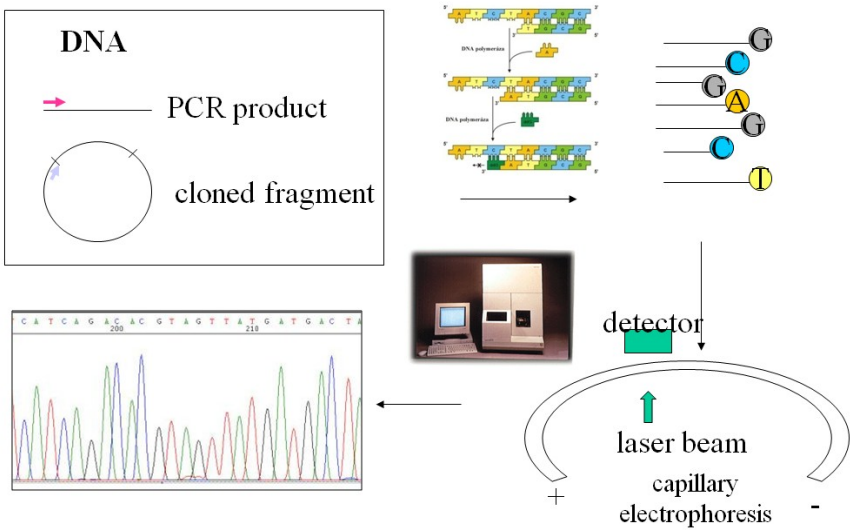
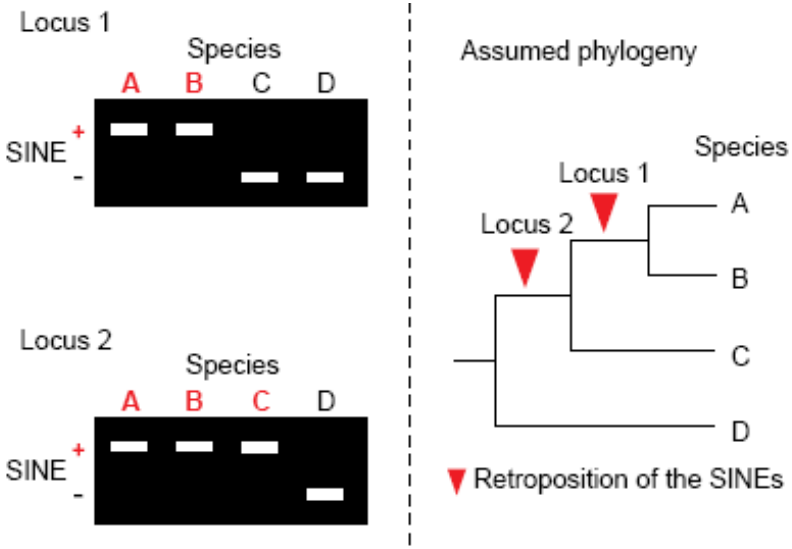
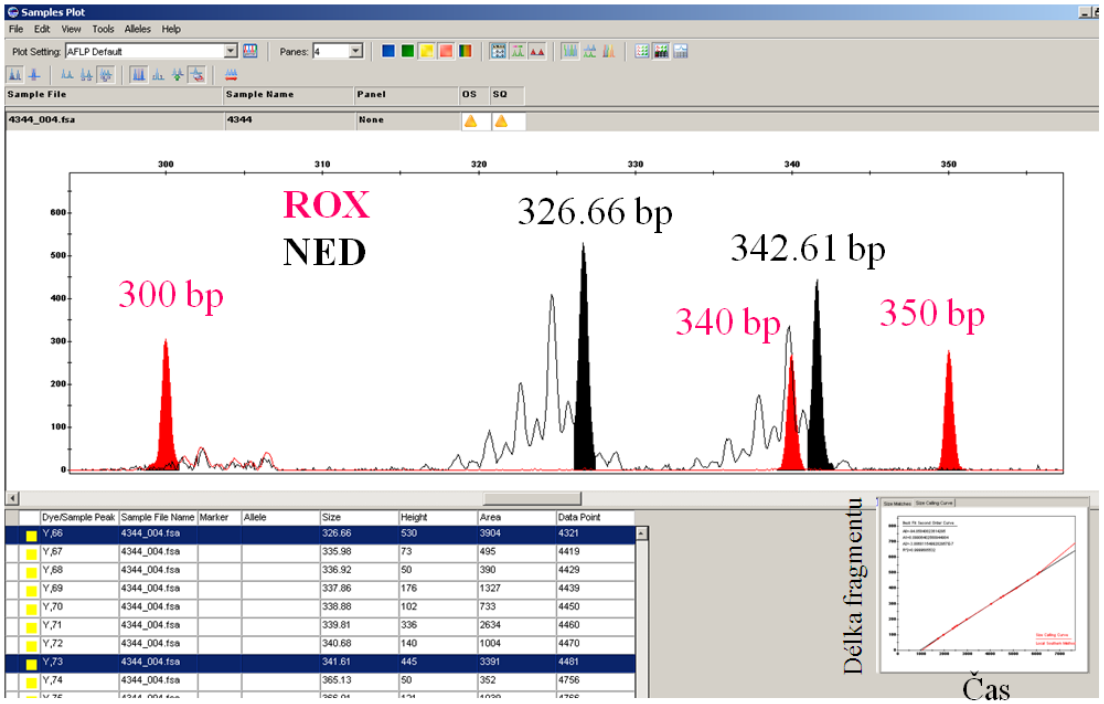
Sekvencování DNA



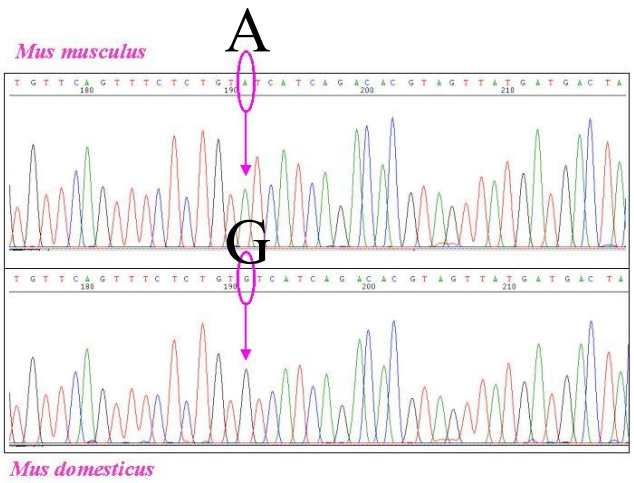
SINE

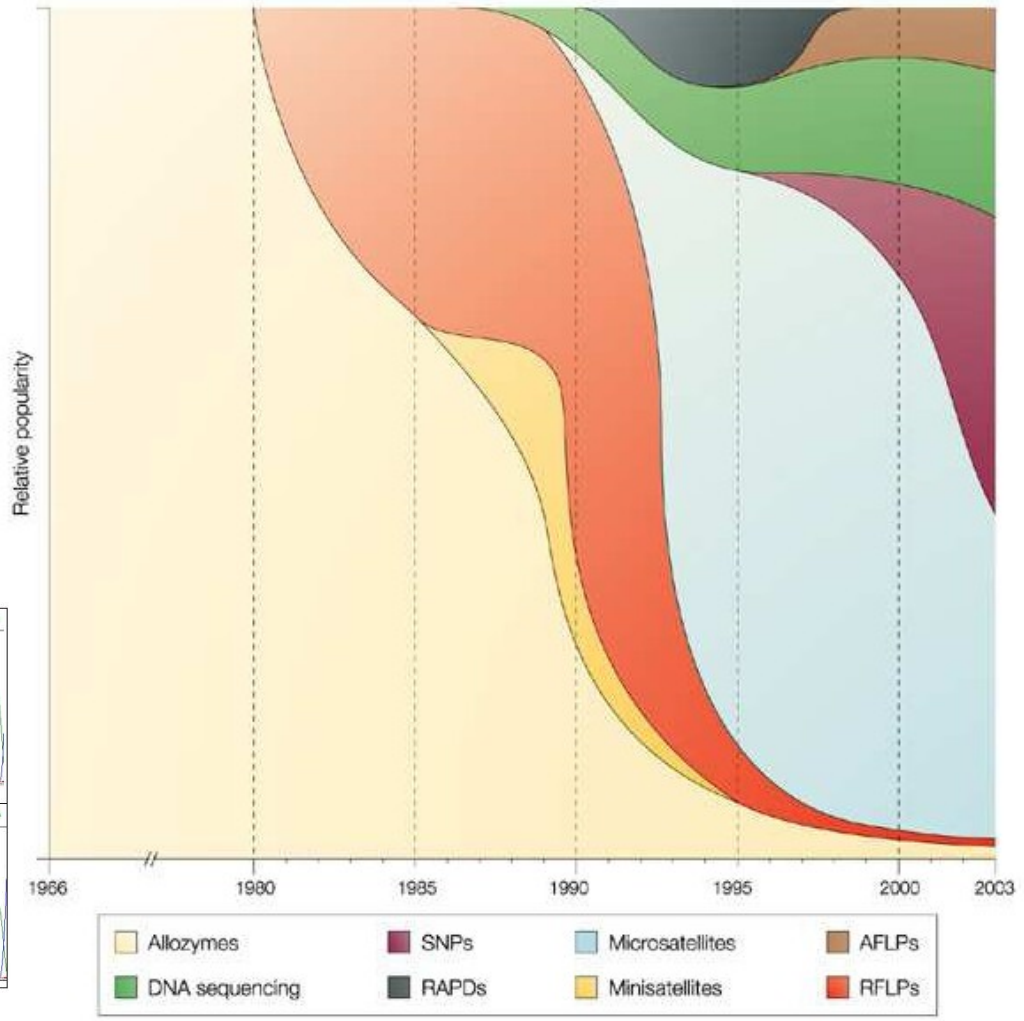


Mikrosatelity



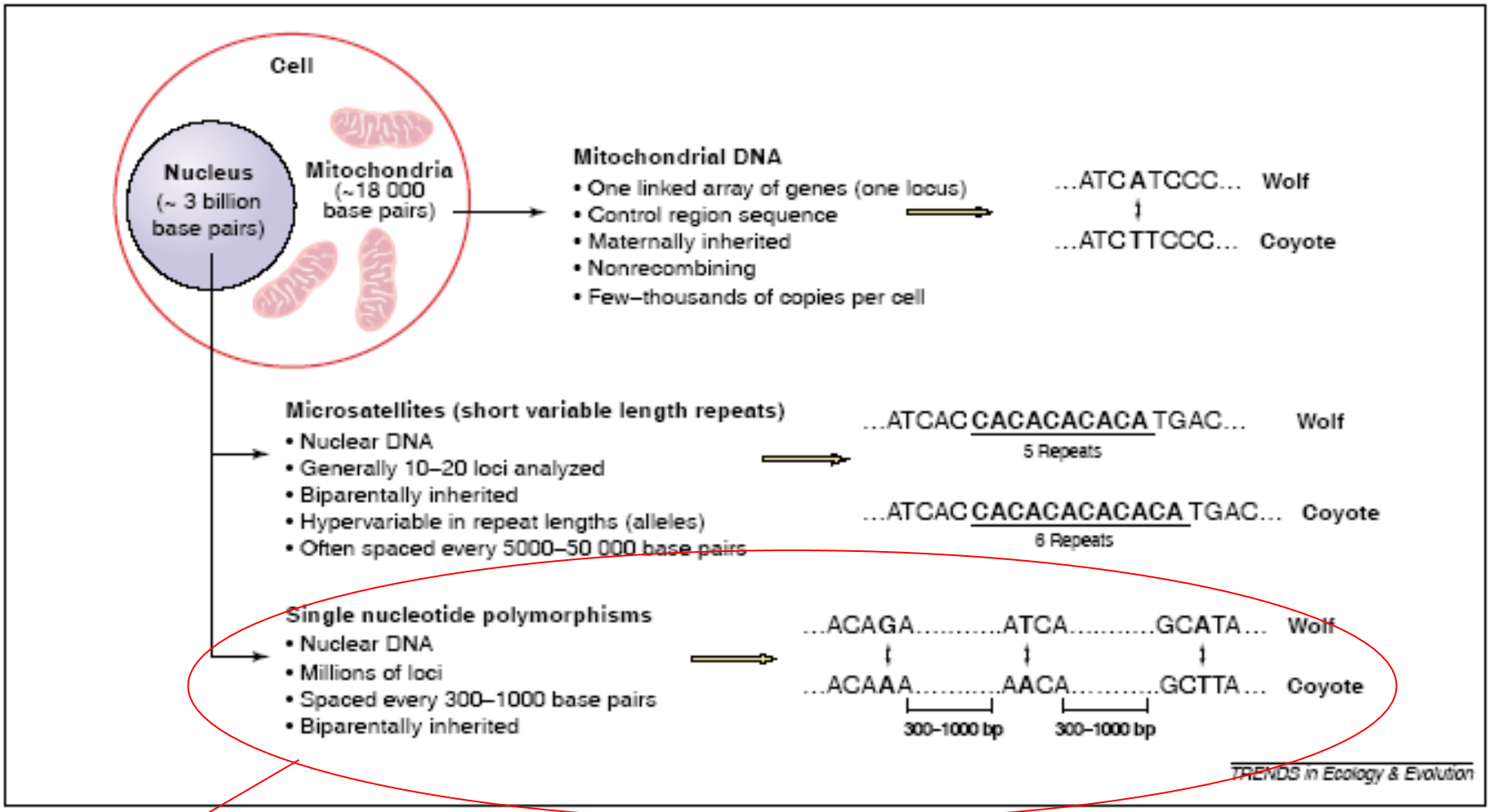
SNP Fashion on the rise





Nature Reviews | Genetics

Single nucleotide polymorphisms (SNPs)

