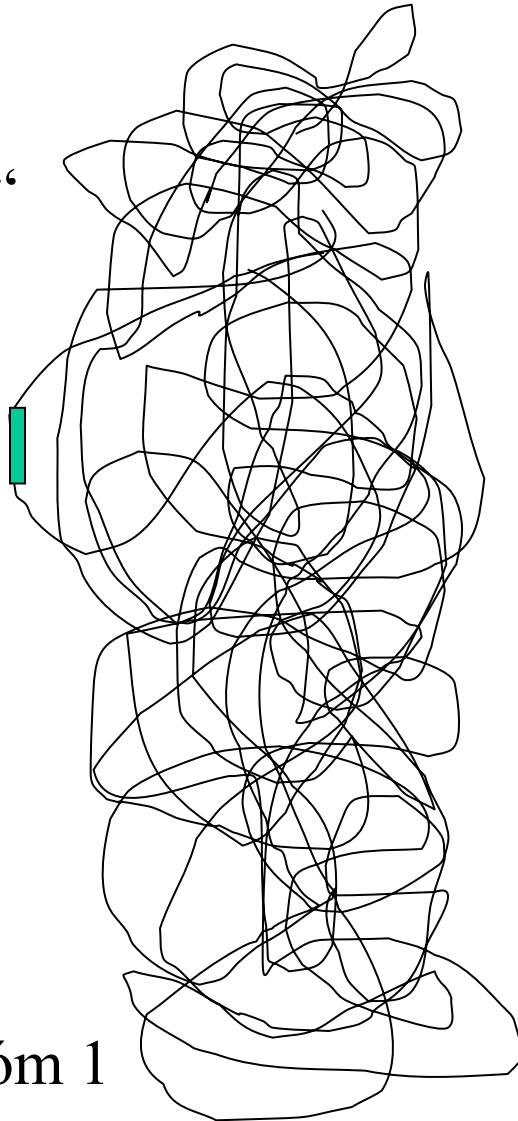


Genotypizace - stanovení genotypu

- stanovení formy (alely, haplotypu) určitého úseku DNA („genetického markeru”)
 - 1) izolace celkové DNA z tkání
 - 2) amplifikace požadovaného úseku DNA (PCR-based methods)
 - 3) studium variability daného úseku (lokus = marker = znak)

