

# Chapter 6

## Using Molecular Techniques to Dissect Plant Genetic Diversity

Linda Mondini and Mario A. Pagnotta

**Abstract** During the past few decades, the employment of molecular markers to discover polymorphisms in DNA has been playing an increasing role in conservation strategies and use of plant genetic resources (PGR). Molecular markers are indispensable tools for determining the genetic variation and biodiversity with high levels of accuracy and reproducibility in short times. Different typologies of molecular markers exist, specific for the different applications in molecular genetic methods. Molecular tools have been successfully applied in the analysis of specific genes and gene pathways, as well as to increase understanding of gene action, to generate genetic maps and assist in the development of gene transfer technologies. Molecular markers have also had a critical role in studies of phylogeny and species evolution, and have been applied to increase our understanding of the distribution and extent of genetic variation within and between species. The main two groups of molecular markers can be classified on the basis of the analysis method used: polymerase chain reaction (PCR) and non-PCR-based. Recently, a new class of advanced techniques has emerged, primarily derived from a combination of earlier, more basic techniques. Advanced marker techniques tend to amalgamate advantageous features of several basic techniques, in order to increase the sensitivity and resolution to detect genetic discontinuity and distinctiveness. The past several years have seen revolutionary advances in DNA sequencing technologies with the advent of next-generation sequencing (NGS) techniques. NGS methods now allow millions of bases to be sequenced in one round, at moderate prices and in very short times. This paper is an overview of the diverse, predominantly molecular techniques, used in assessing plant genetic diversity, discussing

---

L. Mondini (✉) · M.A. Pagnotta  
Department of Science and Technologies for Agriculture, Forestry,  
Nature and Energy (DAFNE), Tuscia University, Via S. C. de Lellis, 01100 Viterbo, Italy  
e-mail: linda.mondini@unitus.it

M.A. Pagnotta  
e-mail: pagnotta@unitus.it

about the most important and recent advances made in molecular marker techniques, their applications, advantages, and limitations applied to plant sciences in order to provide base platform information to the researchers working in the area.

**Keywords** Genetic variation • Biodiversity • Molecular markers • Molecular techniques

## 6.1 Introduction

Unraveling the molecular basis of the essential biological phenomena in plants is crucial for effective and sustainable conservation, management, and efficient utilization of plant genetic resources (PGR). An adequate understanding of existing genetic diversity and how to best utilize it, is of fundamental interest for basic science and applied aspects like the efficient management of PGR. In particular, the improvement of crop genetic resources strictly depends on the continuous introduction of wild relatives, traditional varieties, and the use of modern breeding techniques, requiring an assessment of diversity at some levels in order to select promising varieties.

### 6.1.1 *The Assessment of Genetic Variation*

When studying and measuring diversity, it is imperative to understand what to conserve and/or what is being lost. For the conservation and utilization of PGR, genetic relationships are more important than the taxonomy per se. The gene pools concept, as proposed by Harlan and de Wet (1971), focuses neatly on the relationships between individuals and populations and it is of particular relevance to plant breeders to improve crops (Greene and Morris 2001). The concept is based on the division of the genetic resources into three gene pools: (i) primary gene pool (gene pool 1 or GP 1) to which the crop species, crop wild relatives, and related weedy species belong with crosses yielding fertile hybrids; (ii) secondary gene pool (GP 2) comprises related taxa which are able to hybridize with the crop species but the gene transfer is poor and the progeny are often sterile and not viable; and (iii) tertiary gene pool (GP 3) includes distantly related taxa which do not cross readily in the wild and require anthropogenic assistance in gene transfer and hybridizing through sophisticated techniques, such as embryo culture, grafting, chromosome doubling, and the use of bridging species. We can argue that plant breeding now requires the addition of a quaternary gene pool (GP 4) where gene transfer could take place but only through genetic engineering. Diversity can be measured at the morphological, biochemical, and/or molecular level.

### 6.1.2 Morphological Characterization

Morphological characterization is based on assessing the phenotype, which is the result of genetic and environment interactions, and can be modified to diverse extents by different environmental factors. The capability to respond to environmental pressures without the implication of mutations, known as phenotypic plasticity, can be divided into two main categories: (i) *development flexibility*, which produces the development of different genotypes in different environmental conditions and (ii) *behavioral flexibility*, which comprises all the behavioral elements which allow a temporary adaptation to a particular environmental condition. Genetic variation has been found to contribute significantly to phenotypic variation and produces two main types of characters: (a) *quantitative* characters which are measurable characters and give rise to continuous variability (defined by a Gaussian curve) and (b) *qualitative* characters which are alternative, discontinuous, not defined by a Gaussian curve and producing a type of variability, so-called “discontinuous”. The study of morphological variability is the classical way of assessing genetic diversity. For many species, especially minor crops, it is still the only approach used. Nevertheless, morphological characterization, even if does not require expensive analysis tools, requires large tracts of land for the experiments, making it even more expensive than molecular detection. Moreover, the traits analyzed are often susceptible to phenotypic plasticity; conversely, this allows assessment of diversity in the presence of environmental variation. However, an analysis of genetic diversity based only on agronomic and morphological traits might be erroneous considering that distinct morph types can result from a few mutations.

## 6.2 Cytological Characterization

Cytological markers have been deeply used for the assessment of PGR based on the numbers and morphology of plant chromosomes. Cytological markers include chromosome karyotypes, bandings, repeats, deletions, translocations, and inversions. Mitotic chromosomes permit to analyze the nuclear genome by microscopic means, allowing the observation of its components individually, as well as globally (the karyotype). Karyotypes offer a phenotypic view of the genotype and prior to the application of chromosome banding, distance analysis was done using various numerical and metric values that described the karyotype such as diploid number (number of chromosomes or  $2n$ ) and fundamental number (number of chromosomal arms or  $N_{fa}$ ).

Ploidy levels are sometimes used to compare species, mainly in plants (Saideswara et al. 1989). Polyploidization, widely spread in plant genomes, can result from genome duplication (autopolyploidization) or by hybridization (allopolyploidization). Nevertheless, even if these changes are considered as rare,

convergence in ploidy may not be that uncommon and tending sometimes to revert to the diploid level complicating the understanding of polyploidization patterns. However, a recently developed technique based on genomic in situ hybridization (GISH) represents a powerful tool for investigating the evolution of polyploidy organisms (D'Hont et al. 2002). Chromosome banding is a powerful and routinely used tool to investigate chromosomal homology and comprise differential staining techniques that reveal a succession of bands along the length of a chromosome that vary in width and staining intensity. These bands reflect intrinsic properties of the genome (Sumner 1990) allowing access to information involving both structural (GTG-, RHG-, and CBG-banding) and functional patterns (replication RBG-banding) of chromosomes (Viegas-Pequignot and Dutrillaux 1978). Moreover, the development of in situ hybridization, and in particular fluorescent in situ hybridization (FISH) using chromosome painting probes (Ferguson-Smith 1997), has confirmed the evidence that homology in banding patterns is significantly related to homology in gene content and synteny conservation. The development of powerful molecular cytogenetic and genomic strategies such as FISH, flow-cytometry, and chromosome painting jointly to gene mapping, allows to overcome the limitations of conventional banding analysis (Ferguson-Smith 1997). Based on the hybridization between labeled DNA probes and genomic DNA, in situ hybridization techniques permit the unequivocal confirmation of homology among chromosomes. Therefore, molecular cytogenetics makes it possible to assess homologies between distantly related taxa and this creates new opportunities for determining chromosomal relationships at higher taxonomic levels (Yang et al. 2003).

### 6.3 Biochemical Characterization

Biochemical characterization includes the assessment of seed storage proteins and *allozymes/isozymes*. These techniques use enzymatic functions and are comparatively inexpensive while being powerful methods of measuring allele frequencies for specific genes. However, because there are only a few allozyme systems per species (not more than 30), there are correspondingly few markers. Analyses of allozymes provide an estimate of gene and genotypic frequencies within and between populations. Such data can be used to measure population subdivision, genetic diversity, gene flow, genetic structure of species, and comparisons among species (Spooner et al. 2005). The first experiences in analysis of isoenzymatic polymorphism in natural populations date up to Zouros and Foltz (1987). Since then, isozymes have been heavily employed also in plant studies and particularly for population genetics studies (Brown 1979). Therefore, allozymes have been used in studying out-crossing rates, population structure, and population divergence, such as in the case of crop wild relatives (Hamrick and Godt 1997; Guarino 1999; Volis et al. 2001; Gonzalez et al. 2005).

Among the major advantages of these types of markers are co-dominance, absence of epistatic and pleiotropic effects, ease of use, and low costs even if at

the same time they present some important limitations such as the limited number of polymorphic enzymatic systems available, the fact that enzymatic loci represent the expressed part of the genome which is only a small and not random portion; they are affected by the phenological phase of the plant, and finally that the observed variability may be not representative of the entire genome. Moreover, although these markers permit a high processivity, a comparison of samples from different species, loci, and laboratories is problematic being affected by extraction methodology, plant tissue, and plant stage.

## 6.4 Molecular Characterization

Analyses of genetic diversity are usually based on either allozymes or molecular markers, which tend to be selectively neutral. It has been argued that the rate of diversity loss of these neutral markers will be higher than those that are associated with fitness. In order to verify this, Reed and Frankham (2003) conducted a meta-analysis of fitness components in three populations and in which heterozygosity, and/or heritability, and/or population size were measured. Their findings, based on 34 datasets, concluded that heterozygosity, population size, and quantitative genetic variation, which are all used as indicators of fitness, were all significantly positively correlated with population fitness.

Genetic variability within a population can be assessed through:

1. The number (and percentage) of polymorphic genes in the population.
2. The number of alleles for each polymorphic gene.
3. The proportion of heterozygous loci per individual (Primack 1993).

Protein methods, such as allozyme electrophoresis, and molecular methods, such as DNA analysis, directly measure genetic variation, giving a clear indication of the levels of genetic variation present in a species and/or population (Karp et al. 1996) without direct interference from environmental factors. However, they have the disadvantage of being relatively expensive, time-consuming, and require high levels of expertise and materials in analysis.

The concept of genetic markers is not a new one; in the nineteenth century, Gregor Mendel employed phenotype-based genetic markers in his experiments. Later, phenotype-based genetic markers for *Drosophila melanogaster* led to the founding of the theory of genetic linkage, occurring when particular genetic loci or alleles for genes are inherited jointly. The limitations of phenotype-based genetic markers led to the development of DNA-based markers, i.e., molecular markers. A molecular marker can be defined as a genomic locus, detected through probe or specific starters (primer) which, in virtue of its presence, distinguishes unequivocally the chromosomal trait which it represents as well as the flanking regions at the 3' or 5' extremity (Barcaccia et al. 2000).

