Int.J.Curr.Microbiol.App.Sci (2018) Special Issue-6: 816-825

EXCELLENT PUBLISHERS

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 Special Issue-6 pp. 816-825
Journal homepage: http://www.ijcmas.com

Review Article

Molecular Marker Techniques: A Review

R. R. Dhutmal*, A. G. Mundhe and A. W. More

Sorghum Research Station, Vasantrao Naik Marathwada Krishi Vidyapeeth,
Parbhani-431401 (M.S.), India
*Corresponding author

A B S T R A C T

Several marker techniques have been generated in the last decade starting from the first generation molecular marker, RFLP, that was based on DNA-DNA hybridization and later the invention of PCR gave rise to a second generation of faster and less expensive PCR-based markers followed by the third generation sequence based makers. Following the rise of EST databases and whole genome sequencing projects several functional markers have been developed. However, marker techniques are continuously changing and evolving. Among the various marker techniques that are available, particularly promising are Amplified Fragment Length Polymorphisms, Random Amplified Polymorphic DNA, Microsatellites, Sequence Characterized Amplified Region or Sequence Tagged Sites etc. The most appropriate genetic marker technique to be used will depend on the specific application, the presumed level of polymorphism required, the presence of sufficient technical facilities and know-how, time constraints and financial limitations. Molecular markers have become important tools for a large number of applications ranging phylogenetic analysis, diversity studies, construction of genetic maps, comparative maps, framework maps, framework/region specific mapping, novel allele detections, high-resolution mapping, very fast mapping, region-specific marker saturation, gene tagging, marker-assisted selection, map-based gene cloning, varietal/line identification, hybrid identification, seed testing, fingerprinting, alien gene introduction etc. Thus molecular markers are considered as valuable tools for genome analysis even in crops where the whole genomes are sequenced.

Keywords

Molecular maker technique, REPL, PCR, DNA

Introduction

The era of genomics started with the development of genetic tools like the DNA-based molecular markers that have been extensively used in various fields like taxonomy physiology, embryology, genetic engineering, etc. Markers are the traits which can be used to distinguish or differentiate between the populations under study. Markers are the traits which can be used to distinguish or differentiate between the populations under study. Markers are broadly divided into two major category i.e. morphological and molecular. Morphological markers are those traits which are visible to the naked eyes while molecular markers are those which expresses at the molecular level (i.e. DNA or proteins). A molecular marker is defined as a particular segment of DNA that is representative of the differences at the genome analysis. The discovery of PCR (polymerase chain reaction) was a landmark in this effort and proved to be a unique process that brought about a new class of

DNA profiling markers. The random DNA based markers were the first genomic tools that were developed and used for several purposes in crop plants. Molecular markers offer numerous advantages over conventional phenotype based alternatives as they are stable and detectable in all tissues regardless of growth, differentiation, development, or defense status of the cell.

They are not confounded by the environment, pleiotropic and epistatic effects. Markers are ideal tools for examining the relationships among individuals, populations or phylogenetic taxa. An increasing number of monogenic, race-specific genes showing gene-for-gene interaction have been mapped and agronomically important genes have been correlated using molecular markers. These are widely being used in screening programs for selecting disease resistant clones and introgression of several resistant genes into a particular background.

Most of the above mentioned markers are developed from genomic DNA, and therefore they may arise from both the transcribed regions and the non-transcribed regions of the genome. These DNA-based markers derived from any region of the genome have also been described as RDMS. These markers when used for indirect selection are completely independent of any functional knowledge about the underlying DNA sequences.

Thus even for tightly linked markers, the effectiveness of marker aided selection is greatly diminished by the occasional uncoupling of the marker form the trait during many cycles of meiosis in the breeding program. Also the application of such random markers for selection across populations has been limited. Recently, interest has shifted towards the development of molecular markers from the genes that are responsible for the expression of phenotypic trait variations. These markers from transcribed region of the genome, target the functional polymorphism in the gene sequences and allows selection in different genetic backgrounds which isnot always possible random markers.

Properties of molecular markers

Properties of ideal genetic markers

There are certain unique properties which makes any marker as ideal genetic marker. It should not have any detrimental effect on phenotype, co-dominant in expression, single copy, economic to use, highly polymorphic, easily assayed, multi-functional, highly available (Un-restricted use). Genome-specific in nature (especially when working with polyploids), can be multiplexed and its ability to be automated.

