
3.1 Introduction

The development of restriction fragment length polymorphism (RFLP) marker system amply demonstrated that DNA sequence polymorphisms could be detected and used as molecular markers. It also highlighted the great abundance and genome-wide distribution of DNA-based markers, and the novel opportunities generated by this development in various genetic and other biological investigations. But RFLP technique requires considerable preparatory work, is technically demanding, and involves expensive reagents. Therefore, efforts were made to develop simpler, less expensive, and more convenient marker systems. These efforts led to the development of several polymerase chain reaction (PCR)-based marker systems during the 1990s, which are generally called *second-generation markers*. These markers have virtually replaced the first-generation hybridization-based markers as they require much smaller quantity of DNA of relatively lower quality and are much more user-friendly and amenable to automation. Simple sequence repeat (SSR), amplified fragment length polymorphism (AFLP), and randomly amplified polymorphic DNA (RAPD) markers are some of the widely used PCR-based markers. These marker systems became possible due the development of PCR procedure by Mullis and coworkers for amplification of specific DNA sequences from DNA samples of very high complexity. At the same time, refinements in chemical synthesis of

DNA ensured that PCR primers became readily available at a reasonable price. Finally, continued refinements in PCR technology enabled the PCR to become a routine laboratory technique. As a result, the PCR-based markers became greatly user-friendly and are very popular. Therefore, a brief description of chemical synthesis of oligonucleotides and PCR procedure precedes the discussion of various PCR-based marker systems.

3.2 Oligonucleotides

An *oligonucleotide*, “oligo” for short, is a short DNA fragment of few to several nucleotides (nt). Oligos are usually single-stranded, but they can also be double-stranded. Oligos are ordinarily chemically synthesized using automated oligonucleotide synthesizers. Khorana and coworkers synthesized a complete gene in 1970 using the phosphodiester method of DNA synthesis. This procedure was soon replaced by the more convenient and efficient phosphotriester approach; this method could synthesize up to 10–20 nt long oligos in a few days, and it was automated. But the present-day oligonucleotide synthesizers use the phosphite triester approach of DNA synthesis. This procedure takes 15 min for adding one nucleotide to the growing chain, and oligos as long as 50 nt can be prepared in good yields. It may be pointed out that the chemical synthesis of DNA proceeds from the 3' to the 5' direction

as compared to the progress of DNA replication from the 5' to the 3' direction.

Oligonucleotides have a variety of applications ranging from their use as primers to that as therapeutic agents. Oligonucleotide sequences of 12–20 nt are used as probes in nucleic acid hybridization for various purposes, including detection of DNA sequence polymorphisms. Oligonucleotides of different lengths and with specified/arbitrary sequences are used as primers for amplification of DNA fragments for the various PCR-based marker systems and for producing cDNA from RNA templates. Oligonucleotides are also used for DNA sequencing by DNA synthesis and for chemical synthesis of a complete gene that can be used for genetic transformation. In addition, oligos are used as linkers and adapters for modification of the cut ends of DNA fragments to facilitate their cloning and/or amplification.

3.3 Polymerase Chain Reaction

Kary Mullis (1990) conceived the idea of PCR in 1983 while thinking of novel approaches for DNA sequencing. Mullis and coworkers developed the PCR procedure, and Saiki et al. (1985) reported the first application of this technique. In a matter of few hours, the PCR procedure produces microgram (μg) quantities of DNA copies (up to billion copies) from even a single copy of the desired DNA or RNA segment (the *target sequence*). The DNA segment amplified by PCR is often referred to as *amplicon*. The PCR process has been completely automated and compact thermal cyclers are commercially available.

3.3.1 Generalized Procedure for PCR

PCR uses the following preparations/reagents: (1) a template DNA preparation containing the desired/target sequence, (2) a thermostable DNA polymerase, (3) a pair of ~20 nt long oligodeoxynucleotide primers that are complementary to the two 3' ends of the target DNA fragment, and (4) the four deoxynucleotide triphosphates, viz.,

dATP, dCTP, dGTP, and dTTP. All these reagents are present in a suitable buffer system. The above reaction mixture is subjected to the following three steps (Fig. 3.1) for, usually, 35–40 cycles. The reaction mixture is first heated most often to 94 °C to ensure denaturation of the template DNA. The duration of the denaturation step is usually 2 min in the first PCR cycle, but it is only 1 min in the subsequent cycles. The mixture is then cooled to a temperature that would allow the primers to anneal to their complementary sequences located at the 3' ends of the target DNA segment, i.e., the template DNA. Generally the annealing temperature is between 40 and 60 °C, and the duration of this step is 1 min. Since the primers are used at a much higher concentration than the template strands, they have a much higher chance to anneal with template strands than that for the two complementary strands of template DNA to pair with each other.

In the third and final step, the primers are extended due to the progressive addition of nucleotides to the free 3'-OH groups of the primers and, subsequently, the new strands being synthesized. These reactions lead to the extension of the two primers so that they grow toward each other; as a result, the DNA sequence located between the two primers is copied. The temperature during primer extension step is generally maintained at 72 °C, and the duration of this step is usually 2 min. Taq DNA polymerase is generally able to amplify DNA segments of up to 2 kb. However, it can amplify longer DNA segments provided it is used under certain special reaction conditions. Completion of the extension step completes the first cycle of amplification, and a new cycle begins with the initiation of the denaturation step. Thus, each PCR cycle takes merely 4–5 min.

The extension of primers continues till the strands are separated during the denaturation step of the next PCR cycle. Therefore, the products of primer extension based on the original DNA template, during the first and the subsequent cycles, are ordinarily longer than the target sequence since extension continues beyond the primer pairing sites; such PCR products are called *long product* (Fig. 3.2).

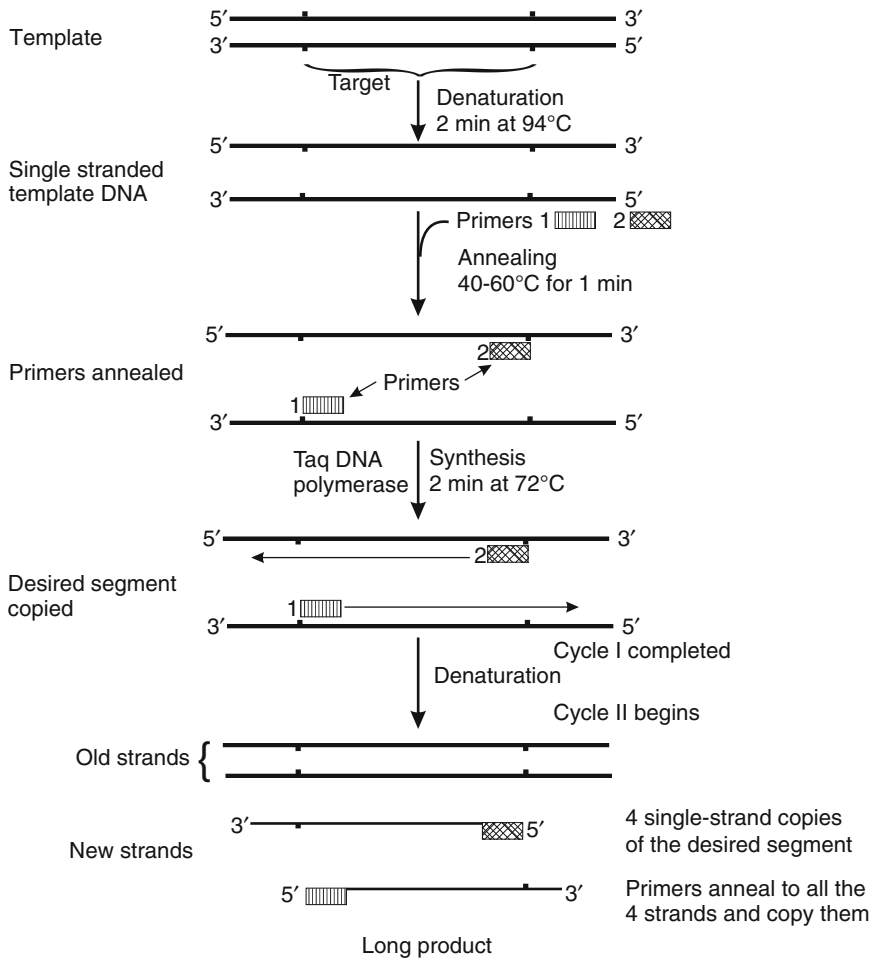


Fig. 3.1 A schematic representation of the three steps performed during the first cycle of PCR and their consequences. Note that the two primers used are complementary to the 3' end sequences of the DNA segment to be amplified. The product of the first cycle is the “long

product.” During subsequent cycles, the long product accumulates linearly, i.e., only 2×40 copies will be produced after 40 cycles of PCR from a single copy of the target segment in the original DNA sample

Fig. 3.2 The correct copy of the target sequence is produced in the second and the later cycles; its number increases exponentially. After 40 PCR cycles, $\sim 2^{39}$ copies are expected to be produced from a single copy of the target segment

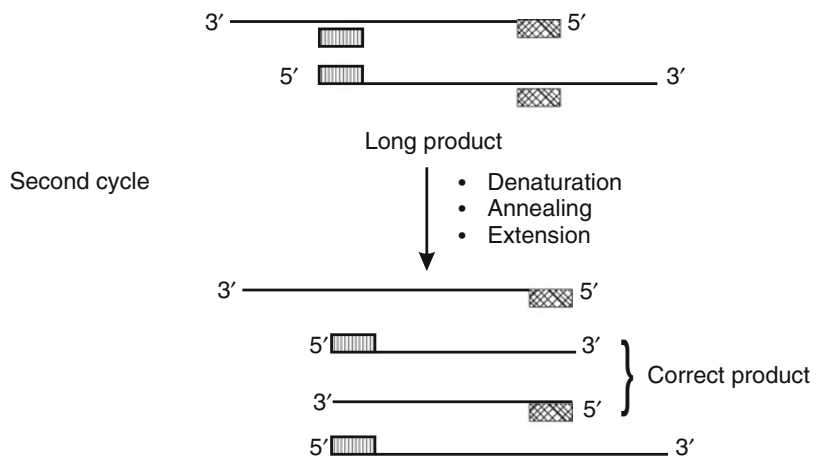


Table 3.1 Factors affecting polymerase chain reaction

Factor	Features
Template DNA	
(A) Natural features	G + C content, complexity, length of regions to be amplified, the composition (e.g., presence/absence of short repeats) of the region to be amplified
(B) Experimental features	Extraction conditions of the DNA, degree of shearing, concentration and copy number of target sequences
Characteristics of thermal block	Temperature accuracy, uniformity of temperature within the tube and between positions in the rack, ramping times
DNA polymerase	Stability, processivity, and concentration
dNTPs	Concentration
MgCl ₂ and KCl	Ionic concentration
Organic compounds	Formamide (for SSR), glycerol, and DMSO
Temperature profiles	Annealing time and temperature, extension time, and number of cycles
Primer	Size, composition, sequence, and purity

During the subsequent cycles, primers will also anneal to the “long products” at primer binding sites located before their 3'-ends. The extension of these primers will yield the correct copy of the target sequence; these copies are known as the *correct product*. Since the “long product” is produced only from the original DNA template, it continues to increase linearly. In contrast the “correct product” is generated from both the types of PCR products so that its number doubles in every cycle, i.e., it increases exponentially. Thus, one PCR cycle increases the number of copies of the target DNA segment by a factor of two in comparison to their number at the beginning of the cycle. As a result, 2^n copies of the target DNA segment are expected to be present at the end of n cycles. But the actual number of copies generated by PCR is lower than, but quite close to, this number. The investigator has only to set the temperature and duration of each step of PCR and the number of cycles to be run in the automated thermal cyclers. After this, the machine carries out all the operations exactly as specified. After the last PCR cycle, the amplification product is separated from the template DNA by gel electrophoresis, removed from the gel, and purified; it can now be used for the specific desired purpose.

PCR is a relatively robust technique when the selectivity of primers allows for stringent annealing conditions. Purity of the template is not important provided no sequence similar to the target and of foreign origin contaminates the

sample. A large number of factors can influence the success of PCR and the nature of PCR products (Table 3.1). Taq DNA polymerase at 1.25 units/25 μ l of reaction mixture would give reproducible results. Taq DNA polymerase (from *Thermus aquaticus*) is perhaps the most commonly used, but Pfu (from *Pyrococcus furiosus*) and Vent[®] (from *Thermococcus litoralis*) polymerases are more efficient. The primer length should be at least 15–17 nt for amplification of the specific desired DNA sequence, and the melting temperature of the two PCR primers should be the same. *Melting temperature* (T_m) of a primer is the temperature at which 50 % of the template-primer duplexes would dissociate into separate strands. The annealing temperature is usually 1–2 °C lower than the melting temperature of the PCR primers, while for RAPD analyses, it is kept ~5 °C lower than the T_m of the primer. In case of RAPD analyses, ~4 μ M primer should be used with ~30 ng template DNA (in 25 μ l reaction mixture) to obtain sharp and reproducible bands.

