Review TRENDS in Plant Science Vol.12 No.3

Almost Forgotten or Latest Practice? AFLP applications, analyses and advances

Heidi M. Meudt[1,2] and Andrew C. Clarke[1,3]

[1] Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Private Bag 11222, Palmerston North 4442, New Zealand
[2] Museum of New Zealand, Te Papa Tongarewa, PO Box 467, Wellington 6140, New Zealand
[3] Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

Amplified fragment length polymorphism (AFLP) DNA fingerprinting is a firmly established molecular marker technique, with broad applications in population genetics, shallow phylogenetics, linkage mapping, parentage analyses, and single-locus PCR marker development. Technical advances have presented new opportunities for data analysis, and recent studies have addressed specific areas of the AFLP technique, including comparison to other genotyping methods, assessment of errors, homoplasy, phylogenetic signal and appropriate analysis techniques. Here we provide a synthesis of these areas and explore new directions for the AFLP technique in the genomic era, with the aim of providing a review that will be applicable to all AFLP-based studies.

AFLP: an established molecular marker technique

The Amplified Fragment Length Polymorphism (AFLP)* technique has come a long way since its publication in 1995 [1], including many technological advances in generating and analysing AFLP data. AFLP has become the method of choice for many studies on plants and, more recently, for animals, fungi and bacteria [2], spanning numerous disciplines in genetics, evolution and ecology. The (predominantly) nuclear origin of AFLPs is attractive because markers derived from the uniparentally inherited organellar genomes (chloroplast and mitochondrial genomes) might not be sufficiently variable, or even appropriate – particularly in plants where processes such as hybridization are important.

AFLPs are generated by complete restriction endonuclease digestion of total genomic DNA, followed by selective PCR amplification and electrophoresis of a subset of the fragments (Box 1), resulting in a unique, reproducible (Box 2) fingerprint (or profile) for each individual (see Glossary) [3]. The markers that make up the fingerprint, although often concentrated in centromeric regions [4,5], are widely distributed throughout the genome, allowing an assessment of genome-wide variation. These anonymous markers consist largely of non-coding DNA [6,7].

AFLP is useful in a wide range of applications including linkage mapping [5,8,9], parentage analysis [10], measuring

Glossary

AFLP genotype: the genetic constitution of an individual inferred from an AFLP fingerprint.
Allele: alternative form of a genetic locus. For a single marker (locus), plus and null alleles, although different, are deemed to be homologous (i.e. possess a common evolutionary origin). When the term 'allele' is used across different markers (e.g. to refer to any visualized fragment), it should be clear that these alleles will be derived from different loci, and therefore will be mostly non-homologous.
Band or peak: an AFLP DNA fragment visualized in a fingerprint (a plus allele). 'Band' is usually applied to a fragment visualized using a gel-based system, and 'peak' to a fragment visualized using a fluorescent system.
Character: a marker that is scored in multiple individuals and included in the data matrix. For AFLP, there are two possible character states: '1' (present in an individual; plus allele) or '0' (absent in an individual; null allele).
Codominant marker: a marker that enables homozygous (*AA*) and heterozygous (*Aa*) states to be distinguished. For many applications, particularly in population genetics, codominant markers are more powerful than dominant markers, enabling allele frequencies to be estimated, and require smaller sample sizes to achieve equivalent analytical power. Codominant markers include microsatellites and single nucleotide polymorphisms (SNPs).
Dominant marker: a marker that is scored as a present (plus) or absent (null) allele. Dominant markers cannot distinguish homozygous (*AA*) and heterozygous (*Aa*) states – a band or peak is present in both cases. AFLPs, RAPDs and ISSRs are mostly dominant markers.
Fingerprint or profile: the complete set of AFLP bands visualized for a given sample. Here we consider 'fingerprint' and 'profile' as synonymous, but the former is sometimes used in a more stringent sense – when the pattern of bands uniquely identifies an individual.
Fragment: a single-stranded AFLP PCR product. Depending on size and strand, not all fragments will be visualized as peaks or bands.
Locus (*pl.* loci): a specific region of the chromosome corresponding to the position of a marker; also the DNA at that position.
Marker: an amplified locus that is identified in AFLP as bands or peaks of equal fragment size across multiple samples. A marker must be polymorphic (i.e. show both plus and null alleles) to be informative. In the AFLP fingerprint, a marker usually occupies a narrow ~1 bp window (bin), so that, across all samples, visualized fragments of approximately equal mobility are treated as homologous alleles derived from a single locus. Homoplasy occurs when, by chance, non-homologous alleles of equal mobility fall into the same bin and, therefore, are treated as a single marker. In some cases, there can be two or more markers in a fingerprint that are derived from a single locus (e.g. a variable microsatellite). Although these markers should be treated as a single, codominant, marker, this is often not possible (see main text).
Null allele: for a given marker, the inferred allele when the band is absent.
Plus allele: an allele present as an amplified DNA fragment (band or peak).

Corresponding author: Meudt, H.M. (heidim@tepapa.govt.nz).

* AFLP® technology is covered by patents and patent applications owned by Keygene N.V.; AFLP is a registered trademark of Keygene N.V. Although 'AFLP' was not originally intended as an acronym [1], we treat it as one here, consistent with widespread usage.

Available online 14 February 2007.

 1360-1385/$ – see front matter doi:10.1016/j.tplants.2007.02.001

genetic diversity [11,12], identifying hybrids [13] and cultivars [14], population genetics [15–17], reconstruction of shallow phylogenies [18–20], population assignment [21], and developing single-locus sequence-characterized amplified region (SCAR) markers [6,22–25]. But, even though AFLP is an established and useful molecular marker technique, how important will it be in the genomic era?

How do AFLPs compare to other genotyping methods?

The three most common techniques for multilocus genomic fingerprinting are AFLP, random amplified polymorphic DNAs (RAPDs), and inter simple sequence repeats (ISSRs). They are PCR-based techniques that use primers to amplify previously uncharacterized DNA fragments and, therefore, can be used on organisms for which there

Box 1. Overview of the AFLP technique: fluorescent AFLPs

Because of its high throughput and high data quality, capillary electrophoresis of fluorescently labelled AFLPs is gradually replacing gel-based systems. However, a survey of recent papers indicates that this gel to capillary transition is slow, probably because of limited funding (set-up costs can be high), access to equipment, know-how, and occasionally purpose (e.g. silver-stained gels are better for isolating and cloning individual fragments).

In Step 1 in Figure I, genomic DNA is digested with a pair of restriction endonucleases (usually *Eco*R I and *Mse* I), producing three species of DNA fragments (i.e. Eco–Eco, Eco–Mse, and Mse–Mse). In Step 2, double-stranded *Eco*R I (Eco) and *Mse* I (Mse) linkers (synthetic DNA adapters) with complementary sticky ends are ligated to the restriction fragments. (Steps 1 and 2 can be performed in the same reaction.) In Step 3, the pre-selective amplification, a subset of all the fragments is amplified, using primers that are complementary to the linker sequences with the addition of one nucleotide (A, G, C or T) at the 3′ end of the primer (usually Eco+A and Mse+C). These 'pre-amp' primers will only prime DNA synthesis of fragments with bases flanking the restriction sites that are complementary to the selective nucleotides of the primers, thus reducing the number of fragments to ~1/16 of the initial amount. In Step 4, the number of fragments is further reduced – to a suitable number to be visualized by electrophoresis – by a second round of PCR (selective amplification), in which the PCR primers have an additional two selective bases (e.g. Eco+ATA and Mse+CAC). The Eco+3 primer is labelled with a fluorescent dye (a 'fluorophore'), so that all strands synthesized from this primer are fluorescently labelled (i.e. one strand from Eco–Mse fragments, both strands from Eco–Eco fragments). Alternative subsets of loci can be amplified by using combinations of primers with different selective bases.

