

# SINE, LINE, etc.

(Shedlock et al. 2004, TREE; Ray et al. 2007, MolEcol)

- **Transposable elements**
- Vytváří kopie (většinou)
- Kopie integrovány na nová místa v genomu
- Obvykle nejsou specificky odstraňovány
- Molekulární fosílie – neexistují homoplasie !!!
- Nesmírně početné
- Člověk – víc jak polovina genomu (ost. druhy – 40-90%)

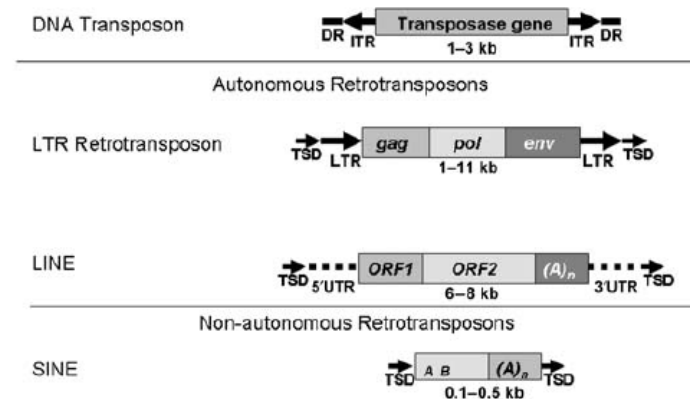


*Objev DNA transpozonů u kukuřice:  
Barbara McClintock*

# Typy transposabilních elementů

- **Kódující své proteiny, autonomní, 1-10 kb**

- **DNA transposony** (cut-and-paste)
- transposasa
- **Retrotransposony** (copy-and-paste)
- **LINE**  
1-2 proteiny, kopie přes RNA
- **LTR retrotransposony**  
5-6 proteinů, také přes RNA

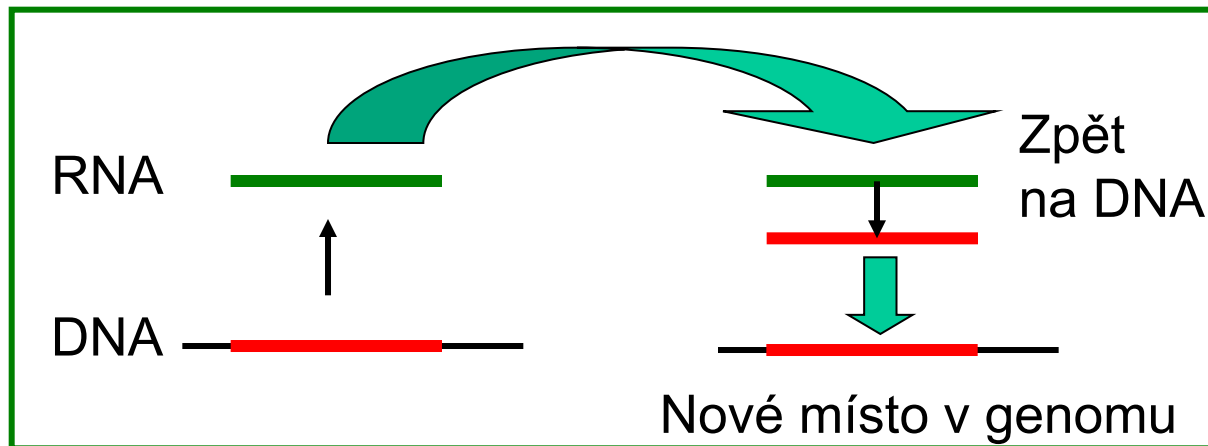


- **Nekódují proteiny, neautonomní, 100-1000 bp**

paraziti předešlých, např. **SINE** (člověk *Alu* – více než 1 milion kopií) – nejčastěji používané v populačních a fylogenetických studiích

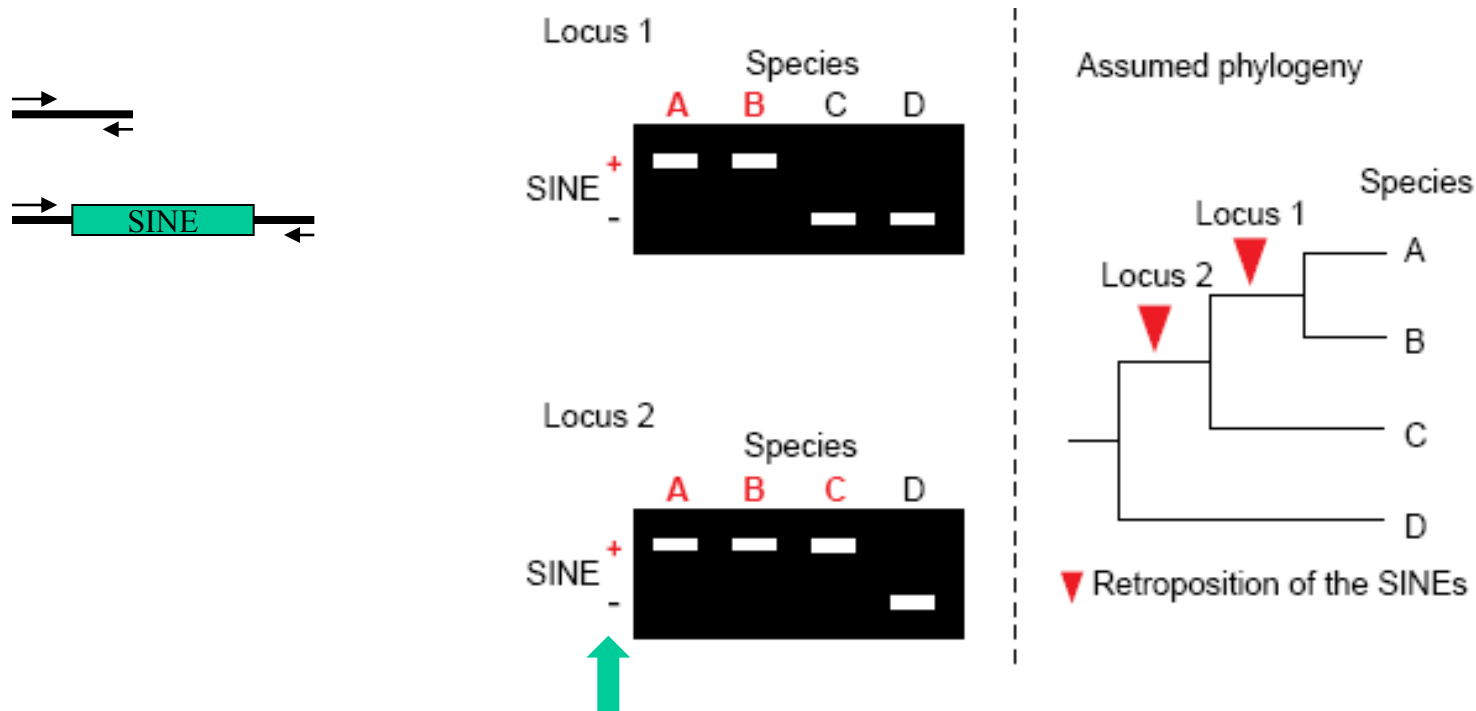
# LINE - mechanismus transpozice

- Kopie přes RNA
- Reversní transkriptáza
- Mašinerii využívají **SINE** (jsou to „paraziti“),  
*Alu* (SINE) a *L1* (LINE) se stejně rychle množí



- **LTR retrotransposony – opět přes RNA, složitější proces**

Velmi nízké riziko homoplázií →  
SINE = vhodné fylogenetické markery  
(spíše historicky, před rozvojem levného sekvenování)

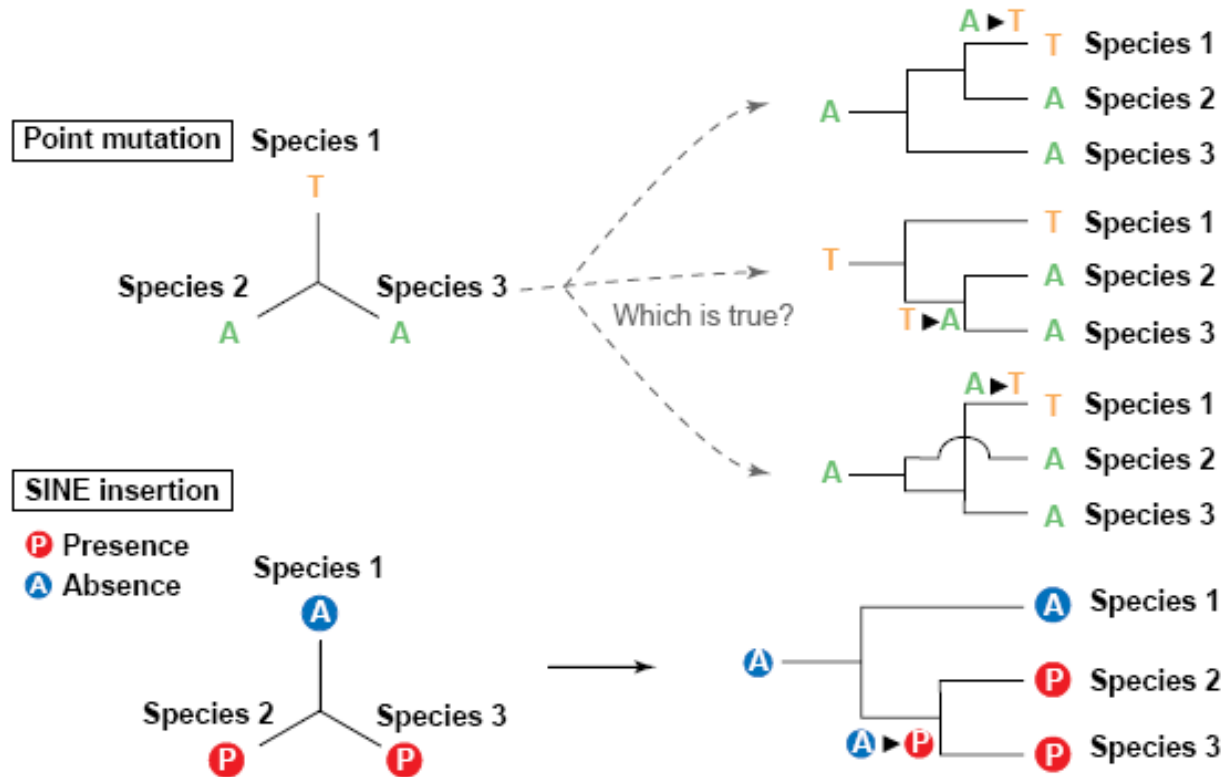


prezence/absence SINE (nikoliv póly elektroforézy)

„single-locus marker“

- PCR amplifikace daného úseku a elektroforéza

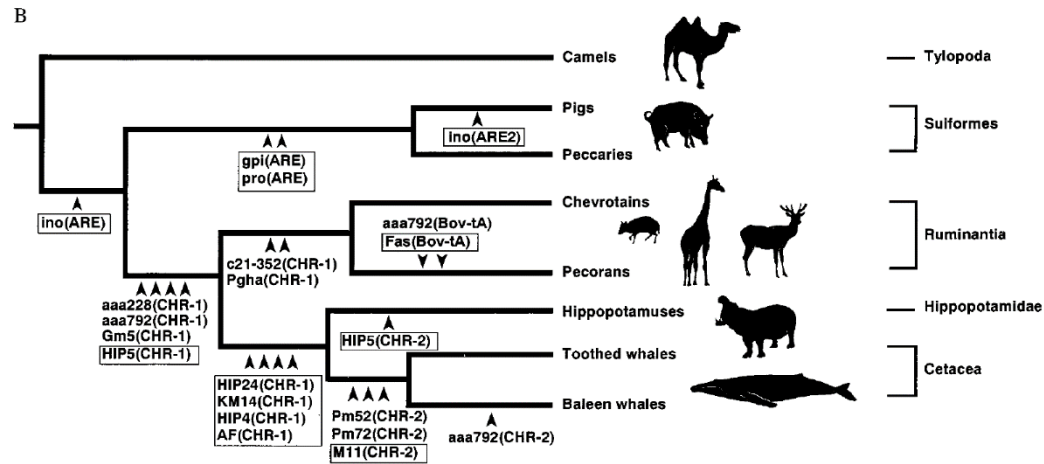
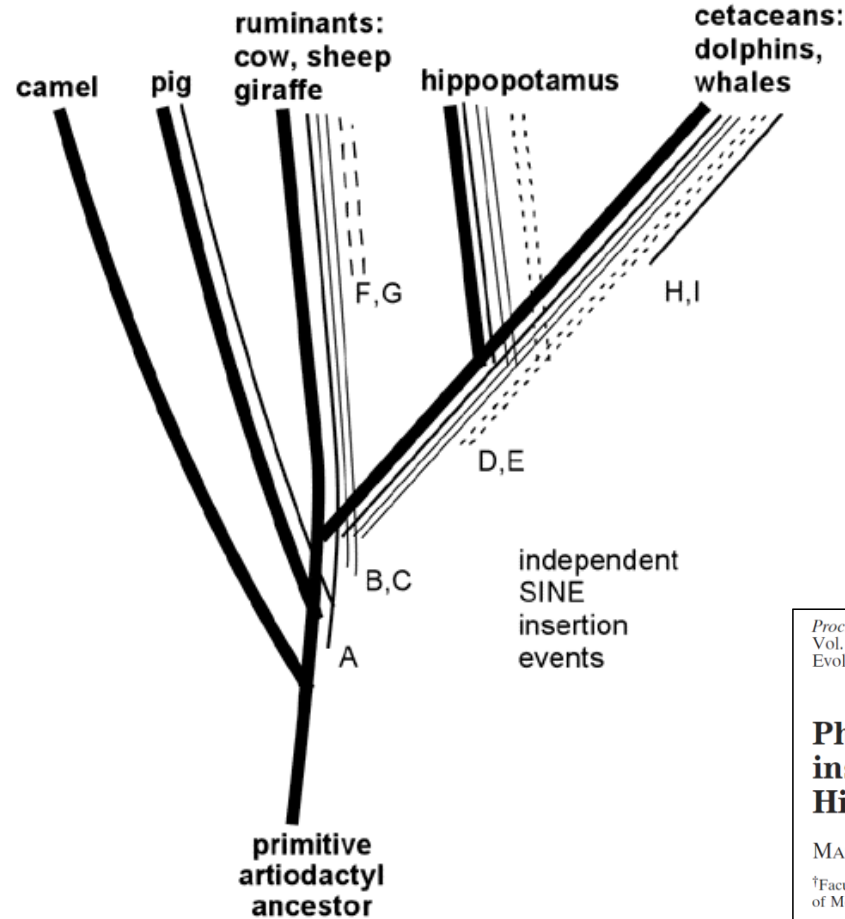
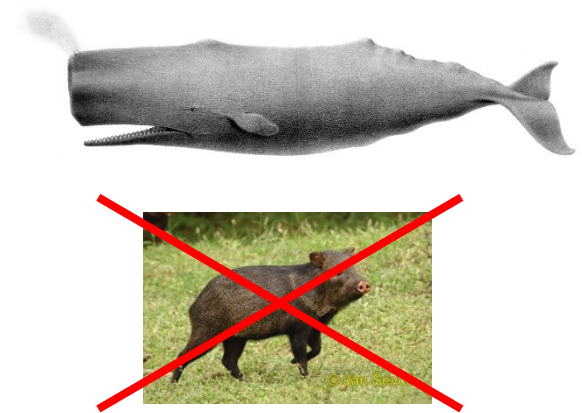
# Neexistují zpětné mutace = výhoda oproti sekvenačním datům



