

## Získavanie dát molekulárne genetickými metódami, štúdium SNP polymorfizmov

Mgr. Civiánová Kristina, doc. RNDr. Knoll Aleš, PhD.

Ústav genetiky

Mendelova zemědělská a lesnická univerzita v Brně



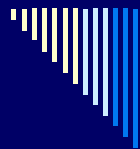
Mnohotvárnosť založená na odlišnostiach v jednotlivých „písmenách“ genetického kódu

## Single nucleotide polymorphisms (SNP)

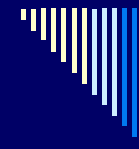
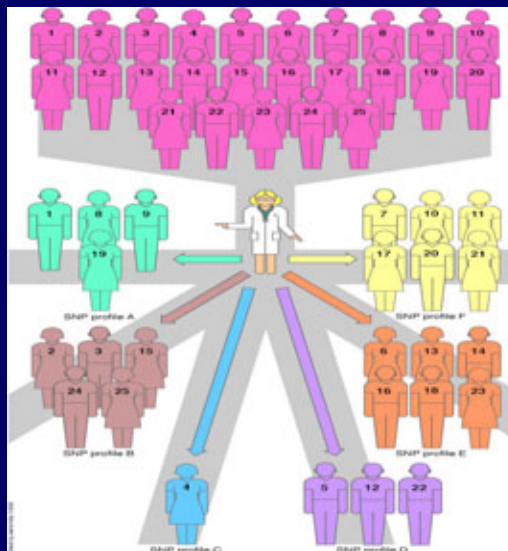
- markery III. typu
- **modifikácia 1 nukleotidu**  
intróny a intergénové oblasti,  
každých **1300 bp**  
v genóme cicavcov početné,  
geneticky stabilné
- **menšia heterozygotita, dialektické**
- **väzbové mapovanie, segregáčnej analýzy, asociačné štúdie, štúdiu variability populácií a rodín, identifikácia jedincov, overovanie rodičovstva**

## Mikrosatelity (MS)

- markery II. typu
- **krátke tandemové repetície**  
zložené z mono, di (300-500 kb)  
tri alebo tetra (3000-5000 kb)  
nukleotidových motívov
- **vysoko polymorfné** ⇒ rôzne alely (5-20)  
Např. (GC)<sub>n</sub>  
n = 5-25 opakovaní  
vysoká informatívnosť  
**multialelické**
- **väzbové mapovanie génov, štúdiu diverzity, určovanie paternity a parentity**

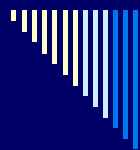


Mnohotvárnosť založená na odlišnostiach v jednotlivých „písmenách“ genetického kódu



## Nejpoužívanéjšie metódy analýzy polymorfizmov

- a) Separáčnej metódy (Elfo)
- b) Denaturácia a hybridizácia
- c) PCR
- d) **Detekcia mikrosatelitov**
- e) **PCR-RFLP**
- f) SSCP
- g) DGGE
- h) **AFLP**
- i) RAPD
- j) **Sekvenovanie / SNaPshot**
- k) Pyrosekvenovanie
- l) **PCR v reálnom čase (real-time PCR)**
- m) **DNA array (biočipy)**
- n) **Aktívne biočipy**



## Metodiky detekcie a genotypizácie SNP

- ☐ vysoko automatizované metódy
- ☐ široké spektrum:

### PCR-RFLP

#### Priame sekvenovanie

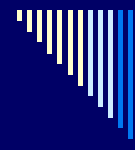
(ABI Prism 3100 Avant Genetic Analyser)

#### SNaPshot- minisekvenovanie

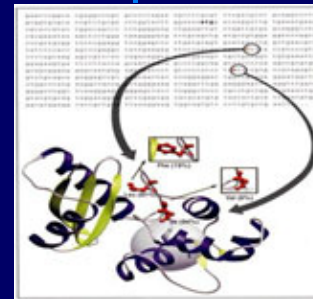
(SNaPshot Multiplex Kit)