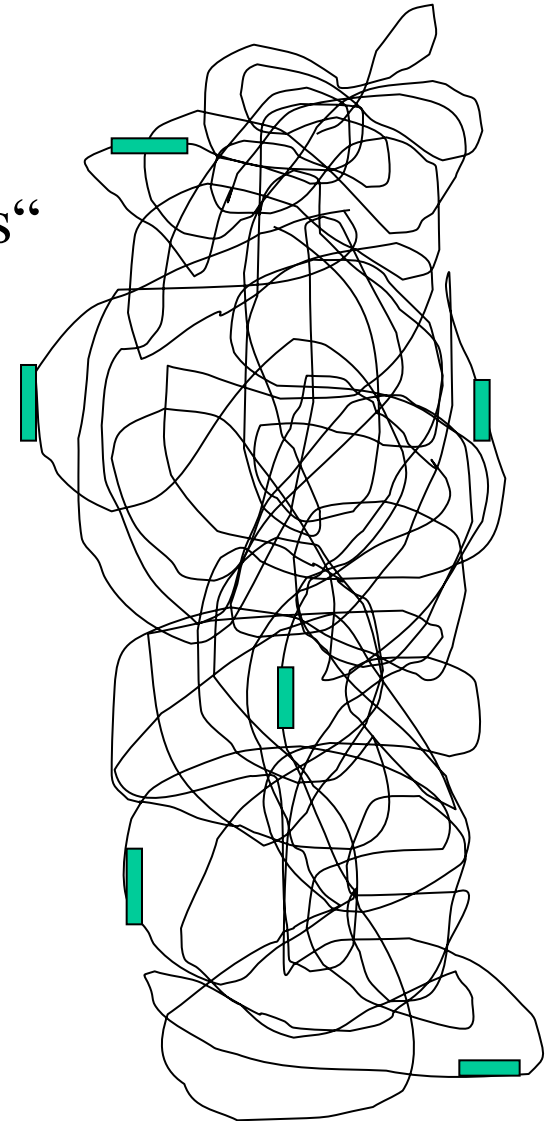
Typy genetických markerů

„single-locus“



Př.: chromozóm 1

„multi-locus“



Typy genetických markerů

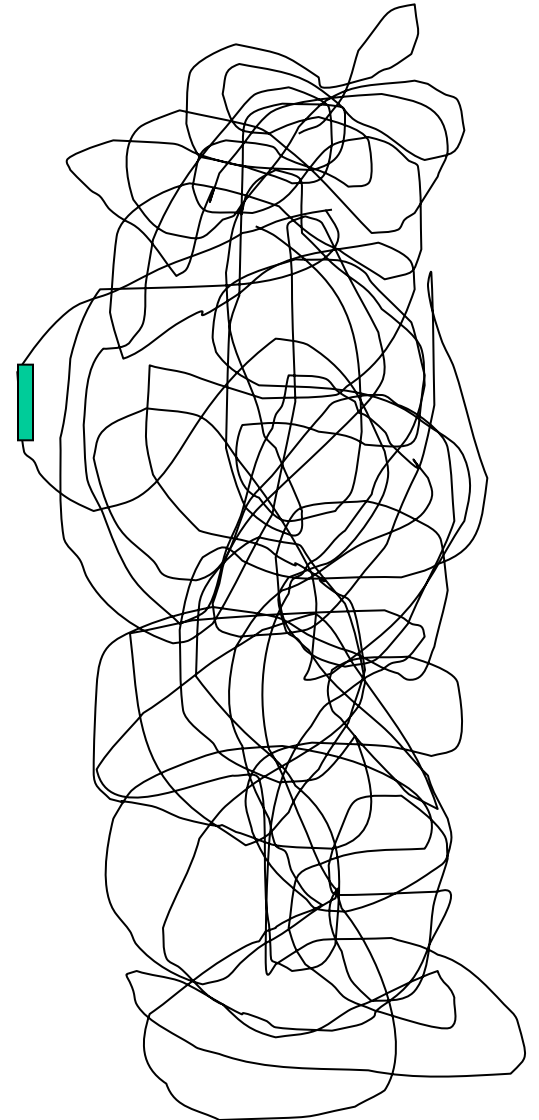
- **dominantní** markery – odliší pouze přítomnost (či nepřítomnost) daného znaku; tj. neodliší obě jeho formy na homologních chromozómech
- **kodominantní** markery – identifikace homologních alel, tj. je možno rozlišit homozygotní a heterozygotní stav (umožňují stanovit frekvenci alel)

