# Using molecular marker technology in studies on plant genetic diversity

## DNA-based technologies

# PCR-based technologies

## Latest strategies

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

## 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)

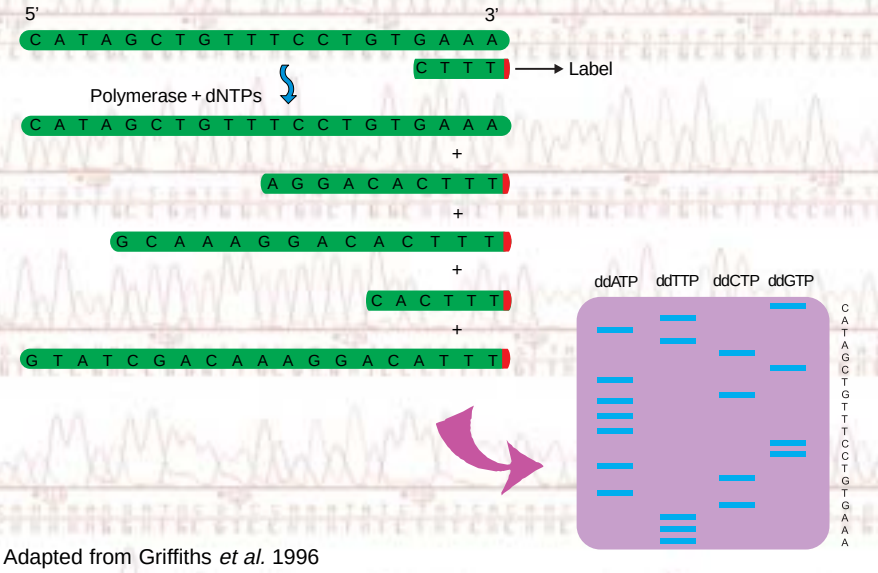



The two methods differ slightly, with the Sanger method (described in slide 6) being easier to automate and, thus, more widely used.

## 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

## A diagram of the Sanger method




Adapted from Griffiths *et al.* 1996





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)

 Latest strategies 8

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

Brown, S.M. and S. Kresovich. 1996. Molecular characterization for plant genetic resources conservation. Pp. 85-93 *in* Genome Mapping in Plants (H. Paterson, ed.). R.G. Landes Company, Georgetown, TX.

# Basic references

Maxam, A.M. and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl Acad. Sci. U.S.A. 74:560-564.

Sanger, F. 1988. Sequences, sequences and sequences. Annu. Rev. Biochem. 57:1-28.

Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl Acad. Sci. U.S.A. 74:5463-5468.

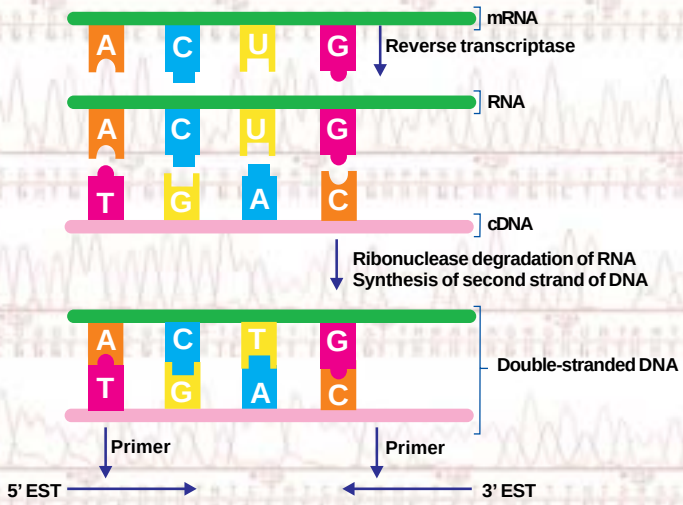
## Expressed sequence tags (ESTs)

- 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




http://www.ostp.gov/NSTC/html/mpgi2001/building.htm

## 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

National Center for Biotechnology Information (NCBI). 2001. ESTs: gene discovery made easier. http://www.ncbi.nlm.nih.gov/About/primer/est.html

## 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

http://bti.cornell.edu/CGEP/CGEP.html





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

## Microarrays: advantages and disadavantages

- 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

Alscher, R. 2001. Grid it: resources for microarray research. http://www.bsi.vt.edu/ralscher/gridit/

Brown, P.O. and D. Botstein. 1999. Exploring the new world of the genome with DNA microarrays. Nature Genet. 21(supp):33-37.