DNA pooling  
Multiplex  
PCR



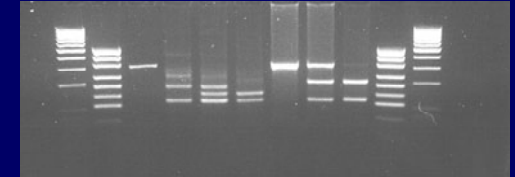
## PCR – RFLP



+/+ 5' GCC TGG GTG **G** ACC ATC CTG 3'  
 ALA TRP VAL GLY THR ILE LEU

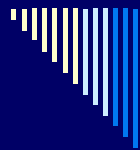
*BstEII*

mut/mut 5' GCC TGG GTG **G** ACC ATC CTG 3'  
 ALA TRP VAL VAL THR ILE LEU



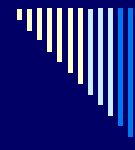
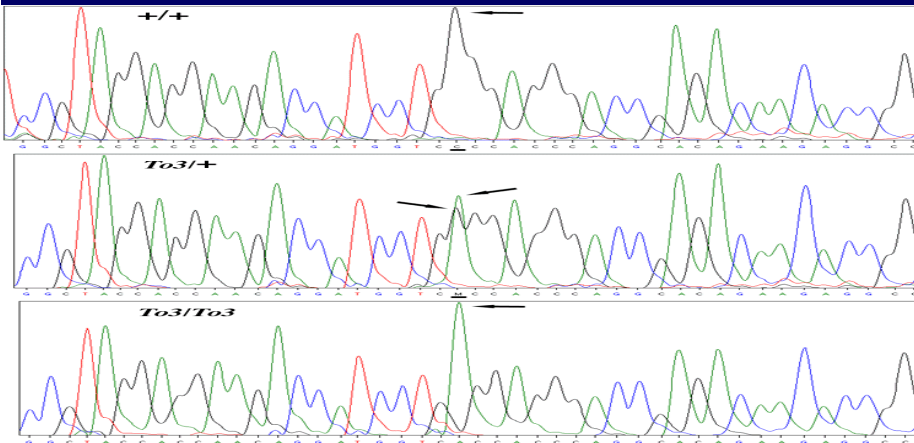
Metodika založená na špecifite sekvencie, ktorú rozoznáva enzým restričná endonukleáza

Substitúcia (transverzia) **G** → **T** v mutovanom kódujúcom vlákne vytvára jedinečné miesto **RES**, ktoré rozoznáva restričná endonukleáza ***BstEII*** a spôsobuje substitúciu **Gly** za **Val** v aminokyselínovom reťazci kódovaného polypeptidu



