Using molecular marker technology in studies on plant genetic diversity

M. Carmen de Vicente (IPGRI) and Theresa Fulton (Institute for Genomic Diversity, Cornell University)
Preface

Biodiversity use and conservation can be made easier and more efficient with the use of molecular markers. Phylogenetic relationships can be determined, redundancies in a germplasm bank can be identified, and new genes can be discovered. Although Using Molecular Marker Technology in Studies on Plant Genetic Diversity aims to promote capacity building and research in plant genetic resources worldwide, it is especially directed to those countries with limited access to up-to-date scientific literature and research technologies.

Using molecular markers is expensive, and limited financial resources must be used in the most judicious manner possible. Therefore, it is of critical importance that scientists do not use this technology simply because they can or because it is the latest technology, but because they thoughtfully chose the most appropriate technology for the biological questions they are tackling.

Our hope is that this online module will give you, its user, the context, knowledge, and tools to make the right decisions when using your resources.
Acknowledgements

The production of *Using Molecular Marker Technology in Studies on Plant Genetic Diversity* was made possible through a collaborative effort between International Plant Genetic Resources Institute (IPGRI) and Institute for Genomic Diversity (IGD of Cornell University). The authors especially thank the following contributors:

The IPGRI-Asia, the Pacific and Oceania Regional Office who, through its Tropical Fruit Trees Project, itself funded by the Asian Development Bank (ADB), partly funded the development of this material.

Félix Alberto Guzmán (IPGRI-Americas Regional Office), for his help in gathering information for the modules, especially in choosing good examples of application of technologies to real genetic diversity studies, and also for his judgement in making the modules clear and comprehensible.

Brian V. Ford-Lloyd (University of Birmingham, UK), who authored the earlier training module produced by IPGRI in 1996 (*Measuring genetic variation using molecular markers*), which helped in the preparation of this extended and updated module. Indeed, we have retained his text in appropriate places.

Pere Arús (Institut de Recerca i Tecnologia Agroalimentàries, IRTA, Spain), for his help with the slides for interpreting isozymes, and for his willingness to lend a series of pictures on laboratory procedures for detecting isozymes. These pictures were included in the respective submodule.

Andrzej Kilian (Centre for the Application of Molecular Biology to International Agriculture, CAMBIA, Australia), for agreeing to our using the schematic illustrations of DArT as they appear on CAMBIA's Web page.

Steve Tanksley (Cornell University, NY), for lending us slides on microsatellites from his teaching collection so we could use them in our modules; and Rebecca Nelson, also from Cornell, for sharing information on simplifying AFLP protocols.

Humberto Gómez Paniagua (Centro Internacional de Agricultura Tropical, CIAT-Colombia and International Crops Research Institute for the Semi-Arid Tropics, ICRISAT-Colombia), for sharing pictures of RAPD profiles for use as examples of good and bad results with this technology; and for giving us plenty of ideas on how to improve the didactic presentation of this product.

Helmer Ayala (Universidad de San Carlos, Guatemala), for the example of AFLP detection with silver staining; Xiaoming Pang (Guizhou University, China), for a picture of microsatellites with silver staining; and Kamel Chabane (International Center for Agricultural Research in Dry Areas, ICARDA, Syria) and Martin Fregene (CIAT, Colombia), for sharing pictures of the application of different molecular markers for use as background in different submodules.

The staff at the Genetic Resources Unit, CIAT (Colombia), for allowing us to take pictures of the isozyme procedures in their laboratory.

Professor Heiber Cárdenas (Biology Department, Universidad del Valle, Cali, Colombia), for testing the training module with five of his students (Iván Andrés González, Olga Lucia Agudelo Henao, Martha Nora Moyano, Fernando Rondón González and Diego Mauricio Villamarín Miranda), and for providing valuable comments for the module's improvement.

Luigi Guarino, Ramanatha Rao, Issiaka Zoungrana, Margarita Baena, Toby Hodgkin, Jan Engels and Elisabeth Goldberg (different IPGRI offices), for their comments and
suggestions on how to improve this product and hopefully make it more useful to our partners.

W. H. Freeman and Company/Worth Publishers (New York) for the permission granted to reproduce modified versions of Figures 4-3, 11-12, 12-3 (a), 12-6, 12-7 (c), 12-7 (d), 14-20, 14-24, 14-29 and 17-10 (a) from the book published by them in 1996 with the title *Introduction to Genetic Analysis* (written by Griffiths et al.).

Elizabeth L. McAdam for her help in manuscript editing and her good suggestions to improve the format of this product.
Foreword

Plant genetic resources are a key component in the sustainable development of agriculture and forestry. They contribute to development by helping to increase food production, eradicate poverty and protect the environment. The loss of genetic resources and, as a consequence, the genetic diversity they represent, is a widespread reality. It is therefore vitally important that we develop adequate and effective strategies to conserve these genetic resources. We need to build further on our knowledge of genetic diversity and introduce novel and powerful approaches that will eventually lead to a cost-effective identification of useful genes in germplasm. An effective use of genetic resources will be an important prerequisite for their sustainable conservation.

Molecular marker technologies are the most advanced and, possibly, the most effective means for understanding the basis of genetic diversity. They are efficient and accurate tools with which genetic variation can ultimately be identified and assessed in a rapid and thorough manner. By applying molecular technologies to approach the biological questions underlying the understanding of genetic diversity, we can make significant progress in the speed and depth at which we attain adequate and appropriate conservation and, thus, genetic resources made available for its use in crop improvement. However, the use of molecular markers is expensive and limited financial resources mean that they should be used in the most judicious way possible.

This set of training modules aims to facilitate capacity building in the use of molecular marker technologies, as it recognises that skilled human resources and institutions capable of “keeping pace with scientific progress” are key to the conservation of genetic resources. The authors are keenly aware of this need, and have therefore developed this set of training modules, of which Using Molecular Marker Technology in Studies on Plant Genetic Diversity is the first, to help build capacity in the use of molecular technologies. With this module, workers in plant genetic resources should be able to make better-informed decisions on the methods to use so they may more readily understand and, therefore, more effectively safeguard the genetic resources that underlie our very existence.

Jan Engels
Director
Genetic Resources Science and Technology Group
IPGRI
What you should know about the module

We feel strongly that when you study plant genetic diversity, you need to know what your goals are, what your limitations are, and what you must accomplish.

We have therefore taken care to discuss:

- The fundamental principles of genetic diversity,
- The qualities of the markers used to measure it and
- The most widely used technologies, including those based on proteins, DNA and the polymerase chain reaction.

Explanatory graphics and photographs illustrate key experimental procedures, and real-life examples are given of applications to particular cases of genetic diversity studies and/or germplasm management. These should help in the use of the module as a useful educational resource, whether as a self-tutorial or incorporated into a university curriculum.

We also compare the various techniques—their advantages and disadvantages, and relative costs of each procedure to help the beginning scientist understand the key components for selecting those procedures most appropriate for a given research.

Because this module was designed for use as a training aid or reference tool, lists of key references, references to extra applications and equipment lists are also given.

The module is intended for scientists with a minimal background in genetics and plant molecular biology, but with a working knowledge of plant genetic resources and issues concerning their conservation and management. We hope that the module will be particularly useful to scientists in developing countries, for whom print materials may be unavailable, expensive, or too quickly outdated. We also hope that it will be useful for science educators who wish to have access to a general overview of current DNA technologies and their possible uses in biodiversity conservation and use.

So that users may select only those sections of relevance or interest to them, we organised the module into complementary yet independent submodules. The exception is the Introduction, which is relevant to all sections and should always be included. In this way, the module, as a whole, can be used as reference for particular protocols, as a refresher or update for the scientist needing to make new research decisions or as a guide for short technical workshops.

Updating and feedback are of critical importance in the very fast evolving fields of molecular genetics (and their associated technologies) and plant genetic resources. We expect to update this product at relatively frequent intervals. To effectively respond to our partners and other users' needs, we would greatly appreciate your giving us feedback on the organisation, content and usefulness of this tool. You can write to us at cdevicente@cgiar.org; tf12@cornell.edu or at our mail addresses:

International Plant Genetic Resources Institute
Via dei Tre Denari 472/a
00057 Maccarese
Rome, Italy

Institute for Genomic Diversity
130 Biotechnology Building
Cornell University
Ithaca, NY 14583
We most wish that the module will be used in diverse ways and that it will, together with a companion module on the analysis of marker data for genetic diversity studies, expected to be out by the end of 2003, bring to many readers, especially those in developing countries with limited access to state-of-the-art technologies, a chance to conduct advanced research in plant genetic diversity, thereby contributing to the world’s knowledge of these valuable resources.

M. Carmen de Vicente
IPGRI

Theresa Fulton
IGD, Cornell University
Objectives of the module

Our goals for the module are that users:

• Understand the basic scientific concepts underlying molecular marker and DNA sequence technologies, and their use with reference to plant genetic resources

• Develop the ability to compare and contrast the advantages and limitations of each technology, and thus make the most appropriate decisions relevant to the users' specific research situations

• Have available a current list of bibliographical resources for each technology
Using molecular marker technology in studies on plant genetic diversity

Introduction
Contents

- Genetic diversity
- Plant genetic resources
- Measuring genetic variation
- Genetic markers:
  - Description
  - Types
  - Desirable properties
- Comparing major techniques:
  - Technologies
  - Costs
Genetic diversity refers to variation in the:
- DNA sequence,
- Amount of DNA per cell or
- Number and structure of chromosomes

Genetic diversity is the result of selection, mutation, migration, genetic drift and/or recombination. All these phenomena cause changes in gene and allele frequencies, leading to the evolution of populations.

Genetic diversity refers to the variation of genes within species, that is, the heritable variation within and between populations of organisms. In the end, all variation resides in the sequence of the four base pairs that compose the DNA molecule and, as such, constitute the genetic code. Other kinds of genetic diversity can also be identified at all levels of organisation in the nucleus, including the amount of DNA per cell, chromosome number and DNA structure.

The generation of new genetic variation occurs continuously in individuals through chromosomal and gene mutations, which, in organisms with sexual reproduction, are propagated by recombination. Genetic variation is also influenced by selection. The consequences of these phenomena are changes in gene and allele frequencies that account for the evolution of populations. Similar situations can occur through artificial selection such as breeding.
Plant genetic resources comprise the present genetic variation that is potentially useful for the future of humankind. These resources include traditional varieties, landraces, commercial cultivars, hybrids, and other plant materials developed through breeding; wild relatives of crop species; and others that could be used in the future for either agriculture or environmental benefits. Hence, plant genetic resources should be conserved, with the ultimate reason being to eventually use them as a source of potentially useful genetic variation.
Measuring genetic variation

Efficient conservation and use of plant genetic resources require thorough assessment of the genetic variation they comprise.

Genetic variation can be measured at two levels:

- Phenotype—the combination of individual traits resulting from a genotype and its interaction with the environment
- Genotype—the particular genetic make-up of an organism

To conserve and use genetic variation, it should first be assessed, that is, the extent and its distribution need to be determined. Variation can be evaluated on the phenotypic and genotypic levels. Assessment of phenotypic variation focuses on morphological traits—those characteristics that define the shape and appearance of a set of individuals. Some of these traits can be considered as 'genetic' if their presence in related individuals is heritable and not dependent on the environment, meaning that they are associated with a particular DNA sequence.

Assessment of genotypic variation is at the level of the DNA molecule responsible for transmitting genetic information. The DNA molecule is composed of nucleotides, which are organised in a double-helix configuration in increasing levels of complexity up to the chromosomal units.
Genetic markers: description

- Genetic markers identify characteristics of the phenotype and/or genotype of an individual
- Their inheritance can be followed through generations

A genetic marker is a measurable character that can detect variation in either a protein or DNA sequence. A difference, whether phenotypic or genotypic, may act as a genetic marker if it identifies characteristics of an individual’s genotype and/or phenotype, and if its inheritance can be followed through different generations.

A genetic trait may not have necessarily observable consequences on an individual’s performance. Sometimes, however, this trait may be linked to, or correlated with, other traits that are more difficult to measure and do affect the individual’s performance. In such cases, these unobservable genetic traits may be used as genetic markers for the linked traits because they indirectly indicate the presence of the characteristics of interest. The two measures can be correlated, using an analysis of inheritance and studying the distribution of the characteristics in both parents and offspring.
Genetic markers: types

- Morphological traits
- Protein (biochemical) markers
- DNA (molecular) markers
Traditionally, diversity within and between populations was determined by assessing differences in morphology. These measures have the advantage of being readily available, do not require sophisticated equipment and are the most direct measure of phenotype, thus they are available for immediate use, an important attribute. However, morphological determinations need to be taken by an expert in the species, they are subject to changes due to environmental factors and may vary at different developmental stages and their number is limited.
Protein (biochemical) markers

- Based on the migrational properties of proteins, which allow separation by electrophoresis
- Detected by specific histochemical assays
- Advantages:
  - Require relatively simple equipment
  - A robust complement to the morphological assessment of variation
- Disadvantages:
  - Subject to environmental influences
  - Limited in number

To overcome the limitations of morphological traits, other markers have been developed at both the protein level (phenotype) and the DNA level (genotype). Protein markers are usually named ‘biochemical markers’ but, more and more, they are mistakenly considered as a common class under the so-called ‘molecular markers’.

Protein markers (seed storage proteins and isozymes) are generated through electrophoresis, taking advantage of the migrational properties of proteins and enzymes, and revealed by histochemical stains specific to the enzymes being assayed.

Detecting polymorphisms—detectable differences at a given marker occurring among individuals—in protein markers is a technique that shares some of the advantages of using morphological ones. However, protein markers are also limited by being influenced by the environment and changes in different developmental stages. Even so, isozymes are a robust complement to the simple morphometric analysis of variation.
DNA (molecular) markers

- Polymorphisms detected in the DNA sequence of the nucleus and organelles

- Advantages:
  - Not subject to environmental influences
  - Potentially unlimited in number
  - Objective measure of variation

- Major disadvantage is the need for technically more complex equipment

DNA polymorphisms can be detected in nuclear and organellar DNA, which is found in mitochondria and chloroplasts. Molecular markers concern the DNA molecule itself and, as such, are considered to be objective measures of variation. They are not subject to environmental influences; tests can be carried out at any time during plant development; and, best of all, they have the potential of existing in unlimited numbers, covering the entire genome.

Many different types of molecular markers with different properties exist, as we show below.
Genetic markers: desirable properties

- Highly polymorphic
- Reproducible
- Codominant
- Evenly distributed throughout the genome
- Discriminating
- Not subject to environmental influences
- Neutral
- Inexpensive
- Easy to measure

A good marker is:

- **Polymorphic**, that is, it is variable among individuals. The degree of polymorphism detected depends on the technology used to measure it.

- **Reproducible in any laboratory experiment**, whether within experimental events in the same laboratory or between different laboratories performing identical experiments.

- **Codominant**. Depending on the type of application, the selected technology must be able to detect the marker’s different forms, distinguishing between homozygotes and heterozygotes (codominant inheritance). A heterozygous individual shows simultaneously the combined genotype of the two homozygous parents.

- **Evenly distributed throughout the genome**. The more distributed and dense genome coverage is, the better the assessment of polymorphism.

- **Discriminating**, that is, able to detect differences between closely related individuals.

- **Not subject to environmental influences**. The inference of a marker’s genotype should be independent of the environment in which the individual lives or its developmental stage.

- **Neutral**. The allele present at the marker locus is independent of, and has no effect on, the selection pressure exerted on the individual. This is usually an assumption, because no data are usually available to confirm or deny this property.

- **Inexpensive**. Easy, fast and cheap in detecting across numerous individuals. If possible, the equipment should be of multipurpose use in the experiment.
Comparing major techniques

<table>
<thead>
<tr>
<th>Markers</th>
<th>Number</th>
<th>Codominant</th>
<th>Polymorphism</th>
<th>Locus Specificity</th>
<th>Technicity</th>
<th>Cost</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isozymes</td>
<td>&lt; 90</td>
<td>Yes</td>
<td>Low</td>
<td>Yes</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>RFLP</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Medium</td>
<td>Yes</td>
<td>High</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>RAPD</td>
<td>Unlimited</td>
<td>No</td>
<td>Medium</td>
<td>No</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>DAF</td>
<td>Unlimited</td>
<td>No</td>
<td>Very high</td>
<td>No</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>AP-PCR</td>
<td>Unlimited</td>
<td>No</td>
<td>Very high</td>
<td>No</td>
<td>Low</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Microsatellites</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Very high</td>
<td>Yes</td>
<td>Low a</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>SCAR</td>
<td>Unlimited ×</td>
<td>Yes/no</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>CAPS</td>
<td>Unlimited ×</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Medium</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>ISSR</td>
<td>Unlimited</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Low/medium</td>
<td></td>
</tr>
<tr>
<td>AFLP</td>
<td>Unlimited</td>
<td>No</td>
<td>High</td>
<td>No</td>
<td>Medium</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Sequencing</td>
<td>Unlimited</td>
<td>Yes</td>
<td>High</td>
<td>Yes/no</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>EST</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Low/medium</td>
<td>Yes</td>
<td>Medium</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>Unlimited</td>
<td>Yes</td>
<td>Very high</td>
<td>Yes</td>
<td>High</td>
<td>High</td>
<td></td>
</tr>
</tbody>
</table>

a When microsatellites have already been identified and primers designed.
b Depending on other markers already available.

Here we compare major techniques that use biochemical and molecular markers to identify genetic diversity. Criteria used to assign levels within each column (yes/no, low/medium/high, etc.) are based on experience and results described in the literature. We cannot provide a number, or even a range, for each item and its technology because results are highly dependent on the species under study. However, the table gives an objective notion of how techniques can be compared among themselves for a given species and the items in the columns.

