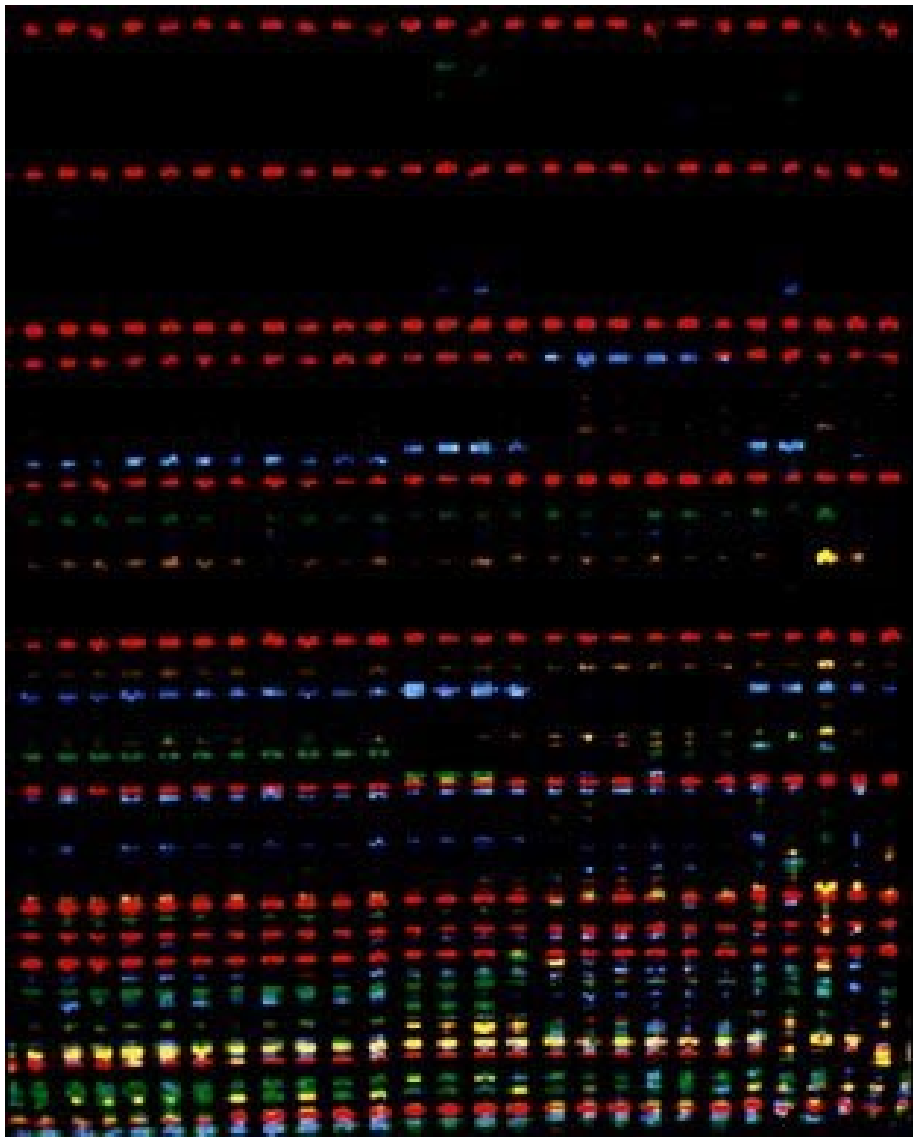


Molecular tools in plant genetic resources conservation: a guide to the technologies

by A. Karp, S. Kresovich, K.V. Bhat, W.G. Ayad and T. Hodgkin



The International Plant Genetic Resources Institute (IPGRI) is an autonomous international scientific organization operating under the aegis of the Consultative Group on International Agricultural Research (CGIAR).

The international status of IPGRI is conferred under an Establishment Agreement which, by January 1997, had been signed by the Governments of Australia, Belgium, Benin, Bolivia, Brazil, Burkina Faso, Cameroon, Chile, China, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Greece, Guinea, Hungary, India, Indonesia, Iran, Israel, Italy, Jordan, Kenya, Malaysia, Mauritania, Morocco, Pakistan, Panama, Peru, Poland, Portugal, Romania, Russia, Senegal, Slovak Republic, Sudan, Switzerland, Syria, Tunisia, Turkey, Uganda and Ukraine.

IPGRI's mandate is to advance the conservation and use of plant genetic resources for the benefit of present and future generations. IPGRI works in partnership with other organizations, undertaking research, training and the provision of scientific and technical advice and information, and has a particularly strong programme link with the Food and Agriculture Organization of the United Nations.

Financial support for the research agenda of IPGRI is provided by the Governments of Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, India, Italy, Japan, the Republic of Korea, Luxembourg, Mexico, the Netherlands, Norway, the Philippines, Spain, Sweden, Switzerland, the UK and the USA, and by the Asian Development Bank, CTA, European Union, IDRC, IFAD, Interamerican Development Bank, UNDP and the World Bank.

The geographical designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of IPGRI or the CGIAR concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries. Similarly, the views expressed are those of the authors and do not necessarily reflect the views of these participating organizations.

Citation: Karp, A., S. Kresovich, K.V. Bhat, W.G. Ayad and T. Hodgkin. 1997. Molecular tools in plant genetic resources conservation: a guide to the technologies. IPGRI Technical Bulletin No. 2. International Plant Genetic Resources Institute, Rome, Italy.

Cover: Gel image highlighting fluorescent AFLP typing. The gel includes 30 individuals (lanes) of cultivated and wild rice screened with three primer combinations (Mse-CAG/Eco-ATG, - AAG, and - ATT) (courtesy of A. Casa, M. Ferreira, S. Mitchell, and S. Kresovich).

ISBN 92-9043-323-X

IPGRI
Via delle Sette Chiese 142
00145 Rome
Italy

© International Plant Genetic Resources Institute 1997

Introduction to the Series

The concept of the Technical Bulletin series was developed by Dr J.M.M. Engels and Ms J. Toll of the then Germplasm Maintenance and Use Group of IPGRI in 1996. The Series as a whole is targeted at scientists and technicians managing genetic resources collections. Each title will aim to provide guidance on choices while implementing conservation techniques and procedures and in the experimentation required to adapt these to local operating conditions and target species. Techniques are discussed and, where relevant, options presented and suggestions made for experiments. The Technical Bulletins are authored by scientists working in the genetic resources area and IPGRI would appreciate receiving suggestions of topics for future volumes. In addition, IPGRI would encourage, and is prepared to support, the exchange of research findings obtained at the various genebanks and laboratories.

Masa Iwanaga

Chair, IPGRI Publications Committee and
Deputy Director General (Programmes)



Contents

Acknowledgements 6

Introduction to this volume 7

1. Introduction 9

2. The need for improved genetic resolution 11

3. A brief review of the basic techniques 13

3.1. Basic tools 13

3.2. Category 1: non-PCR based methods 17

3.2.1. Restriction fragment length polymorphism (RFLP) analysis 17

3.2.2. Variable number of tandem repeats (VNTRs) 20

3.3. Category 2: arbitrary (or semi-arbitrary) primed techniques 21

3.3.1. Multiple arbitrarily primed PCR (MAAP) techniques 21

3.3.2. Amplified fragment length polymorphism (AFLP) 23

3.4. Category 3: site-targeted PCR 25

3.4.1. Sequence-tagged microsatellites (STMS) 27

3.5. Variations or combinations of the basic techniques 28

4. Using the different techniques 29

4.1. The basic technologies 29

4.2. Analysis of molecular data 30

4.3. Application of the techniques 32

4.3.1. Category 1 techniques: Non-PCR-based techniques 32

4.3.2. Category 2: arbitrary (or semi-arbitrary) primed/or multi-locus profiling techniques 33

4.3.3. Category 3: site-targeted PCR sequence techniques 34

4.4. Pre-screening and combinations of techniques 35

5. A framework for selecting appropriate techniques 36

6. Using the decision-making chart 41

References 43

Acronyms 46

Acknowledgements

The authors of this volume wish to acknowledge the valuable support and comments of the participants in the IPGRI Workshop on molecular genetic techniques for plant genetic resources held in October 1995. The support of the Institute of Arable Crops Research, UK (IACR); U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS), the U.S. Agency for International Development (US-AID); and the National Bureau for Plant Genetic Resources, India, is also gratefully acknowledged. IACR, UK receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom. Angela Karp also thanks members of the EU DGXII Biotechnology 'Molecular Screening Tools' project (contract numbers BIO2-920476; BIO2-920486; BIO2-930373; BIO2-930295) for helpful discussions. The assistance of Asha Mulchan (GRST Group, IPGRI) is also gratefully acknowledged.

Introduction to this volume

The dramatic advances in molecular genetics over the last few years have provided workers involved in the conservation of plant genetic resources with a range of new techniques for their work. For the first time techniques are available to analyse variation in plants and animals at the DNA level. Differences in gene sequence can be directly observed and described, a degree of precision previously impossible to achieve. Many of the techniques that have been developed have already been used to study the extent and distribution of variation in species gene pools and to investigate evolutionary and taxonomic questions. They have also shown their value in studies of accession identity and for the detection of novel useful variation.

So far, much of the work on the development and use of molecular techniques has been carried out in developed countries. There are now a number of laboratories in developing countries that have begun to carry out their own programmes but the bulk of the expertise, facilities and capacity remains in the developed world. There is a great need to expand the facilities available in developing countries where much of the genetic diversity that can be examined using molecular techniques is to be found.

The aim of the International Plant Genetic Resources Institute (IPGRI) is to strengthen the conservation and use of plant genetic resources worldwide with special emphasis on the needs of developing countries. Working in partnership with national programmes, research institutions and other organizations, it undertakes research and training and seeks to provide technical advice and information. In October 1995, IPGRI organized a small workshop on the use of molecular techniques in the conservation of plant genetic resources¹. One area of discussion was the considerable range of different molecular techniques available and the ways in which they could best be used. Deciding on which technique would be most appropriate for particular investigations is not always straightforward and depends on a range of different factors including the nature of the problem, the biology of the species and the resources available. The participants at the Workshop recommended that

¹ Molecular Genetic Techniques for Plant Genetic Resources: Proceedings of an IPGRI Workshop, October, 1995 (W.G. Ayad, T. Hodgkin, A. Jaradat and V.R. Rao, eds.). Rome, Italy, 1997.

IPGRI should prepare a publication which provided users with some guidance on the different molecular genetic techniques currently available and their use in addressing some of the key questions faced by plant genetic resources conservation workers. A number of participants have therefore collaborated with IPGRI staff to prepare this publication.

The publication attempts to provide a brief overview of currently available techniques and to outline some of their strengths and limitations. It also provides a framework to assist users in identifying what technique(s) might be most appropriate for their own needs. It is not intended as a laboratory manual of the techniques or as a substitute for the many excellent discussions of the strengths and weaknesses of individual methods that can be found in the literature. Rather, it is a broad survey of the main features of the different techniques and of the factors that conservation workers should bear in mind when initiating a molecular genetic based investigation.

Molecular genetics is a fast-moving field and new techniques are likely to be developed in the future which will have their own strengths and limitations. Users of this publication will want to be able to consider these as well and it is hoped that the principles identified in this publication are sufficiently general to enable them to do this. Although the power of molecular techniques provides tremendous new opportunities for conservation workers, they should not be seen as a substitute for other agromorphological or biochemical studies that provide users with the information they need on the resources conserved. Indeed one of the main objectives of current work should be to link these different components together to provide a more complete understanding of the diversity available and the ways in which it can best be used to enhance agricultural and forestry production and sustainable development.

1. INTRODUCTION

The conservation and use of plant genetic resources are essential to the continued maintenance and improvement of agricultural and forestry production and, thus, to sustainable development and poverty alleviation. Plant genetic resources for agriculture include the reproductive or vegetatively propagated material of (i) cultivars in current use and newly developed varieties, (ii) obsolete cultivars, (iii) farmers' traditional cultivars and landraces, (iv) wild and weedy relatives of cultivated species, and (v) special genetic stocks including elite and current breeders lines, aneuploids and mutants (Frankel *et al.* 1995). Forestry genetic resources are described as the heritable materials contained within and between tree species which have or may have an economic, scientific or social value for people (FAO 1993). The objective of plant genetic resources conservation is to preserve as broad a sample of the extant genetic diversity of target species as is scientifically and economically feasible, including currently recognized genes, traits and genotypes.

Effective conservation of plant genetic resources requires a complementary approach which makes use of both *ex situ* and *in situ* conservation methods to maximize the genetic diversity available for use. The objective of *ex situ* conservation is to maintain the accessions without change in their genetic constitution (see Frankel *et al.* 1995 for an up-to-date and comprehensive text on the conservation of plant genetic resources). The methods used are designed to minimize the possibility of mutation, selection, random genetic drift or contamination. For many crop species and their wild relatives, long-term *ex situ* conservation can be undertaken by storing seeds for long periods at low temperatures and humidities. However, a number of clonally propagated species, such as banana and potato, cannot be conserved in this way, and many species, particularly tropical forest tree species, produce seeds that are 'recalcitrant' and cannot be stored. These groups of species can only be maintained *ex situ* in field genebanks as growing collections of plants, or *in vitro* using tissue culture or cryopreservation (Withers 1992). Whether conserved as seed, *in vitro* or in the field, managers of *ex situ* collections need to maintain the integrity of the accessions conserved and to identify any duplicates. Regeneration will be needed and has to be carried out to ensure that genetic drift, or change in genetic structure of the population, is reduced to a minimum. Characterization to determine the identity of accessions will also be required.

Notes

In situ conservation is considered to be the method of choice for conserving forest species and wild crop relatives and there is increasing interest in the use of *in situ* conservation for crops themselves (on-farm conservation) (Brush 1995). *In situ* conservation allows evolution to continue, increases the amount of diversity that can be conserved, and strengthens links between conservation workers and the communities who have traditionally maintained and used the resources. The development of improved methodologies to support *in situ* conservation is urgent. At the very least, the populations conserved need to be monitored over time to determine how much genetic change is occurring and whether management practices should be modified.

All genetic resources conservation activities require characterization of the diversity present in both the gene pools and the genebanks. In the first instance, this usually involves description of variation for morphological traits, particularly agromorphological characteristics of direct interest to users. This approach has certain limitations: highly heritable traits often show little variation over much of the material studied and trait expression is subject to environmental variation and may be difficult to measure. The genetic information provided by morphological characters is also often limited. These limitations have resulted in the deployment of biochemical techniques such as isozyme and protein electrophoresis (Hunter and Markert 1957) and molecular techniques that analyse polymorphism at the DNA level directly. Characterization for morphological traits cannot, however, be replaced by any of the molecular techniques. The results of molecular or biochemical studies should be considered as complementary to morphological characterisation.

