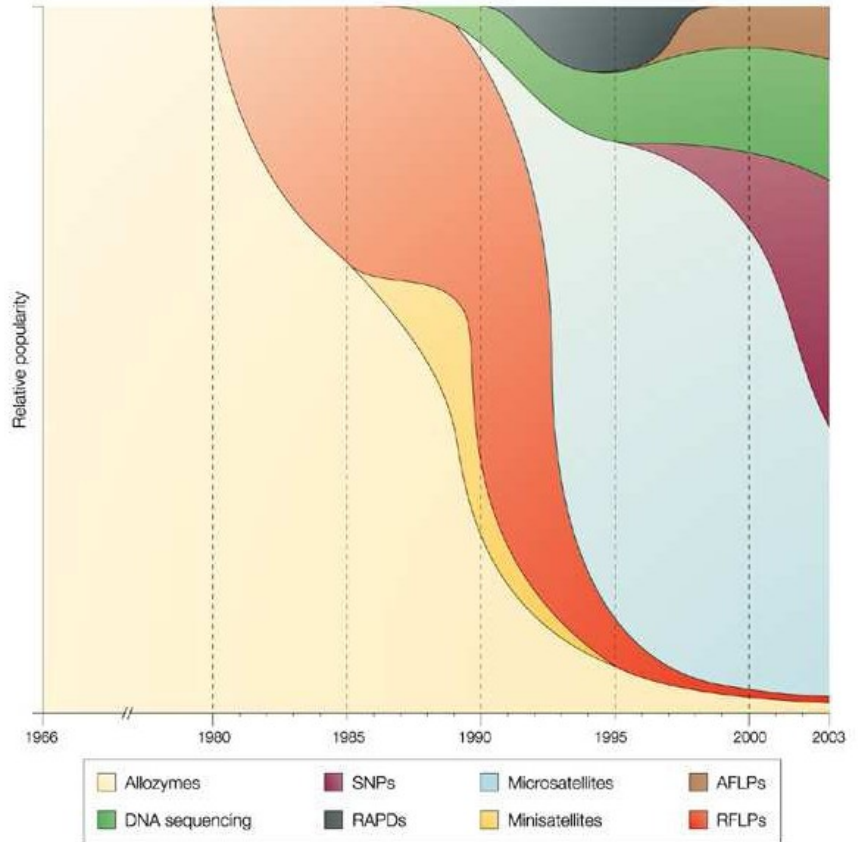
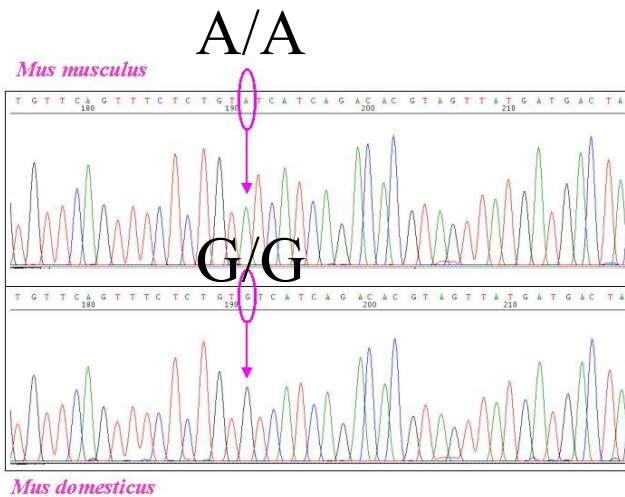
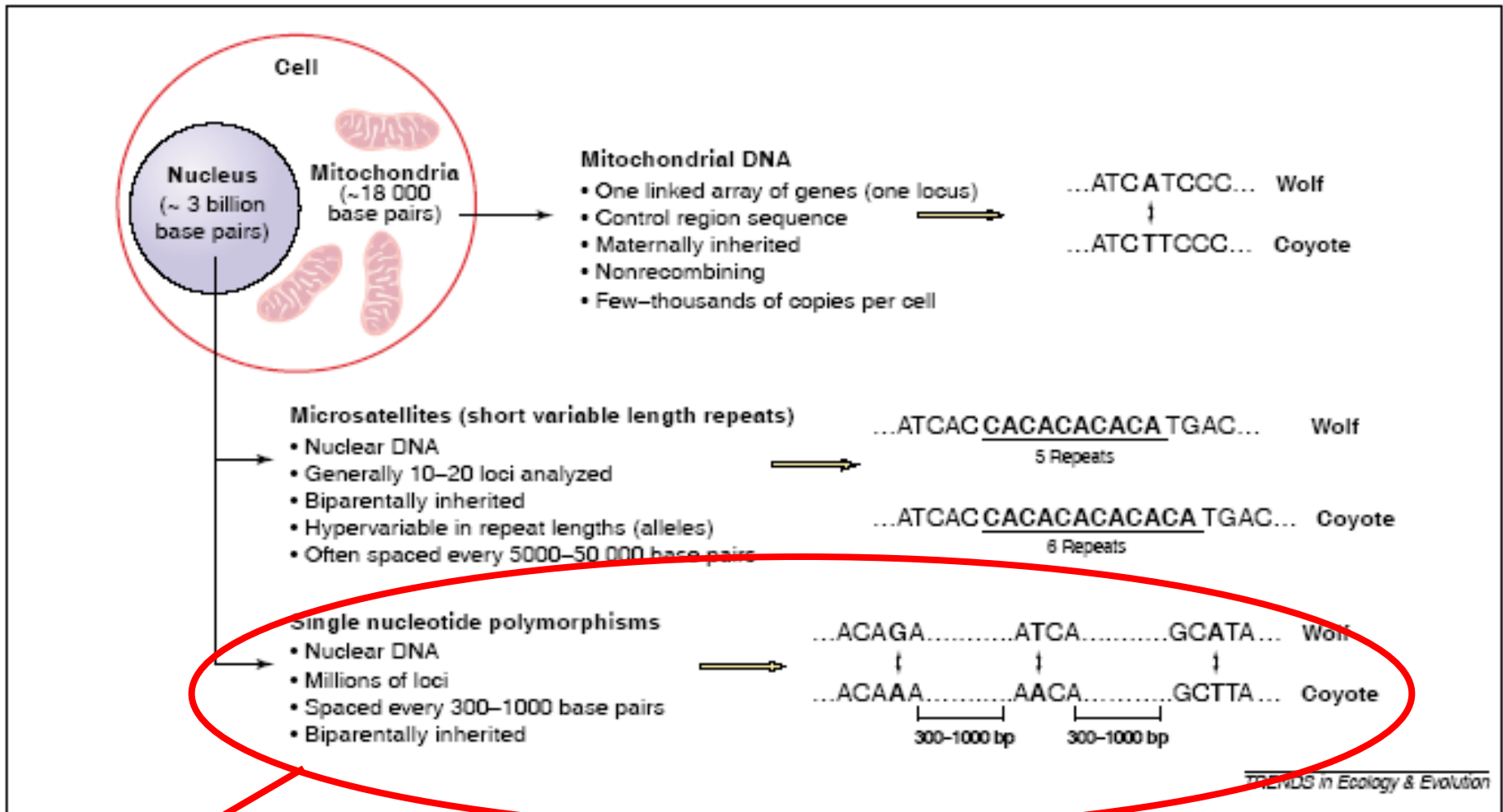


# Single nucleotide polymorphisms (SNPs)



# Single nucleotide polymorphisms (SNPs)



SNPs : nuclear genome (consensus)

# SNPs = single-locus genetic markers

- SNPs (single nucleotide polymorphisms) – sekvenční polymorfismus
- kodominantní – je možné odlišit heterozygota (např. A/T) od homozygota (např. A/A)

CA**A**GTA

TG**G**ACG

CA**T**GTA

TG**C**ACG

**A/T**

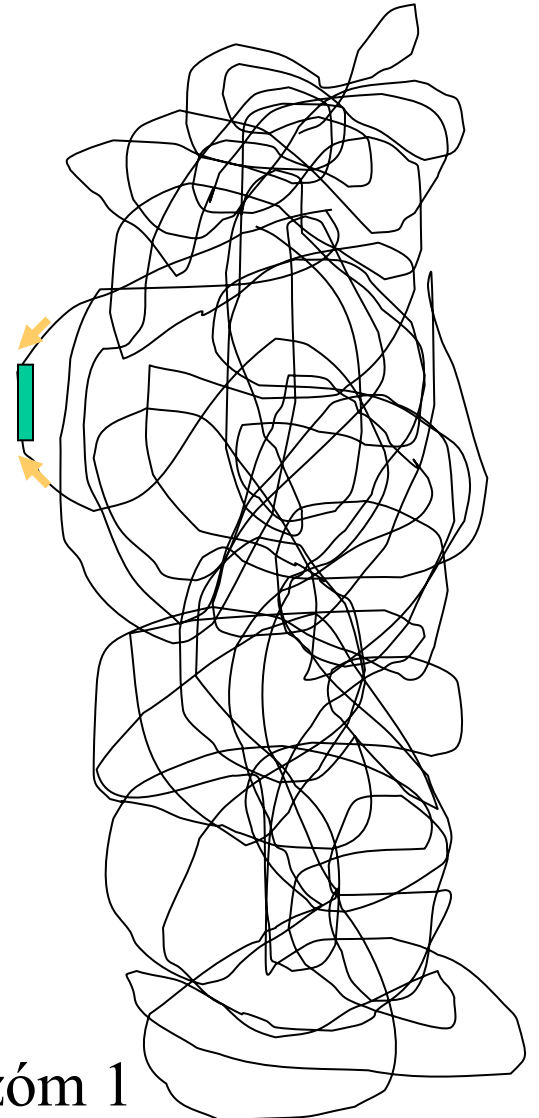
CA**A**GTA

TG**G**ACG

CA**A**GTA

TG**G**ACG

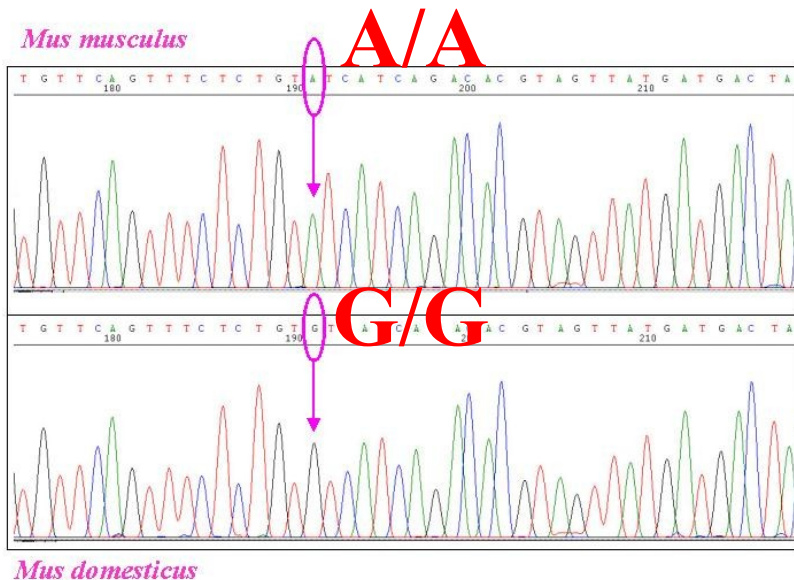
**A/A**



Př.: chromozóm 1

# Příklad informativního SNP znaku

- fixovaný polymorfismus (homozygoti) = **diagnostické SNPs** - využití např. při studiu hybridizací (hybridi = heterozygoti)



## Značení heterozygotů

N = A, C, G, T  
V = G, A, C  
D = G, A, T  
H = A, T, C  
B = G, T, C  
R = A, G  
Y = C, T  
M = A, C  
K = G, T  
S = G, C  
W = A, T

**transice**

**A ↔ G**

Synonymní vs. nesynonymní substituce

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 (+ introgrese částí genomu)
- fylogeografie
- populační genetika (genetická variabilita a struktura, tok genů, identifikace jedinců a vztahů mezi nimi, populační velikost a její změny atd.)
- mutace ve funkčních genech – i záměna jedné aminokyseliny může mít fatální dopad
- genome-wide genotyping – asociace s fenotypem