- jednoduchý mutační model (jednosměrné a nevratné mutace)



Norihiro Okada



*Proc. Natl. Acad. Sci. USA*  
Vol. 96, pp. 10261–10266, August 1999  
Evolution

**Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales**

MASATO NIKAIIDO<sup>†</sup>, ALEJANDRO P. ROONEY<sup>‡</sup>, AND NORIHIRO OKADA<sup>†§</sup>

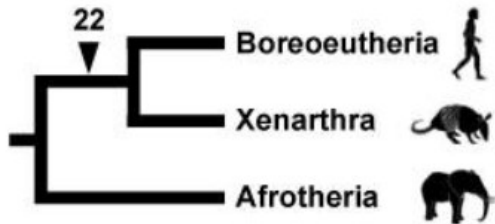
<sup>†</sup>Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Yokohama, Midori-ku, Kanagawa 226-8501, Japan; and <sup>‡</sup>Institute of Molecular Evolutionary Genetics, Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802

Příklad aplikace: kytovci vs. sudokopytníci (hroch je bratr velryby)

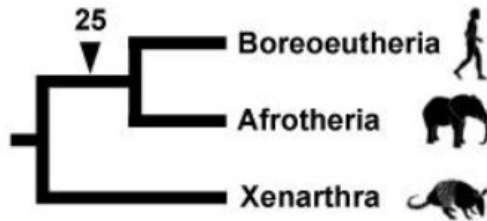
# „Incomplete lineage sorting“

A

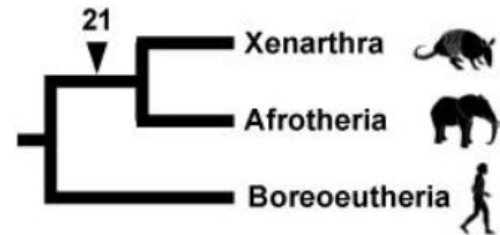
Exafroplacentalia  
(Boreoeutheria + Xenarthra)



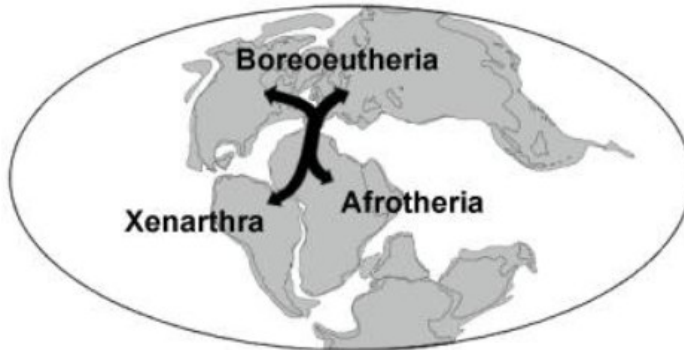
Epitheria  
(Boreoeutheria + Afrotheria)



Atlantogenata  
(Xenarthra + Afrotheria)

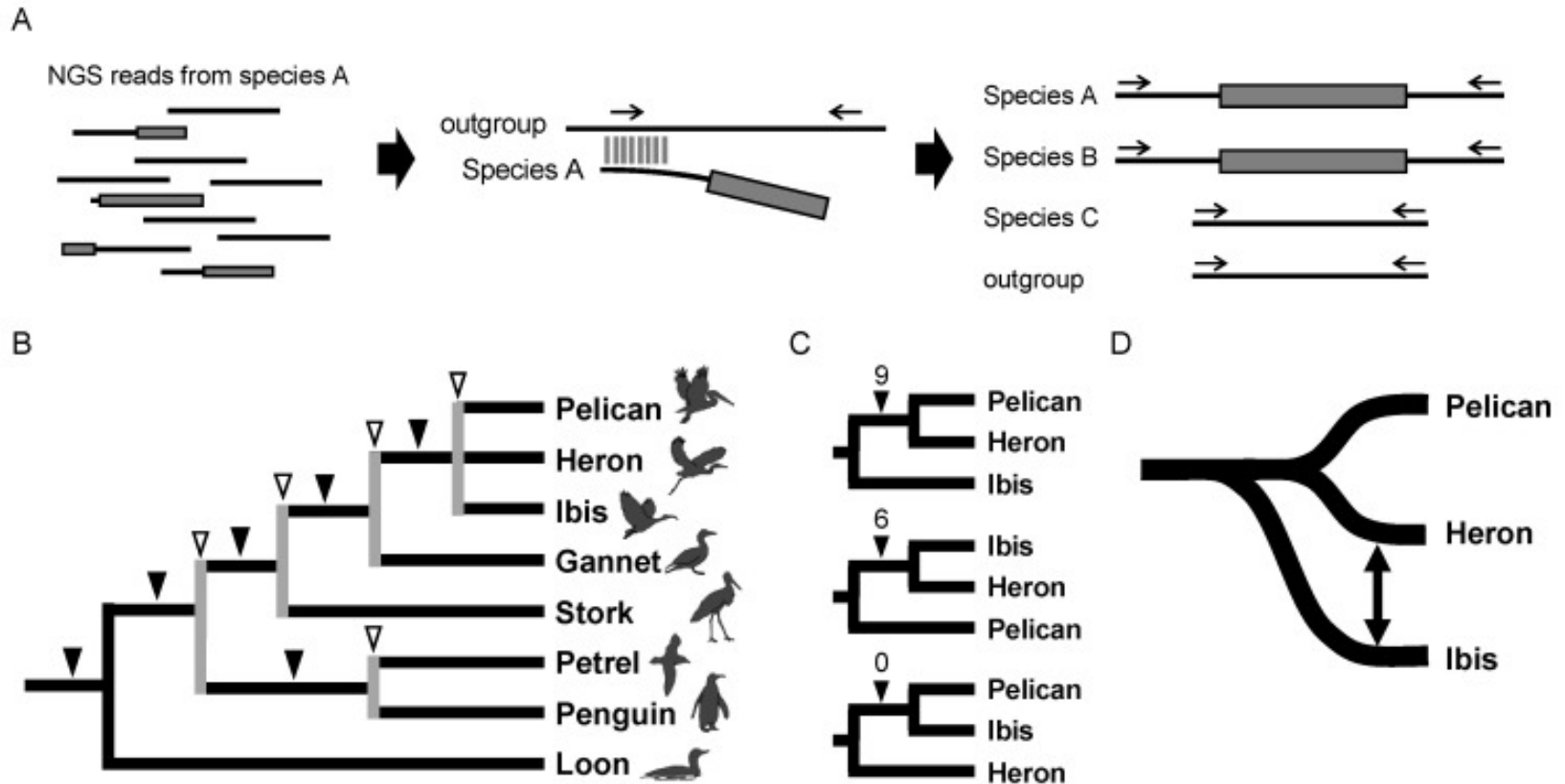


B



- 80 Mya
- ILS (incomplete lineage sorting) – nesoulad mezi lokusy
- předek placentálních savců se rozdělil na tři linie simultánně během krátké doby

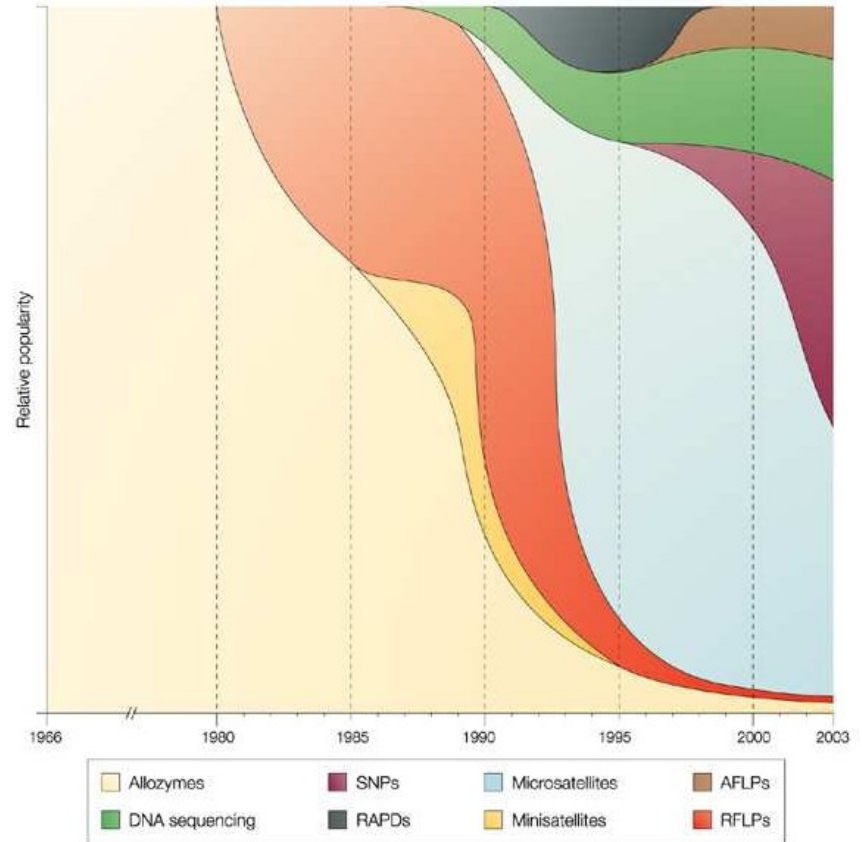
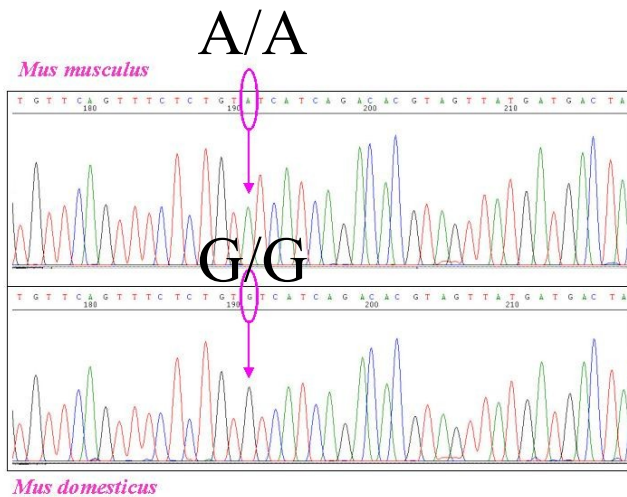
# STRONG (Screening of Transposons Obtained by Next-Generation Sequencing)



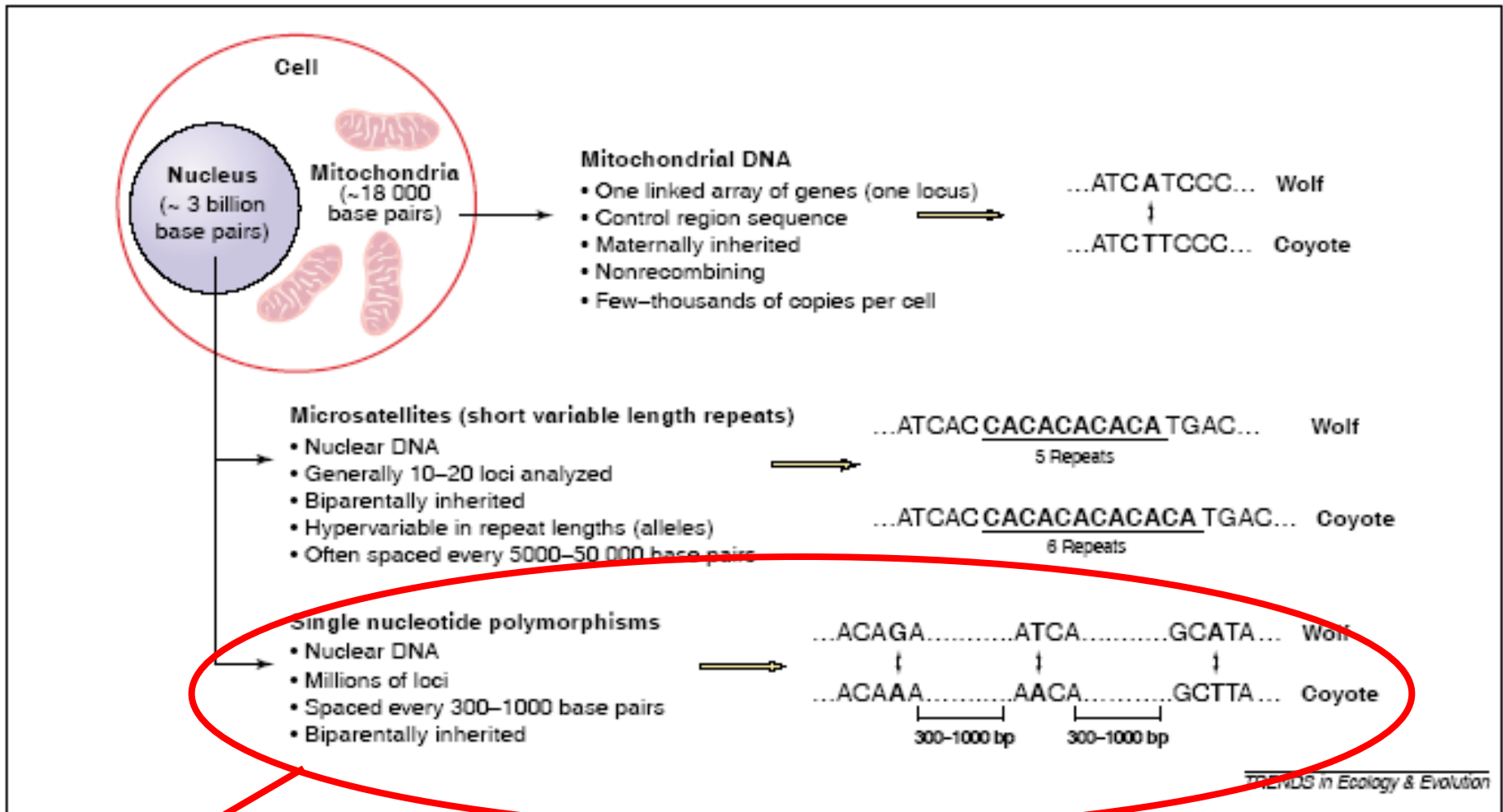
- bílé šipky - ILS na základě dvou SINE
- pelikáni a volavky jsou blízce příbuzní, ale zároveň probíhal tok genů mezi předky volavek a ibisů