Molecular genetics has an important role to play in many aspects of conservation such as characterising plant genetic diversity for purposes of improved acquisition, maintenance and use. With the development of the polymerase chain reaction (PCR), in particular, numerous molecular technologies have been, and still are being, developed, which can be used for the detection, characterisation and evaluation of genetic diversity. These techniques vary in the way that they resolve genetic differences, in the type of data that they generate, in the taxonomic levels at which they can be most appropriately applied, and in their technical and financial requirements.

The application of molecular markers for the resolution of problems of genetic resources conservation is at an early state and requires extensive collaboration among conservation

workers and molecular biologists. This bulletin provides a key to the technologies which can be used and aims to help genebank managers and other conservation workers select the most appropriate techniques for their diversity work, noting any constraints of time, money or other relevant factors.

2. THE NEED FOR IMPROVED GENETIC RESOLUTION

Genebank managers and conservationists concerned with both *in situ* and *ex situ* management try to ensure that they conserve as much as possible of the extant genetic diversity of the species with which they work. The effectiveness with which they do this depends to a large extent on the genetic information available on the germplasm with which they work. Molecular markers provide genetic information of direct value in key areas of conservation both *ex situ* and *in situ*. For *ex situ* conservation the key issues are:

- **acquisition:** Data on the diversity of existing collections can be used to plan collection and exchange strategies. In particular, calculations of genetic distances based on molecular data can be used to identify particular divergent sub-populations that might harbour valuable genetic variation that is under-represented in current holdings.
 - **maintenance:** Genetic data are essential to identify duplicate accessions in order to ensure best use of available resources. Genetic markers are also needed to monitor changes in genetic structure as accessions are generated. Molecular markers provide markers suitable for both of these.
 - **characterization:** The genetic diversity within collections must be assessed in the context of the total available genetic diversity for each species. Existing passport data document the geographic location where each accession was acquired. However, passport records are often missing or incorrect. Molecular markers may extend and complement characterization based on morphological or biochemical descriptions, providing more accurate and detailed information than classical phenotypic data.
 - **distribution to users:** Users of collections benefit from genetic information that allows them to identify valuable traits and types quickly. On a more fundamental level, molecular marker information may lead to the further identification of useful genes contained in collections. Molecular data on diversity may provide essential information to develop core collections (Hodgkin *et al.* 1995) that accurately represent the entire collection.
-

Notes

Molecular markers may therefore be used in four types of measurements needed for effective *ex situ* conservation, all of which are useful in resolving the numerous operational, logistical, and biological questions that face genebanks managers (Kresovich *et al.* 1992). These are:

- **identity:** the determination of whether an accession or individual is catalogued correctly, is true to type, maintained properly, and whether genetic change or erosion has occurred in an accession or population over time;
- **similarity:** the degree of similarity among individuals in an accession or between accessions within a collection.
- **structure:** the partitioning of variation among individuals, accessions, populations, and species. Genetic structure is influenced by *in situ* demographic factors such as population size, reproductive biology and migration.
- **detection:** the presence of particular allele or nucleotide sequence in a taxon, genebank accession, *in situ* population, individual, chromosome or cloned DNA segment.

Those concerned with *in situ* conservation need to ensure that appropriate populations are identified and managed in such a way that they survive and continue to evolve. Their responsibilities can include:

- **location:** the identification of populations which should be conserved based on the genetic diversity present as well as on the value of the resource and the threats to it. Crucial to this is a knowledge of the extent and distribution of genetic diversity in species populations which should optimally include molecular data.
 - **management:** the development of management plans to monitor the changes in target populations over time and ensure their continued survival. The populations maintained *in situ* constitute part of ecosystems and both intra- and interspecific diversity must be maintained over time at appropriate levels.
 - **accessibility:** *in situ* conservation is most commonly of interest in forest genetic resources conservation and that of wild crop relatives but it is also of increasing interest for on-farm conservation of traditional cultivars. Genetic resources conserved in this way remain accessible to the communities who depend on them. Managers need to ensure they are also accessible to other users and that sufficient genetic information is available to assist such users.
-

Within the context of *in situ* conservation, therefore, **identity**, **similarity**, **structure** and **detection** are also important and can be usefully investigated using molecular techniques.

3. A BRIEF REVIEW OF THE BASIC TECHNIQUES

Useful texts which describe the basic principles and procedures of molecular genetics include Weising *et al.* (1995) and Hoelzel and Green (1994) (see also bibliography). A general introduction to measuring genetic variation using molecular markers may also be found as Unit 10.1.4 of IPGRI's training support materials on the Internet (<http://www.cgiar.org/ipgri/training>). This section briefly introduces the techniques of most relevance in plant genetic resources work.

3.1. Basic tools

The detection of genetic variation at the DNA level is made possible through the use of cellular enzymes that act on the DNA molecule in different ways. Among the most significant discoveries in molecular genetics was the identification of restriction enzymes, or restriction endonucleases, that are able to cut DNA in both strands. Each restriction enzyme recognises a unique, specific sequence of, usually, 4-6 base pairs (bp) in length, termed a restriction site, where the enzyme cuts (or restricts) the DNA. In general, restriction sites will occur throughout the genome and, consequently, application of the enzyme to total genomic DNA (restriction of the DNA) results in its conversion into millions of fragments. The frequency of restriction sites will vary depending on both the restriction enzyme and on the genome. Restriction enzymes that cut at sites that are of common occurrence (frequent cutters) in a given genome will result in very large numbers of small fragments, whereas restriction with an enzyme that cuts sites which occur rarely (rare cutters) will result in fewer, larger fragments being formed. The DNA fragments generated from restriction by a specific enzyme will all share in common the same sequence at the ends (i.e. the restriction site, or part thereof, where the cut was made) but will be of different sequence composition between the ends. The different fragments can be separated according to their length (and hence molecular weight) by electrophoresis. In this process, the DNA is loaded into a well at the top end of a flat gel matrix slab composed of agarose, or polyacrylamide. The gel is placed into a special tank where it is immersed in a suitable buffer and an electric current is then passed through the matrix (Fig. 1). During electrophoresis, the smallest fragments

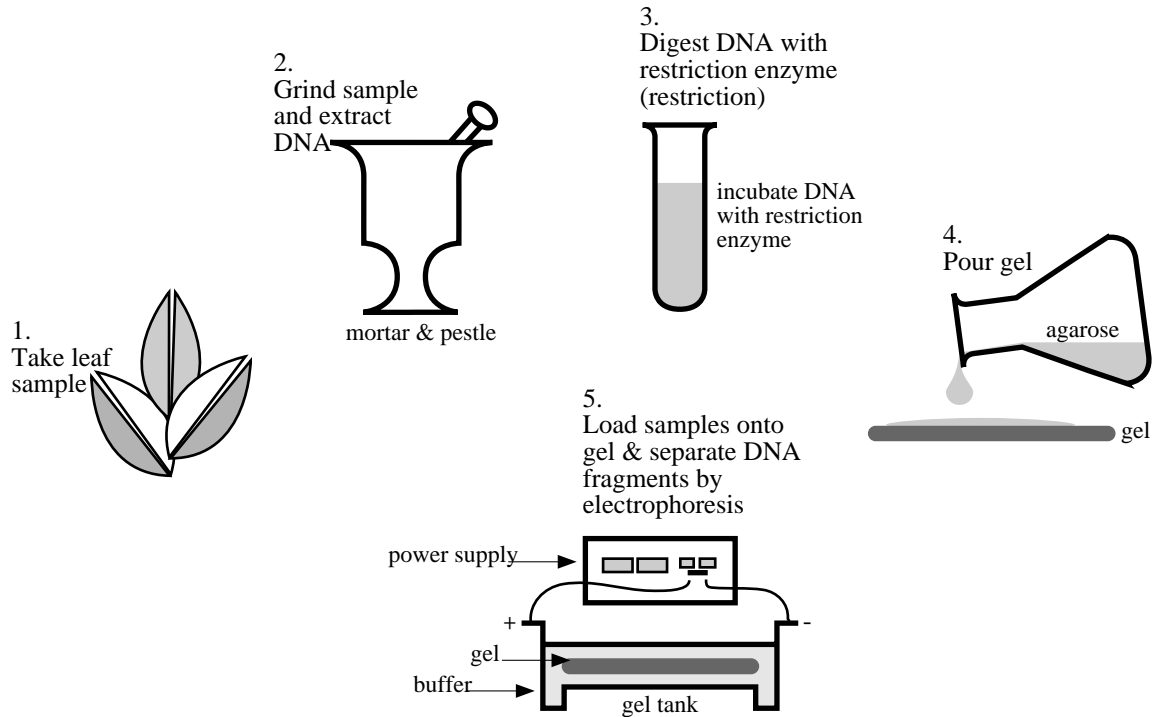


Fig. 1. Restriction and electrophoretic separation of DNA

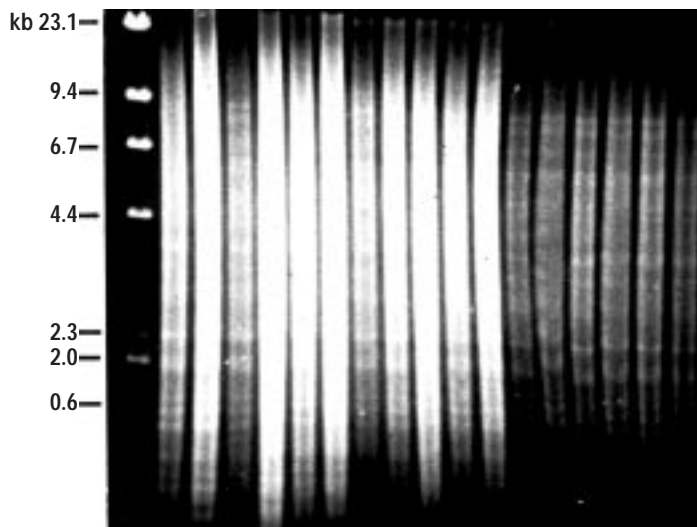


Fig. 2. Restricted digested DNA of barley following electrophoresis (source: A. Karp, IACR-LARS)

move fastest through the gel and will separate towards the lower part of the gel, whereas the larger fragments will move more slowly and will separate at the top end. Several different samples can be run in parallel on the gel, each sample resulting in a track composed of different fragment lengths. These tracks can be visualised by addition of a dye, such as ethidium bromide, to the gel matrix or the loading sample. The dye intercalates with the DNA and can be viewed under ultraviolet light (see Fig. 2).

There is so much DNA in the genome of higher plants and animals that the track of fragments visualised after

electrophoretic separation appears as a smear (Fig. 2). Specialised techniques are therefore required to detect variations in the DNA of two different individuals. A number of techniques can be used for the detection of variations (polymorphisms) in the DNA. Some of these are based upon the initial digestion of the DNA with restriction enzymes, while others depend upon the use of a different enzymatic reaction, known as the polymerase chain reaction (PCR).

The development of the polymerase chain reaction (PCR) was a technological breakthrough in genome analysis since it enabled the amplification of specific fragments from the total genomic DNA. The principle of PCR is very simple. It is based on the function of a copying enzyme, DNA polymerase, which is able to synthesise a duplicate molecule of DNA from a DNA template. The product of duplication of the original template DNA becomes a second template for another round of duplication. Repeated duplications thus lead to an exponential increase in DNA product accumulation (Fig. 3). Even when

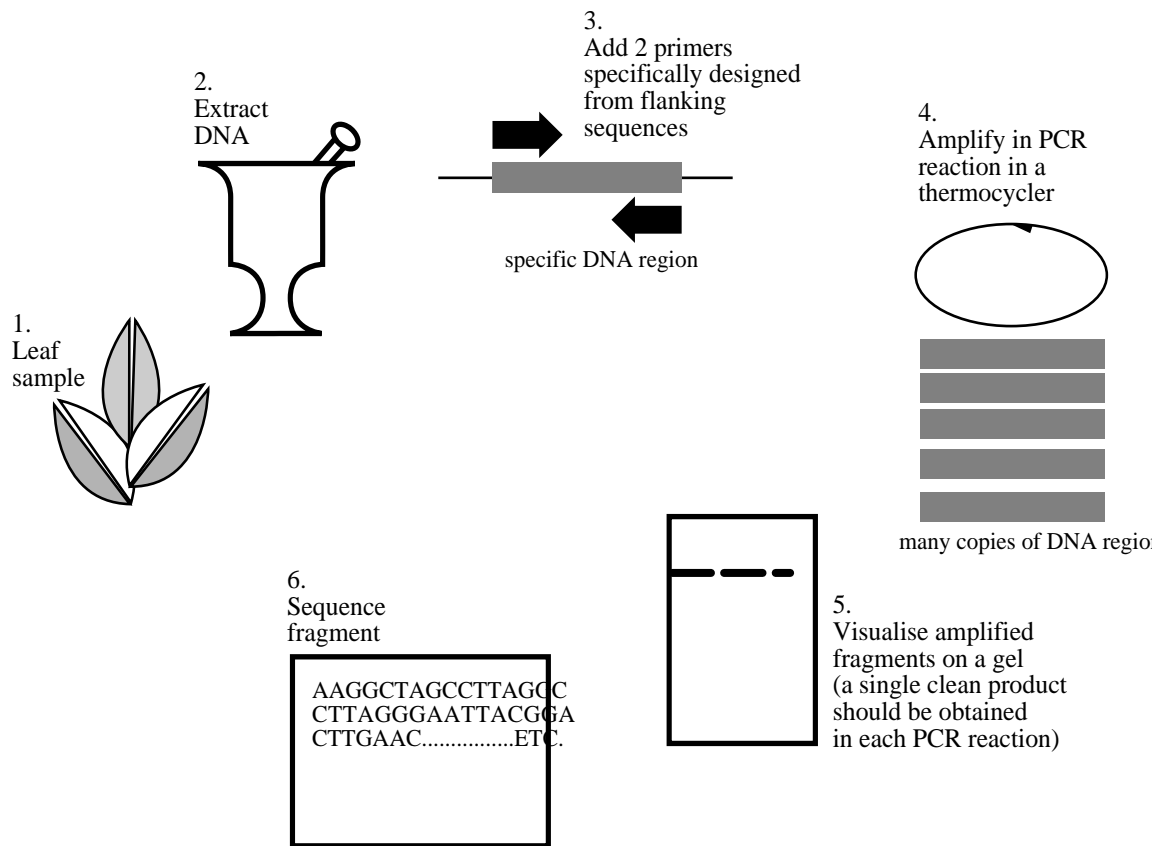


Fig. 3. The polymerase chain reaction (PCR)

Notes

starting from a single DNA molecule, detectable amounts of target DNA are generated by PCR in a few hours.