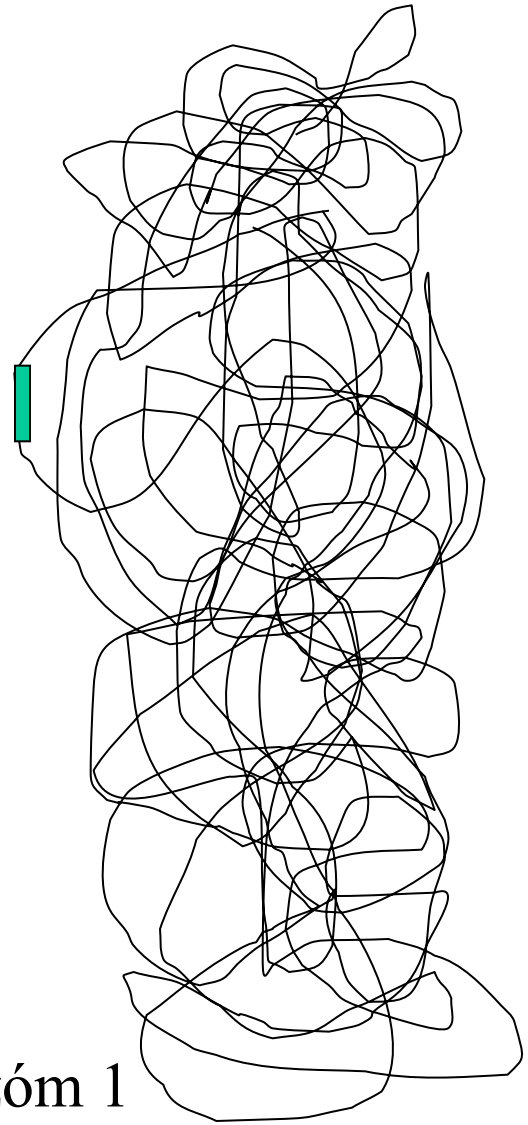


SNPs : nuclear genome (consensus)

Single-locus genetic markers

- SNPs (single nucleotide polymorphisms) – sekvenční polymorfismus

- kodominantní – je možné odlišit heterozygota (např. A/T) od homozygota (např. A/A)



CAAGTA
TGGACG

CATGTA
TGCACG

A/T

CAAGTA
TGGACG

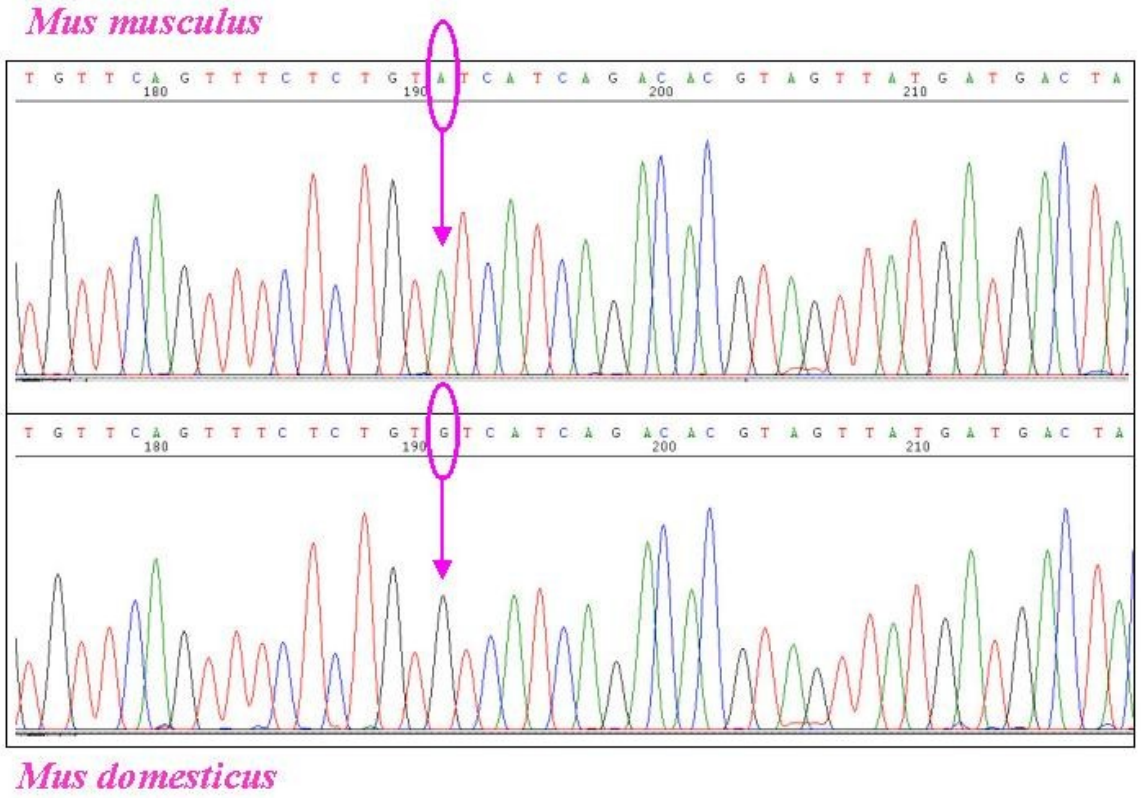
CAAGTA
TGGACG

A/A

Př.: chromozóm 1

Příklad informativního SNP znaku

- fixovaný polymorfismus (homozygoti) – využití např. při studiu hybridizací (hybridi = heterozygoti)



transice
A ↔ G

transition: Pu→Pu or Py→Py transversion: Pu→Py or Py→Pu

Využití SNPs znaků

- obdobné jako u mikrosatelitů
- identifikace druhu (nebo genetické skupiny) - studium hybridizace
- fylogeografie
- populační genetika (genetická variabilita a struktura, tok genů, identifikace jedinců a vztahů mezi nimi, populační velikost a její změny atd.)

Výhody

- početné a rozšířené v genomu (v kódujících i nekódujících oblastech) – milióny lokusů
- 1 SNP cca každých 300-1000 bp
- Mendelovská dědičnost (vs. mtDNA)
- evoluce je dobře popsatelná jednoduchým mutačním modelem (vs. microsatellites)
- jsou analyzovány kratší fragmenty DNA – neinvazivní genetika

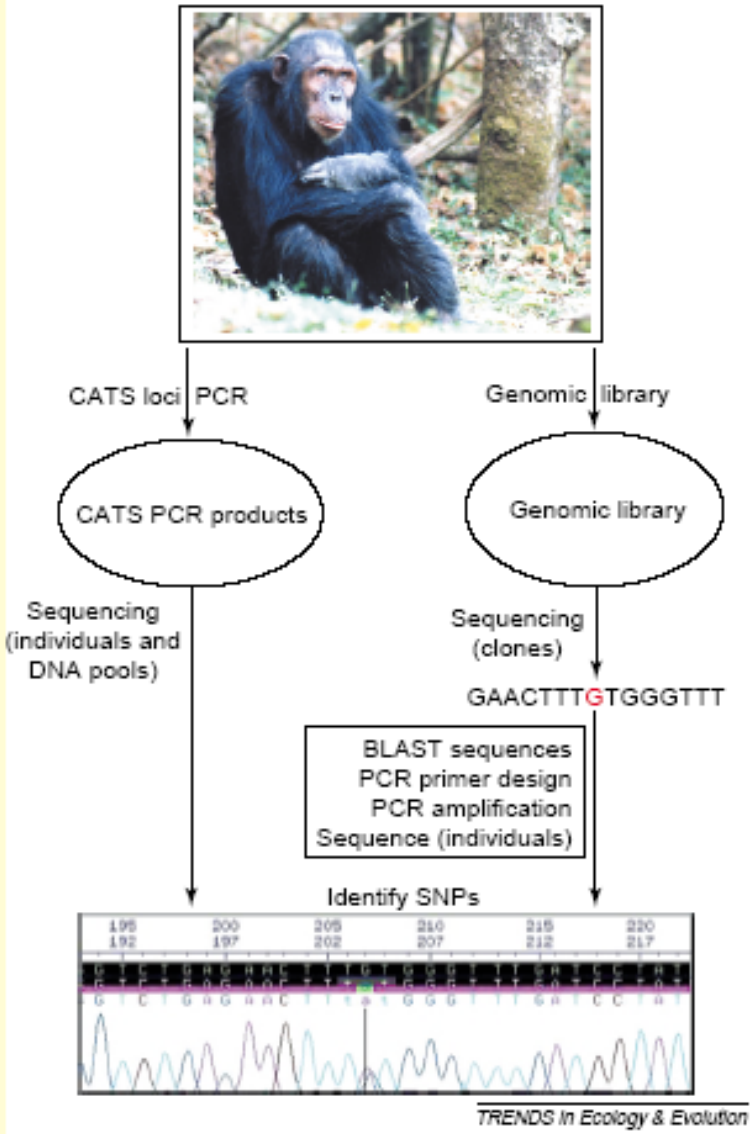
Nevýhody

- „ascertainment bias“ – výběr znaků se provádí na základě jen malého počtu jedinců a nemusí být reprezentativní
- nízká variabilita na lokus (většinou jen 2 alely)
- pro populační genetiku je vyžadován větší počet lokusů (4-10 krát více než u mikrosatelitů)

Metody analýzy

1. Nalezení lokusů („ascertainment“)
2. Genotypizace

Nalezení SNPs

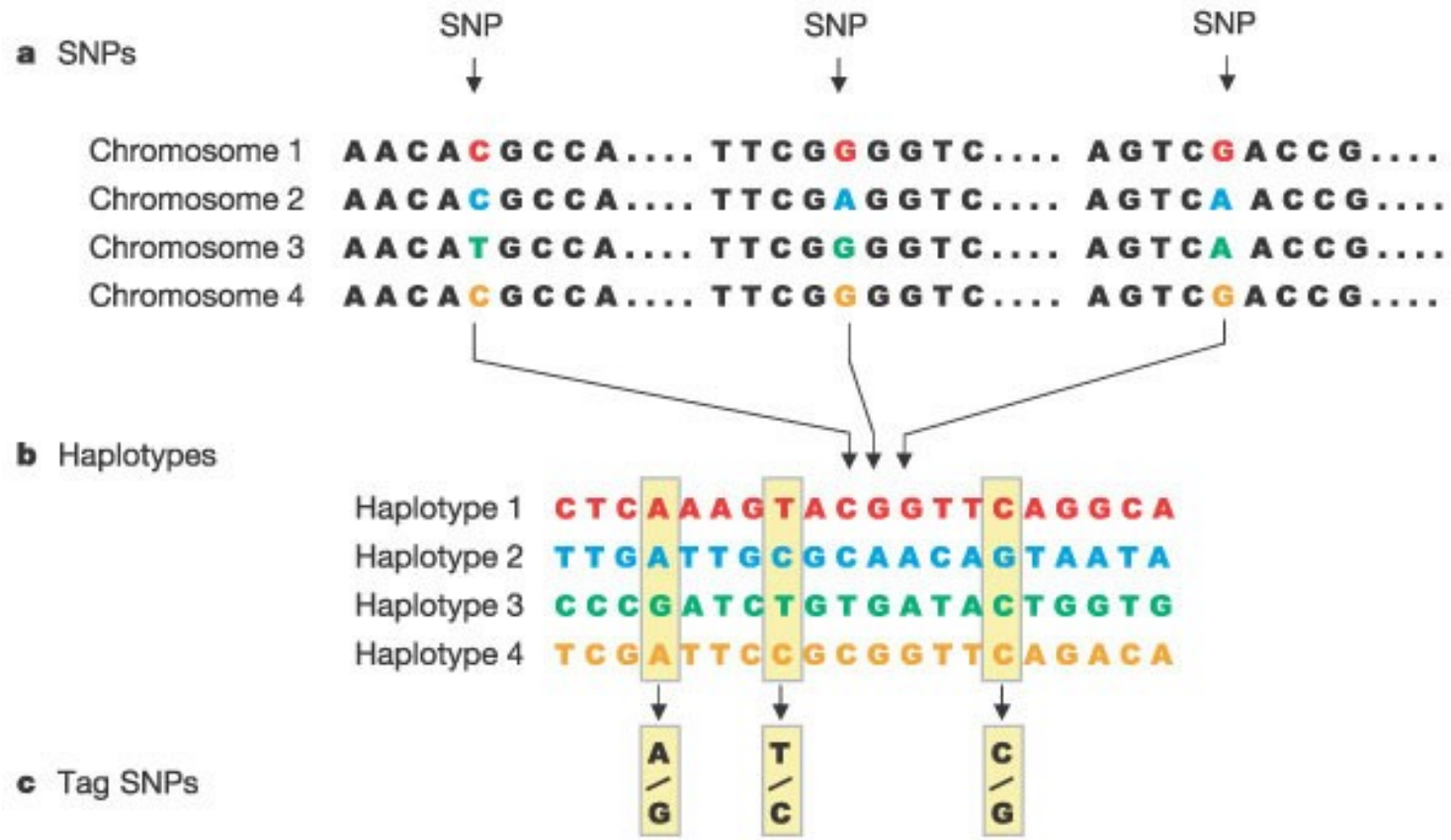


(1) CATS loci = comparative anchor tagged site loci (= cross amplification)

