphisms

TABLE 2: List of analytical programs for measuring molecular (genetic) diversity.

Analytical tools	Data type	Main features	Source links	Reference
Arlequin	RFLPs, DNA sequences, SSR data, allele frequencies, or standard multilocus genotypes.	(i) Estimation allele and haplotype frequencies. (ii) Tests of departure from linkage equilibrium, departure from selective neutrality and demographic equilibrium. (iii) Estimation or parameters from past population expansions. (iv) Thorough analyses of population subdivision under the AMOVA framework and so forth. (v) Current version: Arlequin ver 3.5.1.3.	http://cmpg.unibe.ch/software/arlequin3	Schneider et al. [87] Excoffier et al. [88]
DnaSP	DNA sequence data	(i) Estimating several measures of DNA sequence variation within and between populations (in noncoding, synonymous, or nonsynonymous sites or in various sorts of codon positions), as well as linkage disequilibrium, recombination, gene flow, and gene conversion parameters. (ii) DnaSP can also carry out several tests of neutrality: Hudson et al. [89], Tajima [90], McDonald and Kreitman [46], Fu and Li [91], and Fu [92] tests. Additionally, DnaSP can estimate the confidence intervals of some test-statistics by the coalescent and so forth. (iii) Current version: DnaSP v5.10.01.	http://www.ub.edu/dnasp	J. Rozas and R. Rozas, [93–95] Librado and Rozas [96]
PowerMarker	SSR, SNP, and RFLP data	(i) Computes several summary statistics for each marker locus, including allele number, missing proportion, heterozygosity, gene diversity, polymorphism information content (PIC), and stepwise patterns for microsatellite data. (ii) PowerMarker is also used to compute allele frequency, genotype frequency, haplotype frequency for unrelated individuals, Hardy-Weinberg equilibrium, pairwise linkage disequilibrium, multilocus linkage disequilibrium, consensus trees, population structure, Mantel's test, triangle plotting and visualization of linkage disequilibrium results. (iii) Current version: PowerMarker V3.25.	http://statgen.ncsu.edu/powermarker/	Liu and Muse [97]
DARwin	Single data (for haploids, homozygote diploids, and dominant markers), allelic data, and sequence data	(i) Most widely used for various dissimilarity and distance estimations for different data, tree construction methods including hierarchical trees with various aggregation criteria (weighted or unweighted), Neighbor-Joining tree (weighted or unweighted), Scores method and principal coordinate analysis, and so forth. (ii) Current version: DARwin v5.0.156.	http://darwin.cirad.fr/darwin	Perrier and Jacquemoud-Collet [98]
NTSYSpc	Single data (for haploids, homozygote diploids, and dominant markers), allelic data, and sequence data	(i) Used for clustering analysis, ordination analysis, principal component analysis, principal coordinate analysis, scaling analysis, and comparison of two matrices (Mantel test, Mantel [99] and so forth). (ii) Current version: NTSYSpc version 2.2.	http://www.exetersoftware.com/cat/ntsyspc/ntsyspc.html	Rohlf [100]
MEGA	DNA sequence, protein sequence, evolutionary distance, or phylogenetic tree data	(i) Molecular evolutionary genetics analysis (MEGA) is most widely used for aligning sequences, estimating evolutionary distances, building tree from sequence data, testing tree reliability, and so forth. (ii) Current version: MEGA6.	http://www.megasoftware.net	Kumar et al. [101–103] Tamura et al. [104]
PAUP	Molecular sequences, morphological data, and other data types	(i) Used for inferring and interpreting phylogenetic trees using parsimony, distance matrix, invariants, maximum likelihood methods, and many indices and statistical analyses. (ii) Current version: PAUP version 4.0.	http://paup.csit.fsu.edu/	Swofford [105]

TABLE 2: Continued.