# Výhody

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

# Nevýhody

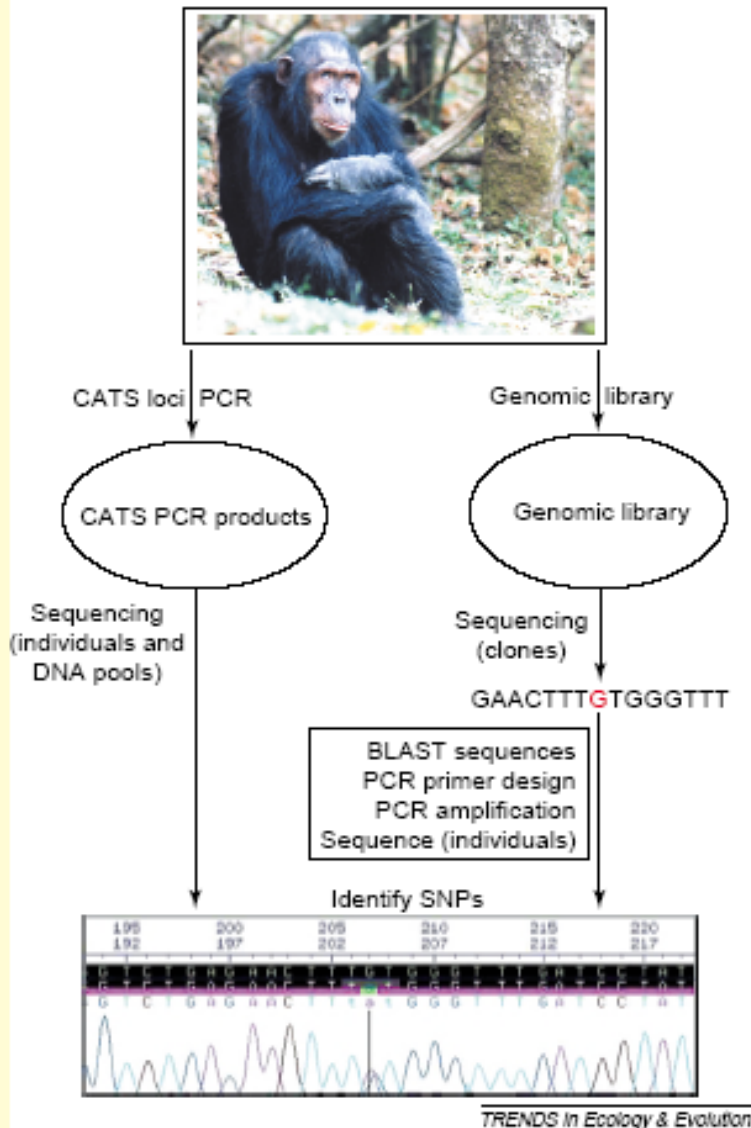
- „ascertainment bias“ – výběr informativních znaků se provádí na základě jen malého počtu jedinců a nemusí být reprezentativní
- nízká variabilita na lokus (většinou jen 2 alely)
- pro populační genetiku je vyžadován větší počet lokusů (4-10 krát více než u mikrosatelitů)

# Metody analýzy

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



# 1. Nalezení SNPs



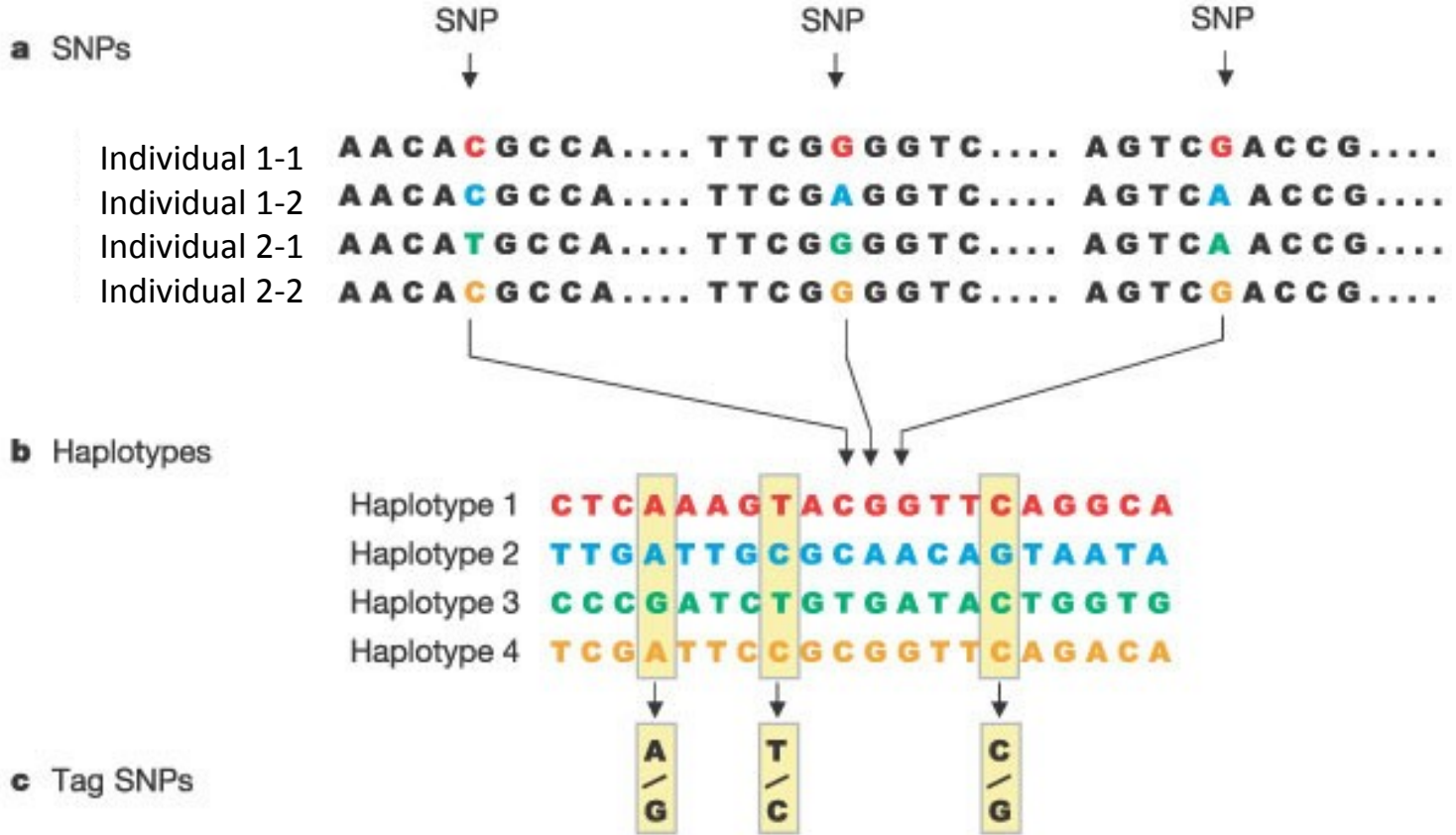
(1) CATS loci = comparative anchor tagged site loci (= cross amplification)

(2) Genomic library = genome restriction + cloning (náhodný výběr klonů – 1 SNP každých 300-1000 bp)

**V současné době: Next-generation sequencing – sekvenování genomu více jedinců a hledání polymorfismů, např. tzv. RAD sequencing (viz další přednášky)**

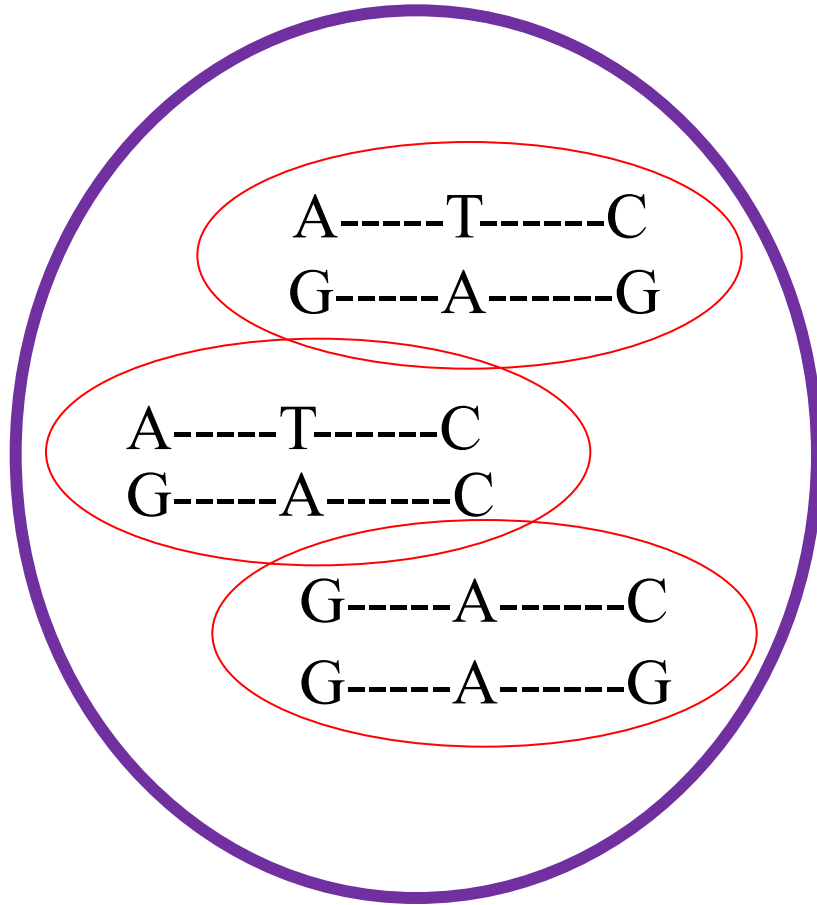
# Analýza NGS dat:

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



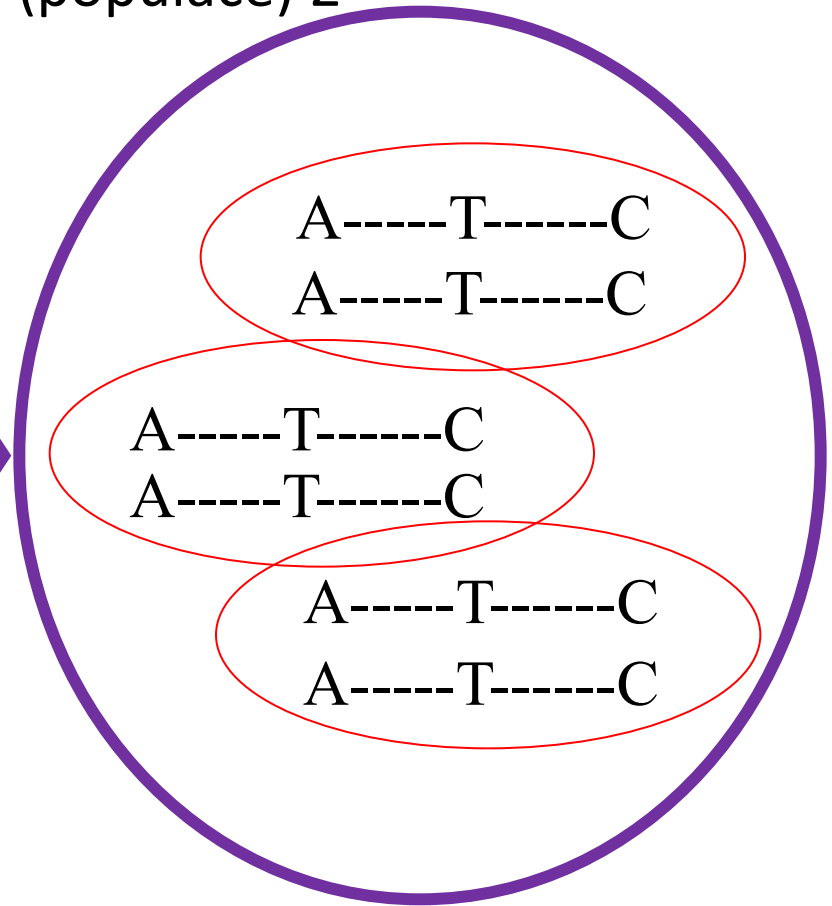
# Ascertainment bias

Druh (populace) 1



Analýza 3 jedinců u druhu (populace) 1  
Tři polymorfní (informativní) SNPs

Druh (populace) 2



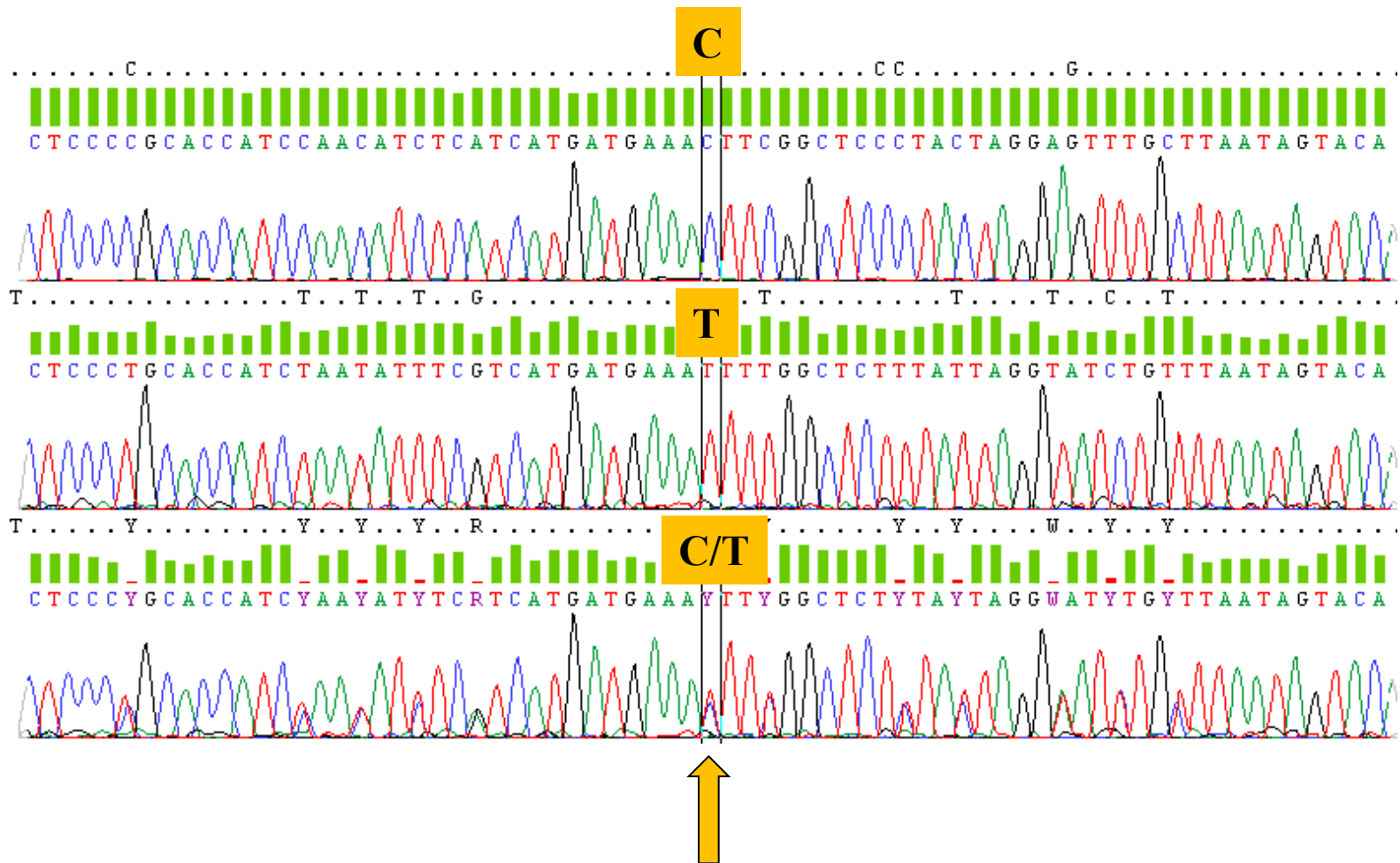
Polymorfismus daných SNPs je druhově  
(populačně) specifický

## 2. SNPs genotyping

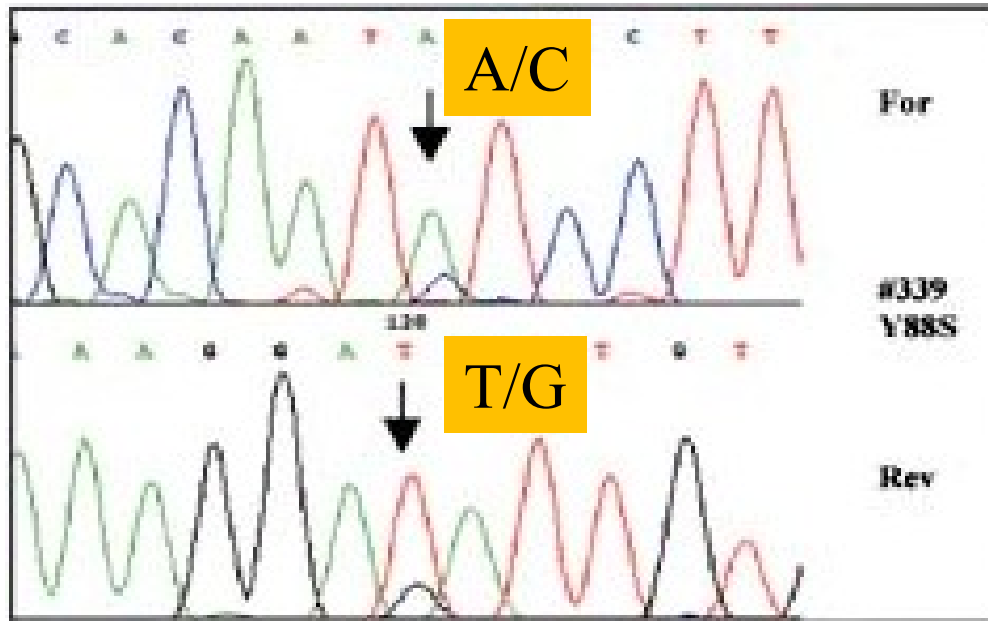
= zjištění genotypu daného jedince

# SNPs genotyping - sekvenování?

Je drahé a nejasné u heterozygotů



# Heterozygotes?



Sekvenování z obou stran - are you really sure?

# SNPs genotyping - klonování a následné sekvenování?

- rozdělení dvou alel (či více u duplikovaných genů)

každý klon obsahuje jen jednu alelu

vector =  
plasmid

izolace vektorů

**!!! cloning – cca 800 Kč**  
**!!! sequencing 1 clone – cca 100 Kč**

PCR product

sekvenování insertů

↑ ligation, transformation



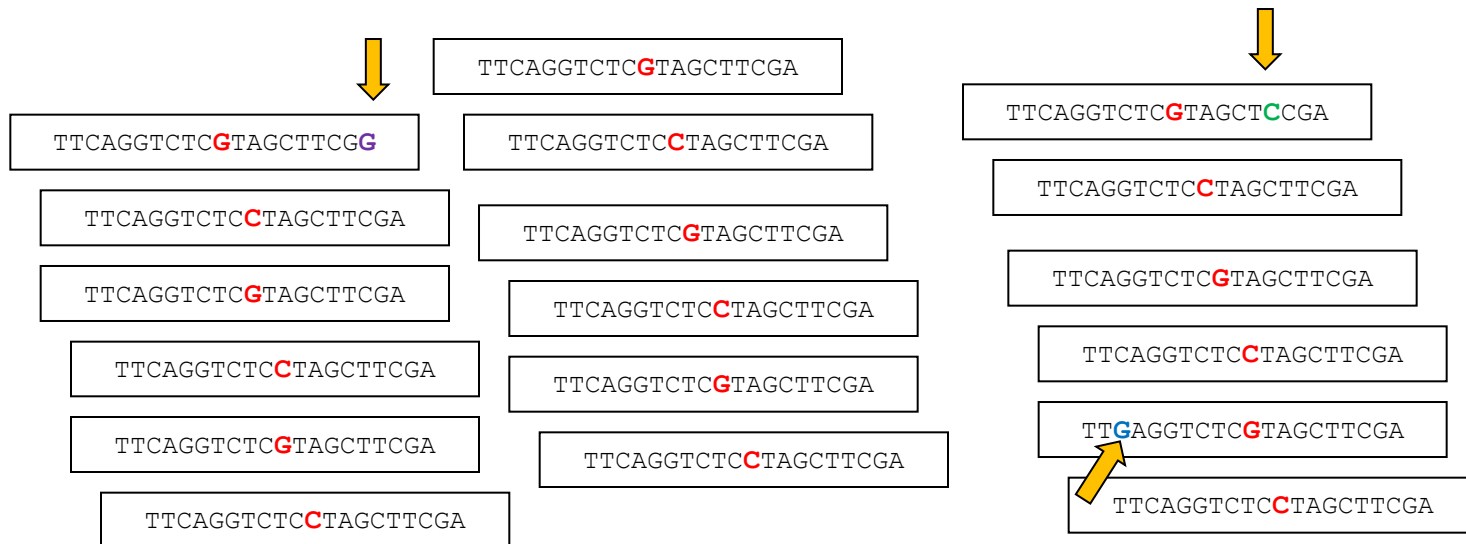
Ex.: heterozygote = two diff. alleles

# PCR is making substitution errors that are visualised by cloning (!)

TTCAGGTCTC**G**TAGCTTCGA

TTCAGGTCTC**C**TAGCTTCGA

... před PCR = heterozygot G/C



## PCR artefacts

(= šum při standardním sekvenování, ale velmi jasné při sekvenování klonů)



# SNPs genotyping

## 1. Old standards (PCR-based)

- RFLP: PCR + štěpení + standardní elfo
- DGGE, TGGE, SSCP: PCR + nestandardní elfo
- původně diagnostika geneticky podmíněných chorob, např. cystická fibróza

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

- HRM: high-resolution melting (real-time 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

„Jelenovi pivo nelej“

„A dál vidí lítat netopýry potentát i lid i vláda“

## 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)

**Restriction site**

**Palindrome**

G T A G G A A T T C A T T T C A C G C A  
C A T C T T A A G T A A G T G C G T

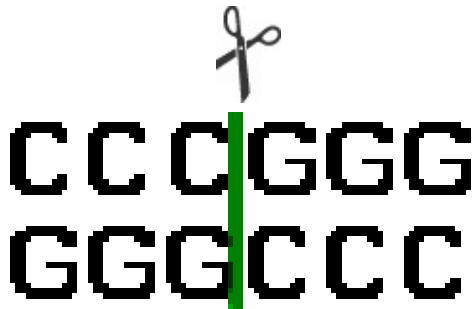
G T A G                      A A T T C A T T T C A C G C A  
C A T C T T A A                      G T A A G T G C G T

**Fragment 1**

**Fragment 2**

# Běžné restriční enzymy

„blunt ends“



**SmaI**  
– blunt end

„sticky ends“



**EcoRI**  
– *Escherichia coli*  
– 5 prime overhang



**PstI**  
– *Providencia stuartii*  
– 3 prime overhang

# SNP genotyping - old standards

## PCR-RFLP

### Allele A

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

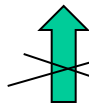


cutting by restriction endonuclease

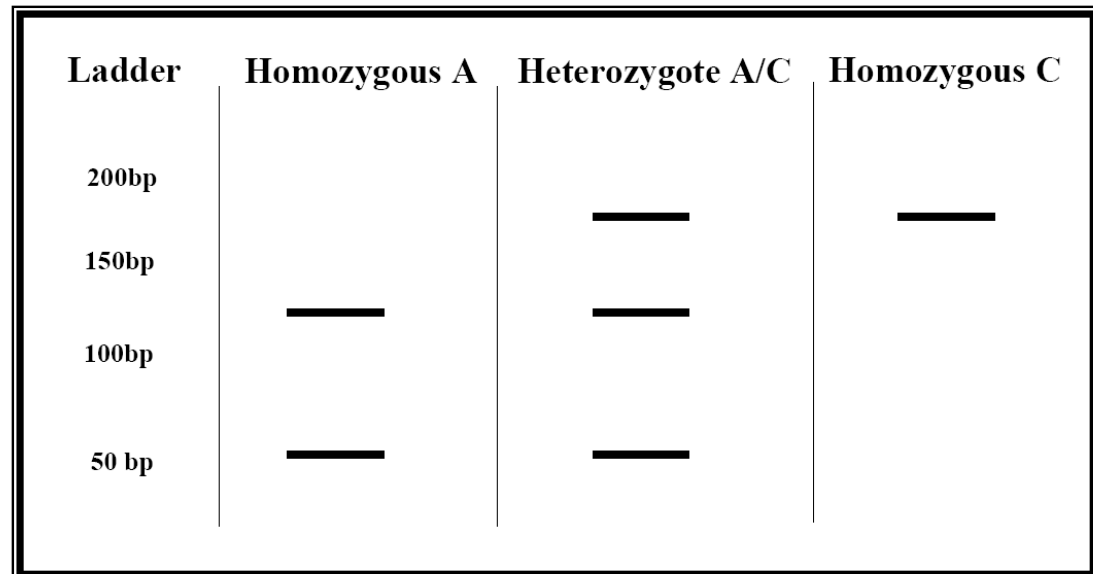
- neumožní nalézt novou variantu daného SNP (odliší pouze 2 formy daného znaku: +/- )

### Allele C

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



no cut



# SNPs genotyping - old standards electrophoresis methods of mutation detection

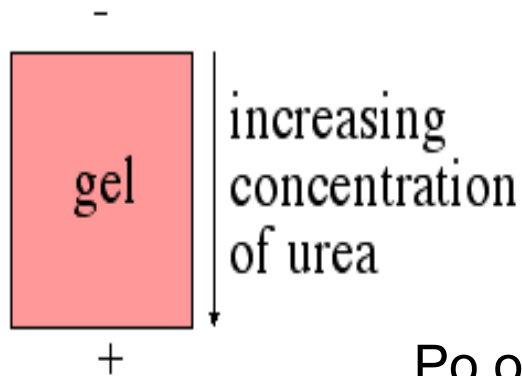
- Thermal gradient gel electrophoresis (**TGGE**)
- Denaturing gradient gel electrophoresis (**DGGE**)
- Single-strand conformation polymorphism (**SSCP**)