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. Molecular characterization is more expensive, but many markers are now known, thus enabling the study of a much larger number of genes that code for plant expression, as well as for other noncoding segments of the chromosomes. Analysis is based on extracting DNA, amplifying it (more often than not, through polymerase chain reaction procedures) and analyzing the resulting gene frequencies and DNA sequences. A molecular marker detects gene sequences at a known location of a chromosome. These markers do not refer to the activity of specific genes, but are directly based on highlighting differences (polymorphisms) within a nucleic sequence in different individuals, as a result of insertion, deletions, translocations, duplications, point mutations, etc.

The seemingly bewildering array of possible approaches is among the first problem faced by newcomers considering the application of these techniques to their own system. A starting point for discerning the different classes of molecular markers can be to consider the different techniques employed. These are based either on restriction-hybridization of nucleic acid or techniques based on polymerase chain reaction (PCR), or both. A further distinction can be obtained through the selection of either multi-locus or single-locus markers.

Multi-locus markers allow simultaneous analyses of several genomic loci, which are based on the amplification of casual chromosomal traits through oligonucleic primers with arbitrary sequences. These types of markers are also defined as *dominant* since it is possible to observe the presence or the absence of a band for any locus, but it is not possible to distinguish between heterozygote ( $a/-$ ) condition and homozygote for the same allele ( $a/a$ ) and attribute different allelic variants at the same locus. By contrast, single-locus markers employ probes or primers specific to genomic loci, and are able to hybridize or amplify chromosome traits with well-known sequences. They are defined as co-dominant since they allow discrimination between homozygote and heterozygote loci.

Advances in the development of molecular marker techniques, powerful tools have been developed so that genetic resources can be accurately assessed and characterized (Table 6.1). Most of these techniques, based on the analysis of information-rich nucleic molecules, provide a reliable estimation of relatedness, phylogeny, and inheritance of genetic characteristics (Caetano-Anolles et al. 1991). Through molecular markers and maps, it is possible to obtain an overall vision on the genes controlling agronomic, morphological, and biochemical traits in plants. Additionally, they become essential for explaining whether existing genetic variability, which is assessed by measuring biochemical factors and morphological traits, is related to genetic diversity analyzed measuring allelic frequencies detected with molecular markers. Through this information it is possible to construct a core collection, which can represent a base for future breeding programs. Hence, in the current scenario, molecular markers become the marker of choice for the study of crop genetic diversity revolutionizing the plant biotechnology.

**Table 6.1** Molecular markers classification

<b>Non-PCR-based techniques</b>	
<i>Restriction-hybridization techniques</i>	
RFLP	Restriction Fragment Length Polymorphism
VNTR	Variable Number Tandem Repeats
REF	Restriction Endonuclease Fingerprinting
<b>PCR derived</b>	
<i>Multiple Arbitrary Amplicon Profiling (MAAP)</i>	
RAPD	Randomly Amplified Polymorphic DNA
DAF	DNA Amplification Fingerprint
AP-PCR	Arbitrarily primed PCR
AFLP	Amplified Fragment Length Polymorphism
SAMPL	Selective Amplification of Polymorphic Loci
ISSR	Inter-Simple Sequence Repeats
SPAR	Single Primer Amplification Reaction
DAMD	Directed Amplification of Minisatellites DNA
<i>Targeted PCR</i>	
<i>Sequence Tagged Sites (STS)</i>	
ARMS	Amplification Refractory Mutation System
ASAP	Arbitrary Signatures from Amplification
ASH	Allele-Specific Hybridization
ASLP	Amplified Sequence Length Polymorphism
ISTR	Inverse Sequence-Tagged Repeats
SSCP	Single Strand Conformation Polymorphism
SPLAT	Single Polymorphic Amplification Test
TGGE	Thermal Gradient Gel Electrophoresis
DGGE	Denaturing Gradient Gel Electrophoresis
<i>Markers based on microsatellite sequences</i>	
SSR	Simple Sequence Repeats
RAHM	Randomly Amplified Hybridizing Microsatellites
RAMPs	Randomly Amplified Microsatellite Polymorphisms
STMS	Sequence Tagged Microsatellite Site
SSLP	Single Sequence Length Polymorphism
MP-PCR	Microsatellite-Primed PCR
RAMS	Randomly Amplified Microsatellites
CAPS	Cleaved Amplification Polymorphic Sequence
SCAR	Sequence Characterized Amplification Regions
SNP	Single Nucleotide Polymorphism
EST-SSR	Expressed Sequence Tags-SSR
<i>Markers based on DNA sequencing</i>	
DART	Diversity Arrays Technology
ASO	Allele Specific Oligonucleotide
CAS	Coupled Amplification and Sequencing

(continued)

**Table 6.1** (continued)

GBA	Genetic Bit Analysis
OLA	Oligonucleotide Ligation Assay
<i>RNA-Based Molecular Markers (RBMs)</i>	
iSNAP	Inter Small RNA Polymorphism
RAP-PCR	RNA Arbitrarily Primed PCR
cDNA-SSCP	cDNA-Single Strand Conformation Polymorphism
cDNA-AFLP	cDNA-Amplified Fragment Length Polymorphism
cDNA-RFLP	cDNA-Restriction Fragment Length Polymorphism
<i>Transposable elements-based molecular markers</i>	
REMAP	Retrotransposon-Microsatellite Amplified Polymorphism
RBIP	Retrotransposon-Based Insertion Polymorphism
IRAQ	Inter-Retrotransposon Amplified Polymorphism
IRAP	Inter-Retrotransposon Amplified Polymorphism
S-SAP	Sequence-Specific Amplification Polymorphisms

## 6.5 The Choice of the “Perfect” Molecular Marker

Due to the rapid developments in the field of molecular genetics and thanks to the novel findings in next-generation sequencing (NGS), a large amount of different techniques have emerged to analyze genetic variation in the recent years.

Unfortunately, there is no single molecular approach for many of the problems facing gene bank managers, and many techniques complement each other, hence the choice of marker typology that “suits me” becomes very difficult. However, some techniques are clearly more appropriate than others for some specific applications like crop diversity and taxonomy studies. In this perspective, the understanding of all features that characterize a molecular marker class is crucial.

Genetic markers can differ with respect to important features such as

- level of polymorphism detected;
- locus specificity,
- genomic abundance,
- reproducibility,
- technical requirements and highly qualified personnel,
- costs, and
- time constraints.

No marker is superior to all others for a wide range of applications and the most appropriate genetic marker strictly depends on the application (Table 6.2). An ideal molecular marker should possess the following features:

1. Be highly polymorphic: necessary condition to assess genetic variability;
2. Co-dominant: able to discriminate between homozygous and heterozygous states in diploid organisms;



3. Frequent occurrence in genome;
4. Provide adequate resolution of genetic differences;
5. Detect multiple, independent, and reliable loci;
6. Selective neutral behaviors: the DNA sequences of any organism are neutral to environmental conditions or management practices, this permits to confer the variation only to a genetic origin;
7. Easy access and fast assay: it must be simple, quick, and inexpensive;
8. High reproducibility: to guarantee robust results among different laboratory and equipments;
9. Requiring small amounts of tissue and DNA samples;
10. Link to distinct phenotypes;
11. Require no prior information about the genome of an organism.

However, it is practically impossible to define a molecular marker which would meet all the above criteria. Hence, the choice of the right marker is based on the capability to associate the different features to the specific application to be undertaken (Weising et al. 1995) (Fig. 6.1). At first, molecular markers can be classified as hybridization-based markers and PCR-based markers. In the former, DNA profiles are visualized by hybridizing the restriction enzyme-digested DNA, to a labeled probe, which is a DNA fragment of known origin or sequence. A PCR-based marker involves in vitro amplification of DNA sequences or loci, using specifically or arbitrarily chosen oligonucleotide fragments (primers) and a thermostable DNA polymerase enzyme (*Taq* polymerase). The amplified fragments are separated electrophoretically and banding patterns are detected by different methods such as staining, autoradiography, or directly sequenced. The primer sequences are chosen to allow base-specific binding to the template in reverse orientation. PCR is extremely sensitive, fast, and reliable. Its application for diverse purposes has opened up a multitude of new possibilities in the field of molecular biology and genetics.

Recently, a new class of advanced techniques has emerged, primarily derived from a combination of the earlier, more basic techniques. These advanced marker techniques combine advantageous aspects of several basic techniques. In particular, the newer methods incorporate modifications in the basic techniques, thereby increasing the sensitivity and resolution in detecting genetic discontinuity and distinctiveness. The advanced marker techniques also utilize newer classes of DNA elements such as retrotransposons, mitochondrial, and chloroplast-based microsatellites, allowing increased genome coverage. Techniques such as Random Amplified Polymorphic DNA (RAPD) and Amplified Fragment Length Polymorphism (AFLP) are also being applied to cDNA-based templates (i.e., sequences of complementary DNA obtained by mRNA retrotranscription) to study patterns of gene expression and uncover the genetic basis of biological responses. With the advent of NGS technologies it is presently possible to analyze high numbers of samples over smaller periods of time.