(2) Genomic library = genome restriction + cloning

Next-generation sequencing – sekvenování genomu více jedinců a hledání polymorfismů

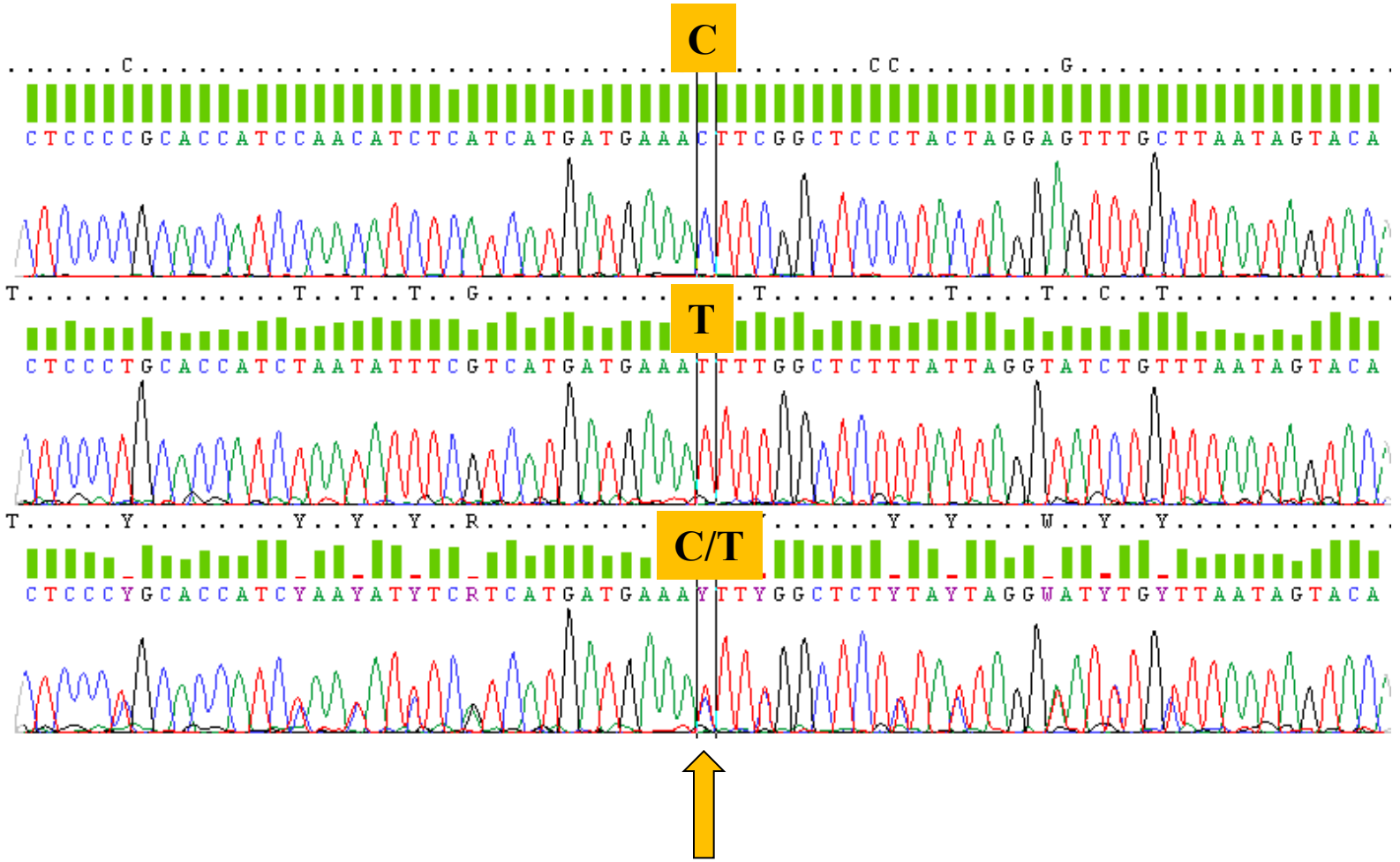
Identifikace různých genotypů u různých jedinců (= homologních chromozómů, tj. variabilita alel)



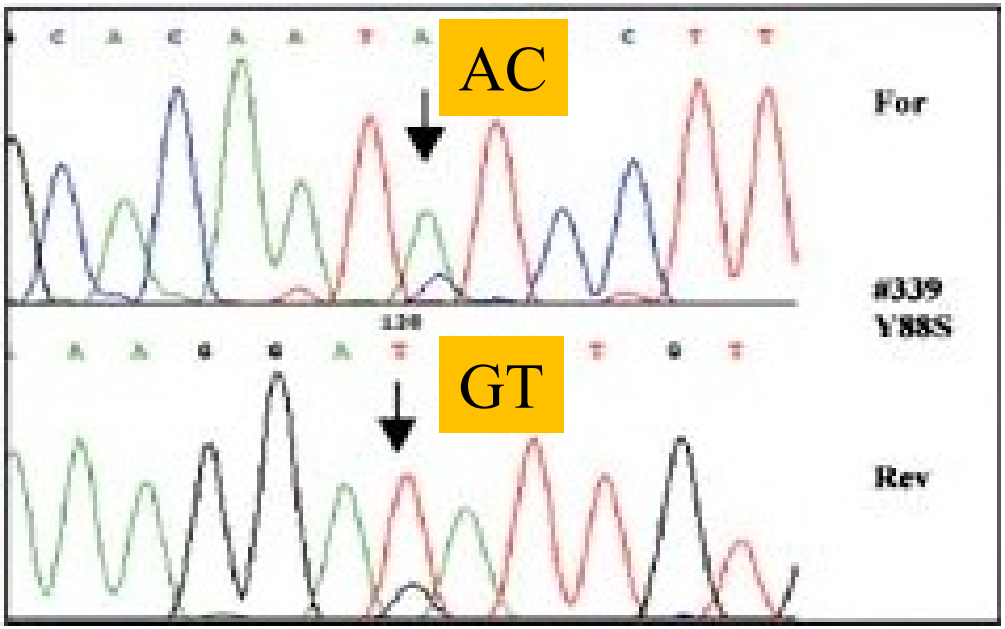
SNPs genotyping

= zjištění genotypu daného jedince

SNPs genotyping – sekvenování?
Je drahé a nejasné u heterozygotů



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

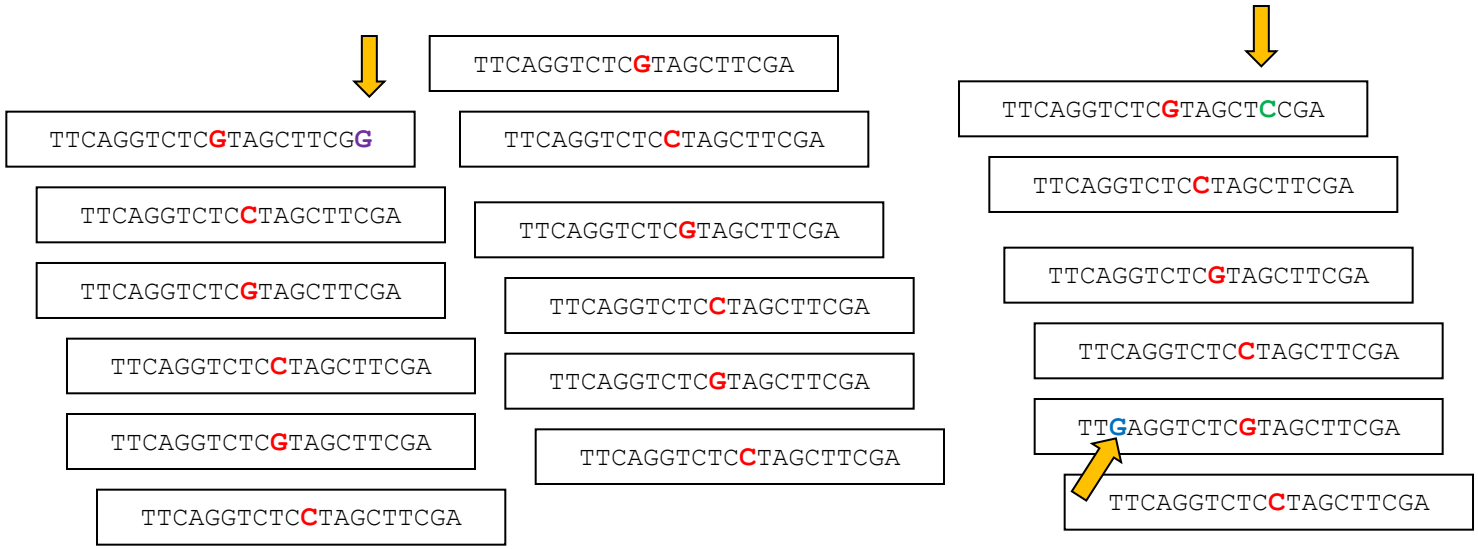
Ex.: heterozygote = two diff. alleles

PCR is making substitution errors that are visualised by cloning

TTCAGGTCTCGTAGCTTCGA

TTCAGGTCTCCTAGCTTCGA

... před PCR = heterozygot G/C



SNPs genotyping

1. **Old standards (PCR-based)**

- RFLP, DGGE, TGGE, SSCP
- HRM: high-resolution melting (real-time PCR)
- původně detekce geneticky podmíněných chorob, např. cystická fibróza

2. **New methods (not based on standard PCR)**

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

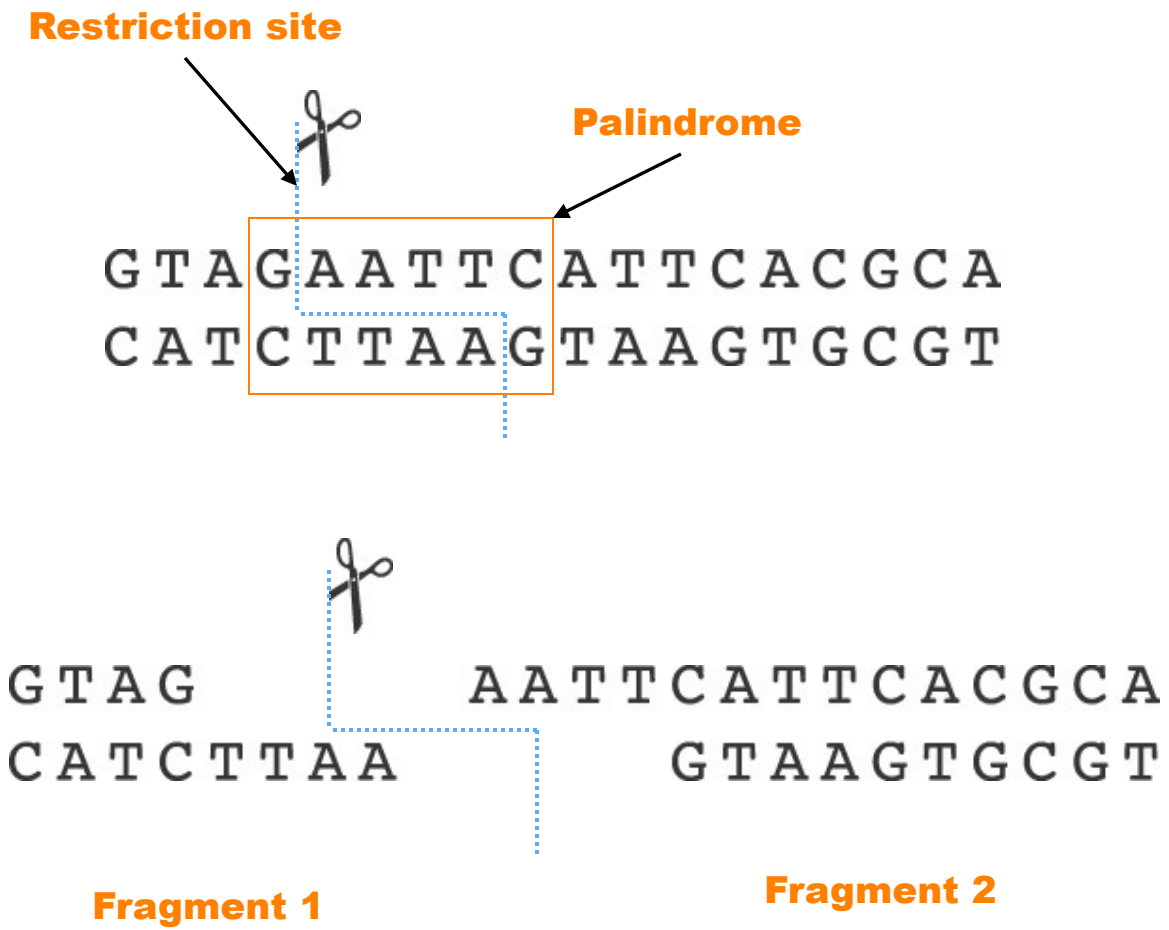
SNP genotyping - old standards

PCR-RFLP

(restriction fragments length polymorphism)