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

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

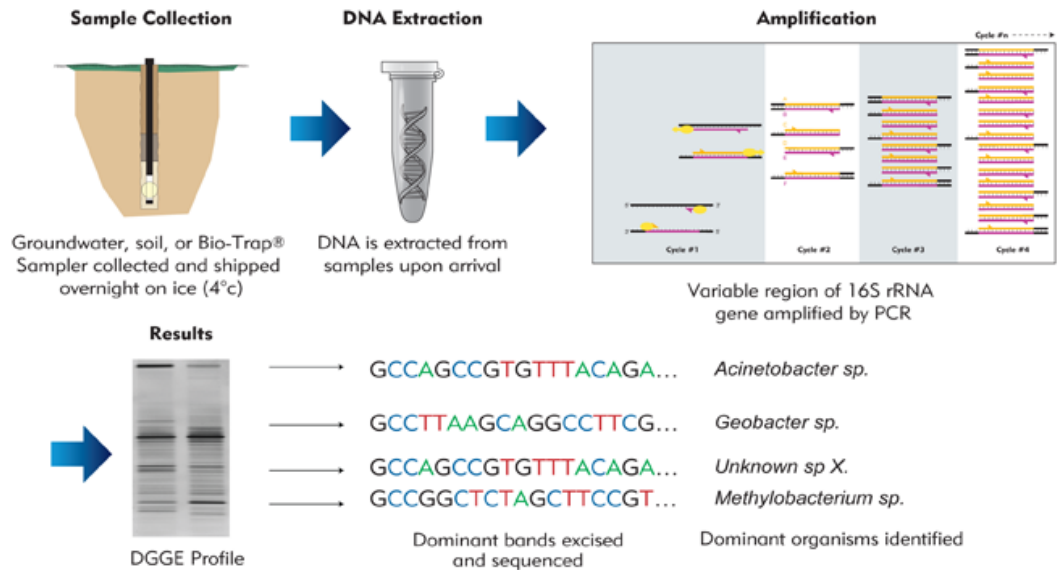
**Krátké PCR fragmenty (200-700 bp) jsou separovány v denaturačním gradientu (PAGE = polyakrylamidový gel)**

→ v určitém bodě začne DNA denaturovat („melting point“) – závisí na sekvenci, tj. každá sekvence denaturuje při jiné koncentraci močoviny



Denaturované fragmenty putují v gelu pomaleji

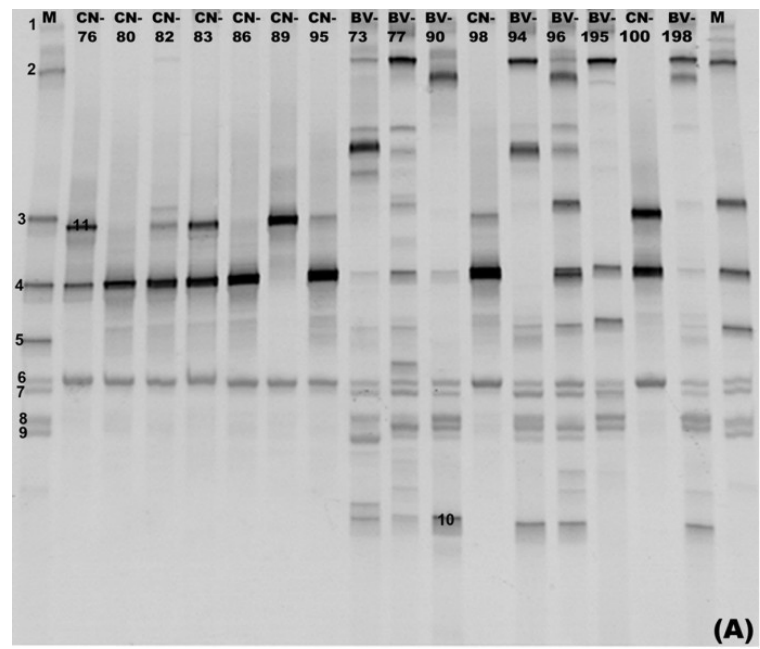
Po obarvení lze vidět rozdílné pozice PCR produktů v závislosti na jejich sekvenci



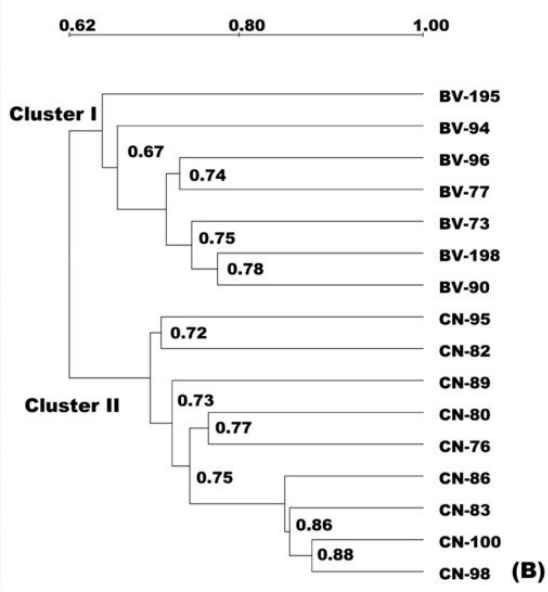
# DGGE v bakteriální metagenomice



Dnes rychle nahrazováno NGS



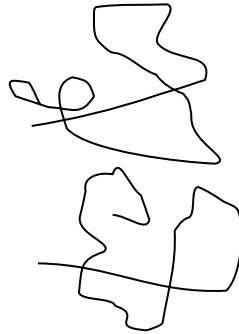
Dendrogram using UPGMA (Dice Coefficient)



# Single strand conformation polymorphism (SSCP)

## Allele 1 - C

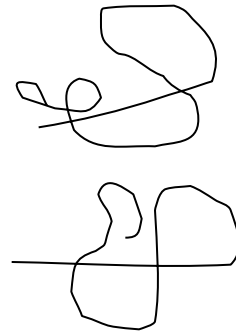
...CGCTT**C**AGG ...  
...GCGAA**G**TCC...



heating - denaturation  
snap-cooling → partial renaturation

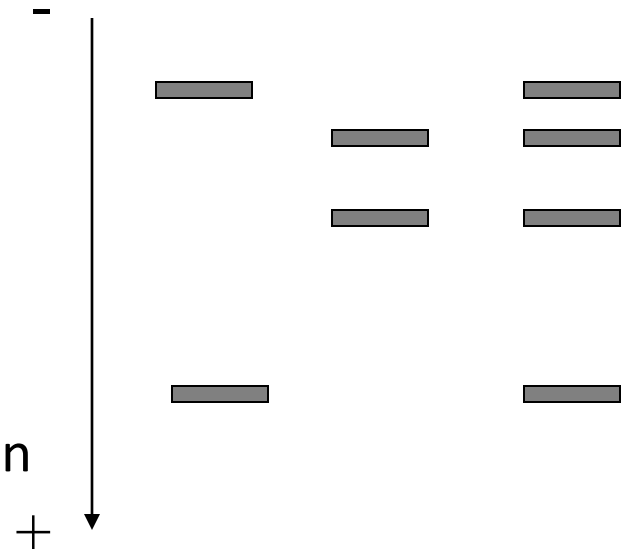
## Allele 2 - A

...CGCTT**A**AGG ...  
...GCGAA**T**TCC...



sequence-specific  
ssDNA conformations

Homo1 Homo2 Hetero

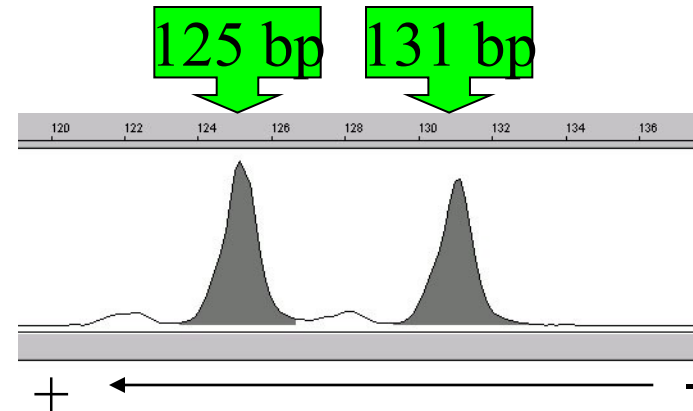


**!!! non-denaturing PAGE**

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?

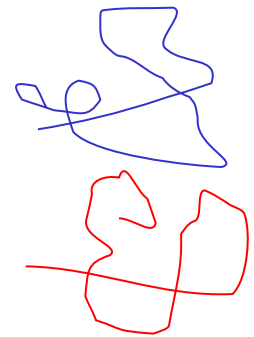
e.g. CAP (conformation analysis polymer)



- well controlled electrophoresis
- two fluorescent labels
- high sensitivity

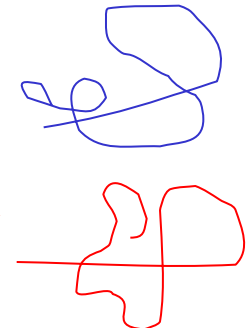
## Allele 1

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



## Allele 2

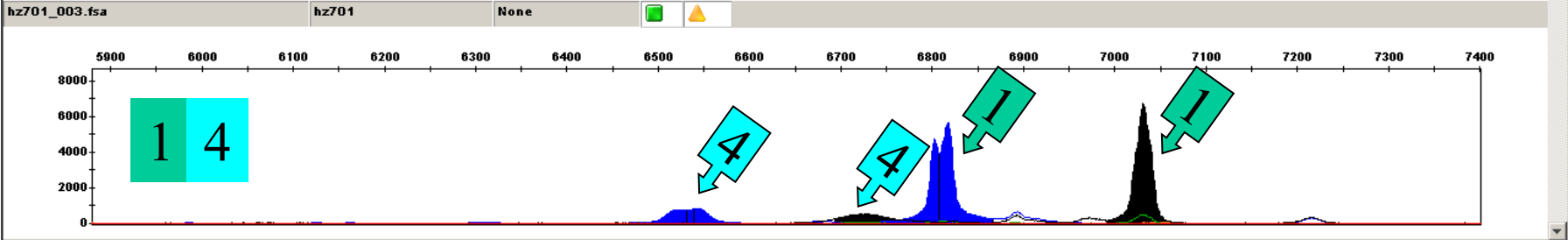
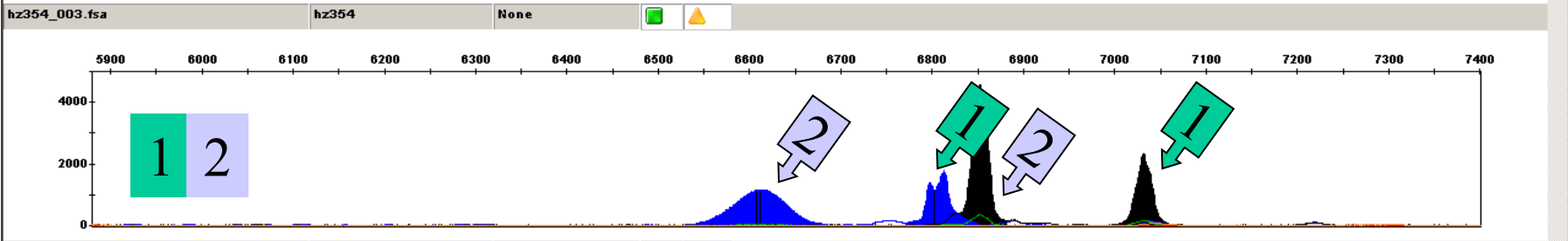
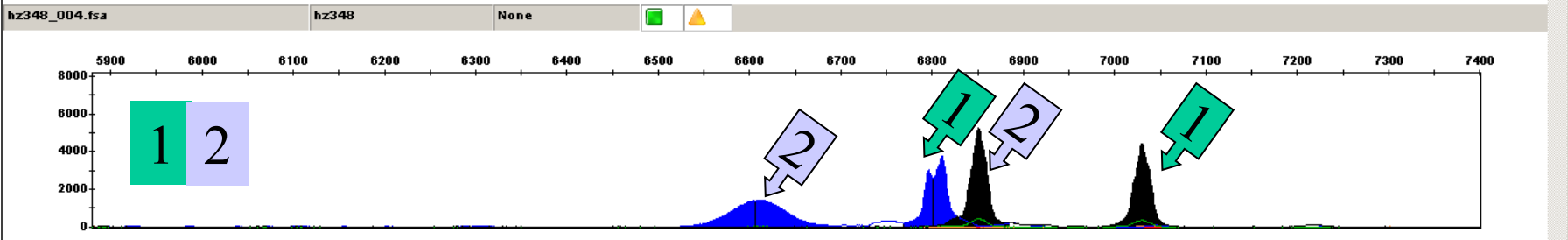
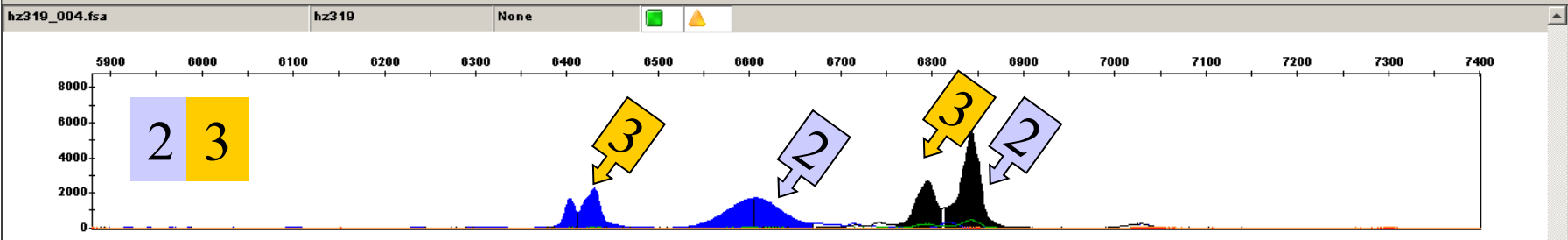
*FAM*... CGCTTAAGG ...  
... GCGAATTCC ...*HEX*