3.3.2 Separation of PCR Amplification Products

DNA fragments/amplicons generated by PCR can be separated by electrophoresis in agarose or acrylamide gels. Agarose gels are easier to make and use, and the electrophoresis system

used for these gels is simpler than that for acrylamide gels. The agarose concentration in the gels depends mainly on the size range of fragments to be separated. An agarose gel of about 1 % can separate fragment of ~300–1,500 bp, and fragments differing in length by about 50 bp can be resolved. Polyacrylamide gels contain a much more uniform pore size than agarose gels and allow separation of DNA fragments with a higher resolution. A gel containing 6 % acrylamide has a fine network formed by polyacrylamide and can separate DNA fragments differing in length by even one or two base pairs. But the maximum fragment length that can be separated using this gel is usually 500 bp. Polyacrylamide gels are suitable for detection of SSR, AFLP, DNA amplification fingerprinting (DAF), and sequence-tagged sites (STS) markers, while agarose gels are well suited for RFLP and RAPD markers. The first-generation automatic DNA sequencers used capillary gel electrophoresis because it afforded automation of filling the capillaries with the polymers as well as loading of the samples. The polymer filled in capillaries of DNA sequencers is similar to polyacrylamide (de Vienne et al. 2003).

3.3.3 Multiplex PCR

Ordinarily, a single primer/pair of primers is used in one PCR reaction set up in a PCR tube to amplify a single target sequence from the given DNA sample. Often amplification of two or more different segments from the same DNA sample may be required, e.g., for analysis of some types of molecular markers. In such cases, a separate PCR reaction will have to be set up for every primer pair because of the difficulties in correct identification of their PCR products. However, if the amplification products of two or more primer pairs can be reliably distinguished from each other, these primer pairs can be used in a single PCR reaction tube; this is known as *multiplex PCR*, and the process is called *multiplexing*. The PCR products from different primers can be reliably separated by gel electrophoresis if their lengths do not overlap. Alternatively,

different primers may be labeled with different fluorophores, and their PCR products can be distinguished on the basis of color differences in their fluorescence emissions. But this approach would require the fluorescence detection system of the first-generation automatic DNA sequencers. It is essential that all the primers used in a multiplex PCR have the same or almost the same melting temperature. This is essential for successful and specific amplification of all the concerned target sequences at the single annealing temperature used for the multiplex PCR. Multiplexing increases the throughput and reduces the cost and effort needed for scoring of markers.

3.3.4 Applications of PCR

PCR has many exciting and varied applications, some of which are as follows. It is used to study DNA polymorphism, including DNA fingerprinting, for which several PCR-based marker systems have been developed. PCR is used to detect the presence of transgenes introduced into organisms either by genetic transformation or hybridization. A variation of the PCR procedure, asymmetric PCR, generates copies of a single strand of the target sequence, which are used for first-generation automated DNA sequencing. PCR is also used for DNA sequencing reaction itself (thermal cycle sequencing PCR). The next-generation DNA sequencing procedures use PCR for in vitro cloning of the DNA fragments being sequenced. The enzyme reverse transcriptase is used along with DNA polymerase in RT-PCR (reverse transcription PCR) to generate DNA copies of RNA. Real-time reverse transcription PCR is used to estimate the initial quantity of the template RNA most specifically, sensitively and reproducibly. Several variations of PCR have been developed for specific applications, including inverse PCR for amplification of sequences flanking the target sequence, anchored PCR amplification of a target segment when the sequence of only one of its ends is known, overlap extension PCR for site-directed mutagenesis in the target segment, etc.

3.3.5 Advantages and Limitations of PCR

PCR is simple, relatively straightforward, very fast (requires only few hours), highly sensitive, and extremely versatile. It can amplify even a single copy of the target sequence present in a DNA sample and generate millions of copies of this sequence. PCR uses nanogram (ng) quantities of DNA, and purity and integrity of the DNA preparation are not critical. Further, even partially degraded DNA can be successfully used for PCR. It is easy to store and relatively cheaper DNA polymerase and does not use radioactivity. However, sequence information for the two ends of the target segment must be known for designing of the primers. In general, segments of only up to 3 kb are amplified, but this length is ideally 1 kb. Taq DNA polymerase lacks proofreading activity so that it cannot remove the errors committed during replication. Further, PCR is sensitive to several inhibitors that may be present in the DNA preparation. The expected exponential amplification continues up to about 20 cycles or so, after which it enters linear phase and soon culminates in a plateau. The PCR procedure can often generate artifacts like “hybrid amplicons” and primer dimers, and it may produce erroneous results due to contaminating DNA. Primer dimers are frequently produced when the two PCR primers have partially complementary 3' termini. They may also arise due to non-template-directed addition of some bases at the 3' ends of the two primers, which may sometimes generate complementary 3' ends in them.

3.4 PCR-Based Markers

PCR-based markers are considered as the *second-generation of molecular markers* and are based on DNA sequence polymorphisms detected by PCR amplification of the sample DNAs. The DNA polymorphisms are reflected in the amplification products from the target regions of the sample DNAs. The PCR procedure

may use a single primer or a pair of primers, and the primers may have either arbitrary or specific nucleotide sequences. The products of amplification are separated by electrophoresis using either an agarose or a polyacrylamide gel and are visualized by staining the gel with either ethidium bromide or silver, autoradiography, or fluorescence detection. The primers used for amplification differ from one marker type to the other and form the basis of the concerned marker systems. These marker systems can be grouped into the following two categories on the basis of the primers used: (1) markers based on arbitrary sequence primers and (2) those based on specific sequence primers. More recently, (3) an intermediate group of techniques has been developed that uses either a combination of specific sequence and arbitrary sequence primers or primers composed of both fixed and arbitrary sequences. In addition, (4) some techniques combine restriction digestion of DNA with PCR amplification, and they together may be regarded as a separate group (Table 3.2). These marker systems have been extensively used for gene/QTL mapping, fingerprinting of plant genetic resources, and breeding materials including commercial varieties, analysis of genetic diversity, and studies on phylogenetic relationships.

3.5 Randomly Amplified Polymorphic DNAs

Williams et al. (1990) reported the procedure for the marker *randomly amplified polymorphic DNAs* that produces fingerprints of virtually any genomic DNA sample within a matter of hours without using radioactive reagents. A single, short (usually, 10 nt long) oligonucleotide with an arbitrary base sequence is used as primer for amplification of sequences from high molecular weight genomic DNAs of the test individuals. This primer acts as both the forward and the reverse primer for the amplification reaction (Fig. 3.3). The single primer would anneal at several sites in the template genomic DNA. Theoretically, for a 10 nt long primer, the binding sites are expected to occur, on an average, every

Table 3.2 A classification of the PCR-based marker systems in common use

Category of marker system	Marker system(s) ^a	Remarks
Arbitrary sequence PCR	RAPD, DAF, AP-PCR, ISSR	Simplest to implement; poor reproducibility
Specific sequence PCR	SCAR, STS, SSR, COS, ITP, IMP	Some to considerable developmental effort; simple to use
Combination sequence PCR	SRAP, TRAP, SCoT, CDDP, S-SAP, REMAP, RBIP, CoRAP, CBDP	Simple to implement (SRAP); database search necessary (TRAP, SCoT, CDDP, CBDP)
Restriction digestion combined PCR	AFLP, CAPSs	Technically more demanding (especially AFLP)

^a*AFLP* amplified fragment length polymorphism, *AP-PCR* arbitrary-primed PCR, *CAPSs* cleaved amplified polymorphic sequences, *CBDP* CAAT box-derived polymorphism, *CDDP* conserved DNA-derived polymorphism, *CoRAP* conserved region amplification polymorphism, *COS* conserved orthologous sequence, *DAF* DNA amplification fingerprinting, *IMP* inter-MITE polymorphism, *ISSR* inter-simple sequence repeat, *ITP* intron-targeting polymorphism, *RAPD* randomly amplified polymorphic DNAs, *RBIP* retrotransposon-based insertion polymorphism, *REMAP* retrotransposon-microsatellite amplified polymorphism, *SCAR* sequence-characterized amplified regions, *SCoT* start codon-targeted marker, *SRAP* sequence-related amplification polymorphism, *S-SAP* sequence-specific amplification polymorphism, *SSR* short sequence repeat markers, *STS* sequence-tagged sites, *TRAP* target region amplification polymorphism

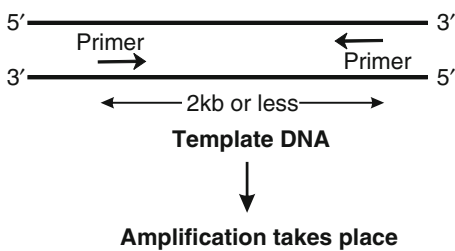


Fig. 3.3 A schematic representation of the RAPD marker system. A single arbitrary sequence primer of, generally, ten nucleotides is used for amplification. Amplification will take place if the primer binds to two sites located on the complementary strands within 2 kb of each other

4^{10} bp or 1,048,576 bp in a DNA strand, assuming a random distribution of nucleotides in the DNA strand (Appendix 3.1). However, exponential amplification can occur only when the primer anneals at two sites within ~2 kb of each other. Further, the two primer molecules should bind to the opposite strands of the template DNA so that their 3' ends face each other; this would occur only when these two binding sites are in the opposite orientation (Fig. 3.3). The reaction conditions are normally so chosen that the number of fragments amplified is less than 20 per reaction (Fig. 3.4). Thus, a very large number

of fragments can be generated by using a relatively small number of different primers. Usually, these fragments would be amplified from different regions of the genome so that several loci can be examined rapidly (Edwards 1998). Many RAPD primers may generate one to three intense bands each, which are polymorphic between the parents of a mapping population. It may be pointed out that only reproducible, intense bands should be used as markers so that the marker genotypes are scored with a degree of reliability.

RAPD method detects high level of polymorphism in plants and does not require large amounts of relatively pure DNA, and prior sequence information about the template genome is not required. It does not involve preliminary work like development of cloned DNA probes, preparation of filters for hybridization, etc., and the procedure can be automated. In addition, RAPD is safe, as it does not use radioactive components. RAPD has been used to construct high-density maps in several crop species like alfalfa, faba bean, apple, etc., in a relatively short time. This marker system has also been used to discover molecular markers linked to the desired genes in crops like tomato, lettuce, and common bean. RAPDs are dominant markers that are scored as “present” or “absent.” When it

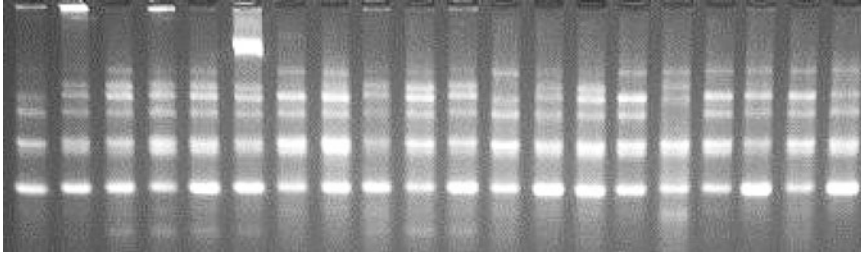


Fig. 3.4 RAPD profiles of 20 pea genotypes generated by the primer HU 12 (TGCTCAGCAG). Genotypes: 1, HUP-2; 2, Rachna; 3, DMR-42; 4, KPMR-551; 5, KPMR-615; 6, KPMR-619; 7, IPF-99-25; 8, VL-40;

9, DMR-46; 10, KPMR-660; 11, IPF-1-17; 12, IPF-1-22; 13, VL-41; 14, KPMR-662; 15, HFP-4; 16, HUDP-15; 17, KPMR-144-1; 18, DDR-49, 19, KPMR-526; 20, LFP-283 (Courtesy Kusum Yadav, Lucknow)

is important to distinguish heterozygotes from homozygotes for a locus, two RAPD markers tightly linked to this locus should be used. Further, one of the two markers should be in coupling phase, while the other marker should be associated in repulsion phase with the target locus. But this strategy will require twice the number of marker assays as that for a codominant marker. In addition, finding of two strategically located RAPD markers is not likely to be an easy task. The reproducibility of RAPD polymorphisms is low and is affected by several factors like primer to template concentration ratio, annealing temperature, and Mg^{2+} concentration (Williams et al. 1990). For example, a change of even $1^\circ C$ in annealing temperature may result in an entirely different profile of RAPD. Further, the amplification may fail due to an experimental error, but this can be scored as the “absence” allele. In many inheritance studies, RAPD markers showed significant deviation from Mendelian ratios possibly due to errors in scoring. The poor reproducibility of RAPD polymorphisms has prevented their widespread application in spite of their other highly attractive features. However, modifications of the RAPD approach have allowed the development of markers systems like SCAR, AP-PCR, RAMPO, etc., and this simple marker system still retains some relevance (Babu et al. 2014).