For each individual, different Eco+3 selective primers can be labelled with different fluorophores, enabling the products from different primer combinations to be pooled for capillary electrophoresis (Figure II). This process is often incorrectly termed 'multiplexing', but the more accurate 'poolplexing' is preferred. The capillary instrument detects fragments present in the spectrum of each fluorophore, producing an electronic profile of relative fluorescence units (RFU) versus fragment size (usually 50–500 bp). Polymorphisms, which are observed as peaks present in some samples and absent in others, are caused by the gain or loss of a restriction site, a change (e.g. SNP) in the selective primer binding site, or a length polymorphism (e.g. indel or variable microsatellite) between the restriction sites [1,3,81]. Profiles from multiple individuals are aligned and scored (see Box 3) based on the presence (1) or absence (0) of a peak, producing a binary data matrix.

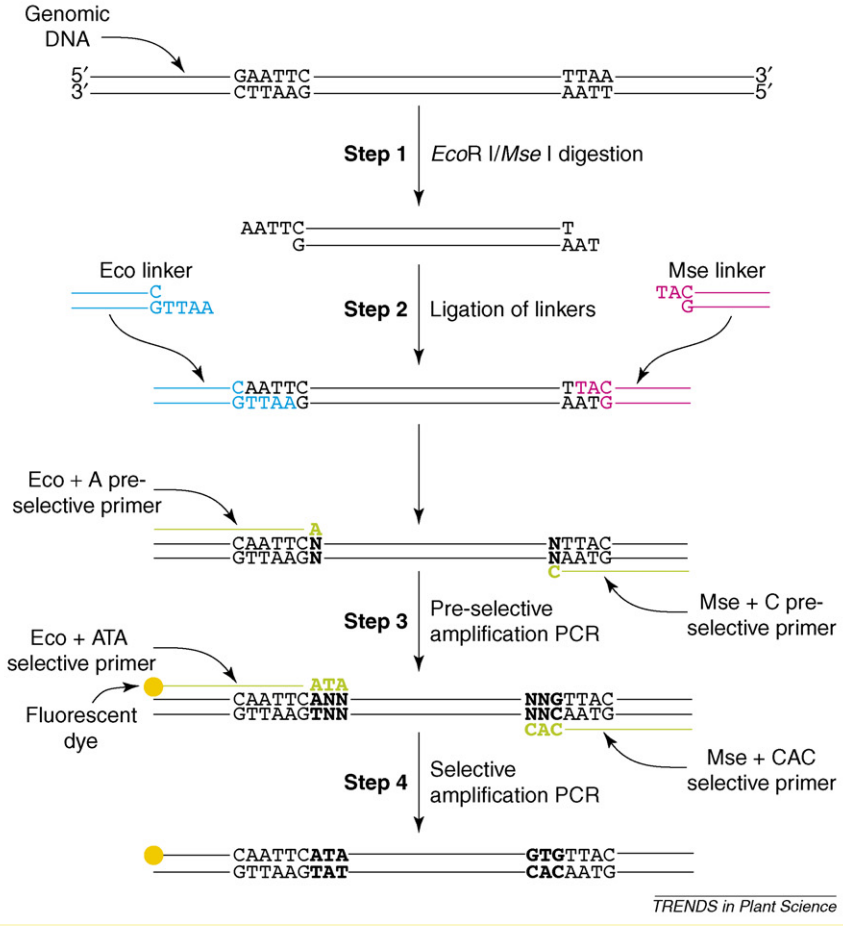

Figure I.

Box 1 cont'd.

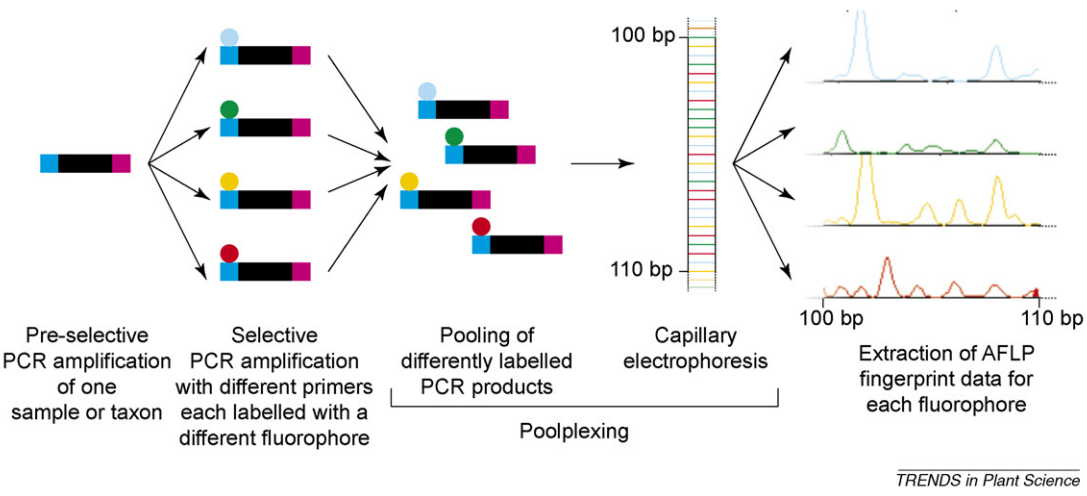

Figure II.

is no *a priori* sequence information [26]. These three techniques vary with respect to data quality, genetic variability and discriminatory power [5,14,22,27,28]. All three techniques show similar patterns of genetic distance and informativeness [12,16,28], particularly for autogamous crops and inbred lines [27]. Nevertheless, in many studies AFLPs outperform ISSRs and RAPDs in their high reproducibility (Box 2), robustness, informativeness, and fewer reported reaction artefacts [14,16,22,27–30].

AFLPs, which like RAPDs and ISSRs are dominant marker systems, can be an attractive alternative to codominant markers such as microsatellites (or simple sequence repeats, SSRs) and single nucleotide polymorphisms (SNPs) [2,10,16,21,26,29,31]. Microsatellites typically comprise a few (5–20) highly informative multiallelic loci with high discriminating capacity. By contrast, dominant marker systems such as AFLP have numerous, genome-wide di-allelic loci that are individually less informative but derive their statistical power from their sheer number [11,16,21,32]. Both types of marker systems are commonly used for linkage analysis [4] and for measuring population genetic structure and diversity [17]. The few explicit comparisons of the relative utility of microsatellites and AFLPs in plants (summarized in Ref. [17]) show that the two systems give congruent results when a sufficiently large number of microsatellite loci are analysed (e.g. 18 microsatellite loci in the study of the fern *Athyrium distentifolium*). In these studies [17], the large number of AFLP fragments generated mean that AFLPs can out-perform microsatellites for discriminating taxa and populations [17,21], although comparisons are difficult when only a few microsatellite loci have been used [17,21,29]. Other variables, such as homoplasy, genomic heterogeneity and population heterogeneity can shift the boundary at which a given number of AFLPs becomes more informative than a given number of microsatellite markers [11,33].

Although the AFLP technique is ideal in many cases, its use should be preceded by careful consideration of factors that determine which marker system is most appropriate for a particular research project. These considerations fall under three broad categories.

(1) Biological

Including the amount of genetic variability (if known), taxonomic breadth (e.g. inter-specific, intra-specific), genome size, occurrence of hybridization, and ploidy [17].

(2) Research questions

Including the application of the technique (e.g. linkage mapping, population genetics, phylogenetics) [16,26,32], whether codominant markers are required, the need for cross-study comparisons, and the life-expectancy of the research project (i.e. cost–benefit analysis of different marker systems through time; AFLPs can be better in the short-term, microsatellites better in the long-term) [17,21,31].

(3) Available resources

Including the quality and quantity of available tissue, previously established genetic resources [e.g. linkage maps, markers established in the same or closely related taxa, sequence data, expressed sequence tag (EST) libraries], and logistics (e.g. available funding, technical difficulty, access to training, and laboratory facilities) [2,16,17,29].

