A review of the Amplified Fragment Length Polymorphism (AFLP) technique in genotyping and DNA fingerprinting studies

[Review Article]

W.M. Muiru¹*, B. Koopmann², A.V. Tiedemann², E.W. Mutitu¹ and J.W. Kimenju¹

¹Department of Plant Science and Crop Protection, University of Nairobi, P.O. Box 29053, Nairobi, Kenya; ²Department of Crop Sciences, Division of Plant Pathology and Plant Protection, University of Göttingen, Grisebachstr. 6, D-37077 Göttingen, Germany.

*Corresponding author e-mail: wmmuiru27@yahoo.com

ABSTRACT

Amplified fragment length polymorphism (AFLP) is a marker based on polymerase chain reaction amplification of restricted fragments ligated to synthetic adaptors and amplified using primers which carry selective nucleotides at their 3’ ends. The technique generates highly reproducible markers from DNA of any organism and allows high resolution genotyping. AFLP has broad applications and has been used to investigate genomes of different complexity from microbes to higher organisms for purposes of species, strains and varieties identification, systematics, pathotyping, population genetics, simple and complex trait mapping, population genetics, construction of linkage and physical maps. In addition, it is being used in medical diagnostics, forensic analysis and microbial typing. AFLP is superior compared to other markers in that it has time efficiency, generates more information, is highly reproducible and has a wide range of applications. The marker has a drawback in that it generates dominant rather than co-dominant markers and can also be expensive if automated systems are used.

Key words: DNA fingerprints, marker, polymorphism, restriction, sequence


INTRODUCTION

A multitude of techniques are available for examining genetic variation between organisms by detecting polymorphisms in nuclear or organellar DNA (Rehner & Uecker, 1995; Madan et al., 1997; Uptmoor et al., 2003). Unlike protein markers, DNA markers are not subject to environmental influences and they exist in unlimited numbers covering the entire genome (Williams et al., 1990; Rita et al., 2002; Kinyua, 2004). Molecular markers have been used to detect genetic variation among a wide variety of organisms including microorganisms, insects, fish, humans, coral and plants (Sheriff et al., 1994; Brown, 1996; Otsen et al., 1996; Madan et al., 1997; Mueller & Wolfenbarger, 1999). The commonly used techniques include amplified fragment length polymorphism, restriction fragment length polymorphism, random amplified polymorphic DNA, single strand conformation polymorphism, sequence characterized amplified region, simple sequence repeats also known as microsatellites, simple sequence polymorphisms and sequence
tagged microsatellites (Freeman et al., 1993; Lin et al., 1996; McDonald, 1997; Majer et al., 1998; Rita et al., 2002; Wagara, 2004).

Molecular markers are based solely on the detection of naturally occurring DNA polymorphisms. These polymorphisms occur as a result of point mutations or rearrangements such as insertions or deletions in the DNA (Williams et al., 1990; Brown, 1996; Mueller & Wolfenbarger, 1999; Uptmoor et al., 2003; Kinyua, 2004). The polymorphism can be detected by scoring the presence or absence of bands in patterns that are generated either by restriction enzyme digestion or DNA amplification procedures or both (Otsen et al., 1999; Kiprop, 2001). Variations in banding patterns reflect the genetic relationship between organisms and therefore the patterns can be considered as genomic fingerprints that allow numerical analysis of the level of genetic diversity and phylogenetic relationships within and between species and to identify particular races and pathotypes (Freeman et al., 1993; Sheriff et al., 1994; Majer et al., 1998; Rehner & Uecker, 1995; Mueller & Wolfenbarger, 1999; Kinyua, 2004).

Some strains, biovars or races of plant pathogenic organisms are difficult to differentiate based on phenotypic/morphological features and hence require the use of molecular techniques (Freeman et al., 1993; Uptmoor et al., 2003). In fungi and other organisms, the non coding regions of the rDNA have been used as variable regions (Madan et al., 1997; McDonald, 1997). The internal transcribed spacers (ITS) of the rDNA can display variation within genera and can therefore be used to differentiate species (Rehner & Uecker, 1995). At intraspecific level, variability in the intergenic spacers sequence (IGS) that separates ribosomal repeat units is variable enough to allow discrimination of related fungi (Sheriff et al., 1994; Brown, 1996; Lin et al., 1996; Nichole et al., 1997; Rita et al., 2002).

**AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP)**

The AFLP technique was developed in the early 1990s by Keygene, a company in Netherlands. It is a convenient, robust and reliable tool that has been used for genetic mapping/genotype identification, taxonomic and population genetic studies, medical diagnostics, in marker assisted selection, forensic analysis and microbial typing (Nichole et al., 1997; Majer et al., 1998; Savelkoul et al., 1999; Kiprop, 2001; Buhariwalla et al., 2005). AFLP has been used to study a wide variety of organisms including bacteria, nematodes, insects, fungi, plants, corals, humans and fish (Vos et al., 1995; Mueller et al., 1996; Mueller & Wolfenbarger, 1999; Terefework et al., 2001; Fanizza, et al., 2003). AFLP involves detection of the presence or absence of restriction fragments produced by restriction enzyme digestion of template DNA and two cycles of PCR (Otsen et al., 1996; Savelkoul et al., 1999; Terefework et al., 2001; Krauss, 2000; Kinyua, 2004).

In most cases, a few primer combinations are needed to generate an adequate number of polymorphic markers. To reveal differences between closely related or inbred individuals, AFLP markers have to be generated with a series of primer combinations (Nichole et al., 1997; Savelkoul et al., 1999; Terefework et al., 2001; Wagara, 2004). The power of AFLP is based upon the molecular genetic variations that exist between closely related species, varieties or cultivars. These variations in DNA sequences are exploited by the AFLP technology such that “finger prints” of particular genotypes can be routinely generated (Vos et al., 1995; Lin et al., 1996; Majer et al., 1998; Buhariwalla et al., 2005; Laurentin & Karlovsky, 2006).

Compared to other molecular markers the advantages of AFLP are that, it generates more information, is highly reproducible, and it has a wide range of applications. For example, it can be used for polymorphism screening (Steiger et al., 2002), quantitative trait loci analysis (Otsen et al., 1996), genetic mapping, identity, parentage and phylogenetic studies. No prior sequence data for primer construction is needed and hence there is no need for prior knowledge about the genomic make up of the organism (Vos et al., 1995). Furthermore, AFLP is not affected by small variations in PCR amplification parameters since these are performed under conditions of high selectivity/stringency (Krauss, 2000). AFLP method has few errors e.g. mispriming and scoring (less than 2%), the markers are highly abundant and distributed throughout the genome (Savelkoul et al., 1999), they
Digestion / restriction of DNA: The genomic DNA is digested simultaneously with two restriction endonucleases consisting of a six base pair cutter and a four base cutter. Restriction endonucleases are protein enzymes that recognize specific nucleotide sequences and cleave both strands of DNA containing those sequences (Kiprop, 2001; Fanizza, et al., 2003; Buhariwalla et al., 2005). The amount of DNA used varies depending on the volume of the reaction mixture with approximately 500 ng being required for a 20 µl reaction mixture. The 4bp recognition sequence enzyme e.g. Mse I cuts frequently whereas the 6bp recognition sequence enzyme e.g. EcoR I or Pst I cuts less frequently since it is a rare cutter (Vos et al., 1995; Laurentin & Karlovsky, 2006). After cleavage, products with 5' or 3' overhanging single stranded ends may be produced (Mueller et al., 1996; Savelkoul et al., 1999; Krauss, 2000; Kinyua, 2004). Recognition sequences for many enzymes are the same on both strands. Such recognition sequences are said to be palindromic. The sites of cleavage of DNA are determined by their nucleotide sequence and the long DNA fragments are broken into discrete sized fragments (Nichole et al., 1997; Mueller & Wolfenbarger, 1999; Wagara, 2004).

The restriction enzymes are used as sticky ends to “glue” DNA fragments from different sources together. For AFLP, complete genomic restriction is necessary to prevent later amplification of uncut fragments (Nichole et al., 1997; Krauss, 2000; Steiger et al., 2002; Fanizza, et al., 2003; Buhariwalla et al., 2005). Complete digestion is achieved by the use of high quality DNA and an excess of restriction enzyme. Factors such as buffer composition, incubation temperature, DNA methylation and star activity due to digestion under non-standard conditions can hinder the activity of restriction endonucleases (Mueller et al., 1996; Savelkoul et al., 1999; Kinyua, 2004).