Analytical tools	Data type	Main features	Source links	Reference
STRUCTURE	All types of markers including mostly used markers like SSRs, SNPs, RFLPs, dArT, and so forth.	(i) A free program to investigate population structure; it includes inferring the presence of distinct populations, assigning individuals to populations, studying hybrid zones, identifying migrants and admixed individuals, and estimating population allele frequencies in situations where many individuals are migrants or admixed. (ii) Current version: STRUCTURE 2.3.4.	http://pritch.bsd.uchicago.edu/software/structure2_2.html	Pritchard et al. [106] Falush et al. [107] Hubisz et al. [108]
fastSTRUCTURE	SNP	(i) An algorithm for inferring population structure from large SNP genotype data. (ii) It is based on a variational Bayesian framework for posterior inference and is written in Python2.x.	http://rajanil.github.io/fastStructure/	Raj et al. [109]
ADMIXTURE	SNP	(i) ADMIXTURE is a program for maximum likelihood estimation of individual ancestries from multilocus SNP genotype datasets. (ii) It uses the same statistical model as STRUCTURE but calculates estimates much more rapidly using a fast numerical optimization algorithm. (iii) Current version: ADMIXTURE 1.23.	https://www.genetics.ucla.edu/software/admixture/	Alexander et al. [110]
fineSTRUCTURE	Sequencing data	(i) A fast and powerful algorithm for identifying population structure using dense sequencing data. (ii) Current version: FineStructure 0.0.2.	http://paintmychromosomes.com/	Lawson et al. [111]
POPGENE	Use the dominant, codominant, and quantitative data for population genetic analysis	(i) Used to calculate gene and genotype frequency; allele number, effective allele number, polymorphic loci, gene diversity, observed and expected heterozygosity, Shannon index, homogeneity test, F-statistics, gene flow, genetic distance, dendrogram, neutrality test, and so forth. (ii) Current version: POPGENE version 1.32,	https://www.ualberta.ca/~fyeh/popgene.html	Francis et al. [112]
GENEPOP	Haploid or diploid data	(i) Used to compute exact tests or their unbiased estimation for Hardy-Weinberg equilibrium, population differentiation, and two-locus genotypic disequilibrium. (ii) It converts the input GENEPOP file to formats used by other popular programs, like BIOSYS [113], DIPLOIDL [114], and LINKDOS [115], thereby allowing communication between them. (iii) Current version: GENEPOP 4.2,	http://genepop.curtin.edu.au/	Raymond and Rousset [116]
GenAIEx	Codominant, haploid, and binary genetic data. It accommodates the full range of genetic markers available, including allozymes, SSRs, SNPs, AFLP, and other multilocus markers, as well as DNA sequences	(i) GenAIEx runs within Microsoft Excel enabling population genetic analysis of codominant, haploid, and binary data. Used to compute allele frequency-based analyses including heterozygosity, F-statistics, Nei's genetic distance, population assignment, probabilities of identity, and pairwise relatedness. (ii) Used for calculating genetic distance matrices and distance based calculations including analysis of molecular variance (AMOVA) [117, 118]; principal coordinates analysis (PCA); Mantel tests [119]; 2D spatial autocorrelation analyses following Smouse and Peakall [120], Peakall et al. [121], Double et al. [122]. (iii) Current version: GenAIEx 6.5.	http://biology-assets.anu.edu.au/GenAIEx/Welcome.html	Peakall and Smouse [123]

RAPD: Randomly amplified polymorphic DNA
RBIP: Retrotransposon-based insertion polymorphism
REF: Restriction endonuclease fingerprinting
REMAP: Retrotransposon-microsatellite amplified
           polymorphism
RFLP: Restriction fragment length polymorphism
SAMPL: Selective amplification of polymorphic loci
SCAR: Sequence characterised amplification regions
SNP: Single nucleotide polymorphism
SPAR: Single primer amplification reaction
SPLAT: Single polymorphic amplification test
S-SAP: Sequence-specific amplification polymorphisms
SSCP: Single strand conformation polymorphism
SSLP: Single sequence length polymorphism
SSR: Simple sequence repeats
STMS: Sequence-tagged microsatellite site
STS: Sequence-tagged site
TGGE: Thermal gradient gel electrophoresis
VNTR: Variable number tandem repeats
RAMS: Randomly amplified microsatellites.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Narain, "Genetic diversity—conservation and assessment," *Current Science*, vol. 79, no. 2, pp. 170–175, 2000.

[2] A. I. Turrent and J. A. Serratos-Hernandez, "Maize and biodiversity: the effects of transgenic maize in Mexico, chapter context and background on wild and cultivated maize in Mexico," *CEC Secretariat*, pp. 1–55, 2004.