Typy genetických markerů

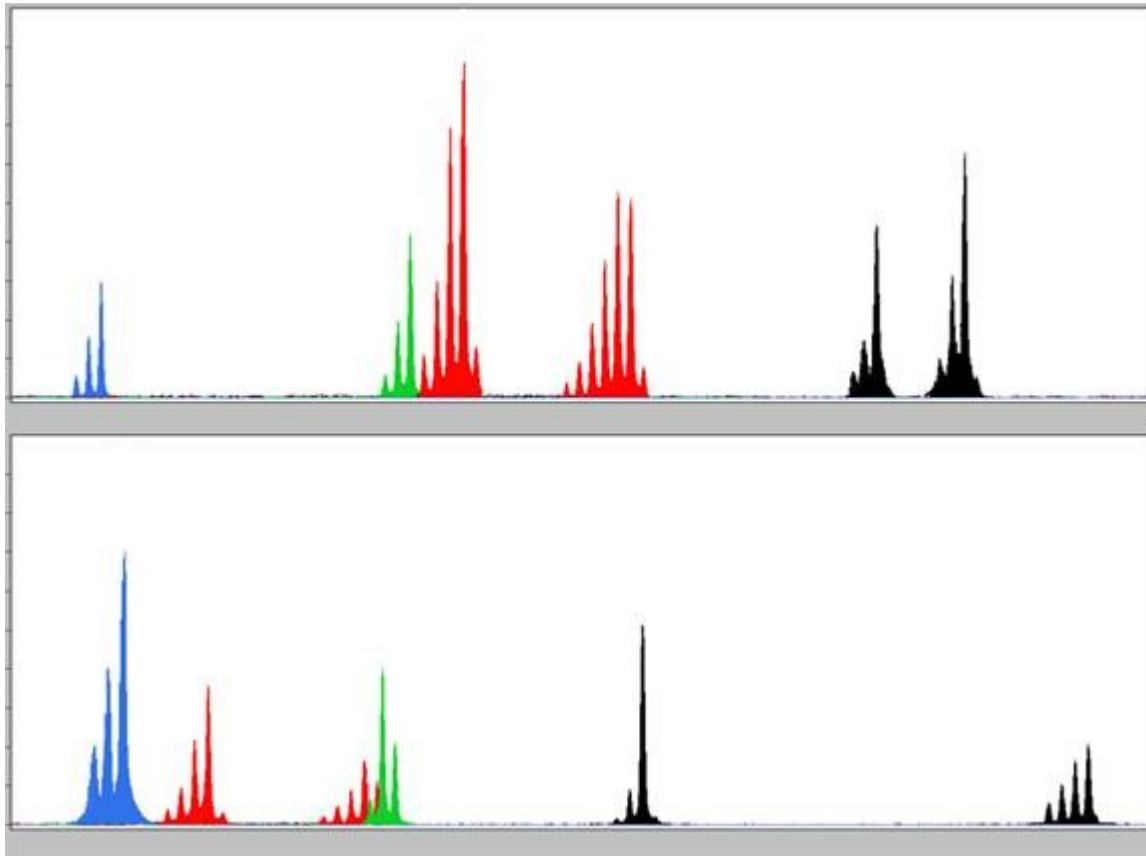
	Single locus	Codominant	PCR	Celková variabilita
Jaderné více-lokusové („nuclear multi-locus“)				
Minisatelitový DNA fingerprints	No	No	No	High
RAPD	No	No	Yes	High
AFLP	No	No	Yes	High
Jaderné jedno-lokusové („nuclear single locus“)				
Alozymy	Yes	Yes	No	Low-medium
Mikrosatelity	Yes	Yes	Yes	High
SINE (LINE)	Yes	Yes	Yes	Low
SNPs (sekvence)	Yes	Yes	Yes	Low-high

Single-locus genetic markers

- kodominantní – možno stanovovat frekvence alel (= lze odlišit homo- a heterozygota)
- **allozomy** a jiné funkční geny - **MM**
- **mikrosatelity** – délkový polymorfismus
- **SNPs** (single nucleotide polymorphisms) – sekvenční polymorfismus
- **SINE, LINE** – inzerce (tj. délkový polymorfismus)