**Table 6.2** Comparison among the most widely used molecular markers in plants

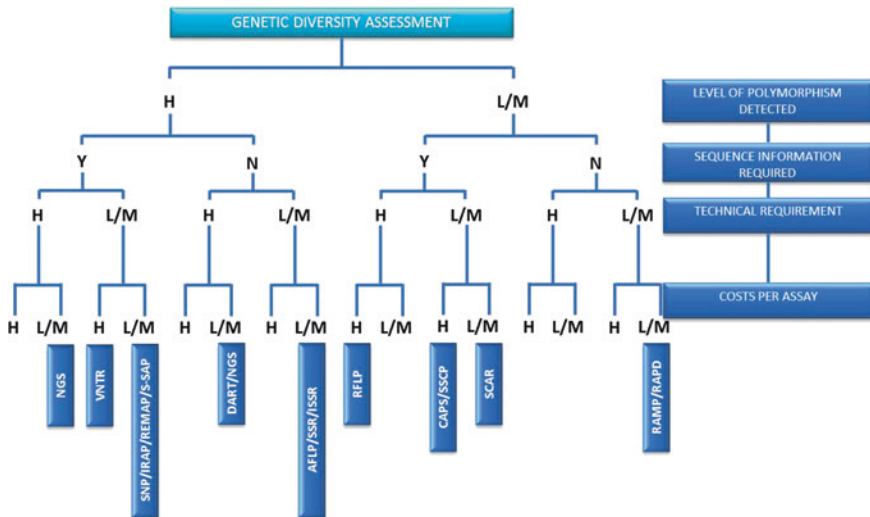
Marker tipology	PD	LS	I	QD	R	SI	A	C	TE	AD	DS
RFLP	M	Y	C	H	H	Y	N	H	M	-High genomic abundance and coverage -Possibility to use filters many times	-Large amount of gDNA required -Laborious and time-consuming -Radioactive labeling -Cloning/characterization of probe
AFLP	H	N	D	M	H	N	Y	M	M	-High genomic abundance -Used across species	-Dominant markers -Laborious interpretation of patterns -Laborious and time-consuming
RAPD	M	N	D	L	L	N	Y	L	L	-High genomic abundance and coverage -Easy and fast	-Dominant markers -Not reproducible -Not across species
SSR	H	N	C	L	M	N	Y	L	L	-High genomic abundance and coverage -Amenability to multiplexing and automation -Presence of multiple alleles	-Medium reproducibility -Not across species -Preliminary sequence information
SNP	H	Y	C	L	H	Y	Y	M	M	-Amenability to automation -Highly polymorphic -Allow cultivar discrimination	-Preliminary sequence information -High costs
DART	H	Y	D	L	H	N	Y	L	H	-Any sequence information required -High-throughput data production -Low costs	-Complex and expensive technical requirement -Primarily dominant markers -Currently used in few species
CAPS	L	Y	C	L	H	Y	Y	H	M	-High reproducibility -Co-dominant markers -Easy and fast respect to RFLP and CAPS	-Low polymorphism detected -Preliminary sequence information
RAMP	M	Y	D	L	M	N	Y	M	M	-High degree of allelic polymorphism respect to RAPD	-Labor intensive technique -Use of radiolabeled primer

(continued)

**Table 6.2** (continued)

Marker topology	PD	LS	I	QD	R	SI	A	C	TE	AD	DS
SSCP	L	Y	C	L	M	Y	N	H	M	-Co-dominant markers -Low quantity of DNA required	-Need of highly standardized protocols -Inability to detect all the mutations present -Low results robustness
IRAP/ REMAP/SSAP	H	Y	D	L	H	Y	Y	L	M	-Great for phylogenetic studies -High genomic coverage -Amenability to multiplexing -Easier interpretation of patterns respect to AFLP	-Preliminary sequence information required for flanking regions -Dominant markers -Low reproducibility -Multi-locus markers
ISSR	H	N	D	L	L	N	Y	L	L	-Quick and easy to use -High reproducibility -Locus specificity	-Preliminary sequence information required
SCAR	M	Y	C	L	H	Y	Y	L	L	-High level of polymorphism -High reproducibility	-Large amount of gDNA required -Laborious and time-consuming -Cloning/characterization of probe
VNTR	H	Y	D	H	H	Y	N	H	M		

*PD* polymorphism degree, *LS* locus specificity, *I* inheritance [dominant (*D*)/co-dominant(*C*)], *QD* quantity of DNA required, *R* reproducibility, *SI* sequence information required, *A* amenability to automation, *C* costs, *TE* technical equipment, *H* high, *L* low, *M* medium, *Y* yes, *N* no



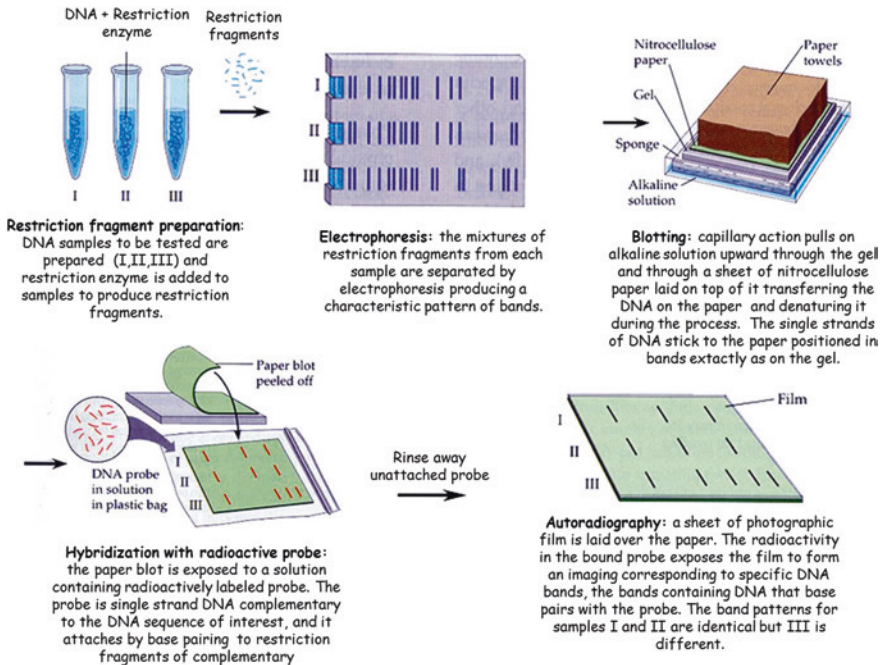
**Fig. 6.1** A rational scheme for choosing the most appropriate molecular genetics analysis strategy. *H* high, *L* low, *M* medium, *Y* yes, *N* no

## 6.6 Non-PCR-Based Techniques

### 6.6.1 Restriction-Hybridization Techniques

Molecular markers based on restriction-hybridization techniques were employed relatively early in the field of plant studies and combined the use of restriction endonucleases and the hybridization method (Southern 1975). Restriction endonucleases are bacterial enzymes able to cut DNA, identifying specific palindrome sequences and producing polynucleotide fragments with variable dimensions. Any changes within sequences (i.e., point mutations), mutations between two sites (i.e., deletions and translocations), or mutations within the enzyme site, can generate variations in the length of restriction fragment obtained after enzymatic digestion.

Restriction Fragment Length Polymorphisms (RFLPs) was the first technology developed which enabled the detection of polymorphisms at the sequence level. The approach comprises the digestion of genomic DNA with restriction enzymes, separations of the resultant DNA fragments by gel electrophoresis, blotting of the fragments to a filter followed by the hybridization with a chemically-labeled DNA probe to a Southern blot resulting in differential DNA fragment profile. The sequences of the probes may be known (e.g., from a cloned gene) or unknown (e.g., genomic or cDNA random cloned fragments) (Fig. 6.2). The combination of specific systems probes/enzymes produces highly reproducible patterns for a given individual and the variation in the restriction profiles between two different individuals occurs when mutations in the DNA sequences change the restriction sites which cannot be recognized by the restriction enzymes. RFLP technique was widely exploited to construct genetic maps and has been successfully applied



**Fig. 6.2** Different steps of restriction fragment length polymorphism technique

to genetic diversity assessment, particularly in cultivated plants (Castagna et al. 1994; Deu et al. 1994) as well as in populations and wild relatives (Besse et al. 1994; Laurent et al. 1994; Bark and Havey 1995).

The RFLP markers are relatively highly polymorphic, co-dominantly inherited, highly replicable, and allow the simultaneous screening of numerous samples. DNA blots can be analyzed repeatedly by stripping and reprobing (usually from eight to ten times) with different RFLP probes. Nevertheless, this technique is not very widely used as it is time-consuming, involves expensive and radioactive/toxic reagents, and requires large quantities of high quality genomic DNA (e.g. 10 µg per digestion). Moreover, the prerequisite of prior sequence information for probe construction contributes to the complexity of the methodology. However, the main problem faced is simply that insufficient level of polymorphism is detectable at the below species level. Nevertheless, RFLPs have been widely used to investigate relationships of closely related taxa (Miller and Tanksley 1990; Lanner et al. 1996), for studies on hybridization and introgression (in particular studies concerning the gene flow between crops and weeds) (Brubaker and Wendel 1994; Clausen and Spooner 1998), for diversity studies (Dubreuil et al. 1996), and as fingerprinting tools (Fang et al. 1997). They have also been successfully employed in gene mapping studies due to their high genomic abundance and random distribution throughout the genome (Neale and Williams 1991). Moreover, RFLP markers were used for the first time in the construction of genetic maps by Botstein et al. (1980). Nevertheless,

a class of molecular markers able to overcome this inconvenience exists. These markers are developed on a particular class of highly variable regions interspersed along the genome and constituted by repeats of short simple sequences. These are known as “microsatellites” and are formed by basic repeat units from 2 to 8 base pairs in length (microsatellites or SSR) up to longer repeats of 16–100 base pairs called “minisatellites”. Being highly hypervariable, RFLP analysis using probes for mini-microsatellites produces multi-locus patterns able to discriminate at the level of populations and individuals. The variation produced derives from changes in the copy number of the basic repeat and the marker class based on this kind of variation is specifically called Variable Numbers of Tandem Repeats (VNTRs). Being highly polymorphic VNTRs have been widely applied for studying within and between population variation, for estimating genetic distances, and for ecological applications (Lynch 1990; Alberte et al. 1994; Antonious and Nybom 1994).

However, like the RFLP approach VNTRs show the same limitations that led to the development of a new set of less technically complex methods known as PCR-based techniques. Nevertheless, when combined with PCR amplification of a specific locus both RFLPs and VNTRs probes have much to offer.

## **6.7 Markers Based on Amplification Techniques (PCR-Derived)**

With the advent of PCR analysis, an increasing number of techniques became available to screen the genetic diversity. In fact, the use of this kind of marker has been exponential, following the development by Mullis et al. (1986) of PCR assay consisting in the amplification of several discrete DNA products, deriving from regions of DNA which are flanked by regions of high homology with the primers. These regions must be close enough to one another to permit the elongation phase producing several discrete DNA products.

The use of random primers overcame the limitation of prior sequence knowledge for PCR analysis, and being applicable to all organisms facilitated the development of genetic markers for a variety of purposes. PCR-based techniques can further be subdivided into two subcategories: (1) arbitrarily primed PCR-based techniques or sequence nonspecific techniques and (2) sequence targeted PCR-based techniques. Based on the first category, two different types of molecular markers have been developed: RAPD and AFLP.