[3] M. Smale, M. Istvan, I. Devra, and Jarvis, "The economics of conserving agricultural biodiversity on farm: research methods developed from IPRI's global project 'strengthening the scientific basis of in situ conservation of agricultural biodiversity,'" in *Proceedings of a Workshop Hosted by the Institute for Agro Botany (IA) and the International Pant Genetic Resource Institute (IPGRI '02)*, Godollo, Hungary, May 2002.

[4] R. E. Evenson and D. Gollin, "Assessing the impact of the Green Revolution, 1960 to 2000," *Science*, vol. 300, no. 5620, pp. 758–762, 2003.

[5] FAO, *The State of the World's Genetic Resources for Food and Agriculture*, FAO, Rome, Italy, 1998.

[6] S. Ceccarelli and S. Grando, "Plant breeding with farmers requires treating the assumptions of conventional plant breeding: lesson from the ICCRDA Barely Program," in *Farmers, Scientists and Plant Breeding Integrating Knowledge and Practice*, D. A. Cleveland and D. Soleri, Eds., CABI International, New York, NY, USA, 2002.

[7] J. Bruinsma, Ed., *World Agriculture: Towards 2015/2030, An FAO Perspective*, Earthscan, London, UK, 2003.

[8] S. di Falco, J. Jean-Paul Chvas, and M. Smale, *Farmers' Management of Production Risk on Degraded Lands the Role of Wheat Genetic Diversity in Tigray Region, Ethiopia*, Environment and Production Technology Division EPTD Discussion Papers no.153, International Food Policy Research Institute, International Live Stock Research Institute (ILRI), International Plant Genetic Resources Institute (IPRGI) and Food and Agriculture Organization of the United Nations (FAO), Washington, DC, USA.

[9] S. L. Pimm, J. L. Gittlaman, G. F. McCracken, and M. Gilpin, "Genetic bottlenecks: alternative explanations for low genetic variability," *Trends in Ecology and Evolution*, vol. 4, pp. 176–177, 1989.

[10] A. S. Jump and J. Peñuelas, "Running to stand still: adaptation and the response of plants to rapid climate change," *Ecology Letters*, vol. 8, no. 9, pp. 1010–1020, 2005.

[11] S. Freeman and J. C. Herron, *Evolutionary Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1998.

[12] M. L. Shaffer and F. B. Samson, "Population size and extinction: a note on determining critical population sizes," *The American Naturalist*, vol. 125, no. 1, pp. 144–152, 1985.

[13] M. E. Gilpin and M. E. Soulé, "Minimum viable populations: the processes of species extinctions," in *Conservation Biology: The Science of Scarcity and Diversity*, M. E. Soulé, Ed., pp. 13–34, Sinauer Associates, Sunderland, Mass, USA, 1986.

[14] G. Barcaccia, E. Albertini, D. Rosellini, S. Tavoletti, and F. Veronesi, "Inheritance and mapping of 2n-egg production in diploid alfalfa," *Genome*, vol. 43, no. 3, pp. 528–537, 2000.

[15] A. H. Paterson, "Making genetic maps," in *Genome Mapping in Plants*, A. H. Paterson, Ed., pp. 23–39, R. G. Landes Company, San Diego, Calif, USA, Academic Press, Austin, Tex, USA, 1996.

[16] P. Winter and G. Kahl, "Molecular marker technologies for plant improvement," *World Journal of Microbiology & Biotechnology*, vol. 11, no. 4, pp. 438–448, 1995.

[17] K. Weising, H. Nybom, K. Wolff, and W. Meyer, *Applications of DNA Fingerprinting in Plants and Fungi DNA Fingerprinting in Plants and Fungi*, CRC Press, Boca Raton, Fla, USA, 1995.

[18] V. Baird, A. Abbott, R. Ballard, B. Sosinski, and S. Rajapakse, "DNA diagnostics in horticulture," in *Current Topics in Plant Molecular Biology: Technology Transfer of Plant Biotechnology*, P. Gresshoff, Ed., pp. 111–130, CRC Press, Boca Raton, Fla, USA, 1997.

[19] R. Henry, "Molecular markers in plant improvement," in *Practical Applications of Plant Molecular Biology*, pp. 99–132, Chapman & Hall, London, UK, 1997.

[20] Y. Djè, M. Heuertz, C. Lefèbvre, and X. Vekemans, "Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers," *Theoretical and Applied Genetics*, vol. 100, no. 6, pp. 918–925, 2000.