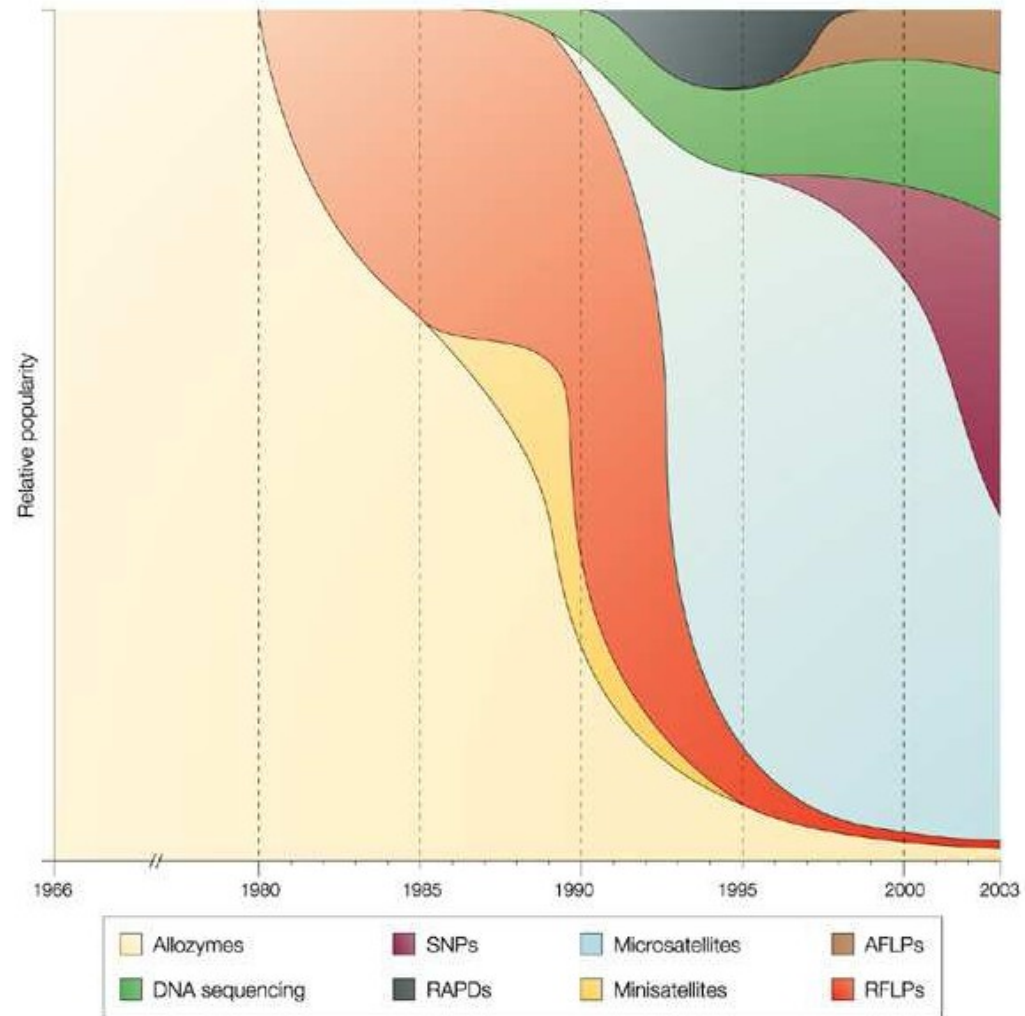


Mikrosatellity



Mikrosatelity byly (a pro něco stále budou) velmi užitečné markery v molekulární ekologii

(i když genotypizační
metody se budou měnit)



Mikrosatelity

- VNTR („variable number of tandem repetitions“), SSR („simple sequence repeats“)
- jednotlivé alely se liší délkou

