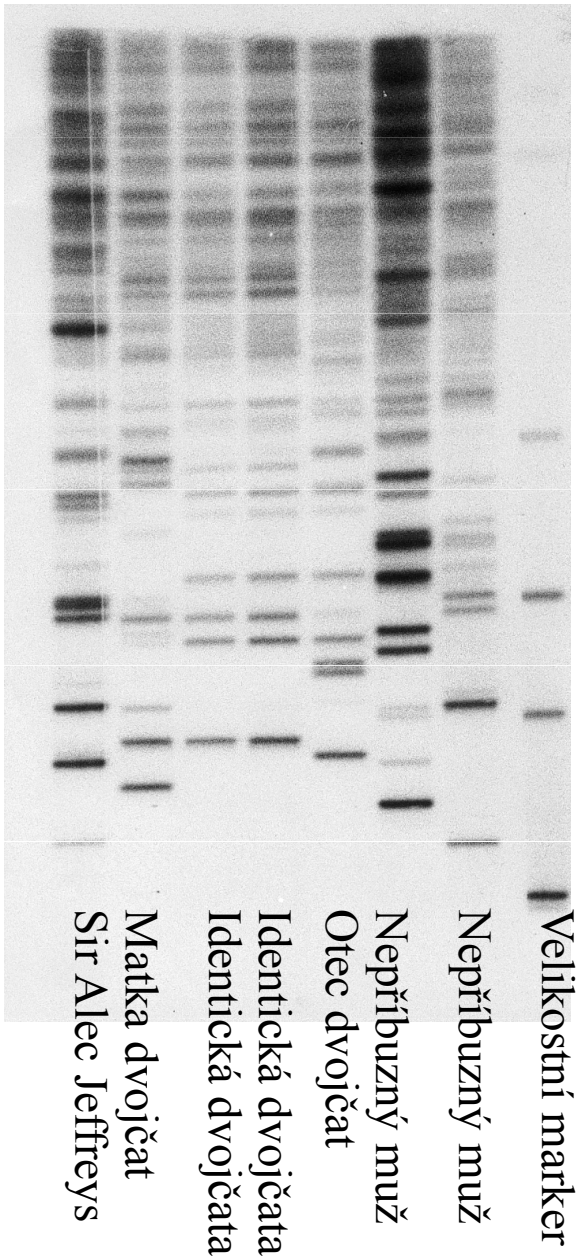
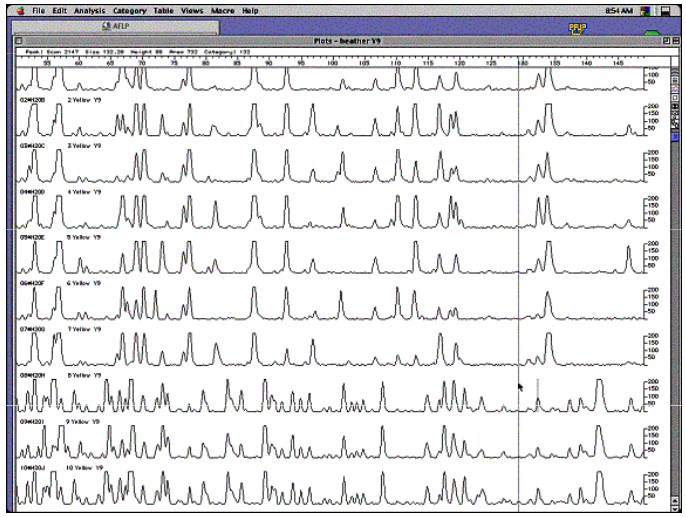
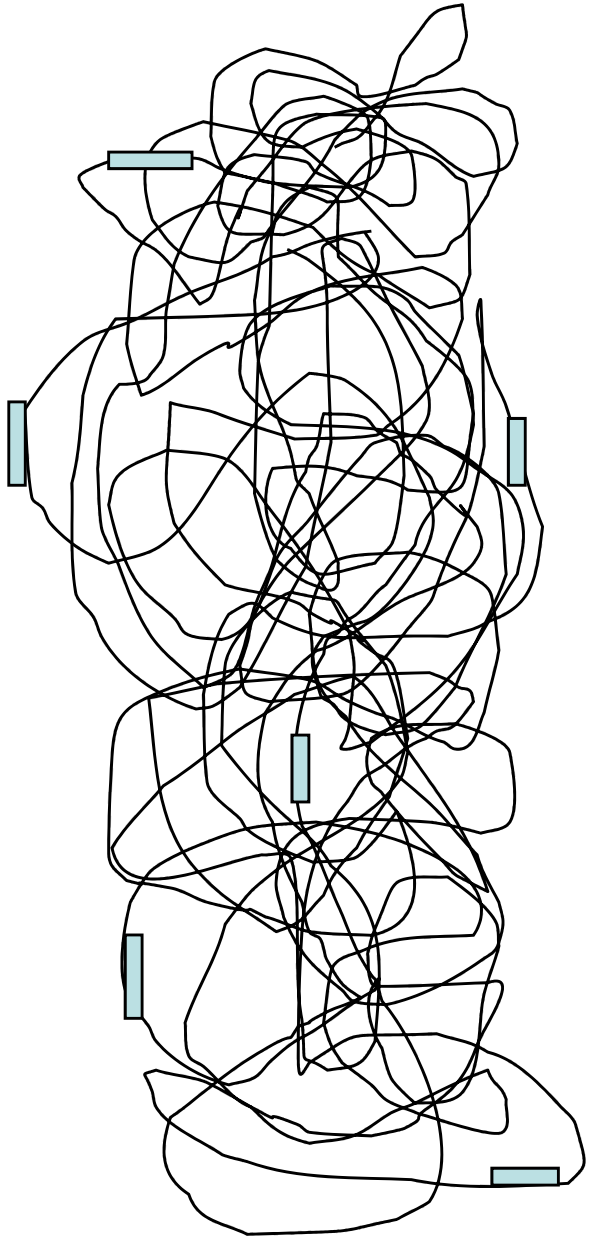


