

Technical Working Party on Automation and Computer Programs TWC/36/9**Thirty-Sixth Session
Hanover, Germany, July 2 to 6, 2018****Original:** English
Date: June 19, 2018

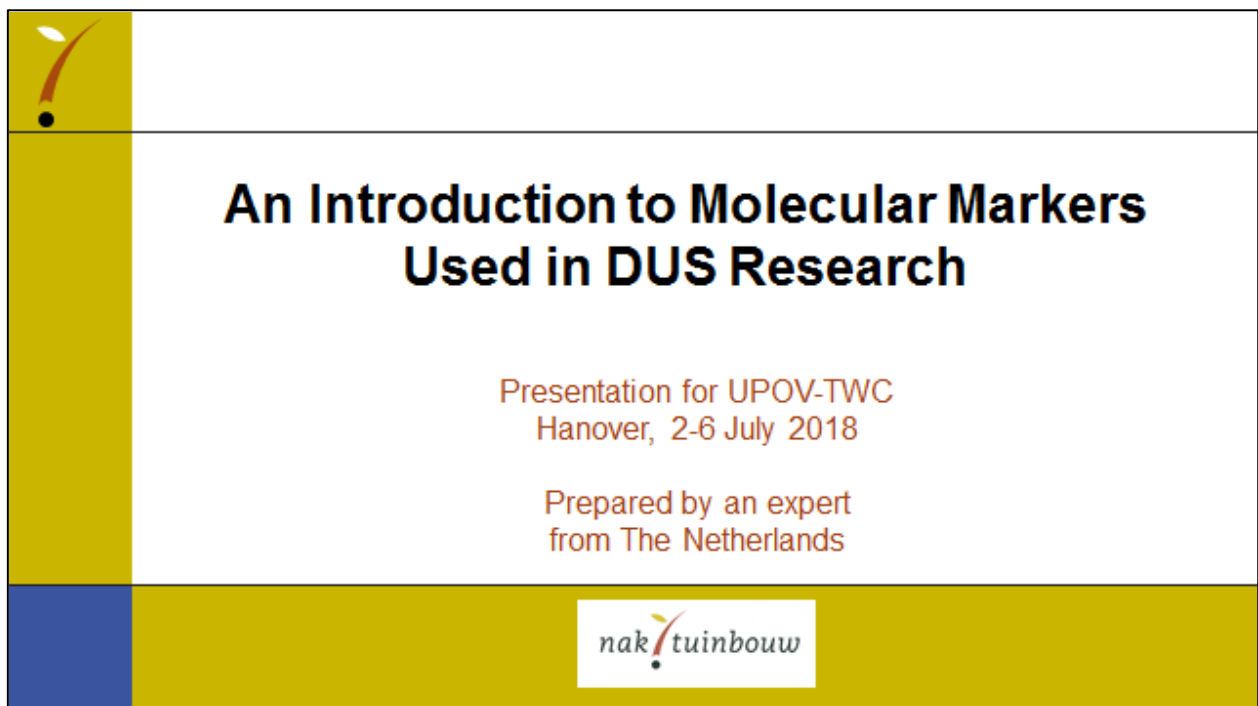
AN INTRODUCTION TO MOLECULAR MARKERS USED IN DUS RESEARCH IN THE NETHERLANDS*Document prepared by an expert from the Netherlands**Disclaimer: this document does not represent UPOV policies or guidance*

The Annex to this document contains a copy of a presentation on “An Introduction to Molecular Markers Used in DUS Research in the Netherlands”, prepared by an expert from the Netherlands, to be made at the thirty-sixth session of the Technical Working Party on Automation and Computer Programs (TWC).

[Annex follows]

AN INTRODUCTION TO MOLECULAR MARKERS USED IN DUS RESEARCH IN THE NETHERLANDS

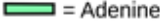




Presentation prepared by an expert from the Netherlands



DNA structure



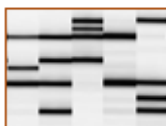
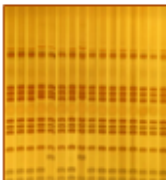
DNA