The AFLP technique can be ideal in the following situations [2,3,26,32]:

- When there is no *a priori* sequence information.
- For intra-specific studies.
- When genomic heterogeneity is high (i.e. when it is necessary to amplify many loci to ascertain an accurate measure of genomic diversity, e.g. outcrossing species).
- When genetic variability is low (i.e. when it is necessary to amplify many loci to locate the few that are polymorphic, e.g. crop species).
- In polyploids.
- When hybridization is occurring.
- For the rapid generation of data.

Box 2. Setting up and implementing an AFLP study

Commercial AFLP kits or DIY?
Commercial AFLP kits (usually from Applied Biosystems or Invitrogen) offer the advantages of convenience and some level of technical support. However, assembling your own set of reagents can reduce the consumables cost by 50%, and is suitable for larger projects and more experienced users [2]. Protocols are readily available on the Web and in the literature (e.g. Refs [2,82]; see: http://awcmee.massey.ac.nz/aflp/AFLP_Protocol.pdf).

DNA
Successful AFLP digests require ~100–1000 ng high molecular weight DNA (i.e. not obviously degraded) that is free of contaminants (e.g. inhibitory compounds and non-target DNA) that could otherwise interfere with the digestion, ligation and amplification steps [1,2,28,56,81]. Commercial DNA extraction kits (e.g. Qiagen DNeasy) can give better quality DNA than some other methods [56]. Researchers should not use DNA samples obtained from different extraction methods because the method can affect the resulting AFLP profile [56]. Use of degraded and/or smaller quantities of DNA (such as that often obtained from herbarium material) can result in poor quality profiles with low reproducibility [2]. Whole genome amplification (WGA) techniques have the potential to enable AFLP fingerprinting in situations where previously insufficient quantities of tissue were available (e.g. herbarium material or small individuals [83]).

Choice of restriction enzymes
The restriction enzymes *Eco*R I (a six base cutter) and *Mse* I (a four base cutter) are used in most AFLP studies (but see Refs [1,28,81]), yielding fragments in an appropriate size range for amplification and electrophoresis. Alternatives to *Mse* I include *Taq* I (which can produce better quality results [2,82]) and *Tru* I (a cheaper isoschizomer of *Mse* I) [2]. *Pst* I, the most common alternative to *Eco*R I, is methylation-sensitive, and although it might be appropriate for differential gene expression and some mapping applications [73], it can have undesirable effects for most other applications of the technique (i.e. when differences in gene expression between samples could affect the AFLP profiles).

Choice of selective primers – length, composition and screening
The number of selective nucleotides on the selective primers should be increased with increasing genome size so that the number of fragments is high enough to maximize resolution but low enough to minimize homoplasy. This ranges from Eco+2–Mse+3 primers for small genomes to Eco+4–Mse+4 for larger genomes [70,81]. Previous AFLP studies on related taxa or those with similar genome sizes provide the best guide for appropriate length selective primers (see http://www.rbgkew.org.uk/cval/homepage.html).

The 'quality' of AFLP profiles varies widely between selective primer combinations partly because of the base composition of the selective primers [2]. Therefore it is necessary to screen potential combinations on a small number of samples (~10) before embarking on the full project [81]. High-quality profiles have well-separated peaks, a high signal-to-noise ratio, a lack of shoulder or stutter peaks, fragments distributed throughout the available size range, and clear polymorphisms. In a screen of 32 primer combinations assessed (subjectively) using these criteria, we found 20% of primer combinations produced profiles suitable for high-throughput genotyping. If a fluorescent system is used, it will normally only be economical to screen different unlabelled primers (e.g. Mse+3) and, therefore, the selective bases of the fluorescent primers (e.g. Eco+3) will need to be decided *a priori* (usually based on a literature survey of successful fluorescent primers).

The number of primer combinations required depends on the application and biological question. It should be determined by measuring the proportional increase in resolving power and decrease in error with the accumulation of data from each additional primer combination [84].

Fluorophores
Fluorescent labelling has dramatically increased the output of AFLP fingerprinting by enabling poolplexing of differently labelled products (up to four, plus a size standard for Applied Biosystems' Genetic Analysers). Choice of fluorophores is largely determined by the available electrophoresis system (gel or capillary) and software because potentially significant problems can occur with non-recommended dyes, including weak fluorescence and interference between emission spectra (spectral 'bleed-through') and absorption spectra of poolplexed fluorophores. Even with recommended set-ups there will be differential amplitude of emission between fluorophores, which can be compensated for by empirically determining the optimum pooling ratio.

Duplication, randomization and reproducibility
Although AFLPs are highly reproducible [3,85,86], replicate or duplicate AFLP profiles – preferably from separate DNA extractions of a single individual – should be generated for at least 5–10% of all samples. These should represent all treatments (e.g. DNA extraction method, position on plate or gel and time stored in refrigerator) [87]. This is crucially important for AFLP because replicates are the only objective measure of quality (unlike DNA sequencing, where correct nucleotides can be determined with a high degree of confidence). The same subset of samples should be included as positive controls in every electrophoresis 'run' to ensure between-run reproducibility, and to act as anchor points to detect errors in sample order (e.g. mistakes in plate orientation). To enable any positional biases to be identified, sample order should be randomized (e.g. order should not reflect evolutionary relationships or DNA extraction method). Samples should be anonymously labelled to prevent any investigator-associated scoring biases. To ensure reproducibility, it is essential to standardize the method and maintain consistency for the duration of the study. In addition to the factors already mentioned, factors such as fluorescent dyes, size standard, laboratory equipment and capillary instrument can affect reproducibility and comparability of AFLP profiles. For example, it is advisable not to change fluorophores mid-project, because different fluorophores have different emission properties, which might make the resulting data incomparable.

Error rates in AFLP data
Quantifying genotyping error rates is an essential component of an AFLP study. Because it is usually not possible to know the 'true' genotype of an individual, error rates cannot be assessed directly but instead must be estimated using replicates [87]. Using replicates, the error rate per locus has been estimated at between 2–5% for AFLP [3,85], but unfortunately this is not explicitly calculated in most empirical studies [87]. There are multiple causes of genotyping errors in AFLPs [59,85,87], including the technical aspects of generating the profiles (e.g. PCR stutter, non-specific amplification), subjectivity or human error in (mis)reading the profiles, and differences in peak mobility and intensity in the fingerprint profiles. Although these errors might not bias the results of the analysis [85], they cause a reduction in the signal-to-noise ratio and hence a loss of resolving power. Several strategies have been proposed to reduce errors in AFLPs [85,87] (see Box 3), and some software has been developed for finding and removing errors from AFLP data [86] (see Table 3 in Ref. [87]).

- When high quality DNA is available.
- Where there are no suitable established markers.
- Where there is access to the appropriate facilities (preferably including a capillary electrophoresis instrument).

Appropriate analysis techniques for AFLP data

Linkage mapping

AFLP data can be linked with other sources of data, including RAPDs, restriction fragment length polymorphisms (RFLPs), and microsatellites to produce linkage

maps in established mapping populations (such as potato, barley, rice and *Arabidopsis* [4]; reviewed in Ref. [30]). AFLP characters that do not deviate from expected 1:1 segregation ratios in a chi-square test are then analysed with specialized software [e.g. MapMaker [34], JoinMap (see http://www.joinmap.nl/)] for linkage analysis. The AFLP data are often complementary to the other data sources, and the resulting linkage maps are better resolved. In some cases, accurate AFLP-derived linkage maps for non-model species can be produced using genetic resources from related model species [8,35].