TTCAGGCACACACA**ATCTCTAGCTTTGA**

27 bp

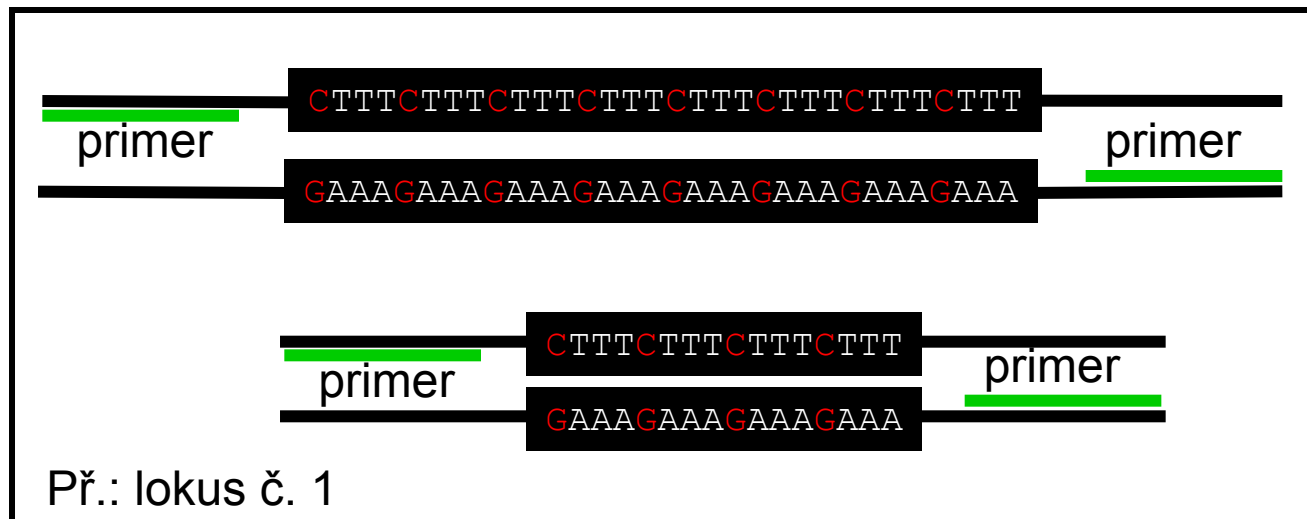
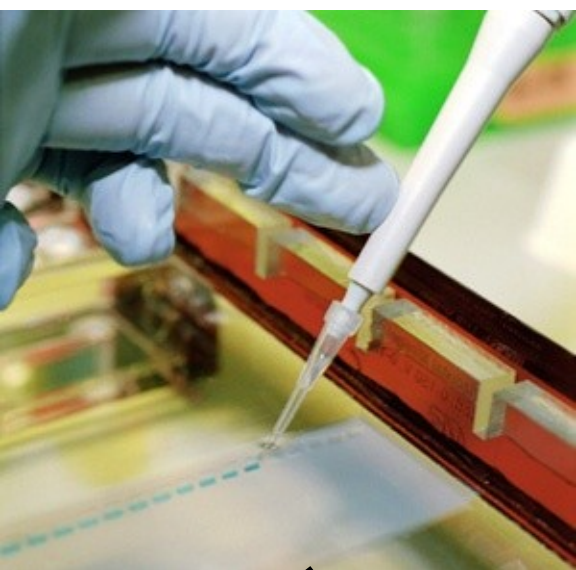
TTCAGGCACACA**ATCTCTAGCTTTGA**