Enzyme Site Recognition

- **Each enzyme digests (cuts) DNA at a specific sequence = restriction site**
- **Enzymes recognize 4- or 6- base pair, palindromic sequences (eg GAATTC)**



Common Restriction Enzymes

EcoRI
- *Eschericha coli*
- 5 prime overhang

PstI
- *Providencia stuartii*
- 3 prime overhang

SNP genotyping - old standards

PCR-RFLP

Allele A

CCGATCA**A**TGCGGCAA
GGCTAGT**T**ACGCCGTT

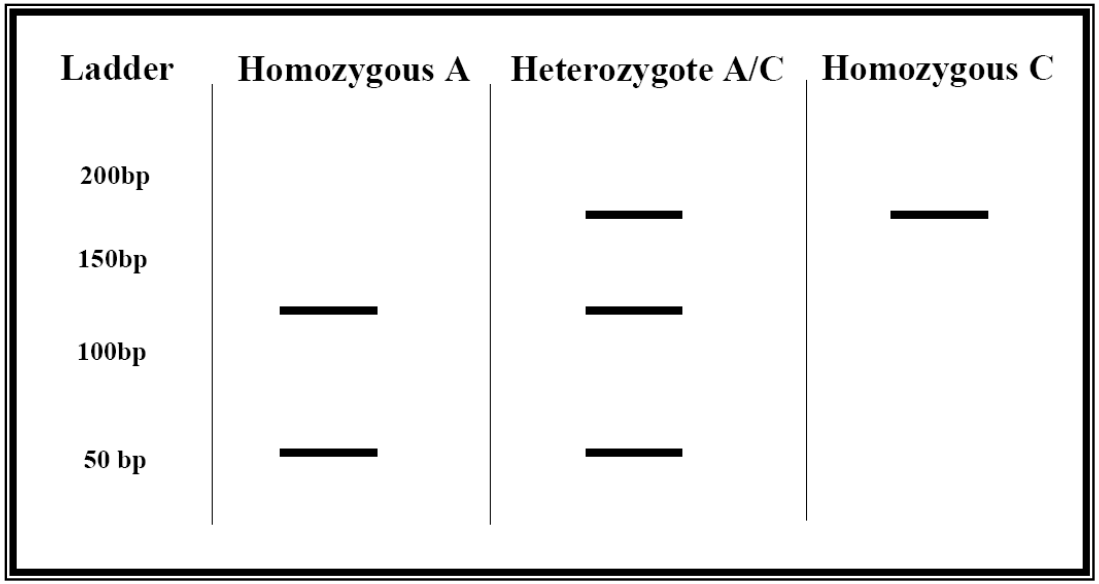
cutting by restriction endonuclease

Allele C

CCGATCA**C**TGCGGCAA
GGCTAGT**G**ACGCCGTT

no cut

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



SNPs genotyping – old standards electrophoresis methods of mutation detection

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

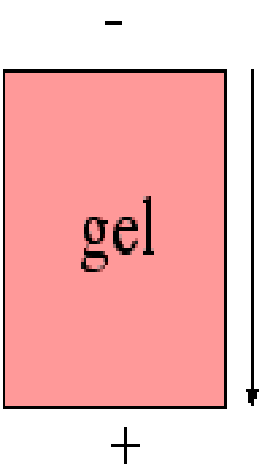
= special electrophoresis methods based on differences in mobility of different DNA sequences

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

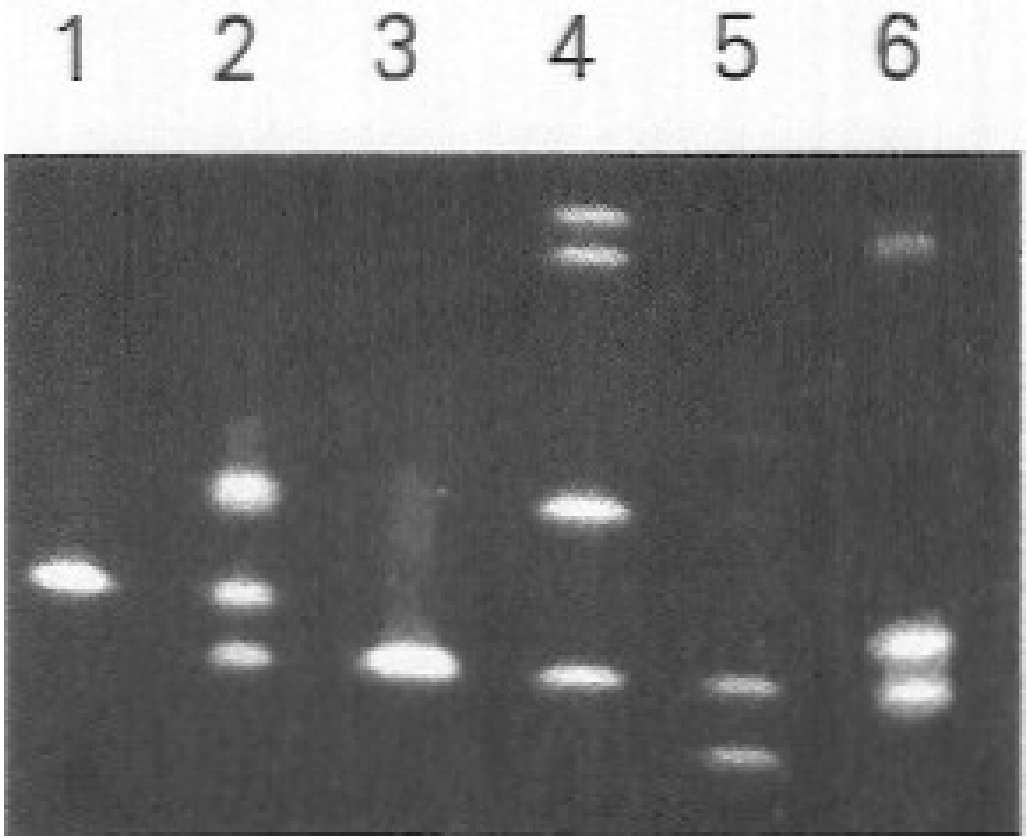


increasing concentration of urea

They retard, and we will see shift in mobility

We will see different shifts in mobility for differing products

Detekce nových mutací – např. v diagnostice genetických chorob nebo při analýzách MHC

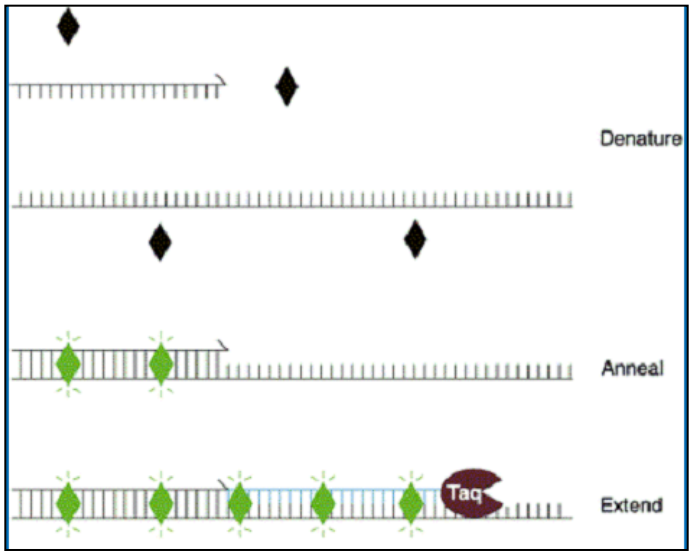


1- normal homozygote
3- homozygous mutations will yield one band on a different position
2, 4, 5, 6 – heterozygous mutations will yield 4 bands (2 homozygous and 2 heterozygous)

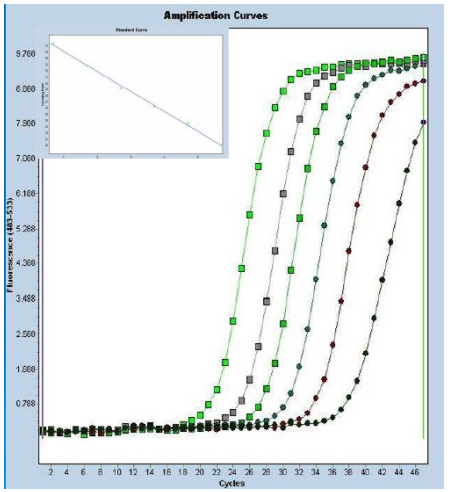
NOT ALL BANDS ARE SEEN !!!!!

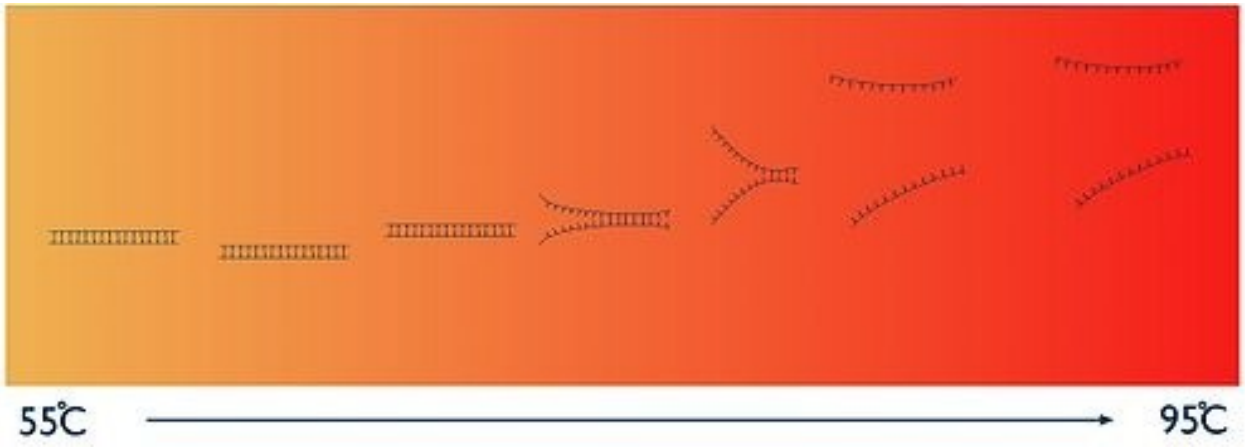
www.leveninc.com/cftr ex.gif

High-resolution melting temperature (HRMT)



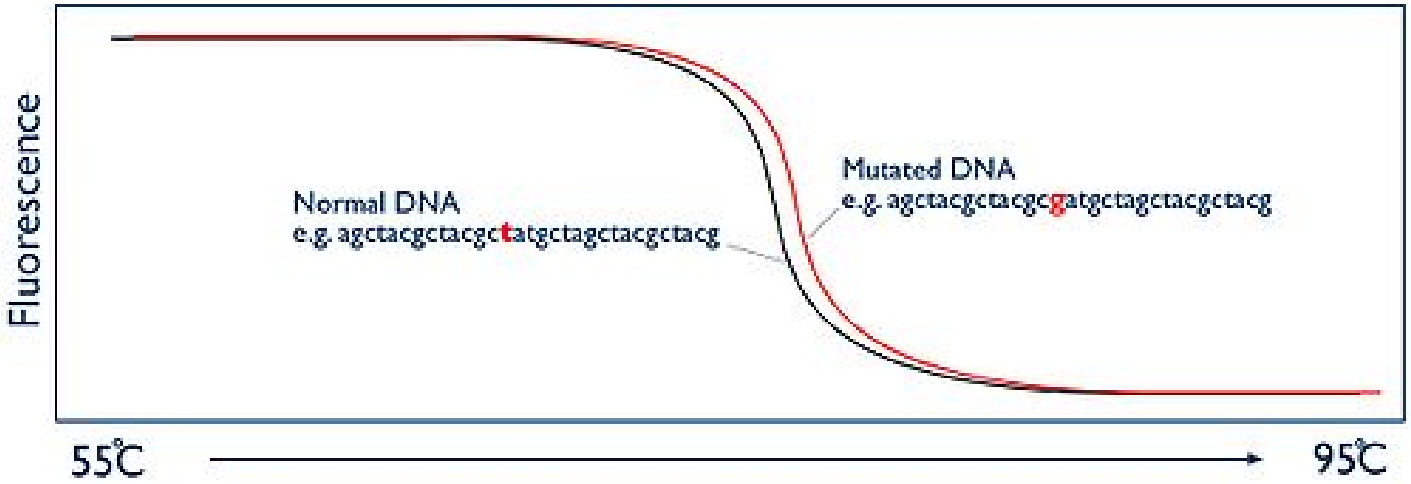
Step 1: real-time PCR = increase of fluorescence