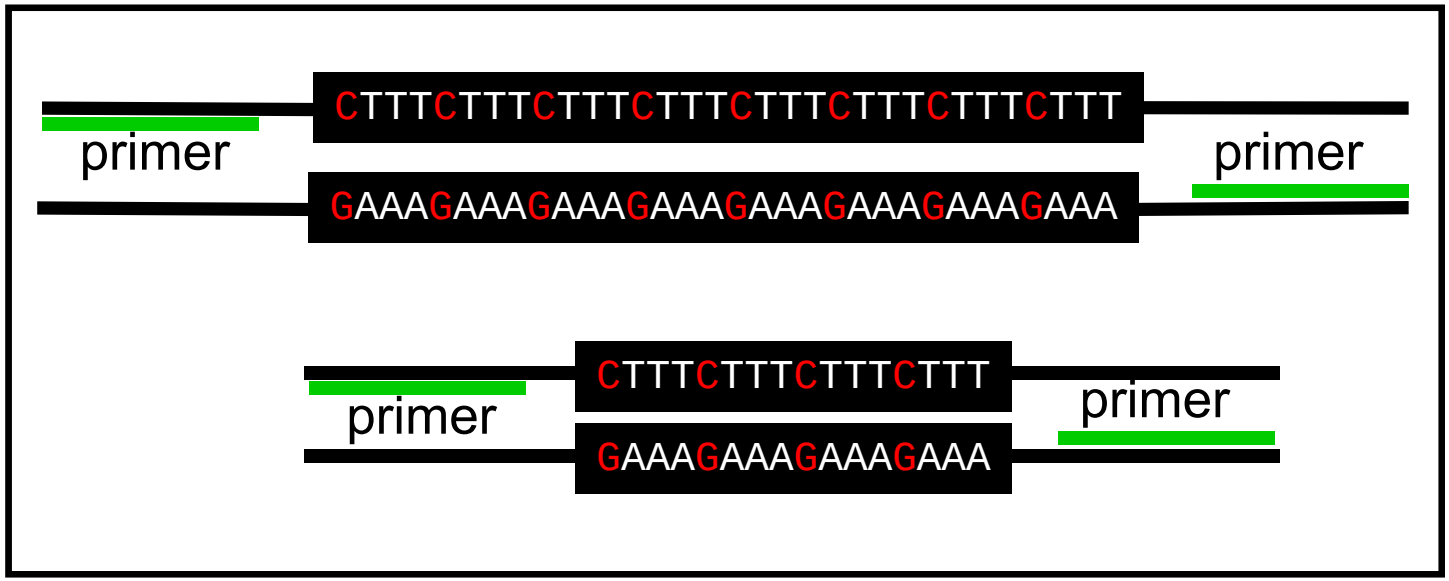
„DNA fingerprinting“

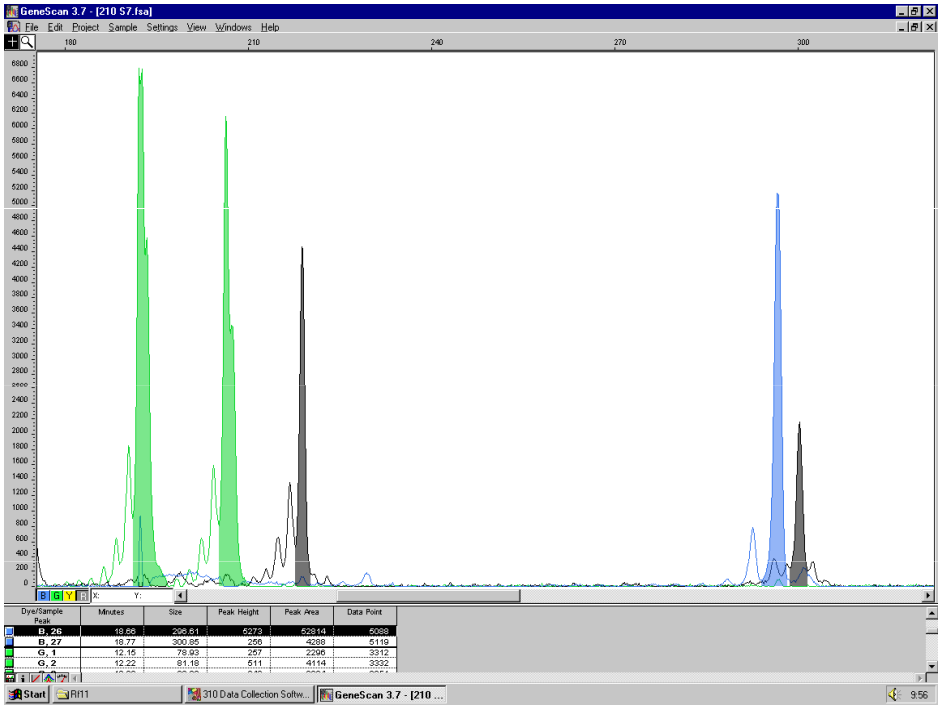


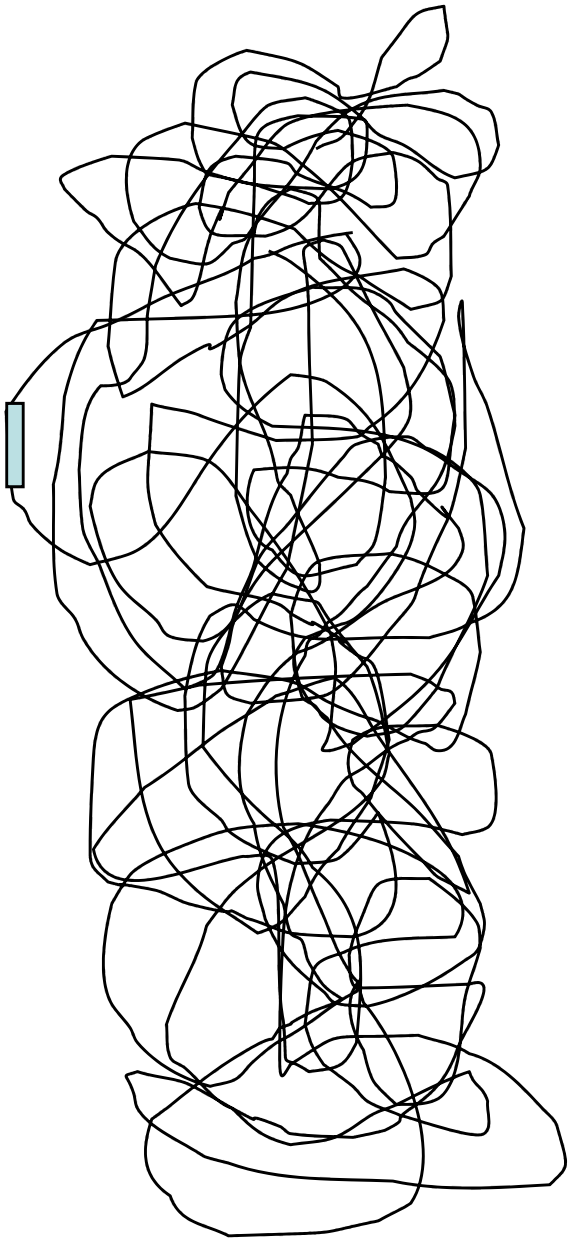
AFLP



Multilocus DNA markery



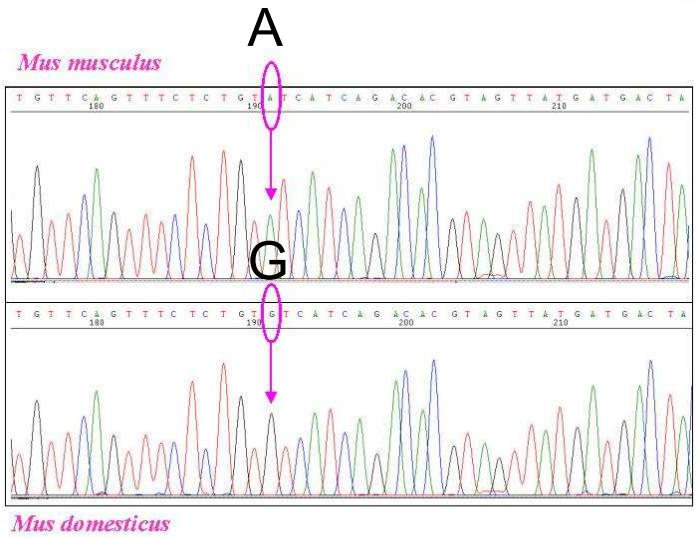


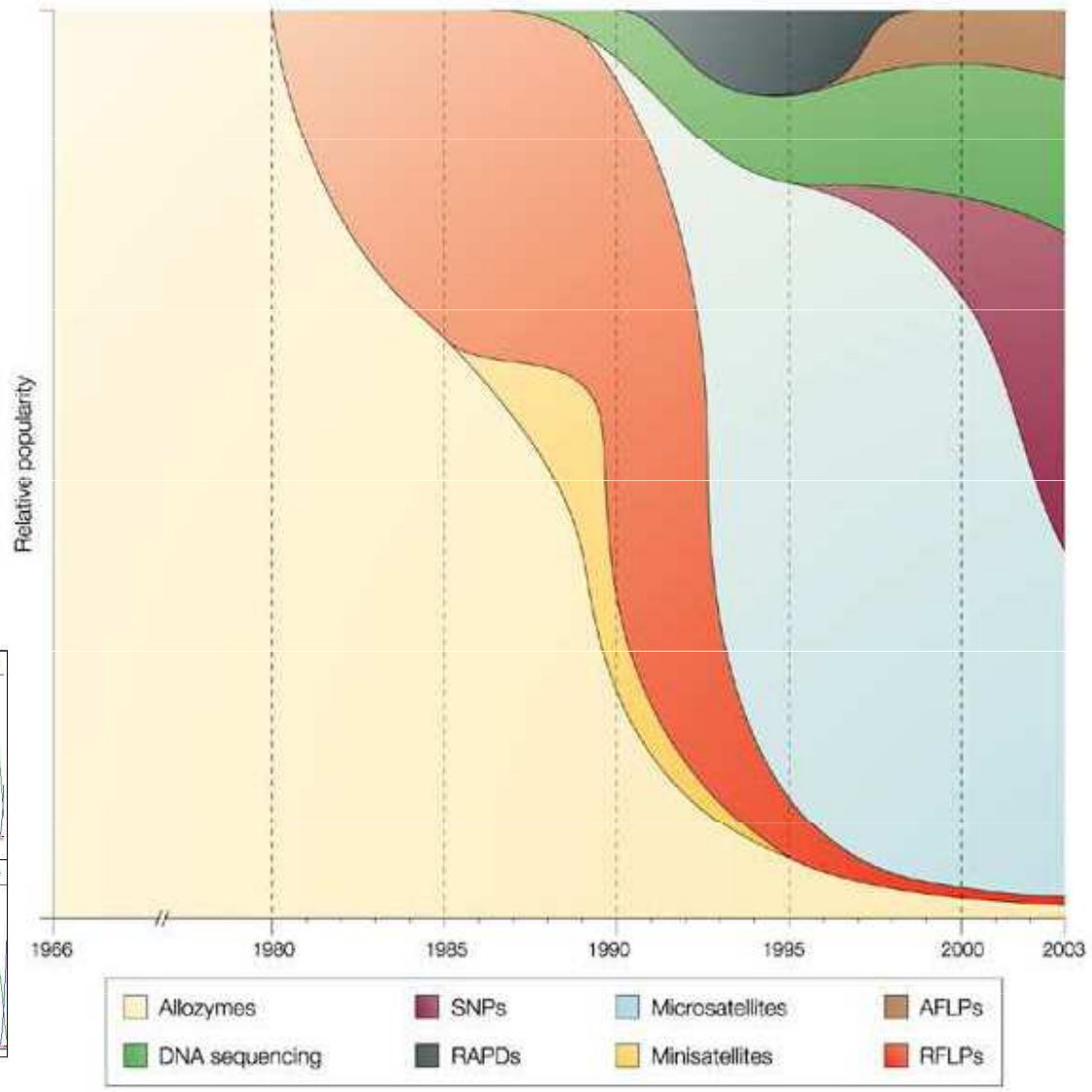


Mikrosatelity

Single-locus DNA markery

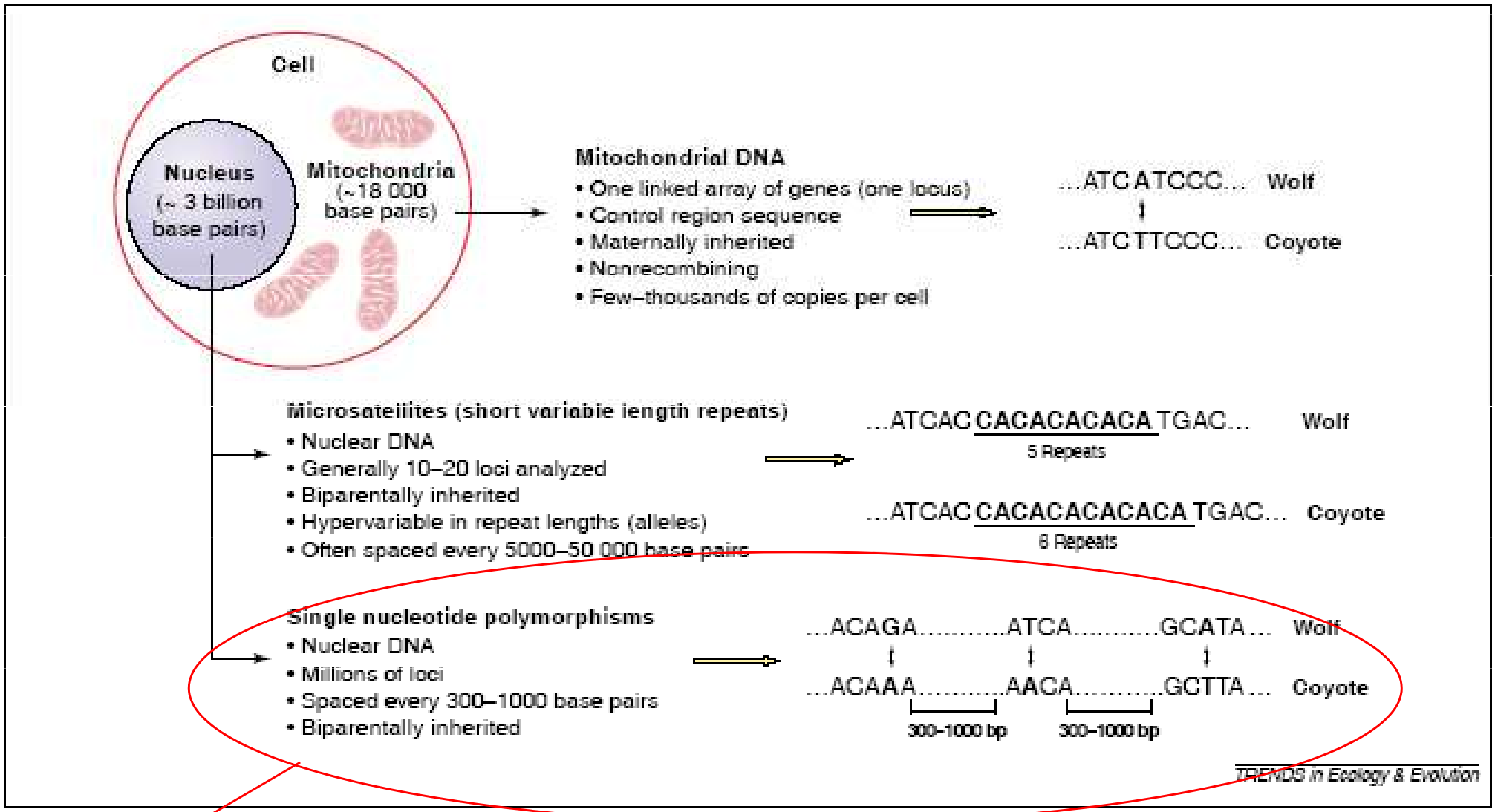
SNP Fashion on the rise





Nature Reviews | Genetics

Single nucleotide polymorphisms (SNPs)

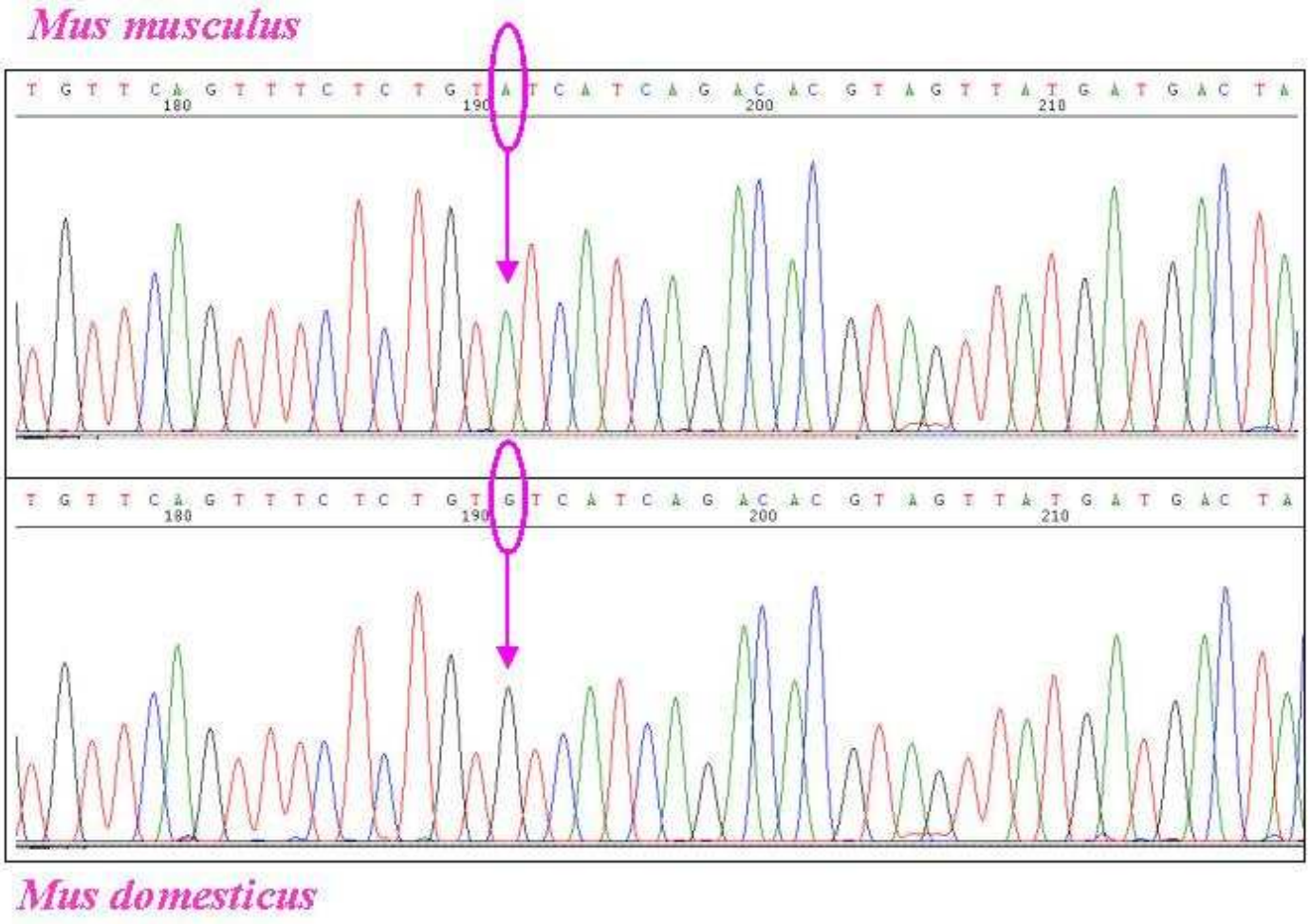


SNPs : nuclear genome (consensus)

How many SNPs are there in humans today?

- Human Mutation rate is ~2.5 x 10^{-8} mutations / site / gen
- ~150 mutations/diploid genome/generation
- 6.3 milliards people in the world = 945,000,000,000 mutations in the world today
- With 3 milliards nucleotides = each nucleotide in the world today is mutated 315 times.

Example of SNP marker



transice
A ↔ G

transition: Pu→Pu or Py→Py

transversion: Pu→Py or Py→Pu

Use of SNPs markers

- species (or genetical group) identification and analysis of hybridization
- phylogeography
- population genetics (genetic variation, individual identification – parentage, relatedness, population structure, population size, changes in population size)

Advantages

- abundant and widespread in many genomes (in both coding and non-coding regions) – milions of loci
- spaced every 300-1000 bp
- biparentaly inherited (vs. mtDNA)
- evolution is well described by simple mutation models (vs. microsatellites)
- shorter fragments are needed – using in non-invasive methods

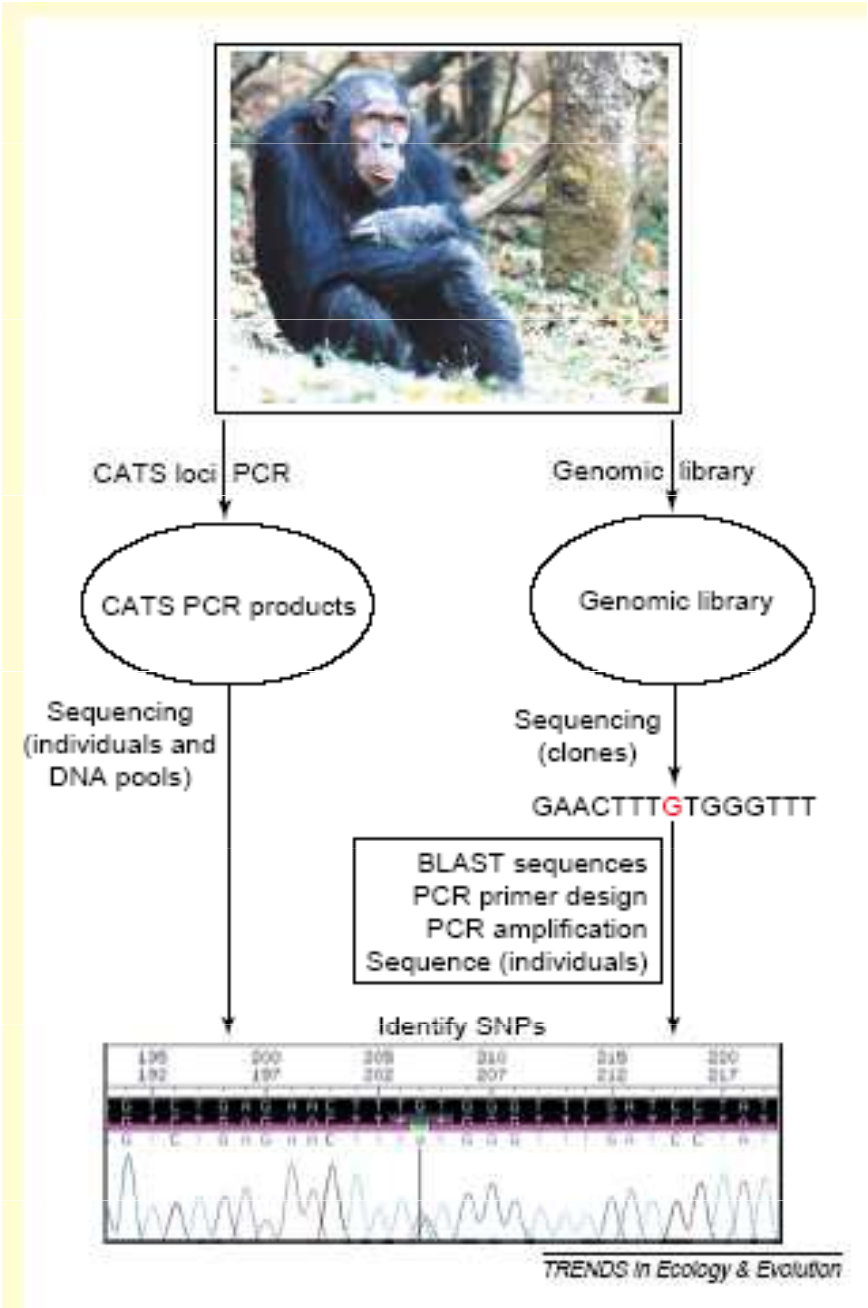
Disadvantages

- ascertainment bias – selection of loci from an unrepresentative sample of individuals
- low variability per locus (usually bi-allelic)
- higher number of loci is needed in population genetic applications (4-10 times more loci)

Methods

1. Locus discovery (ascertainment)
2. Genotyping

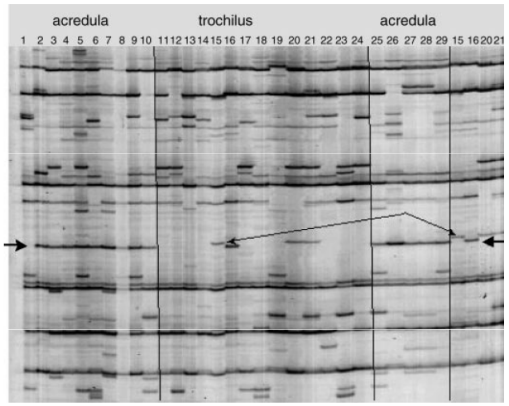
SNPs discovery



CATS loci = comparative anchor tagged site loci (= cross amplification)

Genomic library = genome restriction + cloning

AFLP = alternative to the genomic library construction (provide PCR fragments, can be transformed to informative SNP)



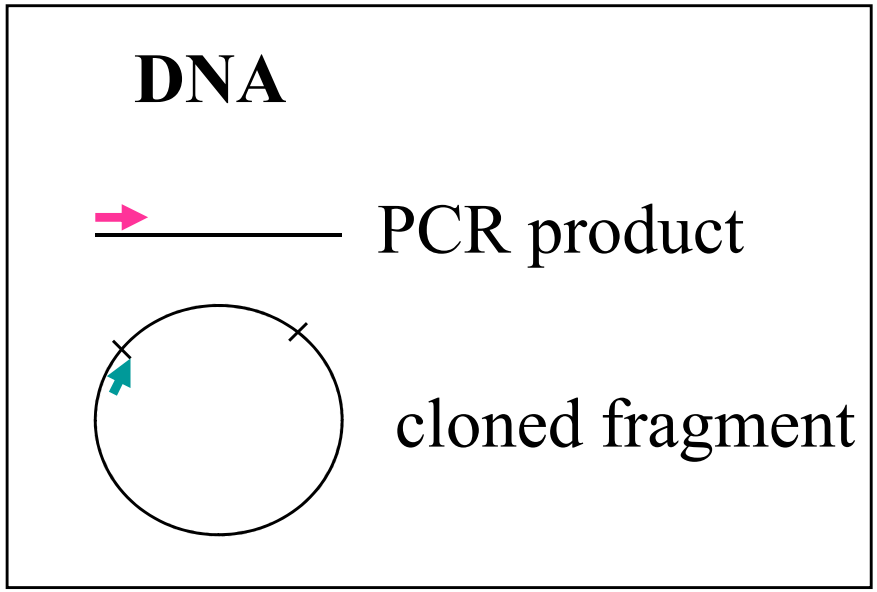
Sequencing

Sekvencování DNA

- Maxam-Gilbertova (chemická) metoda:
  bázově-specifická chem.
  modifikace a štěpení
  fragmentů DNA

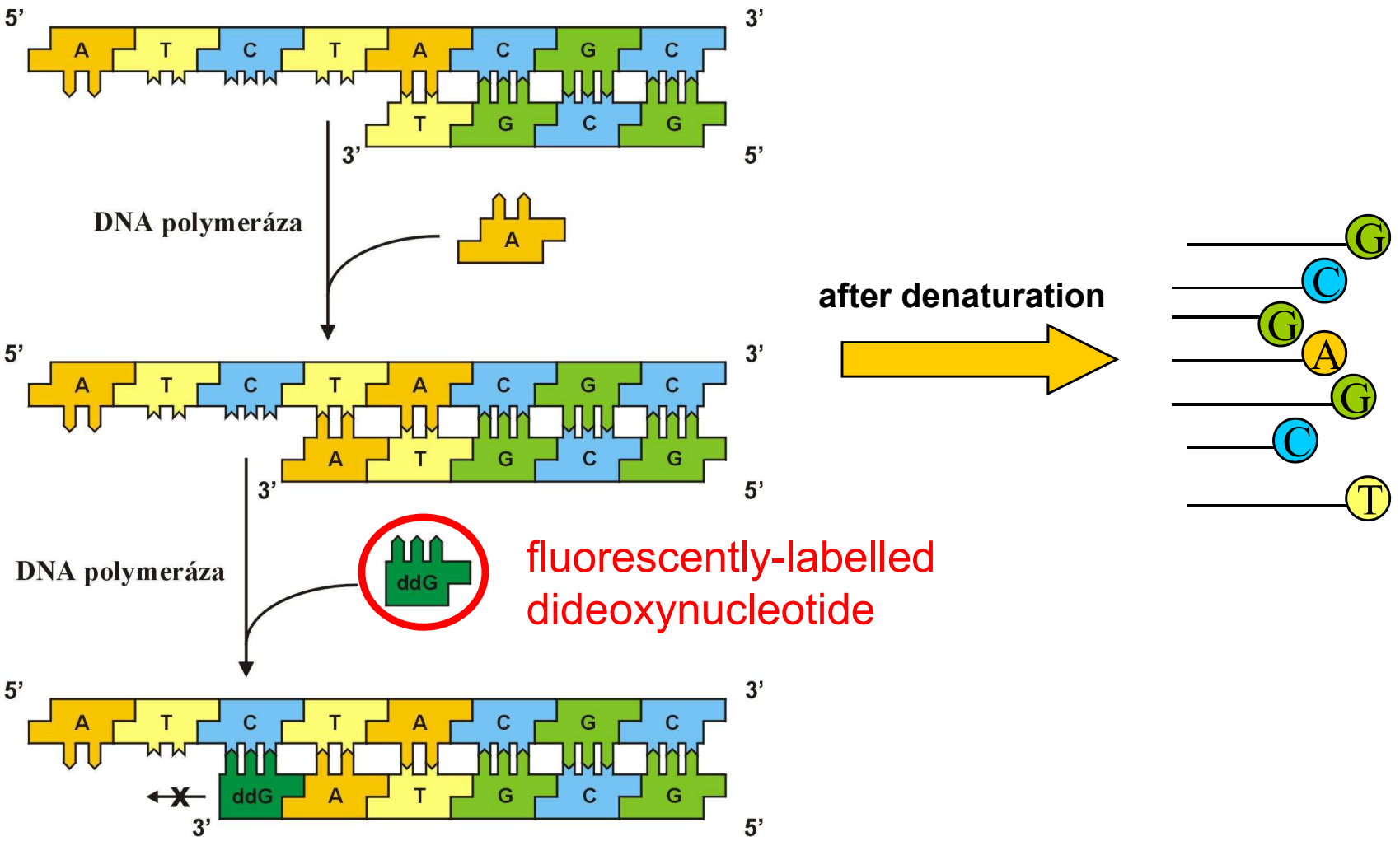
- Sangerova (enzymatická) metoda:
  terminace replikace
  pomocí ddNTP