-  = Adenine
-  = Thymine
-  = Cytosine
-  = Guanine
-  = Phosphate backbone



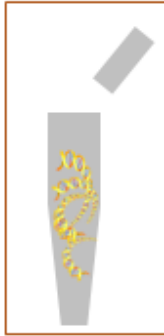
```
TGCGCCCCCTGGCATTCCGGGGGGC
TCATTACAACCCCTCAAGCTCTGCTT
CCTCTCTGCGGACGCGCCTTAAAGA
TTCAGCCCTCAAGCGTAGTAGAATA
CGGTTGGCGTCGCCCCGGACGAA
TCAAGGTTGACCTCGGATGATCATT
GCTCCGGCTCGACTCTCCACCCAA
GTTGCTTTGGCGGGCCGCGTCTC
CGGGGGGGCTGGCCAGCGCCCGCCA
CCAGTCAGTGAACCTTCGAGTCTGA
ACTAAAACCTTCAACAACGGATCTC
ATGAAGAACGCAGCGAAATGCGATA
AGAATTCAGTGAATCATCGAATCTT
CCCCTTGGTATTCCGAGGGGCATGC
TTCAACCTCAAGCTCTGCTTGGTA
CACGGACGCGCCTTAAAGACCTCGG
TCAAGCGTAGTAGAAAACACCTCGC
GTCGCCCGCCGGACGAACCTTGA
GACCTCGGATCATCGAGAAAGTTTCA
TTTCCTTCCGCTGCACGCGCTGGGT
CTGGGTGCTGGGTTCCTCCACTCAA
CGGTGAGGGGCATTTTGGTGGTGGG
GCCTCGTTCGGGCTCGGCAAAATGT
```

Molecular markers: principle



- A marker is a piece of DNA which can be made visible
- Markers can be:
 - locus specific
(well known position on the chromosome)
 - non-specific
(random, no known position)
- Molecular markers reflect DNA composition
- DNA composition varies between genotypes
- Pattern of a set of **molecular markers** is description of **genotype**
- Analogous to a set of **morphological characteristics** being a description of a **phenotype**

Types of molecular techniques

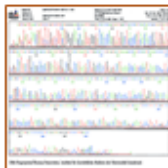
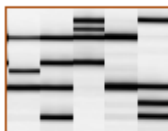
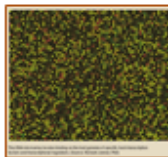


- Non-PCR* based techniques
 - requires isolation of large quantities of pure genomic DNA
 - e.g. RFLP
- PCR based techniques
 - requires only small quantities of DNA
 - e.g. RAPD, AFLP, microsatellites (SSR, STMS), SNP

*PCR = polymerase chain reaction: a technology that amplifies DNA

5

Types of DNA markers



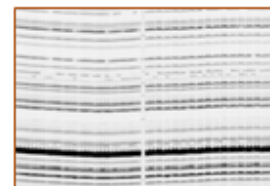
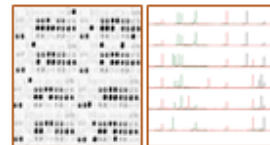
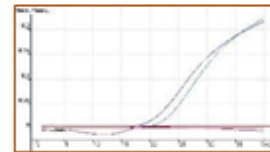
There are many different DNA marker technologies, e.g. RAPD, RFLP, CAPS, VNTR, SCAR etc. etc.

In DUS the most widely used techniques are:

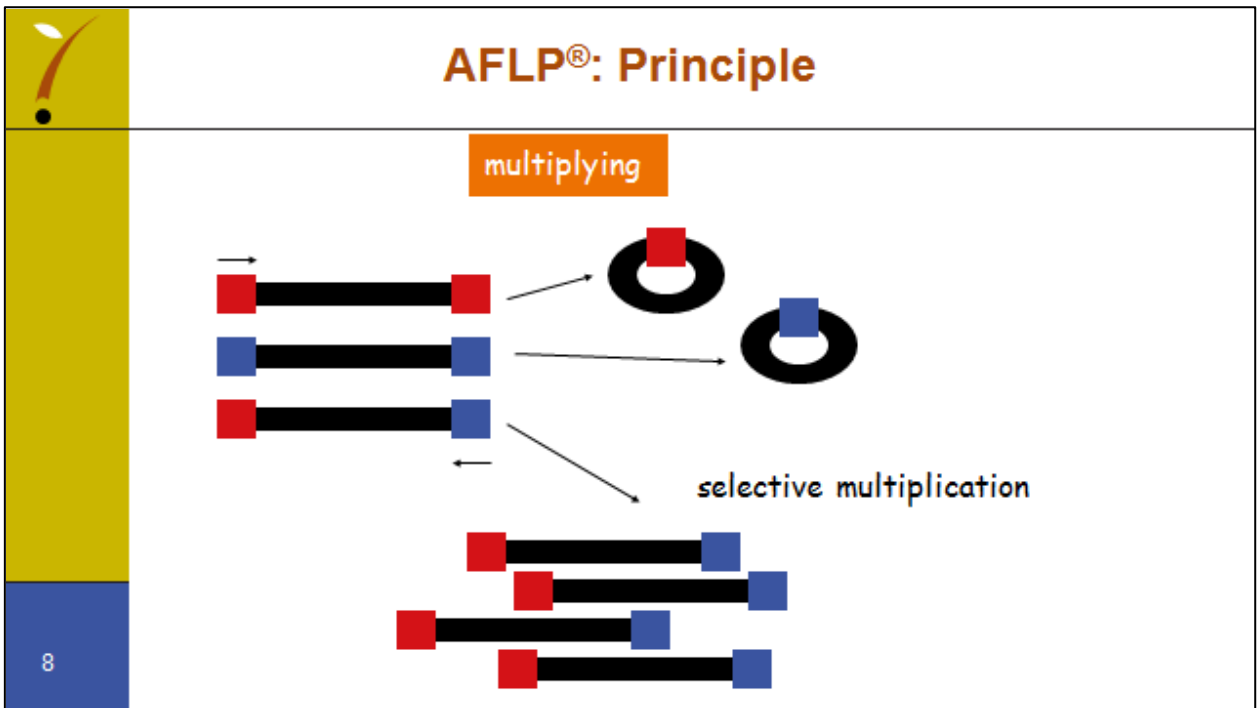
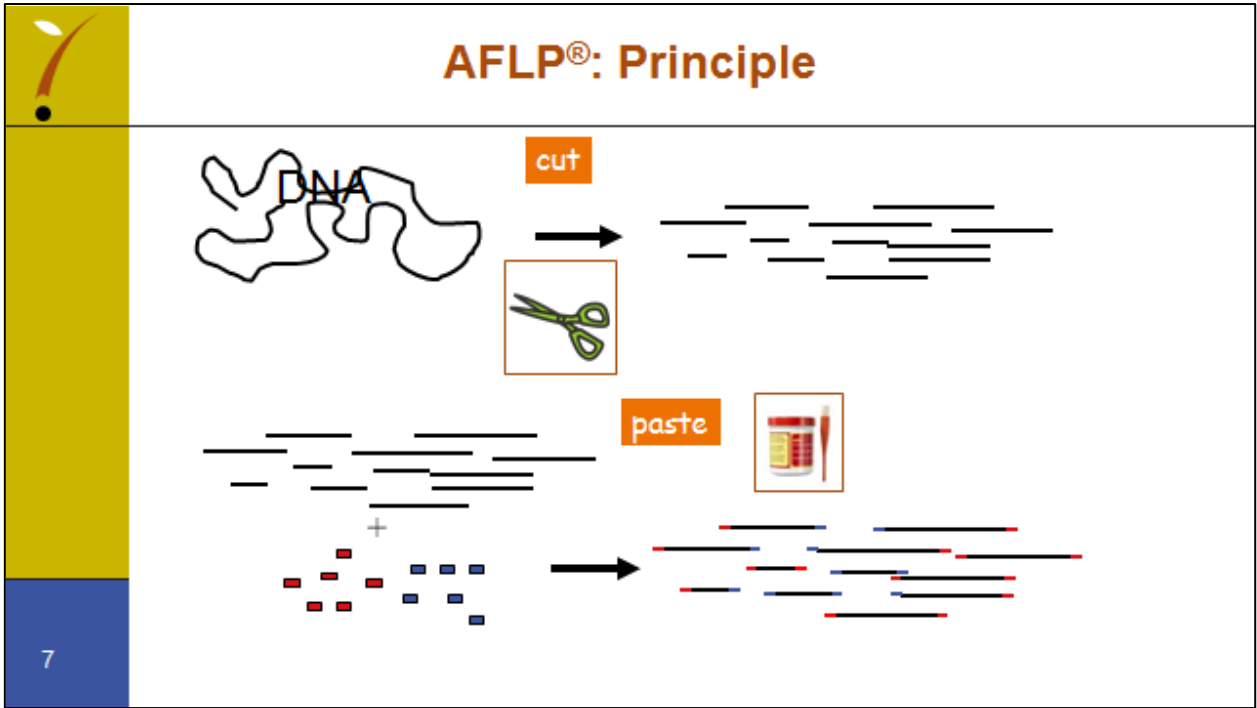
AFLP