# 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

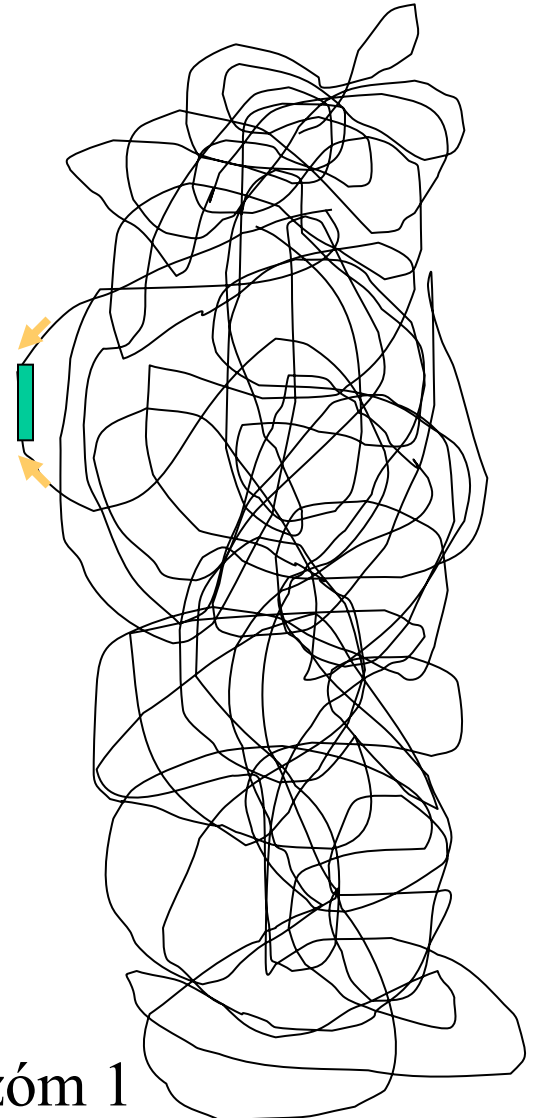
CA**T**GTA  
TG**C**ACG

**A/T**

CA**A**GTA  
TG**G**ACG

CA**A**GTA  
TG**G**ACG

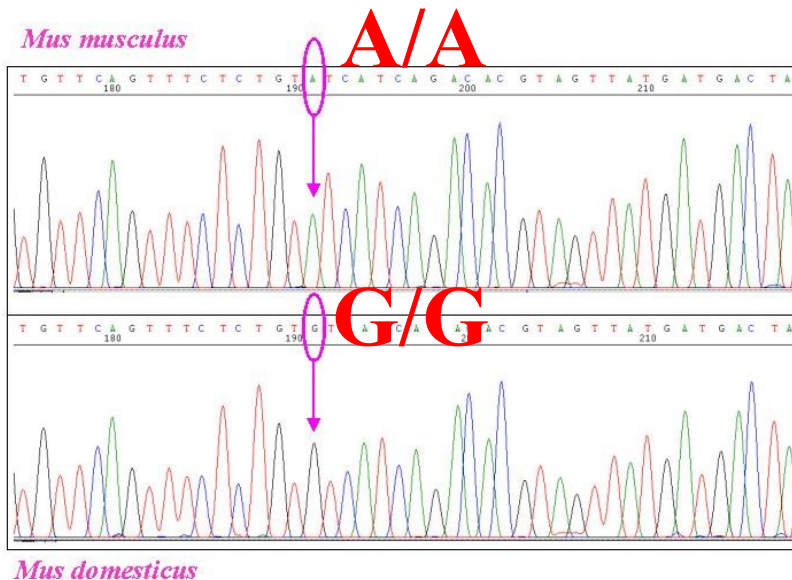
**A/A**



Př.: chromozóm 1

# Příklad diagnostického 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

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 na fenotyp
- 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 (mezi blízce příbuznými druhy)
- Mendelovská dědičnost (vs. mtDNA)
- evoluce je dobře popsitelná relativně jednoduchým mutačním modelem (vs. microsatelity)
- jsou analyzovány kratší fragmenty DNA – neinvazivní genetika

# Nevýhody

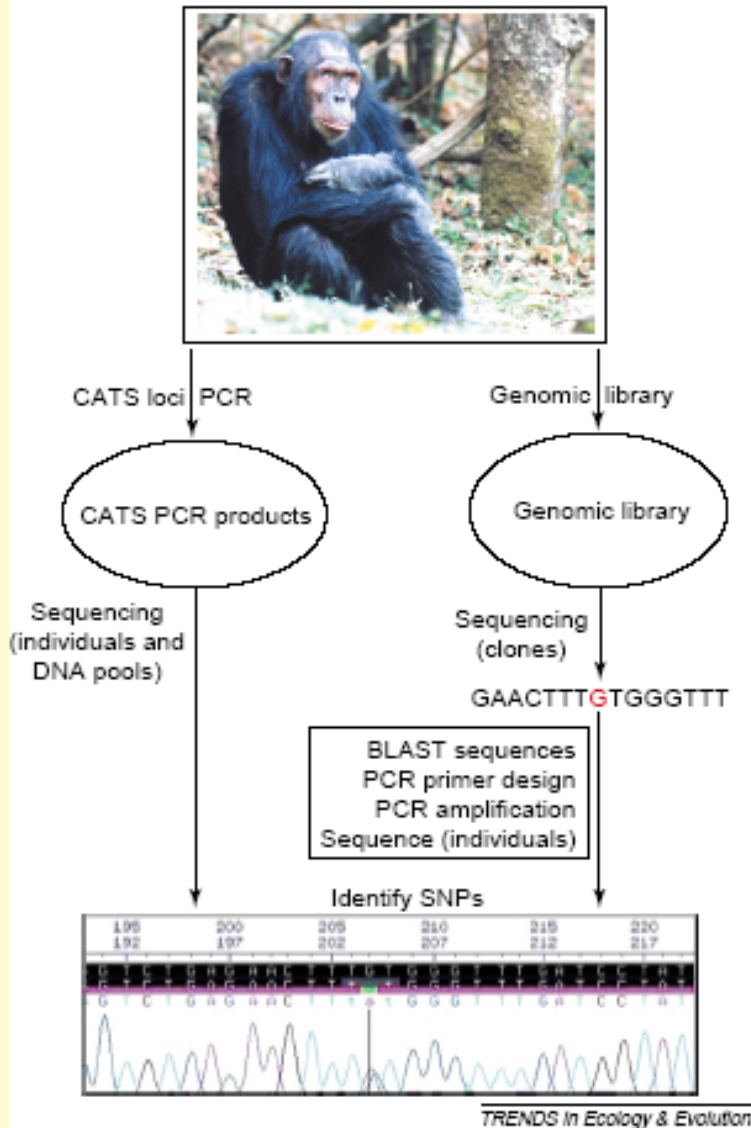
- „ascertainment bias“ – výběr informativních znaků se provádí na základě jen malého počtu jedinců a nemusí být reprezentativní
- nízká variabilita na lokus (většinou jen 2 alely)
- pro populační genetiku je vyžadován větší počet lokusů (4-10 krát více než u mikrosatelitů)

# Metody analýzy

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



# 1. Nalezení SNPs



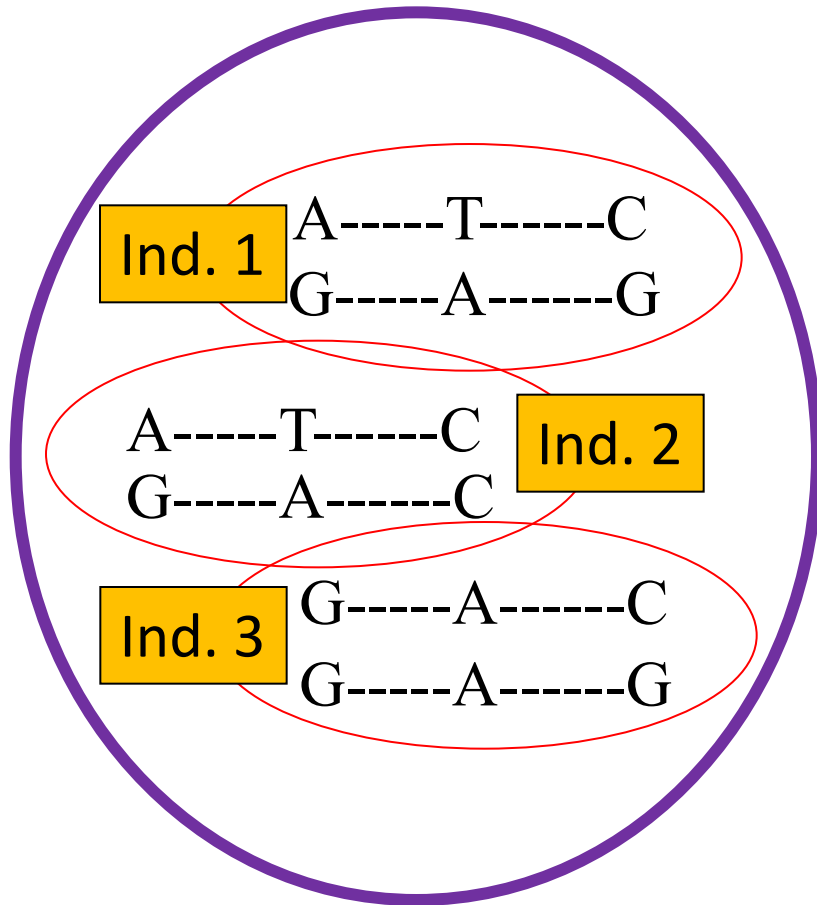
(1) CATS loci = comparative anchor tagged site loci (= cross amplification)

(2) Genomická knihovna = naštěpení genomu + klonování (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)**