[21] S. C. Hokanson, W. F. Lamboy, A. K. Szewc-McFadden, and J. R. McFerson, "Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids," *Euphytica*, vol. 118, no. 3, pp. 281–294, 2001.

[22] M. Jahufer, B. Barret, A. Griffiths, and D. Woodfield, "DNA fingerprinting and genetic relationships among white clover cultivars," in *Proceedings of the New Zealand Grassland Association*, J. Morton, Ed., vol. 65, pp. 163–169, Taieri Print Limited, Dunedin, New Zealand, 2003.

[23] Z. Galli, G. Halász, E. Kiss, L. Heszky, and J. Dobránszki, "Molecular identification of commercial apple cultivars with microsatellite markers," *HortScience*, vol. 40, no. 7, pp. 1974–1977, 2005.

[24] A. Alvarez, J. L. Fuentes, V. Puldón et al., "Genetic diversity analysis of Cuban traditional rice (*Oryza sativa* L.) varieties based on microsatellite markers," *Genetics and Molecular Biology*, vol. 30, no. 4, pp. 1109–1117, 2007.

[25] M. L. Ali, J. F. Rajewski, P. S. Baenziger, K. S. Gill, K. M. Eskridge, and I. Dweikat, "Assessment of genetic diversity and relationship among a collection of US sweet sorghum germplasm by SSR markers," *Molecular Breeding*, vol. 21, no. 4, pp. 497–509, 2008.

[26] V. V. Becerra, C. M. Paredes, M. C. Rojo, L. M. Díaz, and M. W. Blair, "Microsatellite marker characterization of Chilean common bean (*Phaseolus vulgaris* L.) germplasm," *Crop Science*, vol. 50, no. 5, pp. 1932–1941, 2010.

[27] B. C. Y. Collard and D. J. Mackill, "Marker-assisted selection: an approach for precision plant breeding in the twenty-first century," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1491, pp. 557–572, 2008.

[28] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, "Construction of a genetic linkage map in man using restriction fragment length polymorphisms," *The American Journal of Human Genetics*, vol. 32, no. 3, pp. 314–331, 1980.

[29] B. de Martinville, A. R. Wyman, R. White, and U. Francke, "Assignment of the first random restriction fragment length polymorphism (RFLP) locus (D14S1) to a region of human chromosome 14," *The American Journal of Human Genetics*, vol. 34, no. 2, pp. 216–226, 1982.

[30] D. Weber and T. Helentjaris, "Mapping RFLP loci in maize using B-A translocations.," *Genetics*, vol. 121, no. 3, pp. 583–590, 1989.

[31] J. L. Bennetzen, "Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions," *Plant Cell*, vol. 12, no. 7, pp. 1021–1029, 2000.

[32] K. M. Devos, M. D. Atkinson, C. N. Chinoy et al., "Chromosomal rearrangements in the rye genome relative to that of wheat," *Theoretical and Applied Genetics*, vol. 85, no. 6-7, pp. 673–680, 1993.

[33] J. Dubcovsky, W. Ramakrishna, P. J. SanMiguel et al., "Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes," *Plant Physiology*, vol. 125, no. 3, pp. 1342–1353, 2001.

[34] B. C. Y. Collard, M. Z. Z. Jahufer, J. B. Brouwer, and E. C. K. Pang, "An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts," *Euphytica*, vol. 142, no. 1-2, pp. 169–196, 2005.

[35] J. Welsh and M. McClelland, "Fingerprinting genomes using PCR with arbitrary primers," *Nucleic Acids Research*, vol. 18, no. 24, pp. 7213–7218, 1990.

[36] A. Jacobson and M. Hedrén, "Phylogenetic relationships in Alisma (*Alismataceae*) based on RAPDs, and sequence data from ITS and trnL," *Plant Systematics and Evolution*, vol. 265, no. 1-2, pp. 27–44, 2007.

[37] M. Mohan, S. Nair, A. Bhagwat et al., "Genome mapping, molecular markers and marker-assisted selection in crop plants," *Molecular Breeding*, vol. 3, no. 2, pp. 87–103, 1997.

[38] K. Semagn, Å. Bjørnstad, and M. N. Ndjiondjop, "An overview of molecular marker methods for plants," *African Journal of Biotechnology*, vol. 5, no. 25, pp. 2540–2568, 2006.