Genomic abundance

The number of markers that can be generated is determined mainly by the frequency at which the sites of interest occur within the genome. RFLPs and AFLPs generate abundant markers due to the large number of restriction enzymes available and the frequent occurrence of their recognition sites within genomes. Within eukaryotic genomes, microsatellites have also been found to occur frequently. RAPD markers are another abundant class of markers because numerous random sequences can be used for primer construction. If, in addition to genomic abundance, genome coverage is also sought, caution should be taken in marker selection. While some markers are known to be scattered quite evenly across the genomes, others, such as some AFLP markers, sometimes cluster in certain genomic regions.

Level of polymorphism

The resolving power of genetic markers is determined by the level of polymorphism detected which is in turn determined by the mutation rate at the genomic sties involved. Mutation at mini satellite and microsatellite loci, occur due to changes in the number of repeat units of the core sequence, and have been estimated to occur at the relatively high frequency. Higher resolving power is required when samples are more closely related. For example, analyses within species or among closely related species may call for fast evolving markers such as microsatellites. However if the objective is to study genetic relatedness at higher taxonomic levels, AFLPs or RFLPs may be a better choice because co-migrating fast-evolving markers will have less chance of being homologous. A primary guiding principle in marker selection is that more conservative markers (those having slower evolutionary rates) are needed with increasing evolutionary distance and vice-versa.

Codominance of Alleles

Codominance markers are markers for which both alleles are expressed when co-occurring in an individual. Therefore, with codominant markers, heterozygotes can be distinguished from homozygotes, allowing the determination of genotypes and allele frequencies at loci. Codominant markers are preferred for most applications. The majority of codominant markers are single locus markers and hence the degree of information per assay is usually lower compared to the multilocus techniques.

Molecular markers as genomic tools

Several molecular marker techniques were developed from the genome of the crop species as well as from random amplification of the genome. These techniques include the first generation restriction based markers like RFLP followed by the second generation amplification based markers like RAPD, AFLP. SSR, ISSR and the third generation sequence based markers like SNPs. The RFLP markers which are highly polymorphic, reproducible and are co-dominantly inherited but have been used less frequently owing to the laborious procedure involved. RAPDs and ISSRs had been the marker of choice in the nineties and soon interest had shifted to more reliable and reproducible marker systems like AFLPs and SSRs.

Hybridization based markers

In hybridization based markers, DNA profiles are visualized by hybridizing the restriction enzymes digested DNA to a labeled probe, which can be DNA fragment of known origin or sequence. Restriction fragment length polymorphism (RFLP) was the first hybridization based marker system. Later on many markers probes such as microsatellites, mini satellites and STS were developed initially as hybridization based marker. Within the development of easier cloning and sequencing techniques and availability of sequence databases these markers have been converted in to PCR based markers which are easier to assay.

Restriction Fragment Length Polymorphism (RFLP)

The publication of Botstein *et al.,* (1980) about the construction of genetic maps using restriction fragment length polymorphism was the first reported molecular marker technique in the detection of DNA polymorphism. In RFLP, DNA polymorphism is detected by hybridizing a

chemically labeled DNA probe to a Southern blot of DNA digested by restriction endonucleases, resulting in differential DNA fragment profile. Labelling of the probe may be performed with a radioactive isotope or with alternative non-radioactive stains, such as digoxigenin or fluorescein. Probes are generated through the construction of genomic of complementary DNA (cDNA) libraries and hence may be composed of specific sequence of unknown identity (genomic DNA) or part of the sequence of a functional gene (exons only, cDNA). The hybridization results can be visualized by autoradiography (if the probes are radioactively labelled), or using chemiluminesence (if nonradioactive, enzyme-linked methods are used for probe labeling and detection). RFLPs correspond to DNA fragments, usually within the range of 2-10 kb, that have resulted from the digestion of genomic DNA with restriction enzymes. The differential profile is generated due to nucleotide substitutions or DNA rearrangements like insertion or deletion or single nucleotide polymorphisms. The RFLPs markers are relatively highly polymorphic, codominanltly inherited and highly reproducible. RFLPs are applied in diversity and phylogenetic studies ranging from individuals within populations or species, to closely related species. RFLPs have been widely used in gene mapping studies because of their high genomic abundance due to the ample availability of different restriction enzymes and random distribution throughout the genome (Neale and Williams 1991). Botstein *et al.,* (1980) used RFLP markers for the first time in the construction of genetic map. These are very reliable marker in linkages analysis and can determine if linked gene is present in homozygous or heterozygous stage due to their codominant behavior. They also have been used to investigate relationship of closely related taxa, as fingerprinting tools, for diversity studies, and for studies, and for studies of hybridization and introgression, including studies of gene flow between crops and weeds.