The information content of an individual RAPD marker is very low. RAPD markers often originate from repetitive DNA. Therefore, RAPD markers can be used as probes for locus-

specific hybridization only after considerable sequence analysis of the markers. Sometimes, heteroduplex molecules may be formed between allelic RAPD products in heterozygotes, and these may give rise to false polymorphisms (Ayliffe et al. 1994). In addition, co-migrating bands may lack homology, and a single band may contain two or more different amplicons.

3.6 DNA Amplification Fingerprinting

DNA amplification fingerprinting amplifies genomic sequences using a single short oligonucleotide, typically, of 4–6 nt as primer, but primers of up to 15 bases can be used. This produces a range of up to 100 short amplified products of different lengths. The spectrum of products changes with each primer and template combination, but is characteristic for each combination. Fragments can be adequately resolved and visualized by polyacrylamide gel electrophoresis (PAGE) combined with silver staining. DAF uses less stringent conditions for annealing and primer extension reactions than PCR. Temperature variation in the thermocycler block is not as crucial in the case of DAF as it is with conventional PCR. Short extension times are sufficient for complete extension of the short products typically obtained in DAF (Caetano-Anolles et al. 1991). DAF is suitable for DNA fingerprinting of different genotypes.

3.7 Arbitrary-Primed PCR

Welsh and McClelland (1990) reported the procedure for arbitrary-primed polymerase chain reaction (AP-PCR). In *arbitrary-primed PCR*, arbitrary sequence primers of 18–32 nt are used for amplification. It is not likely that even a very large genome will have sequences complementary for an arbitrary sequence of 20 bases or more. Therefore, amplification can occur only when the annealing conditions allow primer–template pairing with mismatches at some base pairs. The first two cycles of PCR are carried out at low stringency, and during the subsequent PCR cycles, a higher stringency (achieved by increased annealing temperature) is used. In this way, up to 100 bands may be generated for each individual, which are separated by PAGE, and scored as “present”/“absent.” The approach is suitable for DNA fingerprinting. Many workers consider AP-PCR to be essentially the same as RAPD, but the two procedures differ in terms of primer length, annealing conditions, number of amplified fragments, and the type of gel used for electrophoresis (de Vienne et al. 2003). This technique has now been refined to permit fragment separation by agarose gel electrophoresis. But AP-PCR is not a popular method as it involves autoradiography.

3.8 Sequence-Characterized Amplified Regions

In 1993, Paran and Michelmore developed the sequence-characterized amplified regions (SCAR) markers from selected desirable RAPD markers. However, this term is often applied for PCR-based markers derived from AFLP and other markers as well. The amplified fragment representing a desirable RAPD marker is eluted from the gel, cloned, and the nucleotide sequences of its two termini are determined. A pair of primers (usually, 20–24 nt long), one forward and one reverse primer, specific for the two terminal sequences is designed. This primer pair is expected to amplify a single fragment and

detect the polymorphism represented by the concerned RAPD marker in a more reliable manner. The primer pairs designed in this manner are tested for their ability to detect the concerned polymorphisms, and the successful primer pairs give rise to SCAR markers. SCAR polymorphisms are generally dominant (scored as “presence” or “absence” of a single unique band), particularly at elevated annealing temperatures (Paran and Michelmore 1993). These markers can be developed into plus/minus arrays to eliminate the need for electrophoresis. Some of the SCAR markers detect length polymorphism either directly or after digestion of the amplified fragment with a suitable restriction enzyme; the latter approach generates a marker system called cleaved amplified polymorphic sequences (CAPS; Sect. 3.14). However, sometimes the SCAR primers fail to detect any polymorphism. In such cases, it becomes necessary to sequence both the alleles of the RAPD fragment and design the two primers based on sequence differences to ensure detection of the polymorphism (Vosman 1998). Thus, SCARs are essentially similar to STS in construction and application. They can be used for physical as well as genetic mapping, comparative mapping, and phylogenetic relationship studies.

3.9 Amplified Fragment Length Polymorphisms

Amplified fragment length polymorphism technology was developed by Zabeau and Vos (1993), and it uses restriction fragments for PCR amplification. It ingeniously combines the restriction digestion of sample DNA step of RFLP system with the PCR technique to generate a robust and highly polymorphic DNA marker system (Fig. 3.5). In the AFLP procedure, 100–500 ng genomic DNA is digested with two restriction enzymes, appropriate adapters are ligated at the ends of the resulting restriction fragments, and a much smaller set of these fragments is selectively amplified by the PCR. Strictly speaking, this marker system does not detect the fragment length polymorphism

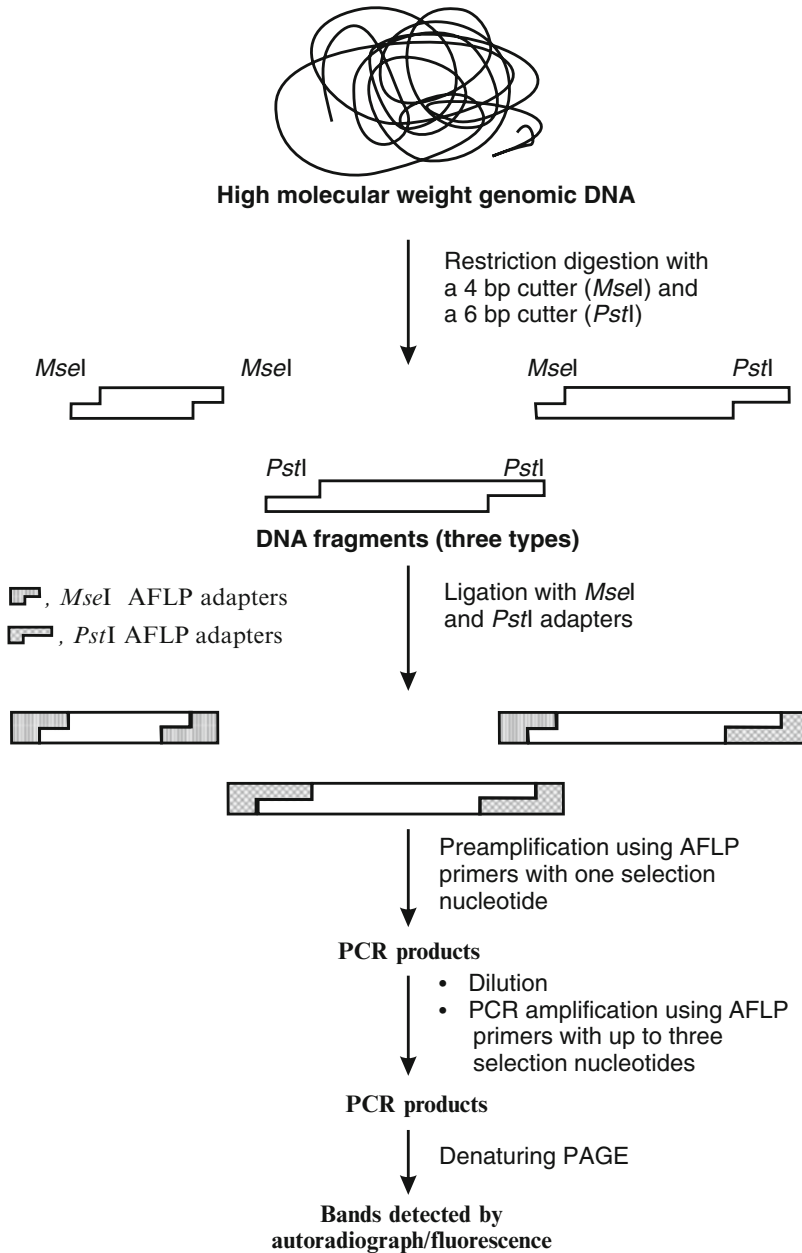


Fig. 3.5 A simplified schematic representation of the two-step AFLP method. Dilution after the preamplification step virtually removes the unamplified fragments.

In the amplification step, the AFLP primer for the 6 bp cutter is labeled with radioactivity or, preferably, fluorescence (Based on Vos et al. 1995; de Vienne et al. 2003)

generated by the restriction enzymes. The restriction enzymes, in essence, only produce the set of restriction fragments from the genomic DNAs in a highly reproducible manner and also provide a dependable strategy for fragment amplification coupled with complexity reduction. The polymorphism detected by the AFLP procedure is

actually generated by the selection nucleotides used in the AFLP primers. A restriction fragment will be amplified only when it has the complementary bases for the selection nucleotides in appropriate positions. On the other hand, a homologous fragment with mismatch at the selection nucleotide sites will not be amplified.

Thus, the polymorphism is generated primarily by differential amplification of the restriction fragments. Therefore, some authors prefer to call this marker system *selective restriction fragment amplification markers*, but *restriction fragment amplification polymorphism* seems to be a better term. A denaturing polyacrylamide gel is used to separate the PCR products, and up to 50–100 bands per sample are obtained. Of these, about 80 % of the bands may be polymorphic and can be used as markers. Therefore, AFLP is regarded as one of the most powerful high-density marker systems that produces ten times more informative markers per analysis than other marker systems and has high reproducibility. Further, prior sequence information is not required for this marker system.

3.9.1 The Procedure of AFLP

In the first step of AFLP procedure, sample genomic DNA is digested with two restriction enzymes (Fig. 3.5). One of these enzymes is a rare cutter, e.g., *PstI* (6 bp recognition sequence, 5'-CTGCA/G); this enzyme does not cut methylated DNA, as result of which it creates a bias in favor of low-copy number fragments. The second enzyme is a frequent cutter, e.g., *MseI* (4 bp recognition sequence, 5'-T/TAA); it is used to produce much smaller (256 bp = 4⁴ bp) fragments from those generated by the first enzyme. This digestion procedure produces the following three types of fragments: (1) Type I fragments have both their ends generated by the rare cutter *PstI* (*PstI-PstI*) and form a small fraction of the total fragments. (2) Type II fragments (*PstI-MseI*) have one end produced by the rare cutter (*PstI*) and the other end generated by the frequent cutter (*MseI*). (3) Type III fragments (*MseI-MseI*), on the other hand, have both their ends generated by the frequent cutter (*MseI*) and are expected to be the most frequent; they are selectively eliminated by the following PCR procedure.

After ligation of adapters to the DNA fragments, their PCR amplification is done in two steps. In the first step, called *preamplification*

step, the samples are amplified using two AFLP primers, each of which has one selection nucleotide each at its 3' end (Fig. 3.5). An *AFLP primer* has the adapter sequence plus one to three arbitrary nucleotides at its 3' end, and the arbitrary nucleotides are called *selection nucleotides*. The inclusion of selection nucleotides reduces the number of fragments that would be amplified by the AFLP primers. For each selection nucleotide added to an AFLP primer, the proportion of amplified fragments would be reduced to 1/16 (= 1/4 × 1/4) of the number of different fragments present in the mixture. In this way, 1/16th of all the three types of fragments present in the mixture will be amplified. The products of the preamplification step are suitably diluted to minimize the fragments that were not amplified in this step. The diluted mixture of the fragments is then used as template for the amplification step, in which each of the two AFLP primers has up to three selection nucleotides at its 3' end. In addition, the AFLP primer corresponding to the ends produced by the 6 bp cutter is labeled with radioactivity or a fluorophore. The AFLP primers and the amplification conditions are so designed that they favor amplification of the type II (*PstI-MseI*) fragments. Denaturing PAGE is used to separate the PCR products, and the bands are detected by either autoradiography or, preferably, fluorescence (Vos et al. 1995). The use of fluorescence-tagged primers permits the analysis of fragments by an automated DNA sequencer, which also enables automated data collection and analysis.