In some restriction reactions, the DNA may be digested simultaneously with the two restriction endonucleases. However, in a case where restriction enzymes require different buffer conditions, restriction is first performed with one enzyme (rare cutter) then the buffer composition is altered before adding the next enzyme for the second restriction (Vos et al., 1995; Krauss, 2000; Terefework et al., 2001; Laurentin & Karlovsky, 2006). In cases where two restriction enzymes have totally incompatible buffers, the digestion with the rare cutter is performed first and the DNA is then recovered (usually by precipitation) and resuspended in the buffer appropriate for the second enzyme (Mueller et al., 1996; Fanizza, et al., 2003). Complete restriction of DNA is indicated by a smear after running restricted DNA on 0.8 % agarose gel (Steiger et al., 2002). Restriction digestion of genomic DNA generates the required substrate for ligation and subsequent amplifications (Otsen et al., 1996; Mueller & Wolfenbarger, 1999; Kiprop, 2001; Uptmoor et al., 2003). The restricted DNA can be stored at -20°C if ligation reactions to be performed later.
Ligation reaction: The resulting restriction fragments are ligated to adapters, e.g. Ecor I and Mse I adapters to generate template DNA for amplification. The adapters are double stranded oligonucleotides consisting of a core sequence and an enzyme specific sequence (Steiger et al., 2002). These adapters are ligated to the ends of the restriction fragments (Savelkoul et al., 1999; Wagara, 2004). The ligation of the adapters to restricted DNA alters the restriction site so as to prevent a second restriction from taking place after ligation has occurred (Krauss, 2000). Adapter ligations can be performed in the presence of restriction enzymes such that any fragment to fragment ligations are immediately re-cleaved by the restriction enzyme (Vos et al., 1995; Otsen et al., 1996; Steiger et al., 2002; Kinyua, 2004).

Pre-selective amplification: Depending on genome size, restriction ligation generates thousands of adapted fragments. Only a subset of these fragments needs to be amplified and pre-selective amplification is done to reduce the complexity of the adapted fragment population (Savelkoul et al., 1999; Krauss, 2000; Kiprop, 2001). This is done with primers complimentary to the adapter sequences, each with one selective nucleotide. The pre-selective PCR amplification is performed under highly stringent conditions to ensure that only perfect matches are primed and elongated (Mueller et al., 1996; Nichole et al., 1997; Steiger et al., 2002; Fanizza et al., 2003; Laurentin & Karlovsky, 2006). A primer extension by one nucleotide reduces the number of amplified fragments by a factor of 4. Because of the high selectivity, primers differing by only a single base pair in the AFLP extension amplify different subsets of fragments (Vos et al., 1995; Wagara, 2004). After the pre-selective amplification, a 1:10 or 1:50 dilution of the PCR products is done using TE buffer. If the previous steps have worked well, a clear DNA band is seen on running the PCR products on 1.2% agarose gel (Mueller & Wolfenbarger, 1999; Buhrivallawa et al., 2005).

Selective amplification: Selective amplification is basically similar to the pre-selective amplification, except that the primers used have 2 - 3 selective nucleotides. The number of amplified fragments is further reduced since only subsets of fragments having matching nucleotides at all the three positions are amplified, which further reduces the complexity of PCR products mixture (Otsen et al., 1996; Mueller & Wolfenbarger, 1999; Terefe-work et al., 2001; Kinyua, 2004). A primer extension of two or three base pairs reduces the number of amplified fragments by a factor of 16 and 64, respectively. Up to 4 selective nucleotides can be included in the selective amplification and the ideal extension lengths vary with the genome size (Krauss, 2000; Steiger et al., 2002). By using combinations of primers with different extensions, a series of AFLP amplifications can thus screen a representative fraction of the genome (Mueller & Wolfenbarger, 1999; Fanizza, et al., 2003).