RFLP: Restriction fragment length polymorphism
RAPD: Random amplified polymorphic DNA
DAF: DNA amplification fingerprinting
AP-PCR: Arbitrarily primed polymerase chain reaction
SCAR: Sequence-characterised amplified region
CAPS: Cleaved amplified polymorphic sequence
ISSR: Inter-simple sequence repeat
AFLP: Amplified fragment length polymorphism
EST: Expressed sequence tag
SNP: Single nucleotide polymorphism
## Costs: how major technologies differ

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Item</th>
<th>Est. costs (USD, 2002), 96 samples</th>
<th>RFLP</th>
<th>SSR</th>
<th>RAPD</th>
<th>AFLP</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA extractions</td>
<td><a href="#">DNA extractions</a></td>
<td></td>
<td>5.40-101.40</td>
<td>5.40</td>
<td>5.40</td>
<td>5.40</td>
<td>96 micropreps vs. 96 large preps</td>
</tr>
<tr>
<td>Centrifuge and/or microcentrifuge tubes</td>
<td>2.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or, 96-well plate</td>
<td>3.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or, 50-ml tubes (large preps)</td>
<td>24.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extraction buffer</td>
<td>2.00-50.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96 micropreps vs. 96 large preps</td>
</tr>
<tr>
<td>Misc. supplies (alcohol, lysis buffer, tips, etc.)</td>
<td>1.00-25.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agarose electrophoresis</td>
<td><a href="#">Agarose electrophoresis</a></td>
<td>23.00 [23.00]</td>
<td>23.00</td>
<td>23.00</td>
<td>[23.00]</td>
<td>[23.00]</td>
<td>[ ] = optional, quality checking only</td>
</tr>
<tr>
<td>Agarose</td>
<td>10.00 (2 gels)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running buffer</td>
<td>13.00 (2 gels)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrylamide gel electrophoresis</td>
<td><a href="#">Acrylamide gel electrophoresis</a></td>
<td></td>
<td>1.50</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrylamide</td>
<td>0.75 (1 gel)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urea</td>
<td>0.75 (1 gel)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running buffer</td>
<td>-a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing gel electrophoresis</td>
<td><a href="#">Sequencing gel electrophoresis</a></td>
<td></td>
<td>4.06</td>
<td>4.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluorescent primers</td>
<td>2.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gel (see Acrylamide gel electrophoresis)</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethidium bromide method of visualisation</td>
<td><a href="#">Ethidium bromide method of visualisation</a></td>
<td>0.20-2.00</td>
<td>0.20-2.00</td>
<td>0.20-2.00</td>
<td></td>
<td></td>
<td>High-density paper vs. polaroid</td>
</tr>
<tr>
<td>Ethidium bromide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Photograph of gel</td>
<td>0.20-2.00 (2 gels)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* = negligible cost.
### Costs: how major technologies differ (continued)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Item</th>
<th>Est. costs (USD, 2002), 96 samples</th>
<th>RFLP(^a)</th>
<th>SSR</th>
<th>RAPD</th>
<th>AFLP</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Taq polymerase</td>
<td>50</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td>1.5 U/rxn. $$ will drop sharply when patent ends</td>
</tr>
<tr>
<td></td>
<td>dNTPs</td>
<td>24</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>primers</td>
<td>0.30-0.40</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR buffer with MgCl(_2)</td>
<td></td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td>74.30</td>
<td></td>
</tr>
<tr>
<td><strong>Restriction enzyme digests</strong></td>
<td>Enzymes</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>~10 units/sample, highly variable prices</td>
</tr>
<tr>
<td><strong>Southern blotting</strong></td>
<td>Buffers (HCl, NaOH)</td>
<td>4 (2 gels)</td>
<td>30.50</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>Cost is reduced with every reuse of the filter</td>
</tr>
<tr>
<td></td>
<td>Nylon membrane</td>
<td>24 (2 gels)</td>
<td>30.50</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Whatman paper</td>
<td>2.50 (2 gels)</td>
<td>30.50</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td>0.30-30.00</td>
<td></td>
</tr>
<tr>
<td><strong>Hybridisation</strong></td>
<td>Buffer</td>
<td>- (^b)</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>Hybridising one probe to 96 samples</td>
</tr>
<tr>
<td></td>
<td>ST DNA (or other blocker)</td>
<td>0.30</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LS labelling mix</td>
<td>0.30</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Radioisotope (^{32})P)</td>
<td>2.50</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miscellaneous reagents, tubes</td>
<td>0.25</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td><strong>TOTALS</strong> (per 96 samples)</td>
<td></td>
<td>62.75-190.25</td>
<td>85.16-110.26</td>
<td>104.40-106.20</td>
<td>83.76-106.76</td>
<td>83.76-106.76</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) = PCR not needed for the RFLP procedure. Probes can be PCR’d to save time, minimal cost

\(^b\) = negligible cost
### Costs: estimates of general costs

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Item</th>
<th>Cost (USD, 2002, est.)</th>
<th>Approx. cost per sample (USD, 2002, est.)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA extractions</td>
<td>Centrifuge and/or microcentrifuge tubes</td>
<td>$25/1000</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or, 96-well plate</td>
<td>$3.50</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extraction buffer</td>
<td>($2/litre)</td>
<td>0.02-0.50</td>
<td>microprep vs.large prep</td>
</tr>
<tr>
<td></td>
<td>Tris</td>
<td>$270/5 Kg</td>
<td>0.02-0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sorbitol</td>
<td>$70/5 Kg</td>
<td>0.02-0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Misc. (EDTA, etc.)</td>
<td>Approx. cost per sample (USD, 2002, est.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Misc. supplies (alcohol, lysis buffer, tips, etc.)</td>
<td>0.01-0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Drill and pestle (optional)</td>
<td>$100</td>
<td>0.01-0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genogrinder (optional)</td>
<td>$8000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leaf crusher (optional)</td>
<td>$500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agarose electrophoresis</td>
<td>Agarose</td>
<td>$365/500g</td>
<td>0.11</td>
<td>6 g/gel, ~40 samples/gel</td>
</tr>
<tr>
<td></td>
<td>Running buffer (Tris, EDTA, NaAc)</td>
<td>$6.50/litre</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Horizontal gel system</td>
<td>$400</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power supply</td>
<td>$400 (2 outlets)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acrylamide gel electrophoresis</td>
<td>Acrylamide</td>
<td>$30/100 ml</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urea</td>
<td>$42/500 g</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vertical gel system</td>
<td>$1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Power supply</td>
<td>$400 (2 outlets)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gel dryer</td>
<td>$1500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequencing gel electrophoresis</td>
<td>Fluorescent primers</td>
<td>$40/1500 samples</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gel</td>
<td>$3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Size standards</td>
<td>$12/gel</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Software for gel analysis</td>
<td>&lt;$100,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sequencing gel equipment</td>
<td>$100,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The costs presented here assume that a basic laboratory is available and set up with glassware, plasticware, heater/stirrers, pH meter, balance, refrigerator, freezer, distilled water, pipettes and tips, centrifuge.

These costs vary between countries, and have been calculated according to U.S. prices. Because costs in many countries will be higher, the ones shown here should be taken as indicative and used to compare different technologies and alternative equipment.

Continued on slide 16
<table>
<thead>
<tr>
<th>Procedure</th>
<th>Item</th>
<th>Cost (USD, 2002, est.)</th>
<th>Approx. cost per sample (USD, 2002, est.)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethidium bromide method of visualisation</td>
<td>Ethidium bromide</td>
<td>$100/10 g</td>
<td>- a</td>
<td>Can be reused, to last a long time</td>
</tr>
<tr>
<td></td>
<td>Transilluminator</td>
<td>$1000-$2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Photography equipment</td>
<td>$3000-12,000</td>
<td></td>
<td>Highly variable system types</td>
</tr>
<tr>
<td></td>
<td>Photograph of gel</td>
<td>$0.10-$1.00</td>
<td>0.0025-0.025</td>
<td>High-density paper vs. Polaroid; ~40 samples/gel</td>
</tr>
<tr>
<td>PCR</td>
<td>Thermocycler</td>
<td>$7000-$23,000</td>
<td></td>
<td>96-well - tetrad (four 96-well plates)</td>
</tr>
<tr>
<td></td>
<td>Taq polymerase</td>
<td>$170/500 units</td>
<td>0.51</td>
<td>1.5 U/rxn. $$ will drop sharply when patent ends</td>
</tr>
<tr>
<td></td>
<td>dNTPs</td>
<td>$250/set of 4</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primers</td>
<td>$15-$25</td>
<td>0.0025-0.0042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PCR buffer with MgCl₂</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restriction enzyme digests</td>
<td>Enzymes</td>
<td>$0.003-$0.30/unit</td>
<td>0.03-3.00</td>
<td>~10 units/sample, highly variable prices</td>
</tr>
<tr>
<td></td>
<td>Incubator</td>
<td>$500-$1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern blotting</td>
<td>Buffers (HCl, NaOH)</td>
<td>$1.00/litre (avg.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nylon membrane</td>
<td>$180/roll</td>
<td>6.00/filter (20x10cm)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Whatman paper</td>
<td>$124/pack 100 large sheets</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blotting system (sponges, tray)</td>
<td>$15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hybridisation</td>
<td>Buffer</td>
<td>$306/10 g</td>
<td>0.30</td>
<td>0.15 ml. 1 ml used/50 ml hybrid’n buffer</td>
</tr>
<tr>
<td></td>
<td>ST DNA (or other blocker)</td>
<td>$125/50 unit</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shaking incubator</td>
<td>$2500</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Radioisotope (³²P)</td>
<td>$120/100 µl</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Misc. reagents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Environmental expenses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Waste removal</td>
<td>$500/month</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shielding</td>
<td>$300/station</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a - = negligible cost
In summary

- Defining strategies for good conservation and use requires assessing the variation found in genetic resources.

- Plant genetic diversity can be measured through genetic markers—morphological, biochemical and molecular.

- No single marker meets all the desired properties.

- Choice of technique depends on the nature of the biological question being addressed.
By now you should know

- What genetic variation is and how it can be measured
- The main advantages and disadvantages of the different types of genetic markers
- What constitutes the desirable properties of genetic markers
Basic references


Protein-based technologies
Protein basics

- Protein-based technologies
  - Isozymes
- DNA-based technologies
- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

II  Protein-based technologies
    Protein basics
    Isozymes
Using molecular marker technology in studies on plant genetic diversity

Protein-based technologies
Protein basics
Contents

› Protein basics

› Protein structures:
  • Primary
  • Secondary
  • Tertiary
  • Quaternary

› Protein functions

› Enzymes:
  • Description
  • Allozymes and isozymes
The information carried by the DNA bases translates into proteins. The DNA molecule is copied into a different type of nucleic acid—the RNA or ribonucleic acid. The RNA moves to the ‘ribosome’, an organelle in charge of making proteins. Every set of three bases in the RNA determines which amino acid is added to the protein molecule in progress. The RNA chain passes through the ribosome until the protein is complete.

This process has been called ‘the central dogma’, as it is the basis of all biological life. Amino acids are bi-functional organic compounds that contain a basic amino group (-NH₂) and an acidic carboxyl group (-COOH). In proteins, 20 different amino acids are commonly found, varying in function and property according to the nature of the R-group. For instance, in alanine, the R-group is (-CH₃), whereas, in cysteine, it is (-CH₂-SH).
In proteins, amino acids are joined together in chains by peptide (amide) bonds that form the molecule’s backbone. A peptide bond is formed between a basic amino group (-NH$_2$) on one amino acid and an acidic carboxyl group (-COOH) on another. The general formula of an amino acid is H$_2$N –CHR –COOH. The R group can be anything from an atom to a complex molecule. The term ‘polypeptide’ simply refers to a long chain of amino acids.

Once the protein chain has been made, it must fold up properly before it can do its job. The structure and folding of each protein is specific. How the amino acid sequence causes the protein to fold is not yet completely understood. A ‘protein’ can be made up of one or more separate polypeptides; it can be made of sheets (amino acid chains lining up together) that contain spiral structures. Clearly, the properties of polypeptides and proteins depend on their amino acid composition.
The secondary structure is the result of local hydrogen bonds being created along the polypeptide backbone. Common structures found are:

- **Alpha-helix**, caused by hydrogen-bonding within the polypeptide chain, for example, muscle proteins.

- **Beta-pleated sheet**, caused by hydrogen-bonding between adjacent polypeptide chains, for example, silk fibroin.
Protein structures: tertiary

The tertiary structure results from interactions between the R-groups in a polypeptide; the non-covalent and covalent bonds.

The tertiary structure results from interactions between the R-groups in a polypeptide, such as non-covalent bonds (hydrogen bonds, ionic bonds, hydrophobic interactions) and weak, covalent bonds (disulphide bonds between cysteine residues).
Protein structures: quaternary

The quaternary structure results from interactions between two or more polypeptide chains to form dimers, trimers, tetramers, etc.

Sometimes a single polypeptide is sufficient for the protein to be active; we then talk of a protein that acts as a monomer. Often, however, two or more polypeptides need to interact to allow a protein to perform its particular function. If this is the case, we talk of a dimer; and so on through trimers, etc.

The quaternary structure results from interactions between two or more polypeptide chains to form dimers, trimers, tetramers, etc. They are held together by hydrogen bonds, ionic bonds and, less commonly, hydrophobic interfaces and inter-chain disulphide bonds.
Proteins functions

- The three-dimensional structure of proteins is a direct result of interactions with their internal environment
- Diversity of protein function is a result of the complexity of protein structure

The three-dimensional structure that proteins have is a direct result of interactions with its internal environment. As a consequence, knowing how proteins are structured tells us a lot about how they perform their tasks in the cell.

For instance, in aqueous environments, the hydrophobic R-groups are positioned towards the protein's interior. Changes in temperature or pH can interfere with the non-covalent bonding, causing disruption in the three-dimensional structure and loss of activity. This process is called ‘denaturation’. Denatured proteins can also clump together to become insoluble in a process called coagulation.

The diversity of protein function, as facilitated by the complexities of protein structure, include (with examples):

- **Structural** (collagen, muscle fibres)
- **Storage** (wheat gliadins, barley hordeins)
- **Enzymes** (hydrolases, transferases, isomerases, polymerases, ligases)
- **Transport** (oxygen transfer with haemoglobin)
- **Messengers** (insulin and certain other hormones)
- **Antibodies** (proteins that bind to specific foreign particles)
- **Regulation** (proteins involved in regulating DNA synthesis)
Enzymes: description

- Enzymes are a particular type of protein that act as catalysts

- Each enzyme is highly specific with regard to the type of chemical reaction that it will catalyse and to the substances (called substrates) on which it will act
Enzymes: allozymes and isozymes

The multiple forms that enzymes take fall into two main classes according to how they are coded:

- **Allozymes**—enzymes coded by different alleles at one gene locus
- **Isozymes**—enzymes coded by alleles at more than one gene locus

Usually, the term ‘isozymes’ refers to both classes.

Polymorphisms are generated by changes in the amino acids that may result in changes in the primary structure of the enzyme.
**In summary**

- Proteins are the primary product of genes.
- Proteins may have different structures, which are closely related to the tasks they perform in the cell.
- Enzymes are a particular type of protein that act as catalysts.
- Enzymes may have multiple forms, belonging to either allozymes or isozymes.
By now you should know

- What a protein is
- The mechanisms involved in shaping proteins into different structures
- What an enzyme is
- The difference between allozymes and isozymes
Basic reference

Protein-based technologies

Isozymes

- DNA-based technologies
- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

Protein-based technologies
Isozymes
Contents

- Detecting isozymes
  - Methodology
  - Gel electrophoresis
- Equipment
- Isozymes in pictures
- Interpreting banding patterns
- Advantages and disadvantages
- Applications
  - *Lysimachia* sp.
  - Lima bean
  - Onion
Detecting isozymes: methodology

- Pre-treating plant material
- Starch or acrylamide gel electrophoresis
- Histochemical staining
- Analysing banding patterns

Crude tissue extracts are subjected to gel electrophoresis and probed with enzyme-specific stains, resulting in a simple banding pattern.

Variation in banding patterns between individuals can be interpreted genetically, as would be done with any other phenotypic marker.
Detecting isozymes: gel electrophoresis (1)

- This process combines the separation of molecules by charge with that by size by applying an electric current.

- The current makes the molecules move through pores in a layer of gel.

- The substance that makes up the gel is selected so its pores are of the right size to separate a specific range of molecule sizes and shapes.

Electrophoresis is a chromatographic technique for separating mixtures of ionic compounds. It has been adapted as a common tool for biochemical analysis.
Detecting isozymes: gel electrophoresis (2)

Proteins are the primary product of genes. When the nucleotide sequence of the DNA changes, so too do the proteins’ banding patterns. Enzyme electrophoresis can directly reveal genetic polymorphism through demonstrating the multiple forms of a specific enzyme.
Equipment

Resources:
- Distilled and/or deionised water
- Reagents

Equipment:
- Refrigerator and freezer
- Power supply units
- Hotplate or microwave
- Thick cotton gloves
- pH meter
- Balances
- Gel units
- Suction and volumetric flasks
Isozymes in pictures

The following photographs illustrate the laboratory steps involved in detecting isozymes
The laboratory technician is pipetting extraction buffer onto a section of leaf. Each sample is placed in a plastic dish that is typically used to weigh chemicals. Other utensils such as glass plates can be used instead, but they must be very clean and unlikely to absorb macerated tissue.
Once the extraction buffer is placed on the leaf sample, the leaf is crushed with a plastic rod to ensure the tissue is well broken and homogenized with the buffer. This must be done as quickly as possible to prevent an increase in temperature. While the samples are being crushed, their plastic holders can rest on ice.
A small paper wick, cut from porous paper, is left on the homogenate, until the paper absorbs the sample.
The heated starch solution is being poured into the gel mould. As the gel cools, it will become jelly-like. It will then be ready for being loaded with samples. Until used, the gel can be stored in a refrigerator.
A starch gel, set in advance, has been cut with a scalpel at about one third up the mould. Paper wicks are now being vertically placed against one of the cut sides with the help of tweezers.
A power supply machine must be used to control the voltage and intensity of the current for electrophoresis. Many models are commercially available, and have comparable properties and capacities. In our example, two electrophoresis units can be simultaneously plugged in. A timer may also be included to halt electrophoresis at a given time.
The gel is placed on an electrophoresis unit inside a cold room or refrigerator. The unit has a buffer container at each side (cathode and anode) and contact between the buffer and gel is achieved with a sponge or piece of mesh cloth. A plastic wrap is placed on top to prevent drying out during the process and ensure good buffer transfer. Samples are loaded and their trace followed, the run being marked by a blue colorant.
When the run is complete (usually a brownish shadow can be seen opposite the loading side of the gel), the gel must be cut into slices. The foreground shows the gel ready for cutting, with a plexiglass stand waiting for the gel's transfer (centre). A container where the slices will be placed can be seen at the photograph’s top edge.
The gel is now on the stand and thin black guides are placed at each side to control the thickness of the slices. A notch is made at one corner (top right) as guide to the samples’ correct location after the gel is stained.
A glass plate is placed over the gel so that pressure can be exerted while cutting, ensuring that the slices are even. A handsaw, fitted with a guitar string, is being used to cut the gel into slices.
Different staining solutions let us visualize different enzymes. Here, a yellowish staining solution is poured into a plexiglass container before a gel slice is transferred.
A slice of gel is being transferred to the container with the desired staining solution. Care must be taken on transferring the slices because their fragility. If a slice does break, the broken pieces can often be rearranged and the enzymes still visualized on the gel.
After a certain period of incubation, a stained gel slice appears similar to the photograph, which shows leaf alcohol dehydrogenase (ADH) of sainfoin (*Onobrychis viciaefolia*), a forage crop.
Interpreting banding patterns

The main issues are:

- The quaternary structure of enzymes (whether monomeric, dimeric, etc.)
- Whether the plant is homozygous or heterozygous at each gene locus
- The number of gene loci (isozymes)
- The number of alleles per locus
- How the genes are inherited

Allozymes are controlled by codominant alleles, which means that homozygotes (all alleles at a locus are similar) can be distinguished from heterozygotes (parents of the individual have contributed different alleles to that locus).

For monomeric enzymes (i.e. consisting of a single polypeptide), plants that are homozygous for a given locus will produce one band, whereas heterozygous individuals will produce two. For dimeric enzymes (i.e. consisting of two polypeptides), plants that are homozygous for that locus will produce one band, whereas heterozygous individuals will produce three because of random association of the polypeptides.

Multimeric enzymes also exist, where the polypeptides are specified by different loci. The formation of isozymic heteromers can thus considerably complicate banding patterns.

These complexities and the importance of correctly interpreting banding patterns, make genetic analysis desirable, even necessary, using progeny analysis (F₁, F₂ and back-cross) of artificial crosses between individuals with known banding patterns.
This example shows the behaviour of monomeric enzymes in crosses. Parents are homozygous for the same allele at the same locus or heterozygotes. In the first case, all F<sub>1</sub> progeny will be homozygous and, as a consequence, a single band will result. If parents are heterozygous, three possible F<sub>1</sub> phenotypes will result.
In crosses, dimeric enzymes may behave in three possible ways. If both parents are homozygous for different alleles at the same locus, all progeny are heterozygous, but because of random association of the polypeptides, three combinations are possible and therefore three bands will be resolved on gel electrophoresis and staining.
Example 3: Forming heteromers (continued)

Dimeric allozyme: two genes with two alleles

In the second example on dimeric allozymes, the enzyme’s two polypeptides are coded by separate loci. Parents do not share any alleles. Random association of the polypeptides can lead to a total of 10 dimers, because both intragenic dimers (alleles at the same locus) and intergenic dimers (alleles from different loci) are formed.
Example 4: Forming heteromers (continued)

Heterozygous tetraploid (one locus, four alleles)

- Monomers

- Dimeric active enzymes

In the example above, a heterozygous tetraploid, with four alleles at one locus, forms ten dimeric active enzymes due to random association among the four monomers.
Advantages and disadvantages

Advantages

- Robust and highly reproducible
- Codominant, i.e. suitable for estimating a wide range of population genetics parameters and for genetic mapping

Disadvantages

- Relatively few biochemical assays available to detect enzymes
- Phenotype-based analysis

Isozyme analysis is, in principle, a robust and reproducible method. In addition, isozymes are codominant markers and are suitable for estimating all population genetics parameters and for genetic mapping.

About 90 isozyme systems have been used for plants, with isozyme loci being mapped in many cases.

The major limitation of isozyme analysis is the low number of markers it provides, because the number of biochemical assays available to detect them is small. Consequently, the percentage of genome coverage is inadequate for a thorough study of genetic diversity.

Another disadvantage of isozyme analysis lies in the markers being based on phenotype. As such, they may be influenced by environmental factors, with differences in expression confounding the interpretation of results. Because differential expression of the genes may occur at different developmental stages or in different tissues, the same type of material must be used for all experiments.
Applications

- Gene flow and/or introgression
- Genetics of populations
- Strategies for *ex situ* conservation
- Crop evolution
- Germplasm evaluation and characterization
- Genetic erosion
- Genetic stability of conserved material

References in purple are explained in detail in the following slides.