3.9.2 Features of AFLP

The observed AFLP polymorphisms may result from mutations either in the recognition sequences of the two restriction enzymes used for digestion of the genomic DNA or in the sequences complementary to the selection nucleotides included in the AFLP primers. In addition, insertions within or deletions from the amplified restriction fragments will also generate polymorphism. AFLP fragments/bands are of random origin, but most of them represent unique

sequences. They are dominant markers, but it is possible to differentiate heterozygous and homozygous genotypes on the basis of intensity of the bands (Staub et al. 1996). The AFLP technique is faster and less labor intensive, and detects a large number of loci that provide far greater information than RFLP procedure. Further, AFLPs are highly reproducible, which is a great advantage over RAPDs. This marker system does not require sequence information, there is no marker development step, and it can be used in any species, including nonmodel organisms. But the AFLP marker system is laborious, time-consuming, technically demanding, and expensive to set up, and it uses restriction enzymes. It requires DNA preparations of high purity (necessary for restriction digestion), the polymorphic information content of the marker system is low (the maximum being 0.5), and in some plant species like sunflower and barley, the AFLP markers tend to cluster in the centromeric regions. AFLP markers can be used for variety/line identification, characterization of germplasm, high-resolution mapping, marker-assisted selection (MAS), and gene cloning. It is still used for genetic studies in crops species, for which little or no reference genome sequence is available. In addition, it can be used for fingerprinting of DNA clones and for identification of contigs (Vos et al. 1995).

3.9.3 Modifications of the AFLP Technique

The AFLP technique has been modified in various ways to achieve specific objectives. One modification of the AFLP procedure, called *sequence-specific amplification polymorphism (S-SAP)*, generates a marker system that is similar to, but more polymorphic than, AFLPs. In S-SAP, the restriction fragments are generated and ligated to the AFLP adapters as usual. But in the amplification step, only one AFLP primer is used, and the other primer is based on a conserved sequence of a transposable element (TE). TEs occur in very high copy number in

plant genomes, and sometimes they may be more frequent in the gene-rich regions. The use of TE-based primers amplifies only those DNA fragments that have the TE sequence. The *transposon display* procedure of van den Broeck et al. (1998) is essentially the same as S-SAP, except that it deliberately uses a hexa-cutter restriction enzyme that cuts within the chosen TE. S-SAP has been used for genetic diversity studies and linkage map construction in several species, including pea, wheat, and cashew.

In another modification, called *sequence-tagged microsatellite profiling (STMP)*, one AFLP primer and one primer based on a SSR sequence (anchored at its 3' end) are used for amplification of the restriction fragments after the preamplification step. This modification takes advantage of the SSR polymorphism without prior sequence knowledge and the efforts required for SSR marker development. STMP markers can later be converted to SSR markers. Another modification of the AFLP technique is called *TE-AFLP (three-endonuclease AFLP)* since three restriction enzymes are used to digest the sample DNA. In addition, two sets of adapters are used for amplification of the fragments. The use of third endonuclease increases the discriminatory power of the technique, and a one step amplification procedure can be used for fingerprinting of even complex genomes. The *MEGA-AFLP (multiplex-endonuclease genotyping approach AFLP)* is based on four or more endonucleases used for digestion of the sample DNA. However, this modification employs only a single pair of adapters for PCR amplification.

The AFLP approach has been adapted for marker genotyping by microarray hybridization as DArT (diversity array technology; Sect. 2.6) or as CRoPS (complexity reduction of polymorphic sequences; Sect. 13.4.2) for SNP (single nucleotide polymorphism) and InDel (insertion/deletion) discovery and genotyping using a new-generation DNA sequencing technology. These modifications are amenable to high-throughput marker genotyping as well as automated data acquisition and analysis.

3.9.4 Conversion of AFLP Markers

An AFLP marker of interest can be converted into a STS marker in the same way as SCAR markers are derived from RAPD markers. DNA from the AFLP band of interest is isolated, reamplified using the same AFLP primers that were used in the amplification step, and the amplification products are sequenced either directly or after cloning. Based on this sequence information, a pair of specific PCR primers is designed for amplification of the concerned DNA fragment. This strategy can generate CAPS, dCAPS, or STS markers.

3.10 Sequence-Tagged Sites

A locus that can be unambiguously defined in terms of flanking primer sequences that are used for its amplification is called *sequence-tagged site* (STS; Olson et al. 1989). The pair of primers for an STS locus, typically, amplifies a single band. STSs can be created in the following four ways:

1. The two ends of a RAPD fragment are sequenced, and, based on this information, a pair of PCR primers is designed for reproducible-specific amplification of the intervening segment; this strategy generates SCAR markers.
2. The two ends of an RFLP or AFLP fragment are sequenced, and specific primers are designed for amplification of the RFLP/AFLP locus.
3. STSs are often created by determining the unique sequences flanking mini- and microsatellite sites. A pair of primers specific for these unique sequences is designed for PCR amplification of each of these sites.
4. Sequences of ~400 bp long fragments of genomic DNA are determined, and primers of about 20 bp may be designed for amplification of about 200–400 bp segments. These primers are tested for PCR amplification using the genomic DNA as template. If a pair of primers amplifies a single product of the correct size, a unique STS has been

identified. In human genome project, about 50 % of the primers created in this way identified unique STSs, which have been useful in creation of contigs required for physical mapping.

Thus, the creation of STS markers requires considerable amount of work, but their application requires merely the knowledge of sequences of the concerned primer pairs.

3.11 Microsatellites or Simple Sequence Repeats

Litt and Luty (1989) introduced the term *microsatellite* to describe the simple sequence fragments generated by PCR. Microsatellite sequences are also known as *short tandem repeats* (STRs), *simple sequence repeats* (SSRs), or *simple sequence length polymorphism* (SSLP). SSRs consist of tandemly repeated sequences of 1–6 bp, of which the dinucleotide repeats (CA)_n, (GA)_n, and (AT)_n are the most frequent and highly polymorphic in eukaryotic genomes. In case of plants, (AT)_n and (GA)_n repeats appear to be more numerous, while (CA)_n repeats constitute one of the most abundant microsatellites in mammals. (The value of *n* may range from 5 to 50 or even more.) Plant genomes also contain trinucleotide and tetranucleotide repeats, and the (AAG)_n and (AAT)_n sequences appear to be the most frequent. The average distance between two loci of a given dinucleotide SSR has been estimated as 30–100 kb. The trinucleotide and tetranucleotide SSR sequences are estimated to show similar distribution patterns. It appears that many microsatellites are uniformly distributed throughout the genome, but in some species like tomato, the SSRs may be clustered around centromeres (see Gupta and Varshney 2000).

Microsatellites differ from minisatellites (Sect. 2.7) in terms of the length of the repeating unit (11–60 bp for minisatellites) as well as the pattern of their distribution in the genome. Microsatellite sequences are almost evenly distributed in the plant genome, while minisatellites are generally confined to the

telomeres of eukaryotic chromosomes (Tautz 1989; Weber and May 1989). Microsatellite sequences are believed to have originated from unique sequences by random base substitutions and/or insertions that generated repeat motifs. Once produced, the repeat sequences expanded most likely due to slippage by DNA polymerase during replication and/or unequal crossing over. Consequently, microsatellite sequences are often highly polymorphic and SSR loci show multiple alleles. For example, in the elite germplasm of soybean, usually, only two alleles per RFLP locus are detected, while in a sample of about 100 elite soybean genotypes some microsatellite loci had up to 26 alleles. It may be reiterated that polymorphism at SSR loci is exclusively due to variation in the number of repeat units and base sequence variation is not involved. SSRs have been exploited to develop the following two types of markers: (1) sequence-tagged microsatellite site (STMS) or, simply, SSR markers, and (2) inter-simple sequence repeat (ISSR) markers.

3.12 Simple Sequence Repeat Markers

The *simple sequence repeat (SSR) markers* are a special version of STS markers, in which a microsatellite locus is amplified using a specific primer pair derived from the unique sequences flanking the SSR locus (Fig. 3.6). Sometimes, these markers are called STMS markers, simple sequence length polymorphisms (SSLPs), and even as microsatellite markers. Each SSR locus is amplified using a specific pair of primers, and the amplification products are analyzed by gel electrophoresis for the identification of different alleles of the locus. Ordinarily, a single SSR locus is amplified from a single DNA sample in each PCR reaction, and the PCR products from a single reaction are analyzed in one gel lane. The unique sequences flanking the SSR loci seem to be conserved within species and even across species within a given genus, but rarely across related genera. Therefore, SSR primers designed on the basis of genome sequence information from one species can be used in a related species as well.

3.12.1 Discovery of SSR Markers

Several innovative approaches have been used for the discovery of SSR loci. Initially, DNA inserts/restriction fragments containing microsatellite motifs may be identified from a genomic library/genomic DNA restriction digest. The genomic library used for this purpose may or may not be enriched for DNA inserts with microsatellites. The identified clones/restriction fragments are sequenced. But when genome sequence data are available, SSR loci can be identified more efficiently by analysis of the genome sequence and expressed sequence tag (EST) databases using data mining software like FASTA. The SSR markers derived from genome sequences are sometimes termed as *genomic SSRs (gSSRs)*, while those developed from ESTs are often referred to as *expressed SSRs (eSSRs)*. For example, one eSSR appears to be present in every 5.46 kb of wheat EST sequence. In addition, SSR markers are also derived from unigene sequences available at <http://www.ncbi.nlm.nih.gov/unigene/>; such markers are often called *unigene-derived microsatellites (UGMs)*. *Unigenes* (unique gene sequences) are a set of nonredundant EST sequences from a given species so that each unigene sequence has a unique identity and genomic location. In each of the above cases, primers specific for the unique sequences flanking the SSR sequences are designed, generally, with the help of a suitable computer program. Care should be taken with respect to the following in designing of the primers: (1) GC content of the primers should be around 50 % (T_m about 60 °C), (2) their 3'-ends should be AT-rich, and (3) the frequency of primer dimer formation should be as low as possible.

3.12.2 Increasing the Throughput of SSR Markers

The cost of SSR analysis can be reduced by the following strategies: (1) pooling the PCR products from two or more separate single primer pair-based reactions and running them in a single

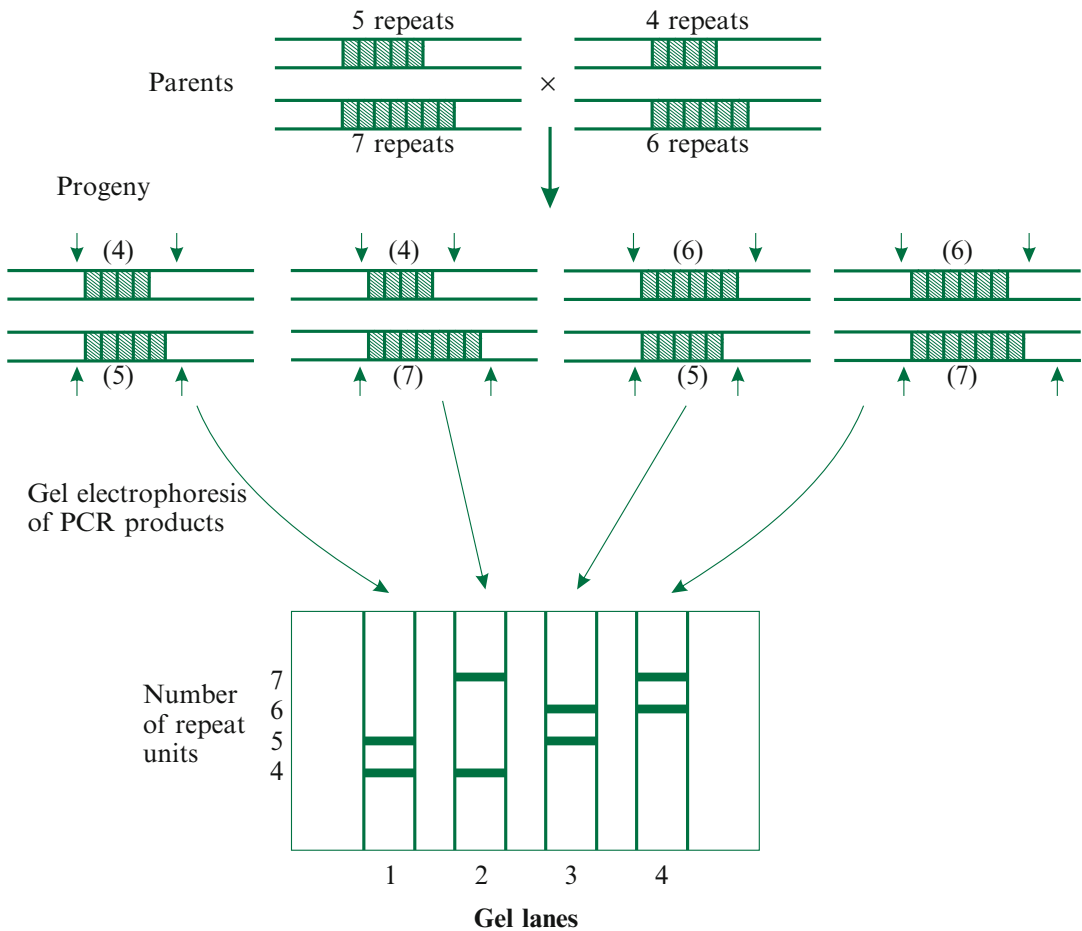


Fig. 3.6 The microsatellite (SSR) marker system: the SSR alleles result from a variation in the number of repeat units. The *arrows* indicate the sites of primer binding for

PCR amplification of the SSR locus. The primers are based on unique sequences flanking the SSR locus

gel lane, (2) using a single PCR reaction tube for simultaneous amplification of two or more SSR loci, or (3) combining the above two approaches. When primer pairs for alleles at two or more SSR loci generate amplification products of different sizes to enable their unambiguous identification, their PCR products can be pooled and used for electrophoresis. If the primers for such SSR loci could be optimized for the same PCR amplification conditions, they can be used together in a single PCR reaction tube for amplification, and the PCR products analyzed in a single gel lane. This strategy, called *multiplex PCR* (Mitchell et al. 1997), leads to a significant reduction in the costs and the time needed for assays. But

when the PCR products from different SSR loci have overlapping range of lengths, they can still be analyzed in a single gel lane by the following procedure. The PCR products from one reaction are loaded in the gel and allowed to run for a suitable period of time. The run is then interrupted, and products of the second PCR reaction are loaded in the gel and the run is resumed. The staggered loading of the PCR products from different reactions would allow the resolution of PCR products of similar lengths (Ribaut et al. 1997).