Sequencing



sequencing reaction
with marked
dideoxynucleotides and
one specific or universal
primer

Sekvencování DNA

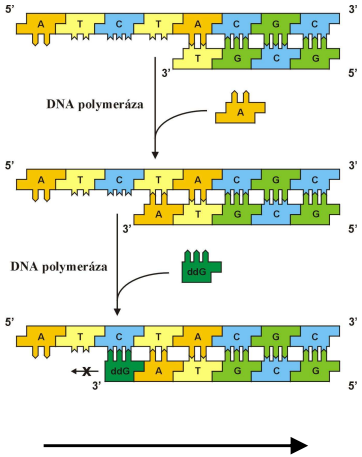


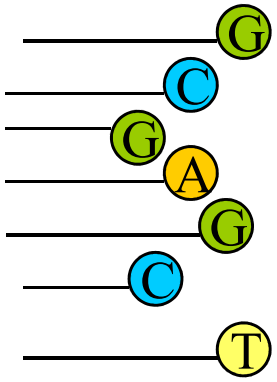
Sequencing

DNA

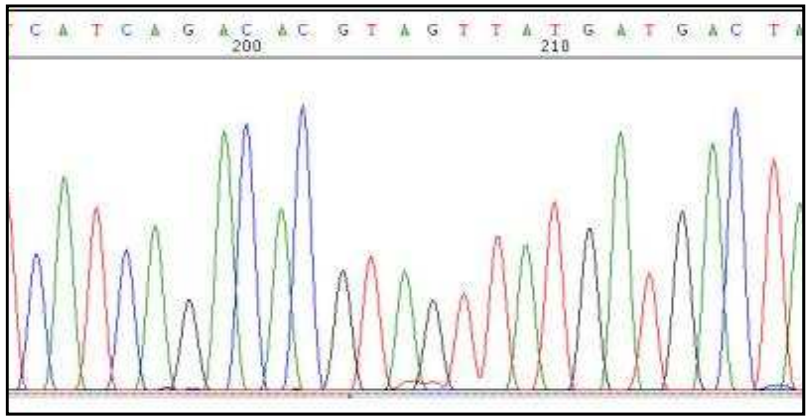
PCR product

cloned fragment





detector

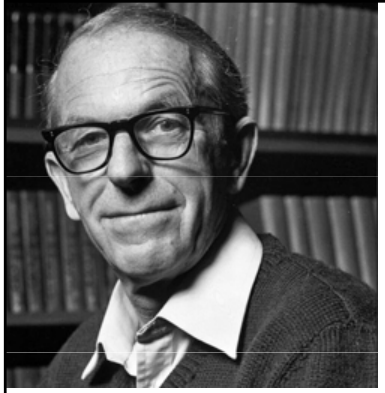


laser beam

capillary electrophoresis

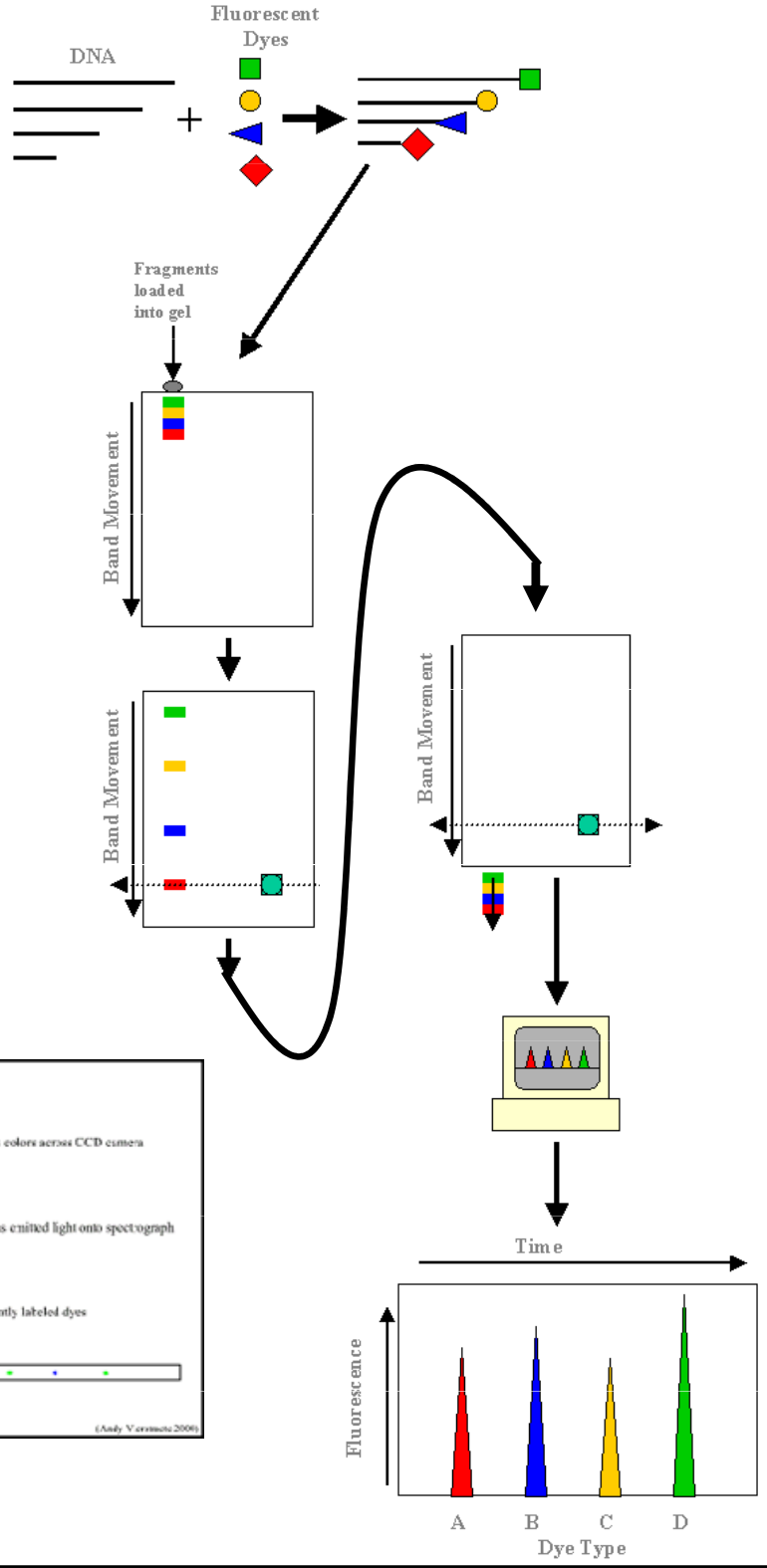
+

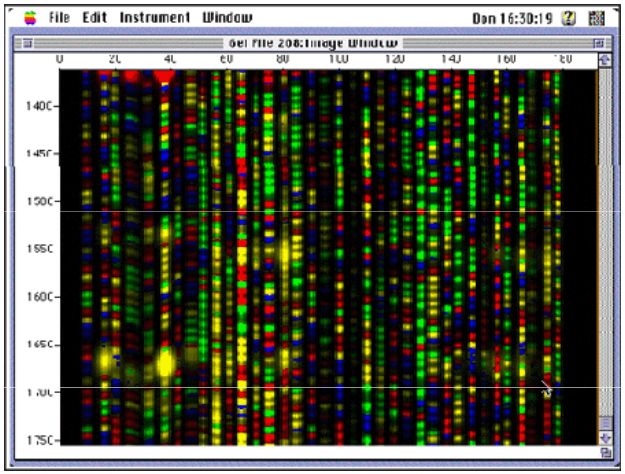
-

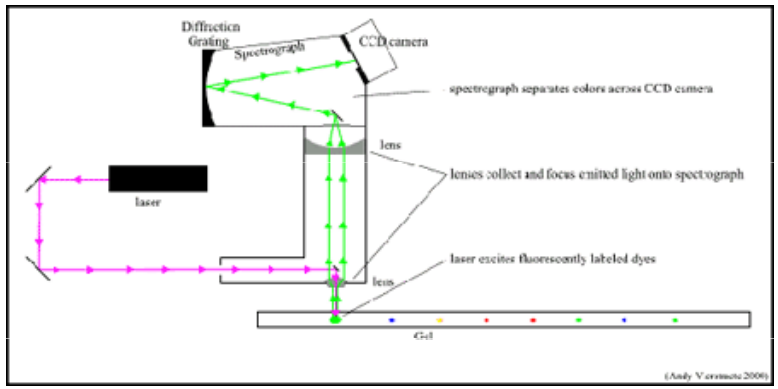


Sangrova dideoxy metoda

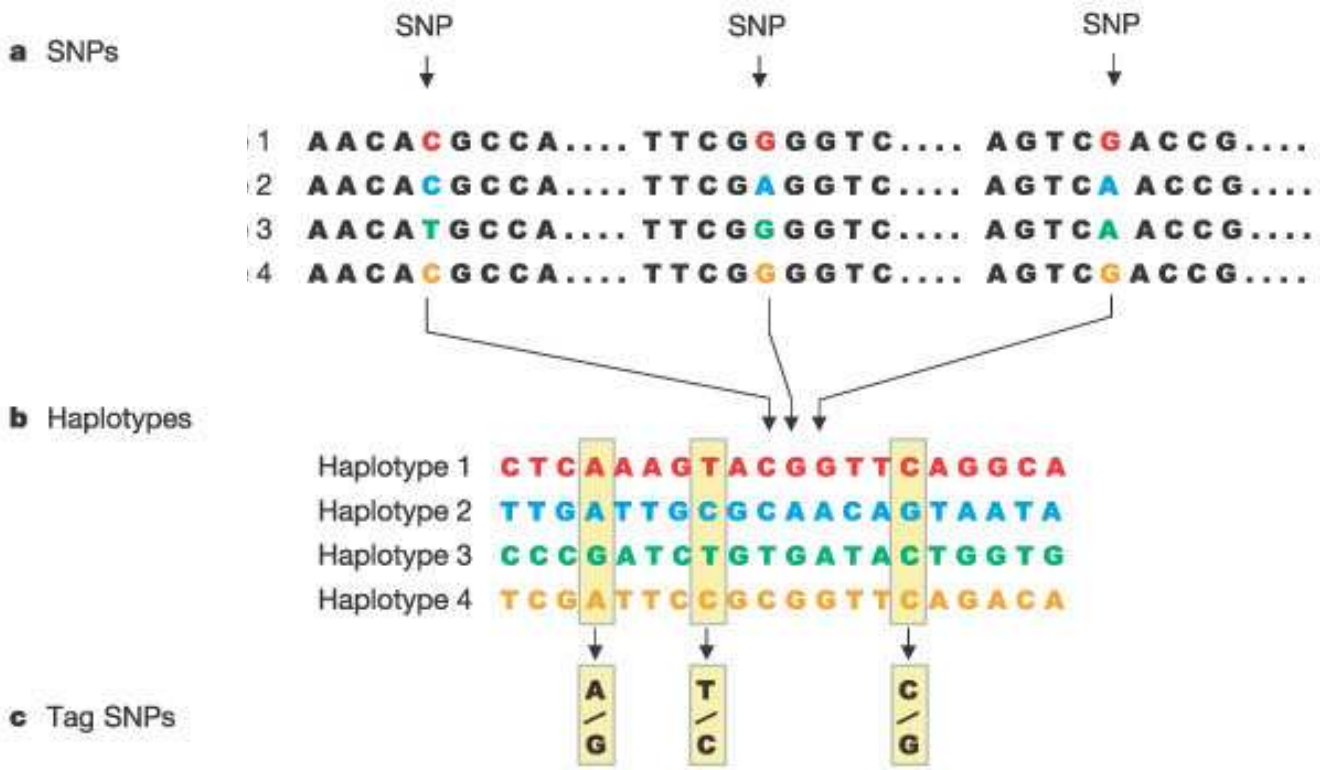
- Sekvence délky 500 – 1000 bp
- 4 kapiláry - destička s 96 vzorky za noc
- Jsou i sekvenátory s 96 kapilárami







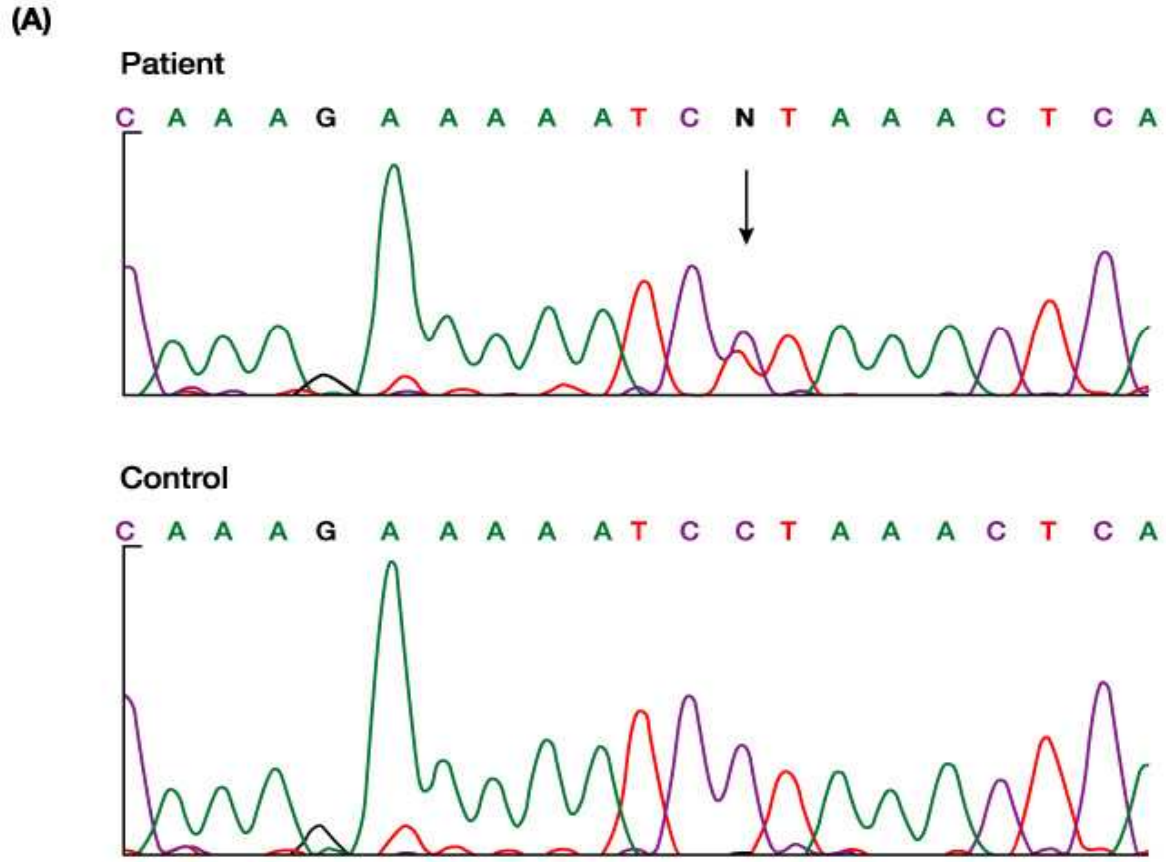
Identifikace různých genotypů u různých jedinců (chromozómů, tj. variabilita alel)



SNPs genotyping

= zjištění genotypu daného jedince

SNPs genotyping – sekvenování?
Je drahé a nejasné u heterozygotů

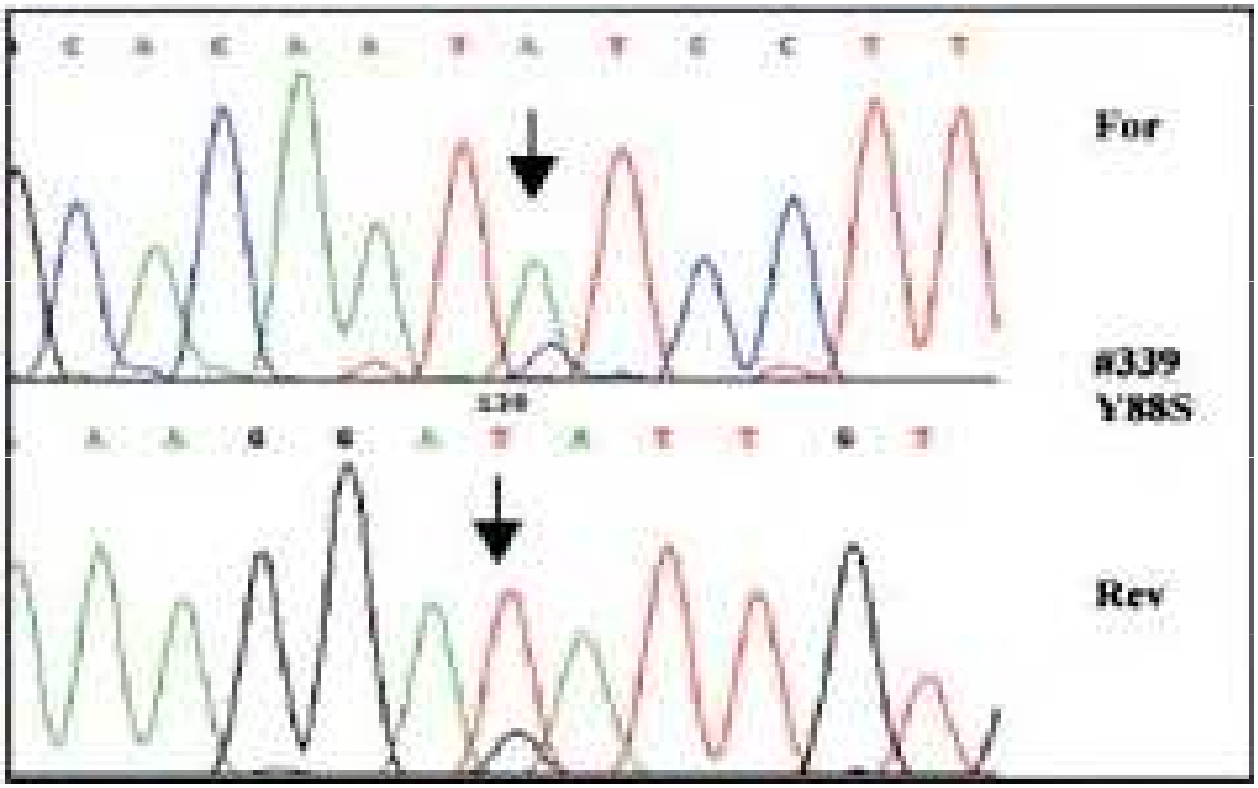


Heterozygote?