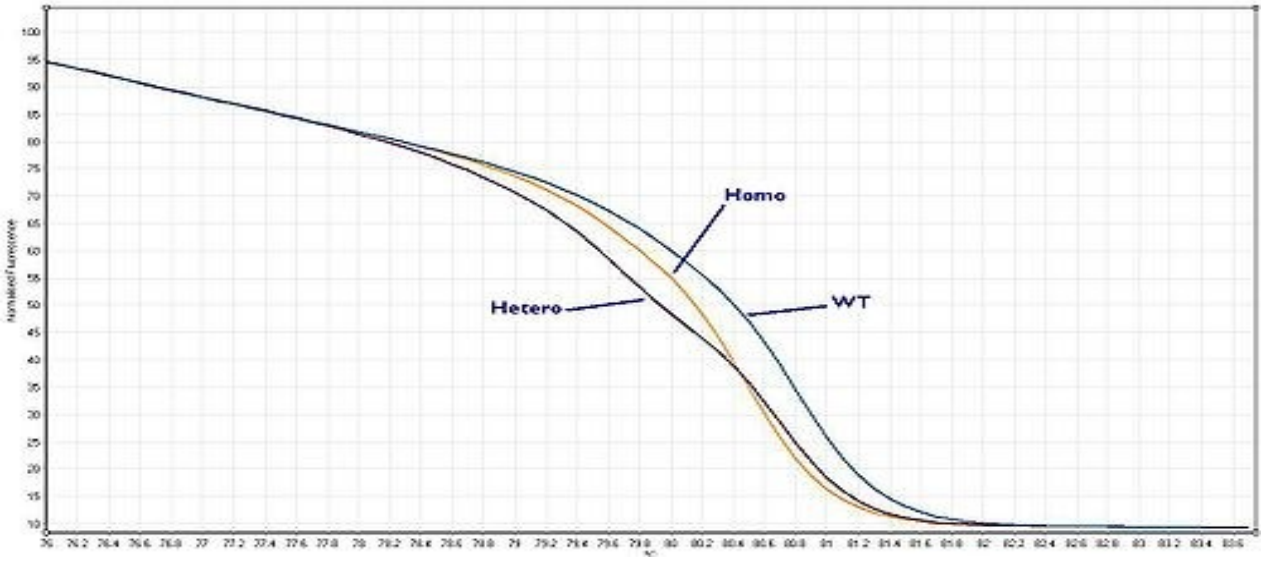


Step 2: measuring melting after PCR = decrease of fluorescence

HRMT genotyping



Detekce heterozygotů

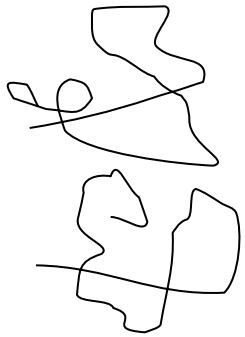


Single strand conformation polymorphism (SSCP)

Allele 1 - C

...CGCTTCAGG ...

...GCGAAGTCC...



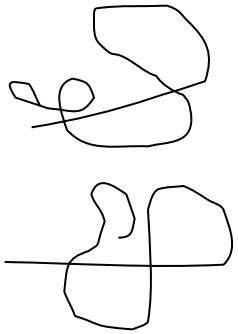
heating - denaturation

snap-cooling → partial renaturation

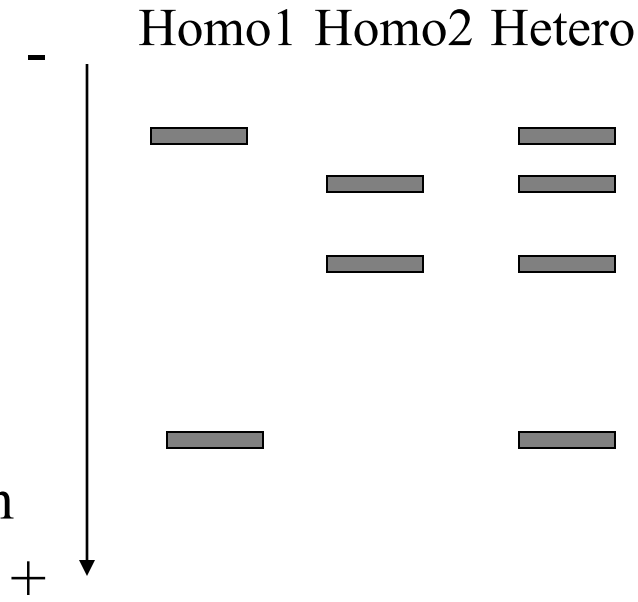
Allele 2 - A

...CGCTTAAGG ...

...GCGAATTCC...



sequence-specific
ssDNA conformations

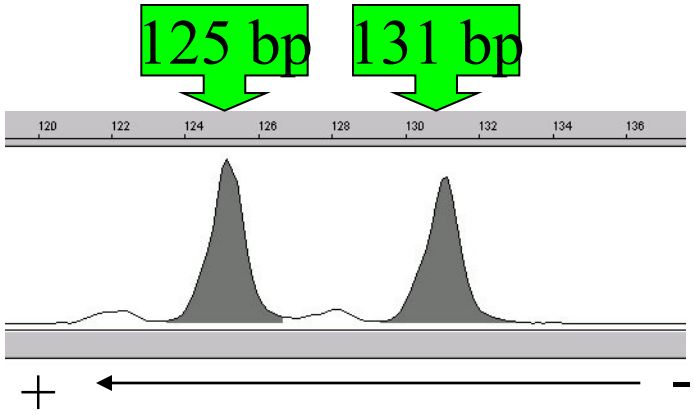


!!! non-denaturing PAGE

radioisotopes
silver-staining
fluorescent dyes (SYBR gold)

Použití automatických sekvenátorů

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



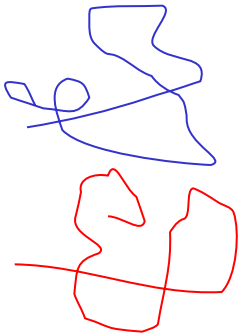
Well controlled electrophoresis parameters, high sensitivity

Použití automatických sekvenátorů

Why not non-denaturing electrophoresis?
e.g. CAP (conformation analysis polymer)

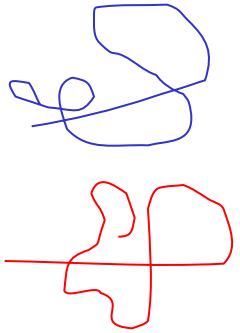
Allele 1

FAM... CGCTTCAGG ...
... GCGAAGTCC *...HEX*



Allele 2

FAM... CGCTTAAGG ...
... GCGAATTCC *...HEX*

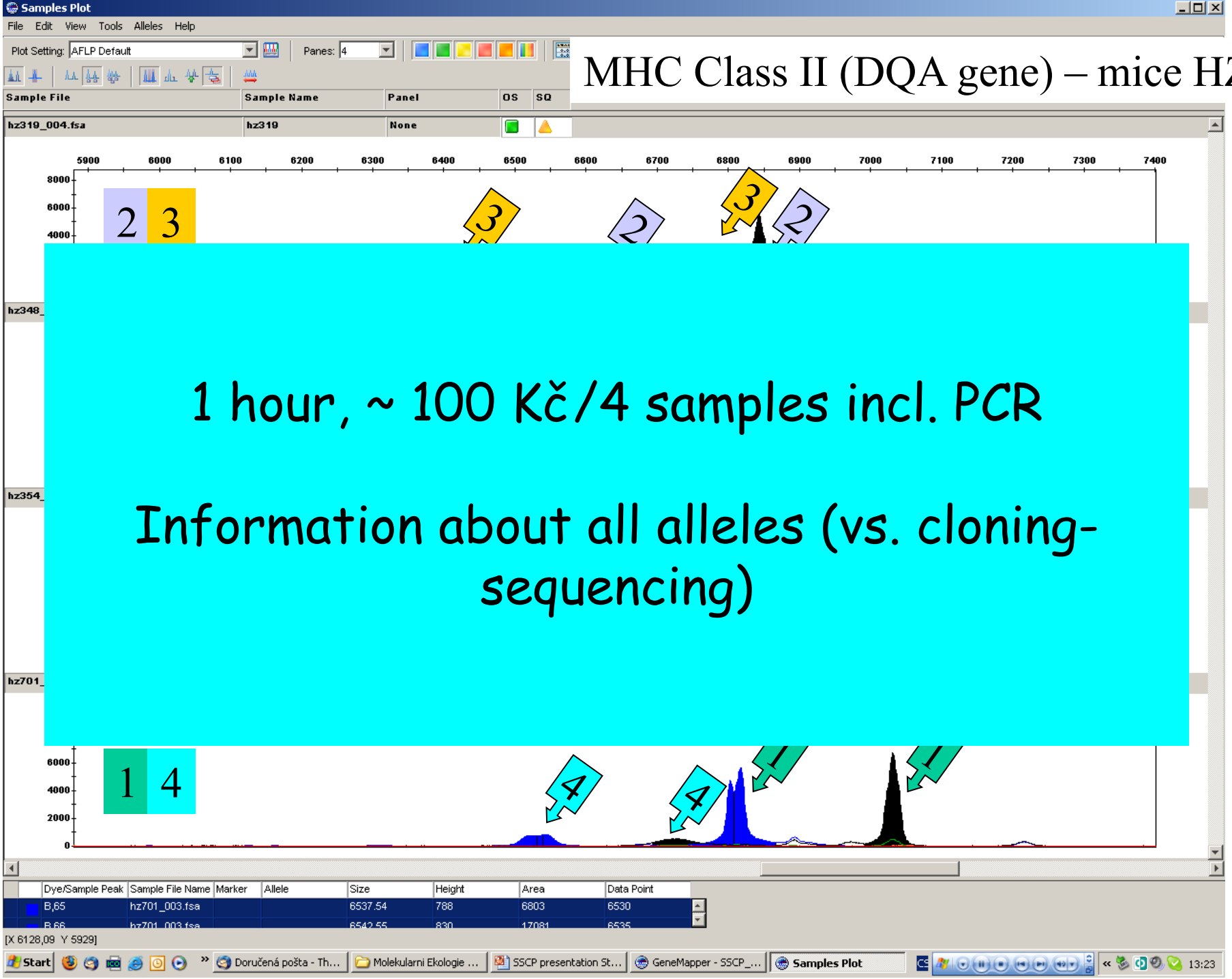


- well controlled electrophoresis

- two fluorescent labels

- high sensitivity

MHC Class II (DQA gene) – mice HZ



1 hour, ~ 100 Kč/4 samples incl. PCR

Information about all alleles (vs. cloning-sequencing)

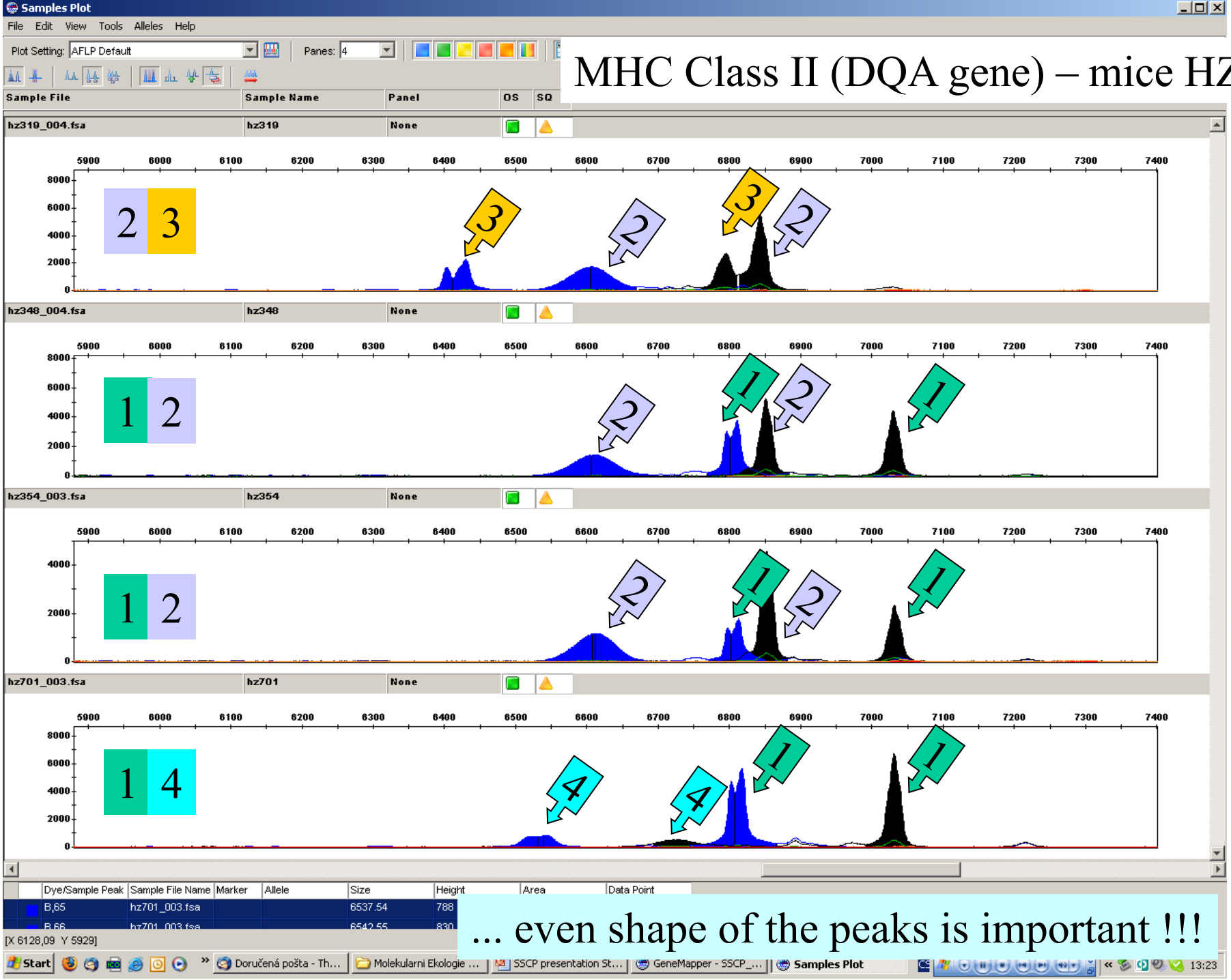
Data analysis

- GeneMapper (Applied Biosystems)
- different „Size Standard" for each temperature
- alignement of more samples
- allows detection of „haplotypes", i.e. short sequences with several SNPs (very useful for e.g. MHC genotyping)