The PCR primers for different SSR loci can be labeled with different fluorescent labels. These primers can be used in a single PCR reaction

when they are optimized for the same PCR conditions. Otherwise, a different PCR reaction would be set up for each primer pair. In either case, the PCR products from three to five different SSR loci can be analyzed in a single capillary of an automated DNA sequencer even when the products from different loci are overlapping. It is possible to use a single capillary or gel lane for the analysis of up to 16 different SSR loci by taking advantage of both differential fluorescence labeling and differences in the lengths of PCR products (Gupta and Varshney 2000; de Vienne et al. 2003). Electronic data collection with automated DNA sequencers and data analysis using software like GenescanTM or GenotyperTM allows reliable fragment size determination and identification of SSR alleles; it also enables separation of native SSR alleles from the products of slippage during PCR amplification. But fluorescent labeling of primers is expensive and increases the cost of assays. The labeling cost can be reduced by using a universal fluorescence-labeled primer like M13(-21) in combination with the normal specific forward and reverse primers. However, the specific forward primer used in this reaction contains the M13(-21) sequence (without label) added to its 5' end. The use of the above set of three primers for amplification labels the PCR product because the labeled M13(-21) primer will be used as primer in the second and subsequent PCR cycles. The cost of assay is reduced because the labeled universal primer is much less expensive than the labeled specific primers. Another approach for reducing the cost is the use of an array tape, in the place of microtiter plate, to drastically reduce the amounts of reagents, consumables, etc. used (Sect. 13.2.7).

3.12.3 Merits of SSR Markers

SSR markers are codominant, highly polymorphic, distributed throughout the genome in most of the cases, and exhibit simple Mendelian inheritance. SSR assay is simple, PCR-based, locus-specific, highly reproducible, amenable to automation, and has medium throughput. The amount of DNA needed per individual is small (~100 ng),

the cost of assay system is low, and the assay can be handled manually. SSR markers are often transferable across different species of the same genus and even across closely related genera (Choumane et al. 2000). *Transferability of SSR markers* means that the primers for SSR markers developed for one plant species can be successfully used in some other, usually, related plant species. For example, the primers designed for *Oryza sativa* were successfully used in wild *Oryza* species and vice versa (Panaud et al. 1996). These markers are highly informative and can distinguish even closely related individuals.

SSR markers have been developed in several crop species. They are widely used for linkage mapping, cultivar identification, germplasm characterization (detection of accession duplications, seed mixtures, outcrossing, and genetic drift), analysis of gene pool variation, and MAS (Powell et al. 1996). SSRs became the “marker of choice” and dominated plant molecular research during the last decade of twentieth century and the first decade of the present century. But their pristine position is under challenge from the more abundant and ultrahigh-throughput SNP markers.

3.12.4 Limitations of SSR Marker System

One of the chief limitations of SSR markers is that their development is technically quite complicated, labor intensive, and costly. This involves construction of a genomic library, preferably, enriched for microsatellite sequences, screening the library with SSR-specific probes, sequencing the positive clones, designing of specific primers, evaluation of the primers for locus-specific amplification, characterization of copy number of the detected polymorphism, and determination of the chromosomal position of each SSR locus. But once the primers for the SSR loci are developed, marker analysis becomes easy and relatively inexpensive (McGregor et al. 2000). SSR markers permit only limited multiplexing and automations and are not

abundant enough to saturate the desired genomic regions. In addition, the cost of automation is relatively high, and often difficulties are encountered in sharing SSR marker data between laboratories due to differences in relative allele sizes detected across different genotyping platforms. Another problem arises due to the presence of null alleles at a proportion of SSR loci (~25 % of the loci in humans). When the specific primers for a SSR locus consistently fail to amplify a detectable product in some individuals, these individuals are said to have the *null allele* of the concerned locus. Null alleles are believed to be generated by mutation in the binding site for one or both of the primers, leading to a failure of amplification. The presence of a null allele at a locus will lead to an underestimation of heterozygosity at that locus (Gupta and Varshney 2000).

3.13 Inter-Simple Sequence Repeats

An *inter-simple sequence repeat (ISSR)* or *inter-SSR PCR* marker is based on a single primer having microsatellite sequence. The ISSR primers amplify the genomic regions flanked by the SSR sequences making up the concerned primers. The primer may consist solely of a microsatellite sequence (*non-anchored primers*) or, more often, a microsatellite sequence plus a short (usually, two nucleotides long) arbitrary sequence either at the 3' or the 5' end of the primer (*anchored primers*). In all these cases, amplification will occur only of such a genomic region that is flanked by the SSR sequence used as primer, and the SSR sequences flanking this region are in reverse orientation. These markers detect variation in the size of the genomic region between the two adjacent microsatellite sequences used as the primer binding sites.

The markers generated by non-anchored primers are called *single primer amplification reactions (SPARs)* or *microsatellite-primed PCR (MP-PCR)*. These markers are useful only when the primers consist of tri-, tetra-, and penta-nucleotide repeats because primers containing dinucleotide repeats generally yield a smear.

MP-PCR appears to offer little advantage over RAPD analysis. Further, fragments of different lengths may be obtained from the same ISSR region as a result of the primer annealing at different positions within the SSR repeats (Caldeira et al. 2002). The use of anchored primers gets around this problem, and it substantially reduces the number of ISSR fragments amplified.

The markers generated by anchored primers have been called *inter-SSR PCR*, *anchored simple sequence repeats (ASSRs)*, *anchored microsatellite-primed PCR (AMP-PCR)*, or *inter-SSR amplification (ISA)*. The region amplified by such primers depends on the anchor position in the primer. If the anchor were attached to the 5' end, the amplified fragment would include the full lengths of the two microsatellite sequences as well as the inter-SSR region. But if the anchors were linked to the 3' end of the primer, only the region between of two SSRs, including the primer, will be amplified (Fig. 3.7).

3.13.1 Modifications of ISSR

The ISSR procedure has been modified in several different ways for achieving the desired objectives. In one modification, a 5' anchored SSR primer can be used in combination with a RAPD primer to yield markers termed as *randomly amplified microsatellite polymorphisms*

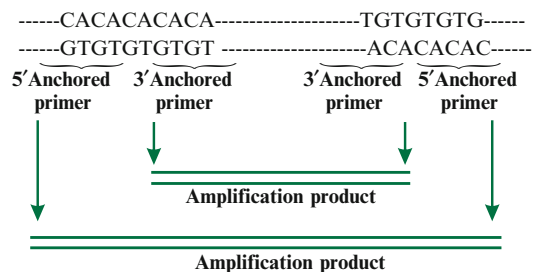


Fig. 3.7 The products generated by 3' and 5' anchored SSR primers. The anchored primers are about 17–32 nucleotides long and have usually two arbitrary bases at their 3' (3' anchored primers) or 5' ends (5' anchored primers) (Based on de Vienne et al. 2003)

(*RAMP* or *RAMPO*; Wu et al. 1994). These markers detect variation in the lengths of the target microsatellite as well as the region between the binding sites of the two primers. The RAPD primer binding site serves as an arbitrary endpoint for the anchored SSR primer-based amplification product. Therefore, the amplification products in RAMPs are greater in number than in the case of AMP-PCR. Since RAPD primers would have melting temperatures (T_m) ~10–15 °C lower than those of anchored SSR primers, the PCR program is so modified that the annealing temperature alternates between high and low (suited for the ISSR and RAPD primers, respectively) during the successive cycles. This approach has been used for genetic diversity studies in some plant species like barley. The amplification products may be digested with a restriction enzyme to yield digested RAMPs (*dRAMPs*) markers that are useful for mapping of genes/QTLs (Becker and Heun 1995).

In another modification, called *selective amplification of microsatellite polymorphic loci* (*SAMPL*), microsatellite-based primers are combined with the AFLP primers in the AFLP procedure to yield markers that are regarded as an improvement over SSRs. In case of *SAMPL*, one AFLP primer with three selective nucleotides and one *SAMPL* primer are used in combination for the amplification step (after the preamplification step) in the AFLP procedure. The best results are obtained when the *SAMPL* primer (18–20 nt long) is based on two different adjacent SSRs and the sequence lying between them; such sequences are known to occur in compound repeats. *SAMPL* primers consisting of a single SSR sequence, in contrast, generate ambiguous and less reproducible results. *SAMPL* bands generate dominant markers, but some of the markers may be codominant (Witsenboer et al. 1997).

Hybridization with a labeled SSR probe, e.g., (CA)₈, (GA)₈, (GTG)₅, (GCGA)₄, may be used to detect polymorphism in the amplification products obtained by using a regular RAPD primer or a 10/15 nt long non-anchored SSR primer. This method has high sensitivity at the intraspecific level, but it uses radioactivity. This

marker is known as *RAMP*, *RAMPO*, *randomly amplified hybridization microsatellites* (*RAHM*), or *randomly amplified microsatellites* (*RAM*). Another marker, termed as *retroposon-microsatellite amplified polymorphism* (*REMAP*), uses a 3' anchored microsatellite primer along with a primer based on the LTR (long terminal repeat) of a retrotransposon for PCR amplification. Consequently, *REMAP* can amplify three different types of DNA fragments: (1) the segments flanked by an LTR at one end and a microsatellite locus on the other, (2) sequences having a microsatellite locus at both their ends, and (3) fragments present between two neighboring insertion sites of the concerned retrotransposon.

3.13.2 Merits and Limitations of ISSR Markers

ISSR markers are more reproducible than RAPD, easy to use, cheap, have high throughput, and yield multiple polymorphic loci. Further, a prior knowledge of the template DNA sequence is not required. Generally, ISSR markers are dominant, but the use of a larger 5' anchor can yield codominant ISSR markers. A major disadvantage of ISSR markers is that they are not highly reproducible, and some primers generate poorly reproducible band patterns.