Amplification based markers

The *in vitro* amplification of the DNA by polymerase chain reaction (PCR) has proven to be revolutionary technique in molecular biology. The amplified products can be visualized on a gel in the form of bands. PCR based markers involve *in vitro* amplification of particular DNA sequences or loci with the help of specially or arbitrarily chosen oligonucleotide primers and a thermo stable DNA polymerase enzyme. The amplified products are separated electrophoretically and banding patterns are detected by different methods such as staining and autoradiography. The PCR based markers can be divided in two categories viz., sequence arbitrary markers and sequence dependent markers.

Random Amplified Polymorphic DNA (RAPD)

The RAPD technique is based on PCR amplification of genomic DNA. It deduces DNA polymorphisms produced by "rearrangements or deletions at or between oligonucleotide primer binding sites in the genome" using short random oligonucleotide sequences (mostly ten bases long) (Williams *et al.,* 1990). As the approach requires no prior knowledge of the genome that is being analyzed, it can be employed across species using universal primers. The major drawback of the method is that the profiling is dependent on the reaction conditions so may vary within two different laboratories. Nevertheless due to the speed and efficiency of RAPD analysis, high-density genetic maps has been

developed in many plant species. Variants of the RAPD technique include Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) which uses longer arbitrary primers than RAPDs, and DNA Amplification Fingerprinting (DAF) that uses shorter, 5-8 bp primers to generate a larger number of fragments. Randomly amplified microsatellite polymorphisms (RAMPO) another variant of RAPD and is a PCR-based strategy where the genomic DNA is first amplified using arbitrary (RAPD) primers and the amplified products are then electrophoretically separated and the dried gel is hybridized with microsatellite oligonucleotide probes.

Several advantages of oligonucleotide fingerprinting, RAPD and microsatellite primed PCR are combined, making it speedy, highly sensitive, detecting higher variability and requires no prior DNA sequence information. This technique has been successfully employed in the genetic fingerprinting of closely related genotypes of several crop species. Multiple Arbitrary Amplicon Profiling (MAAP) is the collective term for techniques using single arbitrary primers.

Amplified Fragment Length Polymorphisms (AFLP)

To overcome the limitation of reproducibility associated with RAPD, AFLP technology (Vos *et al.,* 1995) was developed. It combines the power of RFLP with the flexibility of PCR-based technology by ligating primer recognition sequences (adaptors) to the restricted DNA and selective PCR amplification of restriction fragments using a limited set of primers. The DNA is cut with two restriction enzymes, one being a frequent cutter and the other an infrequent cutter. This is followed by ligation of adapters, including restriction motifs followed by a two-step PCR amplification of selected fragments. The selective amplification uses primers composed of the adapters and 1 to 3 selected nucleotides at the 3' end. It limits the number of fragments to a resolvable range. The PCR-amplified fragments can then be separated by gel electrophoresis and banding patterns visualized. A range of enzymes and primers are available to manipulate the complexity of AFLP fingerprints to suit application. The AFLP banding profiles are the result of variations in the restriction sites or in the intervening region.

The AFLP technique simultaneously generates fragments from many genomic sites (usually 50-100 fragments per reaction) that are separated by polyacrylamide gel electrophoresis and that are generally scored as dominant markers. However, by use of automatic gel scanner heterozygote may be distinguished from homozygote based on band intensity differences, which facilitates the scoring of many AFLPs as codominant markers. The AFLP technique generates fingerprints of any DNA regardless of its source, and without any prior knowledge of DNA sequence. Most AFLP fragments correspond to unique positions on the genome and hence can be exploited as landmarks in genetic and physical mapping.

Selectively Amplified Microsatellite Polymorphic Locus (SAMPL)

This technique is a variant of AFLP technique which amplifies microsatellite loci by using a single AFLP primer in combination with a primer complementary to compound microsatellite sequences, which do not require prior cloning and characterization. It is considered more applicable to intraspecific than to interspecific studies due to frequent null alleles.