Population-based methods

Most other data analysis methods for dominant multilocus markers including AFLPs can be roughly separated into two main groups [36,37]. The first are population-based, and use a comparison of allele frequencies to partition genetic diversity. Calculating allele frequencies (i.e. heterozygosity) from dominant markers is difficult because the presence of a band (or peak) can indicate either the homozygous condition (i.e. with two plus alleles) or the heterozygous condition (i.e. one plus allele and one null allele); therefore the frequency of the null allele must be estimated. This can be accomplished by using a Bayesian approach [38,39], or the inbreeding coefficient and the square root of the frequency of the null homozygote (if Hardy–Weinberg Equilibrium (HWE) is assumed) [40]. For outcrossing species, whose allele frequencies usually do not violate HWE, both of these approaches can yield good estimates of average heterozygosity [38], whereas for species such as self-fertilizing plants, the Bayesian approach (which does not assume HWE) appears to be superior [39]. Other recently developed methods have the potential to improve estimates of genetic diversity by not assuming HWE [41], enabling estimates via a modelling approach in species where the inbreeding coefficient is unknown (e.g. tropical trees [33]), and incorporating uncertainty about the magnitude of inbreeding into estimates [42].

Phylogenetic methods

The second group of data analysis methods are individual-based and assess genetic relationships among the sampled individuals. The discovery that numerous AFLP datasets contain phylogenetic signal [43,44], has recently led to use of AFLP data in phylogenetic reconstruction, particularly for groups of closely related organisms, such as recent species radiations [15,22,43,45,46], ring species [47], and crops and their wild progenitors [48,49]. Intra- and interspecific phylogenetic studies of plants [46,50,51] (as well as other organisms [15,28,30,44]) are becoming increasingly common. Because AFLP markers are sampled throughout the genome, they are likely to uncover rare genetic differences in groups with low sequence variation [44], and have been shown to result in well-resolved trees that are consistent with independent data [22,43,52].

For phylogenetic methods, the binary AFLP matrix (Box 3) is analysed directly, or converted into a distance matrix using dissimilarity measures [37,53,54]. Then, this matrix is subjected to statistical analysis (e.g. principal coordinate analysis [17,47]; analysis of molecular variance, AMOVA [12,51]), and/or tree-building algorithms, including distance-based methods (e.g. neighbour joining [52,55], UPGMA [19,56]) and character-based methods (e.g. parsimony [46,48], Bayesian analysis [45,50]). Robustness of trees based on AFLP data are often assessed by calculating bootstrap support values (e.g. Refs [18,43,45]), using and comparing multiple analysis methods (e.g. Refs [18,45,46,50,51]), and assessing congruence with results from other data sources (e.g. Refs [15,43,55]). Because some dissimilarity measures only take into account shared presences (e.g. Jaccard, Dice, Nei and Li distances), it has been argued [7,54] that they are more appropriate for AFLP data than those that also incorporate shared absences (e.g. Euclidean distance, simple match coefficient); shared absences (null alleles) are particularly susceptible to homoplasy because of the multiple, independent ways in which a fragment can be lost [22,53,54]. Others have far more negative views, suggesting that all the above dissimilarity measures are unsuitable for phylogenetic analysis on most multilocus dominant marker datasets [53], or indeed that all tree-building analyses are inappropriate for AFLP data [36,53]. However, some empirical studies that have examined the tree-like properties of AFLP datasets have been far more encouraging [20,43,49].

The main problem with all tree-building analyses is that, because of homoplasy, measures of genetic dissimilarity between taxa might not be additive and, therefore, they cannot be expected to recover the correct topological relationships. For example, non-homology of shared fragments [52,57], independent losses of a fragment (shared absences) [36,44], the largely dominant nature of markers [53], the presence of ancestral polymorphism [18], and the fact that high similarity between two individuals does not necessarily reflect shared ancestry [22,53], can all contribute to this problem. However, this situation is not unique to AFLP data given that distance- and character-based analyses of DNA sequence data are also plagued by homoplasy and an inability to appropriately correct for multiple substitutions. Instead of avoiding tree-building with AFLP data, splits graph methods (NeighborNet, split decomposition, and consensus and super networks) can be used to detect non-additivity of data and help explore and interpret conflicting signal [49].

Importantly, combining AFLP data with DNA sequence data can result in more robust phylogenies, possibly because of the complementary effect of the different datasets that provide resolution at different depths of the tree [18,43,46] as well as the increase in the total number of characters available for phylogenetic analysis. It is not yet known precisely where, in terms of genetic divergence, the signal-to-noise ratio becomes too low for AFLP data to be informative, although it appears generally accepted that this line is taxonomically somewhere between intraspecific [8,28,37] and intrageneric [19,22,43] comparisons. To this end, some authors have attempted to quantify how informative AFLPs are for phylogenetic purposes by comparing them to internal transcribed spacer (ITS) sequence data [43], and by developing significance tests to identify and discard AFLP profiles that are too divergent to be analysed together [58]. More studies using simulations (e.g. Ref. [54]), theoretical models (e.g. Ref. [53]) and new phylogenetic tools such as splits graph methods (e.g. Ref. [49]) are

Box 3. Data scoring – a crucial step

The generation of a binary (0,1) matrix from raw AFLP data is a challenging process; difficulties include determining which fragments observed in multiple taxa are truly homologous (and therefore should comprise a single character) [58], intensity thresholds above which fragments are scored as 'present', and the treatment of artefactual shoulder peaks associated with larger peaks [87]. The challenge in scoring AFLPs is to maximize the signal-to-error ratio by optimizing parameters such as peak height thresholds (the intensity above which a peak is scored), bin widths and positions (the size range in which all peaks are considered homologous and thus a single character), and the minimum fragment size that is scored.

Figure I shows two common sources of genotyping error in scoring AFLP data – variation in fragment intensity and mobility (see also Ref. [85]). In Figure Ia, if the upper intensity threshold (dashed line) is used, peak A_1 in taxon 1 is scored as present and A_2 in taxon 2 as absent. Although the threshold can be lowered (short dashed line) to include A_2, this causes the same problem to now occur with peaks B_1 and B_2, where B_1 is scored as absent and B_2 as present [81]. In Figure Ib, peaks differ in size by 0.8 bp. The positioning of the bins (1 bp wide) dictates whether these peaks are split into separate characters (yellow) or grouped into a single character (blue).

Data scoring is a crucial step but its associated problems are widely recognized. Scoring is an area with a great deal of research potential, and strategies to identify and limit errors are improving [63,81,85,87]. Capillary electrophoresis systems now enable precise estimates of fragment mobility (size) and fluorescence amplitude, potentially enabling fully automated scoring, which, in contrast to manual scoring, is objective, repeatable and far less time consuming [87]. (Even with automated scoring, some manual checking might be required to detect and discard low-quality profiles and individual markers that are particularly vulnerable to scoring errors [81,85,87]). With increasingly large datasets, automated scoring is often the only feasible option but at present it cannot be completely implemented owing to the limitations of the available software and a lack of experimental and theoretical research on different scoring parameters (Table 1). Most AFLP scoring software allows the control of several parameters (e.g. bin width, amplitude threshold, minimum fragment size). Use of optimized parameter settings significantly increases the resolution and quality of the resulting data matrix by increasing the number of characters and reducing homoplasy (e.g. Ref. [60]), although these settings might need to be determined empirically for each dataset (Barbara Holland, personal communication). Software must be improved to make use of replicates to calibrate and assess the quality of the data, and to integrate an automated method of objectively and systematically choosing the optimal parameter settings for data scoring.