Example: *Lysimachia sp.*

- **Title:**
  Mol. Ecol. 1999. 8: 813-817

- **Objective:**
  To evaluate the genetic diversity of seed accessions of *Lysimachia minoricensis* conserved and provided by 10 European botanical gardens

- **Materials and methods:**
  A total of 158 plants were analysed for 13 enzymes (22 loci)

(continued on next slide)


Example: *Lysimachia sp.* (continued)

- **Results:**
  
  No electrophoretic variation was detected for any of the enzymes assayed (22 loci)

- **Discussion on lack of variation:**
  
  - Electrophoretic techniques resolve only a small portion of genetic variation
  
  - Low sample size? The sample offers a 95.6% probability of detecting any variant allele that existed at an overall frequency of at least 1%
  
  - Is this lack of detectable variation unexpected? Results address more the question of the amount of genetic variation preserved *ex situ* than what the level was before extinction

   (continued on next slide)
Example: *Lysimachia sp.* (continued)

Conclusions:
The exchange system among botanical gardens is not adequate for effective conservation if the genetic variation within a species is underrepresented in the plant collections.
Example: Lima bean

Title:

Objective:
To evaluate genetic diversity and structure within a lima bean (*Phaseolus lunatus* L.) base collection, involving several, widely distributed, wild accessions and landraces

Materials and methods:
Ten enzyme systems were used to analyse 235 lima bean accessions (1-5 seeds each) collected in Latin America and the Caribbean

(continued on next slide)
Example: Lima bean (continued)

Results and discussion:

• Thirteen loci (32 alleles) were found for 10 enzyme systems
• Specific alleles were identified in each gene pool. The dendrogram clearly showed two main clusters:
  • Accessions from the Andes with an Andean seed-protein pattern
  • Mesoamerican and Andean accessions with a Mesoamerican seed-protein pattern
• Both Andean and Mesoamerican landraces were grouped with their respective relatives
• On average, the lima bean showed 76% and 24% of the total diversity, respectively, among and within accessions. Reasons are selfing, occurrence of small populations, and low gene flow

(continued on next slide)
Example: Lima bean (continued)

Conclusions:

• A conservation programme of *P. lunatus* must include wild and cultivated forms from both gene pools
• As the genetic diversity is distributed mainly among accessions, more populations or accessions should be preserved to ensure the retention of allelic and genotypic diversity for both gene pools and botanical forms
Example: Onion

- **Title:**

- **Objective:**
  To evaluate genetic diversity in local populations of onion, using isozymes

- **Materials and methods:**
  - Sixteen local cultivated populations sampled from five countries of West Africa were studied
  - Nine enzyme systems were assayed

(continued on next slide)
Example: Onion (continued)

Results:

• Four systems were polymorphic (ADH, MDH, 6-PGDH, PGI), with a total of nine alleles
• The mean number of alleles found per polymorphic locus was 2.25. In addition, 66.6% of alleles were present in all populations
• Allele adh-a2 was absent in the five populations originating from Burkina Faso. Allele 6-pgdh-a2 was present only in two populations, one from southern Niger and the other from northern Nigeria

(continued on next slide)
Example: Onion (continued)

- **Discussion:**
  - Weak geographical differentiation between populations may have resulted from commercial exchange between countries of the same language.
  - Overall, heterozygotes were few, maybe because of autogamy in the populations sampled or genetic drift resulting from seed production involving tiny quantities of bulbs as progenitors.

- **Conclusions:**
  Similar studies on a continental scale over Africa would be advisable to better understand the species’s variability.
In summary

- Isozyme technology is based on protein extraction, separation by gel electrophoresis and histo-chemical staining
- The equipment needed is simple
- The interpretation of banding patterns is typically codominant and, because of its complexity, may require genetic analysis
- Isozymes as genetic markers are highly reproducible
By now you should know

- Main steps involved in isozyme technology
- Main points for consideration when interpreting banding patterns
- Advantages and disadvantages of isozymes as genetic markers for diversity studies
Basic references


DNA-based technologies
DNA basics

- DNA-based technologies
  - Restriction fragment length polymorphisms (RFLPs)
  - PCR-based technologies

- Complementary technologies

- Final considerations

- Glossary
Using molecular marker technology in studies on plant genetic diversity

III DNA-based technologies
- DNA basics
- Restriction fragment length polymorphisms (RFLPs)
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies
DNA basics
Contents

The DNA molecule:
• Structure and features
• Replication
• Squeezing into the chromosome
• Sequence organization
• Cytoplasmic DNA

DNA technology
• Restriction enzymes
• Nucleic acid electrophoresis
• DNA polymorphism

DNA isolation procedures in pictures
The DNA molecule: structure and features

DNA and RNA are molecules made up of strings of nucleotides

A nucleotide consists of:
- A pentose sugar
- A phosphate group
- A nitrogenous base

The building blocks of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are nucleotides. A nucleotide consists of:

- A pentose sugar; in DNA, it is deoxyribose, and, in RNA, it is ribose
- A phosphate group
- A nitrogenous base, which can be a:
  - Purine base—adenine (A), guanine (G).
  - Pyrimidine base—cytosine (C), thymine (T). In RNA, uracil (U) replaces thymine

The DNA molecule comprises a chain built from four simple building blocks (A, G, C and T) that are assembled to form a double helix. The helix consists of two strands, each with a sugar-phosphate backbone, held together by a weak hydrogen bond between the bases adenine-thymine (two hydrogen bonds) and cytosine-guanine (three hydrogen bonds).

The shapes of A and T, and of C and G are ‘complementary’ and form the reason why DNA may copy itself. Two chains of backbones and bases running in opposite directions (antiparallel) form the double helical structure. The order or ‘sequence’ of these bases along the chain forms the genetic code that carries the precise genetic instructions for the organism to function.
Under certain circumstances (i.e. during cellular DNA replication), the two chains of the DNA molecule separate. The RNA polymerase synthesizes a short stretch of RNA complementary to one of the DNA strands at a particular site, known as the replication start site.

This short section of RNA acts as a primer for the DNA replication to start. New DNA bases come in at the 3' end and adhere to their complementary pair on the template DNA strand. The new bases are then adjoined to make a ‘daughter’ DNA chain. As nucleotides are always added at the 3' end, DNA synthesis occurs from a 5' to 3' direction. This process occurs for each of the original chains of the parent DNA molecule.
A DNA molecule is much longer than a chromosome, so a mechanism is needed to densely fold and pack the DNA fibre.

The mixture of material of which chromosomes are formed is called chromatin, and is the sum of the DNA molecule plus some proteins. In eukaryotes, DNA is condensed with histone and non-histone proteins, and some RNA. Histones are organized into nucleosomes and give the coiled DNA a bead-necklace appearance. Additional coiling of nucleosomes results in a solenoid conformation, with another level of packaging necessary to arrive at the chromosome structure.

Chromosomes are made up of euchromatic regions, lightly packed and containing most of the active genes, and heterochromatic regions, densely packed and apparently inactivating genes by surrounding them. Heterochromatin is often found around centromeres.

The definition of angstrom is modified from Merriam-Webster Online (http://www.m-w.com)
Eukaryotic DNA may be grouped in different types or classes:

- Single-copy, protein-coding genes

- DNA present in multiple copies:
  - Sequences with known function
    - Coding
    - Non-coding
  - Sequences with unknown function
    - Repeats (dispersed or in tandem)
    - Transposons

- Spacer DNA
  Numerous repeats can be found in spacer DNA. They consist of the same sequence found at many locations, especially at centromeres and telomeres. Repeats vary in size, number and distribution throughout the genome, making them highly suitable for consideration as molecular markers.

Reference:

Smaller amounts of DNA are found in the cytoplasm outside the nucleus—in the chloroplasts (cpDNA) and mitochondria (mtDNA). Chloroplasts and mitochondria each have their own unique ‘chromosome’, with several copies. These genes also code for their own translation and transcription of organellar components, and play highly specialized roles in the expression of the phenotype of the organism to which they belong.

Organellar DNA is commonly, but not always, inherited only through the maternal parent, a pattern known as maternal inheritance.

The DNA sequences of cpDNA and mtDNA have their own peculiarities. Plant mtDNA appears to evolve rapidly with respect to gene order, but slowly in nucleotide sequence. Why the accumulation of mutations is slow is not properly understood, but may be a result of the presence of either a highly efficient DNA damage repair mechanism or a relatively error-free DNA replication system. Conversely, the rate of cpDNA evolution usually appears slow, in terms of both primary nucleotide sequence and gene rearrangement.
DNA technology

DNA technology involves the concepts of:

- Restriction enzymes
- Nucleic acid electrophoresis
- DNA polymorphism
Restriction enzymes

- Each restriction enzyme cuts the DNA into defined fragments by acting at specific target sequences
- They form either sticky or blunt ends
- Types of restriction enzymes:
  - Types I and III cut double-stranded DNA outside the target sequence
  - Type II identify 4, 5 or 6 bp sequences and cuts inside the sequence

Bacteria produce restriction enzymes as a defence mechanism against bacteriophages. These enzymes belong to a class that cleave (or cut) DNA at specific and unique internal locations along its length. As a consequence, they are also called endonucleases. These enzymes act as scissors, cutting the DNA of the phages and inactivating them.

Of the three types of restriction enzymes, types I and III cut the double-stranded DNA outside the target sequence. In contrast, type II restriction enzymes identify specific sequences of 4, 5 or 6 base pairs and cut inside this sequence. Because of their features, all three types have become essential for recombinant DNA technology.

Enzymes may cut a given DNA sequence, leaving staggered (or ‘sticky’) ends that allow hydrogen-bonding to a complementary sequence or blunt ends. If two fragments of DNA are cut with the same enzyme, fragments with the same complementary sticky ends will be produced and alternative fragments may attach.

Restriction enzymes are commercially available, usually furnished with the appropriate reaction buffer and information about reaction conditions and temperatures.
Nucleic acid electrophoresis

A method to separate DNA fragments to allow their visualization and/or identification

After digestion with a restriction enzyme, the DNA molecule is converted into a collection of restriction fragments. These fragments may be separated by size by running them through an agarose or acrylamide gel.

To obtain the separation, the mixture of DNA fragments and leftover restriction enzyme is placed in wells formed at one edge of the gel. The gel is then subjected to an electrical field, forcing the migration of DNA fragments according to their size, with large fragments migrating more slowly than short fragments. DNA molecules, negatively charged at neutral pH, migrate towards the anode. Agarose gels (0.8% to 2.0% agarose) are most useful for separating DNA fragments that range in size from 300 to 10,000 bp. Acrylamide gels (3.5% to 20% acrylamide) are most useful for fragments ranging between 20 and 1000 bp in size.

Visualisation of DNA fragments after electrophoresis is achieved by staining with ethidium bromide, a molecule that moves into the bases of the DNA and can fluoresce an orange colour under UV light.

Migration distance is proportional to the logarithm of the number of bases. The actual size of the fragments obtained can therefore be calculated in relation to the mobility of DNA fragments of known size.
DNA polymorphism

- Various events may give rise to variants, more or less complex, in the DNA sequence. Such variants are usually described as polymorphisms.

- Polymorphism is translated into differences in genotype—as evidenced in diverse band profiles when detected with an appropriate procedure—and perhaps phenotype.

- Several events can produce polymorphisms:
  - Point mutations
  - Insertions or deletions
  - Rearrangements
Point mutations occur when a base in the DNA sequence is replaced by another. The length of the DNA sequence does not change.

Point mutations can occur in one base only or in a few bases at the same location. In the diagram above, four bases of the original chromosome sequence (top) are replaced by four alternative bases. Because the original number of the bases does not change, the sequence’s total length does not alter.
Insertions or deletions are the addition or the disappearance of several bases in the DNA sequence. The molecule length changes.

The top half of the diagram illustrates a deletion: some bases are lost and the resulting DNA fragment becomes shorter.

The bottom half of the diagram illustrates an insertion: some bases are introduced into a section of a DNA sequence. The original sequence thus becomes longer according to the number of bases being inserted.
Rearrangements

Chromosomal rearrangements occur through genetic recombination or insertion of transposable elements. The molecule length may or may not change.

Changes in the sequence of the DNA may also occur through rearrangements, such as a segment flipping over. In these instances, although the length of the DNA sequence may not change, its composition could change sufficiently for it to be observed as a polymorphism.
DNA isolation procedure in pictures

The following photographs illustrate various steps of the procedures for isolating DNA.
The laboratory technician is harvesting a few, very young, tomato leaves for a microprep extraction. For a large prep DNA extraction, many more, larger leaves (about 10 g) would be harvested from much older plants.
Leaf tissue (for microprep DNA extractions) and buffer are homogenized in a 1.5-ml microcentrifuge tube, using a drill fitted with a plastic pestle. To increase efficiency, two drills may be used simultaneously. These can be operated by foot pedals, like those used for sewing machines.
Leaf tissue for large prep DNA extractions is homogenized with DNA extraction buffer in standard kitchen blenders. Although not needed for safety, gloves and a laboratory apron may be worn to protect clothing and skin, as the procedure can be messy.
The mixture of leaf tissue and buffer is poured from the blender, through cheesecloth, into centrifuge bottles packed in ice. The cheesecloth is squeezed to get as much liquid as possible while filtering out large pieces of leaf tissue, which are then discarded, together with the filter.
After centrifuging and re-suspending the DNA pellet (which is still green and contains some leaf material), the mixture is transferred to a new tube and chloroform is added. This step should be performed in a ventilation hood, and safety gloves and a laboratory coat worn.
After inverting the tubes to gently mix in the chloroform, the tubes are centrifuged to separate out the DNA.
The lighter layer containing the DNA is now on the top, and can be transferred into a clean tube. The lower layer contains unwanted leaf tissue, cell walls and other residues, and is discarded.
Alcohol is added to precipitate out the DNA, which, on gentle inversion, usually comes together as a string-like substance. The DNA can then be either removed with a hook or spun down. It is then washed and re-suspended.
In summary

- The bricks of DNA are nucleotides, assembled in a double helix to form the DNA chain.
- The DNA molecule becomes increasingly more folded as the nucleotide sequence forms the chromosome.
- DNA is organized in different classes: single copy, multiple copy and spacer.
- Mitochondria and chloroplasts have their own DNA molecule.
- Several events give rise to polymorphisms in the DNA molecule: point mutations, insertions, deletions and rearrangements.
By now you should know

- The components of a nucleotide
- How nucleotides form the DNA double helix
- The different classes of DNA
- The main features of cytoplasmic DNA
- What a restriction enzyme is
- How polymorphisms in the DNA chain are produced
Basic references


Next

DNA-based technologies
Restriction fragment length polymorphisms (RFLPs)

- DNA-based technologies
  - PCR-based technologies

- Complementary technologies

- Final considerations

- Glossary
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies
Restriction fragment length polymorphisms
Contents

- RFLP technology
  - Isolating DNA
  - Restriction digestion and gel electrophoresis
  - DNA transfer by Southern blotting
  - DNA hybridisation
  - Equipment

- RFLP technology in pictures
- Interpreting RFLP bands
- Advantages and disadvantages of RFLPs
- Applications
  - Maize
  - Wheat
  - Scots pine
RFLP technology

RFLP detection relies on the possibility of comparing band profiles generated after restriction enzyme digestion of target DNA. The laboratory steps involved are as follows:

- Isolation of DNA
- Restriction digestion and gel electrophoresis
- DNA transfer by Southern blotting
- DNA hybridisation
  - The procedure
  - The DNA probe
  - Sources of probes
- Equipment

Restriction fragment length polymorphism (RFLP) analysis was one of the first techniques to be widely used for detecting variation at the DNA sequence level. The principle behind the technology rests on the possibility of comparing band profiles generated after restriction enzyme digestion in DNA molecules of different individuals. Diverse mutations that might have occurred affect DNA molecules in different ways, producing fragments of variable lengths. These differences in fragment lengths can be seen after gel electrophoresis, hybridisation and visualisation.
Isolating DNA

- Total DNA is extracted from plant cells
- Alternatively, chloroplast and mitochondrial DNA can be used
- DNA must be clean and of high molecular weight
- Complications:
  - Breakage during isolation
  - DNA degraded by nucleases
  - Joint isolation of polysaccharides
  - Isolation of secondary plant metabolites

Isolating DNA is the first step for many DNA-based technologies. DNA is found either in nuclear chromosomes or in organelles (mitochondria and chloroplasts). To extract DNA from its location, several laboratory procedures are needed to break the cell wall and nuclear membrane, and so appropriately separate the DNA from other cell components. When doing so, care must be taken to ensure the process does not damage the DNA molecule and that it is recovered in the form of a long thread.
Extracted DNA is digested with specific, carefully chosen, restriction enzymes. Each restriction enzyme, under appropriate conditions, will recognise and cut DNA in a predictable way, resulting in a reproducible set of DNA fragments (‘restriction fragments’) of different lengths.

The millions of restriction fragments produced are commonly separated by electrophoresis on agarose gels. Because the fragments would be seen as a continuous ‘smear’ if stained with ethidium bromide, staining alone cannot detect the polymorphisms. Hybridisation must therefore be used to detect specific fragments.
DNA transfer is called ‘Southern blotting’, after E.M. Southern (1975), who invented the technique. In this method, the gel is first denatured in a basic solution and placed in a tray. A porous nylon or nitrocellulose membrane is laid over the gel, and the whole weighted down. All the DNA restriction fragments in the gel transfer as single strands by capillary action to the membrane. All fragments retain the same pattern on the membrane as on the gel.

Reference

The membrane with the target DNA is incubated with the DNA probe. Incubation conditions are such that if strands on the membrane are complementary to those of the probe, hybridisation will occur and labelled duplexes formed. Where conditions are highly stringent, hybridisation with distantly related or non-homologous DNA does not happen. Thus, the DNA probe picks up sequences that are complementary and 'ideally' homologous to itself among the thousands or millions of undetected fragments that migrate through the gel.

Desired fragments may be detected after simultaneous exposure of the hybridised membrane and a photographic film.
To detect the subset of DNA fragments of interest from within all the fragments generated by restriction digestion, a probe is needed. The DNA probe usually comes from a DNA library (either genomic or cDNA), which is a collection of vectors (e.g. plasmids) that contain a representation of an original DNA molecule cut into pieces. Vectors may be transformed into bacteria and may multiply the piece of DNA they contain many times.

The DNA probe is also converted into a single-stranded molecule, conveniently labelled, using any standard method (e.g. a radioisotope or digoxygenin), and hybridised with the target DNA, which is stuck to the membrane.
DNA hybridisation: Sources of probes (1)

- Nuclear DNA:
  - Genomic libraries
  - cDNA

- Cytoplasmic DNA

The species specificity of many single-locus probes requires that libraries be built when studying new species. However, probes from related genera can often be used.

Sources of DNA probes include:

- *Genomic libraries*—total plant DNA is digested with restriction enzymes and individual fragments cloned into a bacterial or viral vector. Suitable probes are selected from this 'anonymous' library for RFLP analysis.

- *cDNA (complementary DNA) libraries*—mRNA is isolated and transcribed into DNA, using the enzyme reverse transcriptase. The cDNA so obtained is cloned into vectors and used as a library for probes in RFLP analysis.

- *Cytoplasmic DNA*—mitochondrial and chloroplast DNA libraries.

As a result of the species specificity shown by many single-locus probes, genomic or cDNA libraries must often be built for studies on new species. This can be very time consuming. However, given current knowledge about common sequences and genes, probes from related genera can often be used.
DNA hybridisation: Sources of probes (2)

Repetitive sequences or minisatellite-type:

- Basic ‘motif’ of 10 to 60 bp in tandem
- Highly variable between human individuals
- Polymorphisms in the number of repeated units (also called VNTRs)

In plants, probes from an internal repeat from the protein III gene of the bacteriophage M13 have been used to reveal minisatellite sequences.

Repetitive sequences of the minisatellite type also have their particular application in RFLP analysis. They are the repeat sequences of a basic 'motif'. They measure 10 to 60 bp, are found in tandem (i.e. head to tail) and occur at many loci on the genome.

Work on plant minisatellite markers resulted from pioneering studies on the human genome by Jeffreys et al. (1985a, b), which showed minisatellite markers to be highly variable in humans. Because polymorphisms are related to the number of repeated units, the sequences are also called variable number of tandem repeats (VNTRs; Nakamura et al. 1987). A carefully selected probe can detect restriction fragments that represent a large number of loci. The patterns of minisatellite-bearing restriction fragments on film (the so-called 'DNA fingerprint') allow clear discrimination between different individuals.