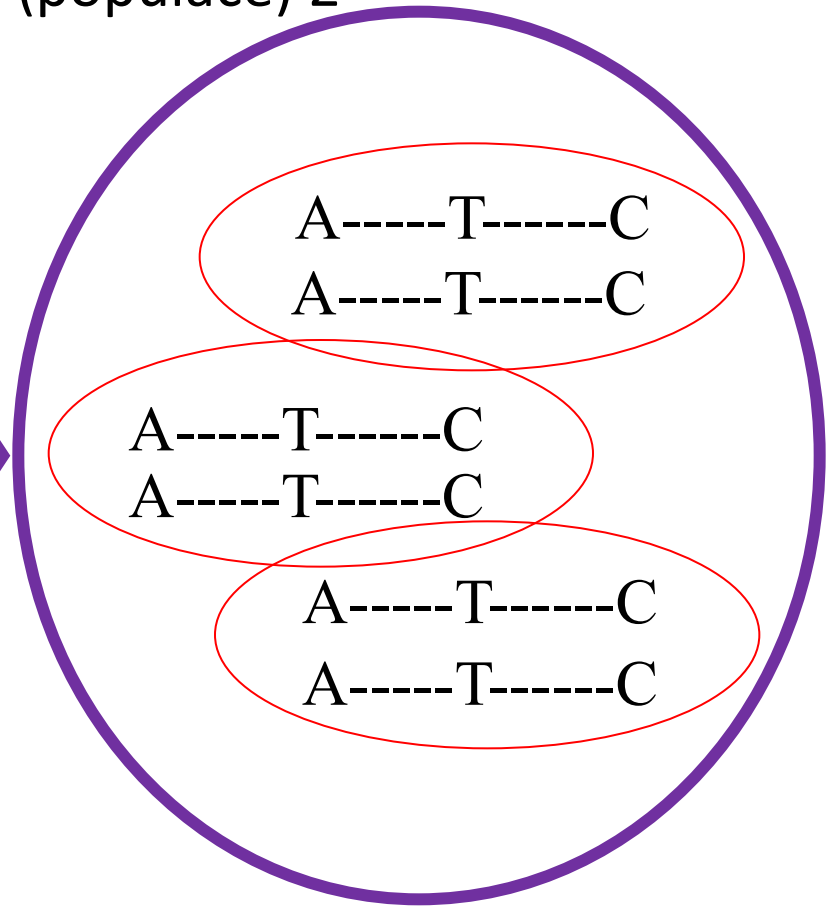
# Ascertainment bias

Druh (populace) 1



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

Druh (populace) 2



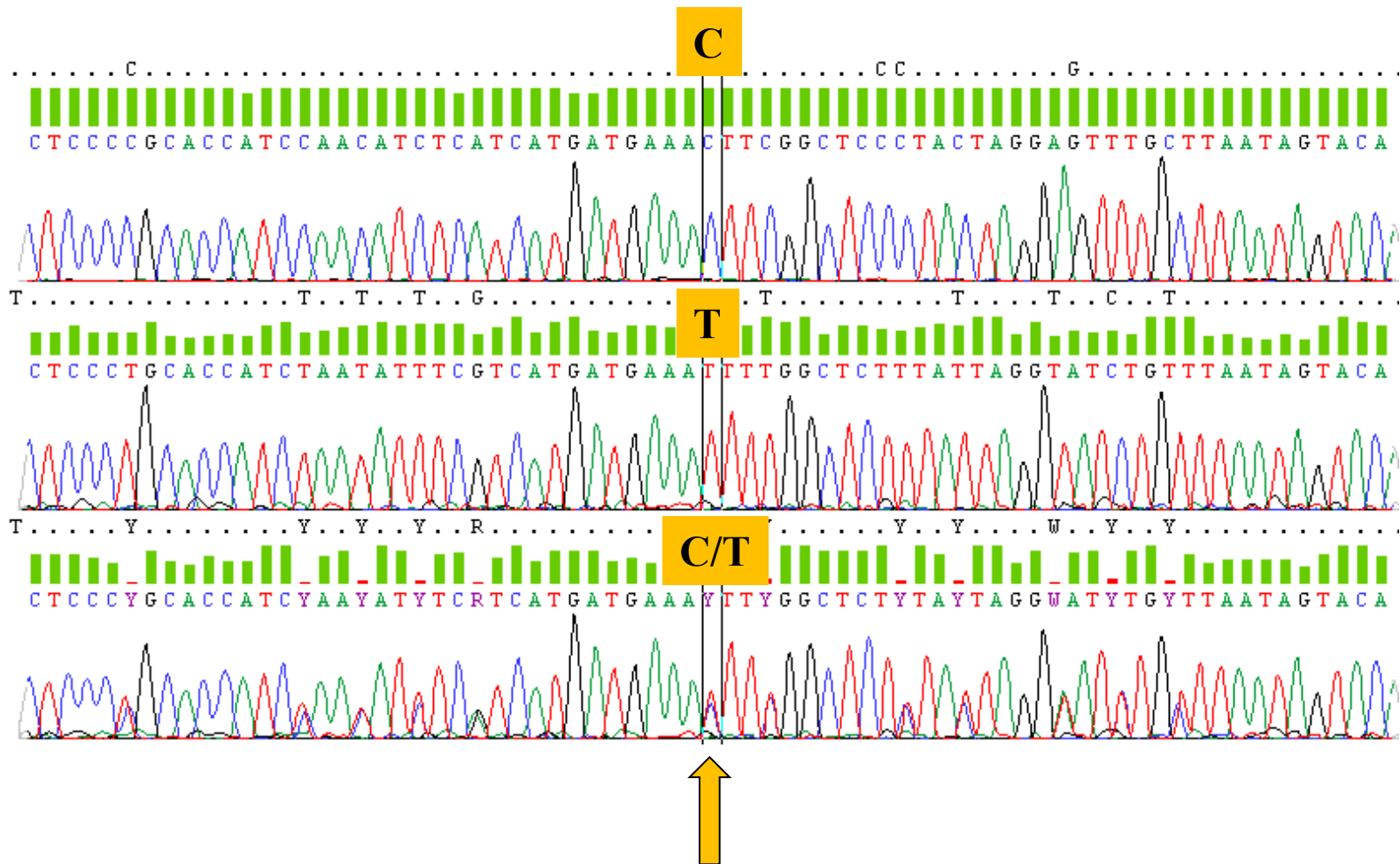
Polymorfismus daných SNPs je druhově  
(populačně) specifický

## 2. SNPs genotyping

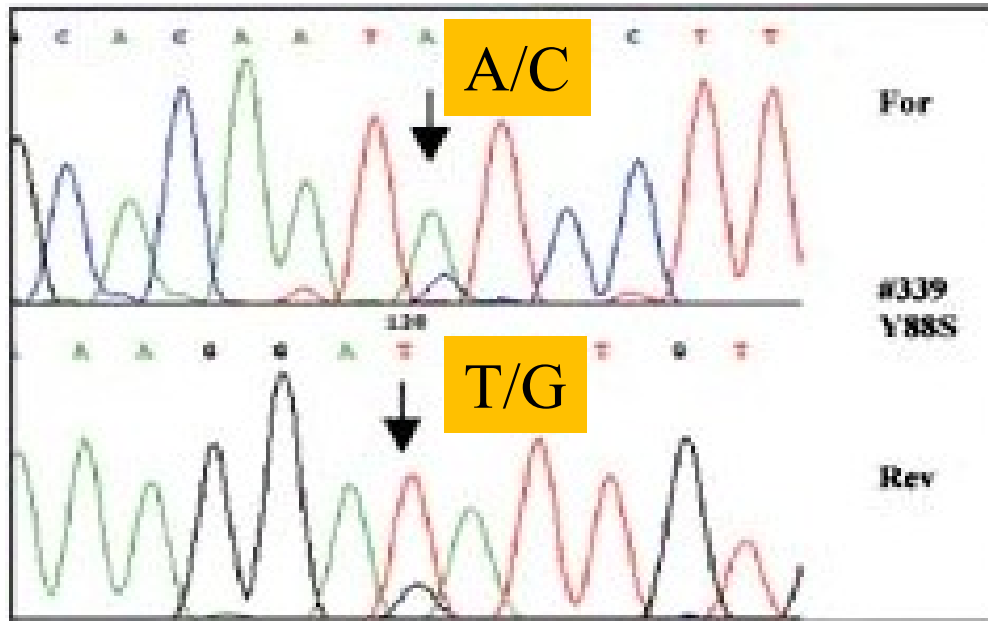
= zjištění genotypu daného jedince

# SNPs genotyping - sekvenování?

## Je drahé a nejasné u heterozygotů



# Heterozygoti?



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

**!!! klonování – cca 800 Kč**

**!!! sekvenování 1 klonu – cca 100 Kč**

PCR product

↑  
ligation, transformation

↓  
sekvenování insertů



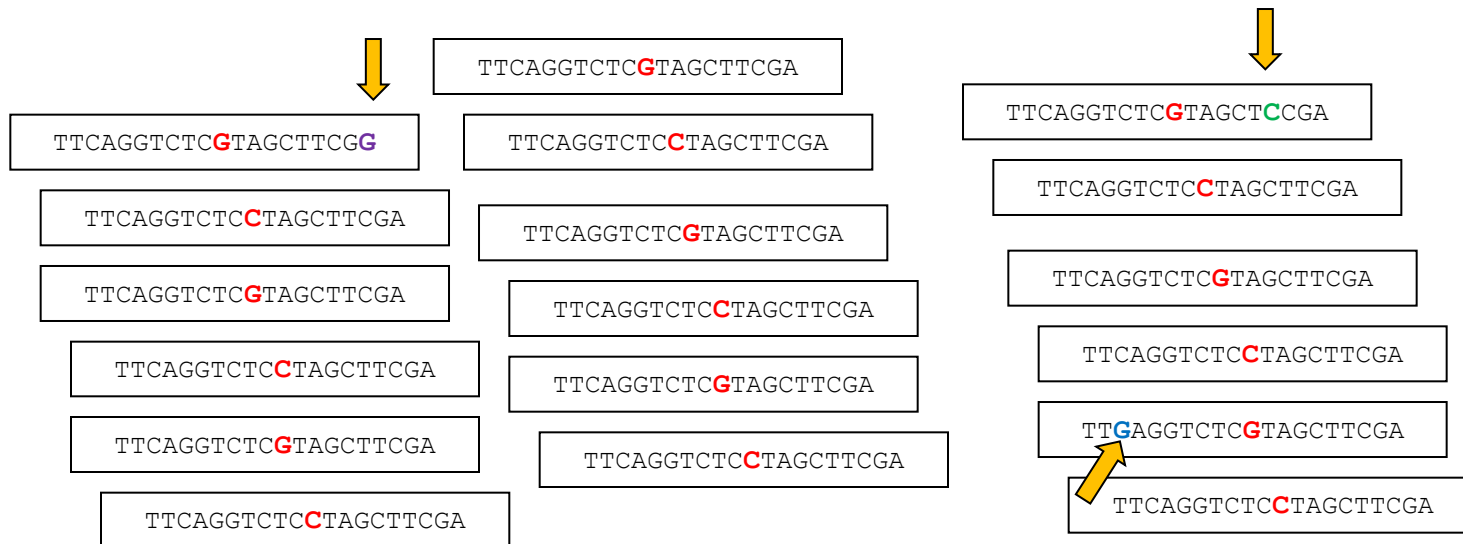
Ex.: heterozygote = two diff. alleles

# PCR vytváří substituční chyby, které jsou vizualizovány klonováním (!)

TTCAGGTCTC**G**TAGCTTCGA

TTCAGGTCTC**C**TAGCTTCGA

... před PCR = heterozygot G/C



## PCR artefakty

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

# SNPs genotypizace

## 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“ (nejsou založeny na standardní 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 genotypizace - „old standards“

„Jelenovi pivo nelej“

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

## PCR-RFLP

(restriction fragments length polymorphism)

Restrikční místa

- každá restrikční endonukleáza štěpí DNA na specifické sekvenci = **restrikční místo**
- enzymy rozpoznávají 4-6 bp dlouhou **palindromatickou sekvenci** (eg GAATTC)

**Restriction site**

**Palindrome**

G T A G G A A T T C A 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

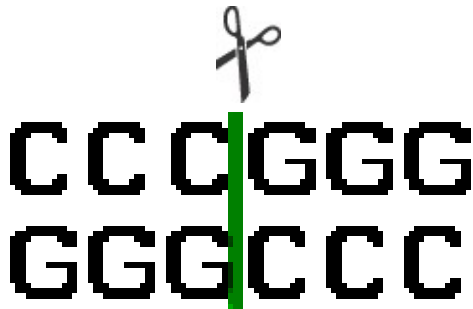
G T A G                      A A T T C A T T C A C G C A  
C A T C T T A A                      G T A A G T G C G T

**Fragment 1**

**Fragment 2**

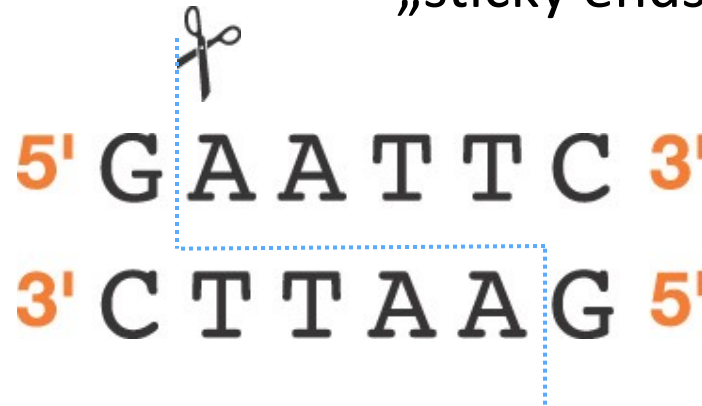
# Běžné restrikční enzymy

„blunt ends“



**SmaI**  
- blunt end

„sticky ends“



**EcoRI**  
- Escherichia coli  
- 5 prime overhang



**PstI**  
- Providencia stuartii  
- 3 prime overhang

# SNP genotypizace - „old standards“

## PCR-RFLP

### Alela A

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

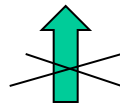


štěpení restriční endonukleázou

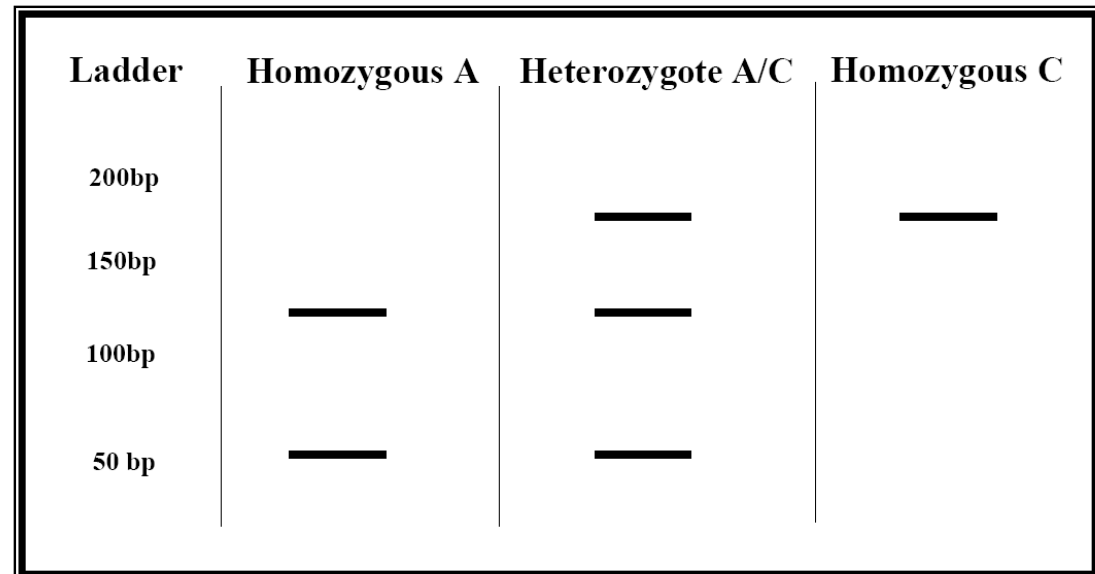
- neumožní nalézt novou variantu daného SNP (odliší pouze 2 formy daného znaku: +/- )

### Alela C

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



neštěpí



# SNPs genotyping - „old standards“ electroforetické metody detekce mutací

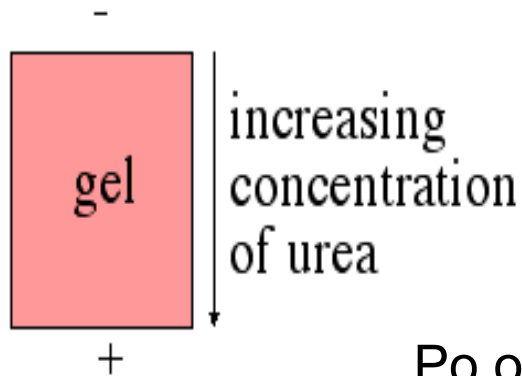
- Thermal gradient gel electrophoresis (**TGGE**)
- Denaturing gradient gel electrophoresis (**DGGE**)
- Single-strand conformation polymorphism (**SSCP**)