### **6.7.1 PCR Arbitrary Priming Techniques**

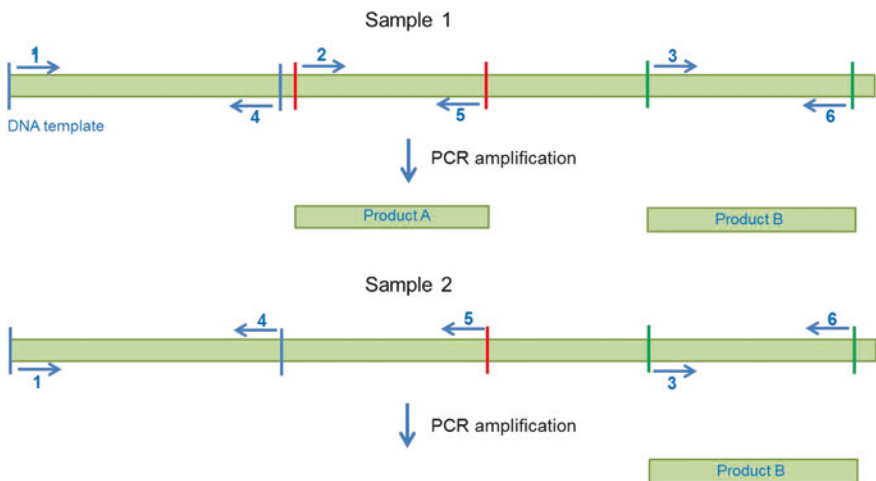
In the first category a number of closely related techniques have been developed and jointly referred to as Multiple Arbitrary Amplicon Profiling (MAAP) (Caetano-Anolles 1994). Even if, among these, RAPD is the most commonly

used, other techniques can be included such as Arbitrary Primed PCR (AP-PCR) (Welsh and McClelland 1990) and DNA Amplification Fingerprinting (DAF) (Caetano-Anolles et al. 1991) differing from RAPDs for primer length, stringency of the conditions, and the method of separation and detection.

### 6.7.2 Random Amplified Polymorphic DNA (RAPD)

RAPDs have been deeply applied thanks to the fact that these kinds of markers do not require DNA probes or any types of sequence information for the design of the specific markers.

RAPDs were the first PCR-based molecular markers to be employed in genetic variation analyses (Welsh and McClelland 1990; Williams et al. 1991). RAPD markers consist of random amplification of genomic DNA using short primers (decamers) and separation of the obtained fragments. The use of short primers is necessary to increase the probability that, although the sequences are random, they are able to find homologous sequences suitable for annealing (Fig. 6.3). Thence, DNA polymorphisms are generated by rearrangements or deletions occurring at or between oligonucleotide primer binding sites along the genome. RAPD-PCR fingerprint has been successfully applied in dissecting genetic diversity among different species. RAPD markers show several advantages: (i) no prior sequence information is needed for designing the primers that can be used for different



**Fig. 6.3** Schematic representation of a Random Amplified Polymorphic DNA (RAPD) reaction. In order to obtain an amplification product, the primers must anneal in the right orientation, pointing toward each other and at a reasonable distance. The *arrows* represent the single primers and the direction indicates the direction in which DNA synthesis will occur. The *numbers* represent primer annealing sites on the target DNA

templates; (ii) RAPDs are simple, quick, and cost-effective especially if compared to RFLP (Williams et al. 1991; Bardakci 2001); (iii) the quantity of DNA to be used is very small being amplified by PCR. At the same time, RAPDs present some not insignificant disadvantages that include: (i) the very low repeatability and reliability of RAPD polymorphic profiles (Vos et al. 1995); (ii) RAPDs, being dominant, cannot be used to distinguish homozygote from heterozygote genotypes in F<sub>2</sub> populations; (iii) nonspecific and therefore non-reproducible binding of primers occurring, insomuch as even a small difference in annealing temperature is sufficient to produce different patterns.

Some variants of RAPD markers have been independently developed named AP-PCR and DAF. They differ from RAPDs essentially in primer length, the stringency conditions, and the method of separation/detection of the fragments. With AP-PCR (Welsh and McClelland 1990), a single primer 10–15 nucleotides long is employed with an initial amplification of two PCR cycles at low stringency. Thereafter, the remaining cycles are carried out at higher stringency by increasing the annealing temperatures.

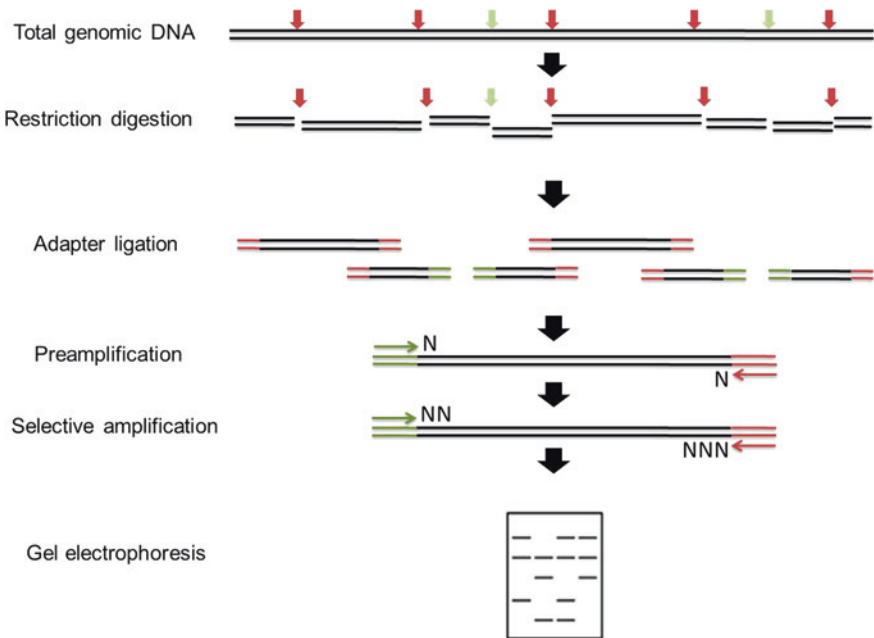
RAPDs have been used for many purposes, ranging from studies at the individual level (e.g., genetic identity) to studies involving closely related species. RAPDs have also been applied in gene mapping studies to fill gaps not covered by other markers (Williams et al. 1990; Hadrys et al. 1992).

Moreover, thanks to the speed and efficiency of RAPD analysis, high-density genetic mapping in many plant species such as faba bean (Torres et al. 1993), alfalfa (Kiss et al. 1993), and apple (Hemmat et al. 1994) were developed in a relatively short times. The RAPD analysis of non-isogenic lines (NILs) has been successfully employed in identifying markers linked to disease resistance genes in common bean (*Phaseolus vulgaris*) (Adam-Blondon et al. 1994), tomato (*Lycopersicon* sp.) (Martin et al. 1991), and lettuce (*Lactuca* sp.) (Paran et al. 1991).

### **6.7.3 Amplified Fragment Length Polymorphism (AFLP)**

Considered an intermediate between RFLPs and RAPDs methodologies, AFLP technique, developed by the Dutch company, Keygene (Zabeau and Vos 1992) combines the power of RFLP with the flexibility of PCR-based technology. AFLP analysis is based on the combination of the main analysis techniques: DNA digestion using restriction endonuclease enzymes and PCR technology. The AFLP protocol consists of DNA digestion using two different restriction enzymes (typically EcoRI and MseI) (Fig. 6.4), ligation of adapters to the extremity of the restriction fragments, DNA preamplification of ligated product using primers complementary to the adapter and restriction site sequences, DNA amplification of a subset of restriction fragments using selective AFLP primers, and separation and detection of the produced patterns, scoring fragments as either presence or absence among samples. The primer pairs used for AFLP usually produce 50–100 bands per assay. The number of amplicons per AFLP assay is a function of the number of selective



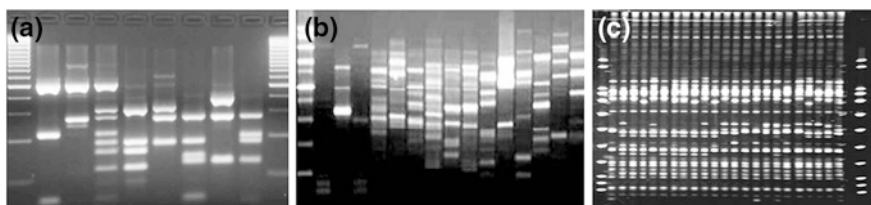


**Fig. 6.4** Different steps of Amplified Fragment Length Polymorphism (AFLP). Genomic DNA is digested with two restriction enzymes and adaptors are ligated to these ends. The first PCR (preamplification) is performed with a single-bp extension, followed by a more selective primer with up to a 3-bp extension. *N* nucleotide

nucleotides in the AFLP primer combination, the selective nucleotide motif, GC content, and physical genome size and complexity. In particular, AFLP polymorphisms can be produced in different ways: (i) insertions, duplications, or deletions inside amplification fragments; (ii) mutations of sequences flanking the restriction sites and complementary to the extension sites of the selective primers enabling possible primer annealing; (iii) mutations in the restriction site able to create or delete it. All these mutations can bring to an appearance/disappearance of a particular fragment or to the modifications (increase or decrease) of an amplified-restricted fragment.

AFLP generates fingerprints of any DNA regardless of its source, and without any prior knowledge of DNA sequence. Most AFLP fragments correspond to unique positions on the genome and hence can be exploited as landmarks in genetic and physical mapping. The technique can be used to distinguish closely related individuals at the subspecies level (Althoff et al. 2007) and can also map genes.

This technique, being PCR based requires no probe or previous sequence information as needed by RFLP. It is sufficiently reliable because of high stringent PCR in contrast to RAPD's problem of low reproducibility. However, the major advantage of AFLPs is the large number of polymorphisms scored. In fact, AFLP seems



**Fig. 6.5** Comparison among different amplification profiles obtained after PCR reactions and staining on ethidium bromide agarose gel: **a** RFLP profile; **b** RAPD profile, and **c** AFLP profile

to be much more efficient than the microsatellite loci in discriminating the source of an individual among putative populations. Similar to RAPD, AFLP analysis allows screening of many loci within the genome in a relatively short time and in an inexpensive way. The weak points of this technique are that this methodology is difficult to analyze due to the large number of unrelated fragments produced and that they are dominant markers.

Nevertheless, their high genomic abundance and generally random distribution throughout the genome make AFLPs a widely valued technology which has been successfully employed for DNA fingerprinting in barley (Becker et al. 1995; Simons et al. 1997), rice (Waugh et al. 1997), in einkorn wheat (Heun et al. 1997), for gene mapping studies (Mackill et al. 1996; Vos et al. 1995; Qi et al. 1998), and for QTL analysis (Powell et al. 1996; Nandi et al. 1997). AFLP markers have been successfully also used for analyzing genetic diversity in some other plant species such as peanut (Herselman 2003), soybean (Ude et al. 2003), and maize (Lübberstedt et al. 2000) (Fig. 6.5).

## 6.8 Sequence-Specific PCR-Based Markers

The alternative approach to arbitrary PCR amplification consists in the amplification of target regions of the genome using specific primers. In particular, with the advent of high-throughput sequencing technologies, abundant information on DNA sequences of many plant species is now available (Goff et al. 2002; Yu et al. 2002; Arabidopsis Genome Initiative 2000).