[39] L. Mondini, A. Noorani, and M. A. Pagnotta, "Assessing plant genetic diversity by molecular tools," *Diversity*, vol. 1, no. 1, pp. 19–35, 2009.

[40] M. Litt and J. A. Luty, "A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene," *The American Journal of Human Genetics*, vol. 44, no. 3, pp. 397–401, 1989.

[41] J. A. L. Armour, S. A. Alegre, S. Miles, L. J. Williams, and R. M. Badge, "Minisatellites and mutation processes in tandemly repetitive DNA," in *Microsatellites: Evolution and Applications*, D. B. Goldstein and C. Schlotterer, Eds., pp. 24–33, Oxford University Press, Oxford, UK, 1999.

[42] D. B. Goldstein and D. D. Pollock, "Launching microsatellites: a review of mutation processes and methods of phylogenetic inference," *Journal of Heredity*, vol. 88, no. 5, pp. 335–342, 1997.

[43] C. Schlotterer, "Microsatellites," in *Molecular Genetic Analysis of Populations: A Practical Approach*, A. R. Hoelzel, Ed., pp. 237–261, IRL Press, Oxford, UK, 1998.

[44] D. C. Queller, J. E. Strassmann, and C. R. Hughes, "Microsatellites and kinship," *Trends in Ecology and Evolution*, vol. 8, no. 8, pp. 285–288, 1993.

[45] M. W. Bruford, D. J. Cheesman, T. Coote et al., "Microsatellites and their application to conservation genetics," in *Molecular Genetic Approaches in Conservation*, T. B. Smith and R. K. Wayne, Eds., pp. 278–297, Oxford University Press, New York, NY, USA, 1996.

[46] J. H. McDonald and M. Kreitman, "Adaptive protein evolution at the *Adh* locus in *Drosophila*," *Nature*, vol. 351, no. 6328, pp. 652–654, 1991.

[47] R. L. Hammond, I. J. Saccheri, C. Ciofi et al., "Isolation of microsatellite markers in animals," in *Molecular Tools for Screening Biodiversity*, A. Karp, P. G. Issac, and D. S. Ingram, Eds., pp. 279–287, Chapman & Hall, London, UK, 1998.

[48] G. K. Chambers and E. S. MacAvoy, "Microsatellites: consensus and controversy," *Comparative Biochemistry and Physiology—B Biochemistry and Molecular Biology*, vol. 126, no. 4, pp. 455–476, 2000.

[49] L. Zane, L. Bargelloni, and T. Patarnello, "Strategies for microsatellite isolation: a review," *Molecular Ecology*, vol. 11, no. 1, pp. 1–16, 2002.

[50] J. Squirrell, P. M. Hollingsworth, M. Woodhead et al., "How much effort is required to isolate nuclear microsatellites from plants?" *Molecular Ecology*, vol. 12, no. 6, pp. 1339–1348, 2003.

[51] Y. Matsuoka, S. E. Mitchell, S. Kresovich, M. Goodman, and J. Doebley, "Microsatellites in Zea—variability, patterns of mutations, and use for evolutionary studies," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 436–450, 2002.

[52] R. Kota, R. K. Varshney, T. Thiel, K. J. Dehmer, and A. Graner, "Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.)," *Hereditas*, vol. 135, no. 2-3, pp. 145–151, 2002.

[53] R. V. Kantety, M. La Rota, D. E. Matthews, and M. E. Sorrells, "Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat," *Plant Molecular Biology*, vol. 48, no. 5-6, pp. 501–510, 2002.

[54] W. Michalek, W. Weschke, K.-P. Pleissner, and A. Graner, "EST analysis in barley defines a unigene set comprising 4,000 genes," *Theoretical and Applied Genetics*, vol. 104, no. 1, pp. 97–103, 2002.

[55] X.-P. Jia, Y.-S. Shi, Y.-C. Song, G.-Y. Wang, T.-Y. Wang, and Y. Li, "Development of EST-SSR in foxtail millet (*Setaria italica*)," *Genetic Resources and Crop Evolution*, vol. 54, no. 2, pp. 233–236, 2007.

[56] S. Senthilvel, B. Jayashree, V. Mahalakshmi et al., "Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags," *BMC Plant Biology*, vol. 8, article 119, 2008.