Applications

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

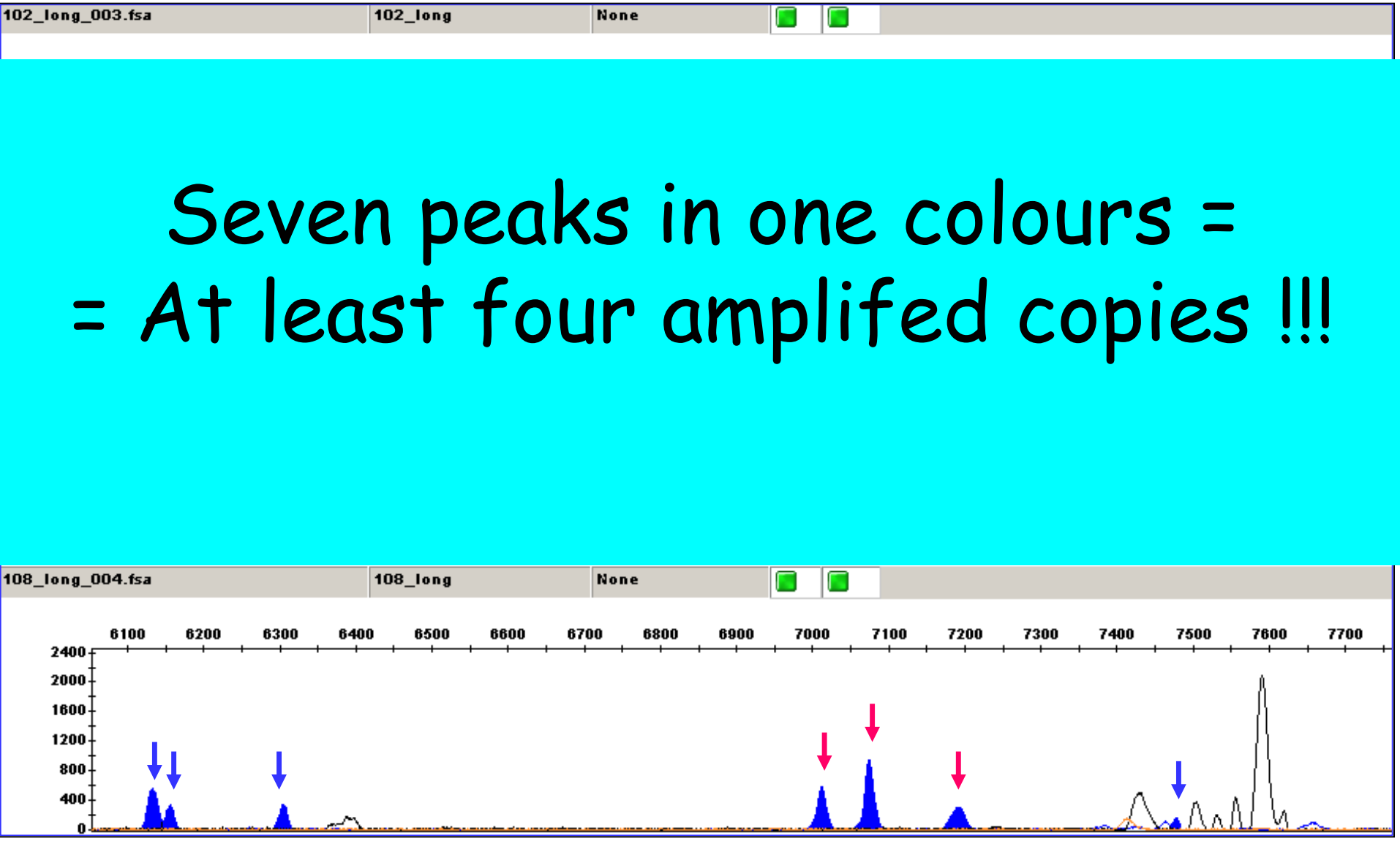
MHC Class II (DQA gene) – mice HZ



... 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)



SSCP of three individuals: - different alleles - same alleles

Carpodacus erythrinus – MHC Class I (Promerová et al. 2009)

Applications

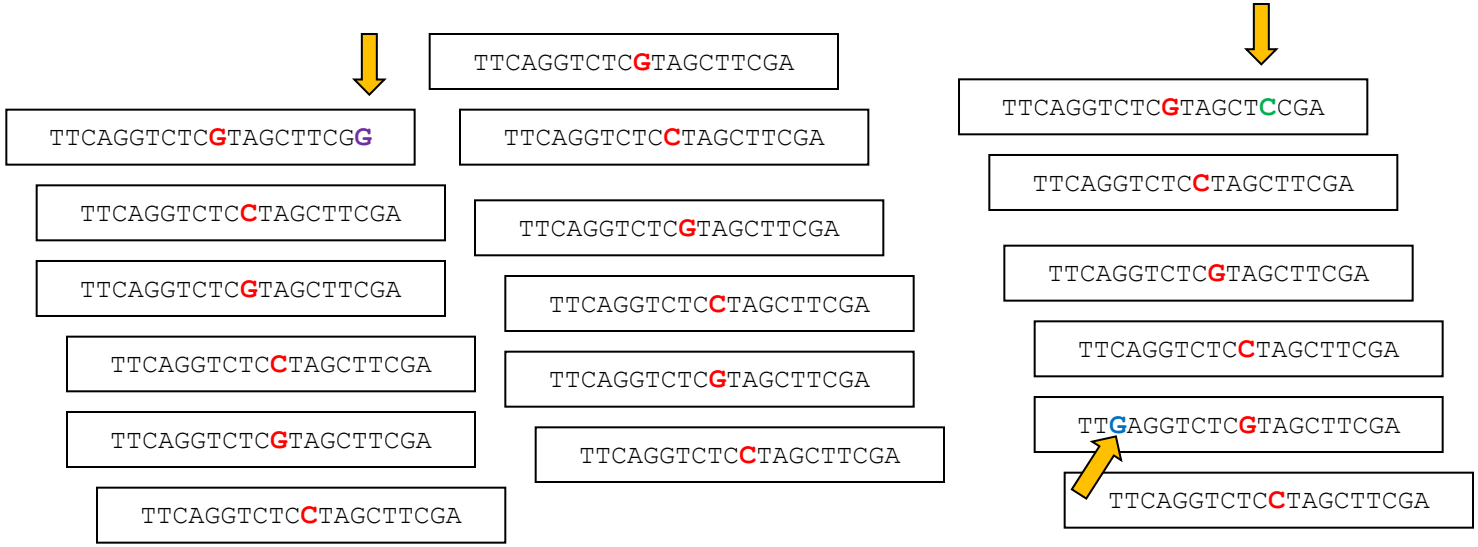
1) Genotyping of codominant markers (e.g. single copy MHC genes)
2) Identification of number of genes (e.g. duplicated MHC genes)
3) Detection of PCR artefacts during cloning

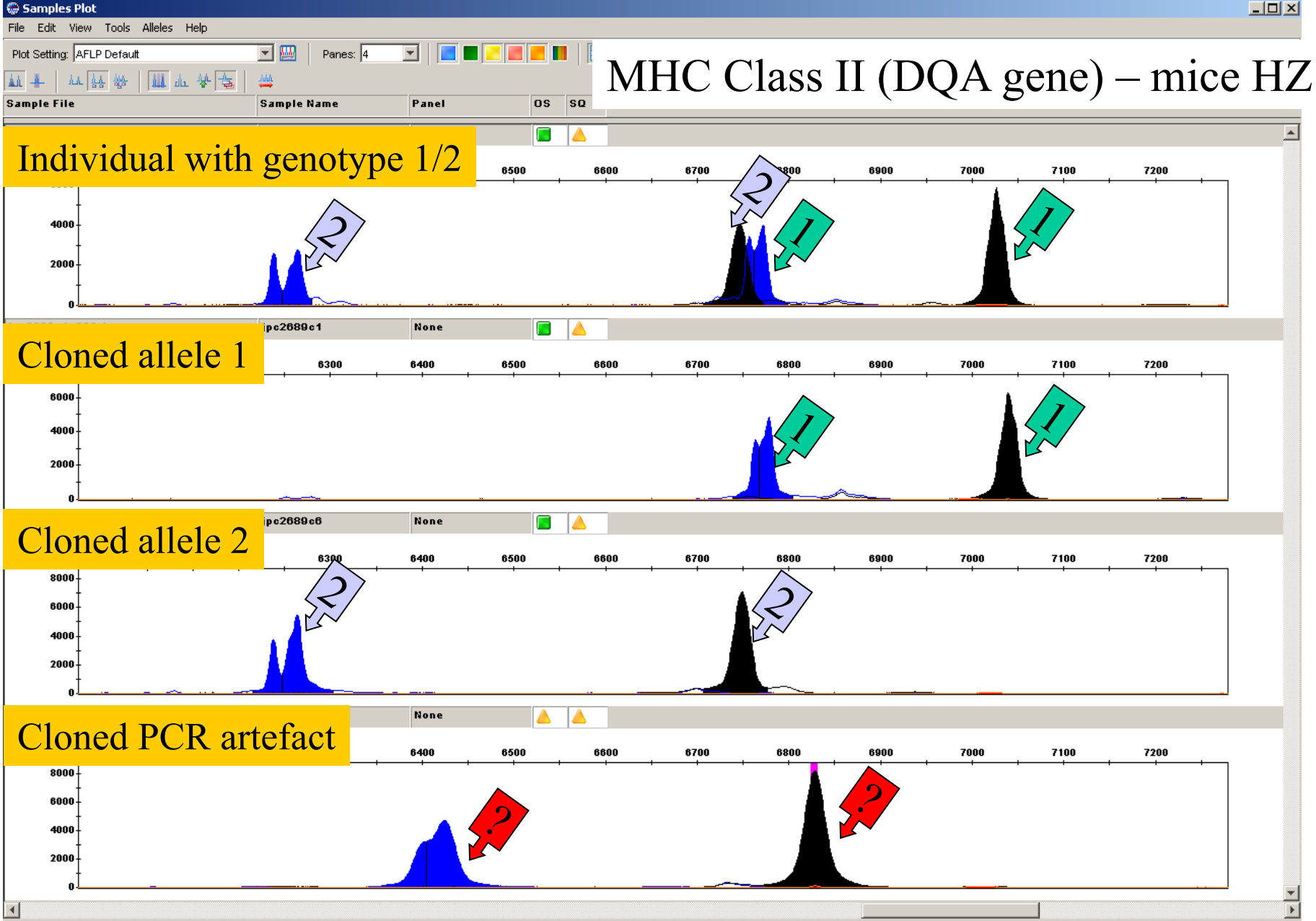
Detection of PCR artefacts during cloning

TTCAGGTCTCGTAGCTTCGA

TTCAGGTCTCCTAGCTTCGA

... před PCR = heterozygot G/C





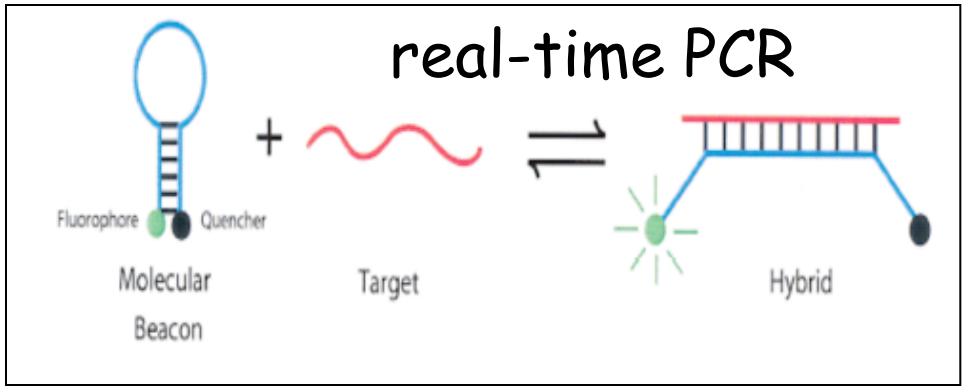
Detection of PCR artefacts during cloning of heterozygotes