The basic concept of PCR was first tested with Klenow polymerase but the real breakthrough came when a thermostable DNA polymerase, *Taq* polymerase, was isolated and purified. This allowed the cycling process to be automated, as only a single addition of enzyme is required.

The DNA target is defined by primer annealing sites. Primers are short stretches of DNA sequence which are complementary to the opposite ends of the target sequence DNA. They anneal to the complementary sequences in the target and thus 'prime' the polymerase amplification. Since both strands of a DNA molecule run in antiparallel (i.e. opposite) orientation, the primer sequences point to each other. The usual distance between the priming sites (and hence the size of the amplified fragment) is between 100 bp and a few kilobases (kb), although the recent development of so-called 'long distance PCR' now allows amplification up to at least 40 kb.

Amplification from virtually any region of a DNA molecule is possible by selecting specific sequences as primers at both flanks of the target region. For such direct-targeted PCR the sequence of these flanks must be known. Whether or not a unique and specific product is obtained depends on the selectivity of the primers that are designed based upon the sequences flanking the target. For sequences that show a high degree of conservation among organisms, degenerate primers (in which the majority but not the exact sequence is complemented) may be sufficient for amplification of the target DNA. In other situations it may only be possible to amplify the unique desired product after very careful design of the primers based on the exact flanking sequences.

PCR reactions can also be carried out with single primers which have not been designed on a known target flanking sequence. In this case, amplification will occur wherever the primer is able to anneal to complementary sequence within the genome. Since the identity of the amplification products is not known, primers of this kind are referred to as arbitrary primers. They can be synthesized or bought from commercial suppliers. Alternatively, single primers based on a known target sequence such as part of a gene, or a microsatellite (see VNTR section below) may be used in the PCR reaction. Such primers are referred to as semi-arbitrary.

Although there are many other enzymes that are used by molecular biologists most techniques for screening diversity are based on the use of either restriction enzymes, PCR or both. In

fact, three basic categories can be identified based upon, firstly, whether the assays are PCR-based or not, and, then secondly, whether arbitrary/semi-arbitrary primers or specifically designed primers for known sequences are used: Category 1: non-PCR based methods; Category 2: arbitrary(or semi-arbitrary) primed techniques, and Category 3: site targeted PCR techniques. The main techniques that fall into these categories are described in greater detail below.

3.2. Category 1: non-PCR based methods

3.2.1. Restriction fragment length polymorphism (RFLP) analysis

In restriction fragment length polymorphism (RFLP) analysis, the DNA is digested with restriction enzymes and the resultant fragments are separated by gel electrophoresis. The restricted DNA fragments are then transferred to a filter by a process termed Southern Blotting (Fig. 4). In this process the gel slab is

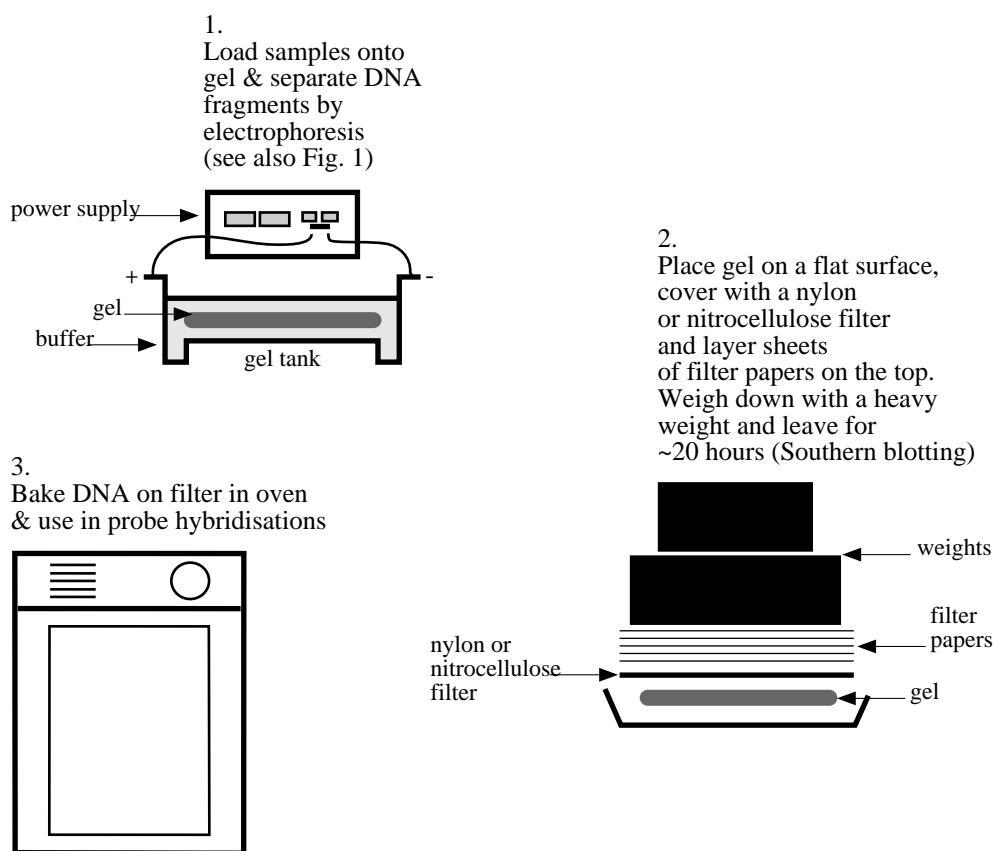


Fig. 4. Transfer from a gel to a filter (Southern blotting)

Notes

placed on a flat surface and a sheet of nitrocellulose or nylon is laid over the gel. On top of the nylon or nitrocellulose filter are then placed layers of filter papers and paper towels, which are weighted downwards by the use of a heavy weight. The blot is left for several hours until the DNA from the gel has transfused over to the nylon or nitrocellulose filter. The latter is then usually baked in an oven so that the DNA is immobilised on the filter.

The filter will now contain the tracks of DNA that were originally separated in the gel matrix, in exactly the same juxtapositions as they were present in the gel. To obtain some information on any difference that may be present in the fragments from different individuals it is necessary to hybridise a "probe" to the filter (Fig. 5). DNA is a double-stranded molecule

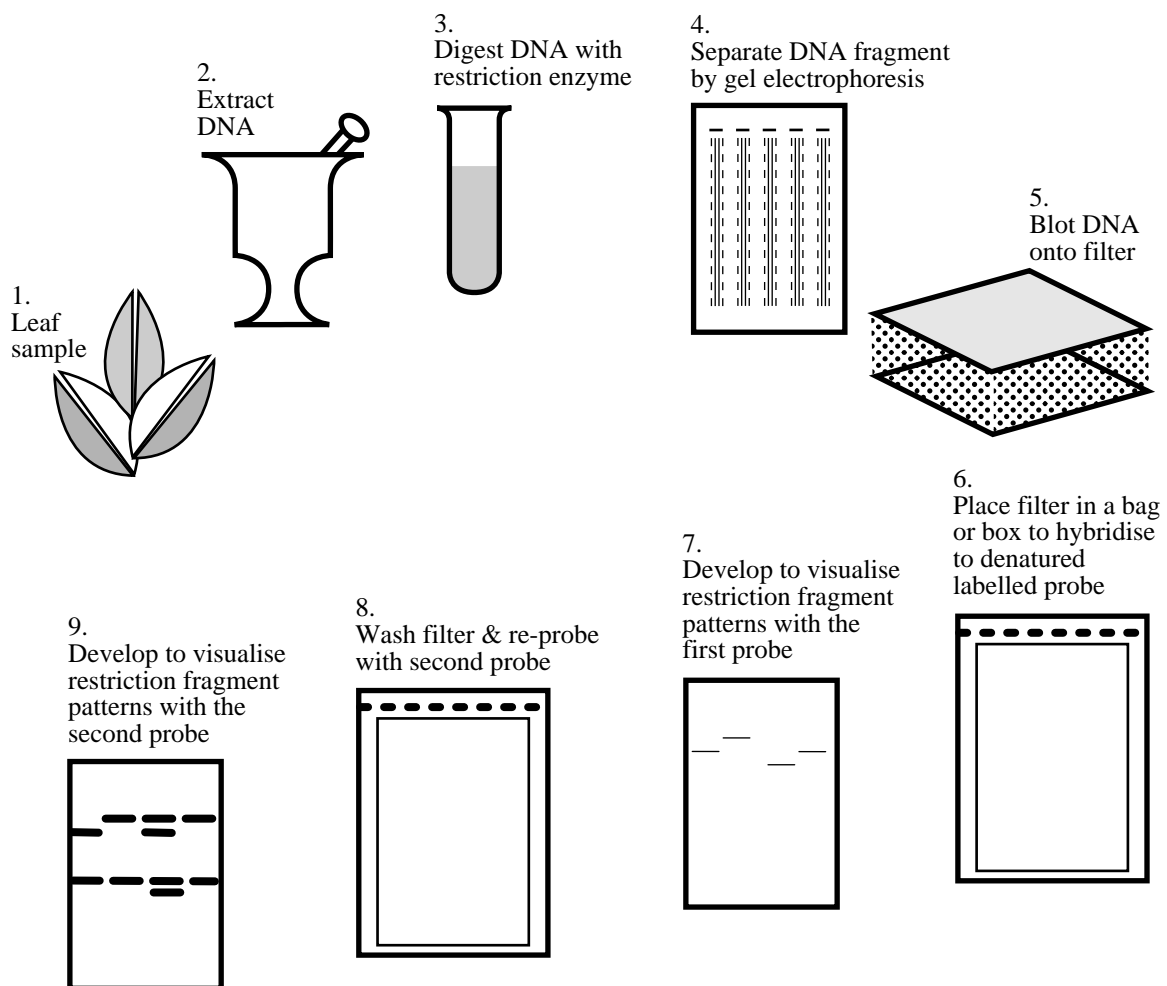


Fig. 5. Restriction fragment length polymorphism

in which the base-pair sequence of one strand is entirely complementary to the sequence of the other. The two strands can be separated (or the DNA denatured) by heat, or alkali hydrolysis which disrupts the hydrogen bonding between the strands. A probe is a short DNA fragment (typically ~800 bp in length). It could be a cloned expressed sequence, an unknown fragment of genomic DNA, or part of the sequence of a cloned gene. To carry out the hybridisation, the probe and filter are brought into contact by placing the filter into a plastic bag or plastic container to which a solution of the probe is added. In order to detect where the hybridisation has taken place it is necessary to make a copy of the probe using radiolabeled nucleotides or nucleotides that are labelled with non-radioactive labels such as digoxigenin. The principle of hybridisation relies on the complementary base pairing of the DNA. For probe-DNA hybridisation to work, the DNA on the filter is denatured and the labelled probe is denatured immediately before adding to the filter. The probe will hybridise to any fragment immobilised on the filter with which it shares sequence complementarity. Once it has hybridised it will remain bonded to the DNA provided that denaturing conditions are avoided. Once hybridisation has taken place (typically over several hours) the filter is washed to remove any excess probe and then, in the dark, an X-ray film is placed against the surface (in the case of radio-labelled probes), or the filter is submersed under a series of different chemicals (in the case of non-radioactive probes). Once the X-ray has been developed, or once the filter has been passed through the correct chemicals, the filter can be examined. Bands will appear only where the probe has hybridised to different fragments, i.e. in the event of sequence complementarity. The pattern obtained is referred to as the restriction fragment pattern (Figs. 5 and 6).

Usually, several (2-4) different restriction enzymes are used to cut the genomic DNA in separate experiments and different

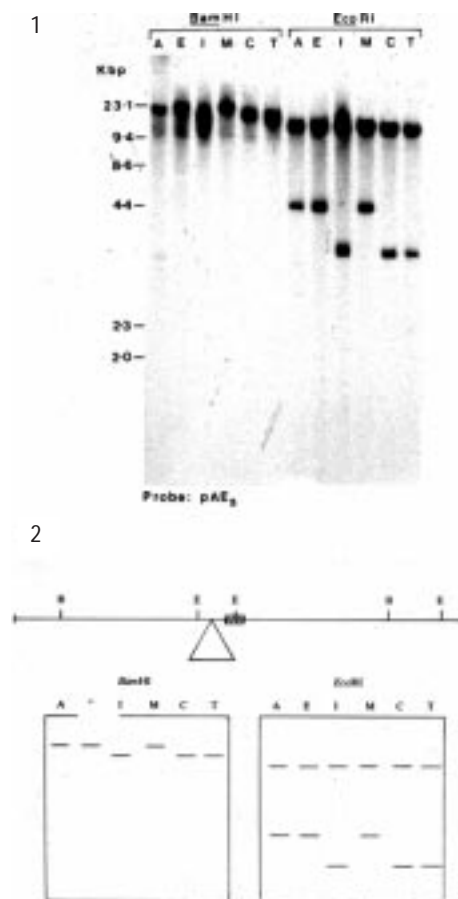


Fig. 6. RFLP analysis of barley cultivars. (1) The results of hybridisation of a single probe to DNA of 6 cultivars (A, E, I, M, C, T) after restriction with the enzymes *Bam*H1 (left) and *Eco*R1 (right). (2) The probed region showing the restriction sites and diagrammatic interpretations of the results. Only one fragment is visible in the tracks after *Bam*H1 digestion which varies in length depending upon the insertion of base pairs, here depicted by the triangle. An internal *Eco*R1 digestion. The smaller fragment varies in length depending on the insertion of the base pairs depicted by the triangle. (source: A. Karp, IACR-LARS)

filters prepared accordingly (Fig. 5). A single filter will then be used in a hybridisation with a single probe. After analysis of the results it is possible to release the probe from the filter and re-use the filter in another experiment with a different probe. In this way the same filter can be probed several times.