Homozygote

Double peak shows heterozygous mutation

Heterozygotes?



Bi-directional sequencing – are you really sure?

SNPs genotyping – klonování a následné sekvenování?
- separation of two (or more in duplicated genes) alleles

each clone contain the only allele

!!! cloning – 1000 Kč
!!! sequencing 1 clone – 150 Kč

ligation, transformation

Ex.: heterozygote = two diff. alleles

SNP genotyping - old standards

PCR-RFLP

(restriction fragments length polymorphism)

Allele A

CCGATCAATGCGGCAA
GGCTAGTTACGCCGTT

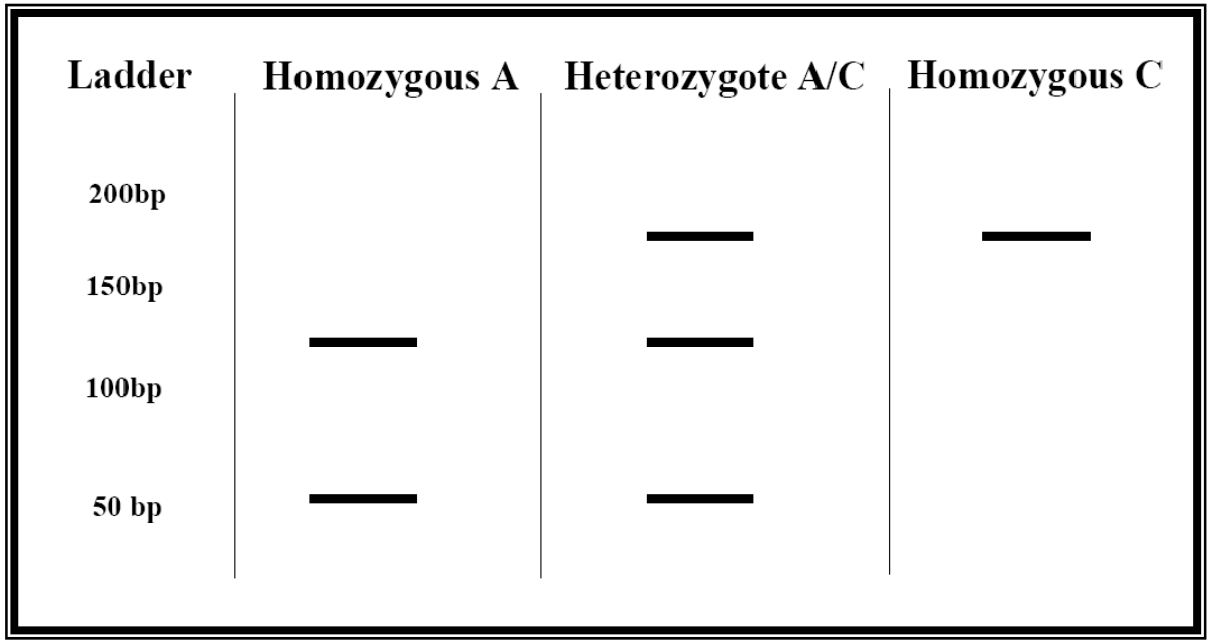
cutting by restriction endonuclease

Allele C

CCGATCACTGCGGCAA
GGCTAGTGACGCCGTT

no cut

- neumožní nalézt novou variantu daného SNP (odliší pouze 2 formy daného znaku: +/-)



SNPs genotyping – old standards
Methods of mutation detection

(comparison of specimen's pattern with pattern of known allelles)

- Thermal gradient gel electrophoresis (TGGE)
- Denaturing gradient gel electrophoresis (DGGE)
- Single-strand conformation polymorphism (SSCP)

- = special electrophoresis methods based on differences in mobility of different DNA sequences

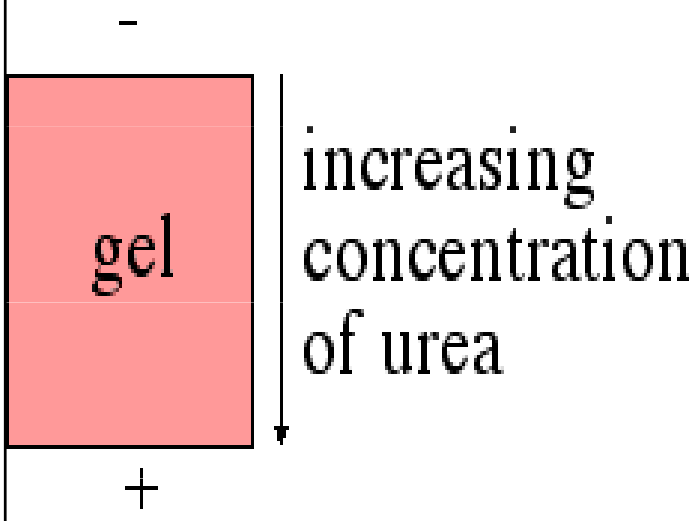
- detekce geneticky podmíněných chorob, např. cystická fibróza

Denaturing gradient gel electrophoresis (DGGE)
(TGGE – podobné, ale gradient teploty)

The small (200-700 bp) genomic fragments are run on a low to high denaturant GRADIENT acrylamide gel

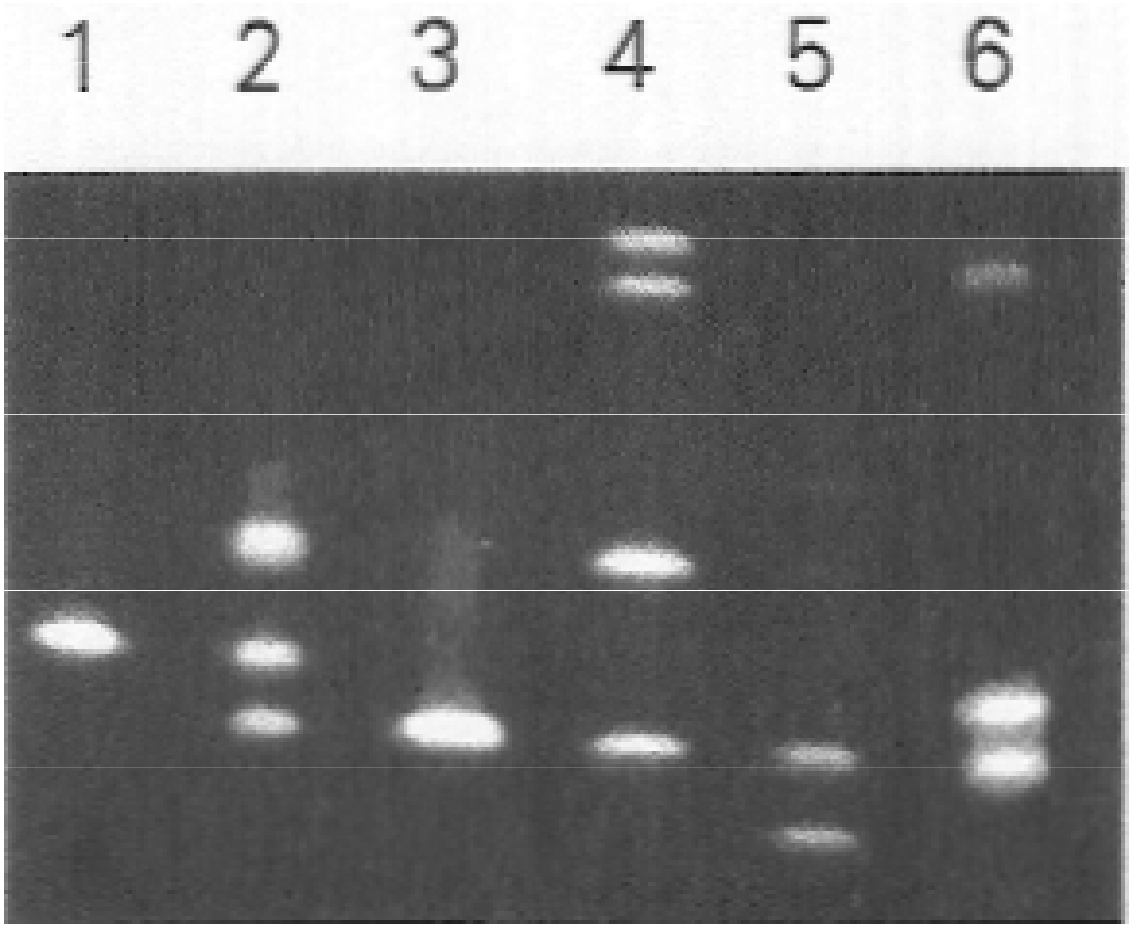
Each fragments move according to molecular weight, but as they progress into more denaturing conditions, each (depending on its sequence composition) reaches

A POINT where the DNA BEGINS TO MELT



They retard, and we will see shift in mobility

We will see different shifts in mobility for differing products



1- normal homozygote
3- homozygous mutations will yield one band on a different position
2, 4, 5, 6 – heterozygous mutations will yield 4 bands (2 homozygous and 2 heterozygous)

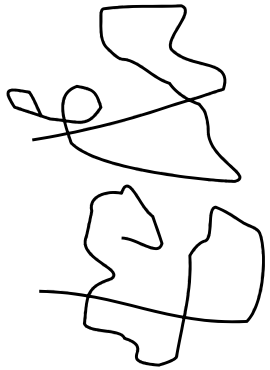
NOT ALL BANDS ARE SEEN !!!!!

www.leveninc.com/cftr_ex.gif

Single strand conformation polymorphism (SSCP)

Allele 1

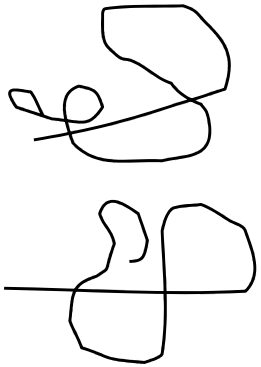
...CGCTTCAGG ...
...GCGAAGTCC...



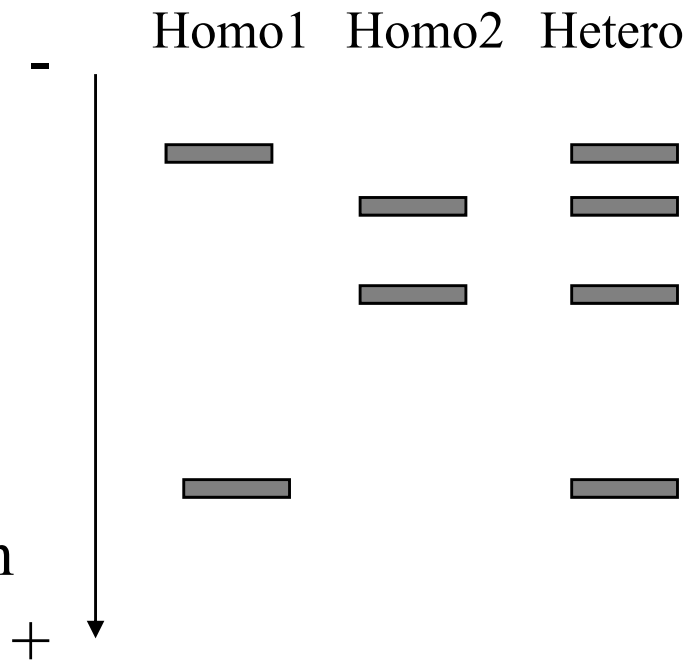
heating - denaturation
snap-cooling → partial renaturation

Allele 2

...CGCTTAAGG ...
...GCGAATTCC...



sequence-specific
ssDNA conformations

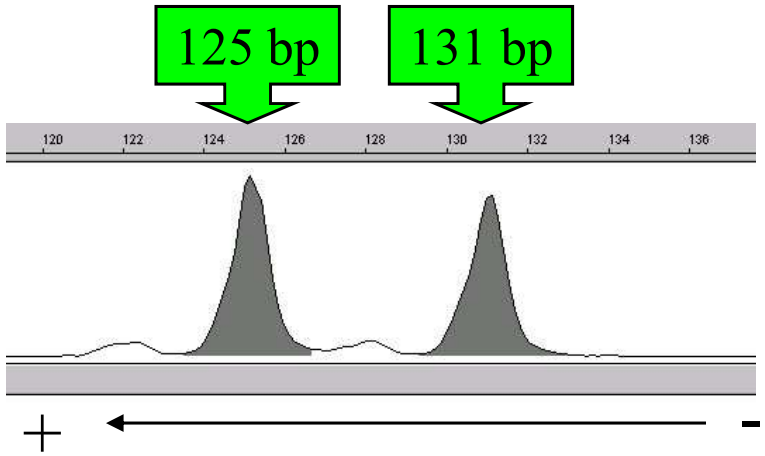


!!! non-denaturing PAGE

radioisotopes
silver-staining
fluorescent dyes (SYBR gold)

The use of automated sequencers

(denaturing polymer POP7 – ssDNA, e.g. microsatellites)



Well controlled electrophoresis parameters, high sensitivity

The use of automated sequencers

Why not non-denaturing electrophoresis?

