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

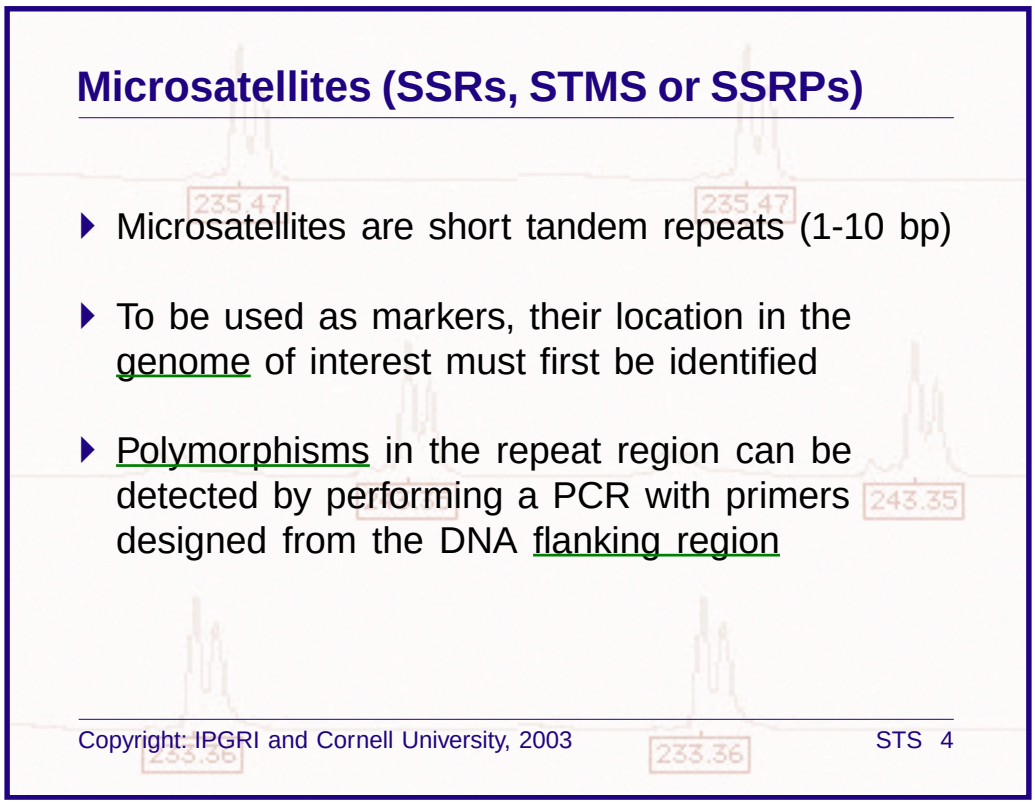
Sequence-tagged sites (STS) as markers

- 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.