Richmond, T. and S. Somerville. 2000. Chasing the dream: plant EST microarrays. Current Opinion Plant Biol. 3(2):108-116.

## Diversity array technology (DArT)

Two steps are involved:

- Generating the array
- Genotyping a sample

 Latest strategies 20

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/

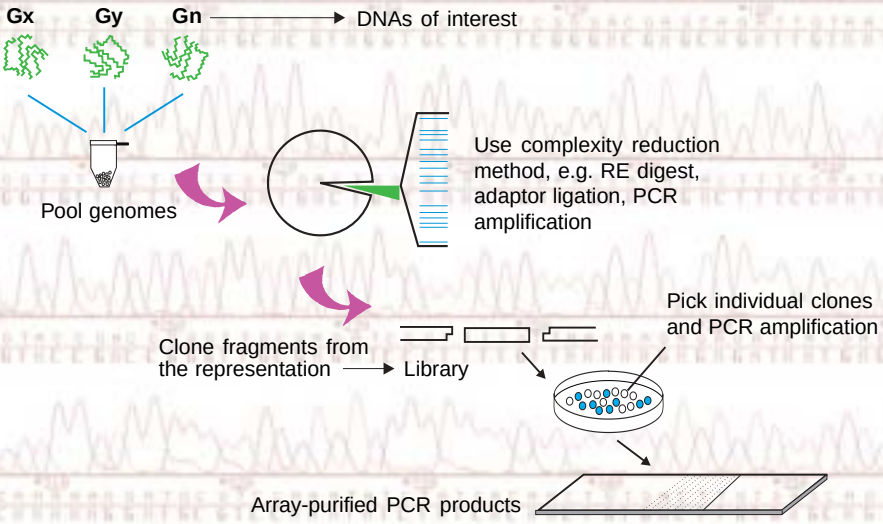
References

CAMBIA. 2000. Enabling innovation. http://www.cambia.org/

## 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

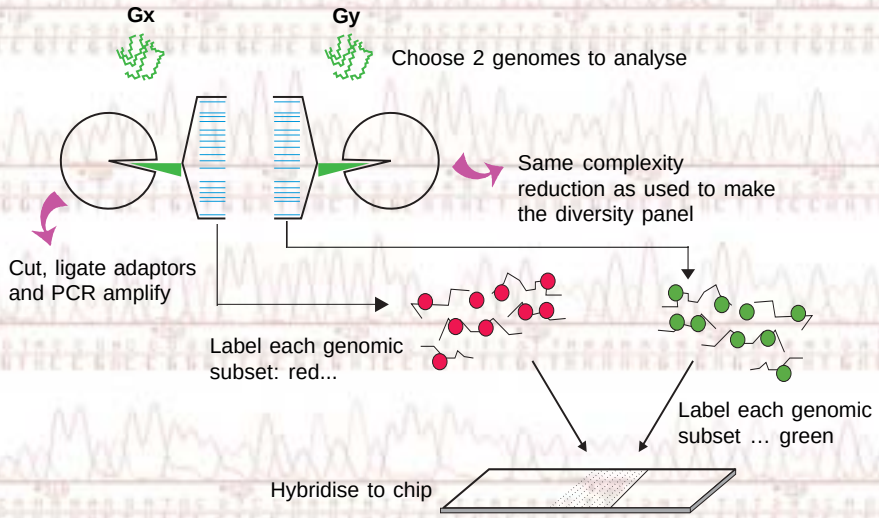
# DArTs: preparing the array (2)