[57] I. Simko, "Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.)," *Journal of Heredity*, vol. 100, no. 2, pp. 256–262, 2009.

[58] M. Slatkin, "Isolation by distance in equilibrium and non-equilibrium populations," *Evolution*, vol. 47, no. 1, pp. 264–279, 1993.

[59] P. K. Gupta, S. Rustgi, S. Sharma, R. Singh, N. Kumar, and H. S. Balyan, "Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat," *Molecular Genetics and Genomics*, vol. 270, no. 4, pp. 315–323, 2003.

[60] I. Eujayl, M. K. Sledge, L. Wang et al., "Medicago truncatula EST-SSRs reveal cross-species genetic markers for *Medicago* spp," *Theoretical and Applied Genetics*, vol. 108, no. 3, pp. 414–422, 2004.

[61] Y. G. Cho, T. Ishii, S. Temnykh et al., "Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.)," *Theoretical and Applied Genetics*, vol. 100, no. 5, pp. 713–722, 2000.

[62] K. D. Scott, P. Eggler, G. Seaton et al., "Analysis of SSRs derived from grape ESTs," *Theoretical and Applied Genetics*, vol. 100, no. 5, pp. 723–726, 2000.

[63] I. Eujayl, M. E. Sorrells, M. Baum, P. Wolters, and W. Powell, "Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 399–407, 2002.

[64] K. Chabane, G. A. Ablett, G. M. Cordeiro, J. Valkoun, and R. J. Henry, "EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley," *Genetic Resources and Crop Evolution*, vol. 52, no. 7, pp. 903–909, 2005.

[65] D. Jaccoud, K. Peng, D. Feinstein, and A. Kilian, "Diversity arrays: a solid state technology for sequence information independent genotyping," *Nucleic Acids Research*, vol. 29, no. 4, article E25, 2001.

[66] D. Jaccoud, K. Peng, D. Feinstein, and A. Kilian, "Diversity arrays: a solid state technology for sequence information independent genotyping," *Nucleic Acids Research*, vol. 29, no. 4, article e25, 2001.

[67] M. Akbari, P. Wenzl, V. Caig et al., "Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome," *Theoretical and Applied Genetics*, vol. 113, no. 8, pp. 1409–1420, 2006.

[68] L. Zhang, D. Liu, X. Guo et al., "Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers," *BMC Genetics*, vol. 12, article 42, 2011.

[69] P. Wenzl, J. Carling, D. Kudrna et al., "Diversity Arrays Technology (DArT) for whole-genome profiling of barley," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9915–9920, 2004.

[70] S. Lezar, A. A. Myburg, D. K. Berger, M. J. Wingfield, and B. D. Wingfield, "Development and assessment of microarray-based DNA fingerprinting in Eucalyptus grandis," *Theoretical and Applied Genetics*, vol. 109, no. 7, pp. 1329–1336, 2004.

[71] A. H. J. Wittenberg, T. Van Der Lee, C. Cayla, A. Kilian, R. G. F. Visser, and H. J. Schouten, "Validation of the high-throughput marker technology DArT using the model plant Arabidopsis thaliana," *Molecular Genetics and Genomics*, vol. 274, no. 1, pp. 30–39, 2005.

[72] L. Xia, K. Peng, S. Yang et al., "DArT for high-throughput genotyping of Cassava (*Manihot esculenta*) and its wild relatives," *Theoretical and Applied Genetics*, vol. 110, no. 6, pp. 1092–1098, 2005.

[73] S. Yang, W. Pang, G. Ash et al., "Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology," *Theoretical and Applied Genetics*, vol. 113, no. 4, pp. 585–595, 2006.

[74] W. Gilbert, "DNA sequencing and gene structure Nobel lecture, 8 December 1980," *Bioscience Reports*, vol. 1, no. 5, pp. 353–375, 1981.

[75] M. Margulies, M. Egholm, and W. E. Altman, "Genome sequencing in microfabricated high-density picolitre reactors," *Nature*, vol. 437, pp. 376–380, 2005.

[76] J. Shendure, G. J. Porreca, N. B. Reppas et al., "Molecular biology: accurate multiplex polony sequencing of an evolved bacterial genome," *Science*, vol. 309, no. 5741, pp. 1728–1732, 2005.