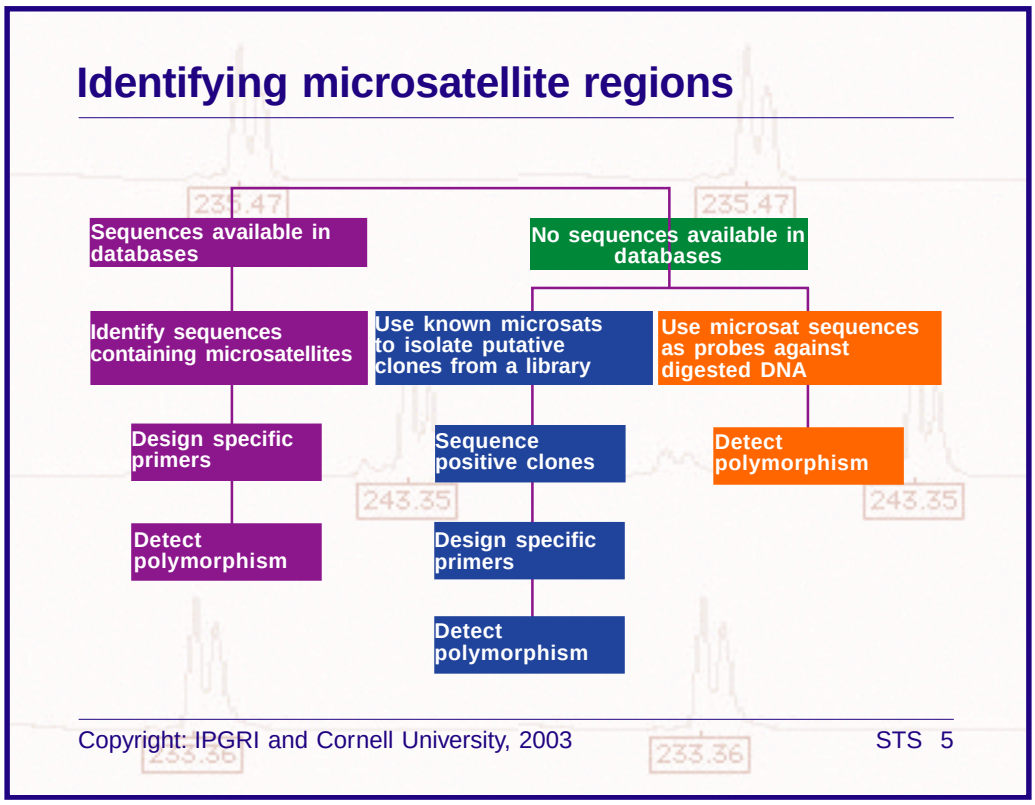


Microsatellites are also called simple sequence repeats (SSRs) and, occasionally, sequence-tagged microsatellite sites (STMS) or simple sequence repeat polymorphisms (SSRPs). They are by far the most widely used type of STS.

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.

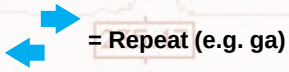
Reference

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.

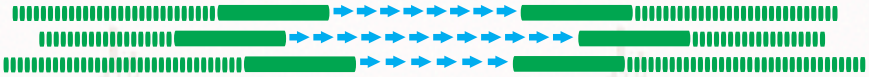


As do areas of the genome high in repeats, SSRs tend to cluster at the centromeres and telomeres. However, this problem can be solved by developing SSRs from EST libraries, which are gene rich and more evenly distributed.

Structure

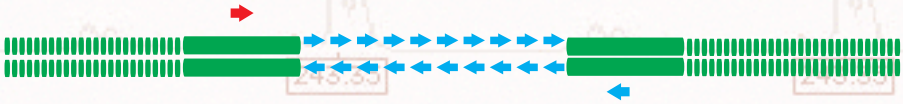


- The number of repeats is highly variable among individuals



Selecting primers

- Design primers ( ) 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

 STS 8

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

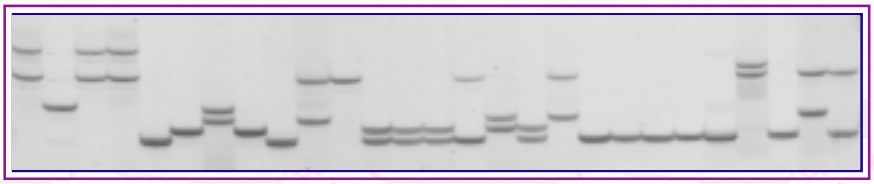
Dean, R.E., J.A. Dahlberg, M.S. Hopkins, S.E. Mitchell and S. Kresovich. 1999. Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with Simple Sequence Repeat (SSR) markers. Crop Sci. 39:1215-1221.

Staining with ethidium bromide

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.

Staining with silver nitrate



Copyright: IPGRI and Cornell University, 2003 STS 10

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.

Equipment

- 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

Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman and J. Doebley. 2002. Microsatellites in *Zea* - variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. 104:436-450.

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.

Cipriani, G., M.T. Marrazzo, R. Marconi, A. Cimato and R. Testolin. 2002. 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.

Dean, R.E., J.A. Dahlberg, M.S. Hopkins, S.E. Mitchell and S. Kresovich. 1999. Genetic redundancy and diversity among 'Orange' accessions in the U.S. national sorghum collection as assessed with Simple Sequence Repeat (SSR) markers. Crop Sci. 39:1215-1221.

Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman and J. Doebley. 2002. Microsatellites in *Zea* - variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. 104:436-450.

Smith, J.S.C., S. Kresovich, M.S. Hopkins, S.E. Mitchell, R.E. Dean, W.L. Woodman, M. Lee and K. Porter. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci. 40:226-232.

Westman, A.L. and S. Kresovich. 1999. Simple sequence repeats (SSR)-based marker variation in *Brassica nigra* genebank accessions and weed populations. Euphytica 109:85-92.

Applications: example of sorghum

- Title:
  Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci. 2000. 40:226-232
- 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

(continued on next slide)

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

Applications: example of black mustard

- 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

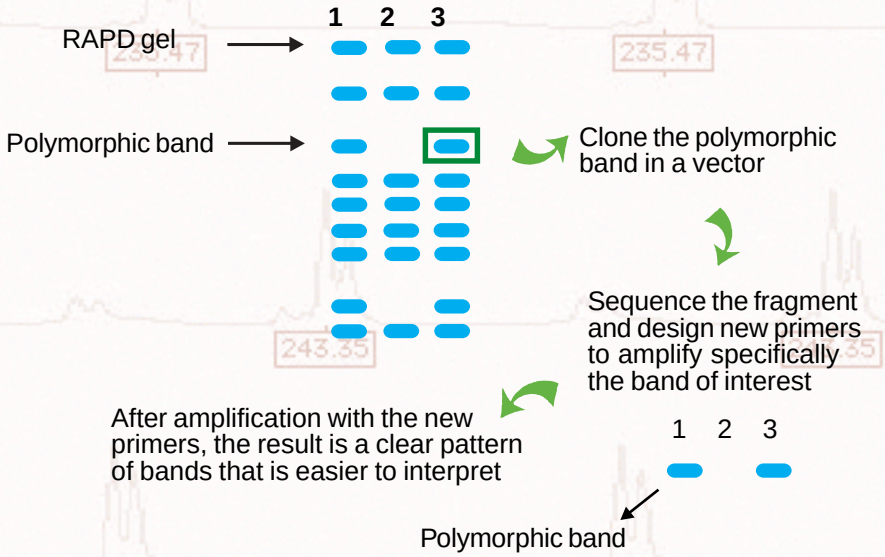
Reference

Paran, I. and R.W. Michelmore. 1993. Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985-993.