Specific probe/enzyme combinations will give highly reproducible restriction fragment patterns for a given individual, but variation between individuals can arise when mutations alter the sequence in the restriction sites (thus preventing the enzyme from cutting) or the DNA sequence in the fragment lengths between them (by the creation of new restriction sites, or the insertion or deletion of base pairs resulting in an alteration of the fragment length between the sites) (Burr *et al.* 1983; Evola *et al.* 1986; Helentjaris *et al.* 1985).

RFLPs are highly reproducible and the same probe/enzyme combination on the same samples will give exactly the same results even when carried out in different laboratories. They are also codominant markers in that the different allelic variant bands are visible in the heterozygotes, enabling all three genotypic classes to be distinguished. Provided suitable probes are available, the technique can be applied immediately. However, a good supply of probes that can reliably detect variation is required and finding probes that can detect polymorphisms at the cultivar or population level can be a problem in some species. If it is not possible to utilise probes from other related species (i.e. heterologous probes), new probes must be isolated from cDNA or genomic DNA libraries, which requires additional skill and investment of resources. RFLPs are time-consuming and they are not easy to automate. Once probe/enzyme combinations have been selected, throughput will depend on the number of gels that can be run each day in the laboratory in question. RFLPs require high quantities of good quality DNA (e.g. 10 µg per digestion) and where polyphenol or polysaccharide contamination reduces DNA yields, or where only very limited amounts of source material are available, this requirement alone may preclude their application.

3.2.2 Variable Number of Tandem Repeats (VNTRs)

Interspersed within the genomes of higher organisms are hypervariable regions which are comprised of tandemly repeated DNA sequences. There are two classes: 'microsatellites', or simple sequence repeats (SSRs), where the basic repeat unit is around 2-8 base pairs in length, and 'minisatellites' for longer repeat units of around 16-100 base pairs. Hybridisation to

restricted DNA with micro- or minisatellite probes gives multilocus patterns which can resolve variation at the levels of populations and individuals (Beyermann *et al.* 1992). The variation results from changes in the number of copies of the basic repeat and is often referred to as Variable Number of Tandem Repeats (VNTRs). VNTR loci are, in principle, codominant markers, but in RFLP analysis they often behave as dominant markers (Arens *et al.* 1995).

3.3. Category 2: arbitrary (or semi-arbitrary) primed techniques

3.3.1 Multiple arbitrarily primed PCR (MAAP) techniques

With the advent of PCR, techniques became available which overcome many of the limitations of probe-hybridisation-based methods RFLPs. Among these, a subset of closely related techniques was developed simultaneously which involves the use of a single 'arbitrary' primers in a PCR reaction, the result of which is usually the amplification of many discrete DNA products. Each product will be derived from a region of the genome that contains two short segments which share sequence similarity to the primer and which are on opposite strands and sufficiently close together for the amplification to work. Techniques of this kind have been collectively termed multiple arbitrary amplicon profiling (MAAP) (Caetano-Annollés 1994) (Fig. 7). The most commonly used version is RAPD analysis (Random Amplified Polymorphic DNA) in which the amplification products are separated on agarose gels in the presence of ethidium bromide and visualised under ultraviolet light (Williams *et al.* 1990). AP-PCR (Arbitrary primed PCR) (Welsh and McClelland 1990) and DAF (DNA Amplification Fingerprinting) (Caetano-Annollés *et al.* 1991) differ from RAPDs principally in primer length, the stringency conditions and the method of separation and detection of the fragments. Polymorphisms are detected as the presence or absence of bands and mainly result from sequence differences in the primer binding sites. The enormous attraction of these techniques is that there is no requirement for DNA probes or sequence information for primer design. The procedure involves no blotting or hybridising steps. The technique is quick, simple and efficient and requires only the purchase of a thermocycling machine and agarose gel apparatus. It requires small amounts of DNA (10 ng per reaction), sample throughput can be high and the procedure is automatable. It is absolutely critical, however, to maintain strictly consistent reaction conditions in order to achieve

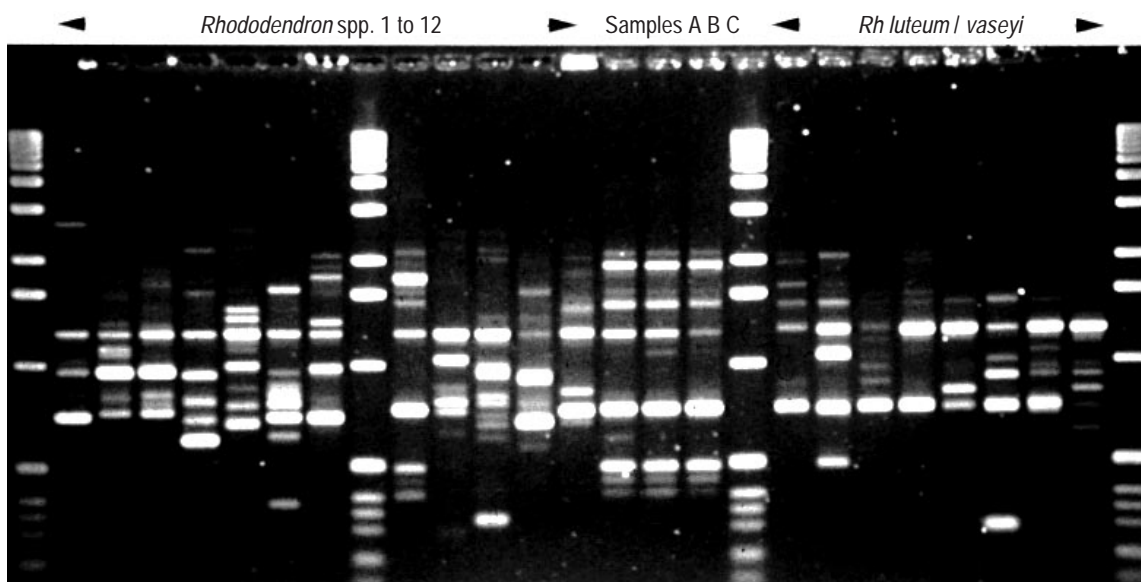
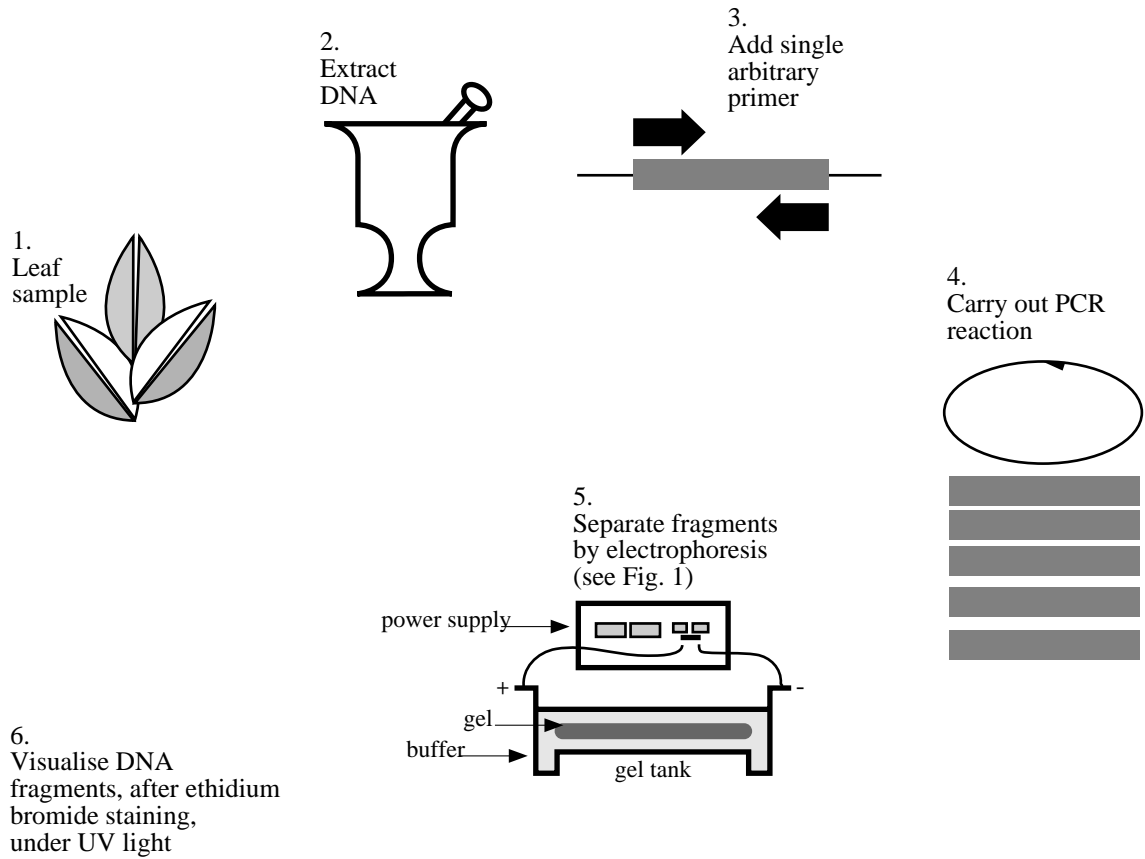


Fig. 7. Multiple arbitrary amplicon profiling (MAAP) showing analysis of different *Rhododendron* spp. using RAPDs (source: A. Karp, IACR-LARS)

reproducible profiles. In practice, band profiles can be difficult to reproduce between (and even within) laboratories, if personnel, equipment or conditions are changed. A further important limitation (discussed later) is that data quality is limited because MAAP gives dominant markers (heterozygosity is not discernible), bands may sometimes consist of comigrating products and band identities are difficult to assign.

3.3.2 Amplified Fragment Length Polymorphism (AFLP)

A more recently developed method, which is equally applicable to all species and is highly reproducible (Vos *et al.* 1995), is termed Amplified Fragment Length Polymorphism (AFLP). AFLP combines restriction digestion and PCR (Fig. 8). The first step involves restriction digestion of the genomic DNA with two specific enzymes, one a rare cutter and the other a frequent cutter. Adaptors are then added to the ends of the fragments to provide known sequences for PCR amplification. These adaptors are necessary because the restriction site sequence at the end of the fragments is insufficient for primer design. Short stretches of known sequence are added to the fragment ends through the use of a ligase (joining) enzyme. If PCR amplification of the restricted fragments was then carried out, all the fragments would be amplified which, under current technology, would not be resolvable on a single gel. Primers are thus designed so that they incorporate the known adaptor sequence plus 1, 2 or 3 additional base pairs, (any one out of the four possible: A,G,C or T). PCR amplification will only occur where the primers are able to anneal to fragments which have the adaptor sequence plus the complementary base pairs to the additional nucleotides. The additional base pairs are thus referred to as selective nucleotides. If one selective nucleotide is used, more fragments will be amplified than if two are used, and even fewer fragments will be amplified with three selective nucleotides. For technological reasons, addition of more than three selective nucleotides results in some non-specific PCR amplification. Normally two separate selective rounds of PCR are carried out. In the first round only one selective nucleotide is used, whereas in the second round the same selective nucleotide plus one or two additional ones are used. In practice this results in between 50-100 fragments being amplified, which can be separated on a polyacrylamide gel by electrophoresis.

The amplified products are normally visualised after exposure to X-ray film, where radiolabelled primers are used, but the technique has been adapted to fluorescent, non-radioactive and silver staining procedures, and has been automated. AFLPs

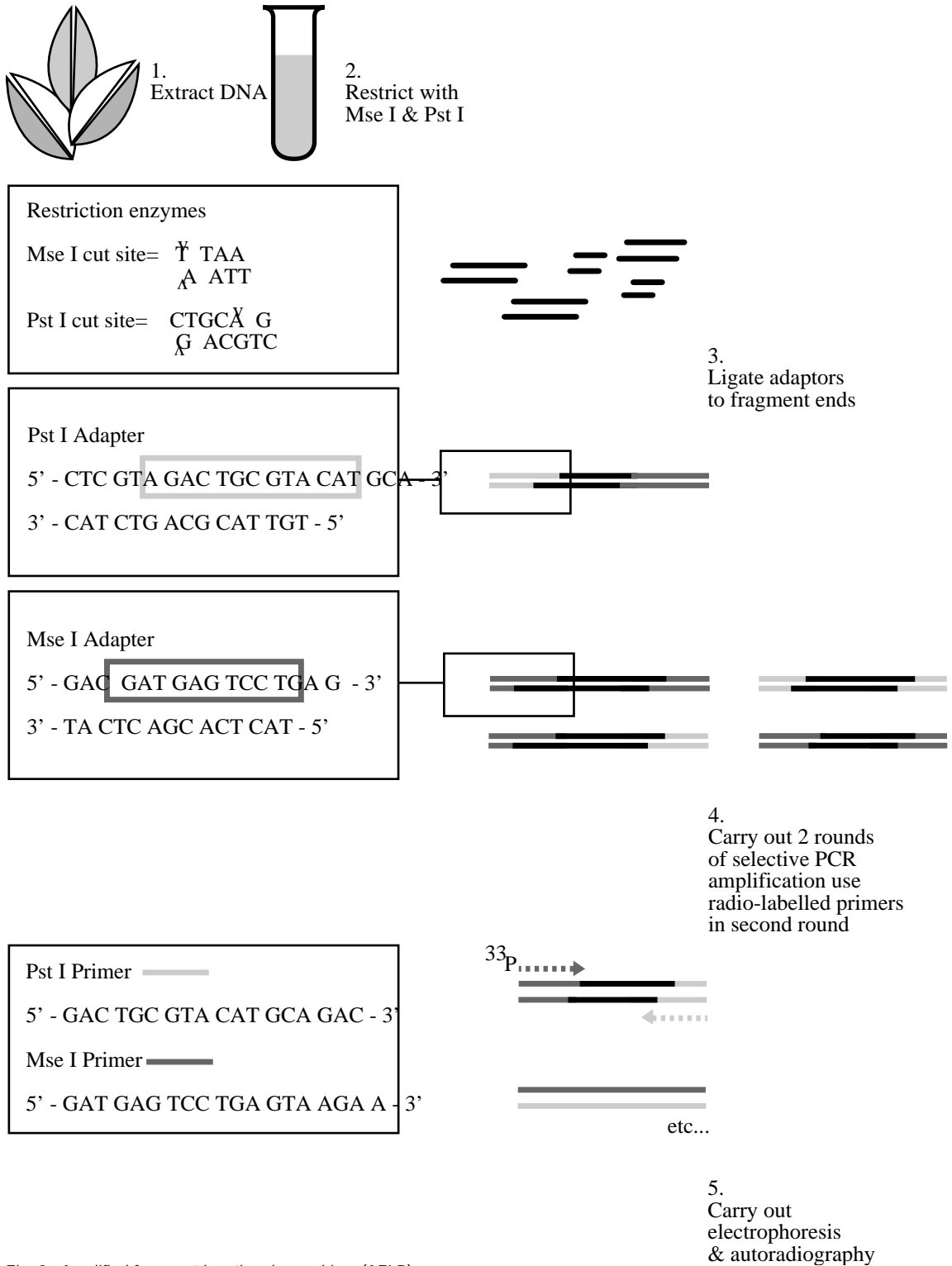


Fig. 8. Amplified fragment length polymorphism (AFLP)

provide an effective means of detecting several polymorphisms in a single assay (Fig. 9). All the evidence so far indicates that they are as reproducible as RFLPs. They require more DNA (0.3-1.0 µg per reaction) and are more technically demanding than RAPDs, but their automation and the recent availability of kits means that the technology can be brought in at a higher level. Using gel scanners, heterozygotes can be identified, otherwise AFLPs are dominant markers.