[77] K. McKernan, A. Blanchard, L. Kotler, and G. Costa, "Reagents, methods, and libraries for bead-based sequencing," US Patent Application 20080003571, 2006.

[78] E. R. Mardis, "Next-generation DNA sequencing methods," *Annual Review of Genomics and Human Genetics*, vol. 9, pp. 387–402, 2008.

[79] X. Zhou, L. Ren, Q. Meng, Y. Li, Y. Yu, and J. Yu, "The next-generation sequencing technology and application," *Protein and Cell*, vol. 1, no. 6, pp. 520–536, 2010.

[80] J. Shendure and H. Ji, "Next-generation DNA sequencing," *Nature Biotechnology*, vol. 26, no. 10, pp. 1135–1145, 2008.

[81] M. L. Metzker, "Sequencing technologies the next generation," *Nature Reviews Genetics*, vol. 11, no. 1, pp. 31–46, 2010.

[82] M. Nei and W. H. Li, "Mathematical model for studying genetic variation in terms of restriction endonucleases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 10, pp. 5269–5273, 1979.

[83] P. Jaccard, "Nouvelles researches sur la distribution florale," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 44, pp. 223–270, 1908.

[84] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Science Bulletin*, vol. 38, pp. 1409–1438, 1958.

[85] S. A. Mohammadi and B. M. Prasanna, "Analysis of genetic diversity in crop plants—salient statistical tools and considerations," *Crop Science*, vol. 43, no. 4, pp. 1235–1248, 2003.

[86] A. Karp, S. Kresovich, K. V. Bhat, W. G. Ayad, and T. Hodgkin, "Molecular tools in plant genetic resources conservation: a guide to the technologies," IPGRI Technical Bulletin 2, International Plant Genetic Resources Institute, Rome, Italy, 1997.

[87] S. Schneider, D. Roessli, and L. Excoffier, *Arlequin: A Software for Population Genetics Data Analysis, 2000; Version 2.000*, Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva, Switzerland, 2000.

[88] L. Excoffier, G. Laval, and S. Schneider, "Arlequin (version 3.0): an integrated software package for population genetics data analysis," *Evolutionary Bioinformatics Online*, vol. 1, pp. 47–50, 2005.

[89] R. R. Hudson, M. Kreitman, and M. Aguadé, "A test of neutral molecular evolution based on nucleotide data," *Genetics*, vol. 116, no. 1, pp. 153–159, 1987.

[90] F. Tajima, "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism," *Genetics*, vol. 123, no. 3, pp. 585–595, 1989.

[91] Y.-X. Fu and W.-H. Li, "Statistical tests of neutrality of mutations," *Genetics*, vol. 133, no. 3, pp. 693–709, 1993.

[92] Y.-X. Fu, "New statistical tests of neutrality for DNA samples from a population," *Genetics*, vol. 143, no. 1, pp. 557–570, 1996.

[93] J. Rozas and R. Rozas, "DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics

parameters from DNA sequence data," *Computer Applications in the Biosciences*, vol. 11, no. 6, pp. 621–625, 1995.

[94] J. Rozas and R. Rozas, "DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis," *Computer Applications in the Biosciences*, vol. 13, no. 3, pp. 307–311, 1997.

[95] J. Rozas and R. Rozas, "DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis," *Bioinformatics*, vol. 15, no. 2, pp. 174–175, 1999.

[96] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.

[97] K. Liu and S. V. Muse, "PowerMaker: an integrated analysis environment for genetic maker analysis," *Bioinformatics*, vol. 21, no. 9, pp. 2128–2129, 2005.

[98] X. Perrier and J. P. Jacquemoud-Collet, DARwin software, 2006, http://darwin.cirad.fr/darwin.

[99] N. Mantel, "The detection of disease clustering and a generalized regression approach.," *Cancer Research*, vol. 27, no. 2, pp. 209–220, 1967.

[100] F. J. Rohlf, *NTSYSpc: Numerical Taxonomy System, Version 2.1*, Exeter Publishing, Setauket, NY, USA, 2002.

[101] S. Kumar, M. Nei, J. Dudley, and K. Tamura, "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences," *Briefings in Bioinformatics*, vol. 9, no. 4, pp. 299–306, 2008.