SSR

SNP

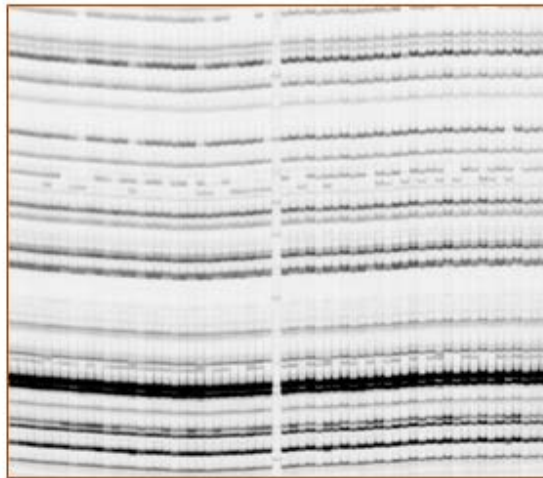


6



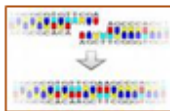
AFLP®: Principle

Separation of different fragments



9

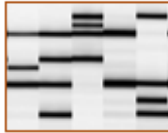
AFLP®: Principle



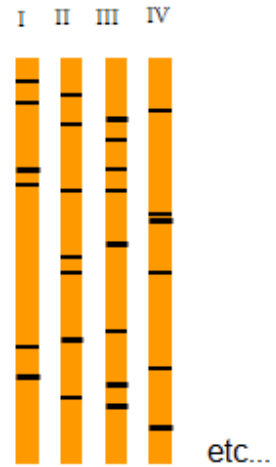
- Restriction enzymes can be adjusted (in type as well as number) for each plant species
- The products resulting from specific restriction enzymes etc. reflect the genetic composition of a variety
⇒ different varieties give different combination of products
- The combined result of more restriction enzymes etc. will increase information on variation between varieties

10

AFLP®: Principle

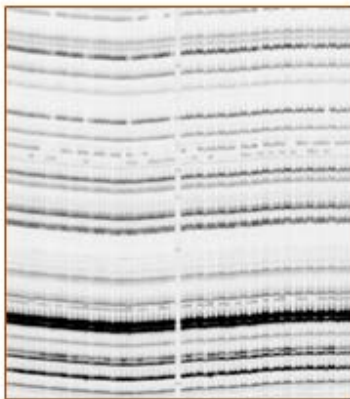


- Markers are distributed throughout the chromosomes, wherever the enzymes can cut
- Markers cannot be allocated to a specific position in the genome
- It is not known where the amplified fragment originated

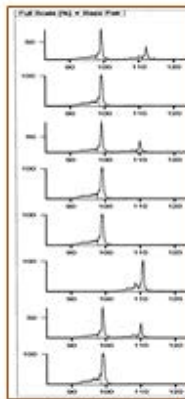


AFLP®: Different detection systems

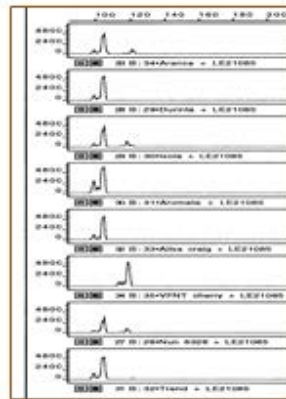
Licor



ALF express

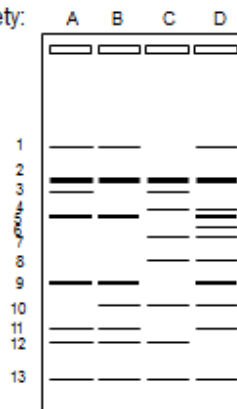


ABI 377



AFLP[®]: From gel (or peaks) to data matrix

variety:



	1	2	3	4	5	6	7	8	9	10	11	12	13
A	1	1	1	0	1	0	0	0	1	0	1	1	1
B	1	1	0	0	1	0	0	0	1	1	1	1	1
C	0	1	1	1	0	0	1	1	0	1	0	1	1
D	1	1	0	1	1	1	1	1	1	1	1	0	1

13

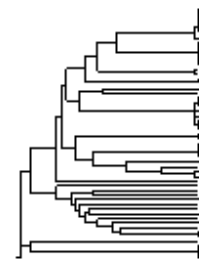
AFLP[®]: from matrix to graph

	1	2	3	4	5	6	7	8	9	10	11	12	13
A	1	1	1	0	1	0	0	0	1	0	1	1	1
B	1	1	0	0	1	0	0	0	1	1	1	1	1
C	0	1	1	1	0	0	1	1	0	1	0	1	1
D	1	1	0	1	1	1	1	1	1	1	1	0	1