CAP (conformation analysis polymer) – Applied Biosystems

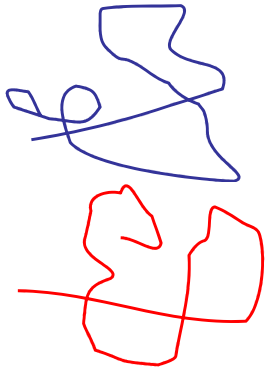
- well controlled electrophoresis

- two fluorescent labels

- high sensitivity

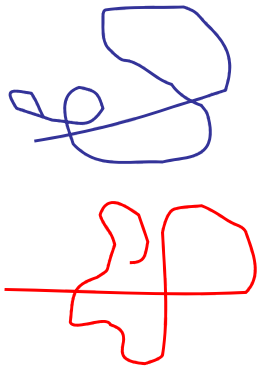
Allele 1

FAM... **C**GCTT**C**AGG ...
... GCGAA**G**TC**C** ***...HEX***

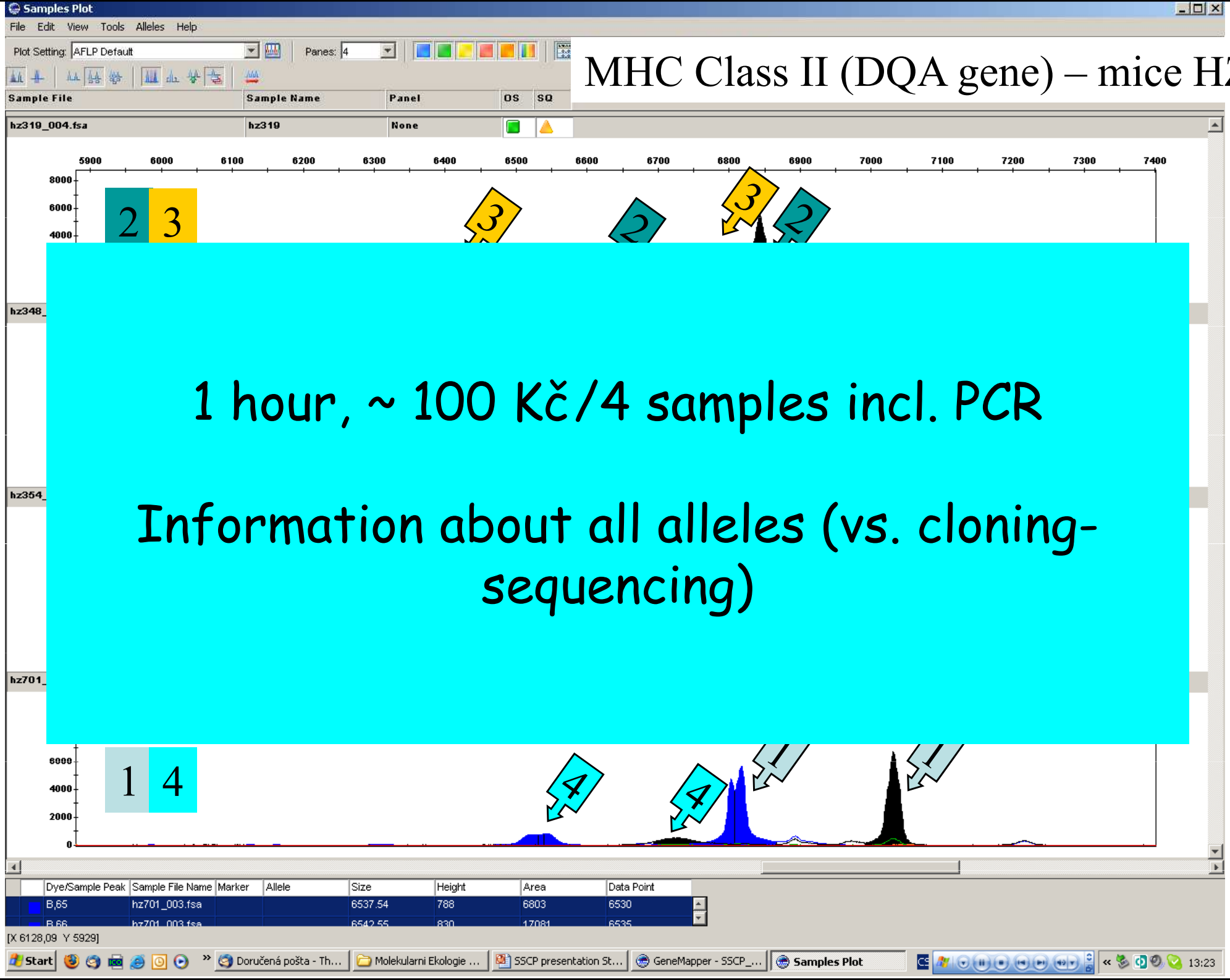


Allele 2

FAM... **C**GCTT**A**AGG ...
... GCGAA**T**TC**C** ***...HEX***

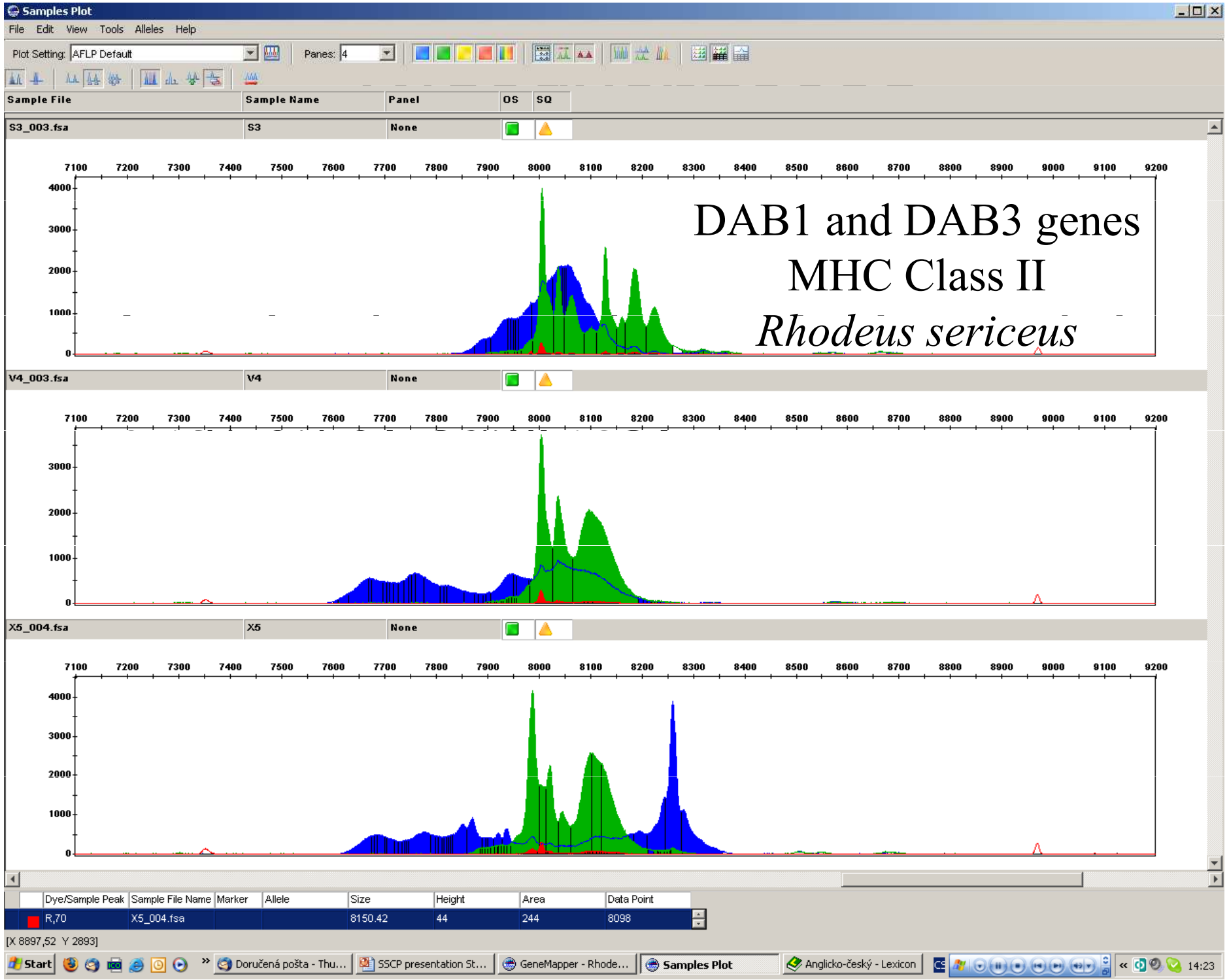


MHC Class II (DQA gene) – mice HZ



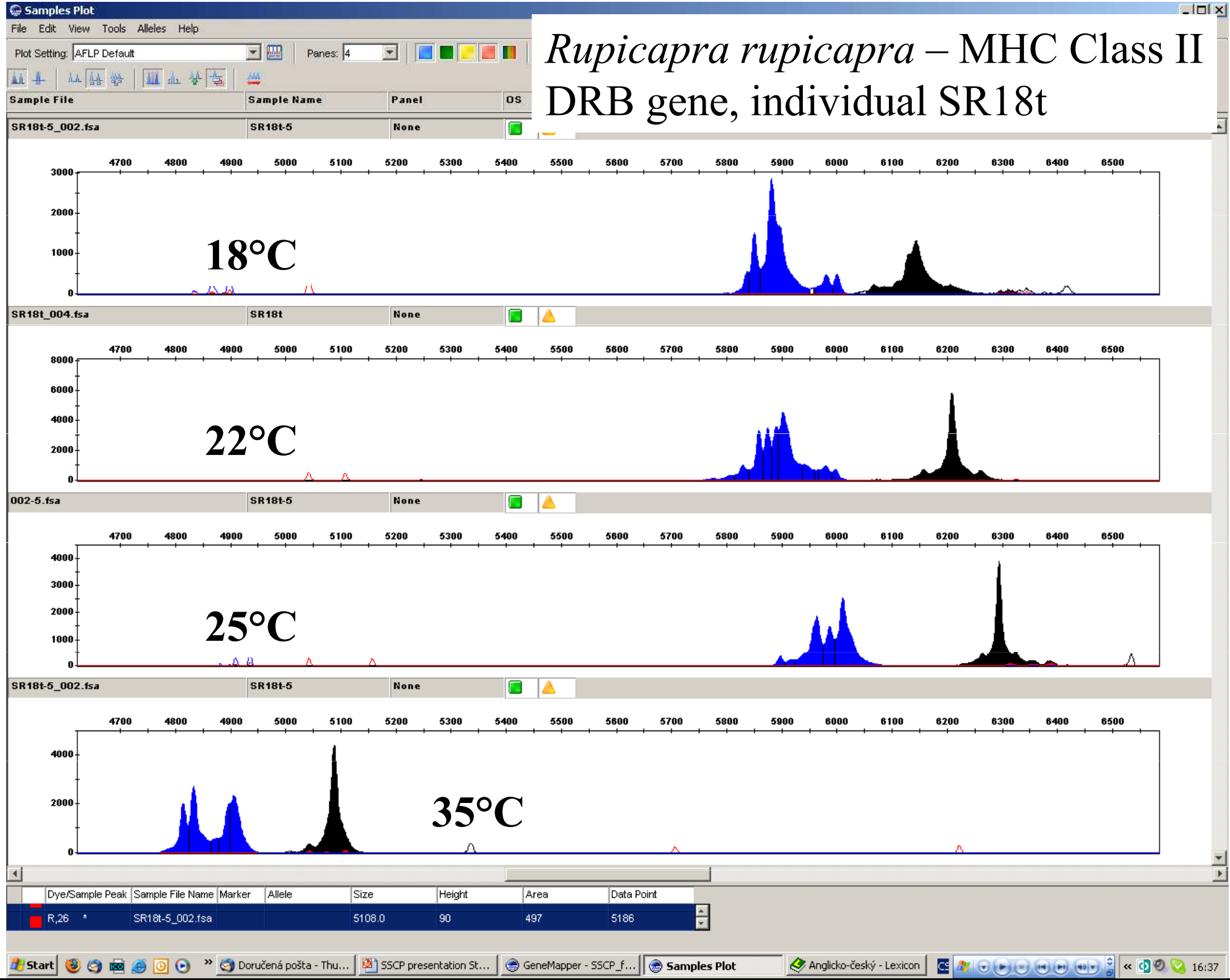
Advantages of CE-SSCP

- high throughput (when using 4, 16, or 96 - capillary sequencer) – time and money saving
- no need of gel preparation and autoradiography
- distinction of two DNA strains by two colour-labeling (usually FAM and HEX)
- potential of multiplexing – not yet used !!!


Samples Plot
File Edit View Tools Alleles Help
Plot Setting: AFLP Default
Panes: 4
Sample File
Sample Name
Panel
OS
SQ
S3_003.fsa
S3
None
V4_003.fsa
V4
None
X5_004.fsa
X5
None
DAB1 and DAB3 genes
MHC Class II
Rhodeus sericeus
Dye/Sample Peak
Sample File Name
Marker
Allele
Size
Height
Area
Data Point
R,70
X5_004.fsa
8150.42
44
244
8098
[X 8897,52 Y 2893]
Start
Doručená pošta - Thu...
SSCP presentation St...
GeneMapper - Rhode...
Samples Plot
Anglicko-český - Lexicon
14:23

Disadvantages

- need for electrophoresis optimisation (running temperature, sieving matrix, dilution of samples)
- „complex" patterns in some sequences
- alleles with the same pattern may rarely occur
- it is necessary to test several run temperatures



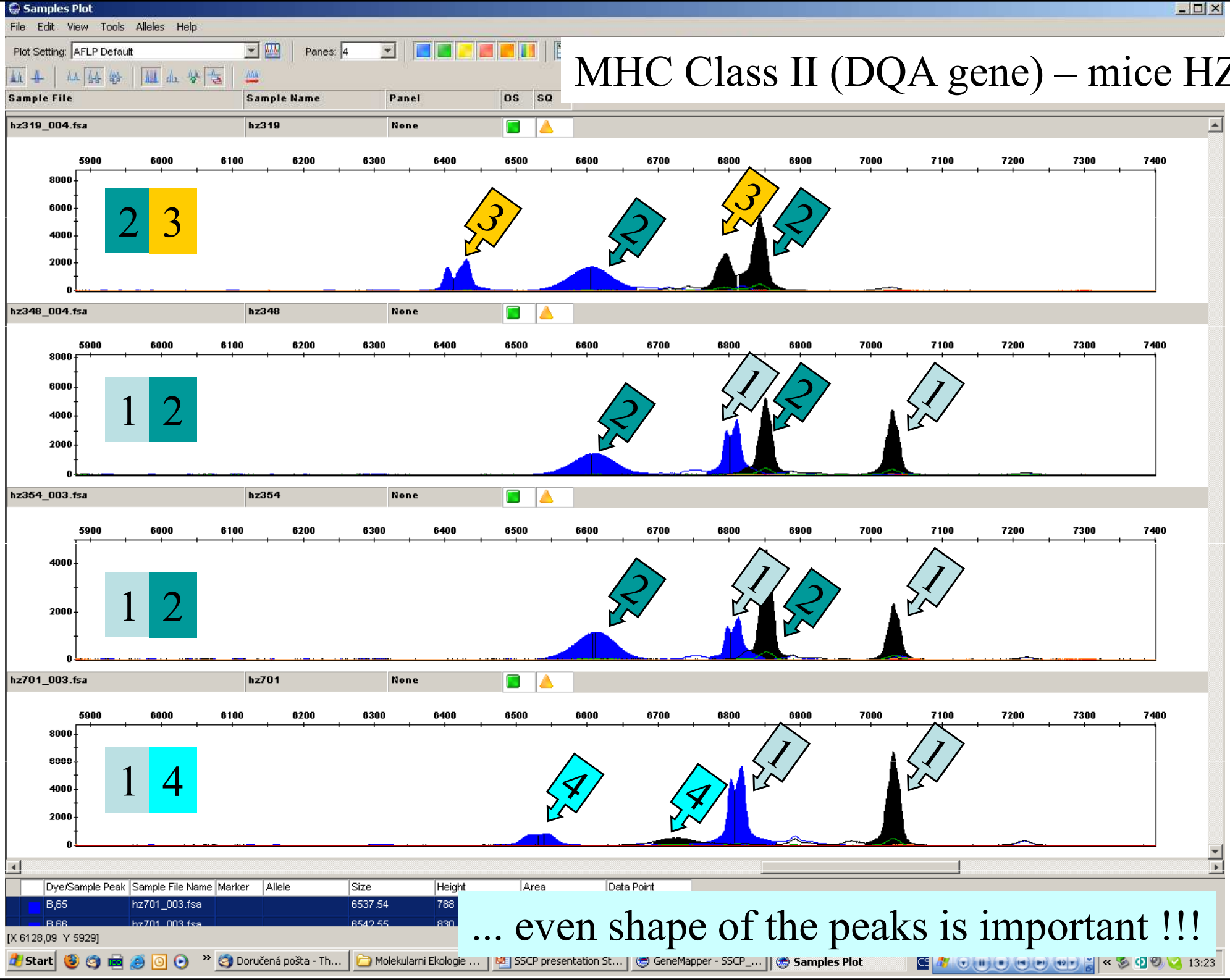
Rupicapra rupicapra – MHC Class II DRB gene, individual SR18t

Data analysis

- GeneMapper (Applied Biosystems)
- different „Size Standard" for each temperature
- alignement of more samples

Applications

1) Genotyping of codominant markers (e.g. single copy MHC genes)



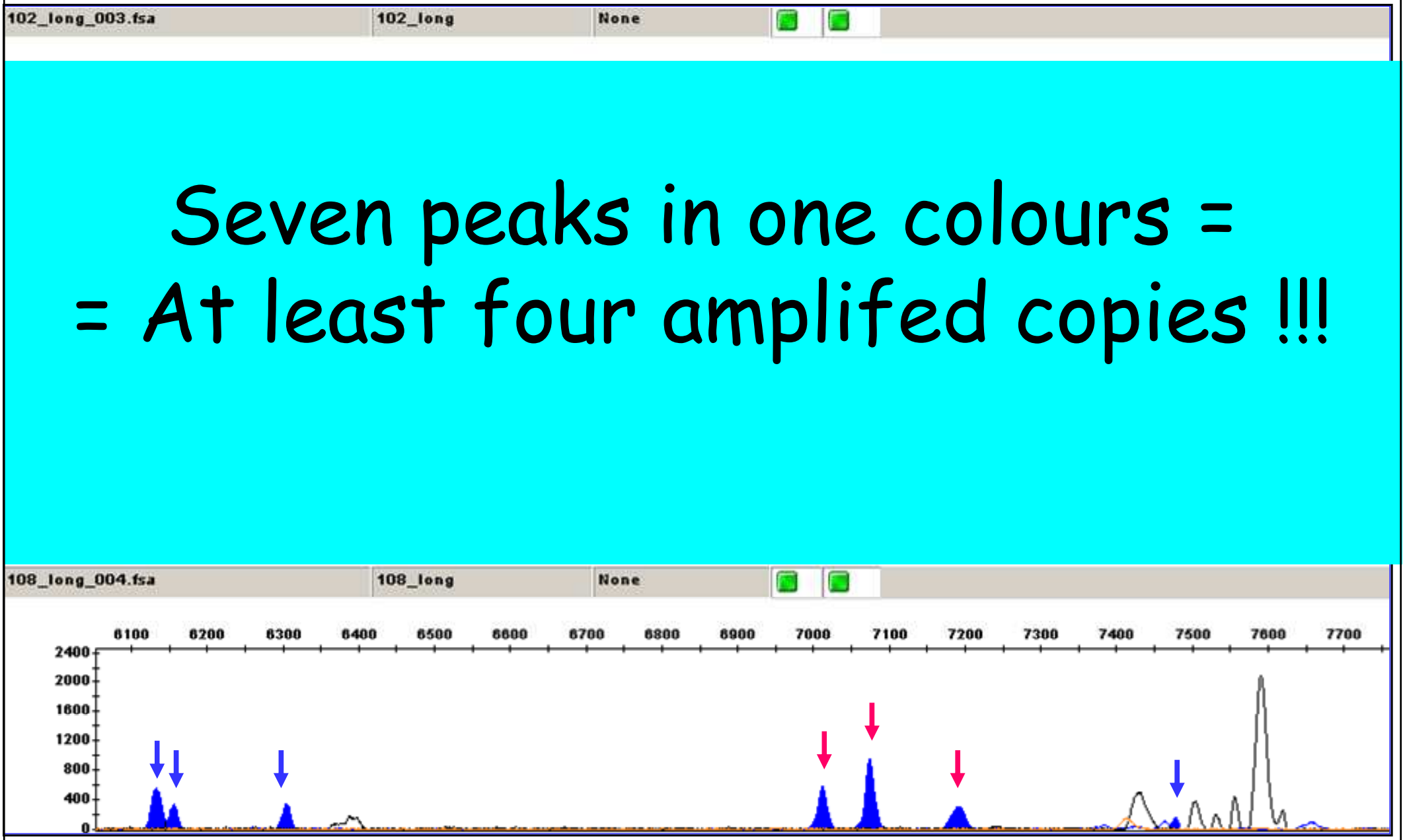
MHC Class II (DQA gene) – mice HZ

... even shape of the peaks is important !!!