### 6.8.1 Expressed Sequence Tags (EST)–SSR

Expressed Sequence Tags (ESTs) are single-read sequences produced from partial sequencing of a bulk mRNA pool that has been reverse transcribed into cDNA (Putney et al. 1983). High-throughput sequencing produces information on thousands of ESTs and the new sequences are promptly accessible in the different

databases, increasing the growing information on gene expression. EST libraries provide a snapshot of the genes expressed in the tissue at the time of, and under the conditions in which, they were sampled (Bouck and Vision 2007). Despite the several advantages that these kinds of markers show, however, EST-SSRs are not without weak points. At first, the possibility to have null alleles, which compromise the amplification due to primer site variation, resulting in the lacking of visible amplicons. The second that being cDNA lacking of introns, unrecognized intron splice sites could disrupt primer annealing sites making impossible the amplification. Lastly, being EST-SSRs placed within genes and thus more conserved across species, they may be less polymorphic than anonymous SSRs. However, on the contrary, many advantages derive from the fact that ESTs are an inexpensive source for identifying gene-linked markers with higher levels of polymorphism, which can also be applied to closely related species in many cases (Cordeiro et al. 2001; Vasemagi 2005; Karaïskou 2008).

### ***6.8.2 Microsatellite-Based Marker Technique***

Microsatellites or Simple Sequence Repeats (SSR) are sequences constituted by sets of repeated motifs found within eukaryotic genomes (Dietrich et al. 1992; Bell and Ecker 1994; Morgante and Olivieri 1993). These sequences comprise basic short motifs (generally between 2 and 6 base pairs long) tandemly repeated several times. Thence, the polymorphisms associated with a specific locus are due to the variation in length of the microsatellite sequence depending on the number of repetitions of the basic motif. The flanking regions of the repeated sequences are mostly conservative and the repetition motifs are highly variable between different species and even different individuals of the same species. In fact, microsatellite assays permit to identify extensive interindividual length polymorphisms during PCR analysis of unique loci using discriminatory primers sets.

Variations in the number of tandemly repeated units are mainly due to polymerase strand slippage occurring during DNA replication where the repeats allow matching, via excision or addition, of repeats (Schlotterer and Tautz 1992). Being the polymerase slippage more probable with respect to point mutations, microsatellite loci tend to be hypervariable.

Microsatellites are among the most used genetic markers for different advantages: (i) they show co-dominant inheritance, (ii) are highly widespread into the genome, (iii) are highly sensible to detect an enormous extent of allelic diversity, (iv) are easy to use and highly reproducible, and (v) different microsatellites can be multiplexed in PCR and automation is possible. However, the development of microsatellites requires preventive and extensive knowledge of DNA sequences. Moreover, sometimes they tend to underestimate genetic structure measurements, hence they have been developed primarily for agricultural species, rather than wild species. Nevertheless, they are not free from disadvantages because: (i) they are time-consuming and expensive to develop; (ii) the heterozygotes may be misclassified as

homozygotes when null alleles occur because of mutations in the primer annealing sites; (ii) stutter bands may complicate accurate scoring of polymorphisms, and even if microsatellite markers are able to identify neutral biodiversity, nevertheless do not provide information about functional trait biodiversity.

The main molecular markers based on assessment of variability generated by microsatellites sequences are: Sequence Tagged Microsatellite Site (STMSs), Simple Sequence Length Polymorphism (SSLPs), Single-Nucleotide Polymorphisms (SNPs), Sequence Characterized Amplified Region (SCARs), and Cleaved Amplified Polymorphic Sequences (CAPS). Moreover, some new markers have recently emerged and are being used in the evaluation of PGR; these include high-density SNP arrays, whole-genome sequencing, and DNA barcoding.

In the main, microsatellite markers detect a high level of polymorphism and being very informative are currently used for population genetics studies due to the capability to be suitable both for the individual level and for closely related species. Microsatellite markers have proven useful for assessment of genetic variation in germplasm collections (Mohammadi and Prasanna 2003). The trend analysis of SSR repeats in genes of known function has permitted to use these markers' typology for association studies with phenotypic variation and biological function (Ayers et al. 1997). Several studies have demonstrated the usefulness of SSRs for estimating genetic relationship and for the detection of functional diversity in relation to adaptive variation (Eujayl et al. 2001; Russell et al. 2004). Microsatellites have been successfully applied also in gene mapping studies (Hearne et al. 1992; Morgante and Olivieri 1993; Jarne and Lagoda 1996).

### ***6.8.3 Single Nucleotide Polymorphisms (SNPs)***

The complications found to fully automate microsatellite genotyping and the advent of NGS has renewed the interest of the scientific community in a new type of marker named SNPs. SNPs are the most abundant molecular markers in the genome and consist of single nucleotide variations in genome sequence. SNPs polymorphisms derive from single nucleotide substitutions (transitions/transversions) or single nucleotide insertions/deletions. They are widely dispersed throughout the genomes with a variable distribution among species and are usually more prevalent in the noncoding regions of the genome where their effects are neutral. Nevertheless, when an SNP occurs within the coding regions, it can generate either synonymous mutations that do not alter the amino acid sequence but also non-synonymous mutations resulting in an amino acid sequence changing (Sunyaev et al. 1999). Synonymous changes can modify mRNA splicing generating phenotypic differences (Richard and Beckman 1995). Moreover, a group of associated SNP loci located on a certain region of the chromosome can form one SNP haplotype. SNPs, distributed in both coding and noncoding regions of genomes, represent key players in the process of population genetic variations and species evolution (Sylvänen 2001).

The majority of SNP genotyping analyses are based on: allele-specific hybridization, oligonucleotide ligation, primer extension, or invasive cleavage (Sobrinho et al. 2005). These kinds of markers can be easily detected using traditional PCR and sequencing, High Resolution Melting (HRM) technology, microchip arrays, and fluorescence technology. These genotyping methods are particularly attractive for their high data throughput and for their suitability for automation.

SNPs can be considered as the third-generation molecular markers coming after RFLPs and SSRs (Peter 2001). To date, SNP markers are not yet routinely applied in gene banks activity, in particular because of the high costs involved, even if they have been successfully applied to investigate genetic variation among different species (Brooks et al. 2010; Amaral et al. 2008). On the contrary, SNP analysis has revealed to be particularly useful for cultivar discrimination in crops where it is difficult to find polymorphisms. SNPs may also be used for a wide range of purposes, including population structure, genetic differentiation, and construction of ultra high-density genetic maps to saturate linkage maps in order to locate relevant traits in the genome. For instance, a high-density linkage map developed in *Arabidopsis thaliana* was completed only after SNP markers development (Cho et al. 1999). Moreover, linkage disequilibrium (LD) among different SNPs can be utilized for association analysis. Furthermore, SNPs can produce information concerning population diversity and evolution (origins, differentiation, and migrations) via SNP haplotypes among different populations. Compared with previous markers, SNPs show the following advantages because they are:

- abundant and widely distributed throughout the entire genome;
- characterized by a high genetic stability, excellent repeatability, and high accuracy;
- they lend to automation and fast high-throughput genotyping;
- being co-dominant are able to distinguish heterozygote from homozygote alleles.

#### **6.8.4 SNP Markers and Whole-Genome Sequencing**

One disadvantage of SNP markers consist in the low level information obtained respect to the highly polymorphic microsatellite markers. Nevertheless, this inconvenience can be compensated employing a higher numbers of markers (SNP chips) and whole-genome sequencing (Werner et al. 2002, 2004). Thanks to the improvement of sequencing technology with the advent of high-throughput sequencing, whole-genome/gene sequencing has permitted the detection and characterization of genetic diversity among individuals. Nowadays, it can be considered the most straightforward method producing more complete information on the genetic variation among different populations going to detect all the variations within the genome. However, even if a problem with whole-genome sequencing consists in the development of a high-throughput data analysis platform, the in-depth analysis of NGS data, extensively produced by genetics and genomics studies, has strongly increased the accurate calling of SNPs and genotypes thanks also to the development of recent statistical methods able to improve and quantify the

considerable uncertainty associated with genotype calling. Before the advent of NGS, SSR markers were developed using the time-consuming and laborious construction of genomic libraries, starting from recombinant DNA with the consequent isolation and sequencing of clones containing the SSRs. Zalapa et al. (2012) have demonstrated the power of NGS for developing SSRs in plants in a review focusing on their work on cranberry and several other studies where SSRs were developed using Sanger, 454, and Illumina platforms.

## 6.9 Markers Based on Other DNA Typology

Ribosomal RNA (rRNA) represents another kind of nuclear genome and, due to the fact that some regions of rRNA are well preserved in eukaryotes, has been extensively employed to study genetic diversity. rRNA genes are placed on the specific chromosomal loci, and organized in tandem repeats which can be repeated up to thousands of times. A particular feature of rRNA, which could explain its wide application, consists in the contemporary presence of regions that are highly conserved throughout eukaryotic evolution providing very useful genetic tools and other regions called “Internal Transcriber Spacers” (ITS) that are highly variable and hence can be used to detect polymorphisms at intraspecific level.

Other highly informative approaches exist, based on organelle microsatellite sequences detection. Due to their uniparental mode of transmission, chloroplast (cpDNA) and mitochondrial genomes (mtDNA) allow to detect different patterns of genetic differentiation with respect to nuclear alleles (Provan et al. 1999a, b). Consequently, in addition to nuclear markers, other marker typologies based on chloroplast and mitochondrial microsatellites have also been developed. The cpDNA, which is maternally inherited in most plants, can be considered an additional tool for within-species genetic variation analysis (Ali et al. 1991; McCauley 1994) and has proved to be a powerful tool for phylogenetic studies. Thanks to its good level of conservation within the genome, CpDNA has been employed widely for studying plant populations through the use of PCR–RFLP and PCR sequencing approaches (McCauley 1994), in the detection of hybridization/introgression (Bucci et al. 1998), in the analysis of genetic diversity (Clark et al. 2000), and in obtaining the phylogeography of plant populations (Parducci et al. 2001; Shaw et al. 2005). On the contrary, mitochondrial DNA in plants, being quantitatively scarce, is unsuitable for studying phylogenesis and genetic diversity.