25 bp

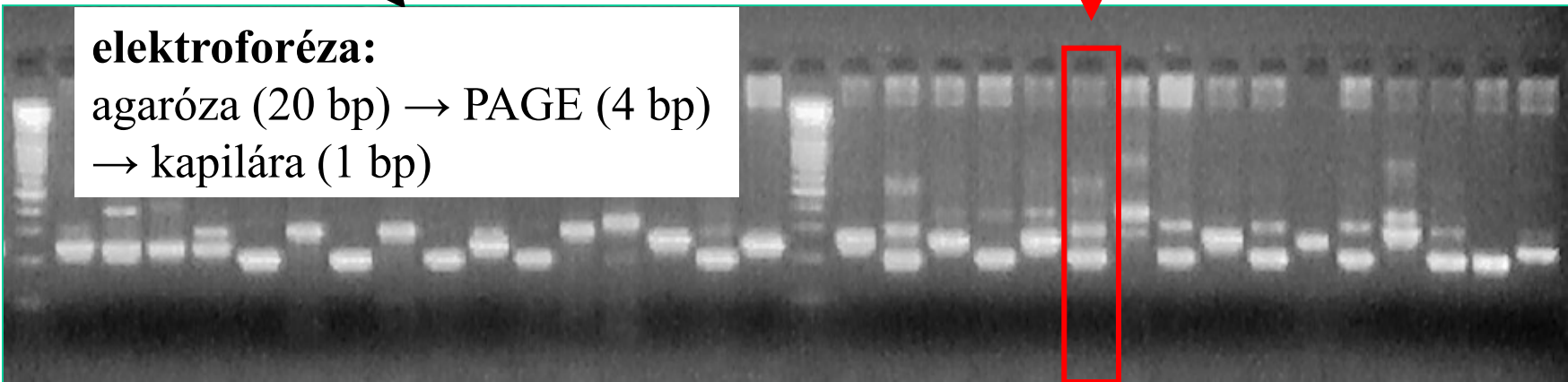
genotyp diploidního jedince: **25/27**

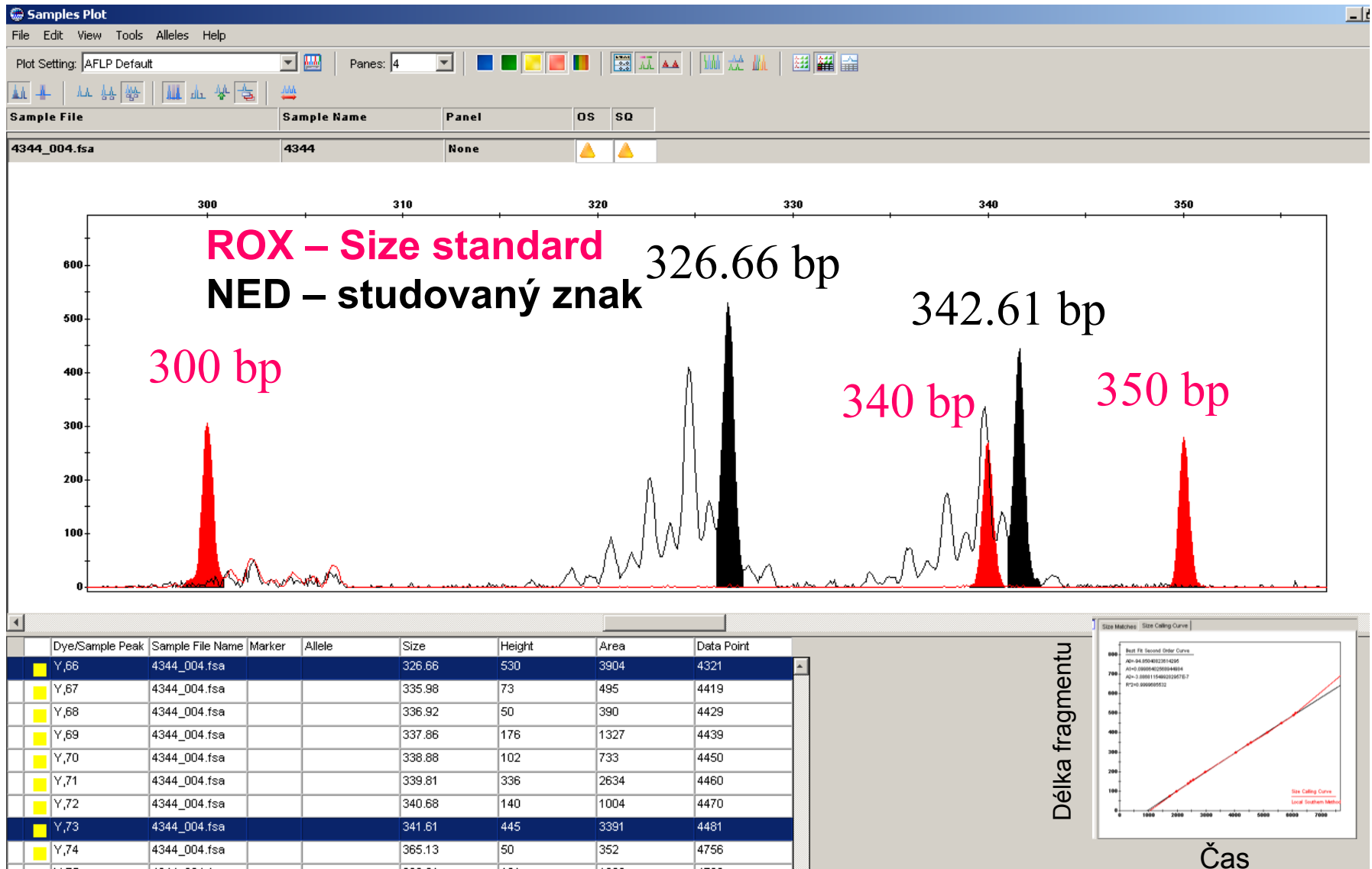
Mikrosatelity

- 1-6 (nejč. 2-4) bp motiv
- početné po celém genomu
- vysoká úroveň polymorfismu (běžně 15 alel v populaci)