3.14 Cleaved Amplified Polymorphic Sequences

The *cleaved amplified polymorphic sequences* (*CAPSs*) detect length polymorphism generated by restriction digestion of specifically amplified PCR products from different genotypes. They are often called *PCR-RFLP* since they were developed for easy genotyping of RFLP markers using gel electrophoresis, following PCR of the target regions (Williams et al. 1991; Fig. 3.8). Therefore, *CAPSs* are codominant markers. They result from alterations in the recognition sites, located within the amplification products, for the respective restriction enzymes. The restriction enzymes used for *CAPS* analyses should

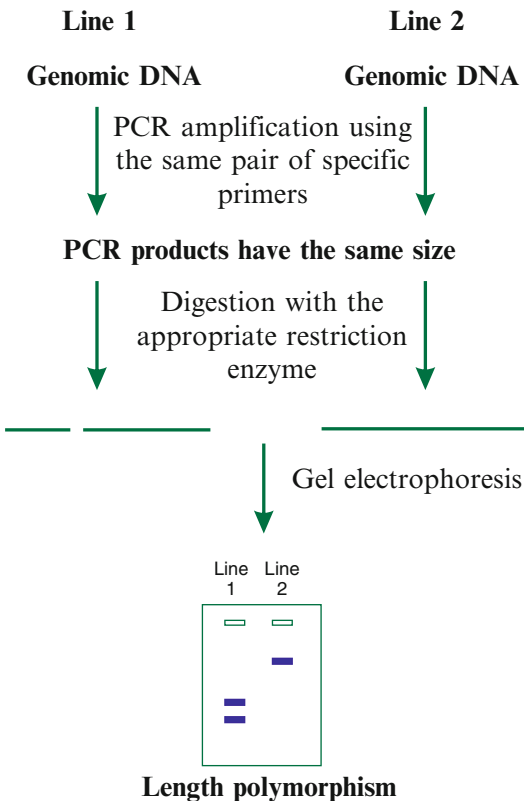


Fig. 3.8 A schematic representation of cleaved amplified polymorphic sequence (CAPS) marker system

have 4 bp recognition sequences since they are much more likely to have recognition sites within the amplification products of ~0.5–2 kb. This technique is useful when the amplified DNA fragments are large, fail to reveal polymorphism among genotypes, and contain a SNP within the recognition site for a restriction enzyme. The CAPS approach is preferable to the standard RFLP analyses. But the use of restriction enzymes for CAPS analysis adds to the assay cost and makes this marker system unsuitable for high-throughput analysis and automation.

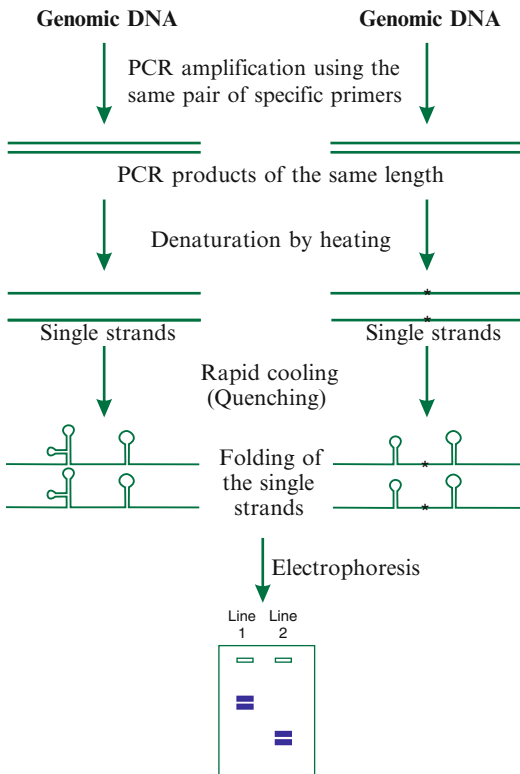
In a variation of the CAPS method, called *derived cleaved amplified polymorphic sequence (dCAPS)* or *mismatch PCR-RFLP*, one of the PCR primers generates in the PCR product a recognition site for a restriction enzyme. This primer is so designed that it contains one of the SNP alleles and one or more mismatches with the target template DNA sequence. These

mismatches together with the SNP allele generate a restriction site in the PCR product of this allele, but not in that of the other allele. The concerned restriction enzyme is used for digestion of the PCR products, and the SNP alleles are deduced from the restriction fragments generated from them (Michaels and Amasino 1998). The dCAPS method is simple, relatively inexpensive, and would be useful for scoring known SNP alleles and for positional cloning of new plant genes.

3.15 Single-Strand Conformation Profile/Polymorphism

Single-strand conformation profile/polymorphism (SSCP) is detected as differential movement of single-stranded DNA molecules, representing identical genomic regions from different individuals of a species (Orita et al. 1989). The DNA fragments used for SSCP analyses are generally obtained by PCR amplification. The differential migration of the single-stranded fragments results from differences in their secondary structures. The secondary structures of the single strands result from folding and internal complementary base pairing in short regions. The base pairing produces short double-stranded regions that stabilize the folding pattern as well as contribute to the secondary structures. The internal base pairing would depend on the base sequence. Therefore, the differences in conformations of the single-stranded molecules would reflect the differences in their base sequences.

The detection of SSCP involves heating the solution of a double-stranded DNA molecule to 95 °C so that the two strands of the DNA molecules become separated. This denatured DNA solution is now quenched, i.e., cooled very rapidly. As a result, the complementary strands do not get sufficient time to pair with each other. Instead, the single strands fold onto themselves, and internal base pairing in short regions leads to the formation of characteristic secondary structures (Fig. 3.9). The differences in secondary structures of the single strands are



The complementary strands form two different bands due to slightly different mobility in the gel

Fig. 3.9 A schematic representation of single-strand conformation polymorphism (SSCP) for discrimination between PCR products of identical lengths from the same genomic region of two lines differing for a mutation in this region (Based on de Vienne et al. 2003). * Mutation

detected by acrylamide gel electrophoresis under non-denaturing conditions. Fluorescence-labeled primers may be used for amplification of the target sequence to facilitate detection of the bands after electrophoresis. It has been estimated that in DNA molecules of up to 200 bp, 100 % of the differences in base sequence are revealed by SSCP. However, as the length of DNA duplex increases, the percentage of sequence differences detected by SSCP decreases.

The two strands of a DNA duplex usually generate slightly different secondary structures. Therefore, two bands will be observed in homozygotes (Fig. 3.8), and the heterozygotes would exhibit four bands. But each single strand of some DNA molecules can form more than one

slightly different semi-stable conformation leading to the formation of multiple bands in homozygotes. SSCP procedure is useful for rapid screening of sequence differences among amplification products, when precise information about the sequence differences is not needed. This procedure is simpler and more convenient than CAPS, which requires restriction digestion of the PCR product, and D/TGGE (denaturing/temperature gradient gel electrophoresis), where a precise control of the electrophoresis conditions is necessary. SSCP has been used for mapping and genetic studies in plants only to a limited extent (de Vienne et al. 2003). The major disadvantages of SSCP are labor-intensive and costly marker development and the lack of automation.

3.16 Denaturing/Temperature Gradient Gel Electrophoresis

Denaturing/temperature gradient gel electrophoresis (D/TGGE) reveals differences in the movement of double-stranded DNA molecules from the same genomic regions of different individuals of a species. These DNA molecules are obtained by PCR amplification. Short stretches within a DNA duplex would differ from each other in terms of melting temperature, which depends on their base composition. For example, AT-rich stretches would have lower melting temperatures than GC-rich regions. As a result, the two strands of a DNA duplex will begin to separate earlier in AT-rich stretches than in GC-rich stretches, when the DNA molecules are subjected to increasingly denaturing conditions, e.g., during denaturing/temperature gradient gel electrophoresis. This property is exploited for the detection of sequence differences among PCR products from different individuals of a species.

The PCR products from different individuals are loaded in separate wells of an acrylamide gel. Preparing the gel with a denaturing agent, e.g., urea and formamide, can create the denaturing conditions during electrophoresis; this agent is added in a gradient of increasing concentration

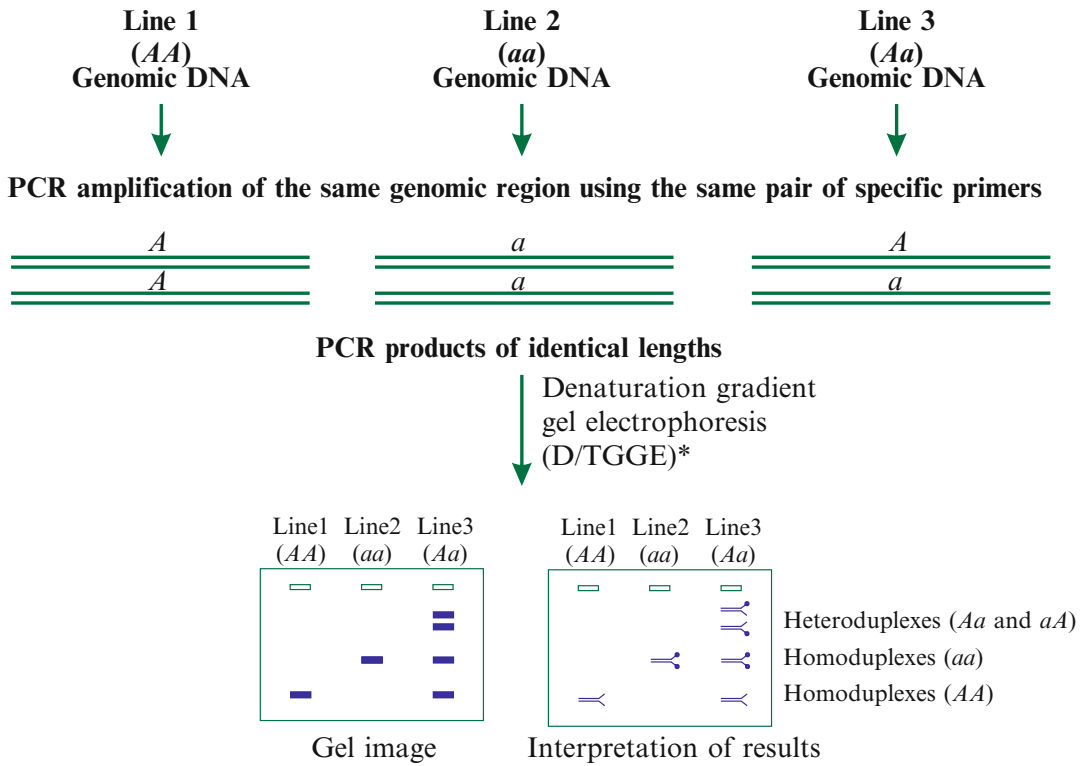


Fig. 3.10 A schematic representation of denaturing/temperature gradient gel electrophoresis (D/TGGE) to distinguish between PCR products of the same length but differing for a mutation. The *dots* at the ends of DNA strands from a line signify the presence of mutation. * the PCR products are denatured, followed by renaturation prior

to D/TGGE in order to definitely heterozygotes due to the formation of heteroduplexes. PCR products migrate as duplexes in the gel till one of their ends melts to produce a branched structure and prevents further migration. Heteroduplexes form the branched structure earlier than the homoduplexes (Based on de Vienne et al. 2003)

starting from the loading wells. Alternatively, a normal acrylamide gel may be used and an increasing temperature along the gel during electrophoresis can create the denaturation gradient. The PCR products initially migrate in the gel as double-stranded molecules. As they migrate farther in the gel, they meet stronger denaturing conditions, and soon their least stable regions begin to melt. At some point in the gel, one end of the molecule would become single-stranded; this would produce a branched structure that does not migrate any further in the gel (Fig. 3.10). In most cases, a difference of even a single base pair in the least stable region of a DNA molecule of <300 bp would lead to a difference in the mobility of the molecule, and the variant molecule would form a different band in the gel (de Vienne et al. 2003).

D/TGGE permits identification of all heterozygous individuals by a simple step at the end of PCR. After the last PCR cycle, the denaturation step is implemented and is followed by renaturation of the PCR products; this would lead to the formation of two heteroduplexes in addition to the two homoduplexes in all the heterozygotes. The two heteroduplexes will be produced by association of each of the two strands of one allele with its complementary strand from the other allele. Heteroduplexes have considerably lower melting temperatures than the homoduplexes so that they do not migrate very far in the gel and form distinct slow moving bands. The heteroduplex bands are easily detectable even when the bands in the two homozygotes are not distinguishable (Fig. 3.9; de Vienne et al. 2003).

D/TGGE is a delicate technique and requires considerable care. The gradient must be chosen on the basis of stability features of the fragment to be analyzed, and the slope and the limits of the gradient must be carefully determined, preferably, by using suitable software. In any case, the preparation of gradient gels is time-consuming as well as prone to technical errors. In addition, the size of individual DNA fragments will determine the amount of denaturant to which they will be exposed. As a result, small DNA fragments would migrate to the bottom of the gel and might even be eluted from the gel, before they encounter sufficient amount of the denaturant for causing differences in mobility. These difficulties are overcome by temperature gradient gel electrophoresis. In case a PCR product does not have two distinct regions differing in stability, a GC clamp may be attached to the molecule. A *GC clamp* is a stretch of about 30 bp containing only GC bases. The GC clamp can be appended to the 5' end of the PCR primer used for amplification of the fragments to be analyzed. If the base sequences of the variants of the concerned fragments were precisely known, their migration can be modeled to facilitate quick screening of the variants (de Vienne et al. 2003).