Applications

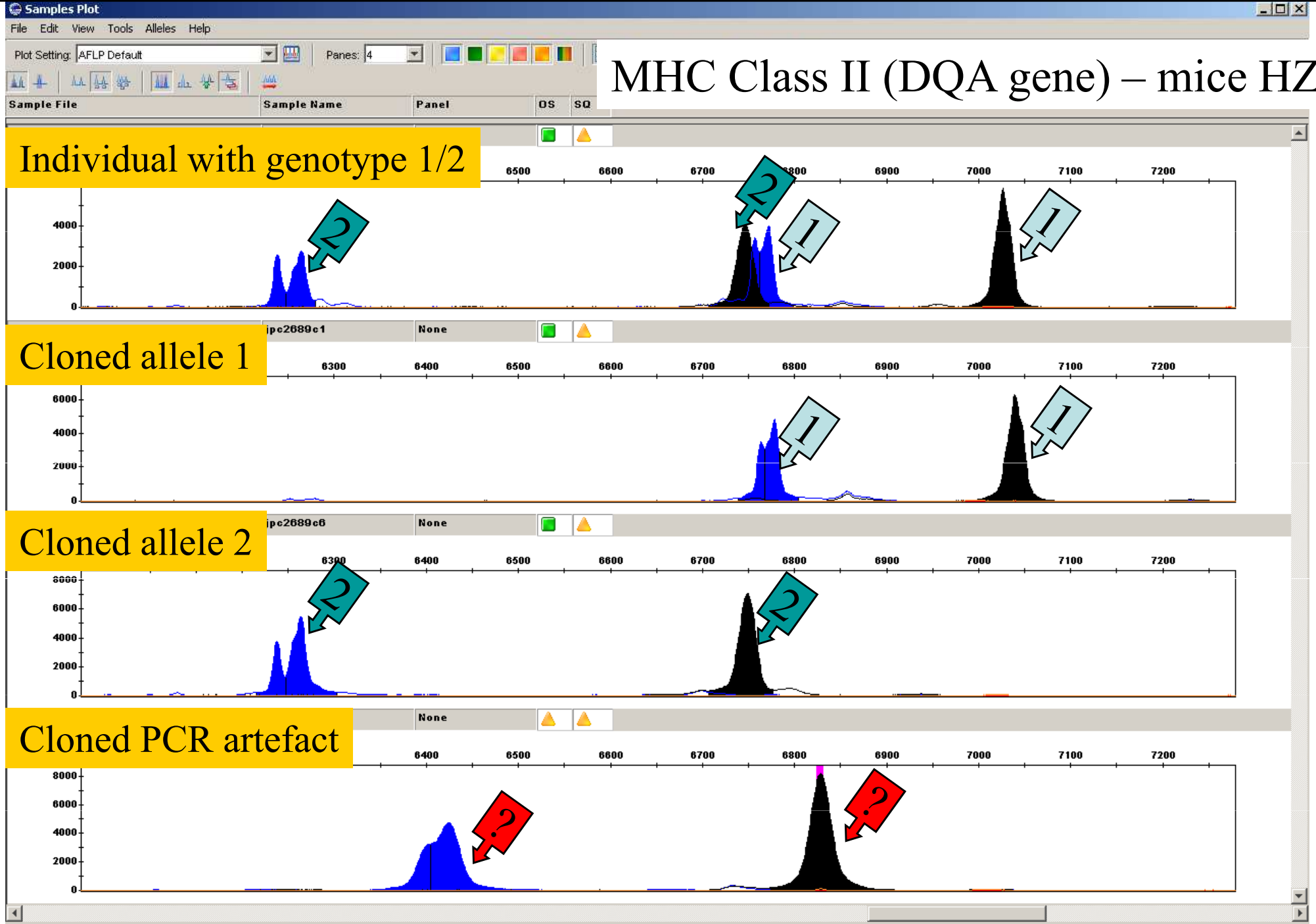
1) Genotyping of codominant markers (e.g. single copy MHC genes)
2) Identification of number of genes (e.g. duplicated MHC genes)

Seven peaks in one colours =
= At least four amplifed copies !!!



SSCP of three individuals: - different alleles - same alleles

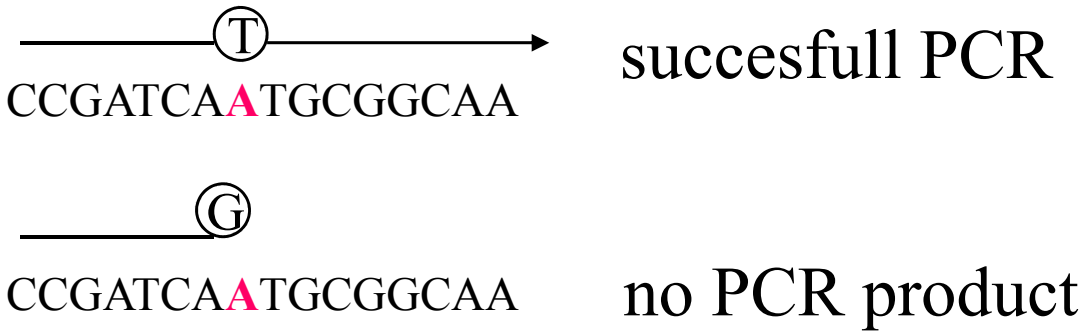
Carpodacus erythrinus – MHC Class I (Promerová et al. 2007)


Samples Plot
File Edit View Tools Alleles Help
Plot Setting: AFLP Default
Panes: 4
Sample File
Sample Name
Panel
OS
SQ
MHC Class II (DQA gene) – mice HZ
Individual with genotype 1/2
Cloned allele 1
Cloned allele 2
Cloned PCR artefact
None
Detection of PCR artefacts during cloning of heterozygotes
[X 7235,83 Y 2209]
Start
Samples Plot
17:08

SNP genotyping – new methods

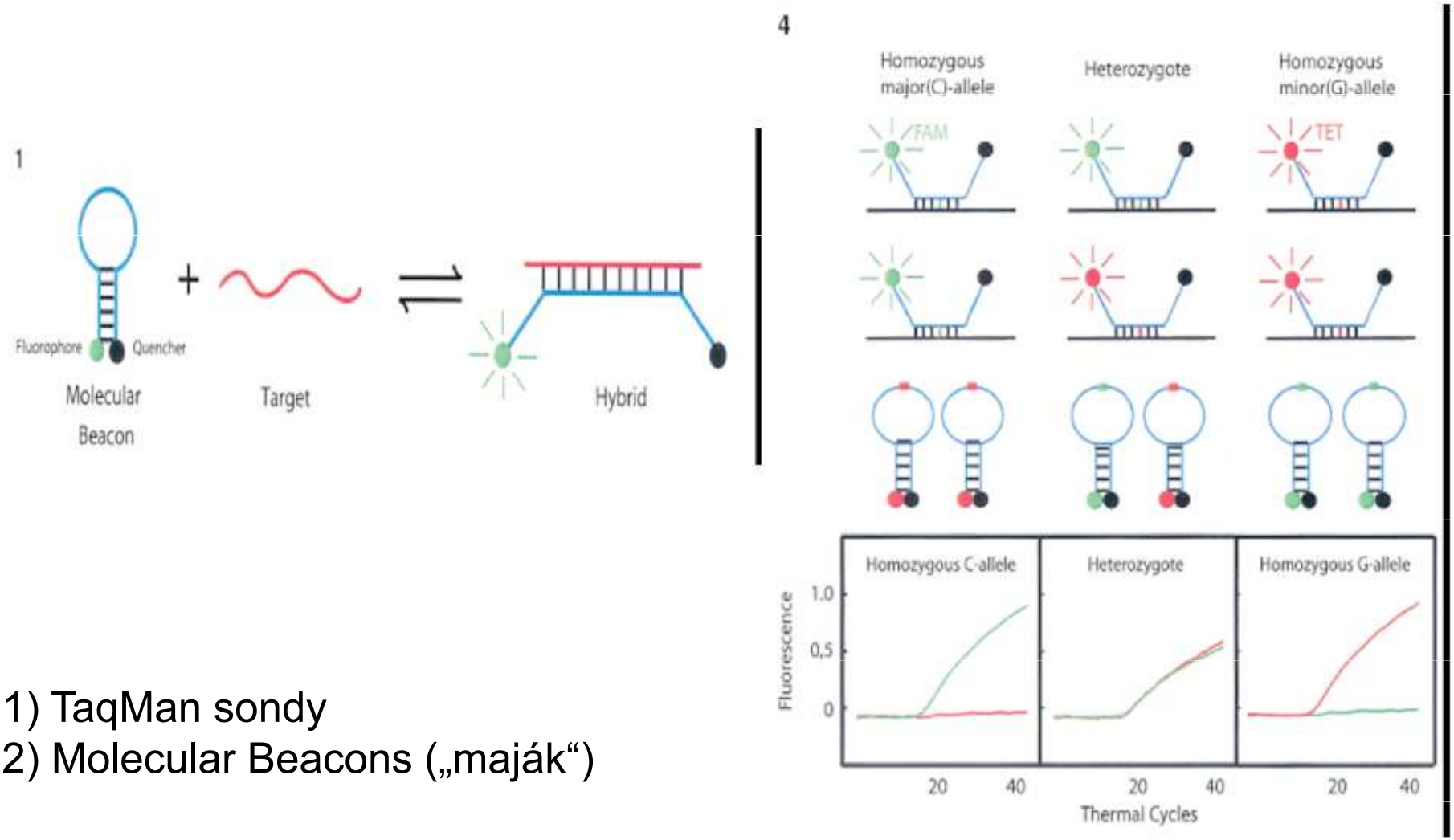
1. ASPE: allele-specific primer extension
2. real-time PCR se specifickými sondami (TaqMan, molecular beacon)
3. SBE: single base extension
4. SNP microarrays (GeneChip method)

ASPE: allele-specific primer extension



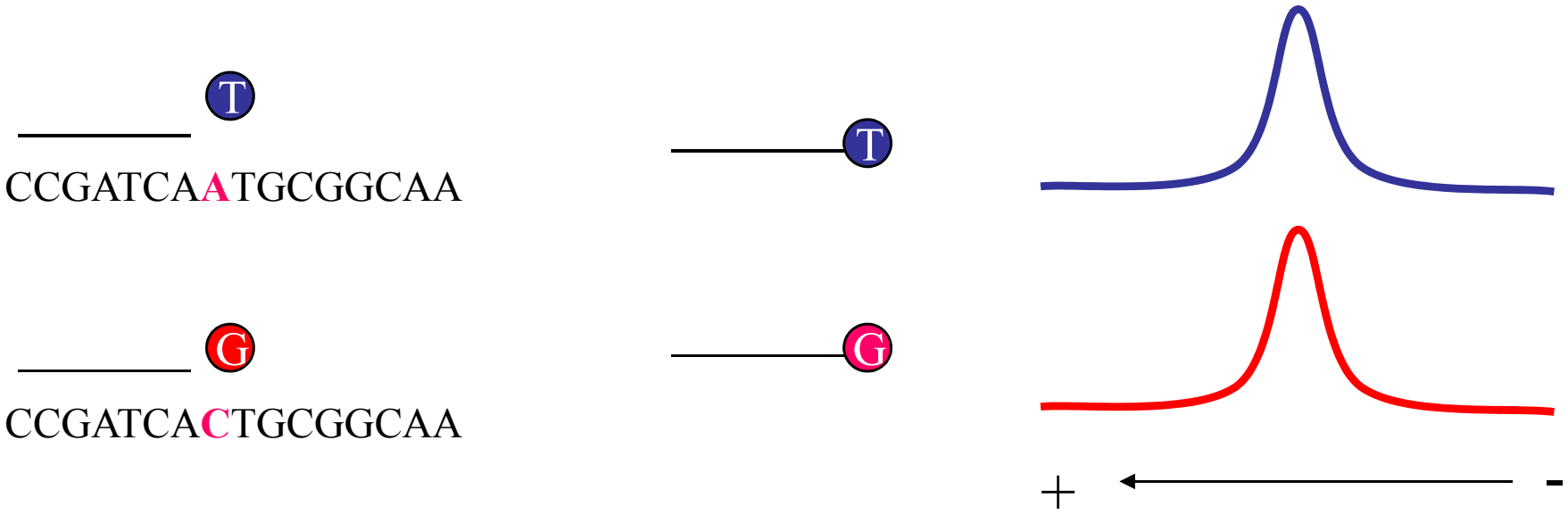
- primer extension with deoxy nucleotides and highly specific polymerase enables allele-specific amplification
- 3' terminal nucleotide of the two primers contains the SNP nucleotide
- two PCRs with specific primers are necessary

Real-time PCR se specifickou sondou



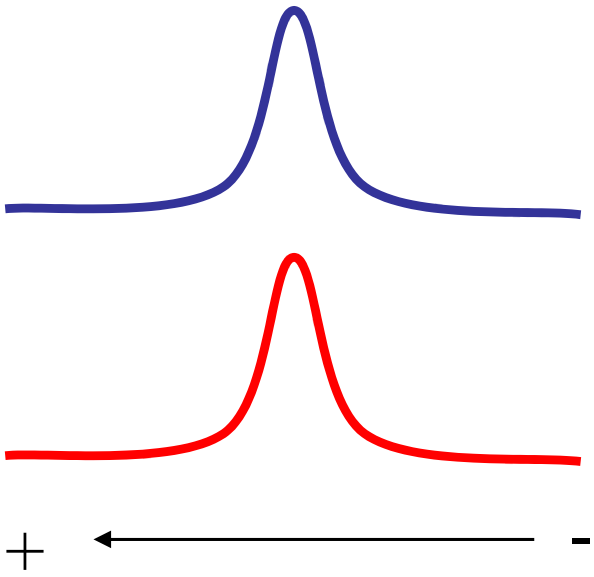
1) TaqMan sondy
2) Molecular Beacons („maják“)

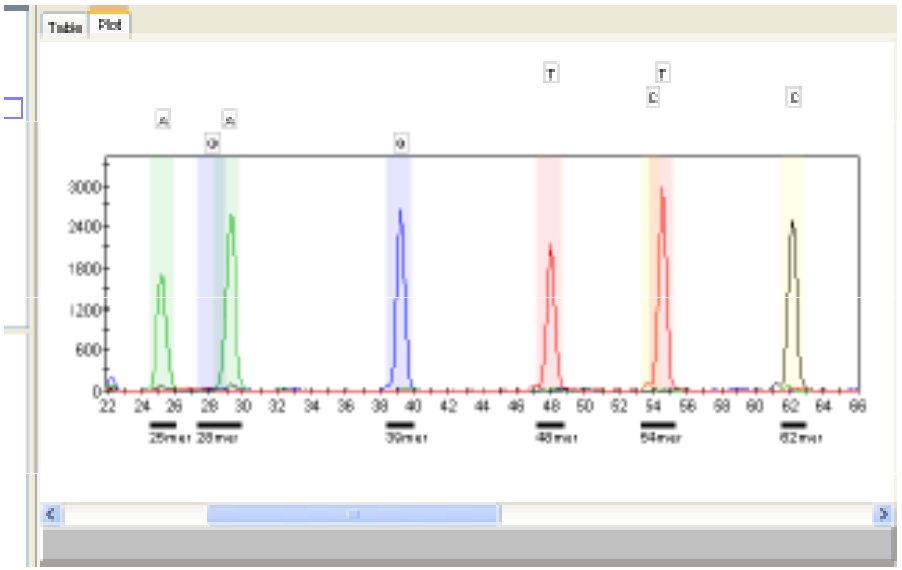
SBE: single base extension



- only one dideoxy nucleotide is added to the primer
- detekce různými metodami

Detection or SBE products





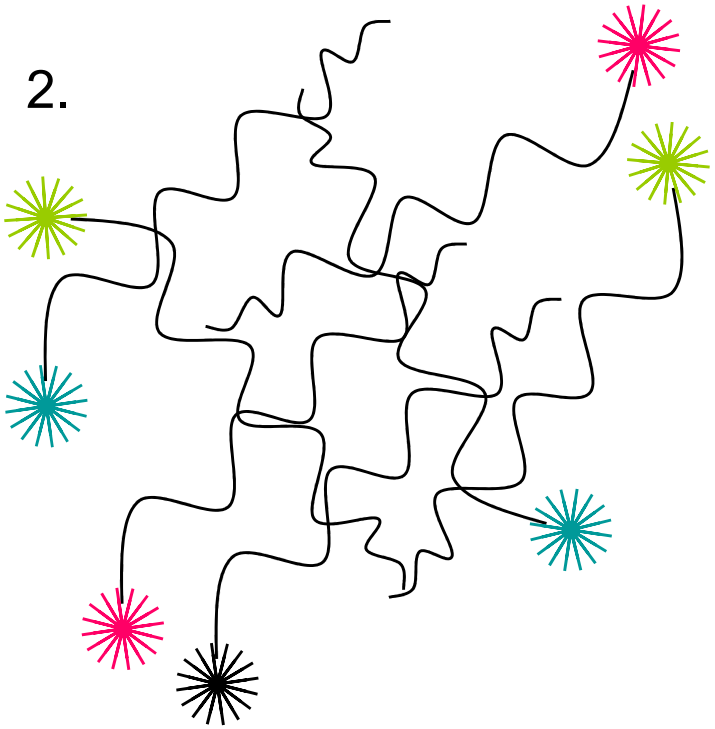
„multiplex version“ – různě dlouhé primery

electrophoresis in a capillary
SNaPShot Multiplex Kit (Applied Biosystems)

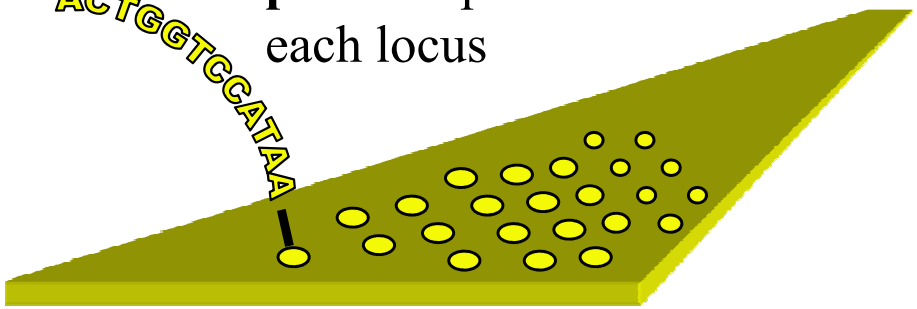
Microarray detection of SBE products

1. **tag** – specific for each locus

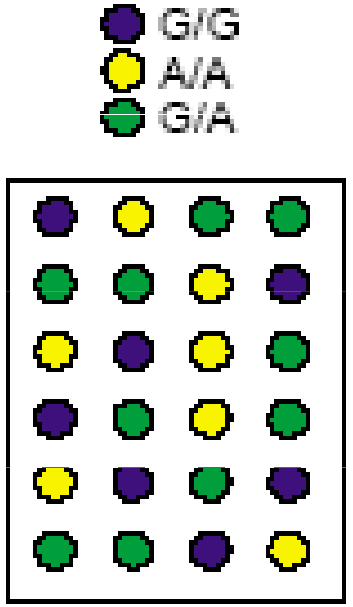
2.