Single-Strand Conformation Polymorphism (SSCP)

SSCP relies on intra strand (single strand) differences (conformation) in DNA of different sequence. SSCPs are DNA fragments of about 200-800 bp amplified by PCR using specific primers of 20-25 bp. Gel electrophoresis of single-strand DNA is used to detect nucleotide sequence variation among the amplified fragments. The electrophoretic mobility of single-strand DNA depends on the secondary structure (conformation) of the molecule, which is changed significantly with mutation. Thus, it provides a method to detect nucleotide variation among DNA samples without having to perform sequence reactions. In SSCP the amplified DNA is first denatured, and then subject to non-denaturing gel electrophoresis.

Inter Simple Sequence Repeats (ISSR)

ISSRs are DNA fragments of about 100-3000 bp located between adjacent, oppositely oriented microsatellite regions. ISSRs are amplified by PCR using microsatellite core sequences as primers with a few selective nucleotides as anchors into the non-repeat adjacent regions (16-18 bp). Primers based on repeat sequences, such as (CA) n, can be made with a degenerate 3'-anchor, such as (CA) 8 RG or (AGC) 6TY. An unlimited number of primers can be designed for various combinations of di-, tri-, tetra and pentanucleotides with an anchor made up of few bases. The resultant PCR amplifies the sequence between two SSR, yielding multilocus marker system useful for fingerprinting, diversity analysis and genome mapping. About 10-60 fragments from multiple loci are generated simultaneously, separated by gel electrophoresis and scored as the presence or absence of fragments of particular size. PCR products may be separated on agarose gel with ethidium bromide visualization, or polyacrylamide gel with silver staining techniques. The ISSR marker technique is nearly identical to RAPD expect that ISSR primer sequences are designed from microsatellite regions and the annealing temperatures used are higher which makes them more robust. The advantage of greater band amplification is augmented. Techniques related to ISSR analysis are Single Primer Amplification Reaction (SPAR) that uses a single primer containing only the core motif of a microsatellite, and Directed Amplification of Mini satellite region DNA (DAMD) that uses a single primer containing only the core motif of a mini satellite. The main advantage of ISSRs is that no sequence data for primer construction are needed and requires low quantities of DNA.

Simple Sequence Repeat markers (SSR) or microsatellites

Microsatellites are short tandem repeats of 1-6 base pairs. Di-, tri- or tetranucleotide repeats like (CA)n, (AAT)n or (GATA)n are widely distributed throughout the genome of the plants and animals and developed as markers by deducing the flanking regions of the microsatellite. Their polymorphism lies in the variation of the number of repeats, probably because of errors during replication. If nucleotide sequences in the flanking regions of the microsatellite are known, specific primers (generally 20-25 bp) can be designed to amplify the microsatellite by PCR. Microsatellite and their flanking sequences can be identified by constructing a small-insert genomic library, screening the library with a synthetically labelled oligonucleotide repeat and sequencing the positive clones. Alternatively, microsatellites may be

identified by screening sequence databases for microsatellite sequence motifs from which adjacent primers may then be designed. Polymerase slippage during DNA replication, or slipped strand mispairing, is considered to be the main cause of variation in the number of repeat units of a microsatellite resulting in length polymorphisms. The advantages of SSRs include codominance of alleles, high genomic abundance in eukaryotes and random distribution throughout the genome, with preferential association in low-copy regions. It requires low quantities of template DNA (10-100 ng per reaction) and the use of long PCR primers ensures the reproducibility of microsatellites permitting exchange between researchers. SSR assays are more robust than random amplified polymorphic DNA (RAPD) and more transferable than AFLPs. Multiple microsatellites may be multiplexed during PCR or gel electrophoresis if the size ranges of the alleles of different loci do not overlap thus decreasing the analytical costs. Furthermore, the screening of microsatellite variation can be automated, if the use of automatic sequencers is an option.