### 6.9.1 RNA-Based Molecular Markers (RBMs)

Biological responses and the developmental programming in organisms are crucial phenomena, thence the analysis of mechanisms which control their genetic

expression are essential. This has led to the development of markers derived from transcribed/expressed regions of genomes. The greatest advantage of RBMs is that, being derived from the expressed regions of the genome, the generated fragments can easily be associated with phenotypic traits becoming a key tool for genetic mapping studies of Quantitative Traits Loci (QTLs). On the contrary, these markers should be used with caution in such studies aiming to detect genetic variation in natural populations because they may be under selection. RNA-based markers, designed on coding regions of the genome characterized by a good level of conservation, are also expected to be transferable between related species and genera. Among PCR-based marker techniques, inter small RNA polymorphism (iSNAP) is the most recent and is based on endogenous noncoding small RNAs consisting of 20–24 nucleotides that are ubiquitous in eukaryotic genomes where they play important regulatory roles, representing an excellent source for molecular marker development. This technique is highly reproducible and feasible for automation and it has been successfully applied for genome mapping and for genotyping. Nevertheless, a negative point is that being based on the expressed portion of the genome, it could be also affected by phenological plant stage and environmental conditions. Other techniques such as cDNA–SSCP, cDNA–AFLP, cDNA–RFLP, and RAP–PCR are used for differential RNA studies, using selective amplification of cDNA. These techniques are efficient for the identification of common and rare transcripts and for studying genome-wide gene expression (Xiao et al. 2009) and can also be used to identify differences in the expression of different genes under various stress conditions (Song et al. 2012). Moreover, another RBM technique exists consisting in EST–SSR markers where thanks to the recent increase in the availability of EST data, have been developed in a number of plant species groups (La Rota and Sorrells 2004). Technically, EST–SSR is identical to common genomic (gSSR) microsatellites in terms of amplification and detection but differs in primer development and the locations of the primers being generated from the transcribed portion of the genome.

### ***6.9.2 Transposable Elements-Based Molecular Markers***

Transposable elements (TE) are mobile DNA sequences which can change their positions in the genome. Based on their excision mechanism, TEs can be divided into Class I (retrotransposons), commonly called ‘copy-and-paste’ elements, and Class II (DNA transposons), or ‘cut-and-paste’ elements (Finnegan 1989). In particular, LTR retrotransposons are elements surrounded by long terminal repeats (LTRs) that do not code for any protein and contain the promoters and terminators for transcription. These regions provide the basis for primer binding sites in many techniques. Retrotransposons represent an excellent basis for the development of markers due to their dispersion (Katsiotis et al. 1996; Suoniemi et al. 1996), ubiquity (Flavell et al. 1992; Voytas et al. 1992), and prevalence in plant genomes; for this reason most TE-based markers utilize Class I retrotransposons.

Even if transposon insertions can be deleterious for host genomes, transposons are actually considered crucial for adaptative evolution favoring the rearrangement of the genomes and the acquiring of novel traits (Miller et al. 1997; Agrawal et al. 1998; May and Dellaporta 1998; Girard and Freeling 1999; Gray 2000). Despite their great contribution to the genome structure, size, and variation, only recently retrotransposons have received attention for the assessment of genetic diversity (Gynheung et al. 2005) where retrotransposons can be used alone or in combination with other markers, such as AFLPs and SSRs. Retrotransposon-based molecular analysis relies on amplification using a primer corresponding to the retrotransposon and a primer matching a section of the neighboring genome. To this type of class of molecular markers belong: Sequence-Specific Amplified Polymorphism (S-SAP), Inter-Retrotransposon Amplified Polymorphism (IRAP), Retrotransposon-Microsatellite Amplified Polymorphism (REMAP), Retrotransposon-Based Amplified Polymorphism (RBIP), and finally, Transposable Display (TD).

## **6.10 Optimization of Molecular Marker-Based Analysis: Multiplex PCR**

Through multiplex PCR system it is possible to contemporarily detect multiple target sequences using simultaneous amplification reactions (James et al. 2003). Multiplex PCR presents many advantages being more sensitive, fast, and easy to perform. The multiplex-ready PCR technology provides several enabling advances in marker genotyping reducing assay costs, increasing information throughput and permitting automation. It requires limited sample concentration, makes it possible to obtain more information per unit of time and using standardized protocols, economizes on reagents, enzyme, buffers and labor, streamlines data analysis, and has a high tolerance to variation in the concentration and quality of DNA samples. Moreover, multiple-tube amplification permits to avoid allelic dropout consisting in an erroneous classification of one locus as homozygous due to the chance amplification of only one of the two heterozygous alleles, and false alleles due to reaction contaminations, PCR slippage artifacts, or other causes (Taberlet et al. 1996, 1999; Broquet and Petit 2004). However, multiplex PCR reactions require several devices such as uniformity in product abundance, especially for simultaneous SSRs and SNPs genotyping, and differential sizes of the amplification fragments obtained in order to connect a specific allele to the marker that characterizes it. In particular, multiplex amplifications using fluorescence detection show high power of discrimination in a single test and permit to jointly analyze up to 10 different genomic loci. This technique has been successfully applied in high-throughput SNP genotyping, gene deletion, mutation, and linkage analysis.

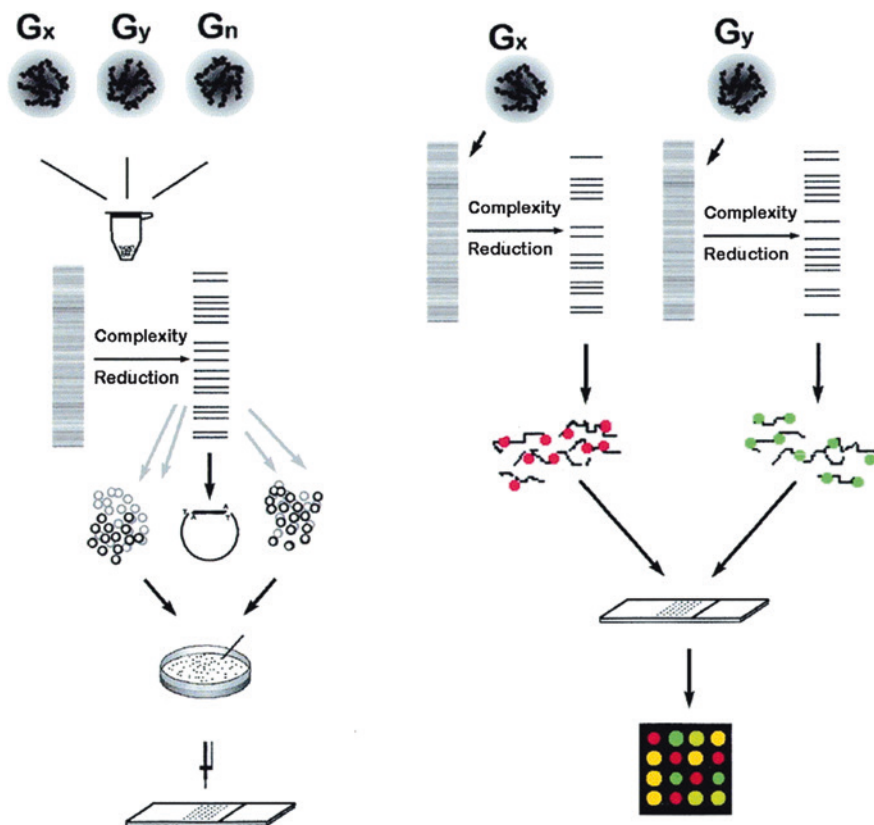


## 6.11 DNA Barcoding Markers

With the advent of practical computer technologies applied to genetic studies, such new identification technologies have been developed to facilitate the analysis in the presence of an increasing number of samples. Among these, barcoding system is an automatic scanning identification tool that has been applied by biological taxonomists to species classification, referring to a DNA barcode. In particular, a DNA barcode is a short DNA sequence deriving from a standardized region of the genome used for identifying species. DNA barcoding permits using a large-scale screening of one or more reference genes, to assign an unknown individual to an exact specie, and enhance discovery of new species (Hebert et al. 2003; Stoecklem 2003). In this perspective, public libraries of DNA barcodes linked to named specimens are available (Tautz et al. 2002; Hebert et al. 2004). Compared with time-consuming and inefficient traditional morphological classification (Huang et al. 2007), DNA barcoding presents several advantages being very fast and having a high accuracy of 97.9 % (Hajibabaei et al. 2006). On the contrary, in DNA barcoding technique, the genome fragments are difficult to obtain and being relatively conserved have no enough variations.

## 6.12 Diversity Arrays Technology (DArT)

DArT is a genotyping technology developed to overcome some of the limitations of other molecular marker technologies such as RFLP, AFLP, and SSR (Akbari et al. 2006). DArT represents a fast and cost-effective alternative method to time-consuming hybridization-based techniques, characterizing simultaneously several thousand loci in a single assay. DArT has been successfully applied to genotyping polyploid species with large genomes, such as wheat. This technology generates whole-genome fingerprints by scoring the presence/absence of DNA fragments in genomic representations and acts by reducing the complexity of a DNA sample to obtain a “representation” of that sample. DArT technology consists of several steps: (i) library creation, (ii) microarray of libraries onto glass slides, (iii) hybridization of fluoro-labeled DNA onto slides, (iv) scanning of slides for hybridisation signal, and (v) data analysis (Fig. 6.6). Among the methods used for DNA complexity reduction, the main method consists of a combination of restriction enzyme digestion and adapter ligation, followed by amplification even if an infinite range of alternative methods can be used. DArT markers for new specie are produced by screening a library deriving from a genomic representation prepared starting from a pool of DNA samples that embrace the diversity of the specie. Thanks to the use of the microarray platform, the discovery process results as more efficient being all markers scored simultaneously, and for each reduction method an independent collection of DArT markers can be assembled on a separate DArT array. The number of markers to use for the analysis of a given species



**Fig. 6.6** Schematic drawing of DARt pipeline. G<sub>x</sub>, G<sub>y</sub>, and G<sub>n</sub> represent DNA from three different individuals in the reduction step to obtain single genomic DNA

is only dependent on the level of genetic variation within the species (or gene pool) and the number of complexity reduction methods screened. DARt technology was originally developed in rice due to its small genome (430 Mbp) (Jaccoud et al. 2001) and subsequently applied to several other crops. To date, DARt has been successfully applied for genetic mapping and genetic diversity analysis, also to species characterized by large genomes such as wheat and barley, (Mochida et al. 2004; Wenzl et al. 2004) up to the 16,000 Mbp of the hexaploid genome of bread wheat (Akbari et al. 2006).

### 6.13 Next-Generation Sequencing Technologies

In the past decade, the emergence of NGS technologies has deeply changed all the genetics disciplines that depend on DNA sequence data. NGS technologies have revolutionized and increased the capabilities of traditional Sanger sequencing

method (Sanger et al. 1977), allowing millions of bases to be sequenced in one round at a fraction of the cost. NGS techniques can be distinguished into three main types: sequencing by synthesis, sequencing by ligation, and single-molecule sequencing.