The end sequences of each adapted fragment contains the adaptor sequence and the remaining part of the restriction sequence (Mueller & Wolfenbarger, 1999; Fanizza, et al., 2003). These known end sequences serve as priming sites in the subsequent AFLP-PCR (Kiprop, 2001). Ligation reaction is performed by adding T4 DNA ligase into the restricted DNA together with the adapter ligase solution and incubating for 2 hours at 20°C. After the ligation, the DNA ligase enzyme is deactivated by incubating at 65°C for 10 minutes followed by a 1:10 dilution with T.E. buffer (Nichole et al., 1997; Krauss, 2000; Uptmoor et al., 2003; Laurentin & Karlovsky 2006). The diluted and ligated mixture can then be used for pre-amplification or stored at -20°C until further use. The ideal number of bands generated by AFLP should range between 50 – 100 since too many bands cause smears (Vos et al., 1995). The number of bands depends on the selective nucleotides in the primers and the complexity of the genome (Wagara, 2004; Laurentin & Karlovsky, 2006). When using capillary systems in scoring, the complexity of the banding pattern can be reduced by radioactively labeling the primer directed against the rare cutter adaptor e.g. Ecor I adaptor sequence, since Mse I – Ecor I and Ecor I – Ecor I fragments are a more limiting subset of the total DNA fragments (Mueller et al., 1996; Reineke & Karlovsky, 2000, Papa et al., 2005). After selective amplification, formamide dye is added to the PCR products followed by denaturation for 2 – 3 minutes at 85°C and the denatured products are immediately placed on ice (Savelkoul et al., 1999; Kiprop, 2001; Uptmoor et al., 2003).
Scoring AFLP markers: AFLP – PCR products can be separated and scored using a variety of techniques ranging from simple gel electrophoresis to automated genotyping. Agarose gel electrophoresis, though cheap and user friendly, gives the least resolution and is unable to score fragment length differences of less than 10 nucleotides (Vos et al., 1995; Savelkoul et al., 1999; Kinyua, 2004). Polyacrylamide gel electrophoresis, done either manually or with automated sequence provides the maximum resolution of AFLP banding patterns to the level of single nucleotide differences (Sheriff et al., 1994; Nichole et al., 1997; Mueller & Wolfenbarger, 1999; Kiprop, 2001; Terefework et al., 2001; Fanizza et al., 2003; Brugmans et al., 2003; Papa et al., 2005). Capillary sequencers can also be used that register signals in electropherograms which are analysed by computer software thus eliminating the need for manual scoring and analysis (Otsen et al., 1996; Mueller et al., 1996; Steiger et al., 2002; Buhariwalla et al., 2005). Compared to other methods, the use of capillary systems and automated analysis increases data throughput and scoring reliability, thus decreasing the overall experimental error (Papa et al., 2004).

Analysis of AFLP fingerprints: Bands are assigned numbers in relation to their migration distance within the gel. Bands with the highest molecular weight are assigned number one and so on until the band with the lowest molecular weight (Excoffier et al., 1992; Savelkoul et al., 1999). It is assumed that bands of the same molecular weight in different individuals are identical in sequence (Wagara, 2004). Presence or absence of a band is given a score of 1 or 0, respectively (Steiger et al., 2002; Brugmans et al., 2003). Similarity matrices from binary banding data are derived with appropriate software, e.g. NTSYSpc-Numerical Taxonomy and multivariate analysis system for personal computer (Saitou & Nei, 1987; Rohlf 1998; Kiprop, 2001; Steiger et al., 2002; Laurentin & Karlovsky 2006). Estimates of similarity are done using coefficients such as Jaccard and matrices of similarity analysed using the unweighted pair group method with arithmetic averages (UPGMA) clustering method (Sheriff et al., 1994; Mueller et al., 1996; Fanizza, et al., 2003). Dendograms, principle coordinate analysis and genetic distance matrices can be generated using appropriate software to help in interpreting the genetic relatedness of the organisms or strains being studied (Saitou & Nei, 1987; Excoffier et al., 1992; Vos et al., 1995; Savelkoul et al., 1999; Terefework et al., 2001; Brugmans et al., 2003; Wagara, 2004; Buhariwalla et al., 2005; Papa et al., 2005).

Conclusion
The reliability of AFLP coupled with its wide applications makes it a useful marker in pathotyping, systematics, mapping of quantitative trait loci, population genetic studies and in differentiating closely related organisms. Due to its high sensitivity to minor genetic variations, AFLP will remain a key molecular tool, even replacing several other techniques such as RFLP and microsatellites. However, due to its high cost and dominant outcomes, the decision to use AFLP should be made based on the research question being addressed and availability of facilities. Although other codominant markers such as microsatellites allow more powerful population genetic analysis, AFLP is more reliable and easier to use and thus will remain a popular marker in DNA fingerprinting studies.

References