References


Equipment

Resources:
- Distilled and/or deionised water
- Reagents

Equipment:
- Refrigerator and freezer
- Laminar flow hood
- Centrifuge
- Power supply units
- Hotplate or microwave
- pH meter
- Standard balance
- Gel electrophoresis units
- Dark room
- UV transilluminator
RFLP technology in pictures

The following slides illustrate the procedures of the RFLP technique
After agarose has been poured into the gel mould, combs are immediately inserted to form wells and left until the gel hardens. The combs are then removed and the gel placed in an electrophoresis chamber.
Samples of digested DNA, with bromophenol blue dye added, are loaded into the wells with a pipettor.
After electrophoresis, the gel is treated with NaCl to break the DNA double helix bonds and make it single-stranded. This allows later hybridisation with a single-stranded DNA probe.
The blotting tray is first prepared by saturating sponges with NaOH. Safety glasses and gloves are required, and a laboratory coat recommended. The safety regulations of the participating institution should be followed.
Absorbent paper is placed on top of the sponges to prevent direct contact with the gel.
Bubbles between the absorbent paper and sponges are removed by rolling a pipette or a glass rod across the paper. This ensures a complete transfer of the solution all through the gel.
The gaps left between the sponges and tray are covered with strips of plastic sheets to prevent evaporation, which would reduce the efficiency of transfer.
The treated agarose gel is placed on top of the absorbent paper.
Bubbles between the gel and paper are being squeezed out with a glass rod.
Membrane is cut into the appropriate size.
The membrane is placed on top of the gel, then covered with a piece of absorbent paper.
A stack of porous paper such as paper towels or newspaper is placed on the absorbent paper protecting the membrane.
The entire set-up is topped with a weight (here, a bottle of water standing on a piece of glass) to promote good transfer. After some hours the transfer is complete, the blotting paper is taken away, and the membrane stored until hybridisation with the probe.
The process of hybridisation begins. A blocker DNA (to minimise background hybridisation with the membrane) is boiled to denature it to single strands.
The membrane is placed in a plastic container with the appropriate hybridisation solution and the blocker DNA, and pre-incubated.
The labelled probe is added to the container with the hybridisation solution and membrane, and incubated overnight in an oven. The following day, the membrane is removed from the hybridisation set up, and washed with the appropriate stringency solution.

(Note: Institutions vary with respect to required safety practices; please check with your host institution.)
The membrane is then blotted dry and put into a cassette for holding X-ray film.
Inside a dark room, an X-ray film is also inserted into the cassette.
The cassette is wrapped, or sealed with tape, and stored in a freezer until the film is sufficiently exposed, usually 1 to 4 days.
This is an RFLP autoradiogram.
RFLP in pictures: summary

A
digestion

B

A  B
blotting

mut

electrophoresis

hybridization

Mutation = a new restriction site

Probe

Restriction site

Copyright: IPGRI and Cornell University, 2003
Interpreting RFLP bands (1)

A mutation creates a new restriction site within the target region. Two smaller bands are therefore detected on the film.

RFLP technology ideally results in a series of bands on a gel, which can then be scored either for presence or absence of particular bands, or as codominant markers. Differences between genotypes are usually visualised as a diverse pattern of DNA restriction fragments.

The different mutational events responsible for the polymorphisms detected by RFLP analysis are presented in this and following slides.

In the diagram above, a mutation creates a new restriction site in segment A, just at the location of the probe's recognition. As a consequence, the probe will simultaneously hybridise with the two segments created by the enzyme. In segment B, where no mutation has occurred, only one segment will hybridise with the probe. During electrophoresis, the two segments from A will migrate farther through the gel than the segment hybridised on B and, as such, the polymorphism will be observed in the gel as shown in the inset (at right).
Interpreting RFLP bands (2)

A mutation creates a new restriction site between flanking restriction sites, creating a smaller restriction fragment.

In this case, a new restriction site is again created by mutation in segment A. However, the new site appears to one side of the location of the probe's recognition. Hence, only one fragment will hybridise in segment A and one in segment B. The polymorphism will be shown as a shorter fragment hybridised in A than in B. The shorter fragment will also migrate farther in the gel (see inset).
Interpreting RFLP bands (3)

An insertion of a DNA sequence between the flanking restriction sites creates a larger restriction fragment.

If an insertion event takes place between two restriction sites (segment A), the hybridised fragment will be longer. As a result, a polymorphism will be observed between individuals A and B, in that the fragment hybridised in A will be longer than that in B and its migration distance shorter (see inset).
Interpreting RFLP bands (4)

A deletion of a DNA sequence between the flanking restriction sites creates a smaller restriction fragment.

If a deletion occurs between flanking restriction sites, the probe will hybridise to a shorter segment. This will be observed as a polymorphism in the gel with a fragment that migrated farther in the gel for individual A (see inset).
Interpreting RFLP bands (5)

One of the flanking restriction sites is changed or lost through mutation or deletion. Consequently, the restriction fragment is altered.

If only one restriction site remains, no restriction fragment is generated, and no hybridisation can occur. If a new site has been generated by the change, the new fragment will hybridise, and will show a different band pattern than an individual not having the change.
Advantages of RFLPs

- Highly robust methodology with good transferability between laboratories
- Codominantly inherited and, as such, can estimate heterozygosity
- No sequence information required
- Because based on sequence homology, highly recommended for phylogenetic analysis between related species
- Well suited for constructing genetic linkage maps
- Locus-specific markers, which allow synteny studies
- Discriminatory power—can be at the species and/or population levels (single-locus probes), or individual level (multi-locus probes)
- Simplicity—given the availability of suitable probes, the technique can readily be applied to any plant
Disadvantages of RFLPs

- Large amounts of DNA required
- Automation not possible
- Low levels of polymorphism in some species
- Few loci detected per assay
- Need a suitable probe library
- Time consuming, especially with single-copy probes
- Costly
- Distribution of probes to collaborating laboratories required
- Moderately demanding technically
- Different probe/enzyme combinations may be needed
Applications

- Genetic diversity
- Genetic relationships
- History of domestication
- Origin and evolution of species
- Genetic drift and selection
- Whole genome and comparative mapping
- Gene tagging
- Unlocking valuable genes from wild species
- Construction of exotic libraries

References marked in purple are discussed in detail in the following slides.


Example: Maize

Title:

Objective:
To examine the genetic relationships among inbred lines from known heterotic groups and landraces of great historical importance in the development of elite material

Materials and methods:
Sixty-two inbred lines of known heterotic groups and 10 maize populations (about 30 individuals per population) were assayed at 28 RFLP loci (29 different probe/enzyme combinations)

(continued on next slide)
Example: Maize (continued)

Results:
Comparing alleles specific to each type of germplasm showed a deficit of alleles within lines accounting for about 22% of the total allelic richness of the populations:

• Associations among inbreds and populations proved consistent with pedigree data of the inbreds and provided new information on the genetic basis of heterotic groups

• European flint inbreds were revealed to be as close to the north-eastern U.S. flint population studied as to the typical European populations

(continued on next slide)
**Examples: Maize (continued)**

- **Discussion:**
  The populations represent significant reservoirs of diversity, and elite germplasm is not likely to contain all useful alleles.

  *Question:* How can the European heterotic group be closer to the north-eastern U.S. flint 'Compton's Early' populations than to other U.S. populations?

- **Conclusions:**
  Results suggest that a larger set of populations should be studied with molecular markers to develop appropriate strategies for the most effective use of these genetic resources in breeding programs.
Example: Wheat

Title:

Objective:
To compare allelic frequencies in wheat populations that have been subjected to natural selection

Materials and methods:
Two initial populations and six derived subpopulations cultivated for 10 years in contrasting sites were studied at 30 loci

(continued on next slide)
Examples: Wheat (continued)

» Results and discussion:
  Differentiation between subpopulations based on RFLP diversity was highly significant:
  • Allelic frequency variation was found to be much greater than expected under genetic drift only. Selection greatly influenced the evolution of populations
  • Some loci revealed higher differentiation than other. This may have indicated that they were genetically linked to other polymorphic loci involved in adaptation

» Conclusions:
  Variations of allelic frequencies observed for the RFLP markers cannot be explained by evolution under genetic drift only but by direct or indirect effects of selection
Examples: Scots pine

Title:

Objective:
To investigate the geographical structure of mitochondrial DNA variants in western European populations of Scots pine

Materials and methods:
Twenty populations of *P. sylvestris* from Scotland and 18 from continental Europe (an average of 21 individuals per population) were studied, using RFLP analysis of total DNA

(continued on next slide)
Example: Scots pine (continued)

Results:
Three major mtDNA RFLP patterns (mitotypes a, b, and d) were detected:

- Within Spain, all three mitotypes were found. Gene diversity was high, being distributed predominantly among, rather than within, populations. Mitotype d was present only in the southernmost population (Sierra Nevada, Spain)

- Italian populations were fixed for mitotype b. Populations from northern France, Germany, Poland, Russia and southern Sweden were fixed for mitotype a. Populations in northern Fennoscandia (Norway, Sweden and Finland) were fixed for mitotype b

- In Scotland, mitotype a was largely fixed. Mitotype b was present in some polymorphic populations

(continued on next slide)
Example: Scots pine (continued)

Discussion:
- The detection of different mitotypes was unequivocal proof of genetic differences in the mtDNA genome. The diversity of mitotypes in Spanish populations indicates that these either descend from survivors in refugia during the last glaciation or represent Tertiary relics that have survived as isolated populations.
- In Europe, after glaciation, *P. sylvestris* apparently began developing in at least three evolutionary directions, each of which had a different origin—Spain, north central Europe and northern Fennoscandia.

Conclusions:
Studies of maternally inherited mtDNA markers can provide useful insights into the history of species and help define geographical areas with germplasm that may be conserved.
In summary

- The RFLP technology detects length changes in target DNA molecules after restriction enzyme digestion
- RFLP bands are detected by hybridising the target DNA with a DNA probe
- RFLP banding patterns reflect different mutational events at the hybridisation site of the probe or its neighbouring region
- RFLP is a highly robust technology, but time consuming and technically demanding
By now you should know

- Main steps required for detecting RFLPs
- Different sources of probes
- The effect of different mutational events on detecting RFLP banding patterns
- The advantages and disadvantages of the RFLP technology for genetic diversity analysis
Basic references


Next

DNA-based technologies
PCR-based technologies
PCR basics

- DNA-based technologies
  - PCR-based technologies
    - PCR with arbitrary primers
    - Amplified fragment length polymorphisms (AFLPs)
    - Sequences-tagged sites (STS)
    - Latest strategies
- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

IV PCR-based technologies

- PCR basics
- PCR with arbitrary primers
- Amplified fragment length polymorphisms (AFLPs)
- Sequence-tagged sites (Microsatellites, SCARs, CAPS, ISSRs)
- Latest strategies
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies

PCR-based technologies

PCR basics
Contents

› The polymerase chain reaction

› PCR procedures:
  • Steps
  • Cycle 1
  • Cycle 2
  • Cycle 3
  • Conditions for cycling
  • Conditions for the reaction mixture
  • Components
  • Equipment

› PCR technology in pictures
The polymerase chain reaction

PCR is a rapid, inexpensive and simple way of copying specific DNA fragments from minute quantities of source DNA material

- It does not necessarily require the use of radioisotopes or toxic chemicals
- It involves preparing the sample DNA and a master mix with primers, followed by detecting reaction products

Reference

The DNA polymerase, known as 'Taq polymerase', is named after the hot-spring bacterium *Thermus aquaticus* from which it was originally isolated. The enzyme can withstand the high temperatures needed for DNA-strand separation, and can be left in the reaction tube.

The cycle of heating and cooling is repeated over and over, stimulating the primers to bind to the original sequences and to newly synthesised sequences. The enzyme will again extend primer sequences. This cycling of temperatures results in copying and then copying of copies, and so on, leading to an exponential increase in the number of copies of specific sequences. Because the amount of DNA placed in the tube at the beginning is very small, almost all the DNA at the end of the reaction cycles is copied sequences.

The reaction products are separated by gel electrophoresis. Depending on the quantity produced and the size of the amplified fragment, the reaction products can be visualised directly by staining with ethidium bromide or a silver-staining protocol, or by means of radioisotopes and autoradiography.
The PCR steps are all carried out, one after the other, in bouts of cycling. Cycle 1 is as follows:

- During denaturation (about 1 min at 95°C), the DNA strands separate to form single strands.

- During annealing (about 1 min at temperatures ranging between 45°C and 60°C), one primer binds to one DNA strand and another binds to the complementary strand. The annealing sites of the primers are chosen so that they will prime DNA synthesis in the region of interest during extension.

- During extension (about 1 min at 72°C), the DNA synthesis proceeds through the target region and for variable distances into the flanking region, giving rise to ‘long fragments’ of variable lengths.
When the second cycle starts, there are effectively two types of template: (1) the original DNA strands; and (2) the newly synthesised DNA strands, consisting of the target region and variable lengths of the flanking region at the 3’ end. When the latter template is used in this cycle, only the target region is replicated.
In the third cycle, the newly synthesised target region DNA (i.e. without flanking regions) acts as template. The original DNA molecule is still present, and will be until the end of the reaction. However, after a few cycles, the newly synthesised DNA fragment quickly establishes itself as the predominant template.

Cycles are typically repeated 25 to 45 times. Standardisation of the thermocycler's running conditions is essential for the reproducibility of results.
PCR procedures: conditions for cycling

- Complete denaturation of the DNA template
- Optimal annealing temperature
- Optimal extension temperature
- Number of PCR cycles
- Final extension step

In the initial denaturation step, complete denaturation of the DNA template at the start of the PCR reaction is essential. Incomplete denaturation of DNA will result in the inefficient use of the template in the first amplification cycle and, consequently, poor yield of PCR product.

The annealing temperature may be estimated as 5°C lower than the melting temperature of the primer-template DNA duplex. If non-specific PCR products are obtained in addition to the expected product, the annealing temperature can be optimised by increasing it stepwise by 1-2°C.

Usually, the extension step is performed at 72°C and a 1-min extension is sufficient to synthesise PCR fragments as long as 2 kb (kb = kilobase = 1000 bp). When larger DNA fragments are amplified, time is usually extended by 1 min per 1000 bp.

The number of PCR cycles will basically depend on the expected yield of the PCR product.

After the last cycle, samples are usually incubated at 72°C for 5 min to fill in the protruding ends of newly synthesised PCR products.
PCR procedures: conditions for the reaction mixture

Contamination of the DNA must be prevented by:

- Separating the areas for DNA extraction and PCR
- Using sole-purpose laboratory equipment
- Autoclaving and aliquoting
- Adding a control reaction

Some useful tips:

- DNA extraction and PCR reaction mixing and processing should be performed in separate areas.
- Use of sole-purpose vessels and positive displacement pipettes or tips for DNA sample and reaction mixture preparation is strongly recommended.
- All solutions, except dNTPs, primers and Taq DNA polymerase, should be autoclaved. Where possible, solutions should be aliquoted in small quantities and stored in designated PCR areas.
- A good practice, to confirm absence of contamination, is to add a control reaction without template DNA.
# PCR procedures: components

<table>
<thead>
<tr>
<th>Components:</th>
<th>Considerations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sterile deionised water</td>
<td>• Template DNA</td>
</tr>
<tr>
<td>• 10X PCR buffer</td>
<td>• Primers</td>
</tr>
<tr>
<td>• dNTP mix</td>
<td>• MgCl(_2) concentration</td>
</tr>
<tr>
<td>• Primer</td>
<td>• Taq polymerase</td>
</tr>
<tr>
<td>• Taq DNA polymerase</td>
<td>• dNTPs</td>
</tr>
<tr>
<td>• MgCl(_2)</td>
<td></td>
</tr>
<tr>
<td>• Template DNA</td>
<td></td>
</tr>
</tbody>
</table>

Many PCR machines are now available in 48-, 96- or 384-well formats. This, combined with the use of multichannel pipettors, can greatly increase the number of reactions that can be done simultaneously. If several reactions need to be simultaneously prepared, a master mix should be used as follows: water, buffer, dNTPs, primers, MgCl\(_2\) and Taq DNA polymerase in a single tube. This will then be aliquoted into individual tubes.

## Considerations:

**Template DNA.** Nearly any standard method is suitable for template DNA purification. An adequate amount of template DNA is between 0.1 and 1 µg for genomic DNA for a total reaction mixture of 100 µl. Larger template DNA amounts usually increase the yield of non-specific PCR products.

**Primers.** (1) PCR primers should be 10-24 nucleotides in length. (2) The GC content should be 40%-60%. (3) The primer should not be self-complementary or complementary to any other primer in the reaction mixture, to prevent primer-dimer and hairpin formation. (4) Melting temperatures of primer pairs should not differ by more than 5°C, so that the GC content and length must be chosen accordingly. (5) The melting and annealing temperatures of a primer are estimated as follows: if the primer is shorter than 25 nucleotides, the approximate melting temperature is calculated with the formula: \( T_m = 4 \times (G + C) + 2 \times (A + T) \). (6) The annealing temperature should be about 5°C lower than the melting temperature.

**MgCl\(_2\) concentration.** Because Mg\(_2+\) ions form complexes with dNTPs, primers and DNA templates, the optimal concentration of MgCl\(_2\) has to be selected for each experiment. Too few Mg\(_2+\) ions result in a low yield of PCR product, and too many will increase the yield of non-specific products. The recommended range of MgCl\(_2\) concentration is 1 to 3 mM, under the standard reaction conditions specified.

**Taq DNA polymerase.** Higher Taq DNA polymerase concentrations than needed may cause synthesis of non-specific products.

**dNTPs.** The concentration of each dNTP (dATP, dCTP, dGTP, dTTP) in the reaction mixture is usually 200 µM. These concentrations must be checked as being equal, because inaccuracies will increase the degree of misincorporation.
PCR procedures: equipment

- Micropipettes
- Thermocycler
- Electrophoresis units
- Power supply units
- Photographic equipment
PCR technology in pictures

The following photographs demonstrate how PCR technology is carried out
The PCR mixture is prepared on ice. Safety clothing is not required, but may benefit the reaction unless good sterile techniques are practised.
The PCR reactions are loaded into the thermocycler.
The thermocycler is locked shut and programmed.
Different types of thermocyclers exist: the black one is a tetrad, in which 4 sets of 96 samples can be run simultaneously. The four smaller white ones each have a 96-well capacity.
Depending on the size of the PCR bands produced and the discrimination needed, band visualisation can be accomplished through either a regular, horizontal, agarose gel or a vertical acrylamide sequencing gel (see next slide). Here, the products are being run on an agarose gel.
The acrylamide gel may be run in either an independent unit or an automatic sequencer. The preparation of the sequencing gel, although somewhat complicated to set up, is similar for both cases. Here, the glass is being cleaned and wiped before preparing the gel for an automatic sequencer.
Glass units are being inserted into the frame, which will be fitted into the sequencer.
The glass units are clamped into place, and the entire unit readied for gel to be poured in.
Liquid acrylamide gel is poured into the mould. Because acrylamide is a carcinogen, safety clothing must be worn.
Clamps grip the bottom ends of the glass units to prevent the gel leaking.
A comb is inserted into the top of the gel to create the wells into which the samples will be loaded.
Once the gel is set, the unit is placed into the sequencing machine.
Samples are taken up with a multi-pipettor …
…and loaded into the wells of the gel.
This is a close-up from the previous photograph. Note the extremely small size of the wells and the pipettor, making it possible to load large numbers of samples into each gel.
A heating plate is placed against the gel to ensure a constant performing temperature.
Gel buffer is loaded into the top tank…
…and into the bottom tank.
A final look, then the door is shut and the program begun.
In summary

- PCR is a simple technique to obtain many copies of specific DNA fragments.

- Three steps are involved in PCR: denaturation, annealing and extension.

- To ensure success, care should be taken both in preparing the reaction mixture and setting up the cycling conditions.