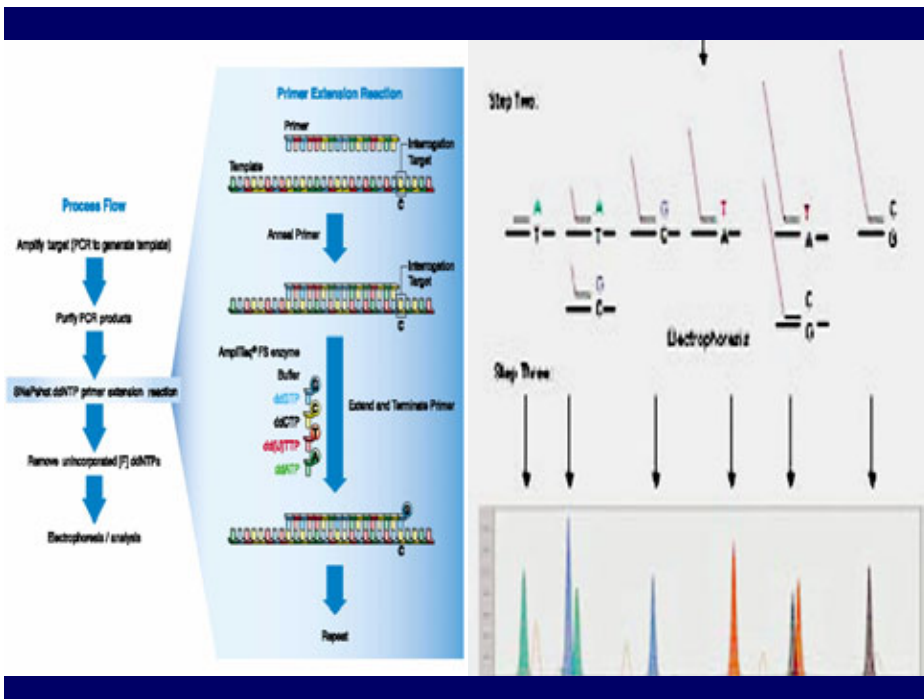
## Sekvenovanie

- stanovovanie priamej sekvencie nukleotidov DNA
- Sangerova di-deoxy reakcia-zmes značených dNTP a ddNTP
- kapilárna elektroforéza na automatickom sekvenátore



## SNaPshot - minisekvenovanie

- ☐ Overovanie SNP polymorfizmov
- ☐ Väzbové mapovanie
- ☐ Asociačné štúdie
- ☐ Umožňuje podrobne analyzovať až desať SNP polymorfizmov so známou lokalizáciou vo viacerých génoch súčasne (1-10 templátov DNA)
- ☐ Je založená na jedno-dideoxy-nukleotidovej elongácii neznačeného primeru
- ☐ Aplikovateľná v prípadoch, kedy nemožno využiť metódu PCR-RFLP, nahradenie sekvenovania

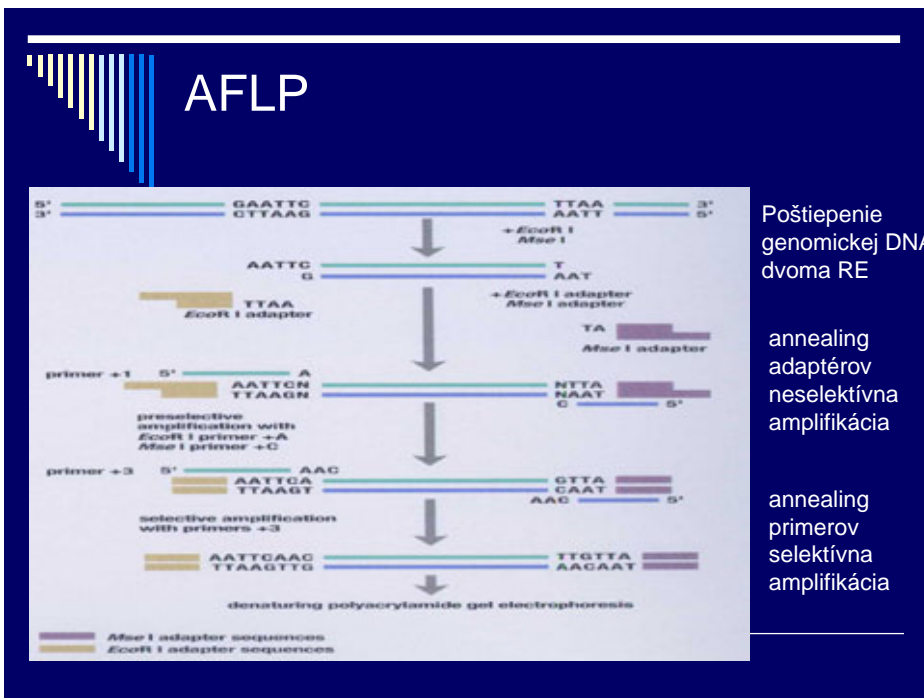


## Mikrosatelity – Fragmentačná analýza

- multiplex PCR – v jednej skúmavke viac mikrosatelitov (9-12)
- kapilárna elektroforéza, fluorescenčná detekcia



rozsah veľkostí pre jednotlivé MS



## Vyhodnotenie dát

## Výpočet frekvencií aliel (RFLP)

- Mendlove zákony, segregácia
- Hardy-Weinbergov zákon ( $F=0$ )