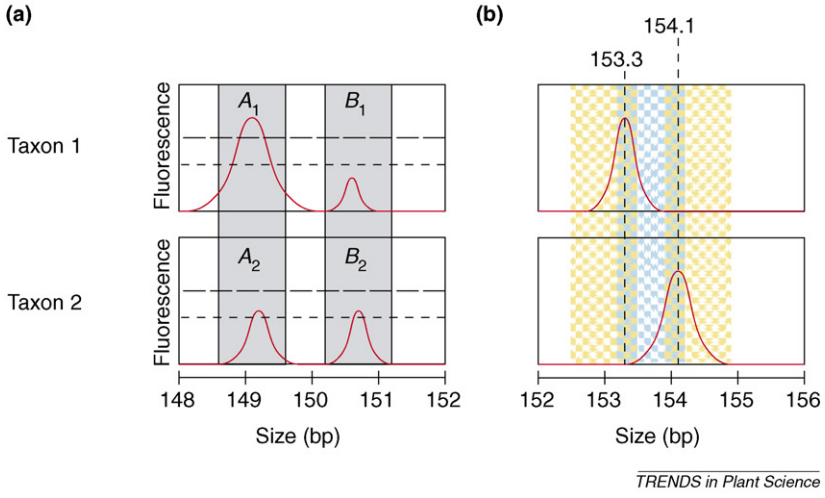

Figure I.

needed to stimulate novel analysis methods that move beyond the standard frequency and phylogenetic analysis methods outlined above.

Assessment of homoplasy in AFLP data

Homoplasy is a major issue in the analysis and interpretation of AFLP data. It occurs when different accessions are incorrectly scored (Box 3) as having a shared character state as a result of either the co-migration of non-homologous fragments (shared '1' character state), or independent losses of a fragment (shared '0' character state). Consequently, homoplasy results in an underestimation of genetic diversity among samples and a loss of resolution in the analyses [59,60]. However, testing the homology of individual AFLP fragments via cloning and sequencing is outside the scope of most studies [44] – even for a small subset of the fragments generated.

A few studies, at both intra- and interspecific levels, have estimated homoplasy in AFLP data by sequencing AFLP fragments. In these studies all or nearly all sequenced fragments have been shown to be homologous with high sequence identity (>99% [44]; >95% [8]), localized in the same areas on genetic maps [8], or identical to fragments predicted by *in silico* analyses [61,62]. It has also been shown that in interspecific studies of *Echinacea* [57] and *Hordeum* [52], homology decreases and homoplasy increases with increasing taxonomic rank, but because taxonomic rank is not equivalent across groups, unless homology is compared to evolutionary distance or genetic divergence it is difficult to interpret the general relevance of these results.

Two different groups independently published a simple experiment that they claimed could quantify homoplasy in AFLP datasets by comparing the banding pattern of an Eco+3–Mse+3 primer pair (e.g. Mse+CAC) to those of the four derived Eco+3–Mse+4 primer pairs (e.g. Mse+CACA, +CACT, +CACC and +CACG). In a study of sugar beet, 87% of the bands from profiles of an Eco+3–Mse+3 primer pair showed the expected pattern of being found in only one of the Eco+3–Mse+4 profiles [59]. The remaining 13% of

bands were found in the profiles of more than one Eco+3–Mse+4 primer pair, which the authors concluded was largely because of size homoplasy. This method also resulted in estimates of 0–5% homoplasy among species of two thistle genera [63]. However, these figures represent a maximum estimate of homoplasy in AFLP data because problems of primer mispairing [1] and poor resolution of bands on gels, which are inherent in the AFLP technique, might also play a role.

Simulations are yet another way to quantify homoplasy: one such study showed that 30% of the fragments amplified during the generation of AFLPs were not actually detected – this masking was attributable to size homoplasy [60]. There also appears to be a negative relationship between homoplasy and fragment size for the two plant species under study (*Phaseolus lunatus* and *Lolium perenne*) [60]. If nothing else, the experimental and theoretical studies described above have established that homoplasy is a real component of AFLP datasets. The quantification of homoplasy in many AFLP datasets both experimentally and via simulation (e.g. Ref. [62]), as well as identification of potential effects that homoplasy might have on results, are key research directions that require further study.

New directions for AFLP

Innovative hypothesis testing using AFLP data

So, what does the future hold for the AFLP technique? Novel uses of AFLP data to test evolutionary hypotheses are continually being developed. In an exciting new development for evolutionary studies, the use of AFLPs is moving beyond standard inferences of relationships (e.g. population and phylogenetic) toward assessing, for example, the role of selection in shaping patterns of divergence in wild populations in animals [64,65] and plants [45,66,67]. For example, studies using a genomic scan approach found a small percentage of AFLP characters (<5%) had significantly high F_{ST} values (i.e. greater genetic differentiation than expected under neutrality) [64–66]. Importantly, these same 'F_{ST} outlier' characters were found in comparisons involving different morphological types across different locations. Because these high levels of differentiation can only be explained by directional selection, these loci (or other linked loci) were implicated in the process of adaptation, divergence and ultimately speciation [64–66,68] – including the first plausible discovery of sympatric speciation in plants [66]. Further study of these outlier loci, together with morphological traits such as via quantitative trait loci (QTL) mapping [65], might provide even more detailed information regarding their number, location and relative effect on divergence. AFLPs are a valuable tool for non-model organisms in particular because the large number of characters required for genomic scans and subsequent hypothesis testing can be generated quickly.

Extraction of codominant data

AFLP fragments of the same size from different individuals that show an obvious difference in intensity can be codominant markers; intensity differences are predicted to be positively correlated with allelic copy number [9,69]. In other words, for a single character, diploid homozygous individuals (*AA*) should have a more intense peak than, and be distinguishable from, heterozygous individuals (*Aa*), which only have one copy of the plus allele [9]. A pair of fragments that differ in size by ~1–5 base pairs (bp) can also be a codominant marker (i.e. a true '*amplified fragment length polymorphism*' caused by an indel or microsatellite; sometimes called an allelic band pair) [4,35]. Specialized algorithms and software packages are capable of finding such markers and scoring them codominantly (Table 1). A subset of all the markers produced in an AFLP profile might be codominant (generally 10–20%, but as high as 75% in one study [7]) [4,35]. Codominant AFLP scoring is currently limited to model and/or commercial crop organisms [35], for which much genetic information already exists that can be used for the accurate identification of these characters. Codominant scoring will be an incredibly powerful tool if it can be routinely performed in non-model organisms and polyploids.

Polyploidy and genome size

Relative to diploids, polyploids generally produce higher numbers of AFLP fragments [19,55,70] with highly complex polymorphisms involving multiple loci and alleles, which makes determination of allele dosages problematic [4,14]. Similarly, organisms with large genomes, which can contain large amounts of repetitive DNA and retrotransposons, tend to have more fragments in their AFLP profiles than those with smaller genomes, but also frequently give rise to profiles with many low-intensity peaks that are difficult to score [19,70] (Box 2). Polyploidy has played a major role in the evolutionary history of many plants and other organisms. Therefore, future research should focus on better understanding how polyploidy and genome size affect AFLP data generation and analysis, experimentally testing if codominant scoring of AFLP data is possible in polyploids, and creating new methodologies for analysis of polyploid AFLP data, particularly when individuals of different ploidy levels are (knowingly or unknowingly) included in the same study (e.g. Ref. [55]).

In silico *AFLP*

The simulation of AFLP data and its use is another area under development with great potential for planning an AFLP study as well as a data exploration tool. *In silico* AFLP aims to provide a means to plan the generation of AFLP data for a given organism and choose the best endonucleases and primers via 'virtual AFLP', reducing the need for expensive and time-consuming trials in the laboratory [62,71]. This is accomplished via a simulation of the process that generates AFLPs (Box 1) on already-sequenced genomes. The result is a 'virtual gel' of the expected fragments and their size, their location in the genome, and even their sequence.