Copyright: IPGRI and Cornell University, 2003
By now you should know

- What the basis of PCR is

- Those items that can affect the performance of the PCR, step by step, and what their effects on PCR results would be
Basic references


DNA-based technologies
PCR-based technologies
PCR with arbitrary primers

▶ DNA-based technologies
  ▪ PCR-based technologies
    ▪ Amplified fragment length polymorphisms (AFLPs)
    ▪ Sequences-tagged sites (STS)
    ▪ Latest strategies

▶ Complementary technologies

▶ Final considerations

▶ Glossary
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies

PCR-based technologies
PCR with arbitrary primers
(RAPDs, DAF, AP-PCR)
MAAP (multiple arbitrary amplicon profiling) is the acronym proposed to cover the three main technologies that fall in this category:

- Random amplified polymorphic DNA (RAPD)
- DNA amplification fingerprinting (DAF)
- Arbitrarily primed polymerase chain reaction (AP-PCR)

Three main techniques fall within the category of PCR-based markers using arbitrary primers: RAPD, DAF and AP-PCR. MAAP is the acronym proposed, but not commonly used, by Caetano-Anollés et al. (1992) to encompass all of these closely related techniques. In this submodule, special attention will be given to RAPD, concluding with a comparison of RAPD with DAF and AP-PCR.

Reference

The random amplified polymorphic DNA (RAPD) technique is a PCR-based method that uses a short primer (usually 10 bases) to amplify anonymous stretches of DNA. With this technique, there is no specific target DNA, so each particular primer will adhere to the template DNA randomly. As a result, the nature of the obtained products will be unknown. The DNA fragments generated are then separated and detected by gel electrophoresis.
Isolating DNA

Total, chloroplast or mitochondrial DNA can be used

- Tiny amounts of DNA are sufficient
- DNA must be clean and of high molecular weight

If minimal quality of DNA is not achieved, the reproducibility of results will be hard to ensure
The PCR components needed are already discussed in the submodule "PCR basics". Only a short primer (usually 10 bp long) is used. Primers with these characteristics are commercially available under different brands. A concentration of MgCl$_2$ is often added to promote amplification of more bands through the reaction. However, care should be taken to find the suitable concentration for each case, to prevent the appearance of non-specific products.

PCR conditions usually include an annealing cycle at a low temperature (about 40°C), thus encouraging primer—DNA annealing and leading to a sufficient number of products. Again, the appearance of non-specific products must be prevented, which can be done by determining the appropriate temperature at which 'ghost' bands will not appear.
**RAPD product detection**

Agarose or acrylamide gel electrophoresis and visualisation with ethidium bromide

RAPDs can be detected by running PCR products through electrophoresis on an agarose or acrylamide gel. In both cases, the gel is stained with ethidium bromide.

The difference obtained by running RAPD products in acrylamide versus agarose lies only in the degree of resolution of bands. In most cases, agarose gel electrophoresis gives sufficient resolution.
Diagrammatic summary

Adapted from Griffiths et al. 1996

Copyright: IPGRI and Cornell University, 2003
Equipment

Resources:
- Distilled and/or deionised water
- Reagents

Equipment:
- Refrigerator and freezer
- Laminar flow hood
- Centrifuge
- Thermocycler
- Power supply units
- Hotplate or microwave
- pH meter
- Standard balance
- Gel electrophoresis units
- UV transilluminator
Interpreting RAPD banding patterns (1)

DNA polymorphism among individuals can be due to:

- Mismatches at the primer site
- The appearance of a new primer site
- The length of the amplified region between primer sites
Interpreting RAPD banding patterns (2)

Because of the nature of RAPD markers, only the presence or absence of a particular band can be assessed. Criteria for selecting scoring bands:

- Reproducibility—need to repeat experiments
- Thickness
- Size
- Expected segregation observed in a mapping population
Interpreting RAPD banding patterns: Example of a bad RAPD gel

This picture shows an image of a bad quality RAPD gel. The bands are fuzzy. Those at the top have a smear starting from the well where the PCR product was loaded and many are observed only with difficulty. The hazy background makes observation difficult—whether, in certain cases, one band is found or two side by side. Certainly, some bands are clear and can be scored, but many other bands are dubious and their interpretation would be highly risky. Such difficulties raise questions of confidence in data collection.
Interpreting RAPD banding patterns: Example of a good RAPD gel

This picture shows an image of a very high quality RAPD gel. Both, presence and absence of most bands are very clear and the background is transparent. The researcher would have no doubts while selecting bands and collecting data from this gel. Consequently, the interpretation of results can be very confident.
Advantages of RAPDs

- High number of fragments
- Simple
- Arbitrary primers are easily purchased, with no need for initial genetic or genomic information
- Only tiny quantities of target DNA are required
- Unit costs per assay are low

Some comments:

- Many different fragments (corresponding to multiple loci dispersed throughout the genome) are normally amplified, using each single primer. The technique is therefore rapid in detecting polymorphisms. Although most commercially produced primers result in several fragments, some primers may fail to give amplification fragments from some material.

- The technique is simple. RAPD analysis does not require expertise to handle hybridisation of DNA or other highly technical activities.
Disadvantages of RAPDs

- Dominant
- Lack of a priori knowledge on the identity of the amplification products
- Problems with reproducibility
- Problems of co-migration

RAPD markers are dominant. Amplification either occurs at a locus or it does not, leading to scores based on band presence or absence. This means that homozygotes and heterozygotes cannot be distinguished. In addition, the absence of a band through lack of a target sequence cannot be distinguished from that occurring through the lack of amplification for other reasons (e.g. poor quality DNA), contributing to ambiguity in the interpretation of results.

Nothing is known about the identity of the amplification products unless the studies are supported by pedigree analysis.

Problems with reproducibility result as RAPD suffers from sensitivity to changes in the quality of DNA, PCR components and PCR conditions, resulting in changes of the amplified fragments. Reproducible results may be obtained if care is taken to standardise the conditions used (Munthali et al., 1992; Lowe et al., 1996).

Problems of co-migration raise questions like 'Do equal-sized bands correspond to the same DNA fragment?'

- The presence of a band of identical molecular weight in different individuals is not evidence per se that the individuals share the same (homologous) DNA fragment.
- A band detected on a gel as being single can comprise different amplification products. This is because the type of gel electrophoresis used, while able to separate DNA quantitatively (i.e. according to size), cannot separate equal-sized fragments qualitatively (i.e. according to base sequence).

References


Applications

- Genetic diversity
- Germplasm characterisation
- Genetic structure of populations
- Domestication
- Detection of somaclonal variation
- Cultivar identification
- Hybrid purity
- Genome mapping

References in purple are explained in detail in the following slides.


Example: Meadow fescue

- **Title:**
  Fertilization and defoliation frequency affect genetic diversity of *Festuca pratensis* Huds. in permanent grasslands. Mol. Ecol. 1998. 7:1557-1567

- **Objective:**
  To assess the genetic variability of meadow fescue, and determine whether fertilisation and defoliation frequency influence genetic variability within natural populations

- **Materials and methods:**
  - Six natural populations and 3 cultivars of *Festuca pratensis* were studied, using 69 RAPD markers (13 primers) and 7 agronomic traits.
  - Samples of natural populations were taken from two unrelated long-term experiments, where treatments had been applied for 11 to 38 years

(continued on next slide)
Example: Meadow fescue (continued)

Results:

- Factor analysis of agronomic traits failed to separate cultivars from the other populations

- Cluster analysis of RAPD data resulted in a clear distinction between cultivars and natural populations

- Genetic variability within cultivars was lower than within natural populations

- Analysis of molecular variance (AMOVA) showed a significant effect of management on genetic variation

(continued on next slide)
Example: Meadow fescue (continued)

- Discussion:
  Significant genetic variation exists within cultivars and natural populations of *Festuca pratensis*. However, fertilisation and high cutting frequency may reduce it.

- Conclusions:
  Unfertilised and rarely cut populations of meadow fescue should be conserved as a gene pool. Further studies of other species are needed to learn more about how management systems influence the diversity of plant communities and the genetic architecture of species.
Example: Bell pepper

- Title:

- Objective:
  To characterize germplasm of *Capsicum* species

- Materials and methods:
  A total of 134 accessions from six *Capsicum* species, maintained at the Asian Vegetable Research and Development Center (AVRDC) genebank, were characterised with 110 RAPD markers (25 primers)

(continued on next slide)
Example: Bell pepper (continued)

Results and discussion:

- Ten pairs of potentially duplicate accessions documented

- Diagnostic RAPDs were identified and employed to improve taxonomic identification. They could also be used to monitor natural and artificial hybridisation

- Three *Capsicum* accessions, misclassified on morphological traits, were re-identified through diagnostic RAPDs

- Three accessions, previously unclassified, were assigned to a species based on diagnostic RAPDs

- *Capsicum annuum* accessions from the genebank did not differ from lines in the breeding program for RAPD variation or diversity
Example: Bell pepper (continued)

Conclusions:

The AVRDC’s pepper breeding program seems to be working with a diversity that is representative of its pepper collection in the genebank. Molecular markers can be useful for rationalising the management of an *ex situ* collection (e.g. identifying duplicates or taxonomic errors).
Example: Modern rose

Title:

Objective:
To evaluate the genetic variability among cultivated rose varieties and investigate the history of rose breeding

Materials and methods:
From 13 horticultural groups, 100 old varieties of cultivated rose were selected for the way they marked successive stages of domestication. They were studied with AP-PCR, using five long (20-mer) primers

(continued on next slide)
Example: Modern rose (continued)

Results and discussion:

- Fifty-eight polymorphic DNA fragments were produced, 55 being informative and discriminatory, allowing differentiation among almost all 100 cultivars.

- A dendrogram showed the relationships between Chinese and European founder roses, hybrid groups of the first and second generations, and the most modern hybrid Teas produced during domestication.

- Principal component analysis demonstrated the occurrence of a continuous gradient of the European/Chinese allele ratio, and considerable reduction in genetic variability over the course of domestication.

(continued on next slide)
Example: Modern rose (continued)

Conclusions:
The effect of selection for a limited array of morphological traits resulted in the retention of only a small number of alleles during the domestication of rose. Rose has much more genetic variation than is so far used. Selecting for different variants to generate varieties with new and valuable aesthetic traits should be possible.
Differences: DAF and RAPD technologies

For DAF:

- Primer concentrations are higher
- Shorter primers are used (5 to 8 nucleotides)
- Two-temperature cycle vs. the 3-temperature cycle used in RAPD
- Highly complex banding patterns

References


Differences: AP-PCR and RAPD technologies

In AP-PCR:

- The amplification is in three parts, each with its own stringency and concentration of constituents
- High primer concentrations are used in the first PCR cycles
- Primers of variable length, and often designed for other purposes, are arbitrarily chosen for use (e.g. M13 universal sequencing primer)

Reference

In summary

- The RAPD technology is based on a simple PCR with a single short arbitrary primer

- RAPD easily produces a significant number of bands, but markers are dominant and reproducibility problems are common

- DAF and AP-PCR are alternative, more complex technologies, of RAPD
By now you should know

- Main features of the RAPD procedure
- Causes of polymorphisms in the DNA molecule that can be detected with RAPD
- Criteria for selecting RAPD bands for genetic diversity analysis
- Advantages and disadvantages of RAPD technology
- Main differences of DAF and AP-PCR with RAPD technology
Basic references


DNA-based technologies
PCR-based technologies
Amplified fragment length polymorphisms

- DNA-based technologies
  - PCR-based technologies
    - Sequences-tagged sites (STS)
    - Latest strategies

- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies

PCR-based technologies

Amplified fragment length polymorphisms (AFLPs)
Contents

- AFLP technology, step by step
  - Four steps
  - DNA digestion and ligation
  - PCRs and detection
  - Summarising the technology
- Equipment
- Interpreting AFLP bands
- Advantages and disadvantages of AFLPs
- Applications
  - Limonium sp.
  - Stylosanthes spp.
  - Indian mustard
AFLP* technology, step by step

Main features:

- A combination of the RFLP and PCR technologies
- Based on selective PCR amplification of restriction fragments from digested DNA
- Highly sensitive method for fingerprinting DNA of any origin and complexity
- Can be performed with total genomic DNA or with cDNA ('transcript profiling')

*The AFLP technique was developed by KeyGene (Netherlands), a private biotechnology company that has filed property rights on the technology. For more information, see KeyGene's home page: http://www.keygene.com

Reference

Four steps

- DNA is digested with two different restriction enzymes
- Oligonucleotide adapters are ligated to the ends of the DNA fragments
- Specific subsets of DNA digestion products are amplified, using combinations of selective primers
- Polymorphism detection is possible with radioisotopes, fluorescent dyes or silver staining
The DNA being examined is digested with two different restriction enzymes, one of which is a frequent cutter (the four-base restriction enzyme) and the other a rare cutter (the six-base restriction enzyme). Various enzyme/primer combinations can be used. MseI and EcoRI are best used in AT-rich genomes as they give fewer fragments in GC-rich genomes.

Specific synthetic adapters for each restriction site are then ligated to the digested DNA. Both the restriction and ligation steps can be performed in a single reaction.
A first PCR (pre-selective) is performed, using oligonucleotide primers complementary to the adapter and restriction sites. A nucleotide is added to the primers to select only a subset of fragments.

Pre-selective amplification products undergo another PCR run, and again a subset of those fragments is selected. Usually, for the second selective amplification, two extra nucleotides are added to the primers.

Fragments are separated by denaturing polyacrylamide ('sequencing gels') or capillary gel electrophoresis.

Let's say an extra nucleotide A is added to pre-selective primers. Hence, only a subset of the fragments of the mixture is amplified (i.e. those in which the restriction site sequence is followed directly by an A). Amplification primers are usually 17 to 21 nucleotides in length, and anneal perfectly to their target sequences.

A second amplification is then carried out, using similar oligonucleotide primers but with two extra bases (e.g. AC). Therefore, only a subset of the first amplification reaction will undergo subsequent amplification during the second round of PCR (i.e. those in which the AC sequence follows the restriction site sequence).

The subset of fragments are analysed by denaturing polyacrylamide gel electrophoresis to generate a fingerprint and DNA bands may be detected, using different methods.

In addition to the advantage of not requiring radioisotopes, fluorescent primers can be loaded as sets of three, each labelled with a different coloured dye, into the same gel lane, thus maximising the number of data points gathered per gel (Zhao et al., 2000).

Reference

Summarising the technology*

Genomic DNA

Digestion with two restriction enzymes

Ligation of adapters

Preamplification:
Primers = adapters + 1 base

Selective amplification:
Primers = adapters + 1 + 2 bases

AFLP profile

*Adapted from KeyGene's Web site: http://www.keygene.com
Equipment

Resources:
- Distilled and/or deionised water
- Reagents

Equipment:
- Refrigerator and freezer
- Laminar flow hood
- Centrifuge
- Thermocycler
- Power supply units
- Hotplate or microwave

- pH meter
- Standard balance
- Vertical gel electrophoresis units
- UV transilluminator
- Automatic sequencer
**Interpreting AFLP bands**

The AFLP technique detects polymorphisms arising from changes (presence or size) in the restriction sites or adjacent to these:

- Different restriction enzymes can be used, and different combinations of pre- and selective nucleotides will increase the probability of finding useful polymorphisms.
- The more selective bases, the less polymorphism will be detected.
- Bands are usually scored as either present or absent.
- Heterozygous versus homozygous bands may be detected, based on the thickness of the signal, although this can be tricky.

The molecular basis of AFLP polymorphisms will usually be caused at the nucleotide level. Single nucleotide changes will be detected when (1) the actual restriction sites are affected; and (2) nucleotides adjacent to the restriction sites are affected, which cause the primers to mispair at the 3’ end and prevent amplification.

Most AFLP markers will be mono-allelic, meaning that only one allele can be scored and the corresponding allele is not detected. At a low frequency, bi-allelic markers will be identified, as a result of small insertions or deletions in the restriction fragments.
This image shows an AFLP gel run in an automatic sequencer. Before loading the gel, samples were labelled with one of three fluorescent dyes (yellow, blue or green). Red marks a control sample that was included with the other samples to monitor the performance of the electrophoresis. Ascertaining the presence or absence of particular bands directly from the gel is almost impossible because of the high number of bands normally obtained through the AFLP procedure, and because fluorescent dyes as such cannot be seen by eye. Bands are determined with a laser, and data collection with the help of specialised computer software.
This image shows an AFLP gel run in a manual device (vertical electrophoresis unit) and stained with silver nitrate. The picture shows that certain regions of the gel are crowded with bands, whereas others are emptier. Collecting data from silver-stained gels can be done by eye, or with the help of a computer after the gel is appropriately scanned.
Advantages of AFLPs

- AFLPs allow a quick scan of the whole genome for polymorphisms
- Because of the large number of bands generated, each marker gives a highly informative fingerprint
- They are also highly reproducible
- No prior sequence information or probe generation is needed
- Extremely useful in creating quick genetic maps
- “Transcript profiling”

The AFLP technology can be applied to any DNA sample, including human, animal, plant and microbial DNA, giving it the potential to become a universal DNA fingerprinting system.

Because of the nature of AFLP primers, the markers obtained are highly reliable and robust, unaffected by small variations in the amplification process.

A typical AFLP fingerprint contains between 50 and 100 amplified fragments, many of which, or even most, may serve as genetic markers.

The generation of transcript profiles using AFLPs with cDNAs is an efficient tool for identifying differentially expressed mRNAs. This tool has several advantages that can be useful for discovering genes in germplasm.
Disadvantages of AFLPs

- AFLPs generate huge quantities of information, which may need automated analysis and therefore computer technology

- AFLP markers display dominance

- In genetic mapping, AFLPs often cluster at the centromeres and telomeres

- They are technically demanding in the laboratory and, especially, in data analysis

A further drawback of AFLP technology is perhaps the lack of guarantee of homology between bands of similar molecular weight (MW), thus creating difficulties for some types of studies such as phylogenetic analyses. However, while non-homologous bands with similar weight are also found with other markers such as RAPDs, they may, in fact, be less common with AFLP technology because gel resolution is very high and, consequently, the likelihood of non-homologous bands being coincidentally of the same molecular weight is low.
Applications

- Genetic diversity assessment
- Genetic distance analysis
- Genetic fingerprinting
- Analysis of germplasm collections
- Genome mapping
- Monitoring diagnostic markers

References in purple colour are explained in detail in the following slides.


Example: *Limonium sp.*

- **Title:**

- **Objective:**
  To compare the performance of AFLPs for the genetic diversity analysis of *L. cavanillesii* and their efficiency against a previous RAPD study

- **Materials and methods:**
  DNA was extracted from 29 wild individuals. These individuals were the same as those employed in a previous RAPD study*. Three primer combinations were selected

Example: *Limonium sp.* (continued)

- **Results:**
  A total of 231 fragments were generated: on average, 223 per individual and 77 per primer combination. Only 6% of AFLP markers were polymorphic. With these, 11 different AFLP profiles could be distinguished in the species, while in a previous RAPD* study, no polymorphic markers were obtained.

- **Discussion:**
  AFLPs proved to be a suitable marker type for gathering the information critical to identifying natural populations at risk and planning recovery strategies for *L. cavanillesii*. The low levels of genetic variability found in *L. cavanillesii* could be explained as a consequence of either its apomictic reproductive system or a recent and severe bottleneck event.

Example: *Stylosanthes* spp.

- **Title:**

- **Objective:**
  To assess genetic variation in two species of the tropical genus *Stylosanthes* Sw.

- **Materials and methods:**
  For analysis, 111 accessions were selected: 59 *S. viscosa* and 52 *S. humilis*, representing the geographical distribution of both species. Five primer combinations were used for *S. humilis* and four for *S. viscosa*

(continued on next slide)
Example: *Stylosanthes* spp. (continued)

Results:

- *S. humilis*: 316 out of 417 total bands were polymorphic (75%). The average similarity among accessions was 0.71, and the average genetic distance was 0.17 (Jaccard's coefficient).

- *S. viscosa*: 312 out of 373 total of bands were polymorphic (83%). The average similarity value among accessions was 0.67, and the average genetic distance was 0.20.

- Cluster analysis and PCA grouped accessions from both species by geographical origin, with some exceptions. One explanation may be the incidence of long-range dispersal events or introductions of genotypes from one area to another by humans.

(continued on next slide)
Example: Stylosanthes spp. (continued)

- **Discussion:**
  These results demonstrate the usefulness of AFLP technology, not only to detect genetic diversity within species of *Stylosanthes*, but also to identify individuals that might have been misclassified.

- **Conclusions:**
  This study shows the ability of AFLP to detect geographic patterns in genetic variation, which information is necessary for developing a strategy for optimal conservation.
Title:

Objectives:
- Study the genetic variation among *B. juncea* accessions
- Identify the AFLP primer combinations that would be informative for varietal identification
- Search polymorphic markers for eventual tagging of putative agronomic traits available in the germplasm studied

Materials and methods:
The study used 30 *B. juncea* accessions, representing 21 established natural populations and 9 synthetic varieties and lines. Selective amplification was carried out with 21 EcoRI/MseI primer pair combinations
Example: Indian mustard (continued)

Results:

• For the 30 genotypes, 1251 fragments were scored. On average, 37 bands per primer combination were polymorphic.