3. **tag-complementary probe -** specific for each locus

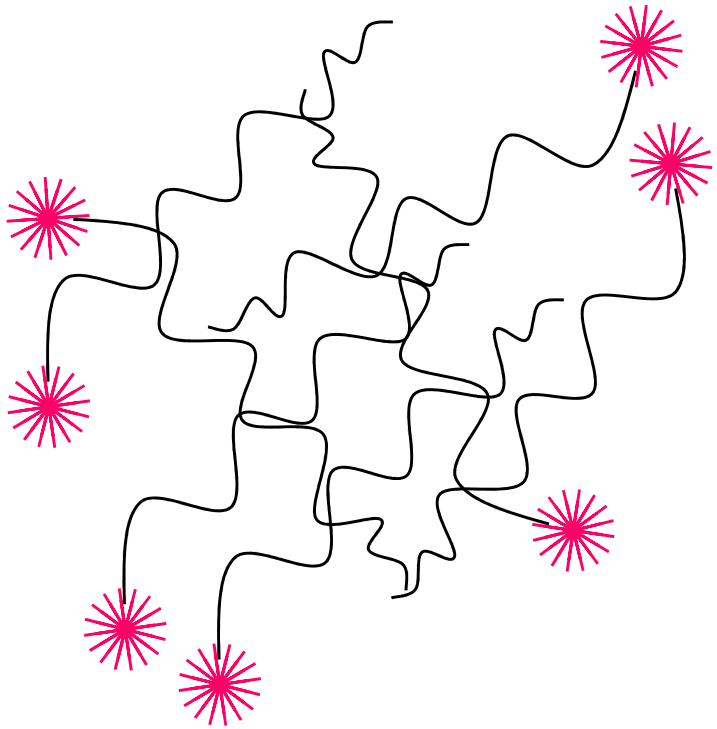


4.

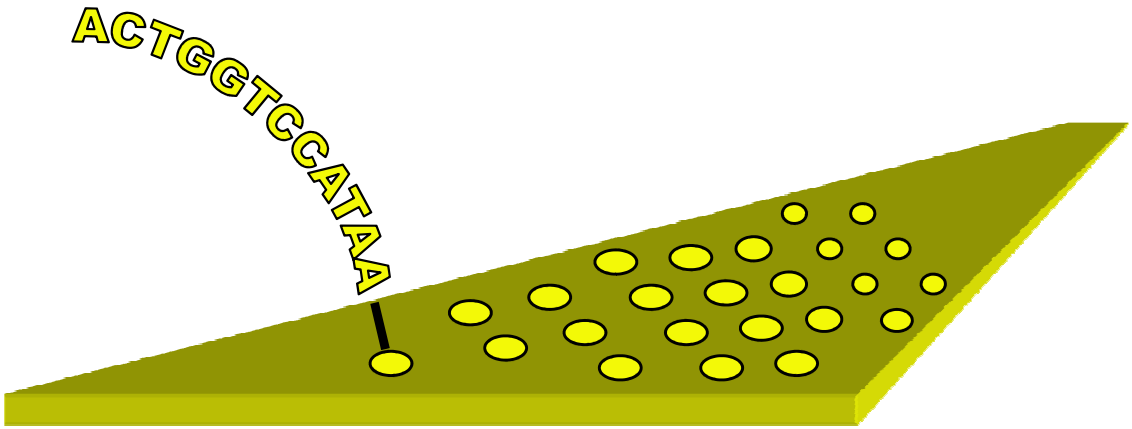


multicolor detection (using of 5’ oligonucleotide tags on SBE primers)

Microarray analysis of SNPs

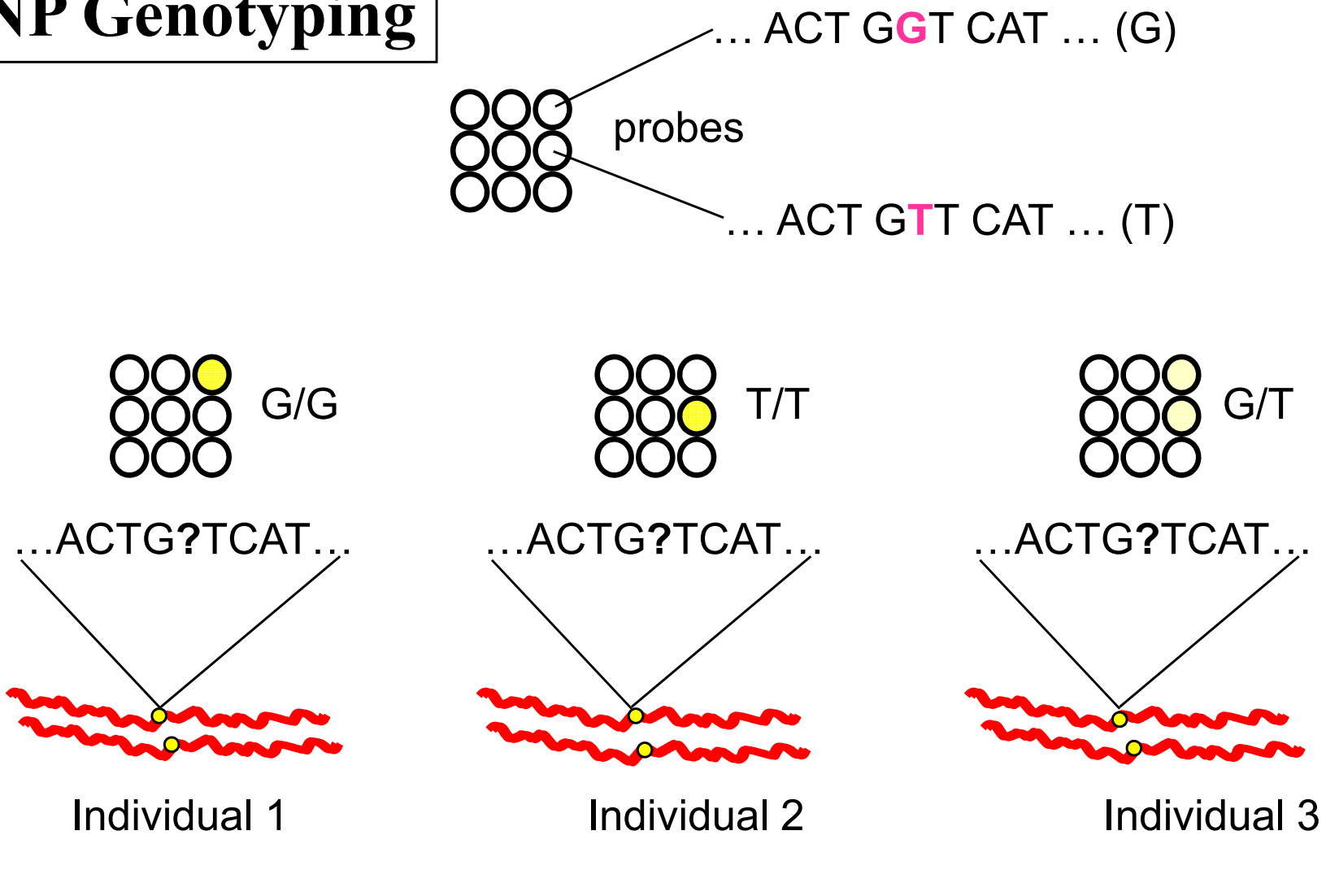


Target

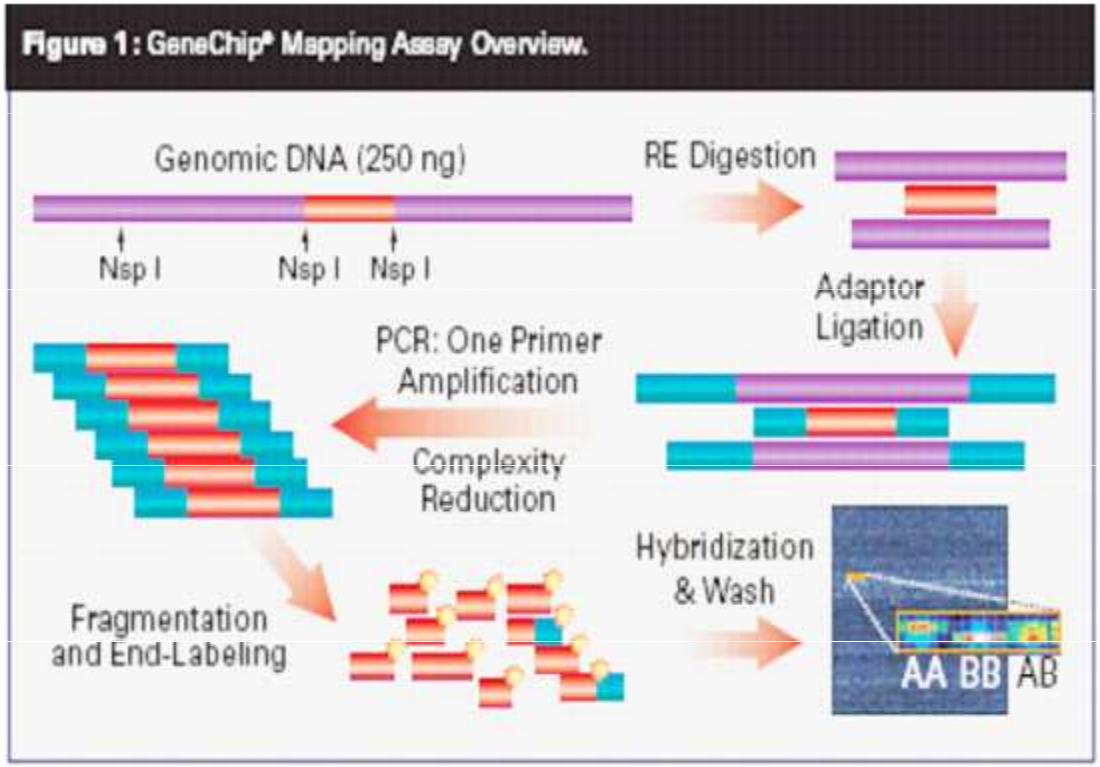


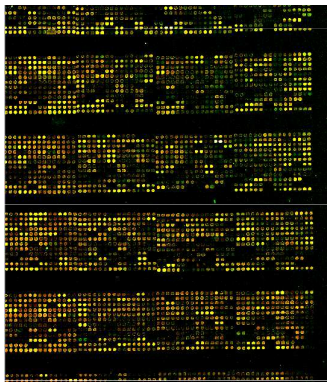
Probe

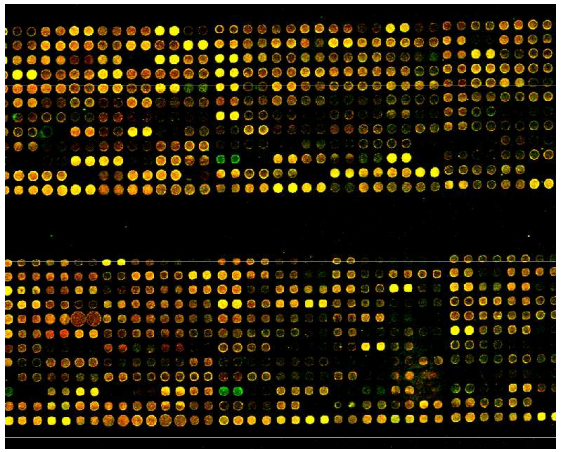
Microarray SNP Genotyping



Detekce: Affymetrix, Illumina aj.







10 – 500 tisíc SNP znaků najednou – „chip technology“

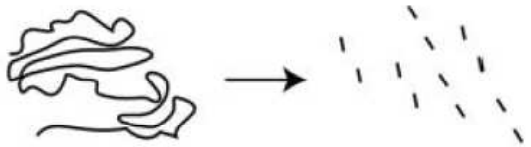
Fees - Whole Genome Genotyping									
Platform	SNP multiplex	# samples per array	# genotypes	array $	reagent $	core fee $	**Project price per sample**	Project price per genotype	volume discount bins
Affymetrix 10K	10,000	1	10,000	185	50	255	**$490.00**	$0.0490	
Affymetrix 50K	50,000	1	50,000	210	50	255	**$515.00**	$0.0103	
Affymetrix 100K (50K x2)	100,000	1	100,000	420	100	510	**$920.00**	$0.0092	
Affymetrix 250K	250,000	1	250,000	470	55	255	**$780.00**	$0.0031	
Affymetrix 500K (250K x2)	500,000	1	500,000	940	110	510	**$1,560.00**	$0.0031	
Affymetrix 500K (250K x2)	500,000	1	500,000	800	110	510	**$1,420.00**	$0.0028	1000-2000 samples
Affymetrix 500K (250K x2)	500,000	1	500,000	700	110	510	**$1,320.00**	$0.0026	2001-5000 samples
Illumina Human-1	109,000	1	109,000	800	na	110	**$910.00**	$0.0083	1-256 samples
Illumina Human-1	109,000	1	109,000	720	na	110	**$830.00**	$0.0076	257-496 samples
Illumina Human-1	109,000	1	109,000	640	na	110	**$750.00**	$0.0069	497-736 samples
Illumina Human-1	109,000	1	109,000	560	na	110	**$670.00**	$0.0061	737-976 samples
Illumina Human-1	109,000	1	109,000	480	na	110	**$590.00**	$0.0054	977+ samles
Illumina HumanHap300	317,000	1	317,000	1100	na	110	**$1,210.00**	$0.0038	1-256 samples
Illumina HumanHap300	317,000	1	317,000	990	na	110	**$1,100.00**	$0.0035	257-496 samples
Illumina HumanHap300	317,000	1	317,000	880	na	110	**$990.00**	$0.0031	497-736 samples
Illumina HumanHap300	317,000	1	317,000	770	na	110	**$880.00**	$0.0028	737-976 samples
Illumina HumanHap300	317,000	1	317,000	660	na	110	**$770.00**	$0.0024	977+ samles
Illumina HumanHapS	240,000	1	240,000	1000	na	110	**$1,110.00**	$0.0046	1-256 samples
Illumina HumanHapS	240,000	1	240,000	900	na	110	**$1,010.00**	$0.0042	257-496 samples
Illumina HumanHapS	240,000	1	240,000	800	na	110	**$910.00**	$0.0038	497-736 samples
Illumina HumanHapS	240,000	1	240,000	700	na	110	**$810.00**	$0.0034	737-976 samples
Illumina HumanHapS	240,000	1	240,000	600	na	110	**$710.00**	$0.0030	977+ samles
Illumina HumanHap550	550,000	1	550,000	1600	na	110	**$1,710.00**	$0.0031	1-256 samples
Illumina HumanHap550	550,000	1	550,000	1440	na	110	**$1,550.00**	$0.0028	257-496 samples
Illumina HumanHap550	550,000	1	550,000	1280	na	110	**$1,390.00**	$0.0025	497-736 samples
Illumina HumanHap550	550,000	1	550,000	1120	na	110	**$1,230.00**	$0.0022	737-976 samples
Illumina HumanHap550	550,000	1	550,000	960	na	110	**$1,070.00**	$0.0019	977+ samles
HumanHap300 + HumanHapS	550,000	1	550,000	1750	na	220	**$1,970.00**	$0.0036	1-256 samples
HumanHap300 + HumanHapS	550,000	1	550,000	1575	na	220	**$1,795.00**	$0.0033	257-496 samples
HumanHap300 + HumanHapS	550,000	1	550,000	1400	na	220	**$1,620.00**	$0.0029	497-736 samples
HumanHap300 + HumanHapS	550,000	1	550,000	1225	na	220	**$1,445.00**	$0.0026	737-976 samples
HumanHap300 + HumanHapS	550,000	1	550,000	1050	na	220	**$1,270.00**	$0.0023	977+ samles

Nové postupy při sekvenování DNA

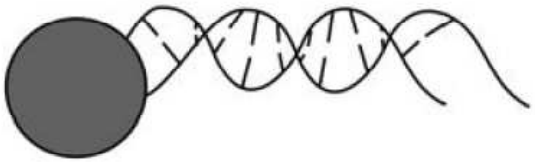
„Next generation" sequencing

(Hudson 2008)

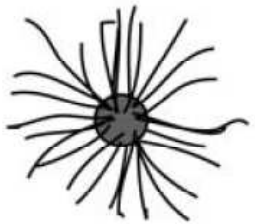
1) Randomly fragment many molecules of target DNA



2) Immoblize individual DNA molecules on solid support

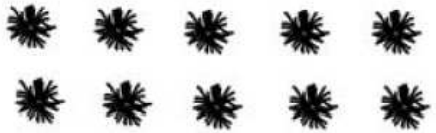


3) Amplify DNA in clonal 'polymerase colony'

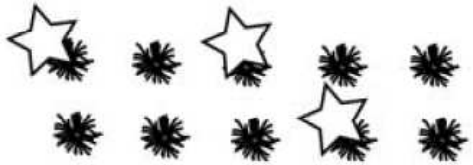


„polonies"
(polymerase colonies)

4) Sequence DNA by adding liquid reagents to immoblized DNA colonies



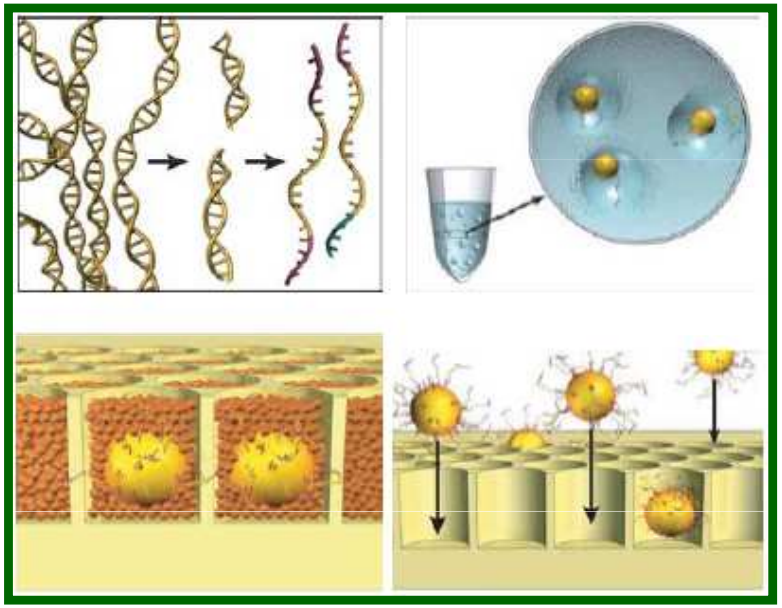
5) Interrogate sequence incorporation *in situ* after each cycle using fluorescence scanning or chemiluminescence