3.17 Sequence-Related Amplification Polymorphism

Sequence-related amplified polymorphism (SRAP) is one of several gene-targeted markers based on PCR amplification (Poczai et al. 2013); many of these markers are described in the following sections. *SRAP* is a simple marker based on open reading frame (ORF) amplification. *SRAP* uses two primers of 17 or 18 nt each, which have, beginning from their 3' ends, three selective nucleotides, followed by a *core sequence* of 4 nucleotides (5' CCGG 3' in the forward primer and 5' AATT 3' in the reverse primer) and a 10 or 11 nt long arbitrary sequence (*filler sequence*) at the 5' end (Fig. 3.11). It is important that different filler sequences are used for the forward and reverse primers. The CCGG core sequence is targeted at exons since exons

5' NNNNNNNNNNCCGGXXX 3'
 Filler sequence Core Selective
 10-11 nucleotides sequence nucleotides

Forward primer

5' NNNNNNNNNNAATTXXX 3'
 Filler sequence Core Selective
 10-11 nucleotides sequence nucleotides

Reverse primer

Fig. 3.11 The forward and reverse primers used for the detection of sequence-related amplification polymorphism (SRAP). Filler sequences of the two primers are arbitrary sequences, but different from each other. The sequence 5' CCGG3' targets exons, while the sequence 5' AATT3' targets introns and promoter regions (Based on Li and Quiros 2001)

are more frequent in GC-rich regions. The AATT core, on the other hand, targets promoters and introns, which are normally AT-rich. The annealing temperature during the initial five PCR cycles is kept at 35 °C; it is set at 50 °C during the next 35 PCR cycles. Denaturing acrylamide gel electrophoresis is used to separate the PCR, and the bands are detected by autoradiography (Li and Quiros 2001).

In recombinant inbred line (RIL) and doubled-haploid (DH) populations of *Brassica oleracea*, SRAP markers were almost evenly distributed over the whole genome. Each primer combination generated many bands of which >10 were polymorphic. About 45 % of the bands represented already known genes that are listed in the GenBank, and 20 % of the bands showed codominance. SRAP method is simple and reliable, has moderate throughput, targets coding sequences, and generates a fair proportion of these markers behave as codominant (Li and Quiros 2001). The codominant markers will be generated by insertions and deletions since they would lead to polymorphism in the amplified fragment size. In contrast, SNPs affecting primer binding would generate dominant markers since they would either allow or prevent fragment amplification. This marker system has been used in several crops including potato, rice, lettuce, and garlic to achieve a variety of objectives, including linkage mapping, identification of

markers linked to useful genes, and genetic diversity analyses.

3.18 Target Region Amplification Polymorphism

The *target region amplification polymorphism (TRAP)* is a PCR-based marker system that involves in silico analysis of the EST database for designing of such primers that detect polymorphism around the desired candidate genes. TRAP uses two primers of 18 nt each; one of these primers is complementary to a sequence of the targeted EST (the *fixed primer*), while the other is an arbitrary primer (Fig. 3.12). The *arbitrary primer* has the same design as that of a SRAP primer (Sect. 3.17): the arbitrary primer may have an AT-rich core (5' AATT 3') and would anneal to an intron or a GC-rich core (5' CCGG 3') and would anneal to an exon. The fixed primer is designed as follows: EST database of the concerned species is searched, the desired EST is identified, and its sequence is used to design an 18 nt long primer with T_m of 50, 53, or 55 °C. The annealing temperature during the first five cycles of PCR is kept at 35 °C, but during the next 35 cycles, it is kept at 50 °C (Fig. 3.12).

Fig. 3.12 A schematic representation of the (a) arbitrary and (b) fixed primers used for detection of the target region amplification polymorphism (TRAP) and (c) the significant features of the PCR amplification (Based on Hu and Vick 2003)

- a**
- 5' NNNNNNNNNNCCGGXXX 3' (Targets exons)
- or
- 5' NNNNNNNNNNAATTXXX 3' (Targets introns and promoters)
- The arbitrary primer (18 nucleotides)**
- b**
- EST database of the species is searched
 - The desired EST sequence is retrieved
 - EST sequence information is used to design a 18-nucleotide long primer with T_m of 50, 53 or 55°C
- The fixed primer**
- c**
- Annealing temperature:**
 During first five cycles, 35°C
 During the next 35 cycles, 50°C
- PCR amplification**

In different plant species, the TRAP technique can generate up to 50 scorable markers of 50–900 bp from a single PCR reaction. The PCR products are resolved by electrophoresis using a 6.5 % polyacrylamide sequencing gel. These markers seem to be reproducible, and an automatic DNA sequencer in conjunction with fluorescent labels can be used for their detection (Hu and Vick 2003). TRAP system is better than SRAP as it yields markers around the target candidate genes, while the latter amplifies from all over the genome. The TRAP method has been used for germplasm characterization, fingerprinting of genotypes, and mapping of genes/QTLs (quantitative trait loci).

3.19 Transposable Element-Based Markers

Transposable elements (TEs) are DNA sequences that move around in the genome. They constitute >50 % of nuclear DNA and generate genetic diversity through insertion into functional genes, excision from various genomic sites, and generation of small structural rearrangements. TEs are classified into *Group I transposons (retrotransposons)* that transpose via RNA intermediates and *Group II transposons* that move as DNA molecules. Some

retrotransposons have long terminal repeats (LTRs), while others lack LTRs. Both these types of retrotransposons are present in plants usually in high copy numbers and are dispersed throughout the genome. There is a great variation in the number of copies and the sites of insertion in the genomes of even closely related species. Several marker systems are based on retrotransposons. Of these, the *sequence-specific amplification polymorphism (S-SAP)* seems to generate the largest number of highly polymorphic markers. S-SAP is an AFLP-like approach that displays as bands the regions between concerned retrotransposon insertion sites and the selected restriction sites (Sect. 3.9.3). In self-pollinated species like pea, S-SAP markers appear to be more informative than AFLP and RFLP markers (Ellis et al. 1998), and they have been used for phylogenetic analyses in pea.

Another approach uses primers based on LTRs of retrotransposons to amplify the region between two neighboring insertions of the element; this is called *inter-retrotransposon amplified polymorphism (IRAP)*. The approach called *retrotransposon-microsatellite amplified polymorphism (REMAP)*, on the other hand, uses one primer based on LTR of a retrotransposon and a second primer representing a microsatellite sequence that may be anchored. REMAP markers detect polymorphism in the genomic fragment flanked by the insertion site of a retrotransposon on one side and a SSR site on the other side. IRAP and REMAP markers are highly polymorphic, and up to 30 bands per individual may be obtained. These marker systems have been used for analysis of genetic relationships within species (Agarwal et al. 2008).

Some other transposable element-based markers are retrotransposon-based insertion polymorphism (RBIP), transposon display (TD), and inter-MITE polymorphism (IMP). The *RBIP* approach is designed to detect retrotransposon insertions at specific sites using PCR amplification (Agarwal et al. 2008). RBIP uses one primer derived from the concerned retrotransposon and a pair of primers derived from the sequences flanking this retrotransposon at the given insertion site. When the primer pair derived from the flanking sequences is used for amplification, a

product would be obtained whenever there is no retrotransposon insertion in the region flanked by the primers. But when the primer based on the retrotransposon is used with a primer specific to one of the flanking regions, a PCR product would be generated only when the concerned region contains the retrotransposon. Polymorphisms can be readily detected by electrophoresis using an agarose gel. Alternatively, a simple dot blot assay using a reference PCR fragment for hybridization may be employed for analysis of the polymorphism. The dot blot assay is amenable to high-throughput automation. This method requires sequence information about the transposable element as well as the regions flanking the concerned insertion site, which involves considerable amount of work. It is perhaps the costliest and the most complicated method for detection of transposon insertions.

IMP markers are an example of markers derived from Group II transposons. The IMP technique is identical to IRAP, except for the use of primers based on MITE-like transposable elements in the place of those derived from retrotransposons. *MITEs (miniature inverted-repeat transposable elements)* are a family of small transposons, which are distributed widely and are plentiful in a number of plant genomes. They are often located in the terminal regions of genes and show considerable polymorphism among inbred lines. The *MITE-AFLP* method is similar to S-SAP as it uses one AFLP primer and one primer based on a MITE element for the amplification step of the AFLP procedure. The MITE-AFLP procedure has been used for studying genetic diversity and analyzing phylogenetic relationships in rice, wheat, and maize.

3.20 Conserved Orthologous Set of Markers

Conserved orthologous set (COS) of genes may be defined as a group of genes that show conservation of sequence as well copy number during the evolution of plant species. The *conserved orthologous set of markers* consists of gene-based markers derived from the conserved orthologous set of genes (Fulton et al. 2002).

The conserved set of genes is identified by computational analysis of genomic and EST sequences from a group of related species along with a well-characterized reference species like *Arabidopsis thaliana* (usually, for dicots) or rice (usually, for monocots). Each gene of the orthologous set has an orthologue in all the species of the group and often even in other distantly related species. Ordinarily, the genes included in the orthologous set are single-copy genes, but low-copy number genes may also be included. The COS gene-based markers are developed by designing a pair of specific primers for each gene set using the highly conserved sequences of exons. These primers may amplify an exonic region of the gene, but the amplified region may include at least one intron. A vast majority of the primer pairs successfully amplify genic regions, and ~90 % of the products show polymorphism. Usually, polymorphism is detected by SSCP, and the bulk (over 60 %) of polymorphisms are due to SNPs, while the rest are due to InDels.

Fulton et al. (2002) analyzed a large database of ESTs from tomato against the *A. thaliana* genome sequence. They identified 1,025 genes that are present in single- or low-copy number in the genomes of both tomato and *A. thaliana* and show high sequence conservation during evolution. They referred to this group of genes as *conserved orthologous set* or *COS markers*. In silico computational analyses and DNA gel blot hybridization were used for identification and evaluation of COS markers. A large fraction of the identified COS markers was concerned with basic metabolic processes like energy generation, biosynthesis, and degradation of cellular components. The COS markers are used for genome evolution studies, comparative mapping among even highly divergent species as well as for physical and linkage mapping of the concerned genes. COS markers have been extensively used to connect the genomes of related species belonging to the same family. Those COS markers that are conserved can be used as hybridization probes for RFLP analyses; this should allow mapping even in such species that do not have either genomic or EST sequence databases. Other COS genes can be used for the development of gene-based markers for

detecting polymorphism among the PCR products using SSCP. The consensus sequences of COS markers can be used as query for homology search of genome sequence databases of other plant species to identify putative orthologous genes in these related species.

The genome sequences of three model species, viz., *A. thaliana*, *Oryza sativa*, and *Populus trichocarpa*, were subjected to comparative analysis; this resulted in the identification of 753 candidates for COS markers. Out of these, up to 359 genes were present in the EST databases of four gymnosperm species. Similarly, the Rosaceae EST databases were compared with single-copy genes of *Arabidopsis* to identify 1,039 RosCOS (COS set for Rosaceae) markers. Out of these, 857 genes were chosen for designing of primers flanking introns so that the PCR product included at least one putative intron. About 91 % of these primers were able to amplify *Prunus* DNA, and 90 % of the PCR products exhibited polymorphism.

3.21 Start Codon-Targeted Polymorphism

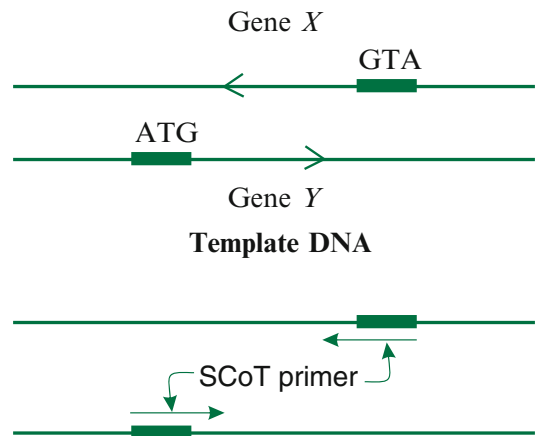
Start codon-targeted (SCoT) polymorphism markers are based on the short conserved sequence surrounding the translation initiation codon or start codon, ATG, of plant genes as reported in various studies (Collard and Mackill 2009a). The SCoT marker system uses a single 18 nt long primer to amplify the sample genomic DNAs, and the amplification products are resolved by agarose gel electrophoresis. The SCoT primer has the following invariant nucleotides: the A, T, G of the start codon (positions +1, +2, +3), G at +4, A at +7, C at +8, and C at +9. The primers also have a variable number of arbitrary nucleotides on the 5' side of the ATG nucleotides. The GC content of the primers may range between 50 and 70 %, and they should differ from each other for at least one nucleotide at their 3' ends. The annealing temperature during PCR is kept at 50 °C, and the primer extension time of at least 2 min is recommended. The SCoT markers are generally highly reproducible, but some primers show poor reproducibility. The amplification products are between two and

six in number, and their lengths range from 200 to 1,500 bp. This marker system is similar to RAPD and ISSR marker systems in respect of the use of a single primer, lack of sequence information requirement, and two to six amplification products in each PCR. But SCoT markers would be based on genic regions as compared to the random genomic regions in the cases of RAPD and ISSR markers. However, some of the SCoT markers may be generated by pseudogenes and even such genes that are situated within transposable elements.