SNP genotyping – new methods

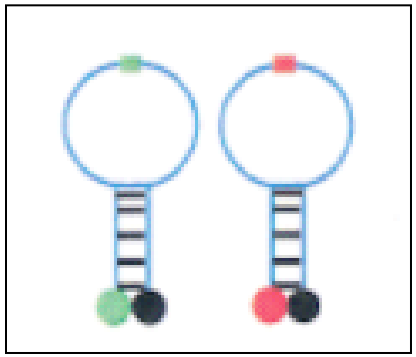
= not based on standard PCR

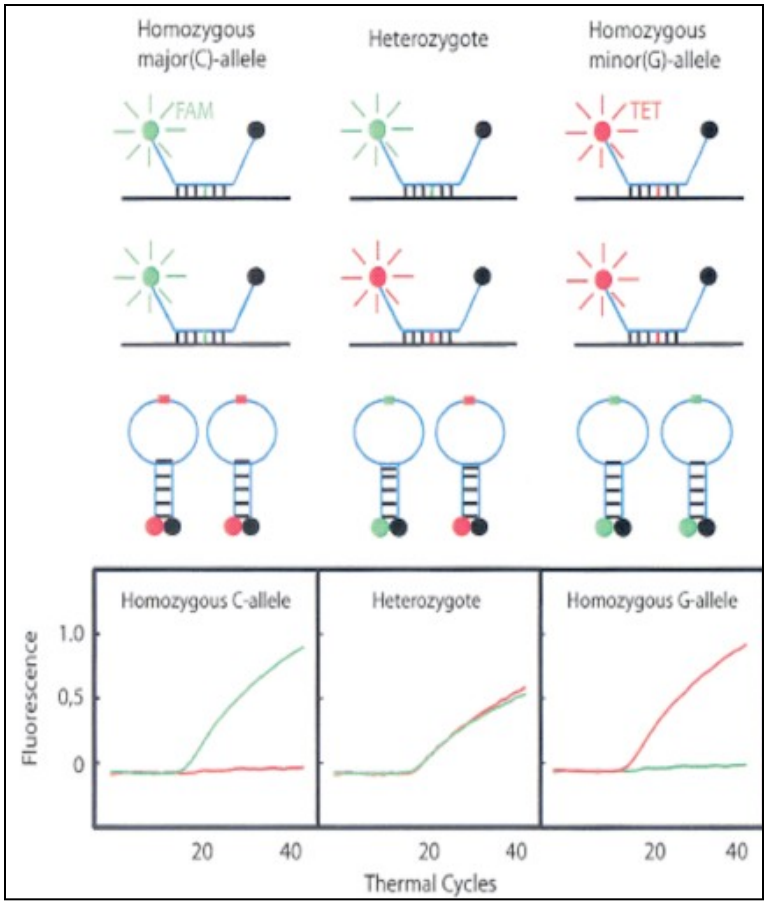
1. real-time PCR se specifickými sondami (TaqMan, molecular beacon)
2. ASPE: allele-specific primer extension
3. SBE: single base extension
4. SNP microarrays (GeneChip method)

Real-time PCR se specifickou sondou



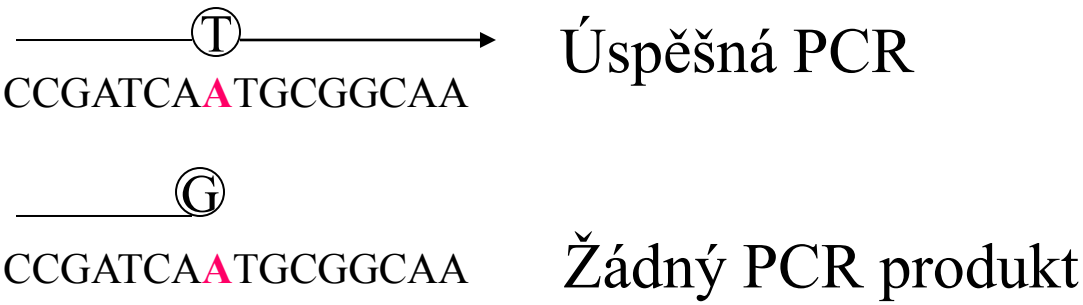
sondy specifické pro jednotlivé alely





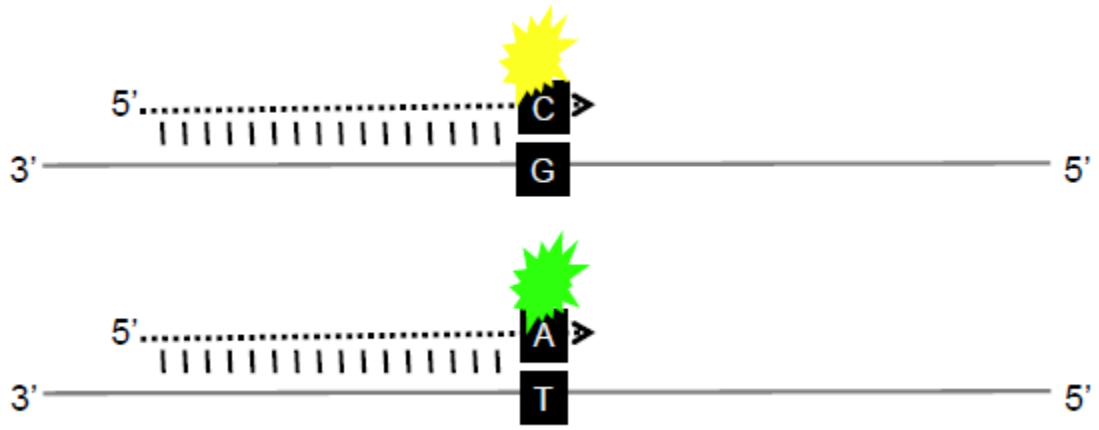
1) TaqMan sondy
2) Molecular Beacons („maják“)

ASPE: allele-specific primer extension



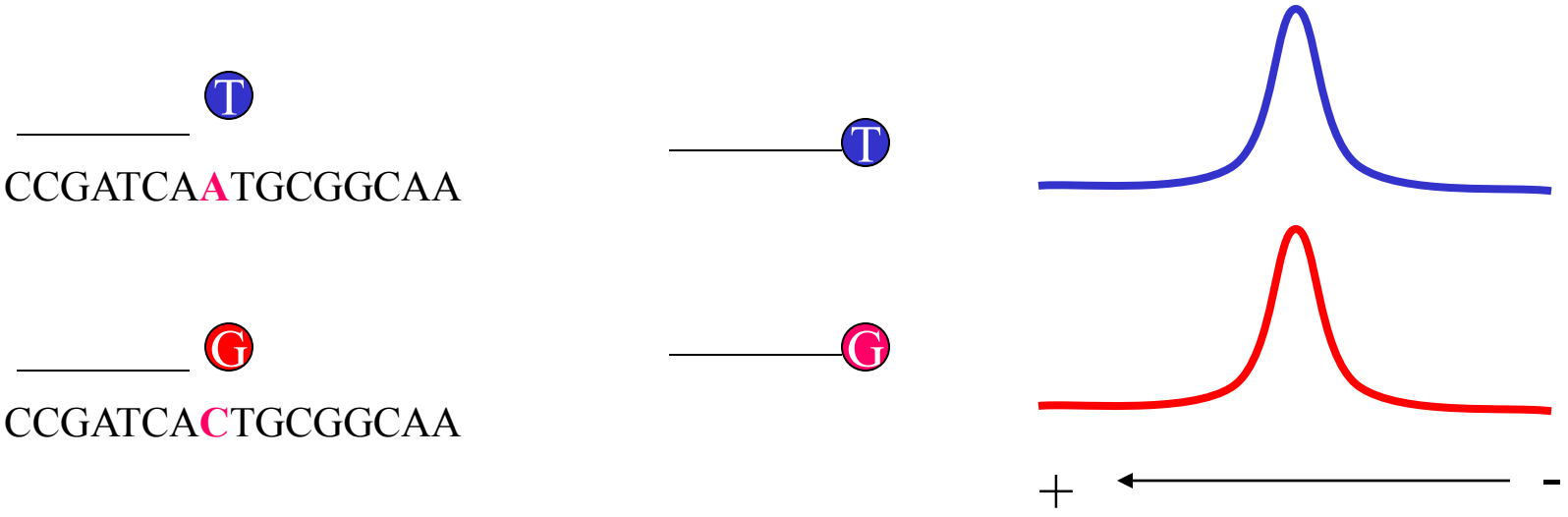
- dvě PCR se specifickými primery
- 3’ terminální nukleotid na primerech je komplementární k SNP nukleotidu
- alelově-specifická amplifikace je umožněna vysoce specifickou polymerázou

ASPE: allele-specific primer extension (automatizovaná verze)



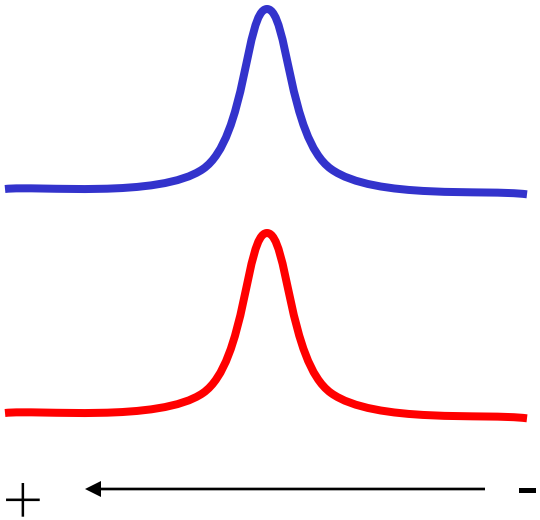
- existují zoptimalizované multiplexy pro modelové druhy (např. člověk 1536 SNPs)
- fluorescenční detekce (Illumina)

(3) SBE: single base extension

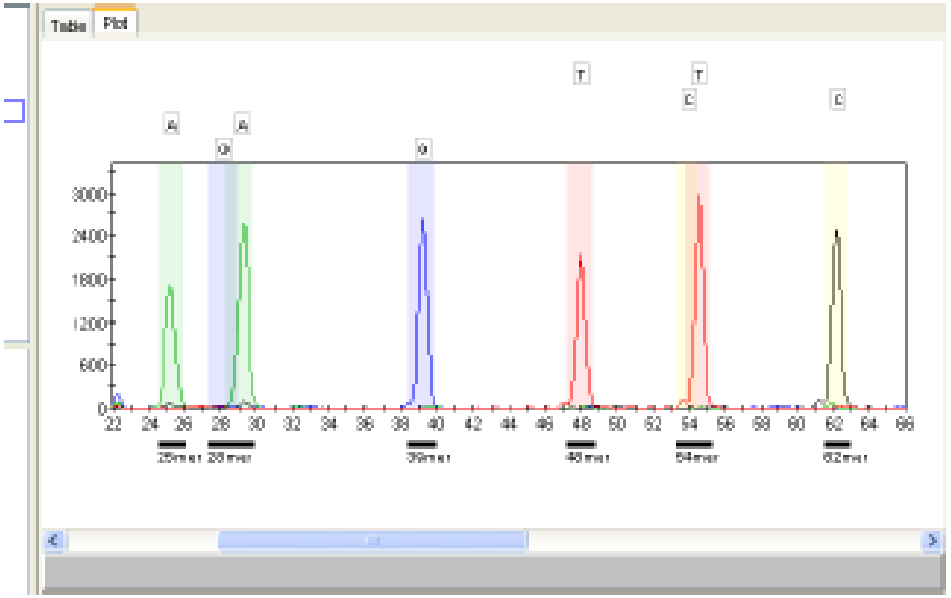


- pouze jeden dideoxynukleotid je přidán k primeru
- detekce různými metodami

Detection or SBE products



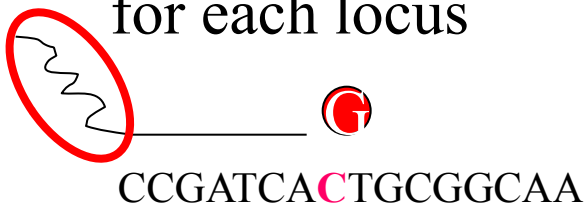
electrophoresis in a capillary
SNaPShot Multiplex Kit
(Life Technologies)