### ***6.13.1 Sequencing by Synthesis***

Like Sanger sequencing, NGS techniques use the emission of chemiluminescence created by nucleotide incorporation during synthesis of the complementary DNA strand by DNA polymerase, to determine base composition. In sequencing by synthesis, DNA is fragmented to obtain the appropriate size, ligated to adaptor sequences, and then amplified to enhance the fluorescent or chemical signal. Templates are then separated and immobilized in preparation for flow-cell cycles. Among the techniques available for sequencing by synthesis the most used are *Illumina* (<http://www.illumina.com>), *Roche 454 pyrosequencing* (<http://www.my454.com>), and *Ion torrent* (<http://www.iontorrent.com>), which differ by read length and in how templates are amplified and immobilized.

### ***6.13.2 Sequencing by Ligation***

This method is based on the use of oligonucleotide probes which differ in lengths and labeled with fluorescent tags depending on the nucleotide types to be determined (Landegren et al. 1988). The DNA template is fragmented and primed with a short, known anchor sequence favoring the probe hybridization and consequently DNA ligase is added. The fluorescent emission is analyzed to determine which probe was incorporated. This process is repeated with different sets of probes to query the DNA template and assess the sequence of nucleotides. Among the methods based on this technique the most used are *SOLiD* (<http://www.appliedbiosystems.com>) and *Polonator G.007 system* (<http://www.azcobiotech.com/instruments/polonator.php>).

### ***6.13.3 Single-Molecule Sequencing***

Single-molecule sequencing (SMS) technique, also called “third-generation sequencing,” is based on the detection of a chemiluminescent signal produced by nucleotide incorporation occurring during DNA sequencing from a single nucleic acid molecule. This method offers several advantages with respect to other NGS methods because it can make use of degraded or low concentrations of starting material and escape from PCR errors due to template amplification.

Presently, the main techniques based on this method are *Helicos Genetic Analysis System* (<http://www.helicosbio.com>) and PacBio RS SMS platform (<http://www.pacificbiosciences.com>).

## 6.14 Conclusion and Prospects

The idea of using gene markers for a variety of purposes in applied genetics, conservation strategies, and genetic diversity assessment is not new. However, until the advent of molecular markers, many of the proposals were technically unfeasible. Molecular analysis of plants has found many applications in plant improvement, in the management of plant production, and in conservation of plant resources. Molecular tools have become key contributors to the management of wild plant populations helping to conserve biodiversity.

Recent dramatic advances in DNA sequencing are now providing cost-effective options for the discovery of very large numbers of markers for any plant species. These developments significantly change the approach to marker discovery and analysis in plants and greatly expand the potential range of application. Advances in biotechnology have resulted in a large variety of molecular marker systems and enhanced opportunities for automation of the majority of the techniques, resulting in a wealth of information. Moreover, due to the developments in the detection techniques, molecular markers are particularly useful in diagnostic applications, such as the screening of samples for the presence or incorporation of favorable traits, the detection of pathogens and diseases in plants, and the screening of plant material for the presence of transgenic elements and jointly with the concept of marker-assisted selection provide new solutions for selecting and maintaining desirable genotypes.

Hence, molecular markers make the prospect excellent for a rapid development of new methodologies for plant genetic diversity dissection that take advantage of the modern techniques.

## References

- Adam-Blondon A, Seignac M, Bannerot H et al (1994) SCAR, RAPD and RFLP markers tightly linked to a dominant gene (Are) conferring resistance to anthracnose in common bean. *Theor Appl Genet* 88:865–870
- Agrawal A, Eastmann Q, Schatz D (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394:744–751
- Akbari M, Wenzl P, Caig V et al (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor Appl Genet* 113:1409–1420
- Alberte RS, Suba GK, Procaccini G et al (1994) Assessment of genetic diversity of seagrass populations using DNA fingerprinting: implications for population stability and management. *Proc Nat Acad Sci USA* 91:1049–1053

- Ali IF, Neale DB, Marshall KA (1991) Chloroplast DNA restriction fragment length polymorphism in *Sequoia sempervirens* D. Don Endl., *Pseudotsuga menziesii* (Mirb.) Franco, *Calocedrus decurrens* (Ton), and *Pinus taeda* L. *Theor Appl Genet* 81:83–89
- Althoff DM, Gitzendanner MA, Segraves KA (2007) The utility of amplified fragment length polymorphisms in phylogenetics: a comparison of homology within and between genomes. *Syst Biol* 56:477–484
- Amaral AJ, Megens HJ, Crooijmans RPMA et al (2008) Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179:569–579
- Antonious K, Nybom H (1994) DNA fingerprinting reveals significant amounts of genetic variation in a wild raspberry *Rubus idaeus* population. *Mol Ecol* 3:177–180
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Ayers NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. *Theor Appl Genet* 94:773–781
- Barcaccia G, Albertini E, Rosellini D (2000) Inheritance and mapping of 2n-egg production in diploid alfalfa. *Genome* 43:528–537
- Bark OH, Havey MJ (1995) Similarities and relationships among populations of the bulb onion as estimated by RFLPs. *Theor Appl Genet* 90:407–414
- Becker J, Vos P, Kuiper M et al (1995) Combined mapping of AFLP and RFLP markers in barley. *Mol Gen Genet* 249:65–73
- Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* 19:137–144
- Besse P, Seguin M, Lebrun P et al (1994) Genetic diversity among wild and cultivated populations of *Hevea brasiliensis* assessed by nuclear RFLP analysis. *Theor Appl Genet* 88:199–207
- Botstein D, White RL, Skolnick M et al (1980) Construction of a genetic map in man using restriction fragment length polymorphisms. *Amer J Hum Genet* 32:314–331
- Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Mol Ecol* 16:907–924
- Brooks SA, Gabreski N, Miller D et al (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for lavender foal syndrome. *PLoS Genet* 6:e1000909
- Broquet T, Petit E (2004) Quantifying genotyping errors in noninvasive population genetics. *Mol Ecol* 13:3601–3608
- Brown AHD (1979) Enzyme polymorphism in plant populations. *Theor Pop Biol* 15:1–42
- Brubaker CL, Wendel JF (1994) Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). *Amer J Bot* 81:1309–1326
- Bucci G, Anzidei M, Madaghiele A et al (1998) Detection of haplotypic variation and natural hybridization in halepensis complex pine species using chloroplast simple sequence repeat (SSR) markers. *Mol Ecol* 7:1633–1643
- Caetano-Anolles G (1994) MAAP: a versatile and universal tool for genome analysis. *Plant Mol Biol* 25:1011–1026
- Caetano-Anolles G, Bassam BJ, Gresshoff PM (1991) DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotech* 9:553–557
- Castagna R, Maga G, Perenzin M, Heun M (1994) RFLP-based genetic relationships of Einkorn wheats. *Theor Appl Genet* 88:818–823
- Cho RJ, Mindrinos M, Richards DR et al (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genet* 23:203–207
- Clark CM, Wentworth TM, Malley DMO (2000) Genetic discontinuity revealed by chloroplast microsatellites in eastern North American *Abies* (Pinaceae). *Am J Bot* 87:774–778
- Clausen AM, Spooner DM (1998) Molecular support for the hybrid origin of the wild potato species *Solanum × rochei*. *Crop Sci* 38:858–865

- Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ et al (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *erianthus* and sorghum. *Plant Sci* 160:1115–1123
- D'Hont A, Paulet F, Glaszmann JC (2002) Oligoclonal interspecific origin of "North Indian" and "Chinese" sugarcanes. *Chromosome Res* 10:253–262
- Dubreuil P, Dufour P, Krejci E et al (1996) Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. *Crop Sci* 36:790–799
- Deu M, Gonzales-de Leon JC, Glaszmann I, Degremont J et al (1994) RFLP diversity in cultivated sorghum in relation to racial differentiation. *Theor Appl Genet* 88:838–844
- Dietrich W, Katz H, Lincoln SE et al (1992) A genetic map of mouse suitable for typing intraspecific crosses. *Genetics* 131:423–447
- Eujayl I, Sorrells M, Baum M, Wolters P, Powell W (2001) Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. *Euphytica* 119:39–43
- Fang DQ, Roose ML, Krueger RR et al (1997) Fingerprinting trifoliate orange germplasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theor Appl Genet* 95:211–219
- Ferguson-Smith MA (1997) Genetic analysis by chromosome sorting and painting: Phylogenetic and diagnostic applications. *Eur J Hum Genet* 5:253–265
- Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trend Genet* 5:103–107
- Flavell A, Smith D, Kumar A et al (1992) Extreme heterogeneity of Ty1-copia group retrotransposons in plants. *Mol Genet Genom* 231:233–242
- Girard L, Freeling M (1999) Regulatory changes as a consequence of transposon insertion. *Dev Genet* 25:291–296
- Goff SA, Ricke D, Lan TH et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100
- Gonzalez AA, Wong A, Delgado-Salinas R, Gepts P (2005) Assessment of inter simple sequence repeat markers to differentiate sympatric wild and domesticated populations of common bean. *Crop Sci* 45:606–615
- Gray Y (2000) It takes two transposon to tango. *Trends Genet* 16:461–468
- Greene SL, Morris JB (2001) The case for multiple-use plant Germplasm collections and a strategy for implementation. *Crop Sci* 41:886–892
- Guarino L (1999) Approaches to measuring genetic erosion. Proceedings of the technical meeting on the methodology of the FAO World Information and Early Warning System on Plant Genetic Resources, Prague June 1999
- Gynheung A, Dong-Hoon J, Ki-Hong J, Sichul L (2005) Reverse genetic approaches for functional genomics of rice. *Plant Mol Biol* 59:111–123
- Hadrys H, Balick M, Schierwater B (1992) Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. *Mol Ecol* 1:55–63
- Hajibabaei M, Janzen DH, Burns JM (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Nat Acad Sci USA* 103:968–971
- Hamrick JL, Godt MJW (1997) Allozyme diversity in cultivated crops. *Crop Sci* 37:26–30
- Harlan JR, de Wet JMJ (1971) Toward a rational classification of cultivated plants. *Taxon* 20(4):509–517
- Hearne CM, Ghosh S, Todd JA (1992) Microsatellites for linkage analysis of genetic traits. *Trend Genet* 8:288–294
- Hebert PDN, Cywinska A, Ball SL et al (2003) Biological identifications through DNA barcodes. *Proc R Soc Biol Sci* 270:313–321
- Hebert PDN, Penton EH, Burns JM (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Nat Acad Sci USA* 101:14812–14817
- Hemmat M, Weeden NF, Manganaris AG et al (1994) Molecular Marker linkage map for apple. *J Hered* 85:4–11