# MHC Class II (DQA gene)

Plot Setting: AFLP Default Panes: 4

Sample File Sample Name Panel OS SQ



	Dye/Sample Peak	Sample File Name	Marker	Allele	Size	Height	Area	Data Point
B.65		hz701_003.fsa			6537.54	788	6803	6530
B.66		hz701_003.fsa			6542.55	830	17081	6535

[X 6128,09 Y 5929]

# Analýza elektroforetogramů

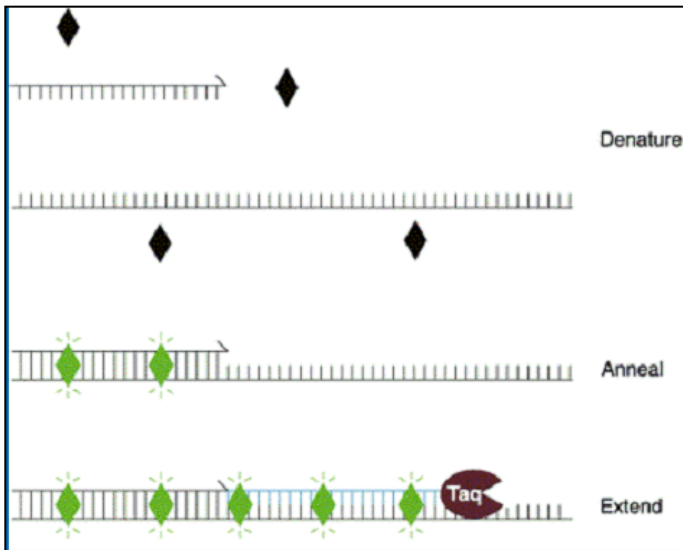
- např. GeneMapper (Applied Biosystems)
- specifický „Size+Conformation Standard“ pro každou teplotu (konformace závisí na teplotě)
- srovnání více vzorků
- umožňuje detekci krátkých odlišných sekvencí s více SNPs (užitečné např. pro genotypizaci MHC, tj. vysoce variabilních genů)
- opět rychle nahrazováno „next-generation sequencing“

# SNP genotyping - new methods

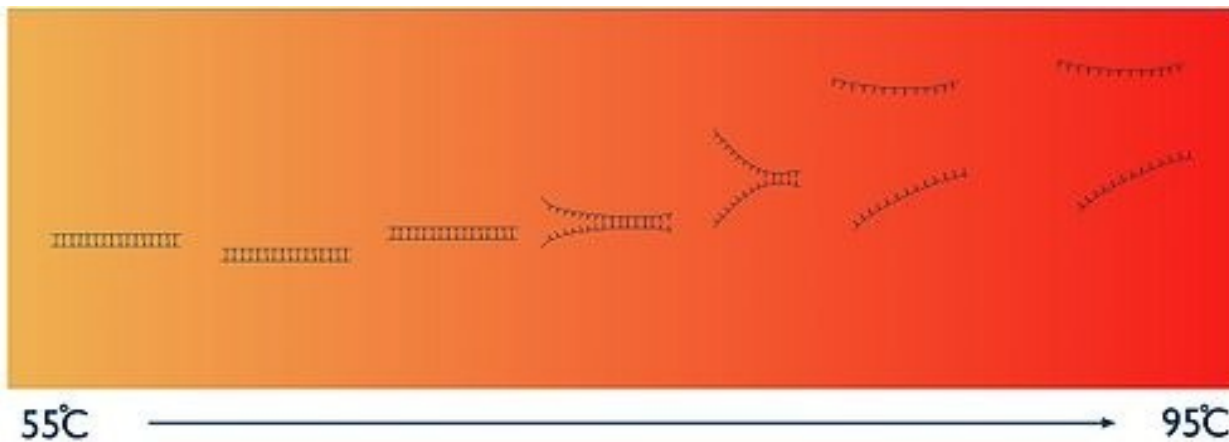
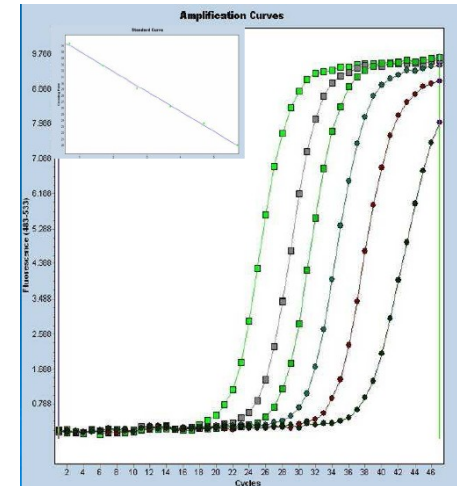
= not based on standard PCR

1. high-resolution melting temperature (HRMT)
  2. real-time PCR se specifickými sondami (TaqMan, molecular beacon)
  3. ASPE: allele-specific primer extension
  4. SBE: single base extension
  5. Alelově-specifická hybridizace
- } mohou využívat tzv. microarrays („SNP chips“)

# 1. High-resolution melting temperature (HRMT)

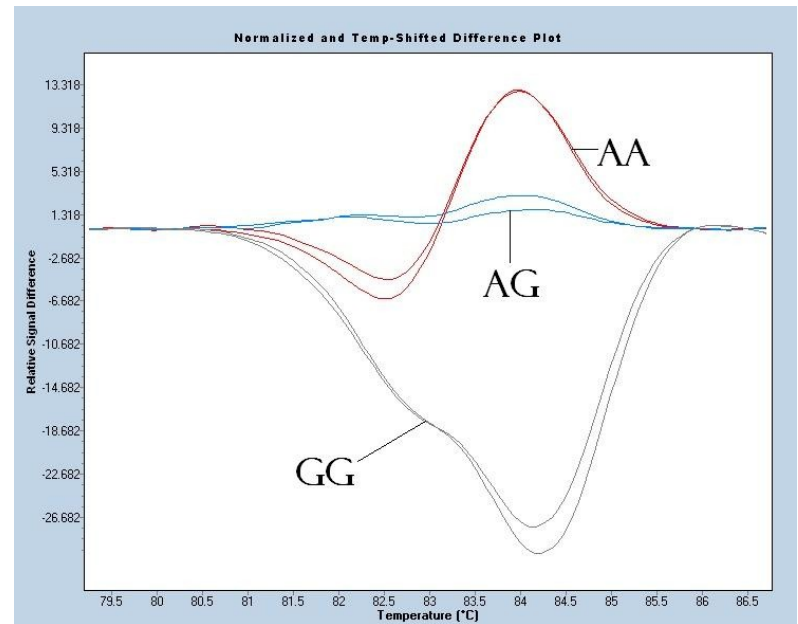
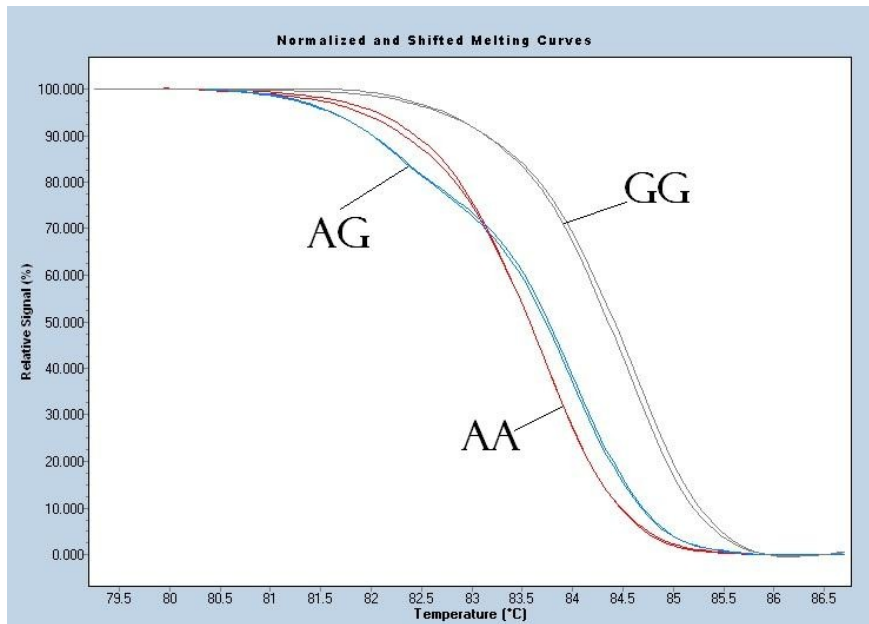


Step 1: real-time PCR = increase of fluorescence



Step 2: measuring melting after PCR = decrease of fluorescence

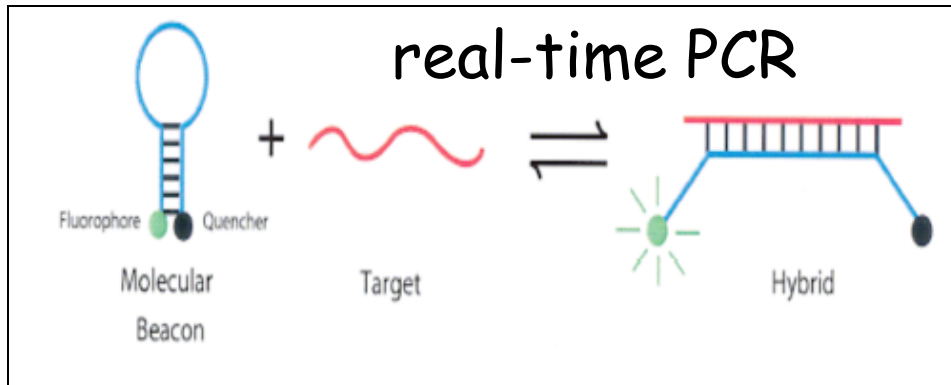
# HRMT genotyping



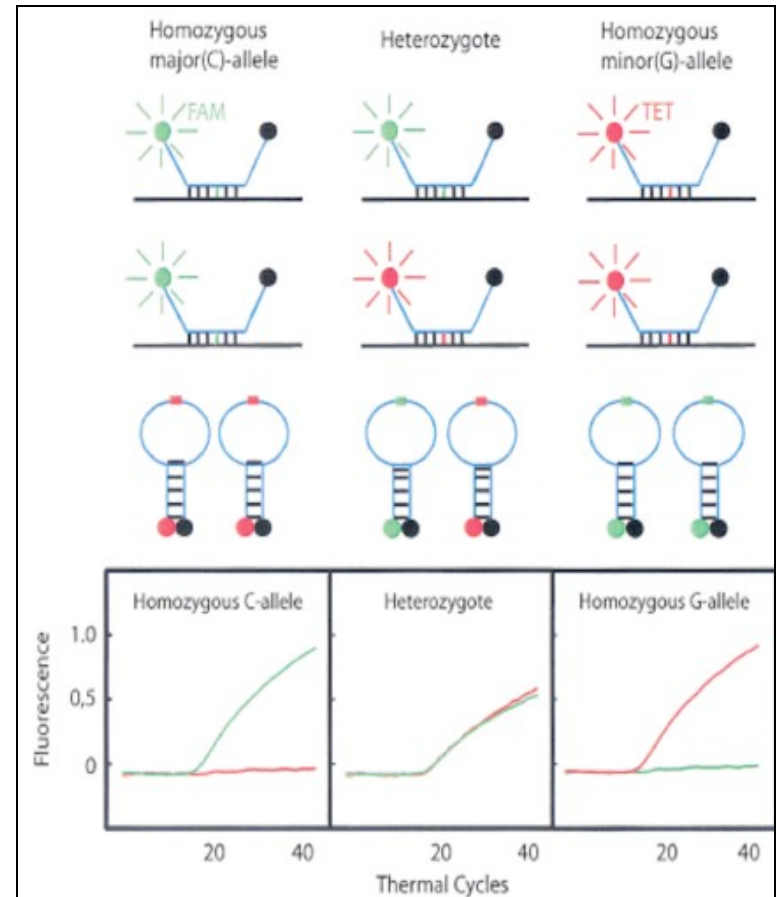
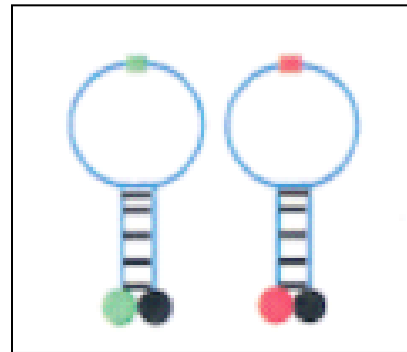
## Detekce heterozygotů