In addition to their usefulness in designing an AFLP experiment, *in silico* AFLPs also show great promise for exploring AFLP data patterns, investigating phylogenetic signal, and developing more comprehensive data analysis and interpretation tools, such as the construction of high-resolution linkage maps [61,62]. AFLP simulation is currently limited to those organisms for which entire

Table 1. AFLP scoring software

Program[a]	Approx. cost (USD)	Score scanned images of gels?	Compatible electrophoresis systems	Compatible fluorophores[b]	Compatible size standard	Normalize profiles between lanes or capillaries?	Automatic binning?	Variable bin width?	Variable amplitude threshold?	Manual editing of profiles?	Codominant scoring?	Website
AFLPMiner 1.0 (BioinforSoft LLC)[c]	≥$1000	×	Any ABI slab gel or capillary systems	Any; up to 5	Any	✓	✓	✓	✓	✓	✓	http://www.bioinforsoft.com/
AFLP-Quantar® (Keygene)[d]	Less than $17 000	✓	Several slab gel systems; ABI, MegaBACE, SpectruMedix capillary systems	Any	Any	✓	✓	✓	✓	✓	×	http://www.keygene.com/
AFLP-Quantar® *Pro* (Keygene)[d]	Less than $29 000	✓	Several slab gel systems; ABI, MegaBACE, SpectruMedix capillary systems	Any	Any	✓	✓	✓	✓	✓	✓	http://www.keygene.com/
CEQ™ AFLP® Dominant Scoring Software (Beckman Coulter)	Part of package[e]	×	CEQ 8000, CEQ 8800, GenomeLab GeXP capillary systems (Beckman Coulter)	WellRed™ D1, D2, D3, D4 dyes (Beckman Coulter)	DNA Size Standard-400 and -600 (Beckman Coulter); WellRed™ dyes (Beckman Coulter)	×	✓	✓	✓	×	×	http://www.beckmancoulter.com/
Cross Checker 2.91 (Jaap Buntjer, Wageningen Uni., The Netherlands)	Free download	✓	N/A (scores digitized slab gel images only)	N/A	Any	×	×	×	✓	✓	×	http://www.dpw.wau.nl/pv/pub/CrossCheck/
GelQuest v1.0 (SequentiX)	$1550	✓	Common (incl. LI-COR) slab gel systems; ABI capillary systems	Any; up to 5	Any	✓	✓	✓	✓	✓	✓ (SA)	http://www.sequentix.de/
GeneMapper® v4.0 (Applied Biosystems)	$12 000	×	ABI 377 slab gel system, ABI 310, 3100, 3130, 3730 capillary systems	Any	ABI LIZ®- and ROX™-based size standards	✓	✓	✓	✓	✓[f]	×	http://www.appliedbiosystems.com/genemapper
GeneMarker® v1.51 (SoftGenetics)	$2600 (academic)	✓	LI-COR slab gel systems; all major capillary systems (e.g. ABI 310 to 3730, Beckman Coulter, SpectruMedix, MegaBACE)	Any	Any	✓	✓	✓	✓	✓	✓ (SA)	http://www.softgenetics.com/
Genographer v1.6 (James Benham and Tom Blake, Montana State Uni., USA)	Free download; open source	×	ABI 373, 377 slab gel systems, ABI 3100, 3700, Beckman Coulter CEQ 2000 capillary systems	Any	Any	✓	×	✓	✓	✓[f]	✓ (SA)	http://hordeum.oscs.montana.edu/genographer/
GenoProfiler 2.0 (Frank You *et al.*, UC Davis, USA)	Free download	×	ABI 3100, 3730 capillary systems	Any; up to 5	Any	✓	×	×	✓	×	×	http://wheat.pw.usda.gov/PhysicalMapping/tools/genoprofiler/

Table 1 (*Continued*)

Program[a]	Approx. cost (USD)	Score scanned images of gels?	Compatible electrophoresis systems	Compatible fluorophores[b]	Compatible size standard	Normalize profiles between lanes or capillaries?	Automatic binning?	Variable bin width?	Variable amplitude threshold?	Manual editing of profiles?	Codominant scoring?	Website
Saga^{MX} AFLP® Software (LI-COR Biosciences)	$6900	✓	LI-COR 4000, 4200, 4300 slab gel systems (LI-COR TIFF)	LI-COR IRDye 700 and IRDye 800 fluorophores	LI-COR size standards	×	SA	✓	×	✓	✓ (SA)	http://licor.com/SagaMX
STRand 2.3.79 (Shayne Hughes, UC Davis, USA)	Free download	✓	ABI 373, 377, MJ BaseStation slab gel systems; ABI 310, 3100, 3700, 3730 capillary systems	Any	Any	×	×	✓	✓	✓	✓	http://www.vgl.ucdavis.edu/STRand/

Abbreviations: SA, semi-automated.
[a]Programs differ in many other features besides those listed; users should fully research software options before committing to a program. All programs automatically create a binary data matrix. Information current as at December 2006.
[b]Fluorophores also need to be compatible with the electrophoresis system.
[c]Available from 2007.
[d]To be released from 2007 at a reduced price.
[e]Requires purchase of CEQ Genetic Analysis System or GenomeLab GeXP Genetic Analysis System.
[f]Profile changes cannot be saved.

genomes are available, and has only been performed in a handful of studies (e.g. Refs [58,61]). In *Arabidopsis*, for example, *in silico* AFLP has been used to quantify homoplasy in AFLP profiles [58] by identifying whether experimentally determined sequences of a sample of AFLP bands actually correspond to those predicted by *in silico* methods [61].

Useful modifications of the technique

Recent modifications of the AFLP technique have lead to the development of other useful methodologies (discussed in Ref. [2]). These include methods for the discovery and development of microsatellite and SNP markers, cDNA-AFLPs [72] and the use of methylation sensitive enzymes [73] for studying gene expression, and domain-directed profiling [74]. Of these, cDNA-AFLPs merit further discussion because they can be used to understand the role of gene expression in determining phenotype. Analysis of quantitative variation in gene expression using cDNA-AFLPs has recently been developed for QTL mapping in *Arabidopsis* [75]. For non-model organisms, cDNA-AFLP techniques, which require no prior sequence information, are a valuable substitute for the microarray tools now available for model organisms (e.g. Ref. [76]). In addition to these modifications, AFLP has also been combined with bacterial artificial chromosome (BAC) DNA pooling in a high-throughput screening method that can result in integrated genetic and physical genome maps (e.g. sorghum [77]) and positional cloning of genes of interest (e.g. disease resistance gene in potato [78]).

Will AFLPs be obsolete?

The genomic era has been heralded with massive advances in whole genome sequencing and, although obtaining the first genome for a species can still be challenging, new pyrosequencing techniques [79] have enabled re-sequencing, at least, to become increasingly routine and relatively cost effective. In addition, new large-scale genotyping techniques, such as diversity arrays technology (DArT [80]; see http://www.diversityarrays.com/), have the potential to replace and surpass AFLP and other PCR-based methods because of their ability to provide comprehensive data from both the genome and transcriptome. In this vein, there are still significant challenges for AFLP, including the scoring of profiles (Box 3), unrealized potential in software (Box 3), and a lack of data deposition databases and standards for methods reporting (Box 4) that, unless resolved soon, might result in a shift away from AFLPs in many studies.

Although some authors might suggest that these new platforms together with outstanding AFLP challenges will cause AFLP fingerprinting to become obsolete (e.g. Refs [26,32,68]), we predict multilocus fingerprinting techniques in general will remain useful for the immediate future. First, many biologists work on species that are low priorities for whole genome sequencing (e.g. non-commercial species) and, therefore, techniques that require no prior sequence information such as AFLP are particularly attractive. Second, DNA sequencing and assembly of complex, large, polyploid, and/or repetitive genomes is still technically difficult. Third, genome sequencing of hundreds or thousands of closely related individuals is not yet economically feasible. Fourth, the AFLP technique can be modified, or combined with new and existing technologies, in useful ways. For example, the Complexity Reduction of Polymorphic Sequences (CRoPS) technique enables the rapid development of SNP and microsatellite markers from pyrose-

Box 4. Methods reporting and data deposition

The reporting of methods for AFLP fingerprinting presents several challenges concerning what should be reported and what can be omitted. The generation of AFLPs is affected by many variables (e.g. various PCR parameters). Describing these in sufficient detail to enable independent replication – although a tenet of scientific practice – would require much more space than journals allow. In general, published AFLP studies do not contain sufficient information to reproduce the exact fingerprints. This might not be problematic because research groups tend to work on a narrow range of taxa and do not necessarily need to replicate the work of others or add to existing datasets. The patterns of relatedness between accessions can usually be reproduced with a small amount of information (i.e. voucher information, restriction enzymes and selective primers used). Description of these three variables, along with labelling and detection methods, scoring parameters, and the number and nature of replicates, should comprise the minimum standards for methods reporting. The results should include, as a minimum, the total number of loci (characters) analysed, percent polymorphism, and some estimation of error and reproducibility. There are many examples of good methods and results reporting (e.g. Refs [17,20,45,51,65]).