3.4. Category 3: site-targeted PCR

The opposite approach to arbitrary amplicon profiling is to design primers to amplify specific regions of the genome (Fig. 3). The targeted amplified product can be compared on an agarose gel to the corresponding product from another individual, but only changes that are many base pairs in length will be detected. Sequencing manually, or using an automated DNA sequencer, will potentially resolve all possible differences and data from the aligned sequences can then be compared. This approach is applicable to extremely small samples, e.g. single pollen grains or tiny leaf fragments.

A number of gel systems, such as TGGE (thermal gradient gel electrophoresis) (Riesner *et al.* 1992), DGGE (denaturing gradient gel electrophoresis), single strand conformational polymorphism (SSCP) (Hayashi 1992) and heteroduplex (HD) formation (White *et al.* 1992), provide sensitive detection assays for sequence variations that can assist in the detection of sequence differences without the need to sequence all the samples. These detection systems are based on the principle of a comparing differences in the stability, or configuration, of the DNA under specific gel conditions. Mutations which change the composition of base pairs in fragments amplified from different individuals may alter stability/configuration and, thus, be detectable as difference in mobility in the gels. They are quite technically demanding and require highly controllable conditions. In the simpler PCR-RFLP, or Cleaved Amplified Polymorphic Sequence (CAPS) procedures the amplified product is digested with a specific restriction enzyme and the products directly visualised on the agarose gel by ethidium bromide staining (Akopyanz *et al.* 1992; Tragoonrung *et al.* 1992; Ghareyazie *et al.*

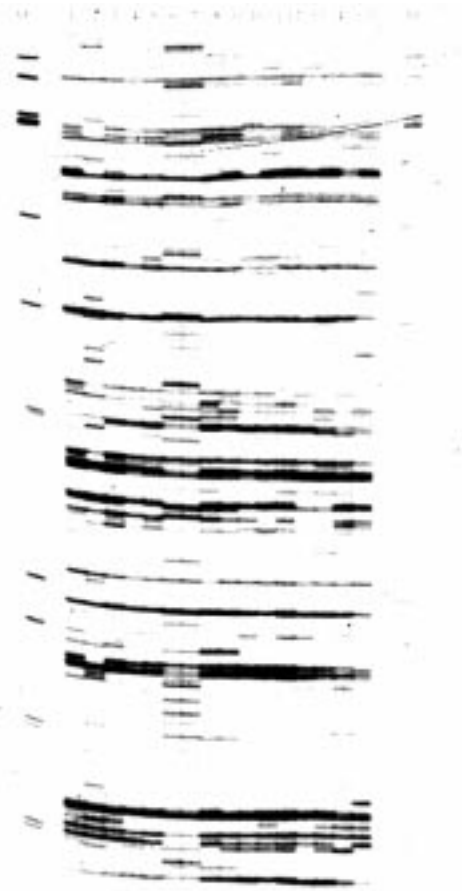


Fig. 9. AFLP analysis of 16 willow (*Salix*) clones by AFLP (source: A. Karp, IACR-LARS)

Notes

1995). This approach is most informative when the restriction sites are mapped within the locus.

Plants possess three different genomes, and therefore three potential sources of sequences for a PCR-targeted approach. The chloroplast genome (cpDNA) is maternally inherited in most angiosperm species and paternally inherited in most gymnosperms. It is highly abundant in leaves and therefore amenable to isolation. The entire cpDNA sequence is known for a few species and appears to be highly conserved in terms of size, structure, gene content and order. Primers are available that will work across broad taxa and can be used for diversity studies at all taxonomic levels (Demesure *et al.* 1995). In contrast, the mitochondrial genome (mtDNA) is less abundant in leaves, there is less background knowledge, fewer probes are available and these have been less well characterised. The high rates of structural rearrangements and the relatively low rates of point mutations mean it is of limited use at interfamily and interspecific levels but the high frequency of rearrangements, which can be easily detected as RFLPs, mean that mtDNA can be very useful for detecting variation at the intraspecific and population levels. Primer pairs for conserved regions of mtDNA sequences are available (Demesure *et al.* 1995). For the nuclear genome, only the rDNA (ribosomal RNA) gene family has been used extensively for diversity studies (Zhang *et al.* 1990). Ribosomal RNA genes are located at specific chromosomal (NOR) loci where they are arranged in tandem repeats which can be reiterated up to thousands of times. Each repeat unit comprises a transcribed region separated from the next repeat by an intergenic spacer (IGS). The transcribed region comprises an external transcribed spacer (ETS), the 18S gene, an internal transcribed spacer (ITS1), the 5.8S gene, a second internal transcribed spacer (ITS2) and the 26S gene. Primer pairs have been designed which will enable amplification of the different regions in a wide range of organisms. These regions evolve at different rates and can thus be used at all taxonomic levels, although in practice it can be difficult to detect sufficient variation at the below-species level.

The advantages of PCR-sequencing approaches are in the quality of the data and the information engendered. The fragment in which polymorphisms are studied is of known identity and, as discussed later, this approach reveals information on phylogenetic relations. However, there are also clear disadvantages. Unless the frequency of variants is high enough for detection by PCR-RFLP, or other sensitive gel assays, sequencing of all individuals is required, which is resource

intensive. The coverage of the genome is highly restricted, often to only one sequence. Although cpDNA and mtDNA primers are available, there are currently few nuclear genes that can be used at the below-species level and the rate at which sequences vary (and therefore the success of this strategy) also appears to differ between genomes. Because of the importance of low copy nuclear markers, numerous efforts are currently being expended towards the identification of universally useful primer pairs (Strand *et al.* 1997). Additional problems, when conserved primers are used for PCR, are contamination by DNA from other organisms and the detection of multiple gene copies and pseudogenes.

3.4.1 Sequence-tagged microsatellites (STMS)

Microsatellites or simple sequence repeats (SSRs) are highly mutable loci which may be present at many sites in a genome. Since the flanking sequences at each SSR may be unique, if SSR loci are cloned and sequenced, primers to the flanking regions can be designed to define a sequence-tagged microsatellite (STMS) (Fig. 10) (Beckman and Soller 1990). There are several important advantages of sequence-tagged microsatellites. They are (usually)

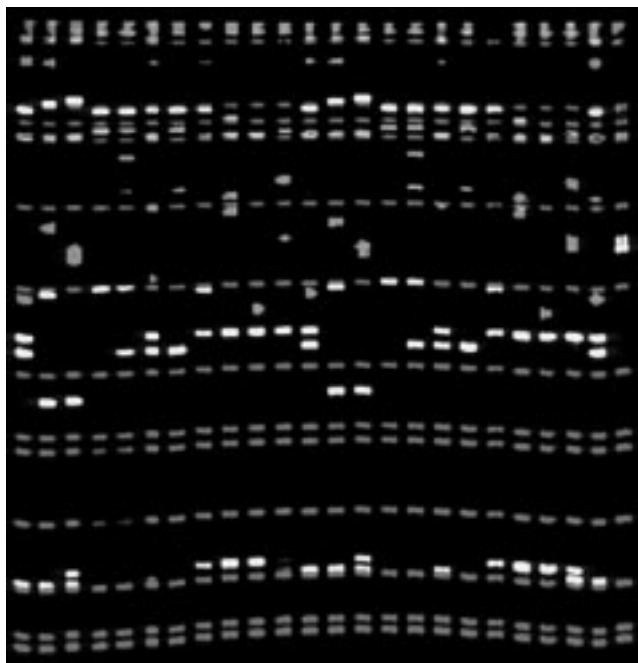


Fig. 10. Sequence tagged microsatellite analysis. Gel image highlighting fluorescent SSR typing. The gel includes 12 individuals (replicated twice) of three cultivated *Brassica* spp. screened with eight SSRs. (source: S. Mitchell, C. Jester, S. Kresovich)

Notes

a single locus which, because of the high mutation rate, is often multi-allelic (Saghai-Marroof *et al.* 1994). They are codominant markers and can be detected by a PCR (non-hybridisation based) assay. They are very robust tools that can be exchanged between laboratories and their data are highly informative (Morgante and Oliveri 1993). Although some changes can be resolved on agarose gels, it is common to distinguish STMS on polyacrylamide sequencing gels where single repeat differences can be resolved and all possible alleles detected. The assay is relatively quick and throughput can be increased by selecting a small number of different STMS with alleles of non-overlapping size ranges and multiplexing either the PCR reactions, or, more easily, the products of the separate reactions, so that all the alleles of the different loci can be run in a single lane on the gel. Multiplexed STMS have also been automated (e.g., Mitchell *et al.*, in press). Unless the investigator is extremely fortunate, however, STMS will not be available for their species of study. Retrieval of microsatellites has not been easy in plants because of their relatively low abundance compared with animal genomes. STMS often show limited cross-transferability to other genera and even to other species within the same genus. An investigator wishing to use microsatellites is thus probably first faced with having to isolate them. Whilst retrieval strategies have now been devised which work with high efficiency (e.g. Edwards *et al.* 1996), STMS development necessitates a considerable investment of time and extra skilled expertise and resources.

3.5. Variations or combinations of the basic techniques

The basic molecular techniques described above can be further refined and also combined in many ways. Sequence tagged site, or STS, is the general term given to a locus defined by its primer sequences. An STS can be created for any site, provided that the locus can be cloned and sequenced. This may be desirable, when for example RFLP probes are being used to test large numbers of samples (e.g. Livneh *et al.* 1992), or when a stable, robust and reliable PCR marker linked to genes controlling a trait of interest is required. Sequence characterized amplified regions (SCARs) are derived from individual RAPD markers (e.g. Paran and Mitchelmore 1993). The RAPD fragments (bands) are cloned, the nucleotide sequences of the terminal ends are determined and used to design primers for specific amplification of the desired fragment. There are also many semi-arbitrary PCR methods: In Directed Amplification of Minisatellite-region DNA (DAMD), VNTR core sequences, such as M13, are used as

primers in PCR reactions (Heath *et al.* 1993). Multi-locus banding patterns usually result. In Single Primer Amplification Reaction (SPARs), the principle is similar but primers are based on the core motifs of microsatellites (e.g. Gupta *et al.* 1994). Again, polymorphic banding patterns are produced. Inter-simple sequence repeat amplification (ISSR) is similar to SPARs but involves the anchoring of designed primers to a subset of SSRs and results in the amplification of the regions between two closely spaced oppositely oriented SSRs (e.g. Kanety *et al.* 1995). Microsatellite primers can also be used in conjunction with AFLPs in a technique referred to as SAMPL (Morgante and Vogel 1994).

4. USING THE DIFFERENT TECHNIQUES

The brief review of the molecular techniques given above is by no means exhaustive and it is important to appreciate that this is a fast-evolving field in which new developments are continually emerging (see for example the abstracts for Plant and Animal Genome V Conference, January 1996 at: <http://probe.nalusda.gov:8000/otherdocs/pg/pg5/allabstracts.html>). Nevertheless, the account above should give an indication of the type and scope of the basic techniques available. A potential user would be forgiven for feeling overwhelmed by the range of possible technologies and their acronyms. However, the different techniques share a number of characteristics in common depending on the general features of the technology, the nature and analysis of the data produced and the ways in which they can be applied.

4.1. The basic technologies

As noted above, three basic categories can be identified based upon, firstly, whether the assays are PCR-based and, secondly, whether arbitrary/semi-arbitrary primers or specifically designed primers for known sequences are used:

- **Category 1:** non-PCR based methods, e.g. RFLP, VNTR (used as probes in genomic hybridisations)
 - **Category 2:** arbitrary (or semi-arbitrary) primed/or multi-locus profiling techniques, e.g. RAPD, DAMD, AP-PCR, ISSR, DAF, SPARs, AFLPs, SAMPL
 - **Category 3:** site targeted PCR techniques, e.g. PCR-SEQUENCING, TGGE, DGGE, CAPS, SSCP, HETERODUPLEX, STMS.
-

Notes

4.2. Analysis of molecular data

It is imperative to understand the different ways that the data generated by molecular techniques can be analyzed before considering their application to diversity studies (Hillis and Moritz 1990; Soltis *et al.* 1992; Avise 1994; Weir 1996). Two main types of analysis will be relevant to genebank curators:

1. analysis of genetic relationships among samples
2. calculation of population genetics parameters, in particular diversity and its partitioning at different levels.

The analysis of genetic relationships among samples starts with the construction of a matrix specifying the character-state of each marker for each sample. A sample will usually be DNA from an individual, but could consist of DNA bulked from a number of individuals (e.g. to represent an accession or taxon). Marker states may be binary, as in the presence or absence of RAPD bands or restriction sites (as revealed by RFLPs and related techniques), or multi-state, as in the nucleotide (A, T, C or G) present at a particular position in a DNA sequence.

This sample \times marker matrix of character-states is then commonly used to construct a sample \times sample matrix of pair-wise genetic distances (or similarities). There are several different ways of calculating the genetic distance (or similarity) between two samples on the basis of the differences between them in the states of a set of genetic markers (e.g. Hendrick 1974), but a commonly used index is Nei's genetic distance (D) (Nei 1973).