= specifické elektroforetické metody založené na odlišné mobilitě stejně dlouhých fragmentů s různou DNA sekvencí

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

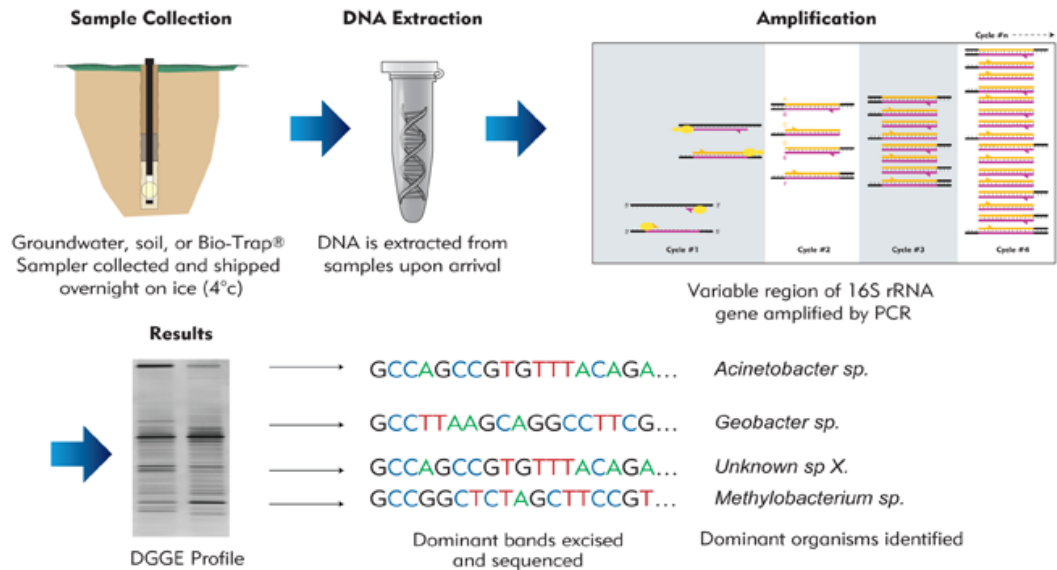
**Krátké PCR fragmenty (200-700 bp) jsou separovány v denaturačním gradientu (PAGE = polyakrylamidový gel)**

→ 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

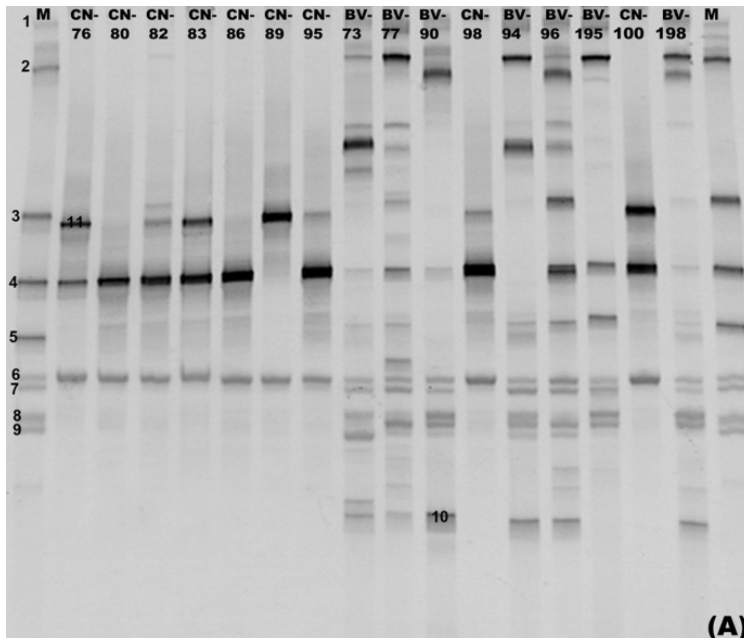
Po obarvení lze vidět rozdílné pozice PCR produktů v závislosti na jejich sekvenci



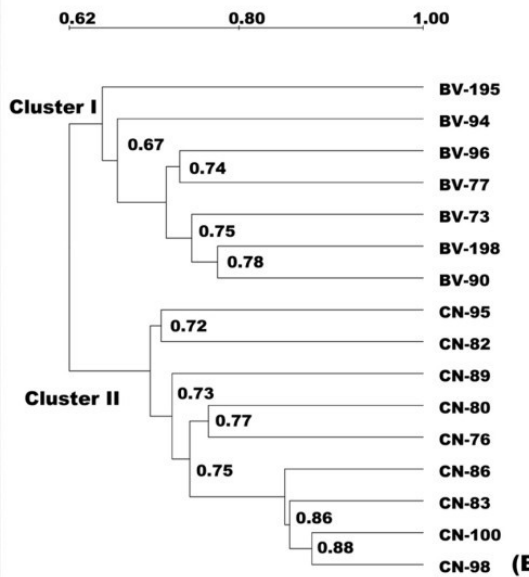
# DGGE v bakteriální metagenomice



Dnes rychle nahrazováno HTS („high-throughput sequencing“)



Dendrogram using UPGMA (Dice Coefficient)

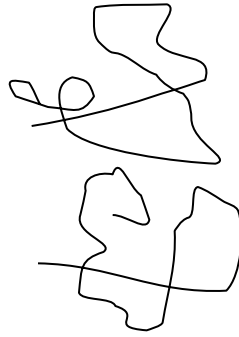


# Single strand conformation polymorphism (SSCP)

## Allele 1 - C

...CGCTT**C**AGG ...

...GCGAA**G**TCC...



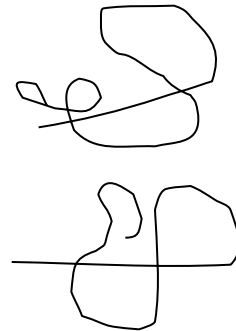
heating - denaturation

snap-cooling → partial renaturation

## Allele 2 - A

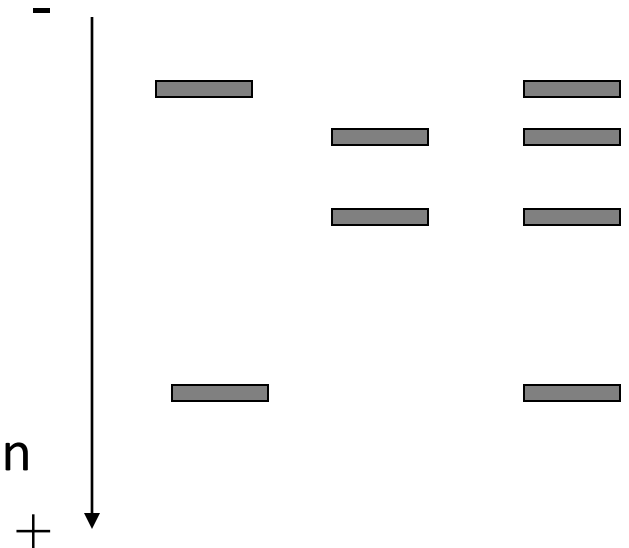
...CGCTT**A**AGG ...

...GCGAA**T**TCC...



sequence-specific  
ssDNA conformations

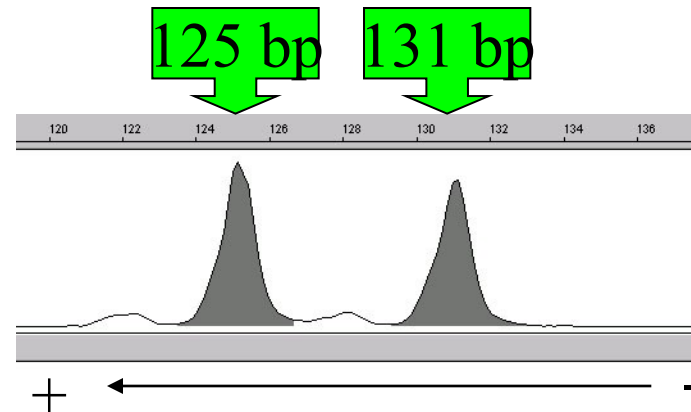
Homo1 Homo2 Hetero



**!!! non-denaturing PAGE**

radioisotopes  
silver-staining  
fluorescent dyes (SYBR gold)

# Použití automatických sekvenátorů (denaturující polymer POP7 - ssDNA, např. mikrosatelity - jeden značený primer)



Vysoká citlivost, dobře kontrolovatelné parametry elektroforézy



# Použití automatických sekvenátorů

Proč nevyužít nedenaturující elektroforézu?

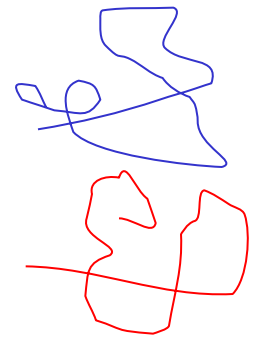
e.g. CAP (conformation analysis polymer)



- well controlled electrophoresis
- two fluorescent labels
- high sensitivity

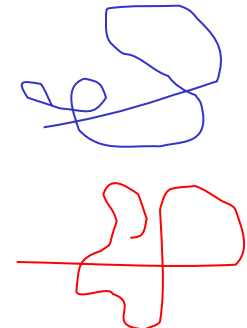
## Allele 1

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



## Allele 2

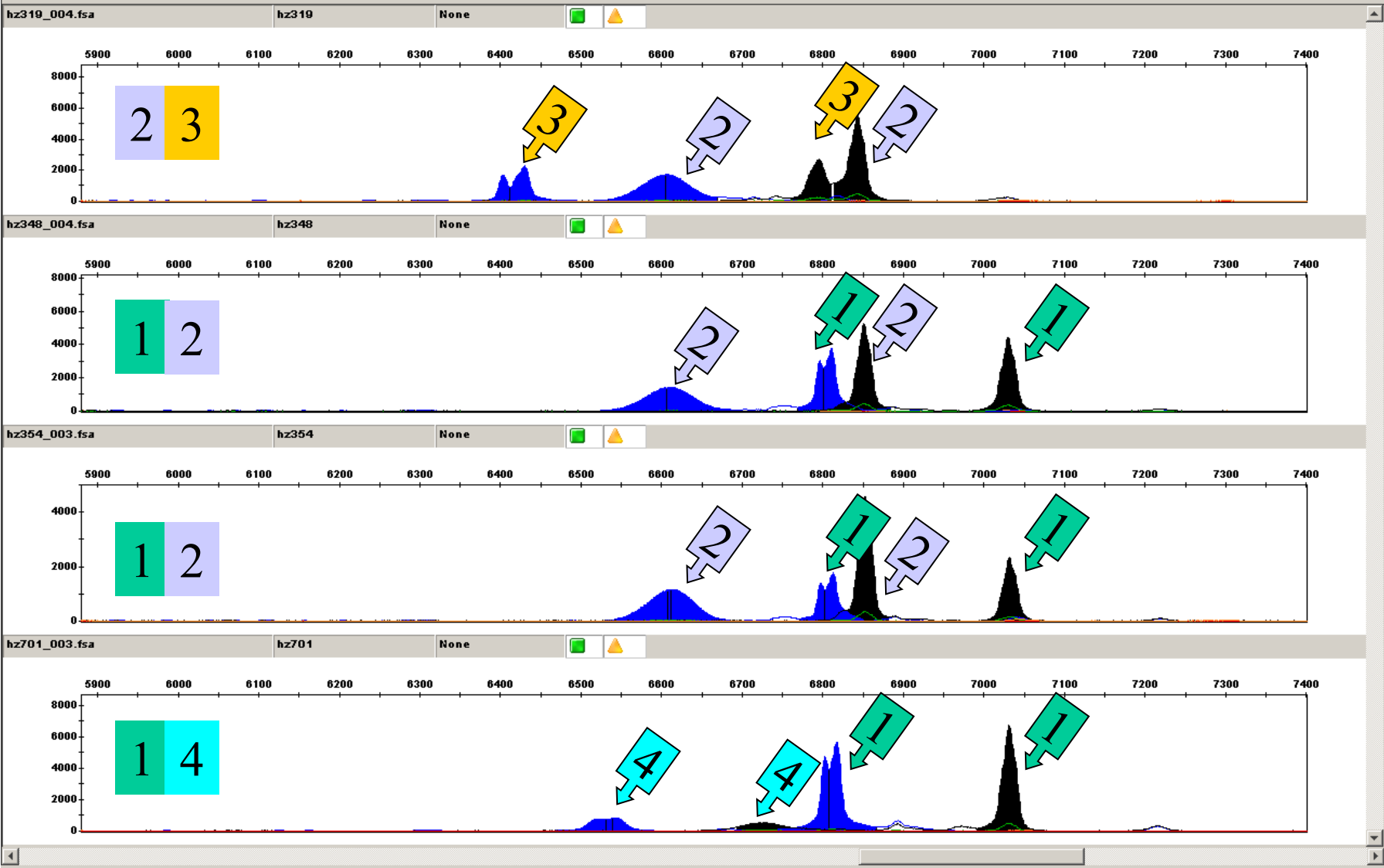
*FAM*... CGCTTAAGG ...  
... GCGAATTCC ...*HEX*



# 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

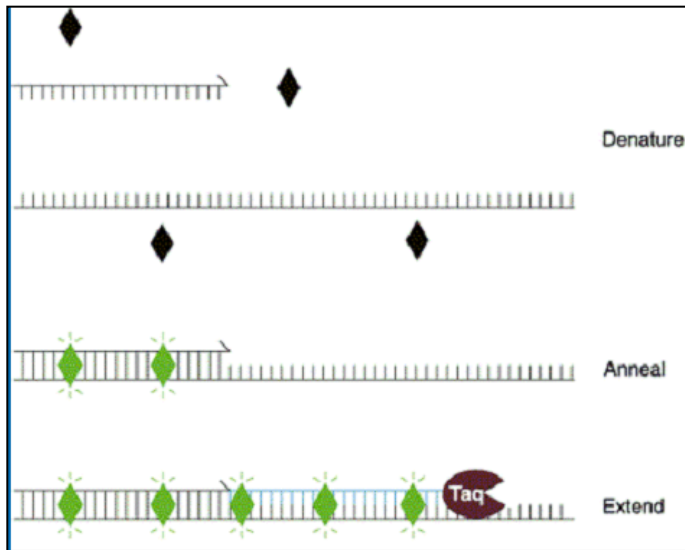
- např. GeneMapper (Applied Biosystems)
- specifický „Size+Conformation Standard“ pro každou teplotu (konformace závisí na teplotě)
- srovnání více vzorků
- umožňuje detekci krátkých odlišných sekvencí s více SNPs (užitečné např. pro genotypizaci MHC, tj. vysoce variabilních genů)
- opět rychle nahrazováno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