Sequence based functional markers

Expressed Sequence Tags (ESTs)

Expressed Sequence Tags are the result of sequencing cDNA clones and the information generated is generally stored in databases. The availability of sequence data of expressed DNA has enabled the development of markers that are physically associated with coding regions of the genome. These sequences can then be used for designing primers to readily generate polymorphic markers. These markers are a part of the cDNA/EST sequences and represent functional component of the genome. The development of these sequences as genetical markers involves the identification of polymorphic sequence that affect the plant phenotype within these genes and validating the associations between these genes and trait variations. Recent studies show that these markers are superior than the random DNA based markers for marker assisted selections. The raw sequence information will also aid in screening for the occurrence of microsatellite sequences (EST-SSR) or single nucleotide polymorphisms (EST-SNP), after which markers can be developed that are targeted to transcribed regions of the genome.

Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms are variations in genome sequence of individuals of a population. They constitute the most abundant molecular markers in the genome and are widely distributed throughout genomes although their occurrence and distribution varies among species. Maize has 1 SNP per 60-120 bp, while humans have an estimated 1 SNP per 1,000 bp. The SNPs are usually more prevalent in the non-coding regions of the genome. Improvements in sequencing technology and availability of an increasing number of sequences in the public domain have made direct analysis of genetic variation at the DNA sequence level possible. High throughput genotyping methods, including DNA chips, allele-specific PCR and primer extension approaches make single nucleotide polymorphisms (SNPs) especially attractive as genetic markers.

cDNA-AFLP

Complementary DNA can also be used as template for subsequent direct marker generation, for example through AFLP

technology. cDNA-AFLP is commonly used in the identification of genetic polymorphisms between contrasting phenotypes like resistant and susceptible individuals under controlled conditions in order to facilitate the construction of linkage maps or to identify candidate resistant genes.

In sugarcane differentially expressed cDNA fragments identified during biotic challenge show great potential as genetic markers for pest and disease resistance. The sequences thus isolated have shown associations with smut resistance and sugarcane mosaic virus resistance.

Sequence-Related Amplified Polymorphism (SRAP)

A simple PCR-based marker technique that targets coding sequences in the genome is Sequence-Related Amplified Polymorphism. SRAP uses forward primers consisting of an unspecific filler sequence of ten bases, the sequence CCGG and three selective nucleotides.

Reverse primers also contain a filler sequence, but are followed by the sequence AATT and three selective nucleotides. The CCGG sequence is used to target GC-rich regions, such as exons in open reading frames, while the AATT sequence on the reverse primers is aimed at AT- rich regions, such as promoters and introns.

The generally conserved nature of exon sequences, combined with the generally variable nature of introns, promoters and spacers, enables SRAP analysis to generate polymorphic bands. In *Brassica oleracea* L. sequence analysis revealed that 45% of the SRAP fragments could be matched with known genes.

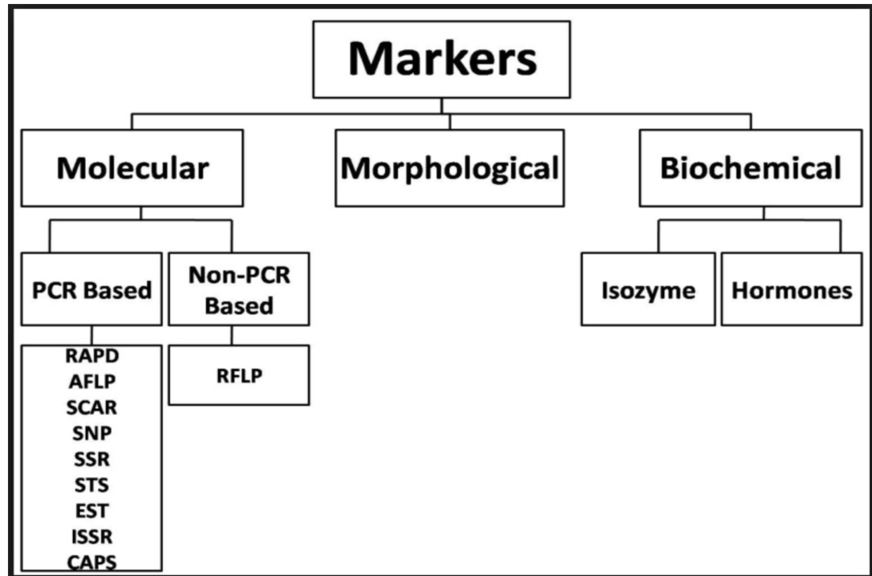
Target Region Amplification Polymorphism (TRAP)

Another related technique that uses EST sequence information is Target Region Amplification Polymorphism (TRAP). For TRAP analysis, a fixed primer designed from a targeted EST sequence is combined with an arbitrary primer having an AT- or CG-rich core sequence. For different plant species TRAP revealed multiple scorable fragments, and the technique may be well suited for determining the genotypes of germplasm and tagging genes for traits of interest. TRAP markers have been used for tagging genes that confer resistance to bacterial, fungal and viral pathogens in common bean.