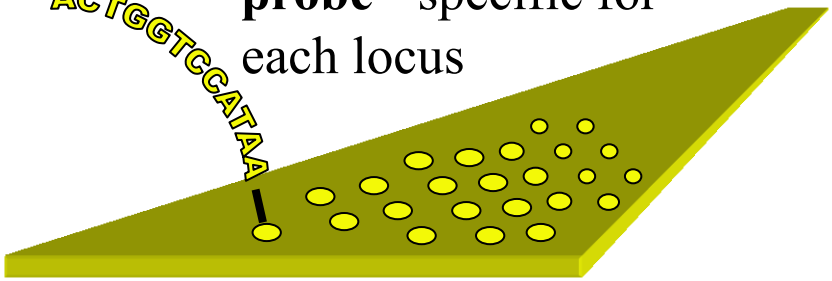
„multiplex version" – různě dlouhé primery, aby bylo možné odlišit různé lokusy

Microarray detection of multiple SBE products

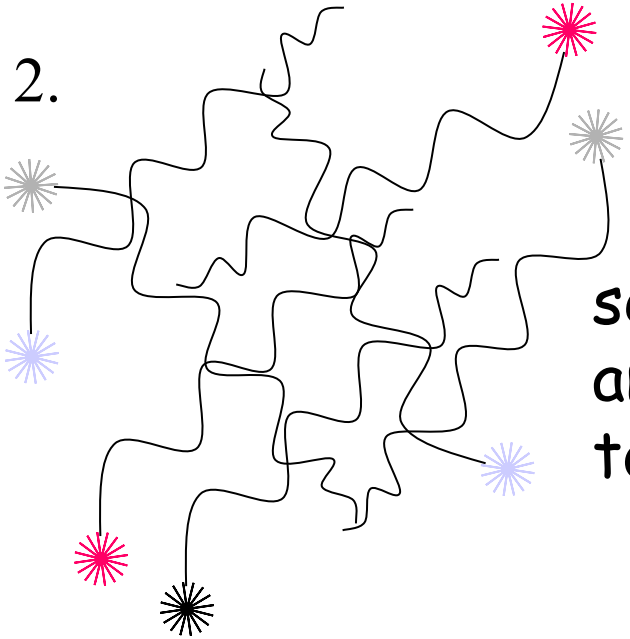
1. **tag** – specific for each locus



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

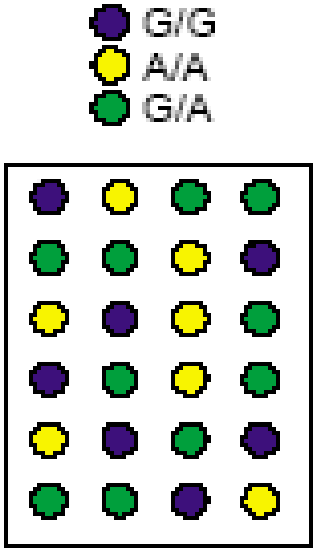


2.



several loci amplified together

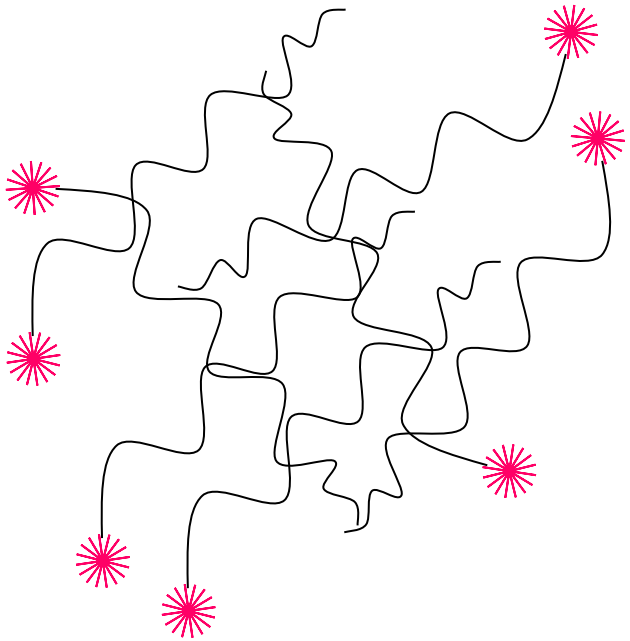
4.



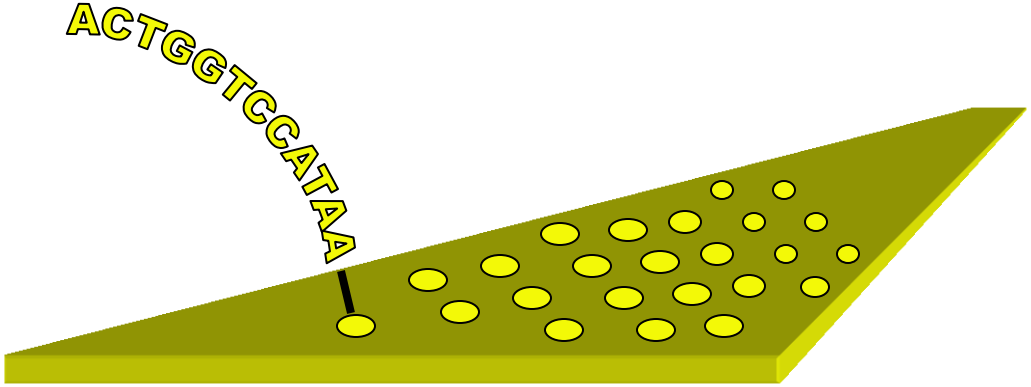
Small-scale "in house" SNP genotyping

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

(4) Microarray analysis of SNPs
(whole genome approach – „chip technology")

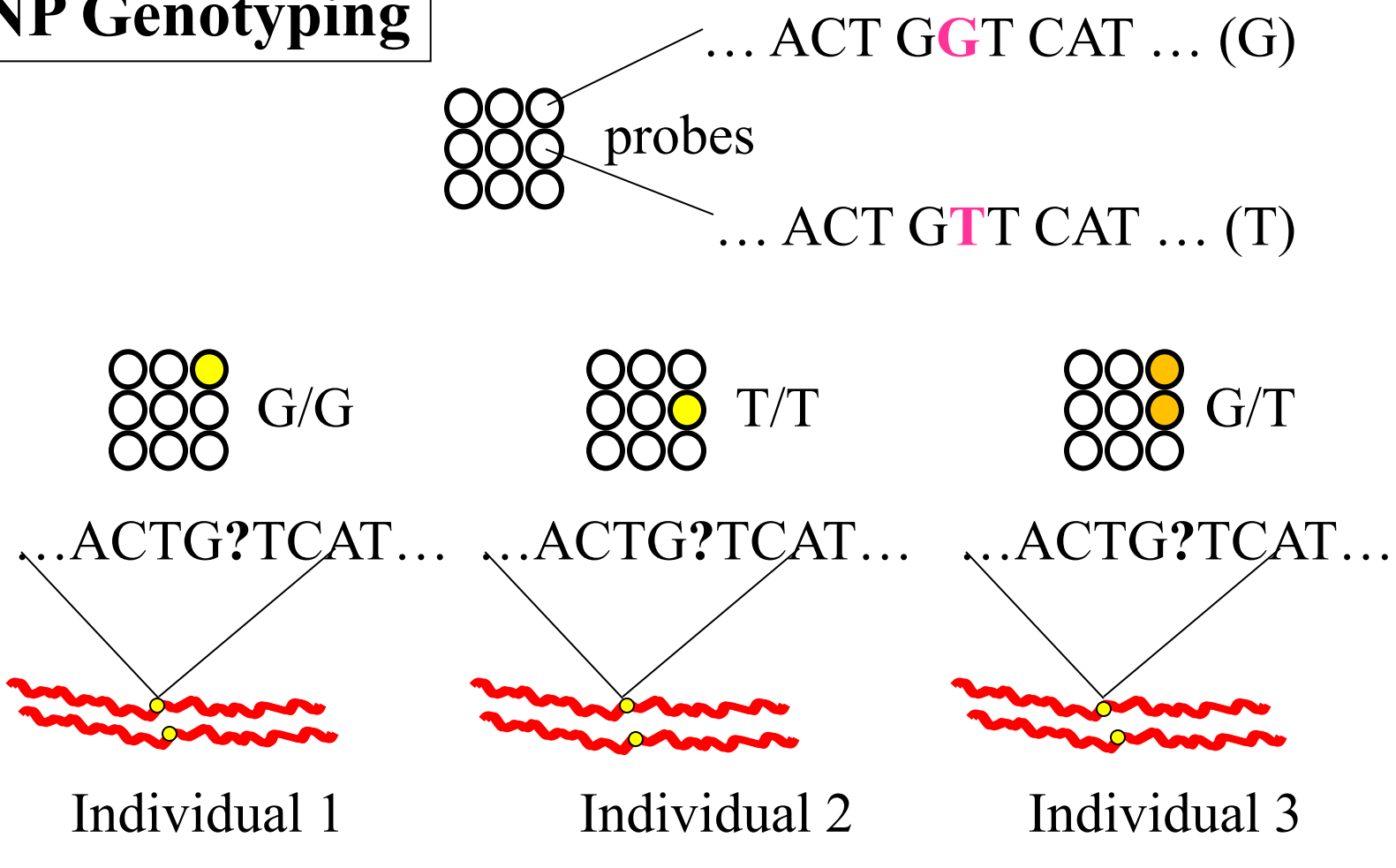


Target (genomic DNA fragmented by e.g. restriction enzymes)

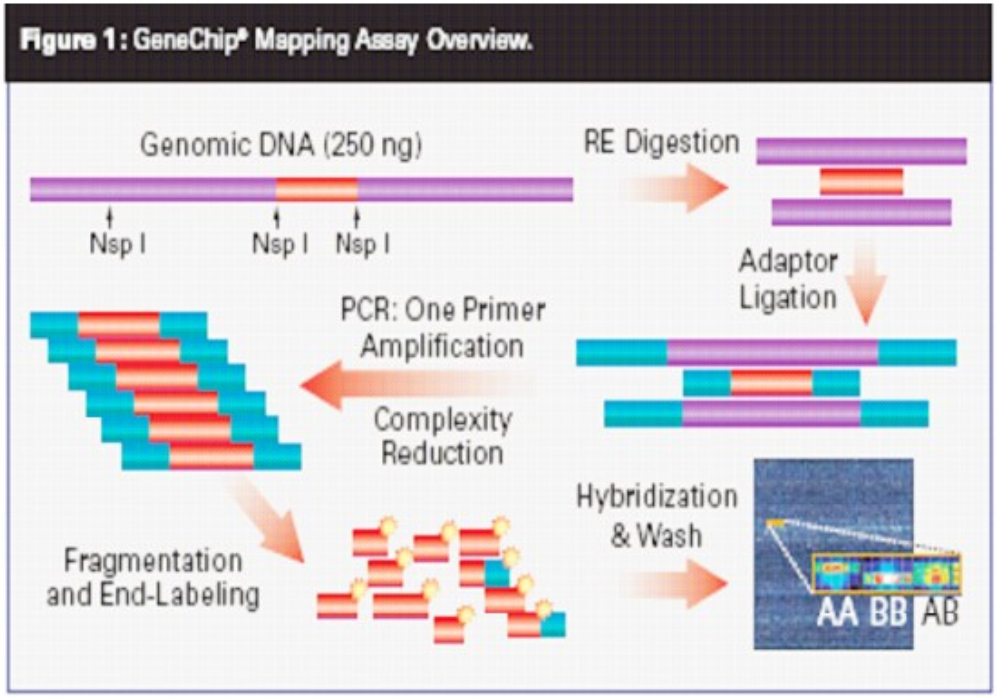


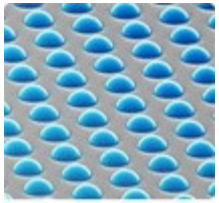
Probe
(specific probes for each allele)

Microarray SNP Genotyping

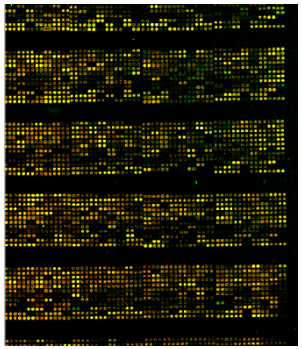


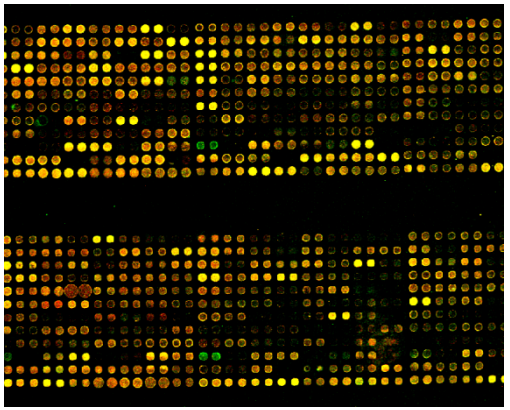
Detekce: Affymetrix, Illumina





BeadArray
(Illumina)





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
Illumi									ples
Illumi									ples
Illumi									mles
Illumi									ples
Illumi									ples
Illumi									ples
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

Použití u příbuzných druhů je možné, ale je tam velmi silný „ascertainment bias“