• No single primer pair could distinguish all 30 genotypes on the basis of the presence or absence of variety specific bands.

• Four primer pairs were found to be most informative and had 100% discriminatory power.

• Cluster analysis based on these four informative primers broadly agreed with results of earlier studies based on morphological traits* and RAPD data**.

(continued on next slide)


Example: Indian mustard  
(continued)

- Discussion and conclusions:
  
  In a previous RAPD* study with 32 primers, an average of 12 polymorphic loci per primer had been obtained. Thus, the AFLP technique provided a higher degree of resolution for discriminating among closely related germplasm than did RAPDs. AFLP analysis therefore has the potential to complement both conventional and other types of molecular marker data.

In summary

- The AFLP technology is based on the selective amplification of restriction fragments after DNA digestion.
- It detects polymorphisms caused by changes in the restriction sites or their neighbouring regions.
- This technology yields a large number of bands per run, but they are dominant and the procedure technically demanding.
By now you should know

- The different steps involved in the generation of AFLPs
- Main considerations for interpreting AFLP bands
- Advantages and disadvantages of AFLPs for genetic diversity analysis
Basic references


DNA-based technologies
PCR-based technologies
Sequences-tagged sites

- DNA-based technologies
  - PCR-based technologies
    - Latest strategies
- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies

PCR-based technologies
  • Sequences-tagged sites
    (Microsatellites, SCARs, CAPS, ISSRs)
Contents

- Sequence-tagged sites (STS) as markers
- Microsatellites (SSRs, STMS or SSRPs)
  - Identifying microsatellite regions
  - Structure
  - Selecting primers
  - Methodology and visualisation
  - Equipment
  - Advantages and disadvantages
  - Applications of SSRs and examples
- SCARs
- CAPS
- ISSRs
Sequence-tagged sites (STS) as markers

- Unlike arbitrary primers, STS rely on some degree of sequence knowledge
- Markers based on STS are codominant
- They tend to be more reproducible because longer primer sequences are used
- Require the same basic laboratory protocols and equipment as standard PCR

Unlike PCR with arbitrary primers, sequence-tagged sites (STS) are primers that are based on some degree of sequence knowledge. These unique, sequence-specific primers detect variation in allelic, genomic DNA. STS have a particular advantage over RAPDs in that they are codominant, that is, they can distinguish between homozygotes and heterozygotes. They also tend to be more reproducible, because they use longer primer sequences.

However, they have the disadvantage of requiring some pre-existing knowledge of the DNA sequence of the region, even if only for a small amount. The investment in effort and cost needed to develop the specific primer pairs for each locus is their primary drawback.

As with RAPDs, using PCR produces a quick generation of data and requires little DNA. All STS methods use the same basic protocols as RAPDs (DNA extraction and PCR) and require the same equipment.
Microsatellites are also called simple sequence repeats (SSRs) and, occasionally, sequence-tagged microsatellite sites (STMS) or simple sequence repeat polymorphisms (SSRPs). They are by far the most widely used type of STS.

SSRs are short tandem repeats, their length being 1 to 10 bp, most typically, 2-3 bp. SSRs are highly variable and evenly distributed throughout the genome. This type of repeated DNA is common in eukaryotes, their number of repeated units varying widely among organisms to as high as 50 copies of the repeated unit. These polymorphisms are identified by constructing PCR primers for the DNA flanking the microsatellite region. The flanking regions tend to be conserved within the species, although sometimes they may also be conserved in higher taxonomic levels.

Reference

As do areas of the genome high in repeats, SSRs tend to cluster at the centromeres and telomeres. However, this problem can be solved by developing SSRs from EST libraries, which are gene rich and more evenly distributed.
Structure

- Repeat (e.g. ga)
- Unique flanking regions

- The number of repeats is highly variable among individuals
Selecting primers

- Design primers (.Comparator) complementary to flanking regions
Methodology and visualisation

- **Methodology:**
  - DNA extraction
  - PCR with primers specific for microsatellite flanking regions
  - Separation of fragments

- **Visualisation:**
  - By agarose gel electrophoresis, using ethidium bromide staining and UV light, or
  - Acrylamide gels using silver staining or radioisotopes, or
  - Through automated sequencers, using primers pre-labelled with fluorescence

- **Data analysis**

PCR product size variation is caused by differences in the number of microsatellite repeat units. SSR polymorphisms can be visualised by agarose or polyacrylamide gel electrophoresis. Microsatellite alleles can be detected, using various methods: ethidium bromide, silver staining, radioisotopes or fluorescence.

If fluorescence-labelled primers are used, and the products are different enough in size and not overlapping, then multiplexing—that is, loading more than one sample per lane—of reaction products can greatly increase the already high efficiency of these markers (Dean *et al.*, 1999, provide a good example).

**Reference**

As mentioned previously, one way of visualising microsatellites is by using agarose gel electrophoresis. This method is appropriate when the alleles are long enough, that is, more than 200-300 base pairs, and the differences among alleles also significant (i.e. more than 10-20 bp).

This picture shows a microsatellite that was run on an agarose gel stained with ethidium bromide. The second and third lanes (the first, very faint, is a marker lane) correspond to the parents, one of which has only one band, and the other two. The heterozygote, thus, has three. In the second parent, one of the bands is much lighter. Because the two bands co-segregate, they are not a result of two loci being in different places, but because two copies of the microsatellite repeats are either separated by an insertion or deletion, or they are located near each other.
Microsatellites can also be analysed after running PCR products through an acrylamide gel stained with silver nitrate. In this picture, individual samples belong to a diploid species and therefore have a maximum of two alleles.
Resources:
- Distilled and/or deionised water
- Reagents

Equipment:
- Refrigerator and freezer
- Laminar flow hood
- Centrifuge
- Thermocycler
- Power supply unit
- Hotplate or microwave
- pH meter
- Standard balance
- Horizontal and vertical gel electrophoresis units
- UV transilluminator
- Automatic sequencer
Advantages and disadvantages

Advantages:
- Require very little and not necessarily high quality DNA
- Highly polymorphic
- Evenly distributed throughout the genome
- Simple interpretation of results
- Easily automated, allowing multiplexing
- Good analytical resolution and high reproducibility

Disadvantages:
- Complex discovery procedure
- Costly

The loci identified are usually multi-allelic and codominant. Bands can be scored either in a codominant manner, or as present or absent.

Because flanking DNA is more likely to be conserved, the microsatellite-derived primers can often be used with many varieties and even other species. These markers are easily automated, highly polymorphic, and have good analytical resolution, thus making them a preferred choice of markers (Matsuoka et al., 2002).

Reference

Applications of SSRs

- Individual genotyping
- Germplasm evaluation
- Genetic diversity
- Genome mapping
- Phylogenetic studies
- Evolutionary studies

References in purple colour are explained in detail in the following slides.


Applications: example of sorghum

- **Title:**

- **Objective:**
  To assess the levels of genetic redundancy in sorghum accessions maintained by the U.S. National Plant Germplasm System

- **Materials and methods:**
  96 individuals (5 plants each of 19 accessions of the line "Orange" and one elite inbred variety) were assayed with 15 SSR markers

(continued on next slide)
Applications: example of sorghum (continued)

Results and discussion:

- Most accessions were genetically distinct, but two redundant groups (involving a total of 5 entries) were found.
- Average heterozygosity values were very low (as expected for a self-pollinated crop). One accession contained a mix of genotypes, indicating some kind of contamination.
- Molecular variance analysis (AMOVA) showed that 90% of the total genetic variation was due to differences among accessions, while 10% resulted from genetic differences between individual plants within accessions.

(continued on next slide)
Applications: example of sorghum (continued)

Conclusions:

- 15 SSR markers provided substantial genetic resolution
- The number of 'Orange' accessions being maintained could be reduced to almost half without a substantial loss in overall genetic variation, thus greatly cutting down on maintenance costs
- The ability to multiplex reactions resulted in savings of 1152 gel lanes, that is, 80% of reagents, and time
Applications: example of olive

Title:
Microsatellite markers isolated in olive (Olea europaea L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theor. Appl. Genet. 104:223-228

Objective:
To assess the efficiency of SSR markers in identifying polymorphisms among olive cultivars

Materials and methods:
36 SSRs were used to assay 12 olive cultivars (4 well known and 8 ancient ones)

(continued on next slide)
Applications: example of olive (continued)

- Results and discussion:
  All except two of the SSR markers showed polymorphism, identifying between 1 and 5 alleles. All cultivars were easily separated from each other

  - Five primer pairs amplified two different loci
  - Six primer pairs were discarded because they yielded unreadable patterns
  - Two varieties, which had been suspected of being identical, were confirmed as such when they showed equal banding patterns at all loci. Another pair of varieties, also thought to be identical, were shown not to be
Applications: example of olive (continued)

Conclusions:

SSRs could easily differentiate between all the varieties, and thus comprised a good tool for fingerprinting. Genetic variability within olive cultivars was also identified, using a very low number of markers. Previous studies that had used AFLPs required many more markers.
Title: Simple sequence repeats (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. Euphytica. 1999. 109:85-92

Objective: To determine the extent and distribution of genetic variation in *B. nigra* or black mustard

Materials and methods: Five SSR markers were used to assay 32 *B. nigra* accessions (including genebank accessions and weed populations) from four regions: Europe/North Africa, India, Ethiopia and North America

(continued on next slide)
Applications: example of black mustard
(continued)

Results and discussion:
- Ethiopian entries formed the most distinctive group
- More than half of the variation was between plants within entries
- European and North American entries contained the most variation, and were generally grouped together
- Unique variants were present in weed populations from North America but variation between populations was not correlated with geographic distance

(continued on next slide)
Applications: example of black mustard

Conclusions:

Despite the belief that little genetic variation existed within *B. nigra*, the SSR markers demonstrated that the species's patterns of variation were consistent with its agricultural history.
Sequence characterized amplified regions (SCARs)

- SCARs take advantage of a band generated through a RAPD experiment
- They use 16-24 bp primers designed from the ends of cloned RAPD markers
- This technique converts a band—prone to difficulties in interpretation and/or reproducibility—into being a very reliable marker

Reference

Steps to obtain SCAR polymorphisms

- A potentially interesting band is identified in a RAPD gel
- The band is cut out of the gel
- The DNA fragment is cloned in a vector and sequenced
- Specific primers (16-24 bp long) for that DNA fragment are designed
- Re-amplification of the template DNA with the new primers will show a new and simpler PCR pattern
Diagram of the SCAR procedure

1. RAPD gel
2. Clone the polymorphic band in a vector
3. Sequence the fragment and design new primers to amplify specifically the band of interest
4. After amplification with the new primers, the result is a clear pattern of bands that is easier to interpret
Advantages and disadvantages

Advantages:
- Simpler patterns than RAPDs
- Robust assay due to the design of specific long primers
- Mendelian inheritance. Sometimes convertible to codominant markers

Disadvantages:
- Require at least a small degree of sequence knowledge
- Require effort and expense in designing specific primers for each locus

Because the primers used are longer than is usual for RAPDs, SCARs are typically more reproducible than the RAPDs from which they were derived. SCARs are usually codominant, although not if one or both primers overlap the site of sequence variation.
Cleaved amplified polymorphic sequence (CAPS)

- This method is based on the design of specific primers, amplification of DNA fragments, and generation of smaller, possibly variable, fragments by means of a restriction enzyme.

- This technique aims to convert an amplified band that does not show variation into a polymorphic one.

Cleaved amplified polymorphic sequences (CAPS) are like SCARs, but with an additional step of a restriction digest to help identify polymorphisms that may not be identifiable from whole PCR products. Both SCARs and CAPS are based on the presence of nucleotide changes or insertions and/or deletions causing differences between the test sequences. One drawback of both is that they detect polymorphism only over a small range of the genome, the area between the primers being typically less than 5 kb.

Reference

Steps for generating CAPS

- A band, DNA, gene sequence or other type of marker is identified as important
- Either the band is detected through PCR (and cut out of the gel, and the fragment cloned and sequenced) or the fragment sequence is already available
- Specific primers are designed from the fragment sequence
- The newly designed primers are used to amplify the template DNA
- The PCR product is subjected to digestion by a panel of restriction enzymes
- Polymorphism may be identified with some of the enzymes

Once a polymorphism is identified with a particular restriction enzyme, the primers may be redesigned, based on the newly generated fragments, to optimise the detection and visualisation of the polymorphism.

Primers, when possible, should be chosen so that the PCR products are likely to include introns. This will increase the chances of obtaining polymorphisms.
In this example, CAPS were generated for two *Arabidopsis* ecotypes.

At the top of the diagram, the three possible genotypes for the experiment are shown: the two homozygous ecotypes (A/A and B/B) and the heterozygote (A/B).

If a standard PCR were to be performed with the primers as drawn (blue arrows), no polymorphism would be detected among the three genotypes.

A restriction enzyme was found that would digest the A fragments twice and the B fragments three times. Consequently, the heterozygote A/B should have a copy of the fragment digested twice and of the fragment digested three times.

A PCR is then performed and the products digested with the specific restriction enzyme already mentioned. Visualisation on an agarose gel showed three fragments for genotype A/A, four fragments for genotype B/B and seven fragments for the heterozygote A/B. The diagram shows only 5 fragments as being observed for A/B, because two (shown by asterisks) of the seven fragments migrate similar distances as other fragments.
Advantages and disadvantages

Advantages:
- Robust assay because specific long primers are designed
- Codominant markers
- Benefit from markers that may have already been mapped
- Identify polymorphisms in markers that were previously not informative

Disadvantages:
- Require at least a small amount of sequence knowledge
- Effort and expense required to design specific primers for each locus
Inter-simple sequence repeats (ISSRs)

- They are regions found between microsatellite repeats
- The technique is based on PCR amplification of intermicrosatellite sequences
- Because of the known abundance of repeat sequences spread all over the genome, it targets multiple loci

Reference

Identifying ISSR polymorphisms

A typical PCR is performed in which primers have been designed, based on a microsatellite repeat sequence, and extended one to several bases into the flanking sequence as anchor points. Different alternatives are possible:

- Only one primer is used
- Two primers of similar characteristics are used
- Combinations of a microsatellite-sequence anchored primer with a random primer (i.e. those used for RAPD)
Designing primers for ISSR polymorphisms

The diagram above presents three different items:

- The original DNA sequence in which two different repeated sequences (CA), inversely oriented, are identified. Both repeated sections are, in addition, closely spaced.

- If primers were designed from within the repeated region only, the interrepeat section would be amplified but locus-specificity might not be guaranteed. In the second row, a PCR product is shown as a result of amplification from a 3’-anchored primer (CA)_n NN at each end of the interrepeat region. CA is the repeat sequence that was extended by NN, two nucleotides running into the interrepeat region.

- Alternatively, anchors may be chosen from the 5’ region. The PCR product in the third row is a result of using primers based on the CA repeat but extended at the 5’ end by NNN and NN, respectively.


Copyright: IPGRI and Cornell University, 2003
Advantages and disadvantages

Advantages:
- Do not require prior sequence information
- Variation within unique regions of the genome may be found at several loci simultaneously
- Tend to identify significant levels of variation
- Microsatellite sequence-specific
- Very useful for DNA profiling, especially for closely related species

Disadvantages:
- Dominant markers
- Polyacrylamide gel electrophoresis and detection with silver staining or radioisotopes may be needed
In summary

- The use of STS as markers relies on some sequence knowledge. They are codominant and highly reproducible.

- SSRs are the type of STS most widely used. Others are SCARs, CAPS and ISSRs.

- SSRs are short tandem repeats, highly variable and evenly distributed in the genome. These features make SSRs a good marker of choice for genetic diversity analyses.

- Polymorphisms in SSRs are caused by differences in the number of repeat units.
By now you should know

- The features of STS as markers
- The different types of STS markers and their contrasting traits
- The steps involved in identifying microsatellites, designing primers and detecting polymorphisms
- The properties of microsatellites for genetic diversity analyses
Basic references


Next

DNA-based technologies
PCR-based technologies
Latest strategies

- Complementary technologies
- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

DNA-based technologies

**PCR-based technologies**

**Latest strategies**

(DNA sequencing, ESTs, microarrays, DArT, SNPs)

Copyright: IPGRI and Cornell University, 2002
Contents

- DNA sequencing
- Expressed sequence tags (ESTs)
- Microarray technology
- Diversity array technology (DArT)
- Single nucleotide polymorphisms (SNPs)
DNA sequencing

- DNA sequencing is the most fundamental measure of diversity because it detects polymorphisms within the DNA's building blocks themselves.
- Data collection can be automated.

DNA sequencing provides the most fundamental measure of diversity, because all markers are derived from polymorphisms in the DNA's building blocks, that is, the nucleotide sequence of a particular DNA segment. Sequencing technology has vastly improved in recent years, and now PCR products (a DNA region amplified in sufficient quantity) can be sequenced directly and targeted to any genomic location of interest. Data collection can be automated.
DNA sequencing: methods

Two main methods exist:

- Maxam-Gilbert
- Sanger (dideoxy sequencing or chain termination)

The two methods differ slightly, with the Sanger method (described in slide 6) being easier to automate and, thus, more widely used.
DNA sequencing: procedures

- The DNA is broken up into fragments, which are then subcloned.
- Each short piece is used as a template to generate a set of fragments that differ in length from each other by a single base.
- Fragments are separated by gel electrophoresis.
- The base at the end of each fragment is identified (‘base-calling’). The original sequence of As, Ts, Cs and Gs is recreated for each short piece generated in the first step.
- The short sequences are assembled into one long sequence.
Differently coloured fluorescent dyes can be used, permitting the separation of all four fragments in a single lane on the gel and greatly increasing efficiency. Automated sequencers can analyse the resulting electropherograms to produce a four-colour chromatogram that shows peaks representing each of the four DNA bases.
DNA sequencing: advantages and disadvantages

- **Advantages:**
  - Results are highly reproducible
  - Maximum amount of information content

- **Disadvantages:**
  - Costs are still high
  - Technically demanding

The results are, of course, highly reproducible and informative. Costs are high, however, and a high level of technical expertise is needed, making this technology unavailable to many researchers. The use of PCR for targeting particular regions of DNA and the availability of automated sequencing machines have reduced the technical difficulties, but the process is still expensive, particularly to set up.
Applications

- Evolutionary studies
- Calculations of genetic variation
- Comparative genomics
- Creating PCR assays (making primers to convert any marker to a PCR-based marker)

Although marker technology is, in general, based on DNA sequence variation, fortunately, a researcher does not necessarily need to know the entire DNA sequence to use molecular markers. Of course, DNA sequencing has many useful applications, but a major drawback, particularly for diversity measurements, is that different genes evolve at different rates. Extrapolating information from particular genes to the species level must therefore be done with care (Brown and Kresovich, 1996).

Reference

Basic references


Expressed sequence tags (ESTs)

- Expressed sequence tags are small pieces of DNA sequence, usually 200 to 500 nucleotides long.
- Generated by sequencing either one or both ends of an expressed gene from a cDNA library.
- This strategy is an extremely efficient way to find new genes.

The number of publicly available plant EST sequences has increased dramatically in the last few years to more than 1,000,000 as of writing (National Plant Genome Initiative Progress Report, December 2001). A list of databases of ESTs for many plants can be found at http://www.ostp.gov/NSTC/html/mpgi2001/building.htm
Designing EST primers

mRNA → Reverse transcriptase → RNA → Ribonuclease degradation of RNA → Synthesis of second strand of DNA → Double-stranded DNA

5' EST → Primer → 3' EST


Copyright: IPGRI and Cornell University, 2003  Latest strategies
ESTs: advantages and disadvantages

Advantages:
- Extremely good as genetic markers
- Codominant
- Sequences can be generated rapidly
- Efficient source of sequences to derive primers for SSRs

Disadvantages:
- Isolation of mRNA may be difficult
- Introns, which may contain important information, are not part of cDNA
ESTs: applications

EST applications are all based on the fact that ESTs originate from segments of gene sequences:

- Comparing gene diversity in different organisms
- Gene evolution studies
- Searching databases for putative orthologues
- Probes for gene expression studies
- Detection of SNPs*

*See section beginning slide 29
Basic reference

Microarray, or ‘chip’, technology

- Microarrays are arrangements of small spots of DNA fixed to glass slides or nylon membranes
- The technology allows monitoring of the whole genome at once
- The underlying principle of chips is base-pairing or hybridisation between short probes and complementary DNA sequences
- Microarrays are constructed using cDNAs (cDNA arrays), genomic sequences or oligonucleotides synthesised in silico (‘DNA chips’)

Genomes can now be analysed on a whole-genome scale, using microarray (also called ‘chip’) technology. This technology is based on hybridisation between short oligonucleotide probes and complementary DNA sequences. Tens of thousands of samples can be immobilised on a tiny glass (more typically) or nylon slide (chip), and can be hybridised more than once with different probes or targets (the terminology is inconsistent on whether the immobilised DNA on the chip should be called the target or probe). More than one probe can be hybridised at a time, for example, to compare differences in expression, by labelling them with different-coloured fluorescent dyes.