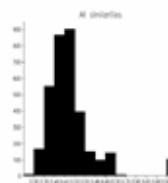
Determination of similarities (pairwise comparison): e.g. Jaccard coefficient

sample ID	2	13	1	12	10	3	6	9	4	8	7	5	
2	100.00												
13	41.67	100.00											
1	34.61	26.08	100.00										
12	34.61	26.08	100.00	100.00									
10	25.00	21.74	52.63	52.63	100.00								
3	30.00	24.00	61.91	61.91	55.00	100.00							
6	34.37	26.93	60.87	60.87	54.55	51.82	100.00						
9	35.71	26.93	54.55	54.55	47.62	30.00	77.27	100.00					
4	35.71	26.93	58.33	58.33	52.17	73.91	83.33	77.27	100.00				
8	36.67	28.00	59.09	59.09	57.14	76.19	86.37	71.43	66.24	100.00			
7	28.57	27.27	57.89	57.89	55.55	77.78	80.00	72.22	84.21	88.23	100.00		
5	37.50	12.50	75.00	75.00	57.14	75.00	77.78	71.43	87.50	85.71	85.71	100.00	
11	34.61	19.05	70.59	70.59	50.00	66.67	78.56	70.59	88.89	82.35	81.25	100.00	100.00

similarity matrix



dendrogram



frequency distribution

14

Similarity indices

A	1	1	1	0	1	0
B	1	1	0	0	1	0
C	0	1	1	1	0	0
D	1	1	0	1	1	1

- There are many similarity indices:
 - Jaccard, Dice, Nei, Cityblock, Euclidian, Simple matching, etc...
 - The choice depends on i.a. the type of data
 - For these binary data we use Jaccard:

$$\text{similarity } J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \quad (\text{distance} = 1 - J)$$

With:

M_{11} = total number where A and B are both 1

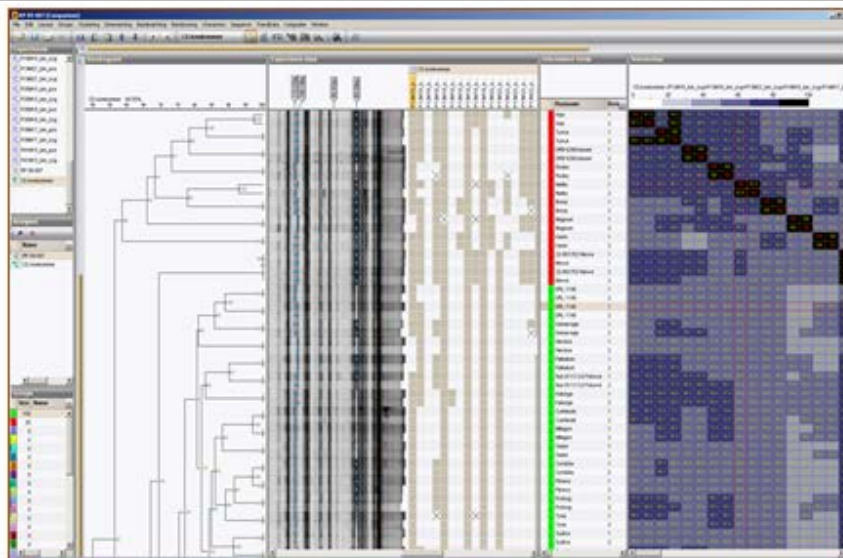
M_{01} = total number where A = 0 and B = 1

M_{10} = total number where A = 1 and B = 0

M_{00} = total number where A and B are both 0

15

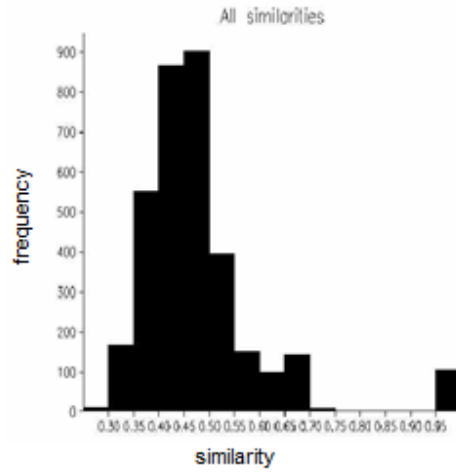
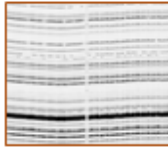
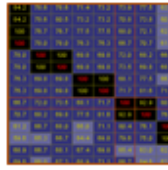
AFLP®: Pair wise comparison (BioNumerics)