# SNP genotypizace - „new methods“

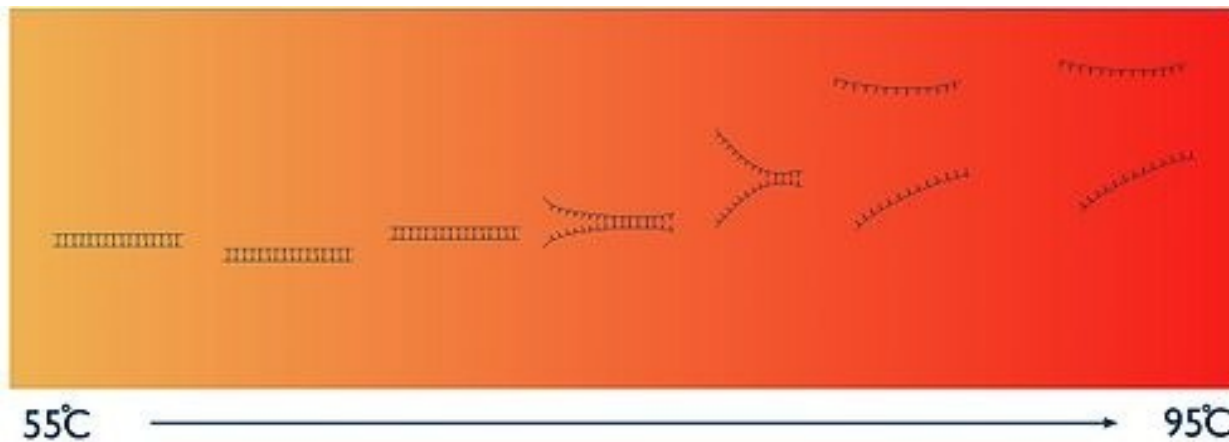
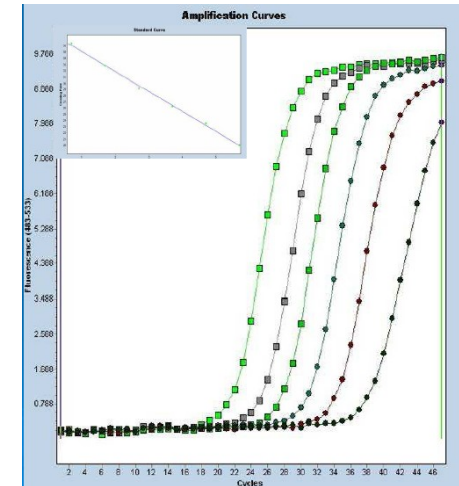
= nejsou založeny na standardní PCR

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)

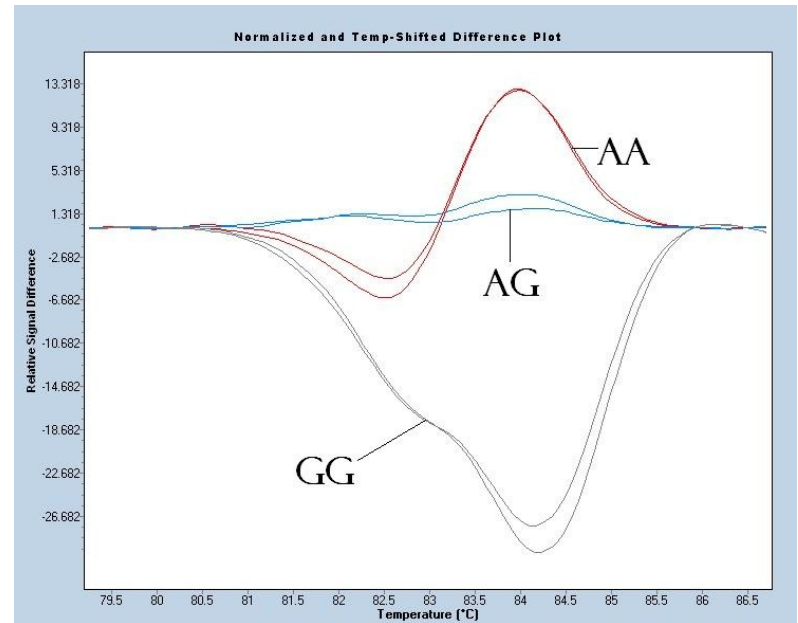
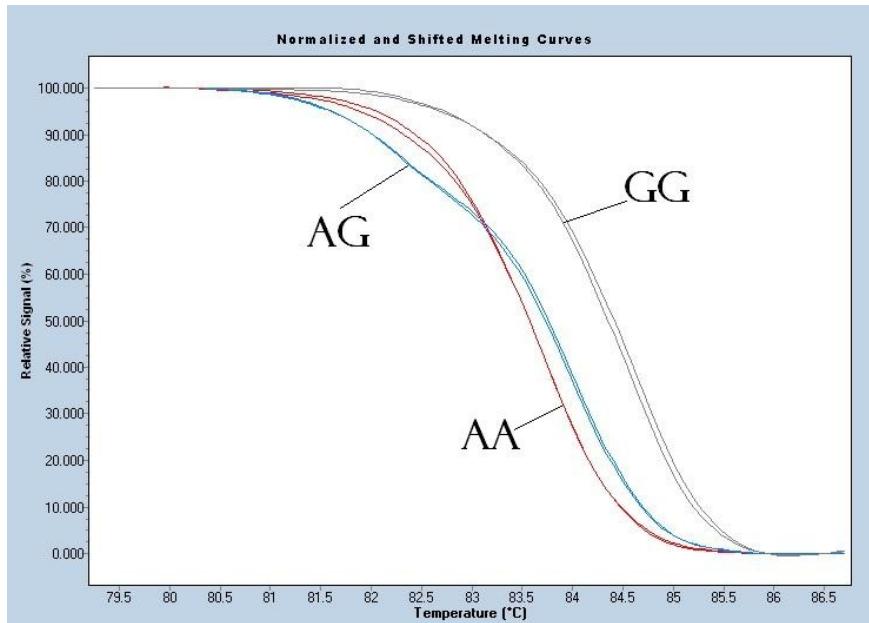


Krok 1: real-time PCR = nárůst fluorescence



Krok 2: měření rychlosti denaturace po PCR = pokles fluorescence (závislý na konkrétní sekvenci)

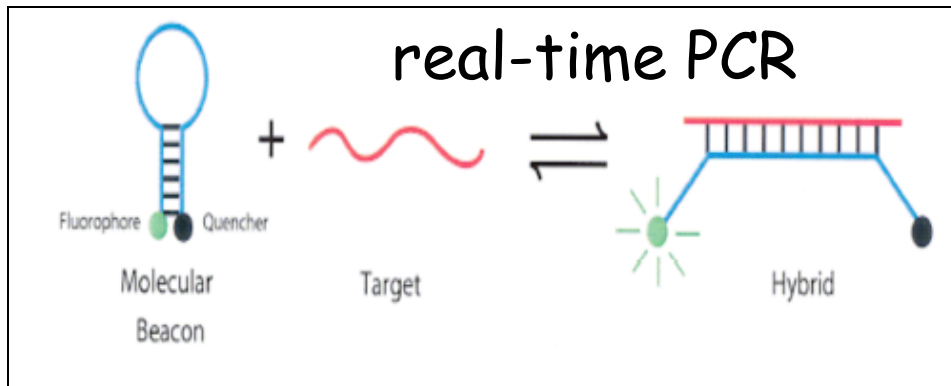
# HRMT genotypizace



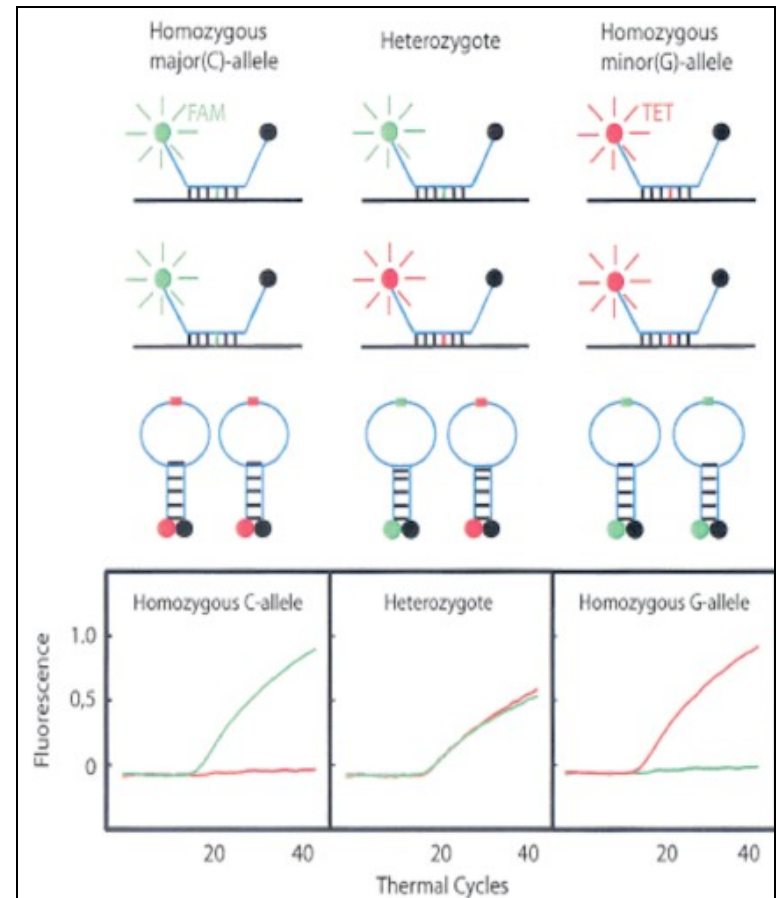
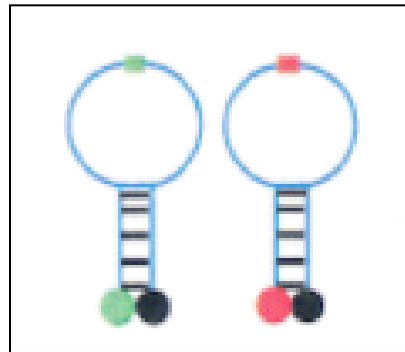
## Detekce heterozygotů

- 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



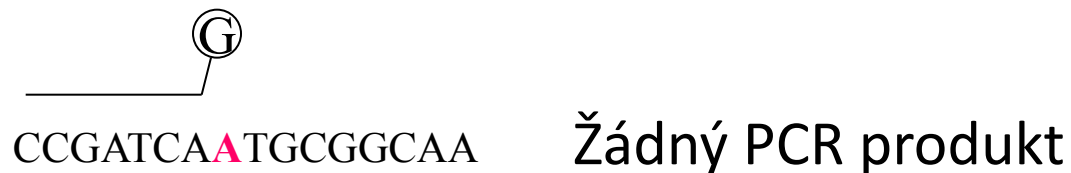
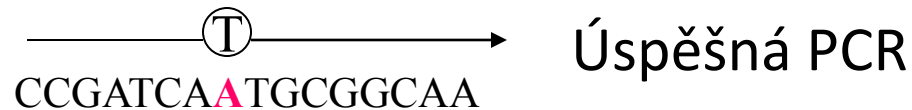
sondy  
specifické pro  
jednotlivé alely



1) TaqMan sondy

2) Molecular Beacons („maják“)

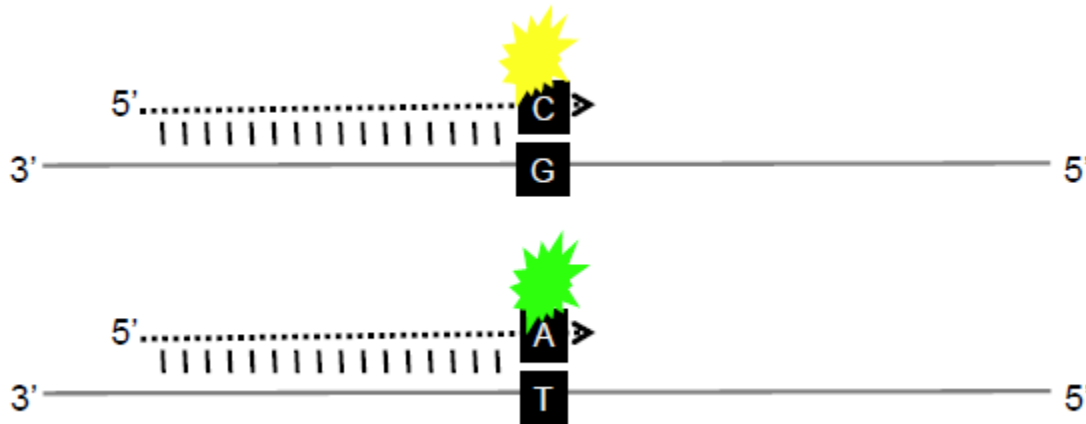
### 3. ASPE: allele-specific primer extension



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

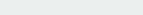
# Kompetitive Allele Specific PCR

## 1) Assay components:

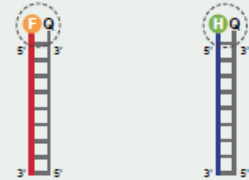
### A) Primer mix

Allele specific forward primers:

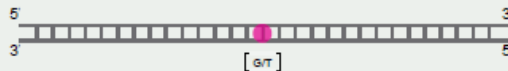


Reverse primer:  
3'  5'