- Herselman L (2003) Genetic variation among Southern African cultivated peanut (*Arachis hypogaea* L.) genotypes as revealed by AFLP analysis. *Euphytica* 133(3):319–327
- Heun M, Schäfer-Pregl R, Klawan D et al (1997) Site of Einkorn wheat domestication identified by DNA fingerprinting. *Science* 278:1312–1314
- Huang J, Xu Q, Sun ZJ (2007) Identifying earthworms through DNA barcodes. *Pedobiologia* 51:301–309
- Jaccoud D, Peng K, Feinstein D et al (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucl Acid Res* 29:25
- James D, Schmidt A, Wall E et al (2003) Reliable detection and identification of genetically modified maize, soybean, and canola by multiplex PCR analysis. *J Agric Food Chem* 51(20):5829–5834
- Jarne P, Lagoda PJ (1996) Microsatellites, from molecules to populations and back. *Trends Ecol Evol* 11(10):424–429
- Karakou N (2008) High degree of transferability of 86 newly developed zebra finch EST-linked microsatellite markers in 8 bird species. *J Hered* 99:688–693
- Karp A, Seberg O, Buiatti M (1996) Molecular techniques in the assessment of botanical diversity. *Ann Bot* 78:143–149
- Katsiotis A, Schmidt T, Heslop-Harrison J (1996) Chromosomal and genomic organization of Ty1-copia-like retrotransposon sequences in the genus *Avena*. *Genome* 39:410–417
- Kiss G, Csanadi G, Kalman K et al (1993) Construction of a basis genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. *Mol Gen Genet* 238:129–137
- La Rota M, Sorrells ME (2004) Comparative DNA sequence analysis of mapped wheat ESTs reveals the complexity of genome relationships between rice and wheat. *Funct Int Genom* 4:34–46
- Landegren U, Kaiser R, Sanders J, Hood L (1988) A ligase-mediated gene detection technique. *Science* 241:1077–1080
- Lanner HC, Gustafsson M, Falt AS et al (1996) Diversity in natural populations of wild *Brassica oleracea* as estimated by isozyme and RAPD analysis. *Genet Resour Crop Evol* 43:13–23
- Laurent V, Risterucci AM, Lanaud C (1994) Genetic diversity in cocoa revealed by cDNA probes. *Theor Appl Genet* 88:193–198
- Lübberstedt TH, Melchinger AE, Dussle D et al (2000) Relationships among early European maize inbreds: Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD, and pedigree data. *Crop Sci* 40:783–791
- Lynch M (1990) The similarity index and DNA finger printing. *Mol Biol Evol* 7:478–484
- Mackill DJ, Zhang Z, Redoña ED et al (1996) Level of polymorphism and genetic mapping of AFLP markers in rice. *Genome* 39:969–977
- Martin GB, Williams JGK, Tanksley SD (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. *Proc Natl Acad Sci USA* 88:2336–2340
- May BP, Dellaporta SL (1998) Transposon sequences drive tissue-specific expression of the maize regulatory gene R-s. *Plant J* 13:241–248
- McCauley DE (1994) Contrasting the distribution of chloroplast DNA and allozyme polymorphisms among local populations of *Silene alba*: implications for studying gene flow in plants. *Proc Nat Acad Sci* 17:8127–8131
- Miller JC, Tanksley SD (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet* 80:437–448
- Miller W, McDonald J, Pinsker W et al (1997) Molecular domestication of mobile elements. *Genetica* 100:261–270
- Mochida K, Yamazaki Y, Ogihara Y (2004) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet Genom* 270:371–377
- Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci* 43(4): 1235–1248
- Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. *Plant J* 3:175–182

- Mullis KB, Faloona FA, Scharf SJ et al (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symp Quant Biol* 51:263–273
- Nandi S, Subudhi PK, Senadhira D et al (1997) Mapping QTLs for submergence tolerance in rice by AFLP analysis and selective genotyping. *Mol Gen Genet* 255:1–8
- Neale DB, Williams CG (1991) Restriction fragment length polymorphism mapping in conifers and applications to forest genetics and tree improvement. *Can J Fors Res* 21:545–554
- Paran I, Kesseli R, Michelmore R (1991) Identification of restriction-fragment-length-polymorphisms and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near isogenic lines. *Genome* 34:1021–1027
- Parducci L, Szmidt AE, Madaghiele A et al (2001) Genetic variation at chloroplast microsatellites (cpSSRs) in *Abies nebrodensis* (Lojac.) Mattei and three neighboring *Abies* species. *Theor Appl Genet* 102:733–740
- Peter G (2001) An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med* 114:204–210
- Powell W, Thomas WTB, Baird E et al (1996) Analysis of quantitative traits in barley by the use of amplified fragment length polymorphisms. *Heredity* 79:48–59
- Primack RB (1993) *Essentials of conservation biology*. Sinauer Associates, Sunderland, MA, USA
- Provan J, Russell JR, Booth A et al (1999a) Polymorphic chloroplast simple-sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Mol Ecol* 8:505–511
- Provan J, Soranzo N, Wilson NJ et al (1999b) Low mutation rate for chloroplast microsatellites. *Genetics* 153:943–947
- Putney SD, Herlihy WC, Schimmel P (1983) A new troponin T and cDNA clones for 13 different muscle protein, found shot gun sequencing. *Nature* 302:718–721
- Qi X, Stam P, Lindhout P (1998) Use of locus-specific markers to construct a high density map in barley. *Theor Appl Genet* 96:376–384
- Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Cons Biol* 17:230–237
- Richard I, Beckman JS (1995) How neutral are synonymous codon mutations? *Nat Genet* 10:259
- Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, et al (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. *Genome* 47:389–398
- Saideswara Y, Rao S, Rao SA et al (1989) New evidence on the phylogeny basic chromosome number in Pennisetum. *Curr Sci* 58:869–872
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* 20:2211–2215
- Shaw J, Lickey EB, Beck JT et al (2005) The tortoise and Hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot* 92:142–166
- Simons G, Van der Lee T, Diergaarde P et al (1997) AFLP-based fine mapping of the *Mlo* gene to a 30-kb DNA segment of the barley genome. *Genomics* 44:61–70
- Sobrinho B, Briona M, Carracedoa A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154:181–194
- Song Y, Wang Z, Bo W et al (2012) Transcriptional profiling by cDNA-AFLP analysis showed differential transcript abundance in response to water stress in *Populus hopeiensis*. *BMC Genom* 13:286
- Southern E (1975) Detection of specific sequences among DNA fragments separated by gel-electrophoresis. *J Mol Biol* 98:503
- Spooner D, Van Treuren R, De Vicente MC (2005) Molecular markers for Genebank Management. IPGRI Technical Bulletin No. 10, 2005
- Stoecklem M (2003) Taxonomy, DNA, and the bar code of life. *Bioscience* 53:796–797
- Sumner AT (1990) *Chromosome banding*. Unwin Hyman, London UK
- Sunyaev S, Hanke J, Aydin A et al (1999) Prediction of nonsynonymous single nucleotide polymorphisms in human disease-associated genes. *J Mol Med* 77:754–760



- Suoniemi A, Anamthawat-Jonsson K, Arna T et al (1996) Retrotransposon BARE-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. *Plant Mol Biol* 30:1321–1329
- Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet* 2:930–942
- Taberlet P, Griffin S, Goossens B et al (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res* 24(16):3189–3194
- Tautz D, Arctander P, Minelli A et al (2002) DNA points the way ahead in taxonomy. *Nature* 418:479
- Torres AM, Millan T, Cubero JI (1993) Identifying rose cultivars using random amplified polymorphic DNA markers. *Hort Sci* 28:333–334
- Ude GN, Kenworthy WJ, Costa JM et al (2003) Genetic diversity of soybean cultivars from China, Japan, North America and North American ancestral lines determined by Amplified Fragment Length Polymorphism. *Crop Sci* 43:1858–1867
- Vasemagi A (2005) Expressed sequence tag (EST) linked microsatellites as a source of gene associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol Biol Evol* 22:1067–1076
- Viegas-Péquignot E, Dutrillaux B (1978) Une méthode simple pour obtenir des prophases et des prométaphases. *Ann Genet* 21:122–125
- Volis S, Mendlinger S, Turuspekov Y et al (2001) Allozyme variation in Turkmenistan populations of wild barley, *Hordeum spontaneum*. *Koch Ann Bot* 87:435–446
- Vos P, Hogers R, Bleeker M, Reijmans M et al (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 23:4407–4414
- Voytas DF, Cummings MP, Konieczny A et al (1992) Copia-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* 89:7124–7128
- Waugh R, Bonar N, Baird E et al (1997) Homology of AFLP products in three mapping populations of barley. *Mol Gen Genet* 255:311–321
- Weising K, Nybom H, Wolff K, Meyer W (1995) In: Arbor A (ed) DNA finger printing in plants and fungi. CRC Press, Boca Raton, p 1–3
- Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res* 18:7213–7218
- Wenzl P, Carling J, Kudrna D et al (2004) Diversity array technology (DArT) for whole-genome profiling of barley. *Proc Natl Acad Sci USA* 101:9915–9920
- Werner M, Herbon N, Gohlke H et al (2004) Asthma is associated with single-nucleotide polymorphisms in ADAM33. *Clin Exp Allergy* 34:26–31
- Werner M, Sych M, Herbon N et al (2002) Large-scale determination of SNP allele frequencies in DNA pools using MALDI-TOF mass spectrometry. *Hum Mutat* 20:57–64
- Williams JGK, Kubelik AR, Livak KJ et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl Acids Res* 18:6531–6535
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1991) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531–6535
- Xiao X, Li H, Tang C (2009) A silver-staining cDNA-AFLP protocol suitable for transcript profiling in the latex of *Hevea brasiliensis* (para rubber tree). *Mol Biotechnol* 42:91–99
- Yang F, Alkalaeva EZ, Perelman PL et al (2003) Reciprocal chromosome painting among human, aardvark and elephant (superorder Afrotheria) reveals likely eutherian ancestral karyotype. *Proc Natl Acad Sci USA* 100:1062–1066
- Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice Genome (*Oryza sativa* L. ssp. indica. *Science* 296:79–92
- Zabeau M, Vos P (1992) Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent EP0534858
- Zalapa JE, Cuevas H, Zhu H et al (2012) Using next-generation sequencing approaches for the isolation of simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot* 99:193–208
- Zouros E, Foltz DW (1987) The use of allelic isozyme variation for the study of heterosis. *Isozymes* 13:1–59