Amplification of a fragment would occur when start codons of two genes are located within a reasonable distance on the complementary strands of the DNA duplex (Fig. 3.13). The SCoT markers are dominant, but few of them may be codominant due to relatively large InDels in the amplified regions; this situation is similar to that for the RAPD markers. These markers can be used for mapping of genes/QTLs, and genetic diversity analyses. A SCoT marker of interest can be converted into a STS marker to make it single band robust marker. The SCoT marker system has the potential to be used for a simplified gene expression analysis with limited resources. The cDNA-SCoT technique was developed for this purpose (Wu et al. 2013).

3.22 CAAT Box-Derived Polymorphism

The *CAAT box-derived polymorphism (CBDP)* marker is a PCR-based marker similar to the SCoT marker as it uses a single primer of 18 nt that targets the CAAT box of the promoter regions of plant genes. The primer has the five-nucleotide CCAAT core flanked by 10–11 filler nucleotides on the 5' side and 2–3 arbitrary nucleotides on the 3' side. Singh et al. (2014a) designed a set of 25 CBDP primers and evaluated them with eight varieties of *Corchorus capsularis* and *C. olitorius*. Most of these primers generated few to several polymorphic bands in the jute varieties and in cotton and linseed as well. The CBDP marker system is similar to the SCoT marker system in many features, including the following. A band



Amplification of SCoT marker amplicon

Fig. 3.13 The principle of SCoT marker system. The template DNA has different genes on the complementary strands of the DNA duplex. The start codons (ATG) of the two genes are within a distance appropriate for amplification (say, up to 1,500 bp). The ATG codons of these two genes should be located as shown in the figure for the region between them to be amplified (Based on Collard and Mackill 2009a)

will be generated when two genes are located on the opposite strands within a distance suitable for PCR amplification. The phrase “two genes” means the CAAT boxes of the promoters of the two genes, in the case on CBDP markers, and start codons of the two genes in the case of SCoT markers. The CBDP markers would be useful for analyses of genetic diversity, DNA fingerprinting for reliable cultivar/germplasm identification, and linkage mapping of genes/QTLs and MAS.

3.23 Conserved DNA-Derived Polymorphism

The *conserved DNA-derived polymorphism (CDDP) markers* are based on conserved DNA regions of a selected set of well-characterized plant genes. For example, Collard and Mackill (2009b) analyzed the sequences of *WRKY*, *MYB*, *ERF*, *KNOX*, *MADS*, and *ABPI* genes to produce several CDDP markers. The above genes are known to participate in abiotic/biotic stress responses or developmental processes. Sequences

of the selected genes present in diverse plant species were obtained from the database and used for multiple sequence alignment analysis by ClustalW program (Sect. 14.3.9) to identify their conserved regions. These conserved sequences were used for designing primers in such a way that their GC contents were over 60 % and a single primer could have up to three degenerate nucleotides. The primers designed in either 5'–3' or 3'–5' direction with respect to the conserved domain sequence generated such markers that were reproducibly polymorphic. The short sequences conserved in the selected genes may be expected to be present at several locations in the plant genome and would serve as binding sites for the CDDP primers. The principle of CDDP markers is similar to that of the SCoT markers and the resistance gene analog markers. The *resistance gene analog markers* are based on primers derived from the conserved regions of genes for disease resistance of plants (Chen et al. 1998). Denaturing polyacrylamide gel electrophoresis is used to separate the PCR products. Each PCR reaction generated from 30 to 130 products, of which 27–47 % showed polymorphism in rice, barley, and wheat.

The CDDP markers are dominant and are scored as “present” or “absent.” CDDP primers generate two to six fragments of 200–1,500 bp in size. This marker system is similar to the SCoT markers (Fig. 3.13) in the use of a single primer for PCR, amplification of genic regions, and the need for genes to be present at proper distance in the complementary strands. In contrast to RAPD, it uses longer primers, much higher annealing temperature (50 °C), and has high reproducibility, except in the case of some primers. CDDP differs from the conserved region amplification polymorphism (CoRAP) in the following ways. The CoRAP procedure uses two primers derived from ESTs for a specific species and requires polyacrylamide gel electrophoresis. In contrast, the CDDP markers use a single primer derived from the sequence of the selected gene present in several plant species, and they are scored by agarose gel electrophoresis. Thus, CDDP markers would target the selected plant genes, including candidate genes where known. These markers can be used for gene/QTL mapping as well as genetic diversity studies.

3.24 Conserved Region Amplification Polymorphism

The *conserved region amplification polymorphism* markers are based on pairs of primers (one fixed and one arbitrary primer) for PCR amplification (Wang et al. 2009b). The fixed primer is derived from the sequence of an EST of a given species extracted from a database like GenBank and targets the coding sequence of the gene. The arbitrary primer contains the core sequence CACGC at the 5' end, followed by 11 arbitrary nucleotides that serve as fillers, and three bases at the 3' end, which serve as selection nucleotides; this scheme is the same as that for the SRAP markers (Sect. 3.17). Since their core sequence is normally found in the introns of plant genes, the arbitrary primers would anneal to the majority of introns. The CoRAP primer pairs are designed for an annealing temperature of 52 °C. These markers are similar to TRAP markers (Sect. 3.18), except for the core sequences of their primers. Thus, the design of fixed primers requires sequence information of the concerned plant species. PCR amplification will occur if the two primers bind within a suitable distance from each other. The amplification products will be polymorphic if the intervening sequences had InDels, as a result of which the PCR products from different individuals/strains would differ in size. The CoRAP markers are codominant and highly reproducible. Each PCR reaction may generate 30–50 fragments of 50–1,000 bp.

3.25 Intron-Targeting Polymorphism

In case of *intron-targeting polymorphism (ITP) markers*, the primers are designed on the basis of the sequences of the conserved regions of exons flanking an intron so that the PCR product includes the intervening intron (Choi et al. 2004). Since the introns are much less conserved than exons, a high proportion of the amplified fragments may be expected to show length polymorphism due to InDels. The ITP primers are derived from the sequences of

known single-copy and low-copy number genes or from those of the ESTs available in the database. The primer pairs are designed to amplify fragments of 200–1,200 bp, which are resolved by subjecting them to agarose gel electrophoresis. The ITP markers are codominant, and the primers are transferable across the species of the same genus and, sometimes, even across genera. ITP markers are generated from genic regions, and some of them might give rise to functional markers. However, the development of ITP markers depends on prior sequence information about several target genes. The ITP markers can be used for genetic diversity analyses.

3.26 RNA-Based Molecular Markers

Several useful markers are derived by analysis of RNA (Poczai et al. 2013). For example, *SSCP analysis of cDNA (cDNA-SSCP)* allows estimation of relative abundance of mRNAs encoded by highly similar homologous genes of polyploid species. *RNA fingerprinting by arbitrarily primed PCR (RAP-PCR)* uses arbitrary sequence primers for fingerprinting of RNAs isolated from a given tissue of different individuals or RNAs obtained from different tissues of a single individual. The sequence polymorphisms detected by RAP-PCR can be used for mapping of genes. The *cDNA-AFLP* technique, as its name suggests, is an AFLP procedure that uses cDNA in the place of genomic DNA as substrate. It can discriminate between such genes that belong to the same gene family and are highly homologous and allow identification of genes related to novel processes, including stress regulation.

Questions

1. Briefly describe the procedure of PCR, and discuss its usefulness in marker development and genotyping.
2. Compare the RAPD and AP-PCR markers. Why are SCAR markers more reliable than RAPD?
3. How is complexity reduction achieved in the case of AFLP procedure? Briefly describe

some of the various modifications of the AFLP procedure.

4. How are SSR markers developed? Why did they become the most widely used marker system before the SNPs became the markers of choice?
5. What are various approaches for increasing the throughput of the SSR marker system?
6. Compare the ISSR and RAPD marker systems. Discuss the applications and limitations of these markers.
7. Compare the various features of SRAP, TRAP, and CoRAP markers and discuss their usefulness in breeding programs.
8. Explain the principles of SSCP and D/TGGE markers and discuss their usefulness in breeding programs.
9. Compare the features and merits of CDDP and SCoT markers. How do they differ from RAPD markers?
10. “Transposons have been used to develop several marker systems.” Discuss this statement with the help of suitable examples.
11. “The PCR technology has facilitated the development of a variety of marker systems.” Discuss this statement giving suitable examples.

Appendices

Appendix 3.1: The Number of RAPD Bands Theoretically Expected from a DNA Sample

The number of RAPD bands theoretically expected from a DNA sample can be estimated on the basis of probability concept. It can be shown that the number of RAPD bands (b) of a given average size (f bp) expected from a genome of known size (N bp) amplified using primers of n nt would be given by the following formula:

$$b = 2Nf/16^n \quad (3.1)$$

The above formula is derived as follows. The probability that a specified base would occur at a given site in a DNA strand will be 1/4 since this

site could have any one of the four DNA bases. It is assumed that the distribution of nucleotides/bases is random, i.e., governed by chance, so that the four DNA bases occur in the DNA molecule in equal proportion. Surely, this assumption is unrealistic, but it is necessary for an easy estimation of the above and similar parameters. Therefore, the probability that the n bases present in a RAPD primer will be found in a DNA strand will be $1/4^n$. Exponential amplification can occur only when a second primer binding site occurs in the neighborhood of the first site; the probability of the two primer binding sites occurring together will be $1/4^{2n}$ or $1/16^n$. Since the template DNA has two complementary strands, the primer binding sites could occur on either strand at a given site. In addition, the two primer binding sites would be separated by f bp, i.e., the RAPD fragment size. Therefore, the probability of two primer binding sites occurring in a DNA duplex of f bp would be $2f/16^n$. If the size of genomic DNA were N bp, the number of expected RAPD fragments of f bp would be $2Nf/16^n$.

According to the above formula, a primer of 10 bases is expected to generate 2 bands in rice, which has the genome size of 450 Mb and

4 bands in tomato that has genome of 950 Mb. Similarly, it would produce 9 bands in maize (genome size, 2,500 Mb) and 19 bands in barley (genome size, 5,300 Mb).

Appendix 3.2: Polymerase Chain Reaction and Randomly Amplified Polymorphic DNAs

PCR was developed for amplification of a specific segment from a DNA sample of high complexity, e.g., human genomic DNA. Subsequently, this procedure was applied to achieve a variety of other objectives, for each of which the procedure was suitably modified. In a general sense, the term PCR signifies repeated replication of a segment of sample DNA by using suitable primer(s) and DNA polymerase. In this sense, all applications of the technique would qualify as PCR. But in a restricted sense, the term PCR signifies amplification of a specific sequence from the sample DNA; this PCR procedure differs in many ways from the other applications of the technique. The various features of PCR (in the restricted sense) and RAPDs are summarized in Table 3.3.

Table 3.3 A comparison between PCR and RAPD procedures

Feature	PCR	RAPD
Amplified region	Specified/known	Random
Prior sequence information of the target segment	Essential for designing the specific primers	Not required; primer sequence is arbitrary
Primer sequence	Complementary to the 3' ends of the two strands of the target DNA segment	Arbitrary; specified by the experimenter
Number of primers	Two: one forward and one reverse	One
Primer length	15–22 nucleotides	10 nucleotides
Primer binding sites	Fixed (due to the primer sequence); usually, one	Random; usually, more than one
Annealing temperature	High (around 65 °C; ~1–2 °C less than the T_m of the primer ^a)	Low (around 36 °C; ~5 °C less than the T_m of the primer ^a)
Annealing conditions	High stringency	Low stringency
Number of amplified fragments/bands	One (usually) or few	Several
Reproducibility	High	Moderate to poor

^a T_m of the primer is the melting temperature of the primer–template duplex. T_m of a DNA duplex can be estimated from the following formula

$$T_m = 4(G + C) + 2(A + T)$$

where $G + C$ and $A + T$ represent the numbers of purine ($G + C$) and pyrimidine ($A + T$) residues in one strand of the DNA duplex