Special software programs generate the data automatically. Microarrays can be used for diagnostics, studying gene expression and gene mapping, among other things. However, the technology is still relatively expensive, especially to set up, and the amount of data generated can be daunting.

For useful references, see Richmond and Somerville (2000) and Brown and Botstein (1999) at the end of this submodule (slide 19).
A hybridised chip

Image courtesy of Mark D’Ascenzo, Boyce Thompson Institute for Plant Research, Center for Gene Expression Profiling, Cornell University.

http://bti.cornell.edu/CGEP/CGEP.html
Microarrays: advantages and disadvantages

Advantages:
• High-throughput technology
• Whole genome scanning
• Allow the discovery of genotype-phenotype relationship

Disadvantages:
• Gene sequence data must be available
• Expensive
• Technically demanding
• Amount and type of data produced requires high-level computing expertise and equipment
Microarrays: applications

- Identification of sequence (gene or gene mutation)
- Determination of expression level of genes
  - Assay of specific genomic DNA sequence abundances
  - Analysis of expression of very large numbers of genes (cDNA arrays)
  - Identification of large numbers of specific DNA markers (e.g. single nucleotide polymorphisms or SNPs) by molecular hybridisation (synthetic oligonucleotide arrays)
Basic references


Diversity array technology (DArT)

Two steps are involved:

- Generating the array
- Genotyping a sample

Diversity arrays, also called DArT, was developed by CAMBIA. It involves a new use of microarrays that does not require sequence knowledge, and thus may become very useful to crop researchers.

All the following slides on DArT have been taken, with the Centre's previous authorisation, from CAMBIA's Web site: http://www.cambia.org.au/

References

DArTs: preparing the array (1)

- Restriction generated fragments representing the diversity of a genepool are cloned. The outcome is called a 'representation' (typically 0.1% to 10% of the genome)

- Polymorphic clones in the library are identified by arraying inserts from a random set of clones and hybridising the array to different samples

- The inserts from polymorphic clones are immobilised on a chip
DArTs: preparing the array (2)

Pool genomes

Use complexity reduction method, e.g. RE digest, adaptor ligation, PCR amplification

Clone fragments from the representation

Library

Pick individual clones and PCR amplification

Array-purified PCR products

Gx  Gy  Gn  →  DNAs of interest

Copyright: IPGRI and Cornell University, 2003  Latest strategies
DArTs: genotyping a sample (1)

- Label the representation (DNA) of the sample with fluorescence and hybridise against the array
- Scan the array and measure, for each array spot, the amount of hybridisation signal
- By using multiple labels, contrast a representation from one sample with a representation from another or with a control probe
**DArTs: genotyping a sample (2)**

1. **Choose 2 genomes to analyse**

2. **Cut, ligate adaptors and PCR amplify**

3. **Same complexity reduction as used to make the diversity panel**

4. **Label each genomic subset: red...**

5. **Hybridise to chip**

6. **Label each genomic subset ... green**
A hybridised DArT chip
DArTs: advantages and disadvantages

Advantages:
- Do not require sequence information
- High throughput
- Fast data acquisition and analysis
- Detects single-base changes as well as insertions and or deletions
- Detects differences in DNA methylation, depending on the enzyme used to generate the fragments
- Sequence-ready clones are generated
- Small DNA sample required
- Good transferability of markers among breeding populations
- Full automation possible

Disadvantages:
- Dominance of markers
- Technically demanding
**DArTs: applications**

- Rapid germplasm characterisation
- Genetic mapping
- Marker-assisted breeding

Reference

Basic References


A newer type of marker that has now been made available through new sequencing technologies is single nucleotide polymorphisms (SNPs). These polymorphisms are single-base substitutions between sequences. SNPs occur more frequently than any other type of marker, and are very near to or even within the gene of interest.

SNPs can be identified by either using microarrays or DHPLC machines.
DHPLC refers to denaturing high-performance liquid chromatography, which is used to visualise SNPs.
The schematic drawing of a single nucleotide polymorphism shows two DNA fragments (top and bottom) sharing the same sequence for 31 base pairs, except one. In position 28, an A-T (top) has changed to a C-G (bottom).
The height of the block represents a stretch of DNA in a chromosome. Each column of small boxes represents the same section of DNA in a different individual per genotype. Each row of yellow or blue boxes represents a single SNP. The blue boxes in each row represent the major allele for that SNP, and the yellow boxes represent the minor allele. The absence of a box at any position in a row indicates missing data.

In this block, 26 common SNPs may be identified. They may be arranged in seven different haplotype patterns (5, 4, 4, 3, 2, 1 and 1 genotypes). The four most common patterns include 16 of the 20 chromosomes sampled. The blue and yellow circles indicate the allele patterns of two SNPs (surrounded by a line), which unambiguously distinguish the four common haplotypes in the block.

Reference

In summary

- Strategies are continuously being developed to improve the detection of polymorphisms
- DNA sequencing allows the detection of variation at even the nucleotide level
- ESTs are powerful tools for detecting diversity within coding regions
- Microarrays and DArT make the simultaneous analysis of many loci possible
- SNPs are single-base substitutions between sequences, and represent the most frequent DNA variant
By now you should know

- The different types of DNA variation that can be detected by sequencing, ESTs and SNPs
- The underlying principle of microarrays and DArT
- The advantages and disadvantages of the newest technologies for analysing genetic diversity
Basic references


Complementary technologies

- Final considerations
- Glossary
Using molecular marker technology in studies on plant genetic diversity

V Complementary technologies

- Denaturing gradient gel electrophoresis (DGGE)
- Thermal gradient gel electrophoresis (TGGE)
- Single-stranded conformational polymorphism (SSCP)
- Heteroduplex analysis
- Denaturing high-performance liquid chromatography (DHPLC)
Using molecular marker technology in studies on plant genetic diversity
Contents

- Introduction
- Denaturing gradient gel electrophoresis (DGGE)
- Thermal gradient gel electrophoresis (TGGE)
- Single-stranded conformational polymorphism (SSCP)
- Heteroduplex analysis
- Denaturing high-performance liquid chromatography (DHPLC)
Introduction

- Various technologies serve to show the presence of sequence differences
- They help narrow down the possibilities of what we need to sequence (is there a polymorphism that might be worth sequencing?)
- But they do not identify what the sequence differences are

When looking at differences in the DNA sequence, we need to be able to separate specific DNA segments from a mixture such as from the whole genome.

Electrophoresis separates molecules in an electrical field on the basis of charge, size and shape. If a DNA molecule is cut into small sections and placed in a well at one end (cathode) of an agarose gel, the DNA fragments will move through the gel towards the anode. Their speed will depend on their individual sizes, so they end up forming bands located at different positions in the gel. The bands can then be visualised with ethidium bromide staining, which causes the DNA to fluoresce under UV light.

The same result is achieved by electrophoresis in an acrylamide gel, the difference being a matter of resolution. The acrylamide is able to discriminate smaller differences in fragment size.

Electrophoresis is a diagnostic procedure that allows us to identify molecules of different sizes. When used as such, electrophoresis is itself a means of showing polymorphisms and, consequently, genetic variation between genotypes.

But electrophoresis can also be useful as a first step towards identifying and isolating specific DNA molecules that, even if the same size, differ in sequence composition.
Denaturing gradient gel electrophoresis (DGGE) (1)

- Permits detecting very small DNA polymorphisms or mutations
- Can be applied to long DNA fragments, measuring hundreds of base pairs
- Does not require previous knowledge on the existence of a polymorphism
- Based on the renaturation properties of DNA strands
- DNA denaturation takes place through electrophoresis
Denaturing gradient gel electrophoresis (2)

- Small DNA fragments are subject to electrophoresis through a polyacrylamide gel under increasingly denaturing conditions (formamide/urea concentrations)
- The DNA 'melts' and becomes single stranded
- DNA molecules change their shape and stop moving
- Differences in DNA sequence are identified

The DNA fragments migrate first as double-stranded molecules. Later, because of the gel's changing composition, the molecules denature and become single stranded, forming a branched structure. This changed structure results in the molecules' diminished ability to move through the gel.

The point at which the DNA melts depends on the nucleotide sequence in the melted region. The final location of the molecules in the gel thus depends on the nucleotide sequence of the fragments.
Denaturing gradient gel electrophoresis (3)

DGGE can be a two-step process:

- Samples are run to separate fragments according to size
- Fragments are separated under DGGE conditions by the ‘melting point’ process

When samples are mixed, double bands indicate sequence differences between bands of the same size.

At least 95% of differences in sequence composition is estimated as being detected with this procedure.

DGGE also serves to distinguish homozygous versus heterozygous genotypes for a particular DNA fragment. To take advantage of this capacity, a cycle of denaturation and renaturation must be conducted after the last PCR cycle. Homoduplexes and heteroduplexes are formed as alleles reassociate. In the DGGE gel, fast-migrating homoduplex combinations will indicate homozygous genotypes. Heterozygous genotypes will show both homoduplex and heteroduplex combinations. Heteroduplexes are formed through mispairing and rapid denaturation in the gel, which will stop the migratory course of these molecules.
Thermal gradient gel electrophoresis (TGGE)

- Similar to DGGE
- Increasingly denaturing conditions are achieved by a temperature gradient instead of by changing reagent concentrations
- Can also be used for analysing single-stranded RNA and proteins
Basic reference

SSCP can distinguish between two very similar DNA sequences only on the basis of the particular shape of their single-stranded structures. In principle, then, even two alleles of the same gene can be discriminated.
SSCP (2)

- The sequence of interest goes through PCR
- Next, the PCR product is denatured at 94°C and rapidly cooled down on ice. Single-stranded molecules do not pair but form stable secondary structures
- The reassociated fragments are then subject to electrophoresis:
  - In an homozygous case, two bands can be observed, each corresponding to a slightly different secondary structure
  - In a heterozygous case, at least four bands can be observed

SSCP is a simple technique, but has at least two major disadvantages:

- The electrophoretic behaviour of the single-stranded molecules is unpredictable, depending very much on temperature and running conditions.

- In the case of long DNA fragments (> 200 bp), the method becomes insensitive to some mutations. In principle, SSCP seems to work better for small insertions and/or deletions.
Basic reference

Heteroduplex analysis

- Two PCR-amplified products are mixed in equal quantities, denatured at 95°C and allowed to cool.
- During cooling, DNA strands reanneal to form heteroduplex DNA.
- Any mismatches in the heteroduplex will cause it to have a different three-dimensional structure, with a reduced mobility that is proportional to the degree of divergence of the sequences.
Basic reference

Denaturing high-performance liquid chromatography (DHPLC): methodology

- This method can detect sequence differences of a single base pair as well as insertions and/or deletions
- Works with crude PCR products and does not require prior DNA sequencing
- Based on the differential elution of homoduplex and heteroduplex DNA when run through a column

DHPLC is a high-performance liquid chromatography method in which DNA fragments are separated according to size and/or presence of heteroduplexes (reannealed DNA strands) during their passage through a gradient in a column.

In double-stranded amplified DNA, nucleotides that are mismatched through mutations and polymorphisms become evident after heteroduplex formation. The presence of these polymorphisms creates a mixed population of heteroduplexes and homoduplexes during reannealing of wild type and mutant DNA. If this mixture of fragments is run under partially denaturing conditions by HPLC, heteroduplexes elute from the column earlier than the homoduplexes because of their lower melting temperature.

Analysis can be performed to detect sequence variation between individuals or determine heterozygosity.
DHPLC: applications

- Finding new mutations and polymorphisms in any DNA fragment or particular gene sequence
- Screening clones to identify candidate fragments for sequencing
- Evaluating the fidelity of amplified fragments
- Diagnosing for the presence of known mutations
Basic reference

In summary

- Several technologies help identify the presence of sequence differences

- While they cannot tell what the variant is, they can help narrow the range of strategies to use for detecting it
By now you should know

The basic principles of:

- Denaturing gradient gel electrophoresis
- Thermal gradient gel electrophoresis
- Single-stranded conformational polymorphism
- Heteroduplex analysis
- Denaturing high-performance liquid chromatography
Final considerations

- Glossary
Using molecular marker technology in studies on plant genetic diversity

VI Final considerations
• Considerations when choosing a technique
• Practical applications
Using molecular marker technology in studies on plant genetic diversity

Final considerations
Contents

- When choosing a technique...
- Practical applications

Copyright: IPGRI and Cornell University, 2003

Final considerations  2
When choosing a technique...

Ask the biological question first:

- What is the problem?
- How many loci and/or alleles are required?
- At what level is discrimination being sought?
- Is the mode of marker inheritance important?

What is the problem being addressed?
This is the most important question. The first step is to know exactly what is the biological question one wants to answer with the research. This is essential for choosing the right technique. For instance, for information on population history or phylogenetic relationships, sequence data or restriction site data should be used.

The number of loci and/or alleles required
Will information from a few loci be sufficient or is greater genome coverage required? Allozymes are limited. AFLP detect high numbers of loci. Where hypervariability is required, the best techniques are those based on single-locus, simple-sequence repeats (e.g. SSR).

Discrimination level
At what taxonomic level is the genetic variation being measured: within populations, between species or between genera? Is the selected method appropriate for detecting the desired level of variation?

Mode of inheritance
Should both homozygotes and heterozygotes be identified? Are codominant markers needed (single-locus RFLPs, allozymes, PCR-amplified microsatellites) or will dominant markers suffice (RAPD, AFLP)? If presence versus absence information is sufficient, then any molecular marker technology can be used; but if information about heterozygotes is needed (e.g. population and diversity structure, knowledge on type of inheritance), then only codominant markers such as isozymes or microsatellites should be used.
When choosing a technique... (continued)

Resources:

- Is good quality DNA important?
- Is the right expertise available?
- Well-equipped laboratory

Costs:
- Equipment
- Consumables

Speed

DNA availability
RFLP analysis requires large amounts of DNA. Most PCR-based methods require only tiny quantities of easily prepared DNA. In many cases, PCR is performed only to amplify the original amount of target DNA.

Expertise required
Techniques involving hybridisation or manual sequencing are technically more demanding. RAPDs or SSRs (once the primers are available) are the least demanding.

Availability of laboratory facilities and equipment
Once the biological questions are set, technical and organisational criteria become important in deciding the technology of choice. For example, having (1) access to a suitably equipped laboratory; (2) money to purchase additional equipment and consumables when necessary; (3) a good grasp of many basic laboratory skills; and (4) a basic knowledge of how to set up an experiment.

Costs
In terms of costs, allozymes are the cheapest; RAPD, RFLP and even AFLP are intermediate, with sequencing being still more expensive. The costs of all types of experiments should be considered, because lack of reproducibility of some markers may, in the end, result in higher costs.

For required skills, an extended visit to another laboratory where the relevant techniques are being used often provides invaluable information for setting up the research. Having relevant contacts who may help in those first steps may also prove invaluable.

Speed
How quickly are data needed, and how much time will the equipment allow? PCR-based methods certainly give fast results when primers are available. Hybridisation-based methods are slower. Conventional DNA sequencing is slow, whereas automated sequencing is faster.
When choosing a technique... (continued)

Additional matters:

- Reproducibility
- PCR versus non-PCR techniques
- Latest strategies

Reproducibility
Are robust methods required? For example, will the markers be exchanged? Is more than one laboratory involved? If so, allozymes, RFLPs, SSRs and sequencing are robust, whereas RAPD is not.

PCR versus non-PCR techniques
PCR-based molecular marker techniques open up numerous possibilities and could be considered first, because of their simplicity. Hybridisation-based techniques are more labour intensive, more demanding technically and require different equipment.

- RAPD is an excellent technique by which to become familiar with PCR. It allows rapid examination of polymorphisms in most, if not all, species of interest, and primers are readily available.

- Other PCR-based markers such as SSR could be applied relatively easily, if primers are already available. More and more, this is the case for many species. Strategies for searching appropriate primers are also improving, and some approaches for searching putative microsatellites rely on sequence databases, circumventing the problem of having to make and screen libraries in the laboratory.

- AFLPs have become a very popular option, although their need for a double PCR and vertical gel electrophoresis makes them more expensive and technically more demanding.

Latest strategies
Costs for sequencing experiments have significantly decreased. Many ESTs are already available for several species. Microarrays, based on either anonymous genomic characterisation or gene expression, are becoming common. Microarray technology is still very demanding, technically and in terms of equipment. Before deciding on it, get acquainted with the techniques, requirements and outputs. A better option might be to consider outsourcing sample analysis and concentrate, instead, on interpreting results and the subsequent decision-making. SNPs are being routinely used in human studies. They are still too expensive for standard applications to genetic diversity studies in plants. Nevertheless, they look at the ultimate level of variation in the DNA sequence the nucleotides, and may well be the future's best molecular marker option when their costs of discovery and application decrease.
Practical applications: genetic diversity studies

- Genetic relatedness and diversity: population genetics
- Studying polymorphism in landraces and cultivars
- Identification of cultivars and taxonomy
- Phylogenetic studies
- Studying domestication and evolution
- Gene flow and introgression
- Comparative mapping

Some of the preceding submodules described real experiments where molecular techniques were applied to answer questions on genetic diversity. Reference lists were also given for the corresponding technology.
Practical applications: germplasm management

- Taxonomic characterisation of germplasm
- Maintenance of collections:
  - Identifying gaps
  - Identifying duplicates
  - Development core collections
  - Assessing stability of conserved material
  - Measuring genetic erosion
- Development conservation strategies

Molecular markers may be used in genebank management to:

- Accurately identify germplasm
- Screen germplasm for use by breeders and other researchers conduct routine maintenance of the genebank, which will be streamlined by identifying duplicates, assessing stability through different rounds of regeneration or multiplication and measuring genetic erosion
- Identify gaps in the collection to plan for future conservation and collection activities and for developing core collections where other data will be complemented by ensuring that the allelic richness of a core will be maximised.

Likewise, molecular data can be used to define conservation strategies, both ex situ (e.g. collecting strategies) and in situ.
Molecular technologies may help promote germplasm use by providing exact data about the genotypic attributes of plants, including crops. Germplasm characterisation offers information about individual genomic composition and, as such, allows breeders to select promising material based on genotype, as well as on phenotype. The construction of molecular linkage maps have opened up the possibility of locating important agronomic traits in crop genomes and, consequently, of selecting germplasm based on the presence of a particular gene of interest. Introgression of genes from ‘donor’ germplasm can thus be followed in subsequent generations, using so-called marker-assisted selection, thus facilitating and accelerating traditional selection trials.

Molecular marker technologies are also used to detect somaclonal variation—which may be useful for breeding—that sometimes occurs after regeneration through tissue culture. They can also help in the routine housekeeping activities of a breeding program, such as keeping track of progenies through pedigree analysis, identifying off-types in seed lots and confirming or disproving hybrid purity.

The latest tools for molecular genetics will, hopefully, speed up breeding procedures through such activities as permitting the quick discovery of useful genes in germplasm collections or correlating genotype with phenotype.
In summary

- The most important criteria for choosing a molecular technique for a genetic diversity study are:
  - The biological question driving the study
  - The resources available versus those required

- Applications of molecular technologies cover all the different aspects of genetic diversity analysis, germplasm management and germplasm use
By now you should know

- The criteria that will help you succeed in selecting and applying molecular technologies to the plant genetic resources of interest
Using molecular marker technology in studies on plant genetic diversity

Glossary

**Accession**: A sample of a crop variety collected at a specific location and time; may be of any size.

**AFLP**: Amplified fragment length polymorphism. A highly sensitive method for detecting polymorphisms in DNA. DNA first undergoes restriction enzyme digestion, then a subset of DNA fragments is selected for PCR amplification and visualisation.

**Allele**: One of the alternative forms of a gene that can exist at a single locus.

**Allele frequency**: A measure of the commonness with which an allele is found in a population, or its proportional share of all alleles of a gene.