16

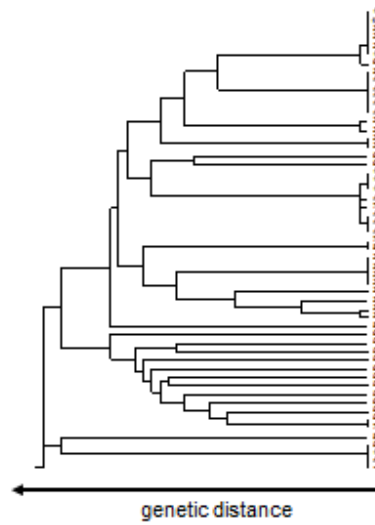
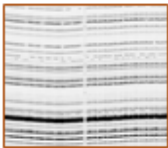
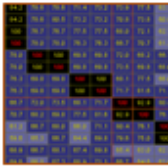
AFLP®: Similarities for variety pairs

Example: Rose



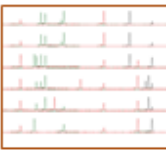
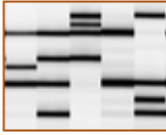
17

AFLP®: Part of a UPGMA dendrogram



18

AFLP®: Characterisation



- Universally applicable
(no sequence info required, suitable for all species)
- Large discrimination power
- Dominant (one 'allele', heterozygotes not visible)
- Multi-locus
(many datapoints distributed over genome)
- Flexible in experimental set-up
- High reproducibility within lab
- Moderate reproducibility between labs
- Suitable for databasing

19

AFLP®: Characterisation



- Technically demanding
- Labor-intensive (expensive)
- Not suitable for automation
- Sensitive for interpretation (dependent on the person)

20

Microsatellites

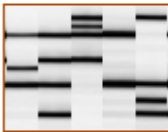


- Synonyms:
 - Sequence Tagged Microsatellite Sites (STMS)
 - Simple Sequence Repeats (SSR's)
 - Short Tandem Repeats (STR's)



21

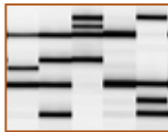
Microsatellites: Principle



- Each marker has a specific location in the genome (*locus, plural=loci*)
- At this location repeats of a specific short sequence occur
- Examples of repeat sequences: TCC, GATA, GT, ...
 - e.g. $(GT)_9 = GTGTGTGTGTGTGTGTGT$
- These locations can be found by developing a primer that finds the flanking regions of the repeat sequence
- Marker=primer+microsatellite

22

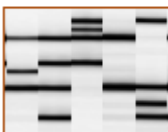
Microsatellites: Principle



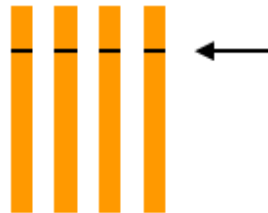
- A specific number of repeats is called an 'allele'
 - e.g. (GT)₉, (GT)₁₀, (GT)₁₂ are different alleles of the same microsatellite marker
- The number of alleles per marker can vary from 2-50(?), depending on the marker
- A diploid variety will show 2 alleles (either identical or different), a tetraploid variety has 4 alleles

23

Microsatellites: Principle



- Potato is tetraploid: it has 4 copies of each chromosome



e.g. chromosome VI:

- The marker on chromosome VI is located on a specific position
- It is exactly known where the amplified fragments originated

24

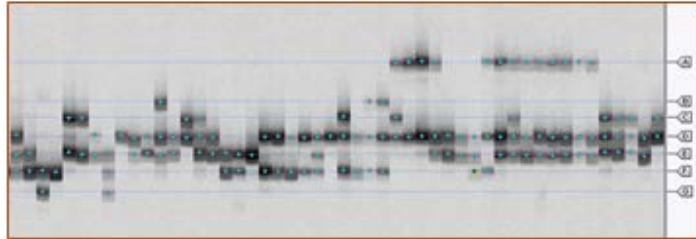
Microsatellites: Principle



- E.g. this marker has 7 different alleles, called A,B,...G



e.g. chromosome VI:



25

Microsatellites: Principle



- E.g. this marker has 7 different alleles, called A,B,...G



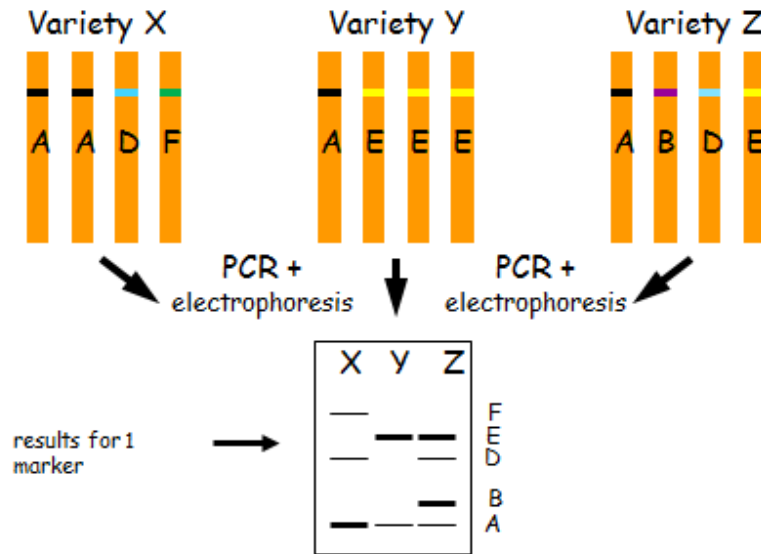
e.g. chromosome VI:

Two copies of chromosome VI carry allele A, one carries allele D and one carries allele F

- Alleles are scored as absent/present:
 - 'allelic phenotype' = ADF
 - ('allelic genotype' would be AADF)

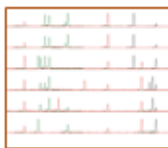
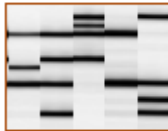
26

Microsatellites: Principle



27

Microsatellites: Detection systems

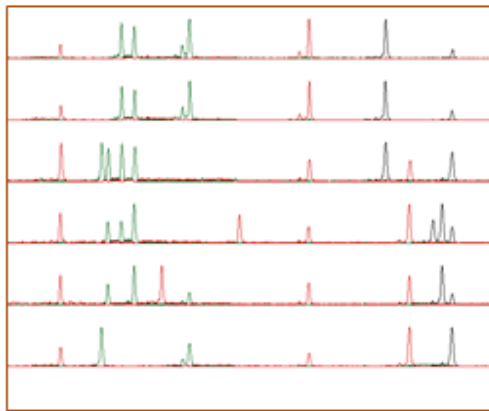


- Agarose/Sequencing gel (band patterns)
 - Radio-active labeling
 - Silver staining
 - Fluorescent label
- Capillary electrophoresis (peak patterns)

28

Microsatellites: Multiplex* analysis

Example: 3 markers - tetraploid species:

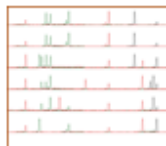


Original variety 1
Mutant of variety 1
Variety 2
Variety 3
Variety 4
Variety 5

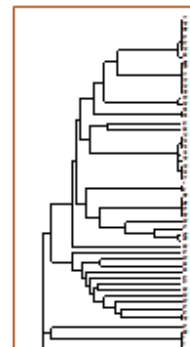
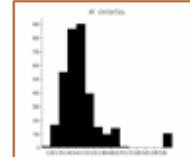
29

*multiplex = more markers analysed together (each marker has a colour label)

Microsatellites: Data handling



- Similar to AFLP data:
 - Convert bands/peaks to data matrix (110001100.....)
 - Calculate similarities
 - Frequency distributions of similarities
 - Dendrograms



30

Microsatellites: Data handling

Example of data scoring of a tetraploid species:

allele	Marker 1						Marker 2				Marker 3								
	a	b	c	d	e	f	a	b	c	d	a	b	c	d	e	f	g	h	i
Variety A	1	0	1	1	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1
Variety B	1	0	1	1	0	0	1	0	1	0	0	1	0	1	1	0	0	0	0
Variety C	0	0	1	0	0	0	0	1	1	1	0	1	0	1	1	0	0	0	1
Variety D	1	0	1	1	0	0	1	1	1	1	0	0	1	0	1	0	1	0	0
Variety E	0	1	1	0	1	0	1	1	0	0	1	0	1	1	0	0	0	1	0

As the number of copies per allele is usually not known, a profile is called an allelic *phenotype* (not *genotype*)

31

Microsatellites: Data handling

Frequency distribution of allelic phenotypes
(300 potato varieties)



marker 1:

4 alleles, 11 phenotypes

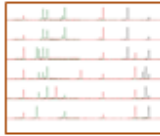
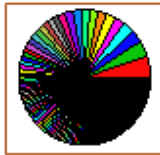


marker 2:

17 alleles, 137 phenotypes

32

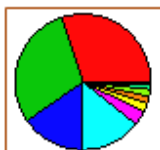
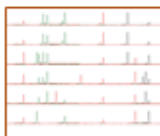
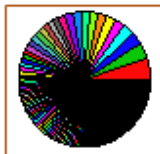
Microsatellites: Characterisation



- Co-dominant (heterozygotes visible)
- Multi-allelic
- Highly discriminative (depending on number of markers)
- Reproducible within and between labs
- Available for many crops (many not freely available)
- Technically relatively easy (quick)
- Suitable for exchange/databases

33

Microsatellites: Characterisation

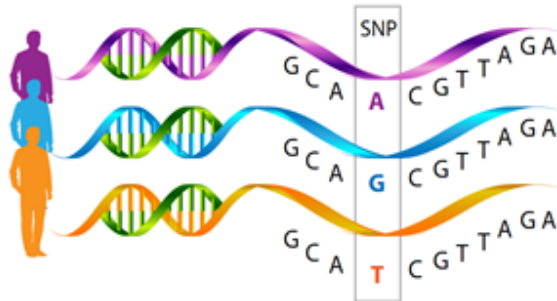


- Markers need to be developed per crop
- Sequence information is required
- High development costs
- Not so many datapoints (loci)
- Sensitive for stutter bands

34

SNP's: Principle

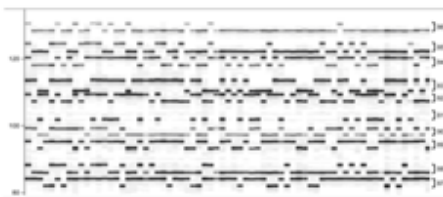
- SNP= Single Nucleotide Polymorphism
- A single base change in DNA sequence



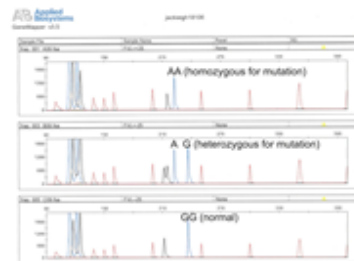
- SNP discovery/development through mutation scanning, hybridization assays, PCR based systems, sequencing ...

35

SNP's: Detection platforms (1)

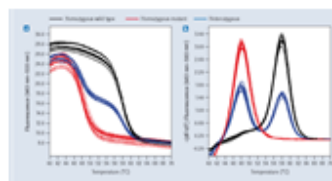
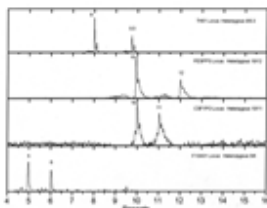


Band patterns on gel



peakpatterns on capillary system

MALDI-TOF mass specrometry



meltcurves
on real time
PCR machine

36

SNP's: Detection platforms (2) Micro arrays



Up to 25.000 SNP's per array

37

SNP's: Data handling



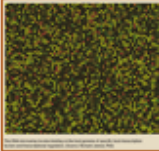
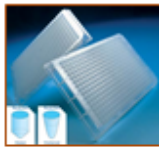
- Scoring of present/absent in genotypes (veg. prop. species)
Data handling similar to SSR and AFLP: similarities etc.
- Description of varieties in terms of SNP frequencies in populations (outcrossing species!)



phylogram illustrating
separate genepools in
ryegrass

38

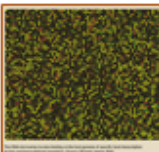
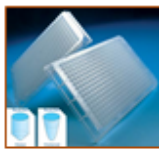
SNP's: Characterisation



- Robust method, highly reproducible
- Suitable for high throughput analyses
- Large numbers of possible markers
- Markers need to be identified per crop
- Technical skills/equipment required
- High development costs (discovery)
- Automated and cost effective screening (detection)
- Suitable for databases
- Suitable for characterization of genotypes as well as populations




39

SNP's: Characterisation



- Robust method, highly reproducible
- Bi-allelic (homozygote/heterozygote)
- Suitable for high throughput analyses
- Large numbers of possible markers
- Markers need to be identified per crop
- Technical skills/equipment required
- Very high development costs (discovery)
- Automated and cost effective screening (detection)
- Suitable for databases
- Suitable for characterization of populations

40

	
	<p><i>Thank you for your attention</i></p> 
41	

[End of annex and of document]