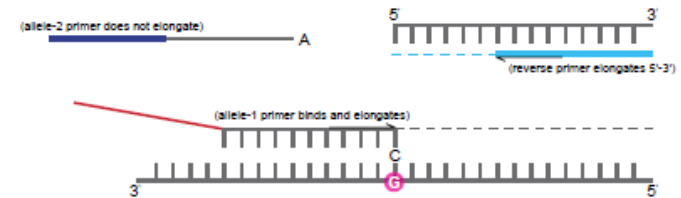
### B) Master mix



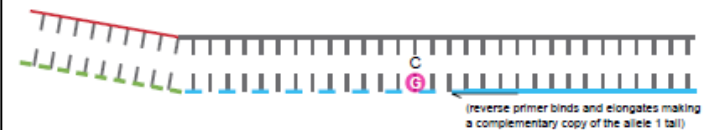
### C) DNA template (sample)



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



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

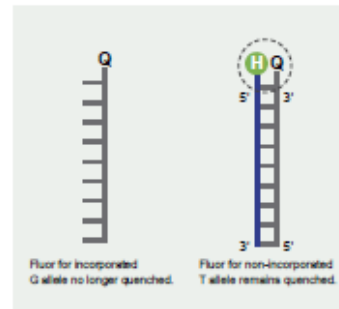


## 4) Signal generation – PCR round 3:








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

# The KASP™ genotyping assay from LGC Genomics



## Cena analýzy (2020)

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

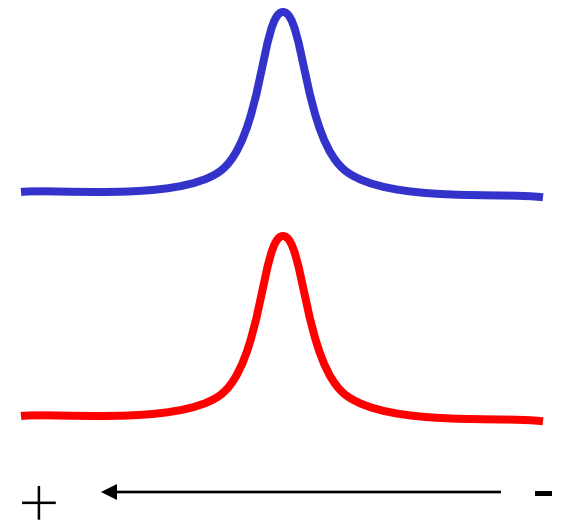
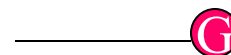
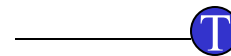
### Single nucleotide polymorphisms