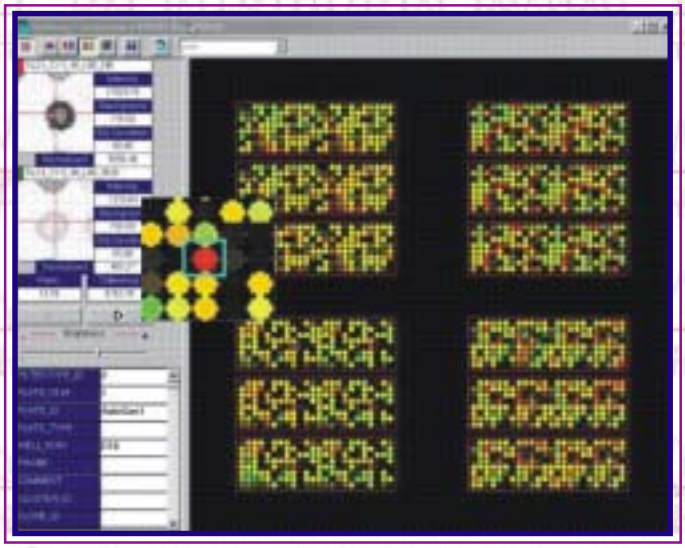
## DArTs: genotyping a sample (1)

- 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)



# 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



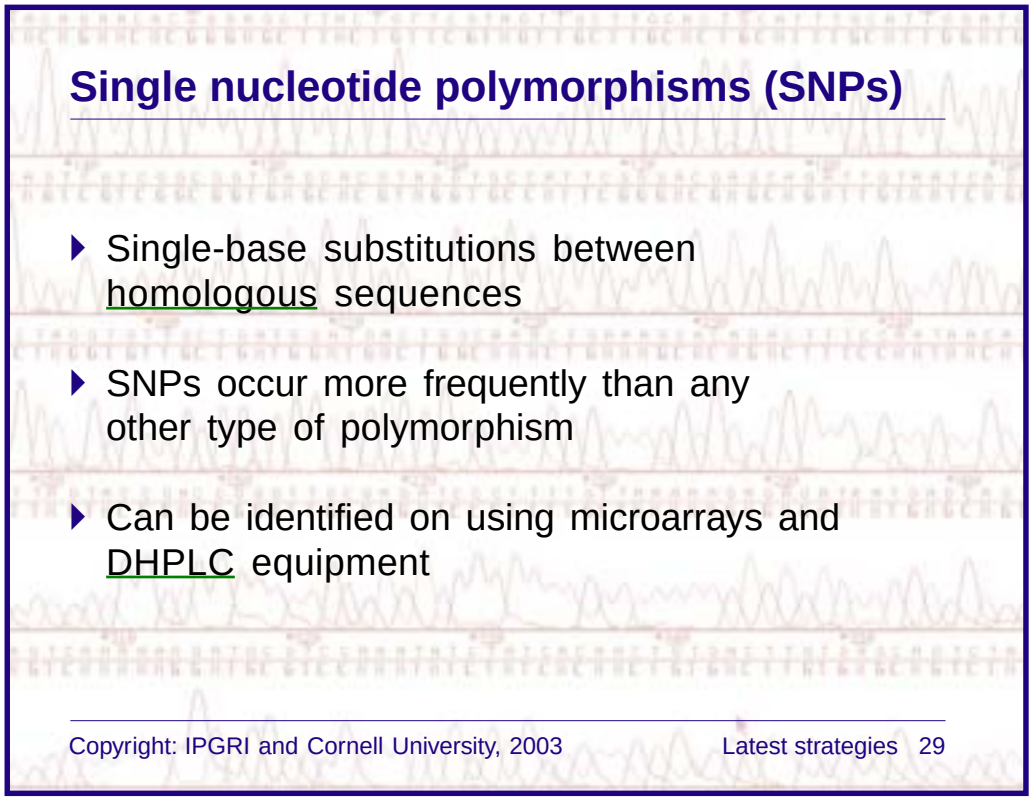

Reference

Jaccoud, D., K. Peng, D. Feinstein and A. Kilian. 2001. Diversity arrays: a solid-state technology for sequence information independent genotyping. Nucleic Acids Res. 29(4):E25.