[102] S. Kumar, K. Tamura, and M. Nei, "MEGA: molecular evolutionary genetics analysis software for microcomputers," *Computer Applications in the Biosciences*, vol. 10, no. 2, pp. 189–191, 1994.

[103] S. Kumar, K. Tamura, I. B. Jakobsen, and M. Nei, "MEGA2: molecular evolutionary genetics analysis software," *Bioinformatics*, vol. 17, no. 12, pp. 1244–1245, 2002.

[104] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731–2739, 2011.

[105] D. L. Swofford, *PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4*, Sinauer Associates, Sunderland, Mass, USA, 2002.

[106] J. K. Pritchard, M. Stephens, and P. Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, vol. 155, no. 2, pp. 945–959, 2000.

[107] D. Falush, M. Stephens, and J. K. Pritchard, "Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies," *Genetics*, vol. 164, no. 4, pp. 1567–1587, 2003.

[108] M. J. Hubisz, D. Falush, M. Stephens, and J. K. Pritchard, "Inferring weak population structure with the assistance of sample group information," *Molecular Ecology Resources*, vol. 9, no. 5, pp. 1322–1332, 2009.

[109] A. Raj, M. Stephens, and J. K. Pritchard, "fastSTRUCTURE: variational inference of population structure in large SNP data sets," *Genetics*, vol. 197, no. 2, pp. 573–589, 2014.

[110] D. H. Alexander, J. Novembre, and K. Lange, "Fast model-based estimation of ancestry in unrelated individuals," *Genome Research*, vol. 19, no. 9, pp. 1655–1664, 2009.

[111] D. J. Lawson, G. Hellenthal, S. Myers, and D. Falush, "Inference of population structure using dense haplotype data," *PLoS Genetics*, vol. 8, no. 1, Article ID e1002453, 2012.

[112] C. Yeh Francis, R. C. Yang, B. J. Boyle Timothy, Z. H. Ye, and X. Mao Judy, *POPGENE Version 1.32, The User-Friendly Shareware for Population Genetic Analysis*, Molecular Biology and Biotechnology Centre, University of Alberta, Alberta, Canada, 1999, http://www.ualberta.ca/~fyeh/.

[113] D. L. Swofford and R. B. Selander, "Biosys-1: a FORTRAN program for the comprehensive analysis for electrophoretic data in population genetics and systematics," *Journal of Heredity*, vol. 72, no. 4, pp. 281–283, 1981.

[114] B. S. Weir, "Intraspecific differentiation," in *Molecular Systematic*, D. M. Hillis and C. Moritz, Eds., pp. 373–410, Sinauer Associates, Sunderland, Mass, USA, 1990.

[115] P. Garnier-Gere and C. Dillmann, "A computer program for testing pairwise linkage disequilibria in subdivided populations," *Journal of Heredity*, vol. 83, no. 3, p. 239, 1992.

[116] M. Raymond and F. Rousset, "ENPOP: population genetics software for exact tests and ecumenicism," *Journal of Heredity*, vol. 86, pp. 248–249, 1995.

[117] L. Excoffier, P. E. Smouse, and J. M. Quattro, "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data," *Genetics*, vol. 131, no. 2, pp. 479–491, 1992.

[118] R. Peakall, P. E. Smouse, and D. R. Huff, "Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss Buchloe dactyloides," *Molecular Ecology*, vol. 4, no. 2, pp. 135–147, 1995.

[119] P. E. Smouse, C. J. Long, and R. R. Sokal, "Multiple regression and correlation extensions of the Mantel test of matrix correspondence," *Systematic Zoology*, vol. 35, no. 4, pp. 627–632, 1986.

[120] P. E. Smouse and R. Peakall, "Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure," *Heredity*, vol. 82, no. 5, pp. 561–573, 1999.

[121] R. Peakall, M. Ruibal, and D. B. Lindenmayer, "Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, Rattus fuscipes," *Evolution*, vol. 57, no. 5, pp. 1182–1195, 2003.

[122] M. C. Double, R. Peakall, N. R. Beck, and A. Cockburn, "Dispersal, philopatry, and infidelity: dissecting local genetic structure in superb fairy-wrens (*Malurus cyaneus*)," *Evolution*, vol. 59, no. 3, pp. 625–635, 2005.

[123] R. Peakall and P. E. Smouse, "GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research," *Molecular Ecology Notes*, vol. 6, no. 1, pp. 288–295, 2012.