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
```

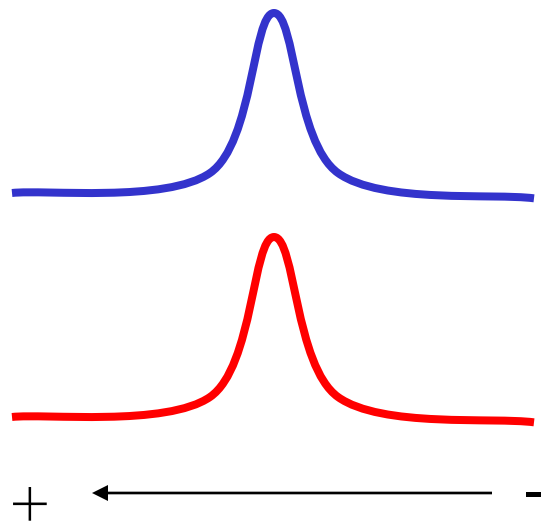
Figure 1. Examples of how to format sequence for KASP assay design for a SNP

# 4. SBE: single base extension

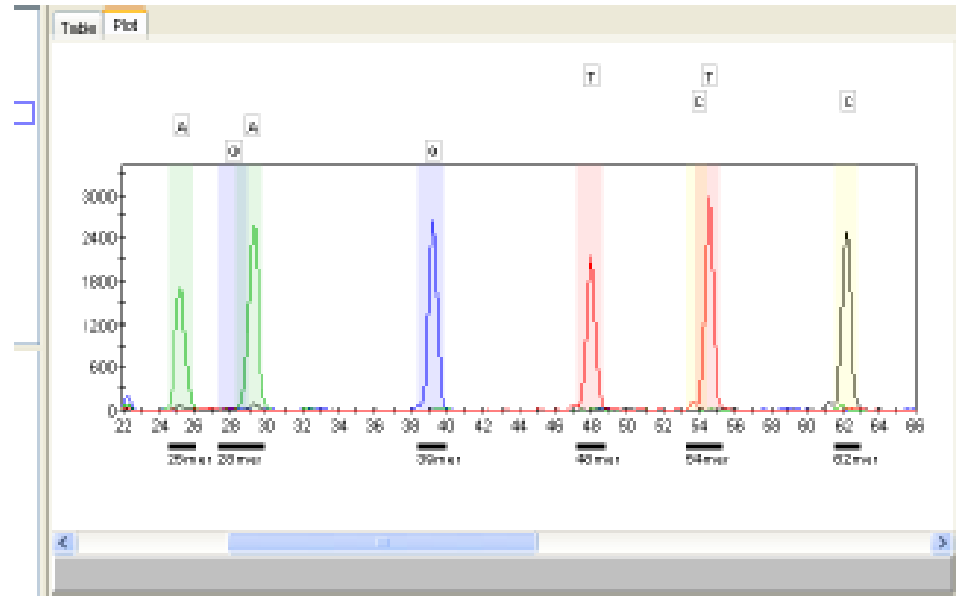


- pouze jeden dideoxynukleotid je přidán k primeru
- detekce různými metodami

# (A) Detekce SBE produktů kapilární elektroforézou

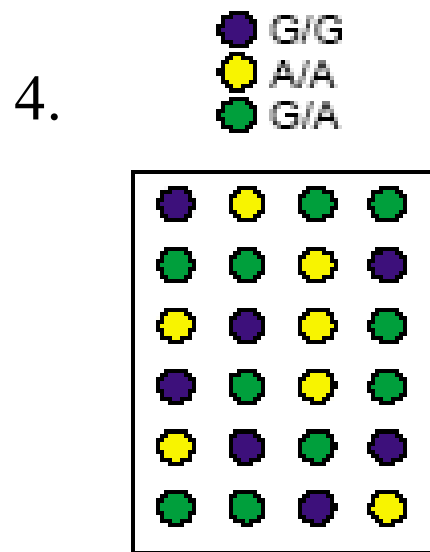
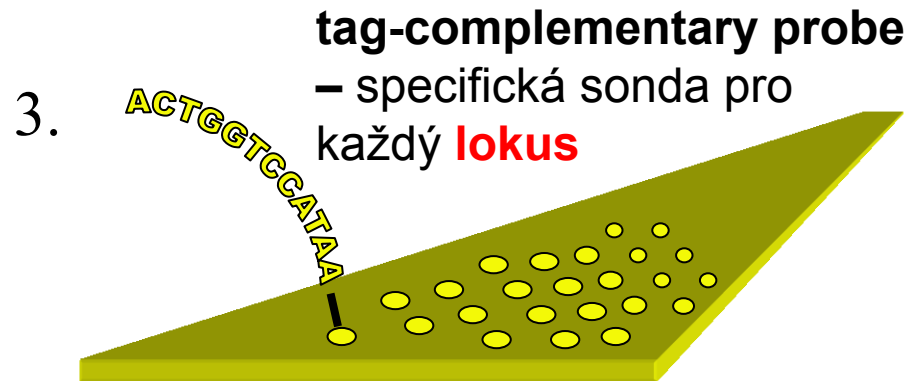
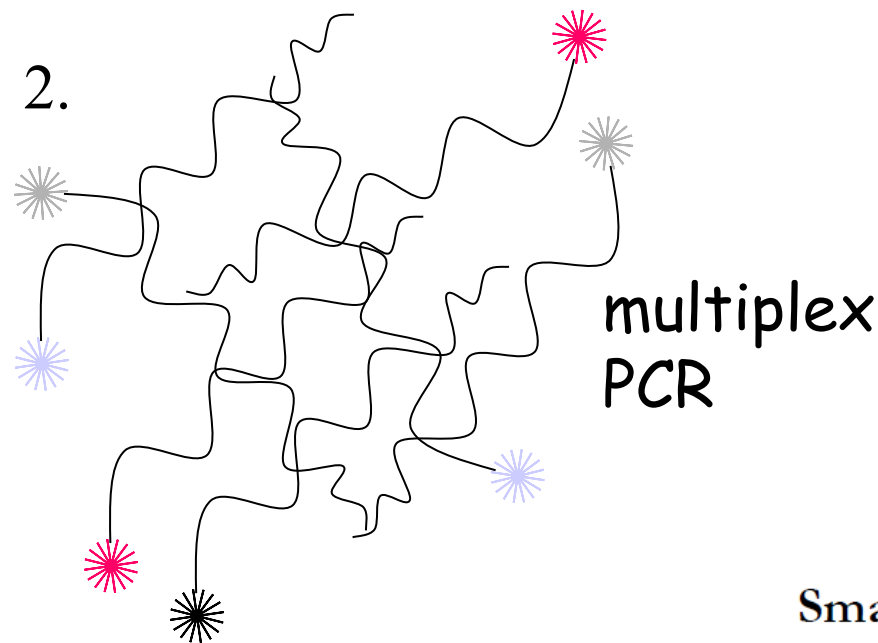
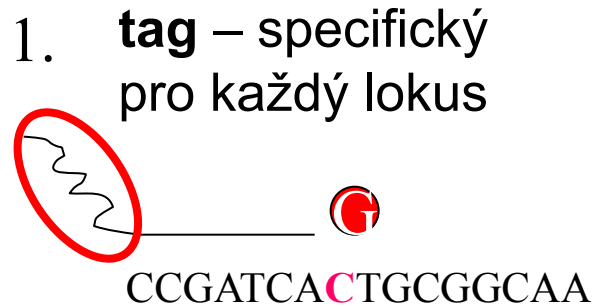


kapilární elektroforéza  
SNaPShot Multiplex Kit  
(ThermoFisher Scientific)



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

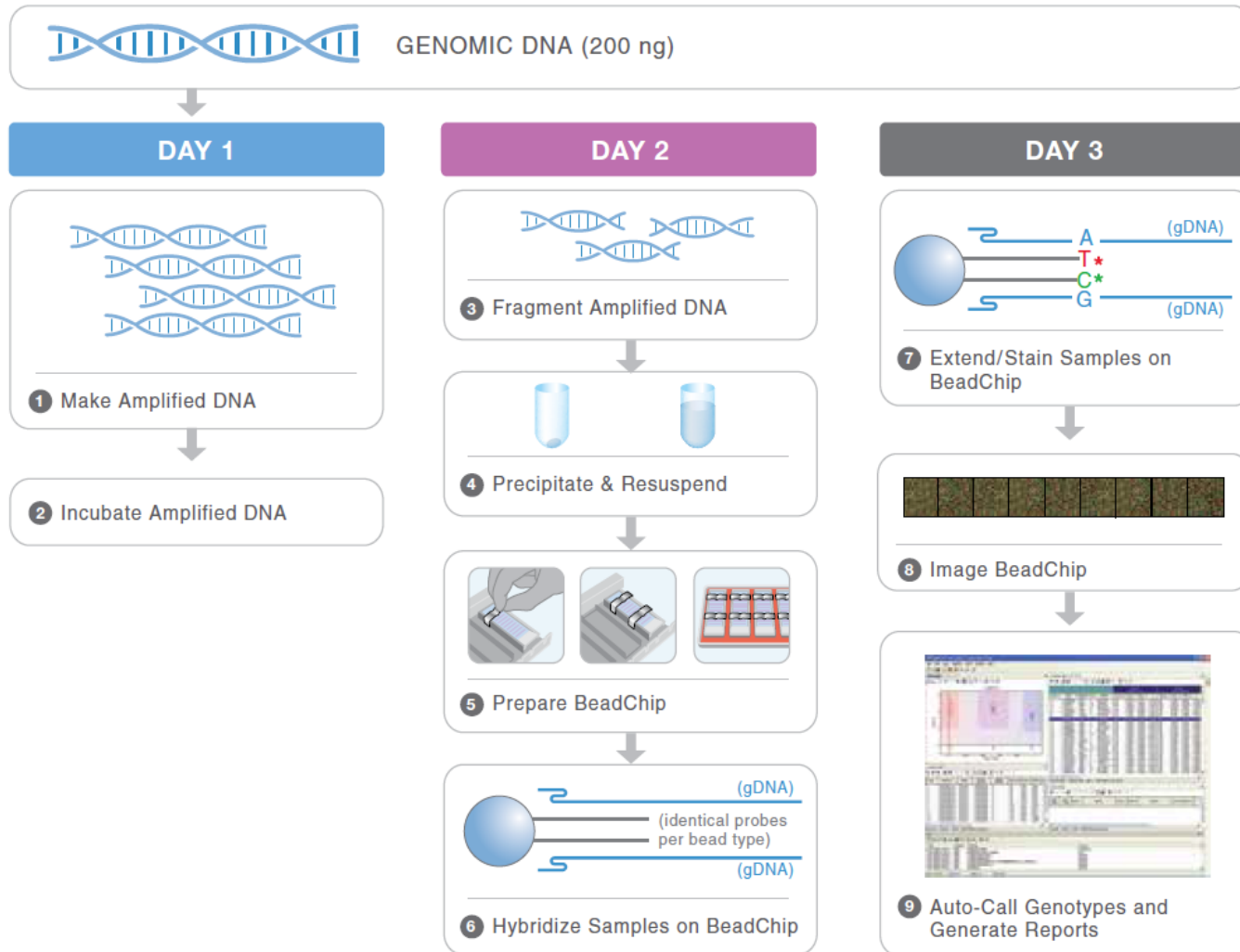
# (B) Detekce SBE produktů přes „microarray“ (tj. hybridizace)



Small-scale “in house” SNP genotyping

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

# Illumina Infinium Bead Chip

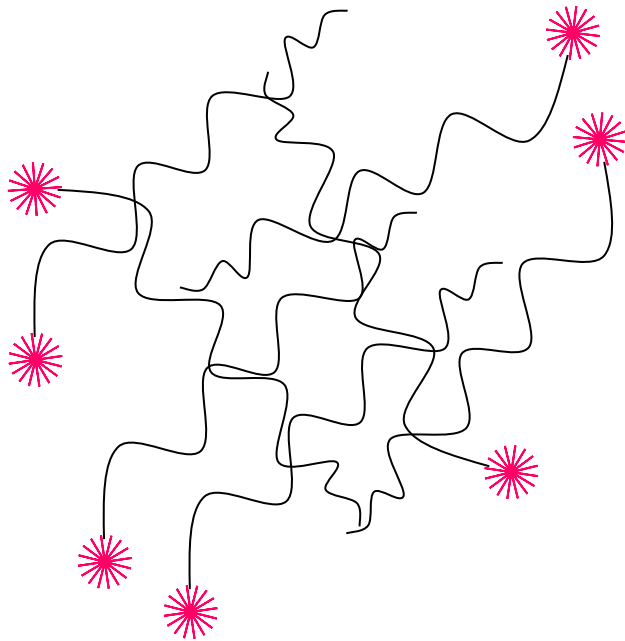


cca 300 000 SNP loci from 200 ng of DNA

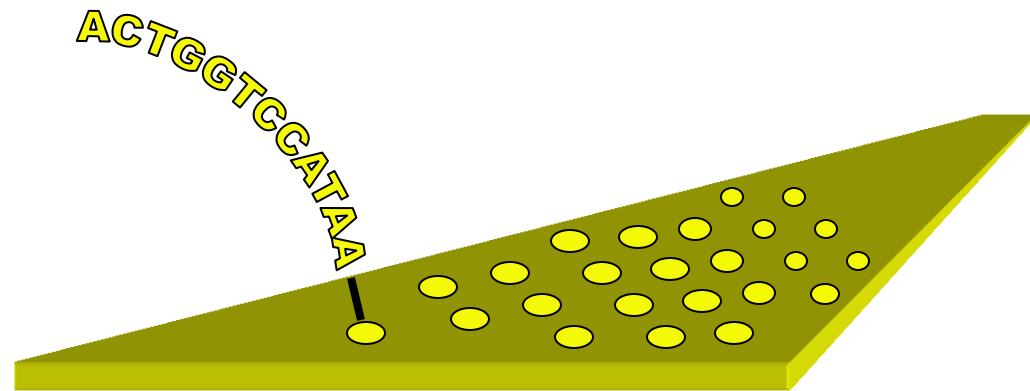
\* Indicates Stain in Red Channel  
\* Indicates Stain in Green Channel

# 5. Alelově specifická hybridizace

## Microarrays - SNPs chips



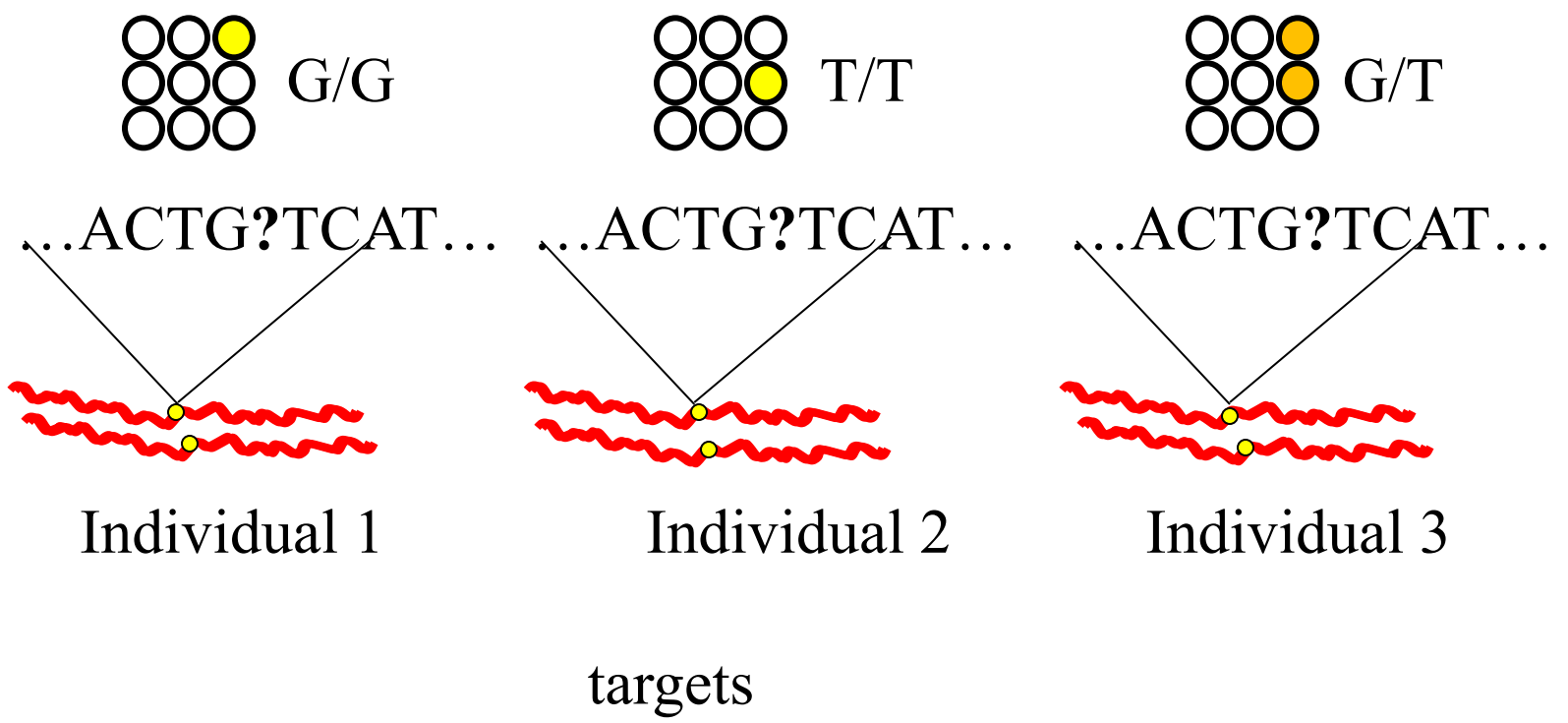
**Target** (genomická DNA  
rozštěpená restričními  
enzymy)



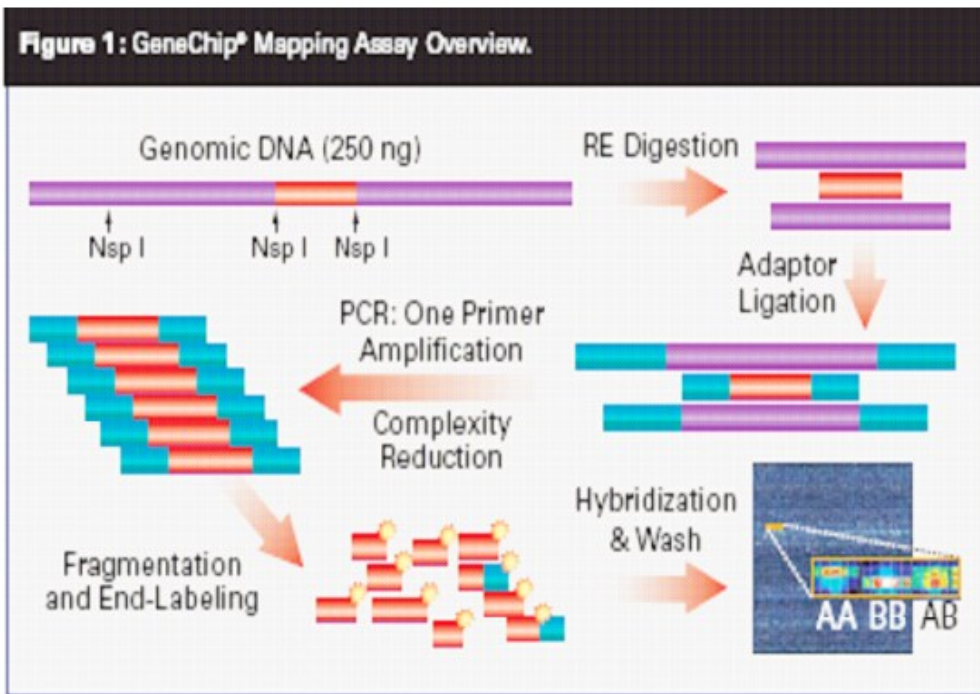
**Probe**  
(specifická sonda pro každou  
**alelu**)



# Microarray SNP Genotyping



# Detekce: např. Thermo Fisher Scientific (Affymetrix)



Affymetrix® Mouse Diversity Genotyping Array

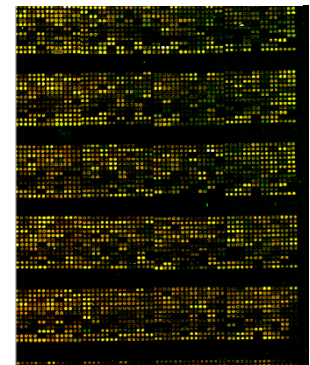


- 10 tisíc – 1 milión SNP znaků najednou – „chip technology“

- 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*



# Může být silný „ascertainment bias“



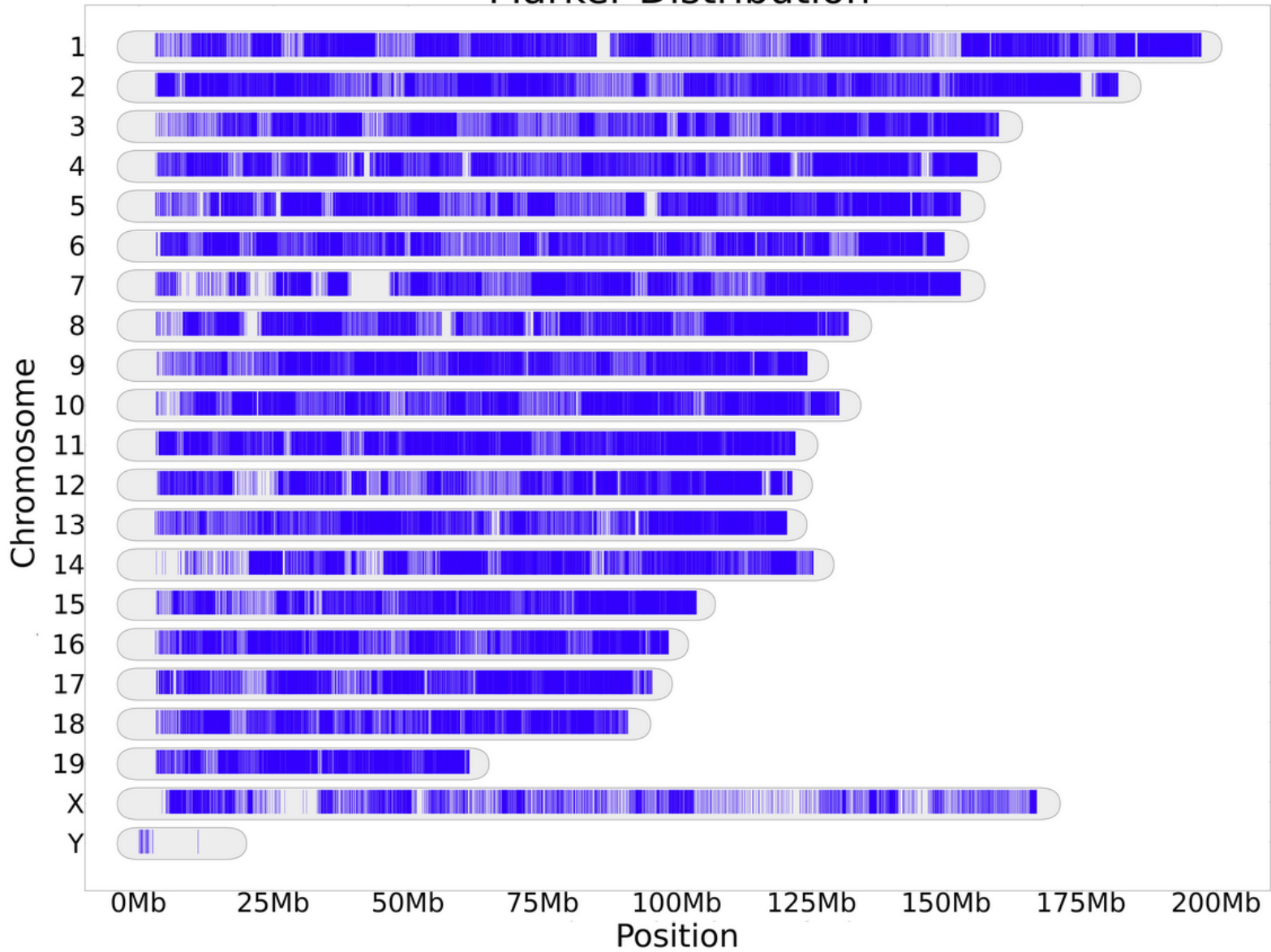
## Performance

- 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)

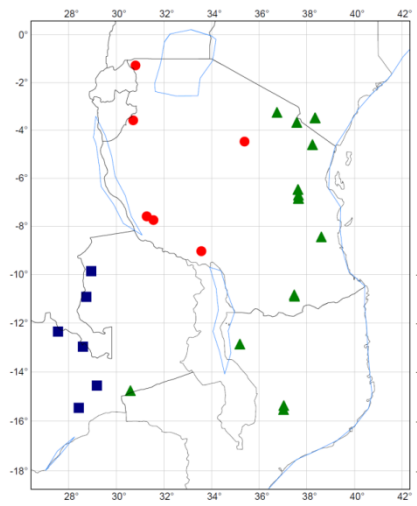
# Marker Distribution



# Použití pro nemodelové druhy („cross-genotyping“)



calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_calls	Score_Call						
MOZ27:MOZ27:VM121	VM121	RS0847	RS0847	J5052	J5052	KE55	KE55	24645	24645	T8_414	T8_414	TA452	TA452	TA453	TA453	TA171	TA171	TA536	TA536	RS1173	RS1173	
42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H	42818H
Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	: Apr 22	2014
Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1	Mega_1
930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571	930571



Studium kontaktní zóny *Mus minutoides* ve východní Africe

1620 informativních lokusů z 5598 lokusů na chromozómu 1 (60 SNPs fixovaných pro dané subpopulace v celém genomu)

jedinec 1  
jedinec 2

# 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