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

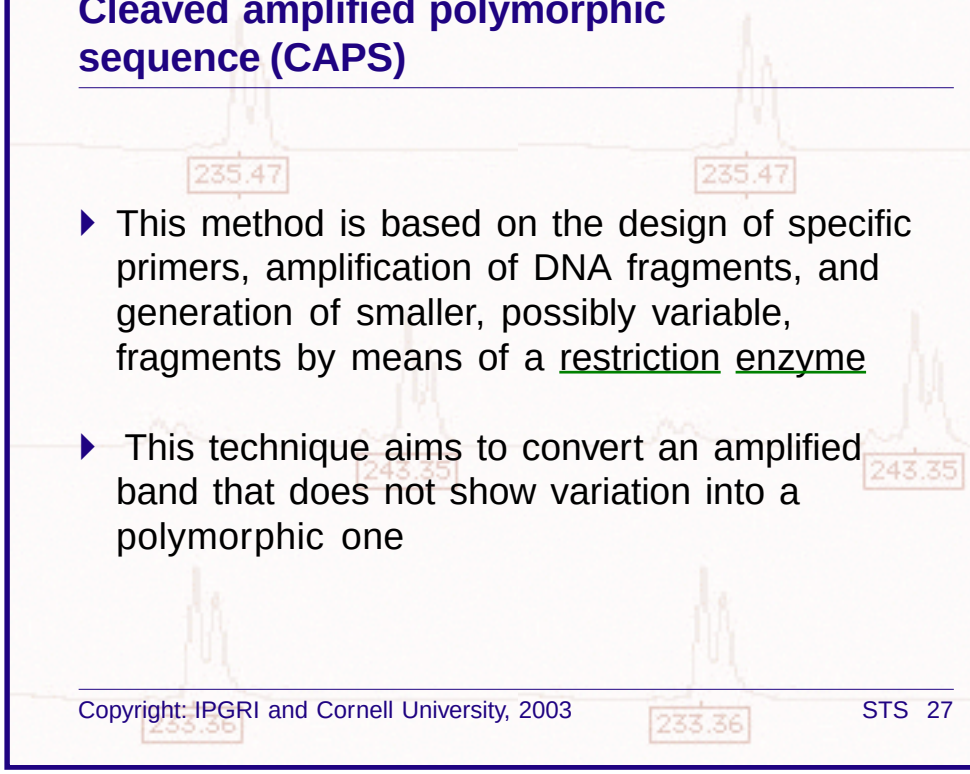


Advantages and disadvantages

- 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

Konieczny, A. and F.M. Ausubel. 1993. A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. Plant J. 4(2):403-410.