- velmi levná a jednoduchá metoda - v podstatě jen qPCR
- vhodná na genotypizaci jednoduchých SNP u velkého množství vzorků

# 2. Real-time PCR se specifickou sondou



sondy  
specifické pro  
jednotlivé alely

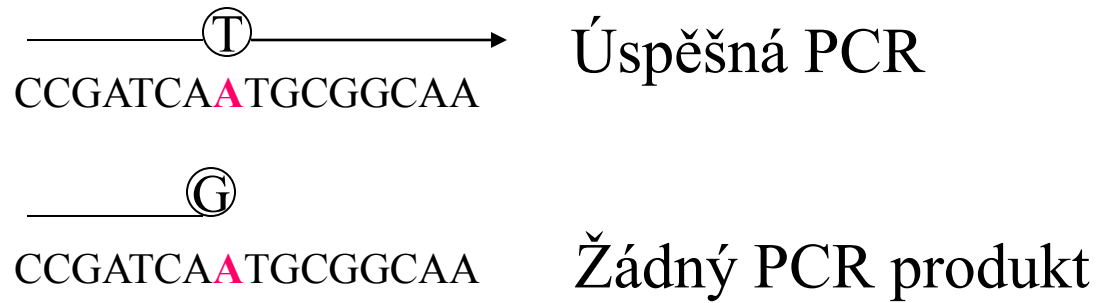


1) TaqMan sondy

2) Molecular Beacons („maják“)

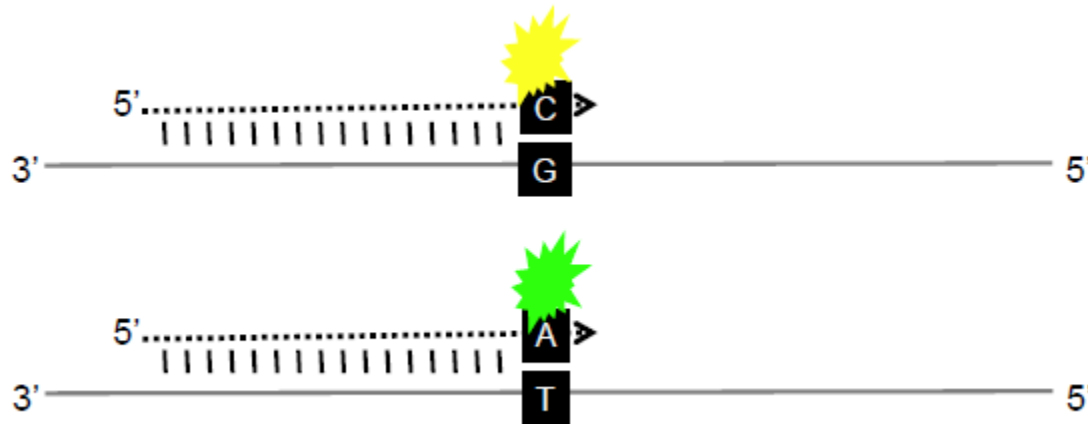


### 3. ASPE: allele-specific primer extension



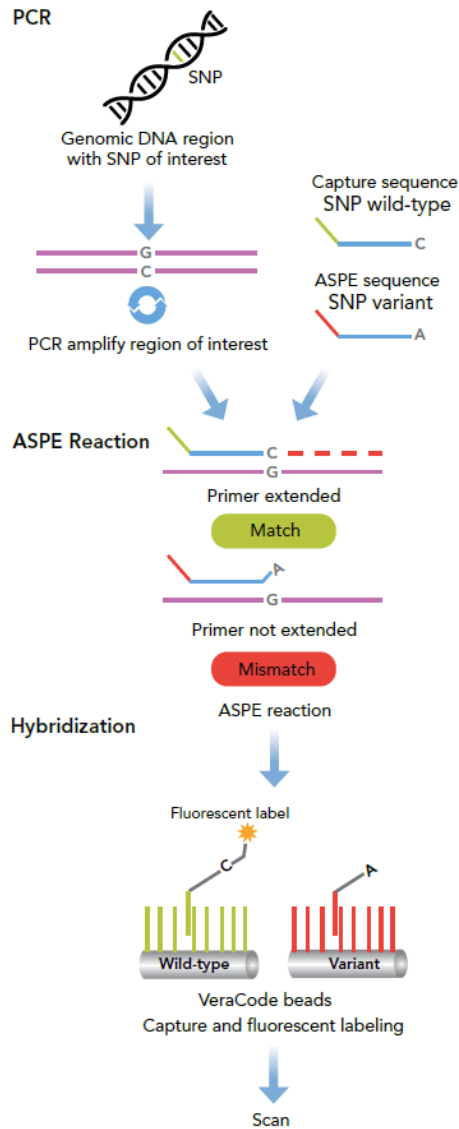
- dvě PCR (každá se specifickými primery k danému SNP)
- 3' terminální nukleotid na primerech je komplementární k SNP nukleotidu
- alelově-specifická amplifikace je umožněna vysoce specifickou polymerázou

# ASPE: allele-specific primer extension (automatizovaná verze)



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

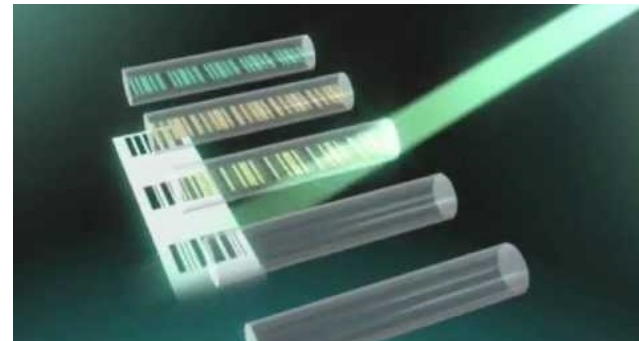
Example: Precise Genotyping Calls Made using ASPE on the VeraCode Platform



# Illumina – GoldenGate ASPE

(web-based VeraCode Assay Designer)

Each VeraCode Capture Bead contains a unique 23-mer oligonucleotide immobilized on its surface. Designing ASPE extension primers that include complementary sequences to these capture oligos allows exclusive targeting of specific beads. Primers that match the targeted sequence will extend preferentially. When a labeled target hybridizes with the complementary sequence on the assigned VeraCode microbead, the target is identified through the embedded holographic digital code.



VeraCode Capture Beads  
fluorescenční detekce

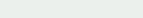
# Kompetitive Allele Specific PCR

## 1) Assay components:

### A) Primer mix

Allele specific forward primers:

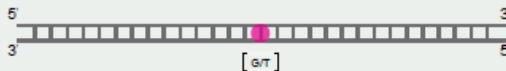


Reverse primer:  
3'  5'

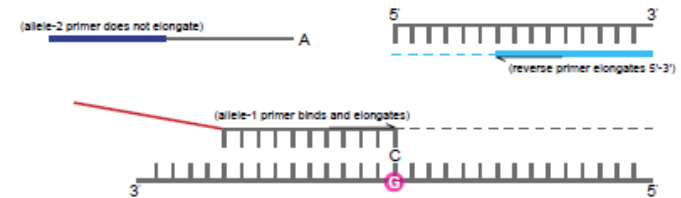
### B) Master mix



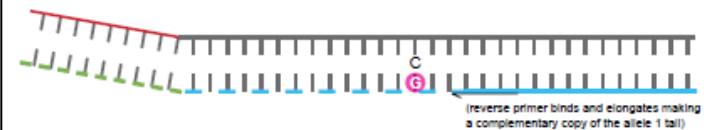
### C) DNA template (sample)



## 2) Denatured template and annealing components – PCR round 1:



## 3) Complement of allele-specific tail sequence generated – PCR round 2:

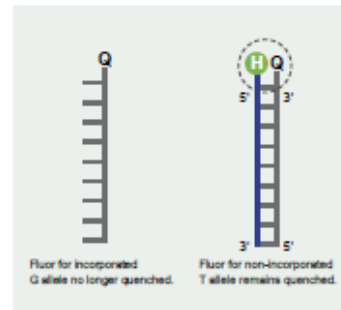


## 4) Signal generation – PCR round 3:








Thermal cycling results in exponential increase in allele-1 amplicon. As PCR continues, an increasing amount of FAM labelled oligo binds to the allele-1 amplicons. Fluorescence occurs as FAM labelled oligo is no longer quenched.



Allelic discrimination achieved through competitive annealing of two allele-specific forward primers, each containing a unique tail sequence that corresponds with a distinctly labelled FRET cassette in the master mix.



## Legend

-  Allele-1 tail FAM™-labelled
-  Allele-2 tail HEX™-labelled
-  Common reverse primer
-  FAM™ dye
-  HEX™ dye
-  Target SNP
-  Quencher

The KASP™ genotyping assay  
from LGC Genomics

# Cena analýzy

## The KASP™ genotyping assay from LGC Genomics



**Small scale study of 15 SNPs genotyped over 96 samples where no Assay on Demand (an alternative type of assay from ABI) SNP exists**

	LGC Genomics cost	ABI Taqman® Assay by Design
SNP assay design costs (validation)	<b>£1,620.00</b>	£6,750.00
Genotyping cost	<b>£701.50</b>	£388.80
Total	<b>£2,321.50</b>	£7,138.80

### Single nucleotide polymorphisms

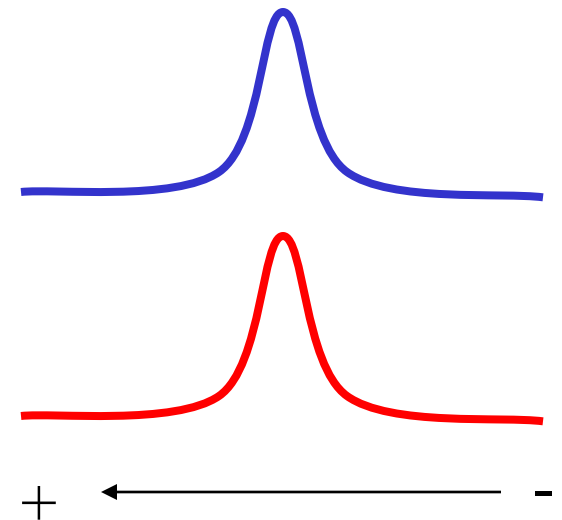
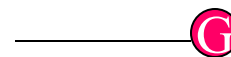
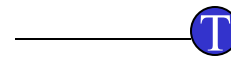
KASP assays can be designed to detect single nucleotide polymorphisms within any organism. The SNP of interest should be submitted within [square brackets] and can either be formatted as [allele1/allele2] or [IUPAC code].

```
CTTAGATCGACAGGTCTAAGAGCTGAAGAGCTAGCTATTAAGTCGAGC[C/G]
AGCTGCTAGACGTCGCAGTCGACACAGCTAGCCTAGGACAAAGTCTCGTG

CTTAGATCGACAGGTCTAAGAGCTGAAGAGCTAGCTGATTAAGTCGAGC[S]
AGCTGCTAGACGTCGCAGTCGACACAGCTAGCCTAGGACAAAGTCTCGTG
```