# Basic References

CAMBIA. 2000. Enabling innovation. http://www.cambia.org/

Jaccoud, D., K. Peng, D. Feinstein and A. Kilian. 2001. Diversity arrays: a solid-state technology for sequence information independent genotyping. Nucleic Acids Research 29 (4): E25.




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.

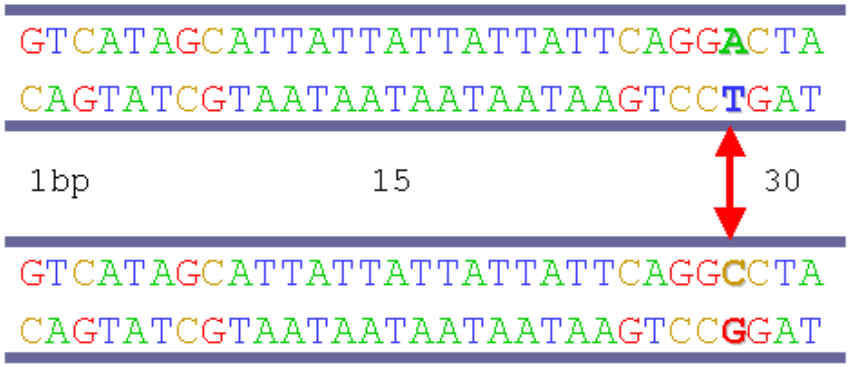
SNPs can be identified by either using microarrays or DHPLC machines.

## DHPLC equipment




DHPLC refers to denaturing high-performance liquid chromatography, which is used to visualise SNPs.

## Interpreting SNPs (1)

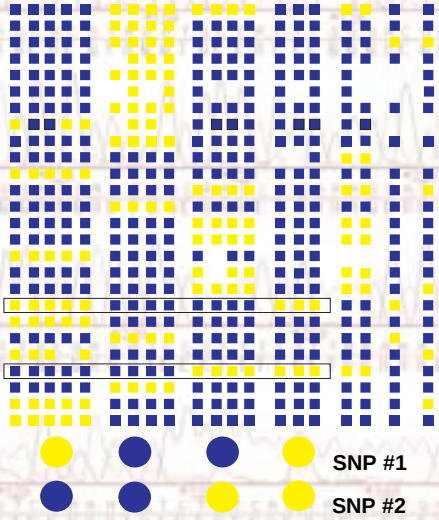



http://bldg6.arsusda.gov/pberkum/Public/sarl/cregan/snps.htm




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).

## Interpreting SNPs (2)




From Patil *et al.*, 2001. Science 294:1719-1723.




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

Patil, N., A.J. Berno, D.A. Hinds, W.A. Barret, J.M. Doshi, C.R. Hacker, and others. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 294:1719-1723.

## 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

Griffiths, A.J.F., J.H. Miller, D.T. Suzuki, R.C. Lewontin and W.M. Gelbart. 1996. An introduction to genetic analysis (6th edn.). W.H. Freeman and Co., NY.

Hajeer, A., J. Worthington and S. John (eds.). 2000. SNP and Microsatellite Genotyping: Markers for Genetic Analysis. Biotechniques Molecular Laboratory Methods Series. Eaton Publishing, Manchester, UK.

USDA-ARS. 1999. The Cregan lab. http://bldg6.arsusda.gov/pberkum/Public/sarl/cregan/snps.htm

Wang, D.G., J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, and others. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280(5366):1077-1082.

## Next

### Complementary technologies

- Final considerations
- Glossary