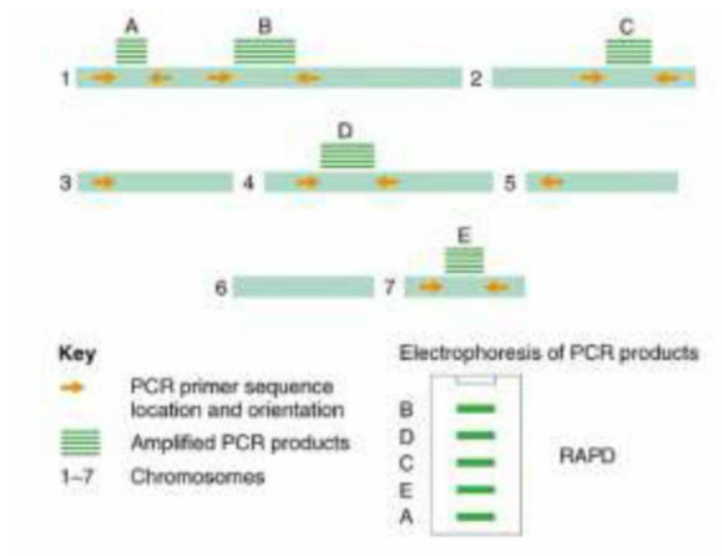
Resistance Gene Homologue Polymorphism (RGHP)

The Resistance Gene Homologue Polymorphism is based on the availability of candidate resistant gene sequences. The use of candidate genes as markers is more effective in the case of disease resistance since many genes involved in the resistant pathways have been characterized. RGHPs target groups of resistance genes by PCR, using primers for conserved domains of resistance genes, such as the Leucine Rich Repeat (LRR) or the Nucleotide Binding Site (NBS), both involved in resistance mechanisms. These RGHPs are then used to identify linkage with known disease resistant loci for use in marker assisted selections as well as to clone the resistant genes. Many RHGPs have been located to chromosome regions containing major R genes as well as QTLs. The cosegregation of RHGPs with major disease resistant genes and quantitative trait loci (QTL) has been reported in several crops species (Pflieger *et al.,* 2001). The disease R-gene database (available on line from the National Center

for Genome Resources web site: http://www.ncgr.org/research/rgenes) facilitates access to R-gene and R-gene-like sequence data collected from public sequence and protein databases. The second database contains information about genes for both pathogen recognition (resistance genes and homologs) and plant defense responses (defense genes). These sequences were mapped on the potato genome and their position was compared to those of resistance QTLs (Nelson *et al.,* 1999).



Random DNA Markers (RDMs)



References

Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. 1980. Construction of a genetic map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32:314-331.

Caetano-Anolles, G., Bassam, B.J. and Gresshoff, P.M. 1991. DNA

amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio Technology, 9:292-305.

Hayashi, K. 1992. PCR-SSCP: A method for detection of mutations. Genetic Analysis Techniques and Applications, 9:73-79.

Li, G. and Quiros, C.F. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica.* Theoretical and Applied Genetics, 103:455^161.

Morgante, M. and Olivieri, A.M. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant Journal, 3:175-182.

Nelson, R., Ghislain, M., Chittoor, J. and Leach, J.E. 1999. Candidate genes for disease resistance. In: Heller S.R. (ed.), International Conference on the Status of Plant and Animal Genome Research VII. Abstracts. San Diego, CA.

Paran, I. and Michelmore, R.W. 1993. Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics 85:985-993.

Pflieger, S., Lefebvre, V. and Causse, M. 2001. The candidate gene approach in plant genetics: a review. Molecular Breeding 7: 275-291.

Vos, P., Hogers, R. M., Bleeker, M., Reijans, T., Lee, M., Homes, A., Frijters, J., Pot, J., Peleman, M., Kuiper, Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23. 4407-4414.

Welsh and McClelland, M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acid Research, 18:7213-7218.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.