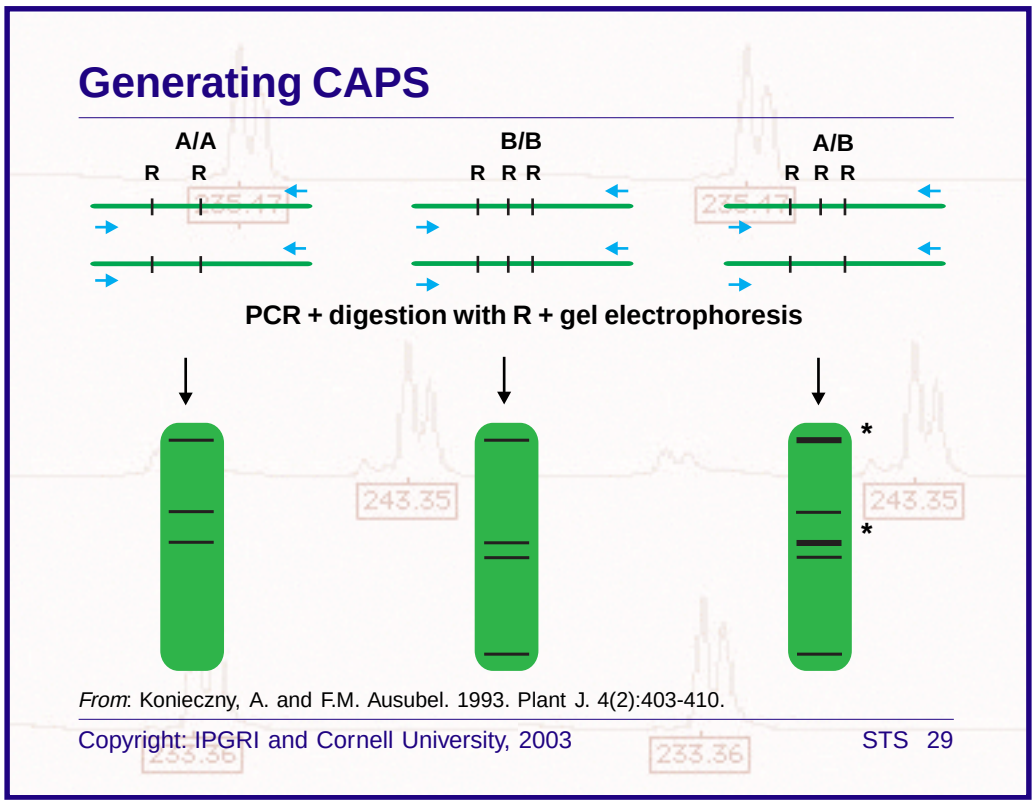
Steps for generating CAPS

- 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

 STS 28

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

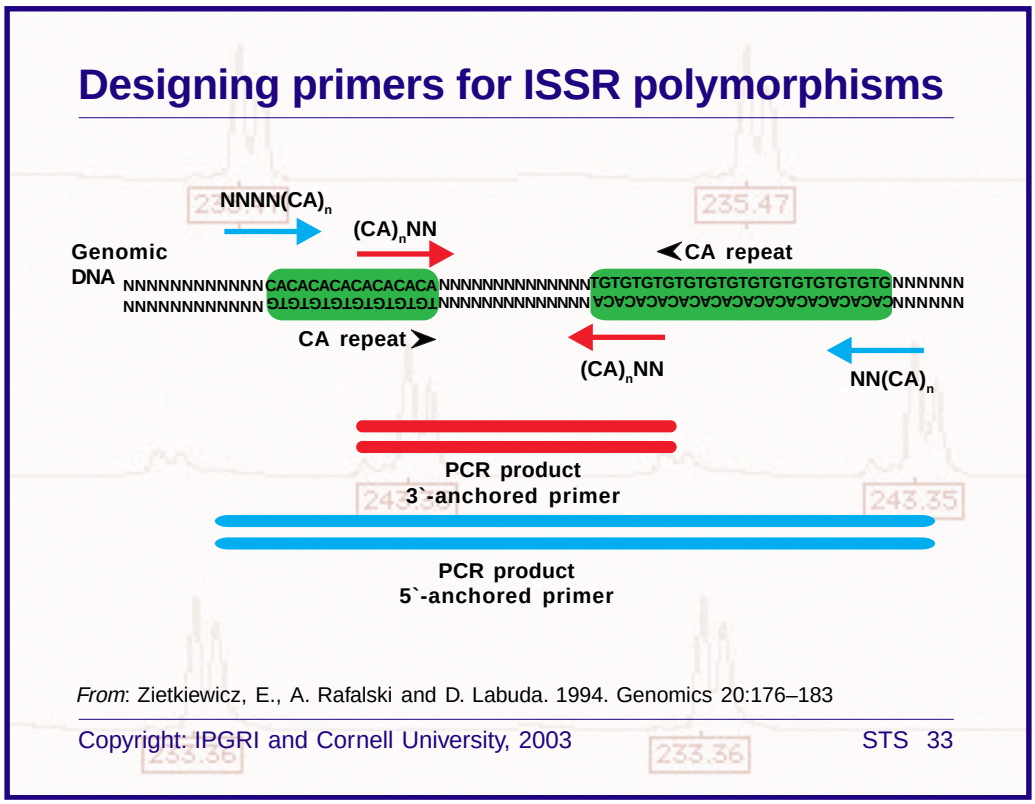
Reference

Zietkiewicz, E., A. Rafalski and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176-183.

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)



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 $(CA)_nNN$ 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.

Advantages and disadavantages

- 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

Ajay, J., C. Apparanda and P.L. Bhalla. 1999. Evaluation of genetic diversity and genome fingerprinting of *Pandorea* (Bignoniaceae) by RAPD and inter-SSR PCR. Genome 42:714-719.

Blair, M.W., O. Panaud and S.R. McCouch. 1999. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). Theor. Appl. Genet. 98:780-792.

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.). RG Landes Company, Georgetown, TX.

Buso, G.S.C., P.H.N. Rangel and M.E. Ferreira. 2001. Analysis of random and specific sequences of nuclear and cytoplasmatic DNA in diploid and tetraploid American wild rice species (*Oryza* sp.). Genome 44:476-494.

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.

Joshi, S.P., V.S. Gupta, R.K. Aggarwal, P.K. Ranjekar and D.S. Brar. 2000. Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theor. Appl. Genet. 100:1311-1320.

Kantety, R.V., X.P. Zeng, J.L. Bennetzen and B.E. Zehr. 1995. Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple sequence repeat (ISSR) amplification. Mol. Breed. 1:365-373.

Konieczny, A. and F.M. Ausubel. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4(2):403-410.

Lowe, A.J., J.R. Russell, W. Powell and I.K. Dawson. 1998. Identification and characterization of nuclear, cleaved amplified polymorphic sequence (CAPS) loci in *Irvingia gabonensis* and *I. wombolu*, indigenous fruit trees of west and central Africa. Mol. Ecol. 7(12):1786-1788.

Morgante, M. and A.M. Olivieri. 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. 3(1):175-182.

Paran, I. and R.W. Michelmore. 1993. Development of reliable PCR based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985-993.

Zietkiewicz, E., A. Rafalski and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20:176-183.

Next

DNA-based technologies

PCR-based technologies

Latest strategies

- Complementary technologies
- Final considerations
- Glossary