Additional publication space should be dedicated to more detailed descriptions of technical advances at the forefront of AFLP data generation (i.e. areas that are being actively researched, such as data scoring and appropriate analysis techniques (e.g. Ref. [81]). Even these apparently obscure technical advances can now be quickly and easily communicated to the scientific community via full-text Web searches of academic literature (e.g. Google Scholar). Improved communication of technical aspects of AFLP fingerprinting could also be achieved with the deposition of data in a centralized, publicly accessible database [17,22]. This would bring together data on taxa studied, good and bad primer combinations, levels of diversity and scoring parameters, benefiting researchers planning an AFLP project. Deposition of binary datasets (and raw data too) would increase accountability to the scientific community, encourage new advances in the generation of AFLP data, and stimulate research on appropriate scoring and analysis methods, including improvements in software. As it is for nearly all DNA sequencing studies, AFLP data deposition should be a requirement for publishing in scientific journals. Similar issues concerning the reporting of methods and deposition of data are being addressed in other areas of biology (see Ref. [88] for review).

quenced AFLP fragments (see http://www.intl-pag.org/14/abstracts/PAG14_W410.html). Given sufficient time, the technical and financial hurdles of whole genome sequencing and microarray technologies will be overcome but, in the meantime, the AFLP technique offers a relatively fast and inexpensive method for genotyping a large number of individuals with a high degree of resolution and without prior genetic information. Far from being 'almost forgotten', AFLP is a highly useful and adaptable technique and, if fostered by the parallel development of new analysis methods, tests and simulations, will continue to be at the forefront in answering important scientific questions in a variety of disciplines for both model and non-model organisms.

Acknowledgements

For performing a literature survey of recent AFLP papers we thank Morore Piripi. For helpful comments on the manuscript we thank Dirk Albach, Jeanne Jacobs, Pete Lockhart, Angela McGaughran, Trish McLenachan, Christian Michel, Mary Morgan-Richards, David Penny, Leon Perrie, Morore Piripi, Kai Stölting, Vaughan Symonds, Steve Trewick, Andrea Weeks, Vincent Woo, and three anonymous referees. For financial support H.M.M. acknowledges the National Science Foundation International Research Fellowship Program, and A.C.C. acknowledges Massey University and the Allan Wilson Centre.

References

1 Vos, P. *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414
2 Bensch, S. and Åkesson, M. (2005) Ten years of AFLP in ecology and evolution: why so few animals? *Mol. Ecol.* 14, 2899–2914
3 Mueller, U.G. and Wolfenbarger, L.L. (1999) AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* 14, 389–394
4 Alonso-Blanco, C. *et al.* (1998) Development of an AFLP based linkage map of L*er*, Col and Cvi *Arabidopsis thaliana* ecotypes and construction of a L*er*/Cvi recombinant inbred line population. *Plant J.* 14, 259–271
5 Saliba-Colombani, V. *et al.* (2000) Efficiency of RFLP, RAPD, and AFLP markers for the construction of an intraspecific map of the tomato genome. *Genome* 43, 29–40
6 Shirasawa, K. *et al.* (2004) Conversion of AFLP markers to sequence-specific markers for closely related lines in rice by use of the rice genome sequence. *Mol. Breed.* 14, 283–292
7 Wong, A. *et al.* (2001) Characterization of AFLP markers in damselflies: prevalence of codominant markers and implications for population genetic applications. *Genome* 44, 677–684
8 Rouppe van der Voort, J.N. *et al.* (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. *Mol. Gen. Genet.* 255, 438–447
9 van Eck, H.J. *et al.* (1995) The inheritance and chromosomal localization of AFLP markers in a non-inbred potato offspring. *Mol. Breed.* 1, 397–410
10 Gerber, S. *et al.* (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Mol. Ecol.* 9, 1037–1048
11 Mariette, S. *et al.* (2002) Sampling within the genome for measuring within-population diversity: trade-offs between markers. *Mol. Ecol.* 11, 1145–1156
12 Nybom, H. (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* 13, 1143–1155
13 Goldman, D.H. *et al.* (2004) Molecular and cytological examination of *Calopogon* (Orchidaceae, Epidendroideae): circumscription, phylogeny, polyploidy, and possible hybrid speciation. *Am. J. Bot.* 91, 707–723
14 McGregor, C.E. *et al.* (2000) A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (*Solanum tuberosum* L.) germplasm. *Euphytica* 113, 135–144
15 Barluenga, M. *et al.* (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439, 719–723
16 Belaj, A. *et al.* (2003) Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. *Theor. Appl. Genet.* 107, 736–744
17 Woodhead, M. *et al.* (2005) Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Mol. Ecol.* 14, 1681–1695
18 Després, L. *et al.* (2003) Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. *Mol. Phylogenet. Evol.* 27, 185–196
19 Kardolus, J.P. *et al.* (1998) The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). *Plant Syst. Evol.* 210, 87–103
20 Perrie, L.R. *et al.* (2003) Evidence for an allopolyploid complex in New Zealand *Polystichum* (Dryopteridaceae). *N.Z. J. Bot.* 41, 189–215
21 Campbell, D. *et al.* (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Mol. Ecol.* 12, 1979–1991
22 Bussell, J.D. *et al.* (2005) Arbitrarily amplified DNA markers as characters for phylogenetic inference. *Perspect. Plant Ecol. Evol. Syst.* 7, 3–26
23 Nicod, J-C. and Largiadèr, C.R. (2003) SNPs by AFLP (SBA): a rapid SNP isolation strategy for non-model organisms. *Nucleic Acids Res.* 31, e19