There are two main ways of analyzing the resulting distance (or similarity) matrix and displaying the results. One is to use Principal Coordinate Analysis (PCO) to produce a 2- or 3-dimensional scatter plot of the samples such that the geometrical distances among samples in the plot reflect the genetic distances among them with a minimum of distortion. Aggregations of samples in such a plot will reveal sets of genetically similar material. Another approach is to produce a dendrogram (or tree-diagram) linking together in clusters samples that are more genetically similar to each other than to samples in other clusters. Clusters are linked to each other at progressively lower levels of similarity until all the samples being analyzed are included in a single cluster. Such Cluster Analysis may proceed according to a range of different algorithms, but some of the more widely used ones include Unweighted Pair Group Method with Arithmetic Averages (UPGMA), Neighbour-Joining Method and Ward's Method. Different combinations of genetic distance/

similarity index and clustering algorithm may give rise to somewhat different dendrograms.

Both PCO and cluster analysis are so-called 'phenetic' methods in that they are based on measures of overall distance or similarity among samples. However, there is another, philosophically quite distinct approach to the analysis of genetic relationships, referred to as 'cladistics'. Cladistic analysis also begins with the sample x marker character-state matrix, and also results in dendrograms, though these are sometimes called cladograms to distinguish them from the phenograms of cluster analysis. The difference is that two samples are placed together in the same cluster (or clade) of a cladogram not on the basis of high genetic similarity between them calculated from all markers taken together, but because they share a particular state of a given marker (or markers). The two approaches are also sometimes distinguished as 'distance' and 'character-state' respectively. Because it is possible to generate many cladograms from a single dataset, due to conflicts among characters, so-called parsimony approaches are used to choose among them. A most-parsimonious cladogram is one that requires the least number of character-state changes. There is a wide range of parsimony algorithms, each with its own data requirements and assumptions. Some require that the polarity of character changes be known, i.e. which character states are ancestral and which derived. Cladograms are reconstructions of phylogenies. RAPD data, because of uncertainty over the identity of bands, is not usually thought suitable for this kind of analysis. Box 1 lists some phylogeny reconstruction software.

Turning now to the measurement of genetic diversity and genetic structure (among and within populations), the F-statistics of Wright (1965, 1978) and the G-statistics of Nei (1973) are commonly employed. Estimates of these statistics are based on allele frequencies, and the most appropriate molecular data for such statistical analyses are clearly those in which allele frequencies can be determined directly,

Box 1. Analysis of molecular data: some distance data and clustering programmes

NTSYS pc (Numerical Taxonomy System): F.J. Rohlf; from Exeter Software, 100 North Country Road, Setauket, NY 11733, USA (price: \$155).

SAS also include various clustering algorithms, although it is more difficult to use than NTSYS. Cary, NC.

BIOSYS-1: D.L. Swofford, 1989; BIOSYS-1, a computer program for the analysis of allelic variation in population genetics and biochemical systematics, release 1.7. Illinois natural history Survey, Urbana, IL, USA.

RAPDistance version 1.03: J.A. Armstrong, R. Gibbs, R. Peakall, G. Weiller, 1995. RAPDistance, Package Manual. Australian National University, Canberra, Australia.

ftp://life.anu.edu.au/pub/molecular_biology/software/rapid103.zip

MIXED PROGRAMMES (include both phenetic and parsimony algorithms).

PHYLIP (Phylogeny Inference Package): J. Felsenstein; from the author, Department of Genetics, SK 50, University of Washington, Seattle, WA 98195, USA (free).

email: joe@genetics.washington.genetics.edu

Web site: <http://evolution.genetics.washington.edu.phylip.htm1>

MEGA (Molecular Evolutionary Genetic Analysis): S. Kumar, K. Tamura, M. Nei, 1993; from Joyce White, Institute of Molecular Evolutionary Genetics, 328 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802, USA (price: \$50). email: imeg@psuvm.psu.edu

Box 2. Analysis of molecular data

MALIGN: W. Wheeler, D. Gladstein; from the authors, Dept of Invertebrates, American Museum of Natural History, Central Park West, 79th Street, New York, NY 10024 5192, USA (price: \$50).

CLUSTAL W: J.D. Thompson, D.G. Higgins, J.T. Gibson, 1994; CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

email: gibson@embl-heidelberg, de.: DesHiggins@ebi.ac.uk
Anonymous ftp: ftp.ebi.ac.uk; and: ftp.bioindian.edu

Parsimony programmes

PAUP (Phylogenetic Analysis Using Parsimony); D. Swofford, Laboratory of Molecular Systematics, Smithsonian Institution, Washington DC 20560, USA (price: \$100).

MACCLADE (Analysis of Phylogeny and Character Evolution): W.P. Maddison, D.R. Maddison; from Sinauer Associates, 108 North Main Street, Sunderland, MA 01375, USA (price: \$100).

HENNIG86: J.S. Farris; from the author; Molekylarsystematiska laboratoriet, Naturhistoriska riksmuseet, S 104 05 Stockholm, Sweden (price: \$50).

DADA and CLADOS: by K. Nixon, L. H. Bailey Hortorium, Cornell University, Ithaca, NY 14853, USA.

RNA (Rapid Nucleotide Analysis) by J.S. Farris, from the author, Molekylarsystematiska laboratoriet, Naturhistoriska riksmuseet, S 104 05 Stockholm, Sweden (price: \$50).

NONA by P.A. Goloboff, from the author, Fundación Miguel Lillo, Miguel Lillo 251, 4000 San Miguel de Tucuman, Tucuman, Argentina (price: \$50).

TREEVIEW by R.D.M. Page, Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, UK.

mail: dpage@udcf.gla.ac.uk

<http://taxonomy.zoology.gla.ac.uk/rod/treeview.htm>

such as RFLPs, STMS and sequence haplotypes. Of these, sequences and restriction site data are unique among molecular markers in providing both frequency and phylogenetic information. Nevertheless, suitable statistical treatments are also available for dominant markers such as RAPDs, though in at least one case population differentiation coefficients based on indirectly estimated RAPD frequencies were not concordant with those based on RAPD frequencies directly estimated from haploid macrogametophytes (Szmidt *et al.* 1996). Careful treatment also needs to be given to difficulties arising from the occurrence of large numbers of alleles at one locus in STMS, and for various sources of sampling error within and among populations (Weir and Cockeram 1984). The software packages in Box 2 may be used to calculate genetic parameters and distances and those in Box 3 can be used for general statistical analysis.

4.3. Application of the techniques**4.3.1 Category 1 techniques:
Non-PCR-based techniques**

RFLPs are codominant markers and allele frequencies, and therefore population statistics can thus be calculated directly for single copy loci. They are useful markers for population studies and diversity classification, provided that sufficient polymorphisms can be detected in the species under study. Unless they are recorded as a combination of probe and restriction site data, RFLPs need to be converted into frequency data which have some limitations (see discussion below). When VNTRs are used as probes in RFLPs, multi-locus profiles are produced which share the same features, and thus applications, described for category 2 techniques below. This is also true for RFLPs in which the probes used are homologous to highly-repeated sequence families where several bands will also occur on a gel with a single probe enzyme combination.

4.3.2 Category 2: arbitrary (or semi-arbitrary) primed/or multi-locus profiling techniques

The data derived from arbitrary primed, AFLP and multi-locus fingerprinting approaches have their strength in distinguishing individuals. Major applications of these approaches are thus in establishing identities, in determining parentage, in fingerprinting genotypes and in distinguishing genotypes below the species level, such as cultivars and clones (e.g. Lu *et al.* 1996; Sharma *et al.* 1996; Tohme *et al.* 1996). The difficulty of achieving robust, repeatable, profiles in arbitrary primed approaches such as RAPDs does, however, make their reliability for 'typing/fingerprinting' questionable. It is likely that RAPDs can be used successfully in a single lab when all the operating conditions can be carefully controlled. However, problems in repeatability suggest that caution should be exercised in including RAPDs' data in databases intended for widespread access and use.

Because of their ease of use and universal applicability, RAPDs, in particular, have been used in all kinds of diversity studies at all taxonomic levels, including population and phylogenetic studies. While this is possible, investigators employing RAPDs (or other category 2 approaches) for such applications should be mindful of the limitations of their data. Category 2 type techniques produce multi-band profiles, in which the number and placement of bands generated vary depending upon the technique and the primers used. These techniques compare different genomes at several points but the identity of these points is not known. In using data from such multi-band profiling procedures it is extremely important to recognize that: (i) they are usually dominant markers; (ii) in the absence of pedigree analysis, the identity of individual bands is not known and there may be uncertainty in assigning markers to specific loci; (iii) the presence of a band of apparently identical molecular weight in different individuals is not evidence that the two individuals share the same homologous fragment, and (iv) single bands can sometimes be comprised of several comigrating amplification products. These limitations in data quality are

Box 3. Analysis of molecular data: general statistical packages

WINAMOVA program version 1.55 provided by L. Excoffier.
 email=excoffie@sc2a.unige.ch
<ftp://acasun1.unige.ch/pub/comp/win/amova.amova.155.zip>
 RAPDistance version 1.03. J. Armstrong, A. Gibbs, R. Peakall, G. Weiller. 1995; RAPDistance, Package Manual. Australian National University, Canberra, Australia.
ftp://life.anu.edu.au/pub/molecular_biology/software/rapd103.zip
 GDA (Genetic Data Analysis), 1996. P.O. Lewis and D. Zaytin. 1996. Genetic Data Analysis: software for the analysis of discrete genetic data. Sinauer Assoc., Sunderland, MA 01375, USA.
 FSTAT J. Goudet. 1995. Fstat version 1.2. A computer program to calculate F-statistics. J. Hered. 86:485-486.
 RAPDFst B.L. Apostol, W.C. Black, P. Reiter, B.R. Miller. 1996. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. Heredity 76:325-334. Anonymous ftp: lamar.costate.edu in directory/pub/wcb4
 MICROSAT and DELTAMU D.B. Goldstein, A Ruiz-Linares, M. Feldman, L.L. Cavalli-Sforza. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Nat. Acad. Sci. USA 92:6720-6727.
 email:minch@lotka.stanford.edu
<http://lotka.stanford.edu/distance.htm>
 GENEPOP version 1.2 M. Raymond, F. Rousset. 1995. GENEPOP (V. 1.2): A population genetics software for exact tests and ecumenicism. J. Hered. 86: 248-249.

Notes

important because they reduce the efficiency of the analytical methods described in the previous section, as assumptions, such as independence (i.e. that the markers do not represent the same or linked mutation), known mutational models, neutrality, non-recombination, etc. are essential facets of the models used. In using RAPDs for population studies, for example, these limitations do not prevent the estimation of allele frequencies necessary for population genetic analysis, but they do reduce the accuracy of such estimation relative to codominant markers such as RFLPs. To achieve the same degree of statistical power using RAPDs (or any other dominant marker system), compared with codominant markers, 2-10 times more individuals need to be samples per locus (Lynch and Milligan 1994). In the use of RAPDs for phylogeny more criteria need to be satisfied to give credence to the analysis (Clark and Lanigan 1993). Similar criticisms can be raised against all category 2 techniques. AFLPs in principle share the same features of data quality as RAPDs but they are also a relatively new technology, about which information is continually being revealed regarding the identity and distribution of bands in the genome. Investigators should thus keep a watchful eye on the latest discussions of this technology in the future.

4.3.3 Category 3: site-targeted PCR sequence techniques

Sequences are the only molecular markers that contain a comprehensive record of their own history. In addition to revealing the groupings of individuals into different classes, appropriate analyses based on sequence data (or restriction site data) can thus provide hypotheses on the relationships between the different categories that are classed together. In contrast, frequency data from category 2 markers (e.g. RAPDs, AFLPs) only provide the means to classify individuals into nominal genotypic categories. It is argued by many that technologies that yield sequence data are, thus, the only appropriate methods for taxonomic studies, indeed for any study in which phylogenetic information is important. This is an important point to grasp for population studies, particularly when the diversity data are used for conservation. In principle, sequencing will allow the determination of which gene sequences, in samples taken from within or between populations, are the most closely related and hence share a most recent common ancestor. For such genealogical relationships (which may be separate from the genealogy of the individuals carrying genes) the influences of genetic factors, such as mutation, are independent from demographic factors such as population size, whereas in the case of markers that provide only frequency information, these factors are

confounded (Milligan *et al.* 1994). This difference is of particular relevance to conservation, where demography (the description and prediction of population growth and age structure) is considered to be as, or more, important than genetic factors (Lande 1988). Because of this very important feature of sequence data, they may be viewed as the ultimate molecular data to be retrieved. The problems with utilising PCR-sequence approaches at the below-species level, such as in population studies or in assessing diversity in cultivars or accessions, however, are that the rate of sequence divergence may simply be insufficient and that horizontal gene transfer (e.g. through introgression) may confound relationships. While a PCR-sequence analysis approach may not distinguish cultivars, category 2 techniques or STMS may. Similarly, in population studies there are only a few primer sets for conserved sequences available that are known from previous studies to reveal polymorphism at suitable rates.

The STMS technique occupies a somewhat intermediary position. It is a PCR-based assay of a single locus with, potentially, an infinite number of alleles. Identity and assignment of alleles are thus not a problem. The markers are codominant so allele frequencies can be determined directly and their rate of change renders them particularly suitable for below-species studies. STMS therefore provide ideal tools for population studies and for assessing diversity among genotypes within species. The problem with STMS concerns the mutational mechanisms by which alleles arise and the occurrence of large numbers of allelic variants. The accuracy with which true homology can be inferred for different genotypes diminishes as genetic distance becomes greater, because of the increasing possibility that different forward and back mutation events may result in alleles of the same size. Phylogenetic inferences are therefore problematic with STMS. Similarly, some population genetics estimates require careful treatment to account for the large numbers of alleles. In both cases, appropriate statistical procedures are being developed. Another problem with STMS is the occurrence of null alleles, as a result of mutation in the primer site. These will not produce a band on a gel and heterozygotes with null alleles can therefore be misclassified as homozygotes.

4.4. Pre-screening and combinations of techniques

Although techniques have been described as separate categories, in applying them to any diversity study two important points should be borne in mind. Firstly, it is desirable to carry out a pre-screen of a sample of individuals before embarking upon a full screening programme. This will enable a preliminary data

Notes

set on the extent and distribution of variation for the species in question to be obtained from which the full experiment can then be better designed. Pre-screening can be performed with quick, low-cost procedures such as RAPDs, or the techniques of choice may be tested to determine the level of polymorphisms they detect. The pre-screen should include accessions or varieties with known pedigrees, some related and some unrelated, in order to test the discriminatory power of the technique being applied. The results can then be used to help select the most appropriate techniques and to determine the number of individuals that should be assayed per collection.