**Allozyme**: An isozyme whose synthesis is controlled by codominant alleles of one gene.

**Amino acids**: Bi-functional organic compounds that contain a basic amino group (-NH$_2$) and an acidic carboxyl group (-COOH).

**Annealing**: Spontaneous alignment of two single DNA strands to form a double helix.

**AP-PCR**: Arbitrarily primed polymerase chain reaction. A technique for amplifying anonymous stretches of DNA, using PCR. Related to RAPD.

**Apomixis**: The production of an embryo in the absence of meiosis. Apomictic higher plants produce asexual seeds, derived only from maternal tissue.

**Arbitrary primer**: A short oligonucleotide primer used in certain PCR methods to initiate DNA synthesis at random locations on the target DNA.

**Autogamy**: Transfer of pollen from the anther of a flower to the stigma of the same flower or, sometimes, to that of a genetically identical flower (as of the same plant or clone). Ability of many plant species to naturally and successfully fertilise within one individual. Also called self-pollination.

**Autoradiography**: A technique where radioactively labelled molecules are visualised through exposure to X-ray film.

**Bacteriophage**: A virus that infects bacteria. Genetically altered forms are used as vectors in cloning DNA.

**Base**: The chemical unit that characterises a nucleotide. In DNA, the bases found are adenine (A), guanine (G), thymine (T) and cytosine (C). In RNA, the bases are adenine, guanine, uracil (U) and cytosine.
**Base pair:** Two nucleotide bases on different strands of a nucleic acid molecule that are held together by hydrogen bonds. Bases can pair in only one way: adenine with thymine (DNA) or uracil (RNA), and guanine with cytosine (DNA). *See also* Complementary sequence

**Base sequence:** The order of nucleotide bases in DNA or RNA.

**Bottleneck:** A brief reduction in population size that usually leads to random genetic drift.

**CAPS:** Cleaved amplified polymorphic sequence, also known as PCR-RFLP, a technique for detecting polymorphisms at a particular locus. A locus undergoes PCR amplification and polymorphisms are detected by differences in restriction fragment sizes between individuals.

**cDNA or complementary DNA:** DNA transcribed from an RNA molecule by an enzyme called reverse transcriptase.

**Centromere:** The constricted region of a chromosome to which spindle fibres attach during cell division.

**Character:** Or trait, an attribute of individuals within a species for which various heritable differences can be defined.

**Chromatin:** Complex of DNA and protein of the interphase cell.

**Chromosome:** A linearly continuous arrangement of genes and other DNA, and associated proteins and RNA.

**Cloning:** In molecular biology: the process of using DNA manipulation procedures to produce multiple copies of a single gene or DNA segment.

**Codominance:** The situation in which a heterozygous individual exhibits the phenotypes of both alleles of a particular gene.

**Complementary sequence:** The sequence of a DNA or RNA strand to which a given nucleotide sequence can bond to form a double-stranded structure, e.g. TAGGAT is the complementary sequence to ATCCTA where A = adenine, C = cytosine, G = guanine and T = thymine. *See also* Base pair.

**cpDNA:** Chloroplast DNA.

**Cytoplasmic inheritance:** Inheritance of the genes found in cytoplasmic organelles (viz. chloroplasts or mitochondria).

**DAF:** DNA amplification fingerprinting. A technique for amplifying anonymous stretches of DNA, using PCR. Related to RAPD.

**DArT:** Diversity array technology.

**Deletion:** A particular kind of mutation involving the loss of some DNA from a chromosome.

**Dendrogram:** Any branching diagram that shows, by means of lines shaped like U's a hierarchy of categories or objects based on the degree of similarity or number of shared characters. Often, the length of each U represents the distance between the two objects being connected.

**Denaturation:** The separation of the two strands of the DNA double helix, or the severe disruption of a complex molecule without breaking the major bonds of its chains.
DGGE: Denaturing gradient gel electrophoresis. A method for separating DNA fragments according to their mobility under increasingly denaturing conditions (usually increasing formamide/urea concentrations). See also heteroduplex analysis, SSCP and TGGE.

DHPLC: Denaturing high-performance liquid chromatography. This method can detect sequence variations of a single base pair.

DNA: Deoxyribonucleic acid, a double chain of linked nucleotides (having deoxyribose as the sugar component), which is the fundamental molecule of which genes are composed.

DNA fingerprint: A unique pattern of DNA fragments as revealed by Southern hybridisation or by PCR.

DNA polymerase: Any enzyme with the ability to synthesise new DNA strands, using a DNA template.

DNA sequence: The order of nucleotide bases in the DNA molecule.

Dominant allele: An allele that expresses its phenotypic effect even when heterozygous with a recessive allele (qv). That is, if A is dominant over a, then AA and Aa have the same phenotype.

Double helix: The structure of DNA first proposed by Watson and Crick, with two linked helices joined by hydrogen bonds between paired bases.

Ecotype: A population or strain of an organism that is adapted to a particular habitat.

Electrophoresis: A technique for separating the components of a mixture of molecules (proteins, DNA or RNA) by size as a result of an electric field within a support gel.

Enzyme: A protein that functions as a catalyst of biochemical reactions.

EST: Expressed sequence tag. A small part of the active part of a gene, made from cDNA. It can be used as a marker, to search the rest of the gene or to locate it in a larger segment of DNA.

Eukaryote: A cell or organism with a distinct, membrane-bound nucleus and other differentiated subcellular components.

ex situ conservation: (1) A conservation method that entails the removal of germplasm resources (seed, pollen, sperm, individual organisms) from their original habitat or natural environment. (2) Keeping components of biodiversity alive outside their original habitat or natural environment. Cf. in situ conservation.

F₁, F₂, F₃, …: A shorthand notation used to denote the different generations involved in breeding experiments. F₁ is the first filial generation, that is, the progeny of the parental cross; F₂ is the second filial generation, that is, the progeny of self-fertilising or intercrossing F₁ individuals, and so on.

Flanking regions: The DNA sequences extending on either side of a specific gene or locus.

Gene: The basic physical and functional unit of heredity, which passes information from one generation to the next. It is a segment of DNA that includes a transcribed section and a regulatory element that allows its transcription.

Genebank: A facility established for the ex situ conservation of individuals (seeds), tissues, or reproductive cells of plants or animals.
Gene flow: The exchange of genes between different but (usually) related populations.

Gene mapping: The determination of the relative positions of genes on a chromosome or plasmid and the distance between them.

Genepool: The sum total of genes, with all their variations, possessed by a particular species at a particular time.

Genetic drift: Change in allele frequency from one generation to another within a population, due to the sampling of finite numbers of genes that is inevitable in all finite-sized populations. The smaller the population, the greater is the genetic drift, with the result that some alleles are lost, and genetic diversity is reduced.

Genetic linkage: The proximity of two or more genes on a chromosome so that they tend to be inherited together.

Genetic marker: An allele, a band in a gel or trait that serves experimentally as a probe to identify an individual or one of its characteristics.

Genetic resources: The genes found in plants and animals that are of actual or potential value to people.

Genome: The entire complement of genetic material in an organism.

Genotype: The specific allele composition of either of the entire cell or, more commonly, of a certain gene or set of genes.

Germplasm: The total genetic variability available to a population of organisms as represented by germ cells, seeds, etc.

Haplotype: A specific allelic constitution at a number of loci within a defined linkage block.

Heredity: The process by which genetic traits are passed from parents to offspring.

Heteroduplex: A double-stranded DNA molecule or a DNA/RNA hybrid where each strand is from a different source.

Heteroduplex analysis: The study of the mobility of heteroduplex DNA under polyacrylamide gel electrophoresis. The reduced mobility of heteroduplex DNA compared with homoduplex DNA is proportional to the degree of divergence of the sequences. See also DGGE, SSCP and TGGE.

Heterozygous gene pair: A gene pair having two different alleles in the two chromosome sets of a diploid individual, for example, A a or A 1 A 2.

Homoduplex DNA: A double-stranded DNA molecule where both strands are from the same source.

Homologous: Corresponding or alike in structure, position or origin.

Homozygous gene pair: A gene pair having identical alleles in both copies of the chromosome set, for example, AA or aa.

Hybrid: Either (1) a heterozygous individual, or (2) a progeny individual from a cross between parents with different genotypes.
Hybridisation: In molecular biology: the binding of complementary DNA and/or RNA sequences to form a double-stranded structure.

Insertion: A type of chromosomal abnormality in which a DNA sequence is inserted into a gene, disrupting the normal structure and function of that gene.

\textit{in situ} conservation: A conservation method that attempts to preserve the genetic integrity of gene resources by conserving them within the evolutionary dynamic ecosystems of the original habitat or natural environment. Cf. \textit{ex situ} conservation

Isozyme: Multiple forms of an enzyme whose synthesis is controlled by more than one gene.

ISSR: Inter-simple sequence repeat. ISSR primers are anchored at their 3' ends to direct the amplification of the genomic segments between the ISSRs.

Landrace: A crop cultivar or animal breed that has evolved with and has been genetically improved by traditional farmers without influence from modern breeding practices.

Library: A collection of DNA clones obtained from one DNA donor.

Ligase: A type of enzyme that can rejoin a broken phosphodiester bond in a nucleic acid.

Ligation: The process of joining two or more DNA fragments together.

Locus (pl. loci): The specific place on a chromosome where a gene or particular piece of DNA is located.

MAAP: Multiple arbitrary amplicon profiling. A collective term for PCR techniques using arbitrary primers.

Marker: An identifiable physical location on a chromosome whose inheritance can be monitored (e.g. gene, restriction enzyme site or RFLP marker).

Melting temperature (Tm): Midpoint of the temperature range over which DNA is denatured.

Microarray: Small spots of DNA fixed to glass slides or nylon membranes. This technology is based on the hybridisation between short oligonucleotide probes and complementary DNA sequences.

Microsatellite DNA: A type of repetitive DNA based on very short repeats such as dinucleotides, trinucleotides or tetranucleotides. See also Repetitive DNA.

Minisatellite DNA: A type of repetitive DNA sequence based on short repeats with a unique common core. See also Repetitive DNA.

mtDNA: Mitochondrial DNA.

Multiplexing: Simultaneously performing several different reactions in the same reaction tube to increase efficiency.

Mutation: A permanent structural alteration in DNA. In most cases, DNA changes either have no effect or cause harm, but occasionally a mutation can improve an organism's chance of surviving and passing on the beneficial change to its descendants.
Nitrogen bases: Molecules that are important components of nucleic acids, composed of nitrogen-containing ring structures. Hydrogen bonds between bases link the two strands of the DNA double helix.

Nuclease: An enzyme that cleaves phosphodiester bonds, which link adjacent nucleotides in DNA and/or RNA. An exonuclease progressively cleaves from the end of the substrate molecule; an endonuclease cleaves at internal sites within the substrate molecule.

Nucleotide: A molecule composed of a nitrogen base, a sugar and a phosphate group. Nucleotides are the building blocks of nucleic acids.

Oligonucleotide: A short segment of DNA that is synthesised artificially.

Pedigree: A simplified diagram of a family's genealogy that shows family members' relationships to each other and how a particular trait or disease has been inherited.

PCR: Polymerase chain reaction. A method for amplifying a DNA sequence in large amounts, using a heat-stable polymerase and suitable primers to direct the amplification of the desired region of DNA.

PCR-RFLP: Alternative name for the technique known as 'cleaved amplified polymorphic sequence' or CAPS qv.

Peptide: Two or more amino acids joined by a peptide bond.

Phenotype: Either (1) the form taken by a trait (or group of traits) in a particular individual; or (2) the detectable external appearance of a specific genotype.

Plasmid: An extra chromosome molecule of DNA that is able to replicate autonomously.

Point mutation: A change in a single base pair of DNA.

Polymer: A molecule having repeated subunits.

Polymerase: General term for enzymes that carry out the synthesis of nucleic acid, using a pre-existing nucleic acid template and appropriate nucleotides (viz. ribonucleotides for RNA and deoxyribonucleotides for DNA).

Polymorphism: The appearance of different forms associated with various alleles of one gene or homologous of one chromosome.

Polypeptide: A protein, which is a chain of linked amino acids.

Primer: A short DNA or RNA fragment annealed to a singled-stranded DNA and to which further nucleotides can be added by DNA polymerase.

Probe: A finite nucleic acid piece that can be used to identify specific DNA segments bearing its complementary sequence.

Protein: A polymer of amino acids joined by peptide bonds and which may comprise two or more polypeptide chains.

RAPD: Random amplified polymorphic DNA. A technique for amplifying anonymous stretches of DNA, using PCR with arbitrary primers.
Recessive allele: An allele whose phenotypic effect is not expressed in the heterozygous state, but is masked by the dominant allele (*qv*).

Recombination: Also known as crossing over. The production of a DNA molecule with segments derived from more than one parental DNA molecule. In eukaryotes, this is achieved by the reciprocal exchange of DNA between non-sister chromatids within a homologous pair of chromosomes during prophase of the first meiotic division. Recombination allows the chromosomes to rearrange their genetic material, thereby increasing the potential of genetic diversity.

Repetitive DNA: A stretch of DNA consisting of multiple repeats of a motif. *See also* Microsatellite DNA, Minisatellite DNA and Satellite DNA.

Restriction enzyme: An endonuclease that will recognise a specific target sequence and cut the DNA chain at that point.

Restriction fragment: A DNA fragment that has been cut by a restriction enzyme.

Restriction site: The specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA.

Reverse transcriptase: An enzyme, requiring a DNA primer, that catalyses the synthesis of a complementary DNA strand from an RNA template.

RFLP: Restriction fragment length polymorphism. Variation between individuals as detected by differences in DNA fragment sizes after restriction digestion.

RNA: An organic acid containing repeating nucleotide units of adenine (A), guanine (G), cytosine (C) and uracil (U), whose ribose components are linked with phosphodiester bonds.

Satellite DNA: Very high repetition (1000 to more than 100,000 copies) of a basic motif or repeat unit (commonly 100 to 300 base pairs) that occurs at a few loci on the genome. *See also* Repetitive DNA.

SCAR: Sequence characterized amplified region. A SCAR is a locus representing a single RAPD fragment that has been sequenced. Primers specific to the locus can be designed and used in PCR amplification.

Segregation: Genetically, it refers to the production of two separate phenotypes corresponding to the two alleles of a gene.

Sequencing: The determination of the order of nucleotides in a DNA or RNA molecule, or of the order of amino acids in a protein.


Somaclonal variation: Variation found in vegetative cells dividing mitotically in culture.

Southern blotting: A procedure in which DNA fragments, separated by electrophoresis, are transferred to membrane filters for detecting specific base sequences by radiolabelled complementary probes. Also known as Southern hybridisation.

SSCP: Single-stranded conformational polymorphism. A method for distinguishing between similar sized DNA fragments according to the mobility of the single-stranded DNA under polyacrylamide gel electrophoresis. *See also* DGGE, heteroduplex analysis, and TGGE.
**SSR**: Simple-sequence repeats. *See* Microsatellite DNA.

**SSRP**: Simple-sequence repeat polymorphism. *See* Microsatellite DNA.

**STMS**: Sequence-tagged microsatellite sites. Primers constructed from the flanking regions of microsatellite DNA, and which can be used in PCR reactions to amplify the repeat region.

**Structural gene**: Any gene that codes for a protein.

**STS**: Sequence-tagged site. A general term given to a marker that is defined by its primer's exact location and order of bases.

**Synteny**: Said to occur where all loci are positioned on the same chromosome. Loci may not appear to be linked by conventional genetic tests for linkage but can still be syntenic.

**Tandem repeat**: Multiple copies of the same base sequence on a chromosome.

**Telomeres**: The distal ends of each chromosomal arm, involved in the replication and stability of linear DNA molecules.

**Template**: A molecule that serves as the pattern for synthesising another molecule, e.g. a single-stranded DNA molecule can be used as a template to synthesise the complementary nucleotide strand.

**TGGE**: Thermal gel gradient electrophoresis. A method for separating DNA fragments according to their mobility under increasingly hot denaturing conditions. *See also* DGGE, heteroduplex analysis and SSCP.

**Transcription**: The synthesis of complementary RNA, using a DNA template.

**Transposable element**: A genetic element that has the ability to move from one site on a chromosome to another. Some resemble, and many originate from, retroviruses.

**Vector**: The agent (e.g. plasmid or virus) used to carry a cloned DNA segment.

**VNTR**: Variable number of tandem repeat. A class of polymorphism characterised by the highly variable copy number of identical or closely related sequences.

**Wild type**: The type or form of an organism or gene that occurs most frequently in nature. Often refers to how organisms or genes are found naturally, that is, in the wild, before researchers induced mutations.
Feedback form

The subject of this module is continuously changing, making periodical updating desirable. For this reason we offer the module in electronic format. To help us with the task of updating, we would appreciate receiving relevant comments and requests from users to further improve the module. Your comments on this tool’s usefulness will also be very helpful.

Below we have asked a few guiding questions for you to answer. You may send your answers to us by fax (no.: 57-2 445 0096), by post to your contact in IPGRI or by e-mail to either cdevicente@cgiar.org or tf12@cornell.edu

Questions:

1. Tell us how you use this tool…

2. Do you think the module:
   a. Provides an overview of the basic scientific concepts underlying molecular marker and DNA sequence technologies and their use in plant genetic resources?
      Yes
      No
      Why?
   
   b. Gives information on how to compare and contrast the advantages and disadvantages of each technology so that the most relevant decisions to a specific research situation are made?
      Yes
      No
      Why?
   
   c. Makes available a list of current bibliographical resources for each of the technologies discussed?
      Yes
      No
      Why?
3. Tell us what you liked and what you would change about the following:
   a. Contents
   b. Spectrum of technologies offered
   c. Criteria to select techniques
   d. Structure within the submodules
   e. Drawings, diagrams, pictures
   f. Other

4. What would you add to the module to make it more useful in your particular situation?
   a. Basic references
   b. Examples of applications
   c. Additional information on basic concepts
   d. More technologies:
      i. Old
      ii. New

5. What other topics would you like to see in this module?

6. Was the module, or parts of it, difficult to download? Why?
The **International Plant Genetic Resources Institute (IPGRI)** is an independent international scientific organization that seeks to advance the conservation and use of plant genetic diversity for the well-being of present and future generations. It is one of 16 Future Harvest Centres supported by the Consultative Group on International Agricultural Research (CGIAR), an association of public and private members who support efforts to mobilize cutting-edge science to reduce hunger and poverty, improve human nutrition and health, and protect the environment. IPGRI has its headquarters in Maccarese, near Rome, Italy, with offices in more than 20 other countries worldwide. The Institute operates through three programmes: (1) the Plant Genetic Resources Programme, (2) the CGIAR Genetic Resources Support Programme and (3) the International Network for the Improvement of Banana and Plantain (INIBAP).

Financial support for IPGRI's research is provided by more than 150 donors, including governments, private foundations and international organizations. For details of donors and research activities please see IPGRI's Annual Reports, which are available in printed form on request from ipgri-publications@cgiar.org or from IPGRI's Web site (www.ipgri.cgiar.org).

The **Institute for Genomic Diversity (IGD)** is located at Cornell University in Ithaca, New York, USA. IGD's mission is to develop, transfer, and provide technologies and educational resources to solve problems affecting the conservation of biodiversity and global food security. Specifically, its goals are: to utilize evolutionary and comparative genomics to solve applied problems; to develop and transfer enabling technologies in genomics and bioinformatics; to provide continuing support for national and international efforts in conservation, agriculture, and crop improvement; and to educate and train at all levels (including but not limited to undergraduate and graduate students, visiting scientists, faculty, teachers).

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, IGD 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 organizations.

Mention of a proprietary name does not constitute endorsement of the product and is given only for information.

**Citation:** de Vicente, M.C. and Fulton T. Using molecular marker technology in studies on plant genetic diversity. Illus. Nelly Giraldo. IPGRI, Rome, Italy and Institute for Genetic Diversity, Ithaca, New York, USA.


© IPGRI and Cornell University, 2003

International Plant Genetic Resources Institute (IPGRI)
Via dei Tre Denari 472/a
00057 Maccarese
Rome, Italy

Institute for Genomic Diversity
130 Biotechnology Building
Cornell University
Ithaca, NY 14853
USA

M. Carmen de Vicente, PhD, is Plant Molecular Geneticist at IPGRI Americas Regional Office, Cali, Colombia. Theresa Fulton, PhD, is Director of Outreach at the Institute for Genomic Diversity of Cornell University, Ithaca, NY.