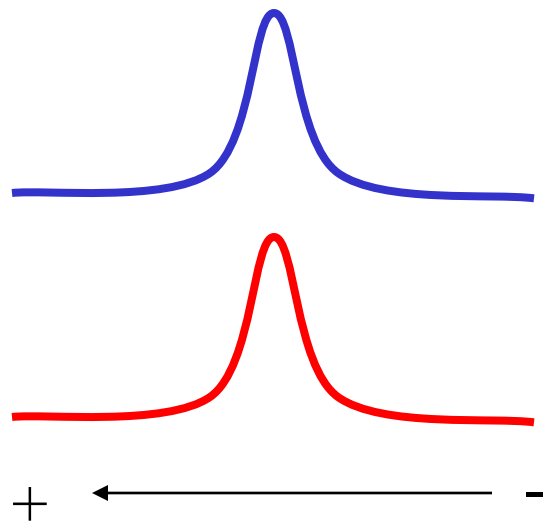
Figure 1. Examples of how to format sequence for KASP assay design for a SNP

# 4. SBE: single base extension

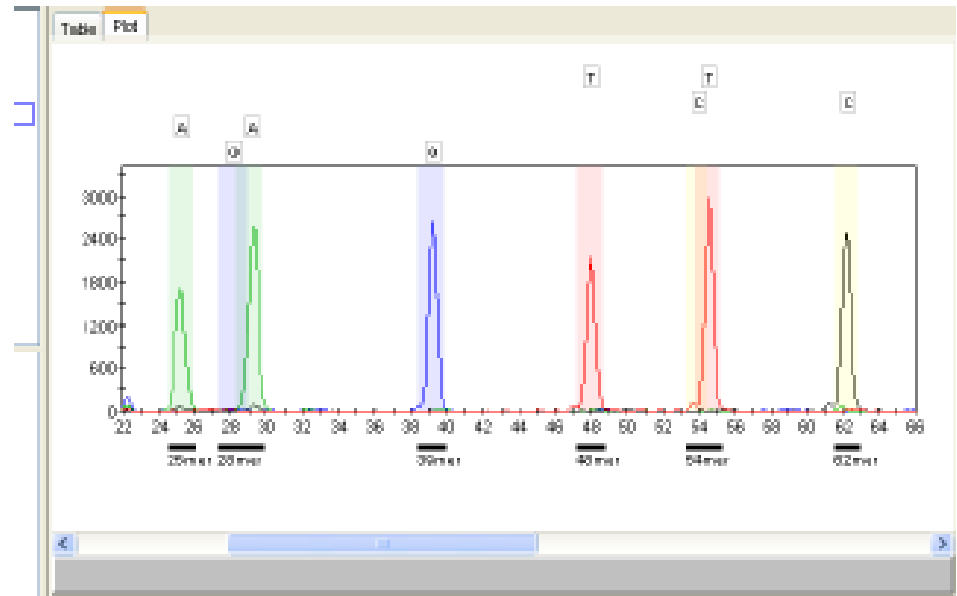


- pouze jeden dideoxynukleotid je přidán k primeru
- detekce různými metodami

# (A) Detekce SBE produktů kapilární elektroforézou



kapilární elektroforéza  
SNaPShot Multiplex Kit  
(Life Technologies)



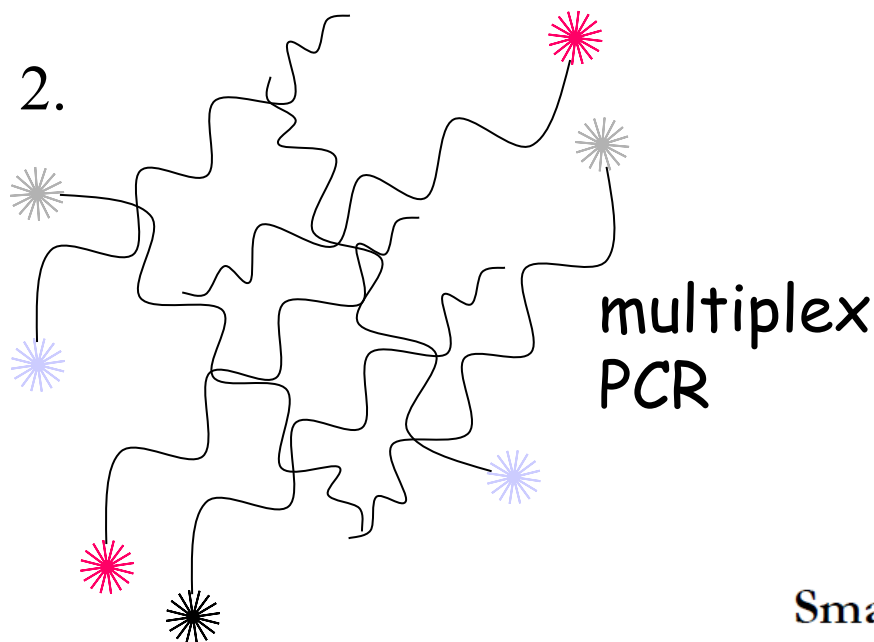
„multiplex version“ - různě  
dlouhé primery, aby bylo  
možné odlišit různé lokusy

# (B) Detekce SBE produktů přes „microarray“ (tj. hybridizace)

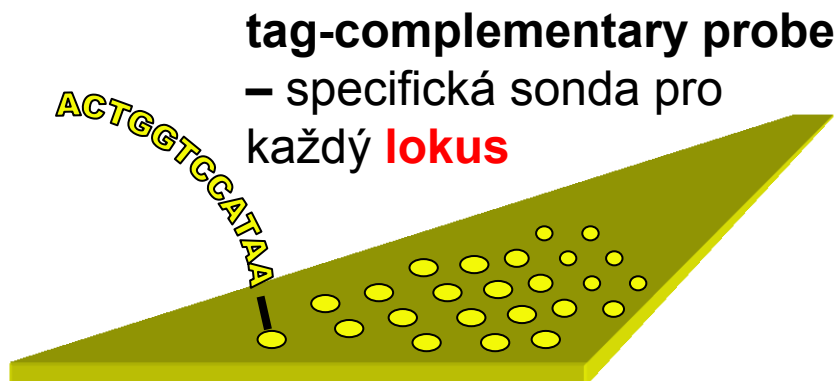
1. **tag** – specifický pro každý lokus



2.

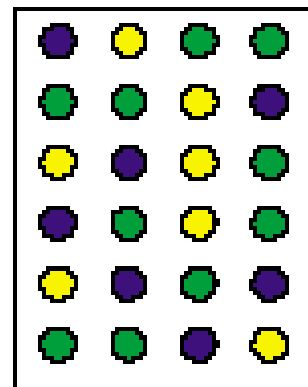


3.



4.

● G/G  
● A/A  
● G/A

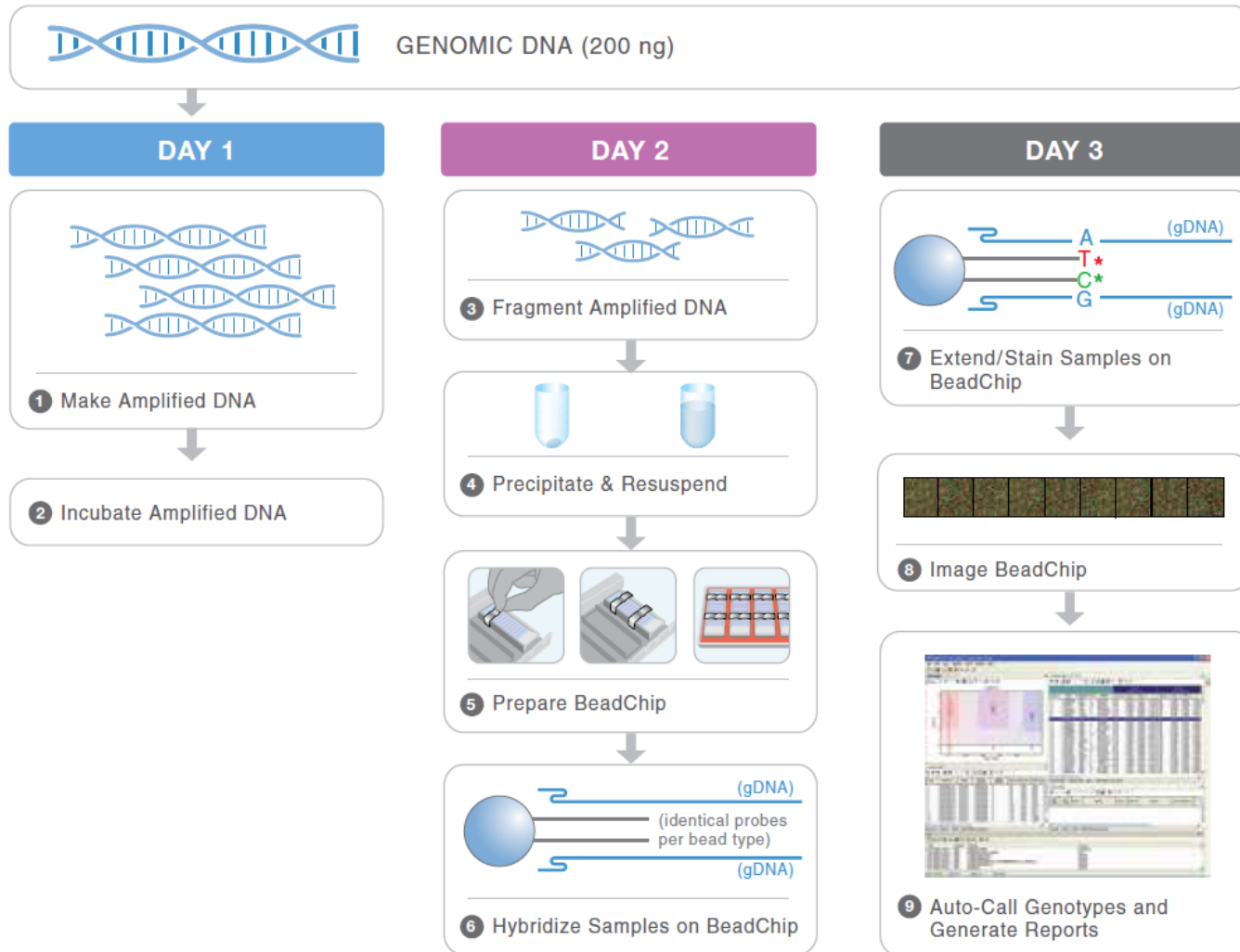


Small-scale “in house” SNP genotyping

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



# Illumina Infinium Bead Chip

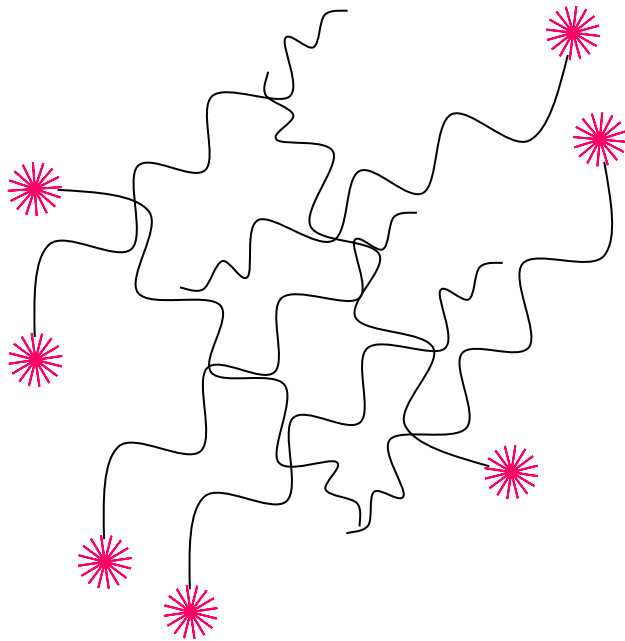


cca 300 000 SNP loci from 200 ng of DNA

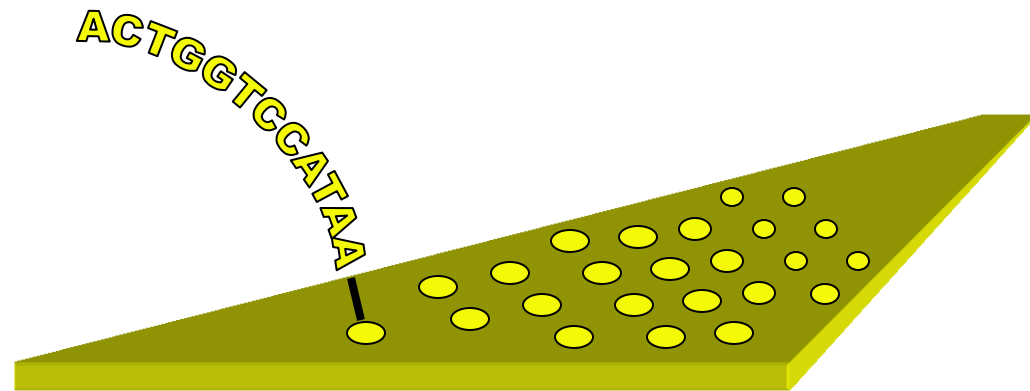
\* Indicates Stain in Red Channel  
\* Indicates Stain in Green Channel