DD, Dd, dd – genotypy

p, q – frekvencia aliel

$$p^2 + 2pq + q^2 = 1$$

- Kodominantná dedičnosť:

$$p = \frac{2n_{DD} + n_{Dd}}{2n} \quad q = \frac{2n_{dd} + n_{Dd}}{2n}$$

- Porovnanie empirických a HW početností pomocou vhodného štatistického testu ( $\chi^2$ )

□ Dôsledky inbrídingu v populácii:

### Bernsteinov – Wrightov zákon

(rozloženie genotypov v populácii, v ktorej sa uplatňuje inbríding)

V imbr. populácii je v porovnaní s panmiktickou je zvýšená frekvencia každého typu homozygotov o  $pqF$   
znížená frekvencia heterozygotov o  $2pqF$

DD	$p^2 + pqF$
Dd	$2pq - 2pqF$
dd	$q^2 + pqF$

Možno dokázať, že i keď sa genotypové frekvencie vplyvom imbrídingu menia, frekvencie aliel zostávajú nezmenené

$$p' = (p^2 + pqF) + \frac{1}{2}(2pq - 2pqF) \quad p' = p \quad F - \text{koeficient inbrídingu}$$

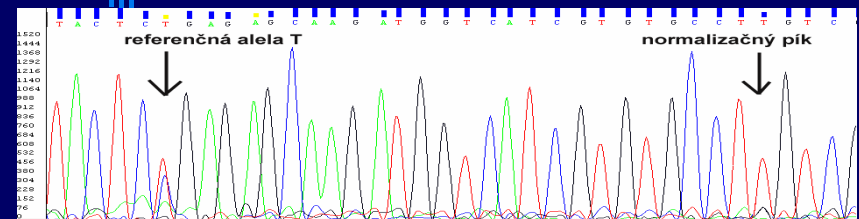
## Výpočet frekvencií aliel (Sekvenovanie)

- Polymorfne miesto –  $N$  (redukovaná výška píku)
- Z relatívnej výšky píkov polymorfnych bází je možné určiť frekvenciu aliel sekvenovaním zmesnej vzorky (DNA pooling)
- Programy

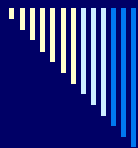
DNA Sequencing Analysis Software – primárna analýza

Sequencher, Seqscape – zrovnávanie, reanalýza

## Algoritmus pre odhad alelových frekvencií:



- Odhad frekvencie aliel v DNA pooling – porovnanie intenzity signálu (výška píku) zmesnej a referenčnej vzorky
- Referenčná vzorka – heterozygot – C/T – vyšší signál pre alelu T – alela T = referenčná alela
- Cca 20 bp 3' alebo 5' smerom od tohto miesta nájdeme pík podobnej veľkosti = normalizačný pík T
- Identifikácia zodpovedajúcich píkov v pooling



- 1a. výpočet pomeru normalizačných píkov pre získanie **normalizačného faktoru**:

$$f = N_{ref}/N_{pool}$$

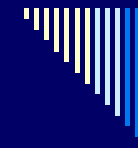
**N<sub>ref</sub>** – výška normalizačného píku v referenčnej vzorke

**N<sub>pool</sub>** – výška normalizačného píku v pooling

- 2a. výpočet **alelovej frekvencie referenčnej alely v pooling** pomerom výšok pík polymorfných aliel:

$$F = c \cdot f \cdot (P_{pool}/P_{ref})$$

F – frekvencia referenčnej alely v pooling, c – konštanta stavu referenčnej vzorky (homozygot = 1, heterozygot = 0,5), f – normalizační faktor, **P<sub>pool</sub>** – výška referenčnej alely v poole, **P<sub>ref</sub>** – výška referenčnej alely v ref.vzorke