24 McLenachan, P.A. *et al.* (2000) Markers derived from amplified fragment length polymorphism gels for plant ecology and evolution studies. *Mol. Ecol.* 9, 1899–1903
25 Brugmans, B. *et al.* (2003) A new and versatile method for the successful conversion of AFLP markers into simple single locus markers. *Nucleic Acids Res.* 31, e55
26 Schlötterer, C. (2004) The evolution of molecular markers – just a matter of fashion? *Nat. Rev. Genet.* 5, 63–69
27 Archak, S. *et al.* (2003) Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. *Genome* 46, 362–369
28 Savelkoul, P.H.M. *et al.* (1999) Amplified-fragment length polymorphism analysis: the state of an art. *J. Clin. Microbiol.* 37, 3083–3091
29 Jones, C.J. *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Mol. Breed.* 3, 381–390
30 Blears, M.J. *et al.* (1998) Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *J. Ind. Microbiol. Biotechnol.* 21, 99–114
31 Squirrell, J. *et al.* (2003) How much effort is required to isolate nuclear microsatellites from plants? *Mol. Ecol.* 12, 1339–1348
32 Sunnucks, P. (2000) Efficient genetic markers for population biology. *Trends Ecol. Evol.* 15, 199–203
33 Kremer, A. *et al.* (2005) Monitoring genetic diversity in tropical trees with multilocus dominant markers. *Heredity* 95, 274–280
34 Lander, E.S. *et al.* (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174–181
35 Assunção, A.G.L. *et al.* (2006) Construction of a genetic linkage map of *Thlaspi caerulescens* and quantitative trait loci analysis of zinc accumulation. *New Phytol.* 170, 21–32
36 Hollingsworth, P.M. and Ennos, R.A. (2004) Neighbour joining trees, dominant markers and population genetic structure. *Heredity* 92, 490–498
37 Robinson, J.P. and Harris, S.A. (1999) Amplified fragment length polymorphisms and microsatellites: a phylogenetic perspective. In *Which DNA Marker for Which Purpose?* (Gillet, E.M., ed.), pp. 1–27, Institut für Forstgenetik und Forstpflanzenzüchtung, Universität Göttingen (http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm)
38 Krauss, S.L. (2000) Accurate gene diversity estimates from amplified fragment length polymorphism (AFLP) markers. *Mol. Ecol.* 9, 1241–1245
39 Zhivotovsky, L.A. (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Mol. Ecol.* 8, 907–913
40 Lynch, M. and Milligan, B.G. (1994) Analysis of population genetic structure with RAPD markers. *Mol. Ecol.* 3, 91–99
41 Hardy, O.J. (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Mol. Ecol.* 12, 1577–1588
42 Holsinger, K.E. and Wallace, L.E. (2004) Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophaea* (Orchidaceae). *Mol. Ecol.* 13, 887–894
43 Koopman, W.J.M. (2005) Phylogenetic signal in AFLP data sets. *Syst. Biol.* 54, 197–217
44 Mendelson, T.C. and Shaw, K.L. (2005) Use of AFLP markers in surveys of arthropod diversity. *Methods Enzymol.* 395, 161–177
45 Ellis, A.G. *et al.* (2006) Evolutionary radiation of "stone plants" in the genus *Argyroderma* (Aizoaceae): unraveling the effects of landscape, habitat, and flowering time. *Evolution Int. J. Org. Evolution* 60, 39–55
46 Pelser, P.B. *et al.* (2003) Phylogeny reconstruction in the gap between too little and too much divergence: the closest relatives of *Senecio jacobaea* (Asteraceae) according to DNA sequences and AFLPs. *Mol. Phylogenet. Evol.* 29, 613–628
47 Irwin, D.E. *et al.* (2005) Speciation by distance in a ring species. *Science* 307, 414–416
48 Spooner, D.M. *et al.* (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14694–14699
49 Kilian, B. *et al.* (2007) Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. *Mol. Biol. Evol.* 24, 217–227
50 Brouat, C. *et al.* (2004) Differentiation in a geographical mosaic of plants coevolving with ants: phylogeny of the *Leonardoxa africana* complex (Fabaceae: Caesalpinioideae) using amplified fragment length polymorphism markers. *Mol. Ecol.* 13, 1157–1171
51 Schönswetter, P. *et al.* (2004) Glacial history of high alpine *Ranunculus glacialis* (Ranunculaceae) in the European Alps in a comparative phylogeographical context. *Biol. J. Linn. Soc.* 81, 183–195
52 El-Rabey, H.A. *et al.* (2002) Speciation and species separation in *Hordeum* L. (Poaceae) resolved by discontinuous molecular markers. *Plant Biol.* 4, 567–575
53 Kosman, E. and Leonard, K.J. (2005) Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol. Ecol.* 14, 415–424
54 Simmons, M.P. *et al.* (2007) A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. *Mol. Phylogenet. Evol.* 42, 528–542
55 Guo, Y-P. *et al.* (2006) Hybrid origin and differentiation of two tetraploid *Achillea* species in East Asia: molecular, morphological and ecogeographical evidence. *Mol. Ecol.* 15, 133–144
56 Benjak, A. *et al.* (2006) Different DNA extraction methods can cause different AFLP profiles in grapevine (*Vitis vinifera* L.). *Vitis* 45, 15–21
57 Mechanda, S.M. *et al.* (2004) Sequence assessment of comigrating AFLP™ bands in *Echinacea* – implications for comparative biological studies. *Genome* 47, 15–25
58 Koopman, W.J.M. and Gort, G. (2004) Significance tests and weighted values for AFLP similarities, based on *Arabidopsis in silico* AFLP fragment length distributions. *Genetics* 167, 1915–1928
59 Hansen, M. *et al.* (1999) Evaluation of AFLP in *Beta*. *Theor. Appl. Genet.* 98, 845–852
60 Vekemans, X. *et al.* (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Mol. Ecol.* 11, 139–151
61 Peters, J.L. *et al.* (2001) A physical amplified fragment-length polymorphism map of *Arabidopsis*. *Plant Physiol.* 127, 1579–1589
62 Rombauts, S. *et al.* (2003) AFLPinSilico, simulating AFLP fingerprints. *Bioinformatics* 19, 776–777
63 O'Hanlon, P.C. and Peakall, R. (2000) A simple method for the detection of size homoplasy among amplified fragment length polymorphism fragments. *Mol. Ecol.* 9, 815–816
64 Wilding, C.S. *et al.* (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* 14, 611–619
65 Bonin, A. *et al.* (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol. Biol. Evol.* 23, 773–783
66 Savolainen, V. *et al.* (2006) Sympatric speciation in palms on an oceanic island. *Nature* 441, 210–213
67 Skøt, L. *et al.* (2002) Molecular genecology of temperature response in *Lolium perenne*: 2. Association of AFLP markers with ecogeography. *Mol. Ecol.* 11, 1865–1876
68 Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4, 981–994
69 Piepho, H-P. and Koch, G. (2000) Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genetics* 155, 1459–1468
70 Fay, M.F. *et al.* (2005) The effects of nuclear DNA content (C-value) on the quality and utility of AFLP fingerprints. *Ann. Bot. (Lond.)* 95, 237–246
71 Qin, L. *et al.* (2001) GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP. *Nucleic Acids Res.* 29, 1616–1622
72 Bachem, C.W.B. *et al.* (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.* 9, 745–753
73 Cervera, M-T. *et al.* (2002) Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. *Mol. Genet. Genomics* 268, 543–552
74 van der Linden, C.G. *et al.* (2004) Efficient targeting of plant disease resistance loci using NBS profiling. *Theor. Appl. Genet.* 109, 384–393
75 Vuylsteke, M. *et al.* (2006) Genetic dissection of transcriptional regulation by cDNA-AFLP. *Plant J.* 45, 439–446
76 Borevitz, J.O. *et al.* (2003) Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* 13, 513–523

77 Klein, P.E. *et al.* (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Res.* 10, 789–807
78 Brugmans, B. *et al.* (2006) Exploitation of a marker dense linkage map of potato for positional cloning of a wart disease resistance gene. *Theor. Appl. Genet.* 112, 269–277
79 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380
80 Wenzl, P. *et al.* (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9915–9920
81 Bonin, A. *et al.* (2005) Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. *Methods Enzymol.* 395, 145–161
82 Papa, R. *et al.* (2005) An improved protocol for the production of AFLP™ markers in complex genomes by means of capillary electrophoresis. *J. Anim. Breed. Genet.* 122, 62–68
83 Gadkar, V. and Rillig, M.C. (2005) Suitability of genomic DNA synthesized by strand displacement amplification (SDA) for AFLP analysis: genotyping single spores of arbuscular mycorrhizal (AM) fungi. *J. Microbiol. Methods* 63, 157–164
84 Medina, R.F. *et al.* (2006) Number of individuals and molecular markers to use in genetic differentiation studies. *Mol. Ecol. Notes* 6, 1010–1013
85 Bonin, A. *et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Mol. Ecol.* 13, 3261–3273
86 van Os, H. *et al.* (2005) SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. *Theor. Appl. Genet.* 112, 187–194
87 Pompanon, F. *et al.* (2005) Genotyping errors: causes, consequences and solutions. *Nat. Rev. Genet.* 6, 847–859
88 Brazma, A. *et al.* (2006) Standards for systems biology. *Nat. Rev. Genet.* 7, 593–605