# 5. Alelově specifická hybridizace

## Microarrays - SNPs chips

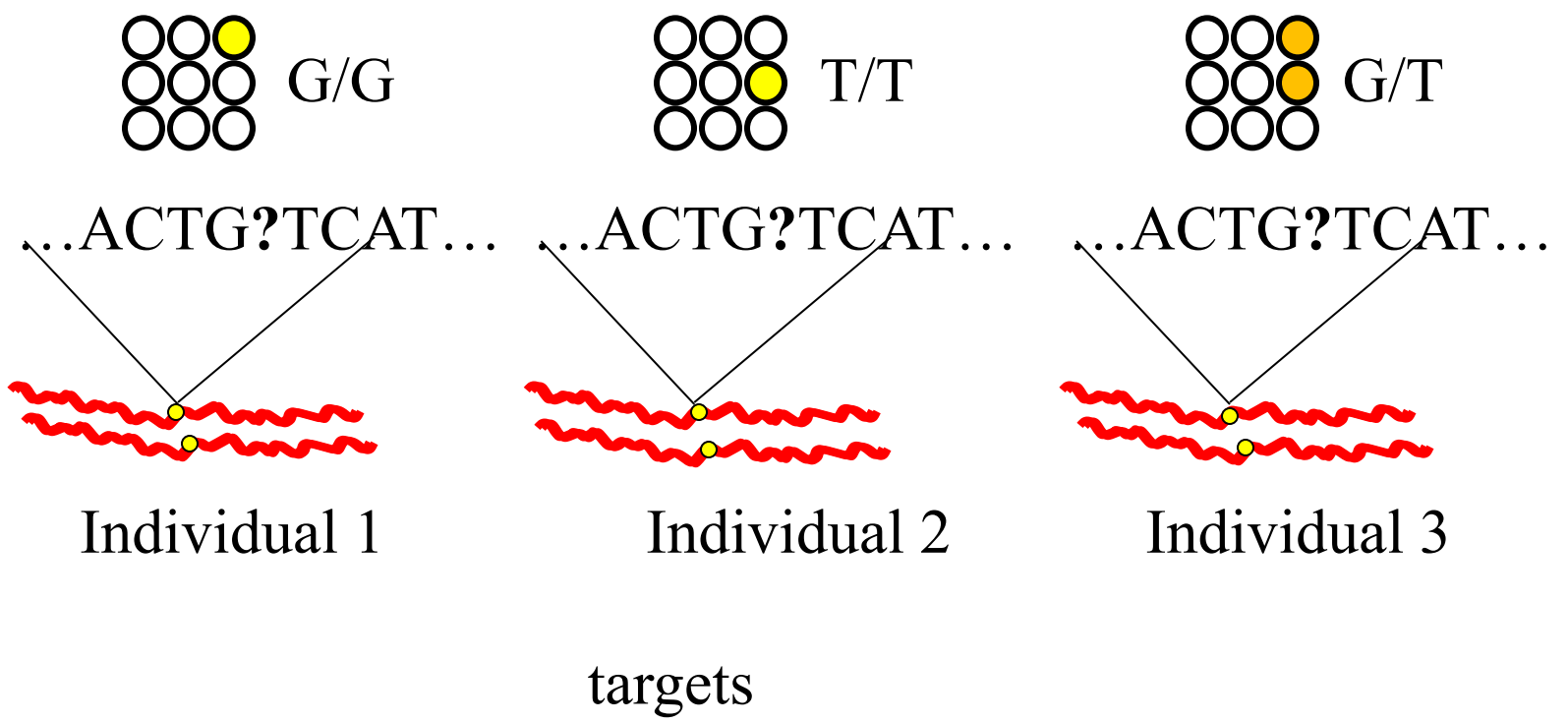


**Target** (genomická DNA  
rozštěpená restričními  
enzymy)



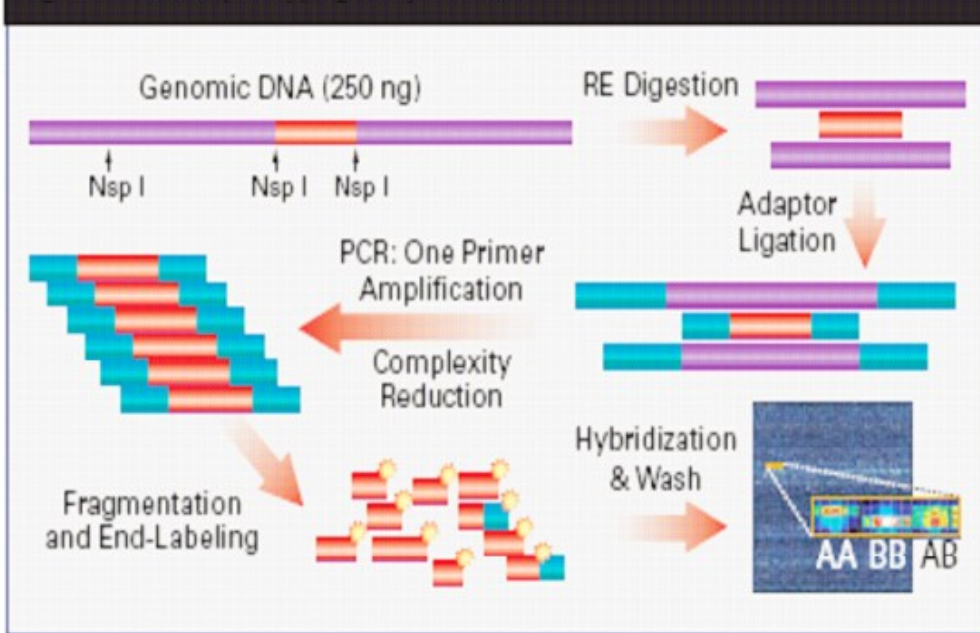
**Probe**  
(specifická sonda pro každou  
**alelu**)

# Microarray SNP Genotyping



# Detekce: např. Affymetrix

Figure 1: GeneChip® Mapping Assay Overview.



Affymetrix® Mouse Diversity Genotyping Array

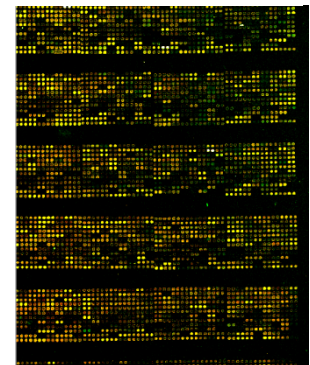


- 10 tisíc – 1 milión SNP znaků najednou – „chip technology“

- např. Mouse Diversity Genotyping Array – 623 tisíc SNPs (je známa pozice každého z nich na genomu)

- je možné si navrhnout vlastní Array

  
Affymetrix®  
Revolutionize life



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+ samples
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 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+ samples
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+ samples
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+ samples

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

# Př. ascertainment bias: MegaMUGA chips



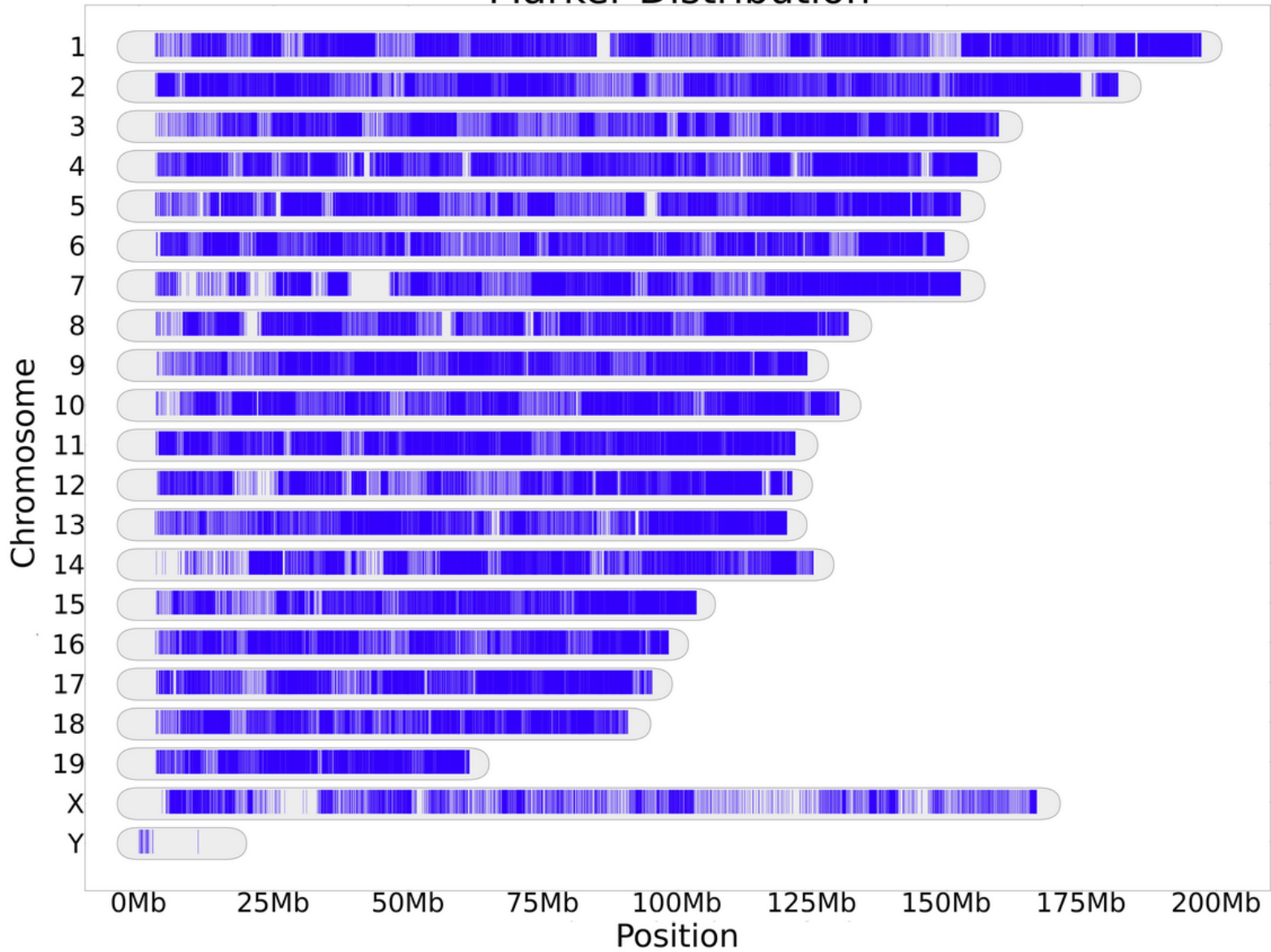
## Performance

- The MegaMUGA array provides robust calls for over 74,000 SNP markers in classical strains with greater than 99.9% concordance (based on over 300 controls).
- Initial analysis in classical inbred strains shows that on average the number of informative SNPs in pairwise combinations is ~20,000. MegaMUGA allows to discriminate between closely related sister strains (e.g. C57BL/6J and C57BL/6CR) and between related wild derived strains of similar origin (e.g. PWK/PhJ and PWD/PhJ; ZALENDE/EiJ and TIRANO/EiJ).
- Genotypes for many inbred strains can be viewed at the UNC Systems Genetics Core Facility website (<http://www.csbio.unc.edu/CCstatus/index.py?run>).

Up to 77,800 SNPs  
**\$90/\$100**  
Per DNA Sample

Využívá Illumina Infinium technologii  
(single base extension)

# Marker Distribution







# Dnes široká škála komerčních možností SNP genotyping pro nemodelové druhy - př. Illumina

Figure 2: Illumina Custom Genotyping Options

Number of Markers	Illumina Assay or Product
3K-1M	Infinium iSelectHD
1K-500K	Infinium Semi-Custom and Add-On Content
96-3,072	GoldenGate on BeadArray
1-384	GoldenGate on VeraCode
48	qPCR on Eco Real-Time PCR System

Illumina products enable a wide range of genotyping experimental designs, depending upon the number of markers.



Platform	iScan System	HiScanSQ System	BeadXpress Reader	Eco Real-Time PCR System
Technology	BeadArray		VeraCode	Real-Time PCR
Assay	InfiniumHD GoldenGate		GoldenGate ASPE	Allelic Discrimination/ High Resolution Melt (HRM)
Product	iSelectHD BeadChips; Custom and Semi-Custom Add-On; GoldenGate Genotyping Assay Kit		VeraCode GoldenGate Genotyping Assay Kit; Universal Capture Beads	Open Platform

ASPE: Allele-Specific Primer Extension

No. of loci: 3 000 – 1 milión

48-384

48

Samples/day 288

288

384