A second point is that because all the techniques suffer some limitations, best results may be obtained by combining more than one approach. For example, STMS and PCR-sequencing or organellar genes (using conserved chloroplast primers) provide a powerful combination for studies of population diversity, differentiation and history, as well as of geneflow and demographic processes (Pope *et al.* 1996).

5. A FRAMEWORK FOR SELECTING APPROPRIATE TECHNIQUES

The final part of this bulletin is a key to the technologies which may help genebank managers and other investigators select the most appropriate technique for their diversity study, given the constraints of time, money or other resources they face. It should be understood that the process outlined is flexible regarding which techniques are most appropriate for which purposes. The aim is to provide a logical framework in which the different methodologies can be assessed.

It is important to appreciate that molecular genetics is a rapidly developing field and frequently technologies are advancing faster than our understanding of their full potential or limitations. New techniques are continually being described and new information about pre-existing techniques is continually altering our understanding and interpretation of data obtained from them. Furthermore, molecular geneticists often disagree about which techniques should be chosen in a given experiment. It is not the case, therefore, that there is only one technique that should be chosen, but rather that there are clear reasons why it is better to choose some techniques compared with others and the limitations of any chosen technique should be recognized.

The decision framework is presented in the form of a chart (Fig. 11) in which the major steps in the decision process can be briefly explained as follows.

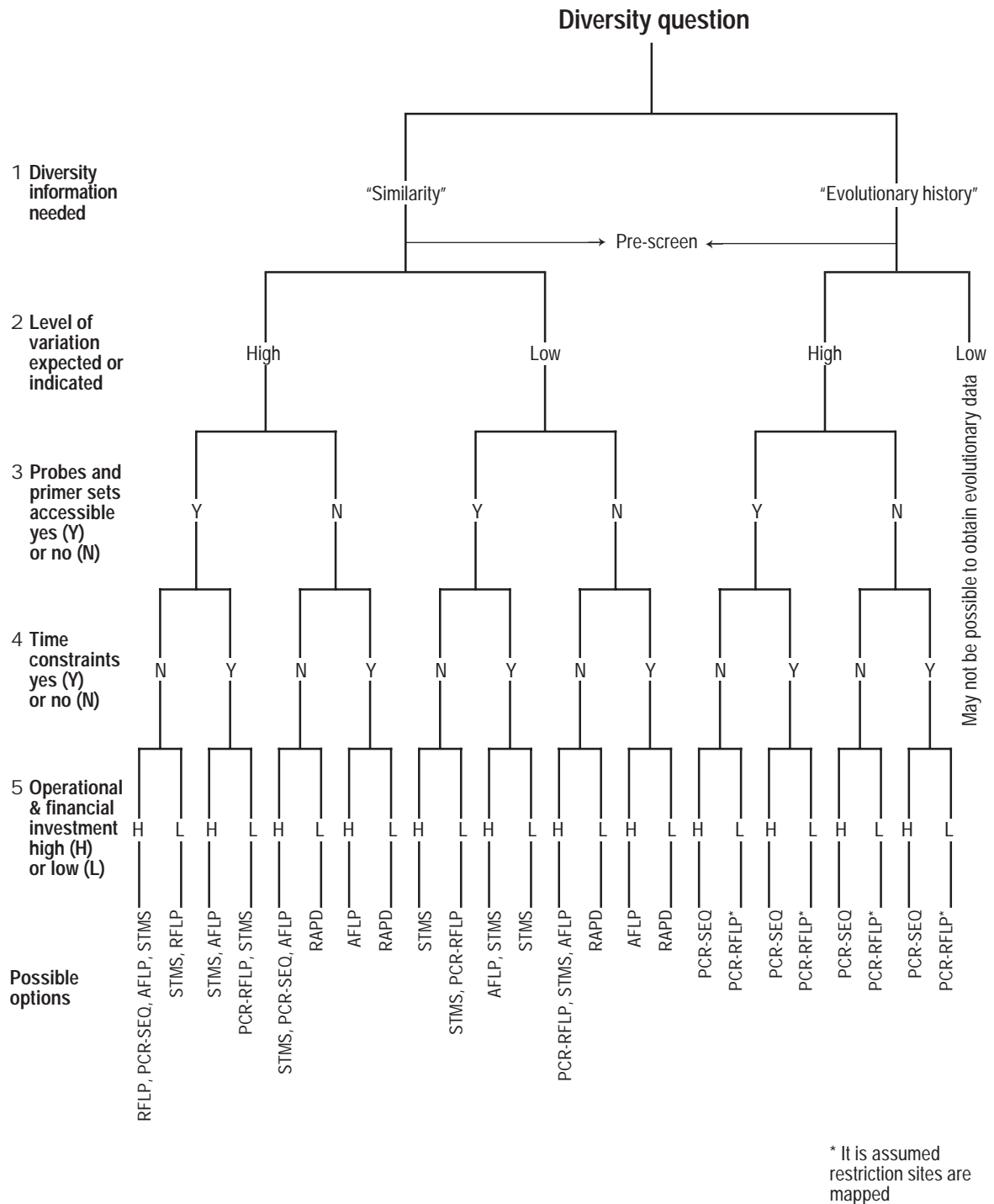


Fig. 11. Decision-making chart for the selection of molecular screening techniques

Notes

Decision 1. What question(s) is (are) being asked in the diversity study?

Application of molecular techniques to diversity questions must take into account whether or not the data derived from a technique provide the right type of information for answering the question being addressed. Foremost, it is important to be clear about what exactly needs to be determined – is it sufficient to know how many different classes are present and how these can be grouped based upon how similar they are, or is it important to know how these different classes are related, whether or not they can be ordered, which classes are ancestral or older and what their evolutionary histories are? If the question is simply how many different classes are present, virtually any technique can be used since all the ones described here give distance measures. The final choice will be determined by other decisions lower down the chart. If, however, the question of interest concerns origins, population history or phylogenetic relationships and it is thus essential that the evolutionary histories of the classes are to be determined, only sequences or restriction site data can be utilized as these can be used to assign different character states.

For many conservation problems, information on the evolutionary history of the material will not be needed. For example, useful information to guide acquisition programmes, monitor genetic integrity, characterize accessions or develop core collections can be obtained from virtually all the techniques. Decisions on which technique to use will depend on other factors. Sequence or restriction site data will be important when the questions asked concern taxa evolution, when knowledge of the evolutionary history of different populations is important or when information is needed on the precise relationship of specific groups of accessions in a genebank such as related cultivars.

Decision 2. What is the anticipated level of polymorphism?

The next point concerns the level of anticipated variation. This will depend upon the taxonomic levels of the material being studied, i.e. different species, subspecies, populations, cultivars and individuals. The closer the relationship, the more necessary it may be to consider highly discriminatory techniques such as AFLPs or STMS. Using isozyme data Hamrick and Godt (1990) have shown that amounts of variation are generally correlated with a variety of life traits, i.e. taxonomic status (gymnosperm/angiosperm, monocot/dicot), life form (annual vs. perennial), geographic range (endemic vs. widespread), regional distribution

(temperate vs. tropical), breeding system (self vs. outcrossing), seed dispersal (gravity vs. wind), mode of reproduction (sexual vs. asexual), successional status (early, mid or late). Much higher levels of genetic diversity are usually found in cross-pollinated species than in self-pollinated ones, in perennials than in annuals and in late successional species. Some annual self-pollinated crops such as soybean (Keim *et al.* 1989; Lin *et al.* 1996) have rather low levels of variation as compared with cross-pollinated ones such as maize or vegetable Brassicas (Shattuck-Eidens *et al.* 1990). If the anticipated level of polymorphism is high, the choice of techniques is very wide (see Table 1), but data analysis may be less problematic with lower resolution techniques such as RFLPs or CAPS since too high a level of polymorphism may result in all samples appearing unique. If, however, the anticipated level is low, the choice should centre on highly discriminative techniques such as AFLPs or STMS (Table 1). A preliminary experiment (pre-screen) is always advisable to ascertain the amount of variation likely to be found in the material under study.

Table 1. Comparative assessment of some of the salient characteristics of different molecular genetic screening techniques

Characteristic	RFLPs	RAPDs	Sequence-tagged SSRs	AFLPs	PCR sequencing
Development costs (\$ per probe)	Medium (100)	Low (none)	High (500)	Low (none)	High (500)
Level of polymorphism	Medium	Medium	High	Medium	Medium
Automation possible	No	Yes/No	Yes/No	Yes/No	Yes
Cost of automation	High	Medium	High	High	High
Repeatability	High	Low	High	Medium	High
Level of training required	Low	Low	Low/Medium	Medium	High
Cost (\$ per assay)	High (2.00)	Low (1.00)	Low (1.00)	Medium (1.50)	High (2.00)
Radioactivity used	Yes/No	No	Yes/No	Yes/No	Yes/No
Samples/day (without automation)	20	50	50	50	20

Decision 3. Are probes/primer sets accessible or not?

The third point relates to whether or not the genebank manager, curator or associated investigators can have access to and use pre-existing sequence information, probes or primer sets or whether these have to be developed *de novo*. No development costs are associated with the deployment of techniques such as RAPDs, AFLPs, ISSR, etc. (Table 1). This is also true for RFLPs if suitable probes are accessible. For STMS or PCR-sequencing, a limited number of specific primer sets are available and more are being

developed but their utility is limited to the respective species for which they were developed, or in some cases, to their relatives. Information on these can be obtained through accessing databases available on the Internet. Examples include Grain Genes and Solgenes (<http://probe.nalusda.gov> (see Box 4)), which contains information on Triticeae genetics and *Solanum* species genetics, respectively. A review of the relevant literature may also provide information on primer sets for STMS and where they can be obtained. If probes or specifically designed primer sets are not available or accessible, the option of developing them (6-12 months including significant financial and technical investment) would have to be considered.

Box 4. An example of information on molecular probes available for cereals on the internet at <http://wheat.pw.usda.gov/graingenes.html>

GrainGenes Probe Repository

The GrainGenes Probe Repository was set up to provide the research community with a resource of DNA probes for plant research. At this time, it is attempted to have chromosome assignments designated for each of the probes within the repository. To date, the probe sources have origins in wheat, oat and barley genomes. Further information about probes included within the repository is also indexed within the GrainGenes genome database.

Probes are distributed free of charge for research under conditions established by the probe originator. Some probes may require communication with the probe originator before shipment of probes from the repository.

Requests for probes should take the following format:

NAME:
 TITLE:
 INSTITUTION:
 EMAIL:
 PHONE NUMBER:
 PURPOSE OF REQUEST:
 CLONES DESIRED: listed by repository id# (not clone name)

[Please do not request more than 25 clones at a time, unless special arrangements are made with the repository center.]

Listings are currently being sorted; future tables will provide probe origin and locations mapped with grain crop genomes. Probes currently available are listed in table format by chromosome assignments:

- Probes Assigned to Chromosome 1
- Probes Assigned to Chromosome 2
- Probes Assigned to Chromosome 3
- Probes Assigned to Chromosome 4
- Probes Assigned to Chromosome 5
- Probes Assigned to Chromosome 6
- Probes Assigned to Chromosome 7
- Unassigned Probes

Information on these can be obtained through accessing databases available on the Internet. Examples include Grain Genes and Solgenes (<http://probe.nalusda.gov> (see Box 4)), which contains information on Triticeae genetics and *Solanum* species genetics, respectively. A review of the relevant literature may also provide information on primer sets for STMS and where they can be obtained. If probes or specifically designed primer sets are not available or accessible, the option of developing them (6-12 months including significant financial and technical investment) would have to be considered.

Decision 4. Are there any time constraints?

The urgency of acquiring the data and the sample sizes that need to be screened must be considered in choosing the molecular genetic screening technique. If time is not a constraint, more informative, accurate and robust techniques such as STMS and PCR-sequencing should be pursued, whereas if there is urgency, immediately applicable techniques should be chosen such as RAPDs and AFLPs, or STMS and PCR-sequencing if primer sets are already available and accessible. In addition to consideration of time constraints, a curator must consider the number of assays against the costs incurred for different molecular genetic techniques (Table 1).

Decision 5. What financial and operational resources can be dedicated to answering the question?

The question must be dealt with both qualitatively and quantitatively

including: the level of technical expertise required to conduct the assay; the supporting infrastructure (facilities and instrumentation); laboratory containment capabilities; availability and cost of reagents; cost per assay; development and operation costs. For situations where operational and financial investments are constrained, techniques such as RAPDs would be an option because of the lower level of skill required, low costs per assay and the ready availability of primers and other requirements (Table 1). If increased accuracy is desired and technical support is available, CAPS, RFLPs or AFLPs (depending upon the question asked) offer better alternatives. Once accessible, STMS are an inexpensive, informative and accurate option for characterising diversity below the species level, whereas PCR-sequencing, where financial and technical skills are not limiting, may be the option of choice for higher taxonomic levels.

Although this may be one of the most important decision-making steps facing the potential user, it is difficult to provide precise data on financial resources necessary or on the operational resources needed. These vary greatly throughout the world and any useful generalization is difficult. Costs are also continually changing as the technologies develop. Thus, while AFLPs are currently expensive and require substantial infrastructural investment, their costs are likely to be reduced over the next few years. Some relative estimates of costs are given in Table 1 but even these are only very approximate. Intending users will benefit from developing close links with established labs in order to learn from their experience. Where possible, at least one staff member should always obtain some hands-on experience through working, at least for a few months, in an established laboratory.

6. USING THE DECISION-MAKING CHART

By working through the decision-making process outlined above, users will arrive at a shortlist of the techniques likely to be most appropriate to the particular studies they envisage. It should be remembered that the options are based on the assumptions that plant tissue availability is not limiting and that large sample numbers are to be studied. Limitations of tissue availability favour PCR-based techniques such as RAPDs, sequencing and STMS, which require very little DNA.

In many cases there will still be more than one technique appropriate for the planned study. For example, studies on the distribution of diversity where there is considerable variation, primer sets are not available and a quick answer is needed could

Notes

be undertaken using either RAPDs or AFLP. Similarly, both AFLP and STMS methods could be used for studies on the genetic integrity of genebank accessions where diversity is rather low and probe and primer sets are available.