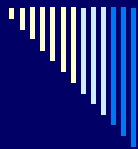


- 1b. Konštantou c vynásobíme pomer výšky referenčnej alely a normalizačného píku v poole a výsledok delíme pomerom výšky referenčnej alely a normalizačného píku v referenčnej vzorke:

$$F = \frac{c \cdot (P_{pool} / N_{pool})}{(P_{ref} / N_{ref})}$$

- 2b. Frekvenciu druhej alely polymorfizmu vypočítame odčítaním frekvencie referenčnej alely od čísla 1.

**Pozn:** V prípade, že sa v susedstve polymorfizmu nenachádza pík T podobnej veľkosti ako je referenčná alela, môže sa za referenčnú alelu považovať pík menšej veľkosti (v našom prípade C).



## Štatistická analýza

### Nukleotidová diverzita (Nei a Li, 1979)

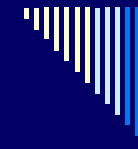
GAGGTGCAACAG  
GCGGTGCAACAG  
GTGGTGAACAG  
GGGGTGAACAG

GAGGTGCAACAG  
GAGGACCAACAG  
GAGGTGCATCAA  
GGGGTGGAAACAG

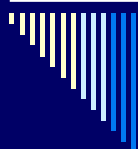
$$\Pi_{ij} = \sum_{ij}^q x_i x_j d_{ij}$$

- $x_i$  – frekvencia i-tej alely
- $x_j$  – frekvencia j-tej alely
- $d_{ij}$  – počet rozdielnych nukleotidov, príp. substitúcií medzi i-tou a j-tou alelou
- q – celkový počet aliel (rozdielne sekvencie)

V populácii s náhodným krížením je  $\pi$  ekvivalentné heterozygotite na nukleotidovej úrovni



- Pre analýzu vhodné – najmä substitúcie
- V praxi – frekvencia menej frekventovanej alely dostatočne vysoká
- Pravdepodobnosť výskytu / vzniku SNP
- Zohľadnenie mutačnej rýchlosti, mutačného tlaku a selekcie v populácii
- Priemerná heterozygotnosť ( $\pi$ )
- Nukleotidová diverzita
- Počet segregačných/polymorfných miest (s)
- Počet analyzovaných chromozómov, aliel, sekvencií (ich dĺžka)
- Vytvorenie vhodného modelu so zodpovedajúcimi parametrami**



# Výpočet frekvencií aliel (SNaPshot, fragmentačná analýza, AFLP)

Štandard veľkostí fragmentov:

GeneScan 120 LIZ Size Standard

GeneScan 500(-250) Size Sstandard (ROX)

Matrice a filtre:

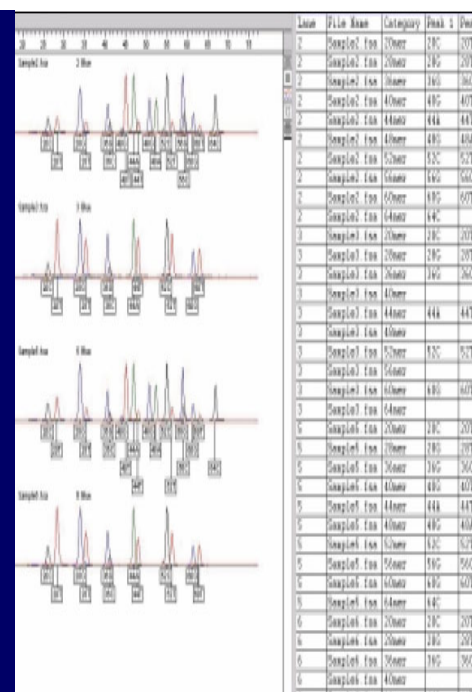
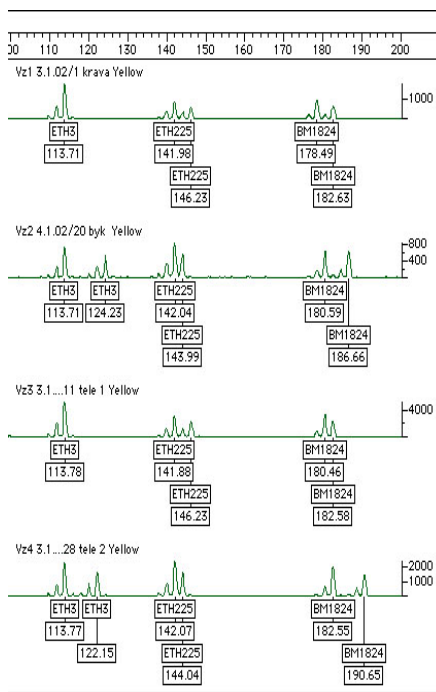
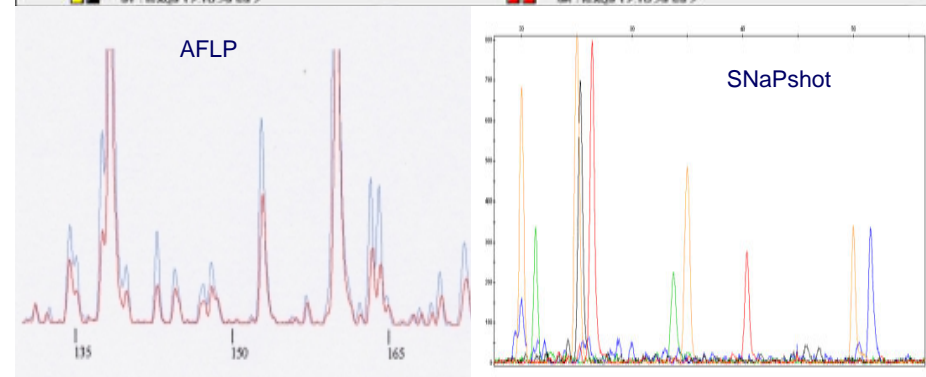
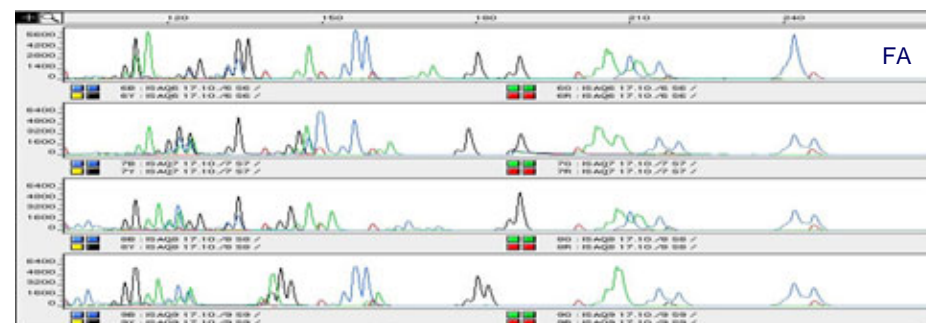
Matrix Standard Set DS-02 (E5), Matrix Standard Set DS-30 (G5), Matrix Standard Set DS-33 (D)...

4-5 fluorescenčných farieb

eliminácia potreby manuálneho „sajzovania“ aliel

„Poloautomatické“ vyhodnotenie: **Genotyper Software v 3.7**  
**GeneMapper**

genotypizácia, sumarizácia dát a výpočet alelových frekvencií



Decorative graphic of vertical bars of varying heights.

## Ďakujeme za pozornosť

[kristinciv@seznam.cz](mailto:kristinciv@seznam.cz)  
[knoll@mendelu.cz](mailto:knoll@mendelu.cz)