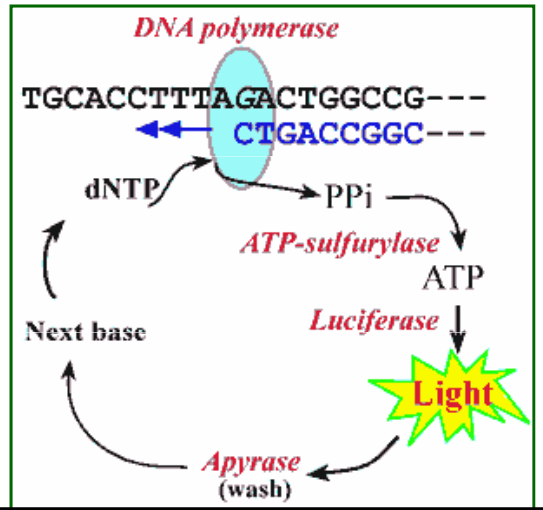


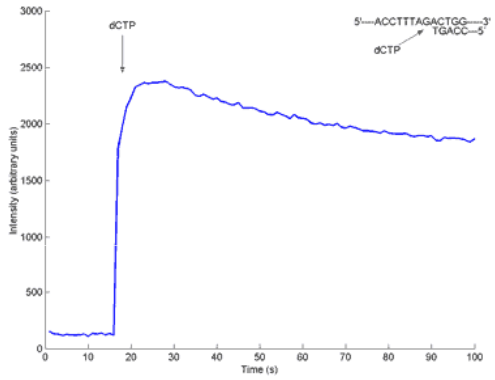
454 pyrosequencing

- emulzní techniky amplifikace
  pikolitrové objemy
- simultánní sekvenování
  na destičce z optických vláken
  detekce pyrofosfátů uvolňovaných při inkorporaci bazí
- První generace GS20
  → 200 000 reakcí najednou
  (zhruba 20 milionů bp)
  dnes FLX → 400 000 reakcí najednou
- Problémy s homopolymery
- Délka jednotlivých sekvencí 100 – 400

1 600 000 well plate



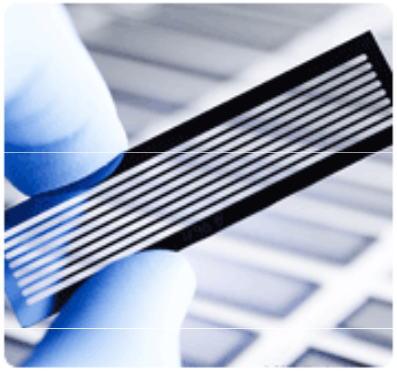




Solexa/Illumina 1G SBS technology

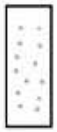
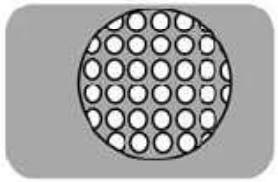
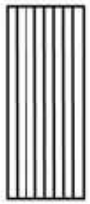
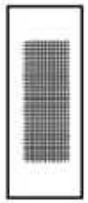
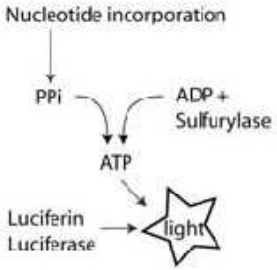
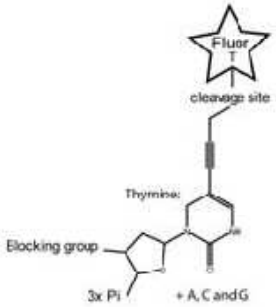
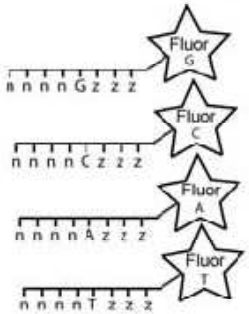
(SBS = sequencing by synthesis)

- 1 Gb (šestinásobek genomu *Drosophily*)
- Výrazně levnější
- Sekvence délky 35 bp
- Flourescence, reversibilní terminátory
- Spíš pro resequencing



SOLiD

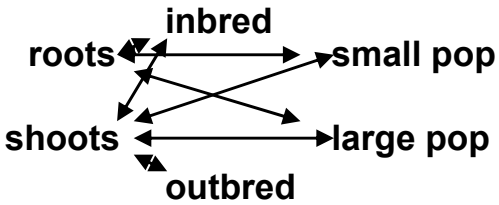
(sequencing by Oligonucleotide Ligation and Detection)

	454 pyrosequencing	Solexa SBS sequencing	Agencourt / ABI SOLiD polony sequencing
	All methods ligate single, randomly sheared DNA molecules to support		
DNA support	 25–36 μm bead	surface of flow cell 	 ~1 μm bead
Amplification	 emulsion-phase PCR	 *in situ* PCR on solid surface	 emulsion-phase PCR
Sequencing surface	 1 600 000 well plate one bead per well	 8-channel flow cell clusters of DNA randomly located	 Single slide imaged in panels beads random
Sequencing chemistry	 pyrosequencing	 reversible-terminator sequencing by synthesis	 Ligation of sequence-specific labeled oligos
Sequence detection	Chemiluminescence (one channel)	Fluorescence (four channel)	Fluorescence (four channel)
Read length and number	100–400 bp $> 2 \times 10^5$ reads	35 bp $\sim 4 \times 10^7$ reads	25 bp (paired) $> 10^7$ reads

Example:

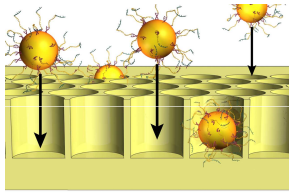
Scabiosa columbaria

Joop Ouborg et al.



Ecogenomics

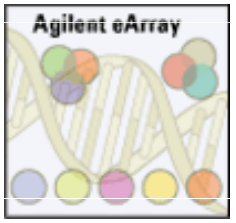
cDNA library

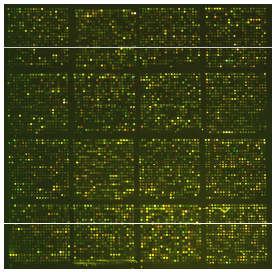


530.000 sequences in one run, leading to ~ 40.000 ESTs

454 LIFE SCIENCES

NCBI





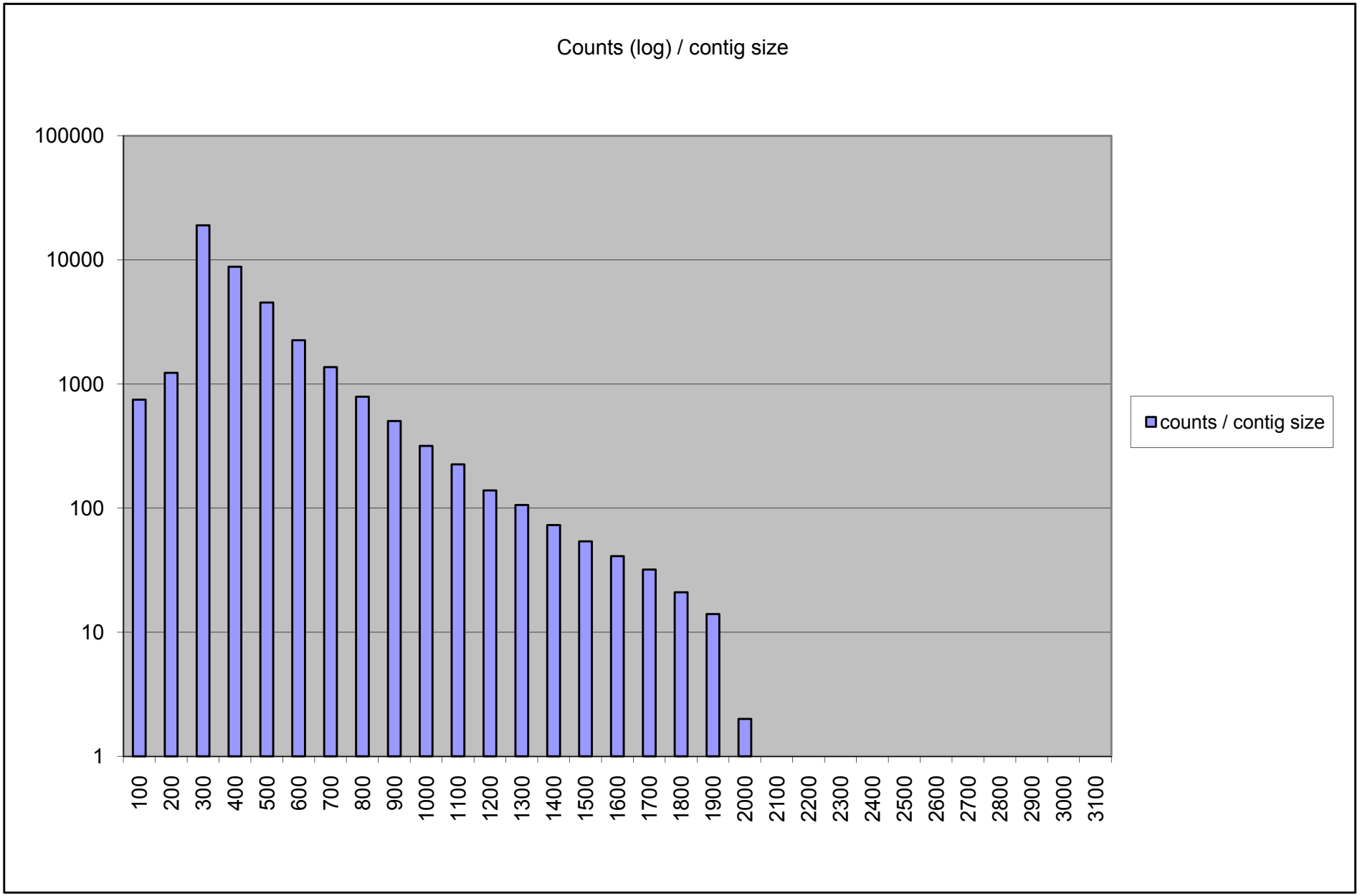
15k – 30k 60-mer microarrays

Agilent Technologies

Experiment 1: transcriptional profiling of inbreeding depression

Experiment 2: transcriptional profiling of genetic erosion

Do transcript profiles of 1 and 2 match????



Total number of reads: **528557** Number of contigs: **40302**

In the next phase:

Annotation of these 40.000+ ESTs

Automated programs available, like **BLAST2GO (http://www.blast2go.de/)**:

just feed a file with the ESTs into the program, and turn it on……

1 week later you will have the results, being:

- Homology with known sequences
- Known function

The sequences may also be searched for:

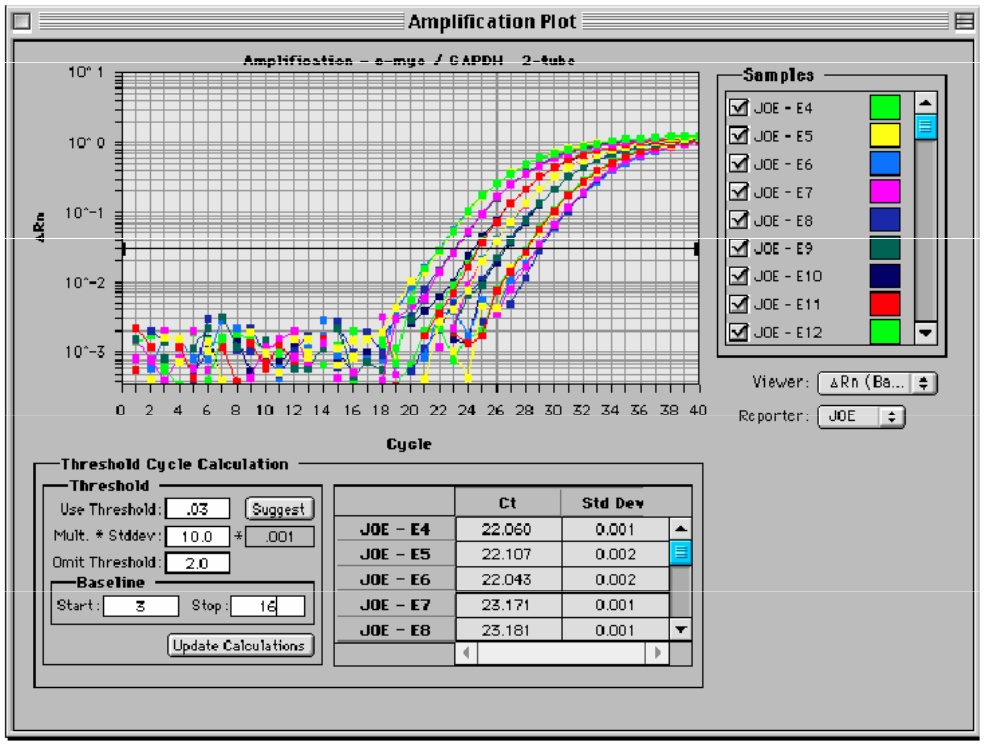
EST-associated SSR markers: MISA (http://pgrc.ipk-gatersleben.de/misa/)

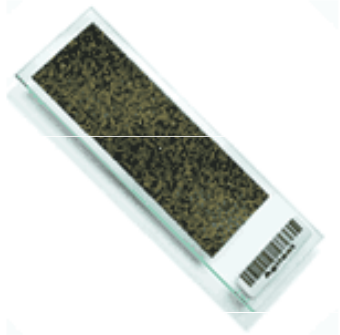
SNP markers: SNP-mining software like PolyBayes (http://genome.wustl.edu/tools/software/polybayes.cgi)

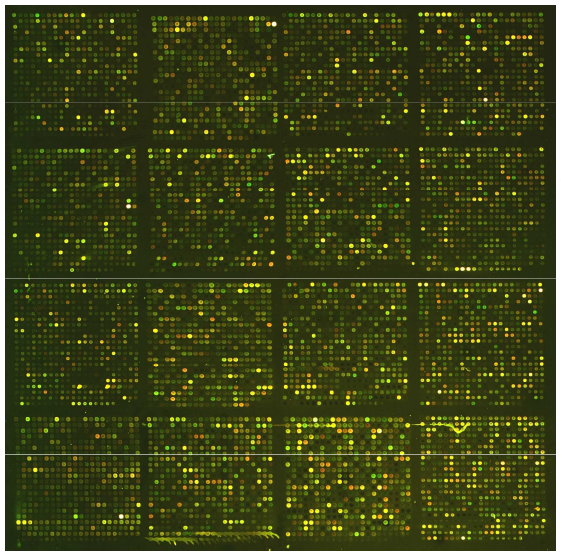
Again by using search software, freeware

Design of quantitative RealTime-PCR methods, based on EST sequences

Design of a Scabiosa specific microarray







Ecogenomics

Expected pay-off:

- Ecogenomic approach to conservation genetics leads to insight in effects of genetic erosion on functional genetic variation
- How does genetic erosion affect evolutionary potential?
- What is the balance between genetic drift and natural selection in effects of habitat fragmentation?
- Are there general inbreeding depression genes, or is inbreeding depression a random phenomenon?
- Which genes are involved in inbreeding depression in different life history stages, and can this explain the non-correlation of IBD between these stages?
- What are the footprints of selection in the genomes of individuals from small and large populations?
- What is the selective value of variation in gene expression?

Ecogenomics

Costs/requirements:

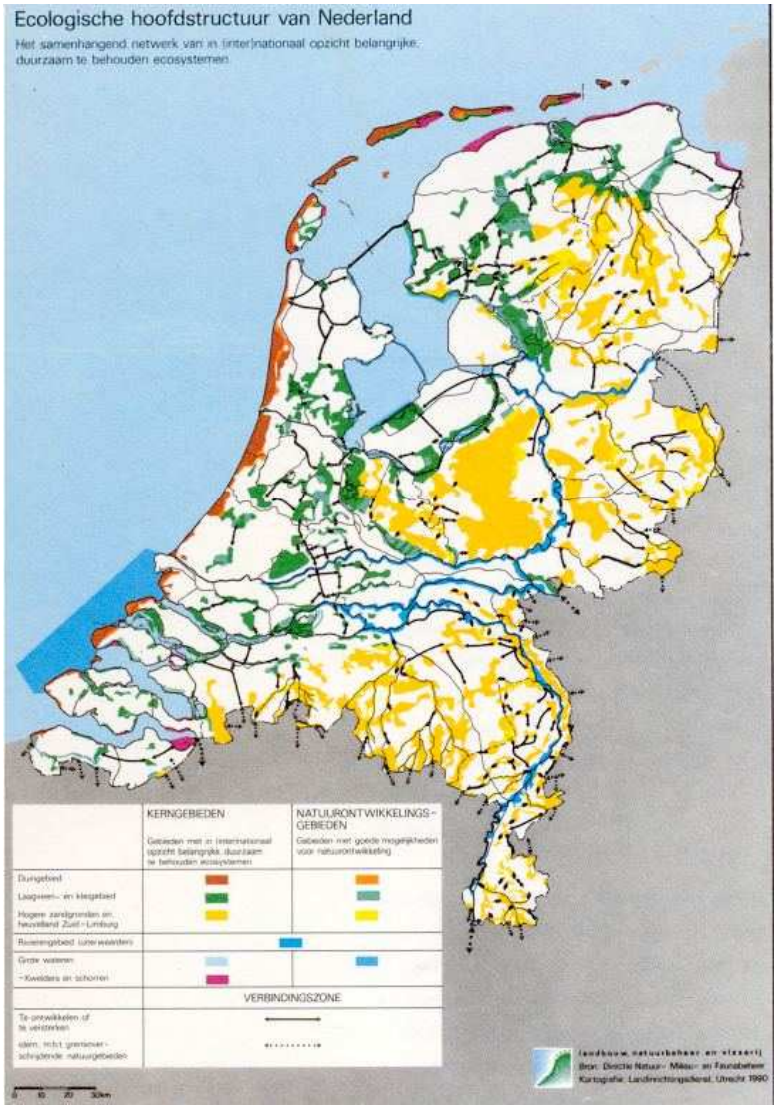
Costs are diminishing continuously

454 FLX-cDNA sequencing : 1 month, 15.000 € (used to be 200.000 € with Sanger technology)

microarray production: 100 € per array
microarray screening: 150 € per array

cheaper options (like SOLEXA technology) are becoming available, at much lower costs

Relative costs of conservation genomics:



Projected costs (but this is almost certain a severe underestimation):

20 billion Euro

That is:

20.000.000.000 Euro

That is equivalent to 40.000.000 microarray runs…….