Over the next few years new techniques will probably be developed which will have their own properties and which can be fitted into this kind of decision-making framework based on their specific characteristics. As information accumulates from the use of the existing techniques for different investigations, it will become possible to make more informed decisions on which to use for what purpose (Powell *et al.* 1996). The precise strengths and limitations of different techniques for answering specific questions and for studying specific genepools will become known. A consideration of particular importance for many genebank managers and users will be the relation between molecular data and that obtained from other biochemical or agromorphological studies.

References

- Akopyanz, N., N. Bukanov, T.U. Westblom and D.E. Berg. 1992. PCR based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res.* 20:6221-6225.
- Apostol, B.L., W.C. Black, P. Reiter and B.R. Miller. 1996. Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. *Heredity* 76:325-334.
- Avise, J.C. 1994. *Molecular Markers, Natural History and Evolution*. Chapman & Hall.
- Beckman, J.S. and M. Soller. 1990. Toward a unified approach to gene mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technol.* 8:930-932.
- Beyersmann, B., P. Nürnberg, A. Weihe, M. Meixner, J.T. Epplen and T. Bvrner. 1992. Fingerprinting plant genomes with oligo nucleotide probes specific for simple repetitive DNA sequences. *Theor. Appl. Genet.* 83:691-694.
- Brush, S. 1995. *In situ* conservation of landraces in centers of crop diversity. *Crop Sci.* 35:346-354.
- Burr, B., S.V. Evola, F.A. Burr and J.S. Beckmann. 1983. The application of restriction fragment length polymorphisms to plant breeding. Pp. 45-59 in *Genetic Engineering Principles and Methods*, Vol. 5 (J.K. Setlow and A. Hollaender, eds.). Plenum, New York.
- Caetano-Anollis, G. 1994. MAAP – A versatile and universal tool for genome analysis. *Plant Mol. Biol.* 25:1011-1026.
- Caetano-Anollis, G., B.J. Bassam and P.M. Gresshoff. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Bio/Technol* 9: 553-557.
- Clark, A.G. and C.M.S. Lanigan. 1993. Prospects for estimating nucleotide divergence with RAPDs. *Mol. Biol. and Evol.* 10:1096-1111.
- Demesure, B., N. Sodzi and R.J. Petit. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Mol. Ecol.* 4:129-131.
- Edwards, K.J., J.H.A. Barker, A. Daly, C. Jones and A. Karp. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. *Biotechniques* 20:758-759.
- Evola, S.V., F.A. Burr and B. Burr. 1986. The suitability of restriction fragment length polymorphisms as genetic markers in maize. *Theor. Appl. Genet.* 71:765-771.
- FAO. 1993. *Conservation of genetic resources in tropical forest management: Principles and concepts*. FAO Forestry Paper 107. FAO, Rome, Italy.
- Frankel, O.H., A.H.D. Brown and J. Burdon. 1995. *The Conservation of Plant Biodiversity*. Cambridge Univ. Press, UK.
- Ghareyazie, B., N. Huang, G. Second, J. Bennet and G.S. Kush. 1995. Classification of rice germplasm, I Analysis using ALP and PCR-based RFLP. *Theor. Appl. Genet.* 91:218-227.
- Goldstein, D.B., A. Ruiz-Linares, M. Feldman and L.L. Cavalli-Sforza. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Nat. Acad. Sci. USA* 92:6720-6727.
- Goudet, J. 1995. Fstat version 1.2. A computer program to calculate F-statistics. *J. Hered.* 86:485-486.
- Gupta, M., Y.S. Chyi, J.R. Romero Severson and J.L. Owen. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple sequence repeats. *Theor. Appl. Genet.* 89:998-1006.
- Hayashi, K. 1992. PCR-SSCP: A method for detection of mutations. *Genetic Analysis: Techniques and Applications* 9:73-79.
- Hamrick, J.L. and M.J.W. Godt. 1989. Allozyme diversity in plants. Pp. 318-370 in *Plant Population Genetics, Breeding and Genetic Resources* (A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir, eds.). Sinauer Associates, Sunderland, MA, USA.
- Heath, D.D., G.K. Iwama and R.H. Devlin. 1993. PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. *Nucleic Acids Res.* 21:5782-5785.
- Helentjaris, T., G. King, M. Slocum, C. Siedestrang and S. Wegman. 1985. Restriction fragment polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. *Plant Mol. Biol.* 5:109-118.
- Hendrick, P.W. 1974. Genetic similarity and distance: comments and comparisons. *Evolution* 29:362-366.
- Hillis, D.M. and C. Moritz. 1990. An overview of applications of molecular systematics. Pp. 502-515 in *Molecular Systematics* (D.M. Hillis and C. Moritz, eds.). Sinauer Associates, Sunderland, MA, USA.
- Hodgkin, T., A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales. 1995. *Core Collections of Plant Genetic Resources*. John Wiley & Sons, UK.
- Hoelzel, A.R. and A. Green. 1994. Analysis of population level variation by sequencing PCR-amplified DNA. Pp. 159-187 in *Molecular Genetic Analysis of Populations* (A.R. Hoelzel, ed.). IRL Press, UK.
- Hunter, R.L. and C.L. Markert. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. *Science* 125:1294-1295.
- Kanety, R.V., X. Zeng, J.L. Bennetzen and E.Z. Brent. 1995. Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. *Mol. Breed.* 1:365-373.
- Keim, P., R.C. Shoemaker and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. *Theor. Appl. Genet.* 77:786-792.
-

- Kresovich, S., J.G.K. Williams, B.A. Schaal, J.R. McFerson and E. Routman. 1992. Characterization of genetic identities and relationships of *Brassica oleracea* L. via random polymorphic DNA assay. *Theor. Appl. Genet.* 85:190-196.
- Lande, R. 1988. Genetics and demography in biological conservation. *Science* 241:14515-1460.
- Lewis, P.O. and D. Zaytin. 1996. Genetic data analysis: software for the analysis of discrete genetic data. Sinauer Assoc., Sunderland, MA, USA.
- Lin, J.J., J. Kuo, J. Ma, J.A. Saunders, H.S. Beard, M.H. MacDonald, W. Kenworthy, G.N. Ude and B.F. Matthews. 1996. Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. *Plant Mol. Biol. Reporter* 14:156-169
- Livneh, O., E. Vardi, Y. Stram, O. Edelbaum and I. Sela. 1992. The conversion of a RFLP assay into PCR for the determination of purity in a hybrid pepper cultivar. *Euphytica* 62:97-102.
- Lu, J., M.R. Knox, M.J. Ambrose, J.K.M. Brown and T.H.N. Ellis. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP-based and PCR-based methods. *Theor. Appl. Genetics* 7:1103-1111.
- Lynch, M. and B.G. Milligan. 1994. Analysis of population structure with RAPD markers. *Mol. Ecol.* 3:91-99.
- Milligan, B.G., J. Leebens-Mack and A.E. Strand. 1994. Conservation genetics: beyond the maintenance of marker diversity. *Mol. Ecol.* 3:423-435.
- Mitchell, S.E., S. Kresovich, C.A. Jester, C.J. Hernandez and A.K. Szewc-McFadden. 1997. Application of multiplex PCR and fluorescence-based semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Sci.* (in press).
- Morgante, M. and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. *Plant J.* 3:175-182.
- Morgante, M. and J. Vogel. 1994. Compound microsatellite primers for the detection of genetic polymorphisms. U.S. Patent Application No. 08/326456.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. USA* 70:321-3323.
- Paran, I. and R.W. Michelmore. 1993. Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.* 85:985-993.
- Pope, L.C., A. Sharp and C. Moritz. 1996. Population structure of the yellow-footed rock-wallaby *Petrogale xanthopus* (Gray, 1854. inferred from mtDNA sequences and microsatellite loci. *Mol. Ecol.* 5:629-640.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S.S. Tingey and J.A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2:225-238.
- Raymond, M. and F. Rousset. 1995. GENEPOP (V. 1.2): A population genetics software for exact tests and ecumenicism. *J. Hered.* 86:248-249.
- Riesner, D.G., U. Steger, M. Wiese, M. Wulfert, M. Heibey and K. Henco. 1992. Temperature-gradient electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction. *Electrophoresis* 13:632-636.
- Saghai-Marooof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang and R. Allard. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. *Proc. Nat. Acad. Sci. USA* 91:5466-5470.
- Sharma, S.K., M.R. Knox and T.H.N. Ellis. 1996. AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theor. Appl. Genet.* 93:751-758.
- Shattuck-Eidens, D.M., R.N. Bell, S.L. Neuhausen and T. Helentjaris. 1990. DNA sequence variation within maize and melon: observations from Polymerase Chain Reaction amplification and direct sequencing. *Genetics* 126:207-217
- Soltis, P.S., D.E. Soltis and C.J. Smiley. 1992. An rbcL sequence from a Miocene Taxodium (bald cypress). *Proc. Nat. Acad. Sci. USA* 89:449-451.
- Strand, A.E., J. Leebens-Mack and B.G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. *Mol. Ecol.* 6:113-118.
- Szmidt, A.E., Wang Xiao Ru and Lu Meng Zhu. 1996. Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* using haploid tissue analysis. *Heredity* 76:412-420.
- Swofford, D.L. and S.H. Olse. 1990. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Systematic Zool.* 39:411-451.
- Thompson, J.D., D.G. Higgins and J.T. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Tohme, J., D.O. Gonzalez, S. Beebe and M.C. Duque. 1996. AFLP analysis of gene pools of a wild bean core collection. *Crop Sci.* 36:1375-1384
- Tragoonrung, S., V. Kanazin, P.M. Hayes and T.K. Blake. 1992. Sequence-tagged-site facilitated PCR for barley genome mapping. *Theor. Appl. Genet.* 84:1002-1008.
- Vos, P., R. Hogers, M. Bleeker, M. Rijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23 :4407-4414.
- Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.
- Weising, K., H. Nybom, K. Wolff and W. Meyer. 1995. DNA Fingerprinting in Plants and Fungi. CRC Press, London.
- Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18:7213-7218.
-

- White, M.B., M. Carvalho, D. Derse, S.J. O'Brien and M. Dean. 1992. Detecting single base substitution heteroduplex polymorphisms. *Genomics* 12:301-306.
- Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.
- Withers, Lyndsey A. 1992. *In vitro* conservation. Pp. 169-200 in *Biotechnology of Perennial Fruit Crops* (F. Hammerschlag and R.E. Litz, eds.). CABI, UK.
- Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395-420.
- Wright, S. 1978. *Evolution and the genetics of populations. IV. Variability within and among natural populations.* Univ. Chicago Press, Chicago, USA.
- Zhang, Q., M.A. Saghai Maroof and R.W. Allard. 1990. Effects of adaptedness of variations in ribosomal DNA copy number in populations of wild barley (*Hordeum vulgare* ssp. *spontaneum*). *Proc. Nat. Acad. Sci. USA* 87:8741-8745.
-

Acronyms

AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrary Primed PCR
BIOSYS	A computer programme for analysis of allelic variation in population genetics and biochemical systematics (see Box 1)
CAPS	Cleaved Amplified Polymorphic Sequence
CLADOS	Parsimony programme for analysis of molecular data from K. Nixon (see Box 2)
CLUSTAL W	Alignment programme for analysis of molecular data from J.D. Thompson <i>et al.</i> 1994 (see Box 2)
cpDNA	Chloroplast DNA
DADA	Parsimony programme for analysis of molecular data from K. Nixon (see Box 2)
DAF	DNA Amplification Fingerprinting
DAMD	Minisatellite-region DNA
DELTAMU	Program for analysis of molecular data by D.B. Goldstein <i>et al.</i> 1995 (see Box 3)
DGGE	Denaturing Gradient Gel Electrophoresis
DNA	Deoxyribonucleic Acid
ETS	External Transcribed Spacer of the Ribosomal RNA gene
F-STAT	Computer programme to analyse F-statistics from J. Gondet 1995 (see Box 3)
GDA	Genetic Data Analysis: Computer programme to analyse discrete genetic data from P.O. Lewis and D. Zaytn 1996 (see Box 3)
GENEPOP	Population genetics software from M. Raymond and F. Ronsset 1995 (see Box 3)
HENNIG86	Parsimony programme for analysis of molecular data from J.S. Farris (see Box 2)
IGS	Intergenic spacer of the Ribosomal RNA gene
ISSR	Inter-simple sequence repeat amplification
ITS	Internal Transcribed Spacer of the Ribosomal RNA gene
MAAP	Multiple Arbitrary Amplicon Profiling
MACCLADE	Analysis of phylogeny and character evolution programme from W.P. Maddison and D.R. Maddison (see Box 2)
MALIGN	Alignment programme for analysis of molecular data from W. Wheeler and D. Gladstein (see Box 2)

MEGA	Molecular evolutionary genetic analysis from Kumar <i>et al.</i> 1993 (see Box 1)
MICROSAT	Program for analysis of molecular data by D.B. Goldstein <i>et al.</i> 1995 (see Box 3)
mtDNA	Mitochondrial DNA
NONA	Parsimony programme for analysis of molecular data from P.A. Goloboff (see Box 2)
NOR	Nucleolus Organizer Region
NTSYS-pc	Numerical taxonomy system from F.J. Rohlf (see Box 1)
PAUP	Phylogenetic analysis using parsimony programme from D. Swofford (see Box 2)
PCO	Principal Coordinate Analysis
PCR	Polymerase chain reaction
PCR-RFLP	Polymerase chain reaction combined with restriction fragment length polymorphism
PHYLIP	Phylogeny inference package from J. Felsenstein (see Box 1)
RAPD	Random Amplified Polymorphic DNA
RAPDFst	Program for analysis of RAPD data by B.L. Apostol <i>et al.</i> 1996 (see Box 3)
rDNA	Ribosomal DNA
RFLP	Restriction Fragment Length Polymorphism
RNA	“Rapid nucleotide analysis”. Parsimony programme for analysis of molecular data from J.S. Farris (see Box 2)
SAMPLE	A technique using microsatellite primers in conjunction with AFLP
SAS	General software package for statistical analysis
SCAR	Sequence characterized amplified region
SPAR	Single Primer Amplification Reaction
SSCP	Single Strand Conformational Polymorphism
SSR	Simple Sequence Repeat
STMS	Sequence-tagged microsatellites
STS	Sequence-tagged site
TGGE	Thermal Gradient Gel Electrophoresis
TREEVIEW	Parsimony programme for analysis of molecular data from R.D.M. Page (see Box 2)
UPGMA	Unweighted Pair Group Method with Arithmetic Averages
VNTR	Variable number of tandem repeat
WINAMOVA	Program for analysis of molecular data by L. Excoffier (see Box 3)