- Mendelovská dědičnost (autosomy) - kodominance
- ideální pro studium populační struktury a příbuzenských vztahů



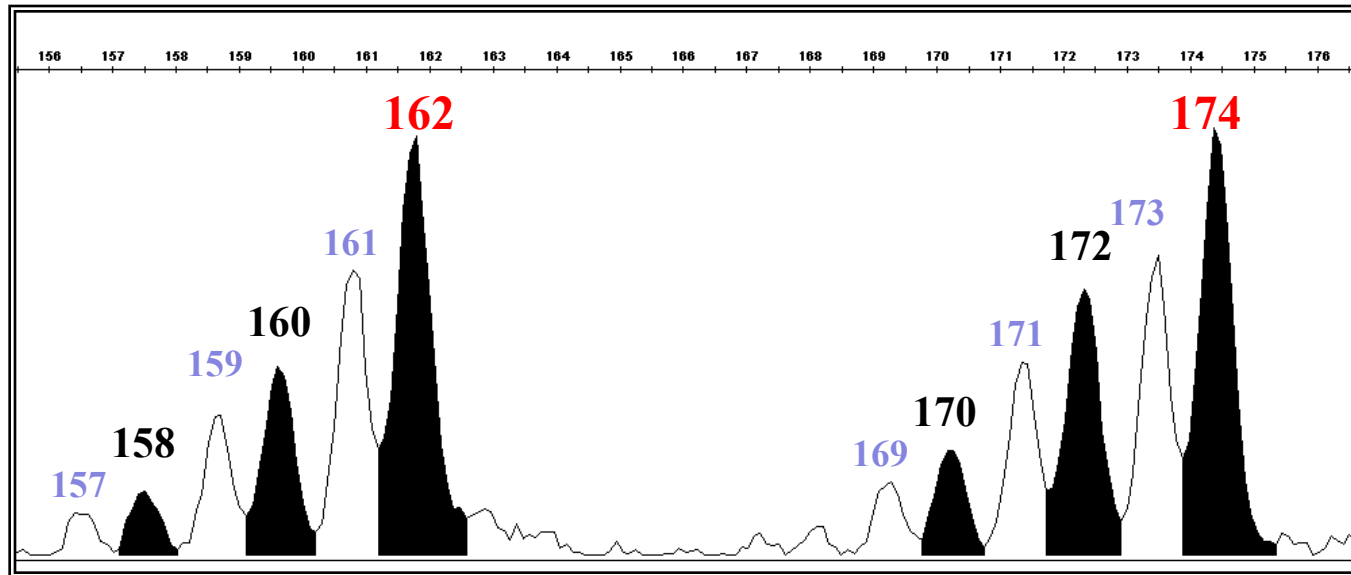
elektroforéza:
agaróza (20 bp) → PAGE (4 bp)
→ kapilára (1 bp)





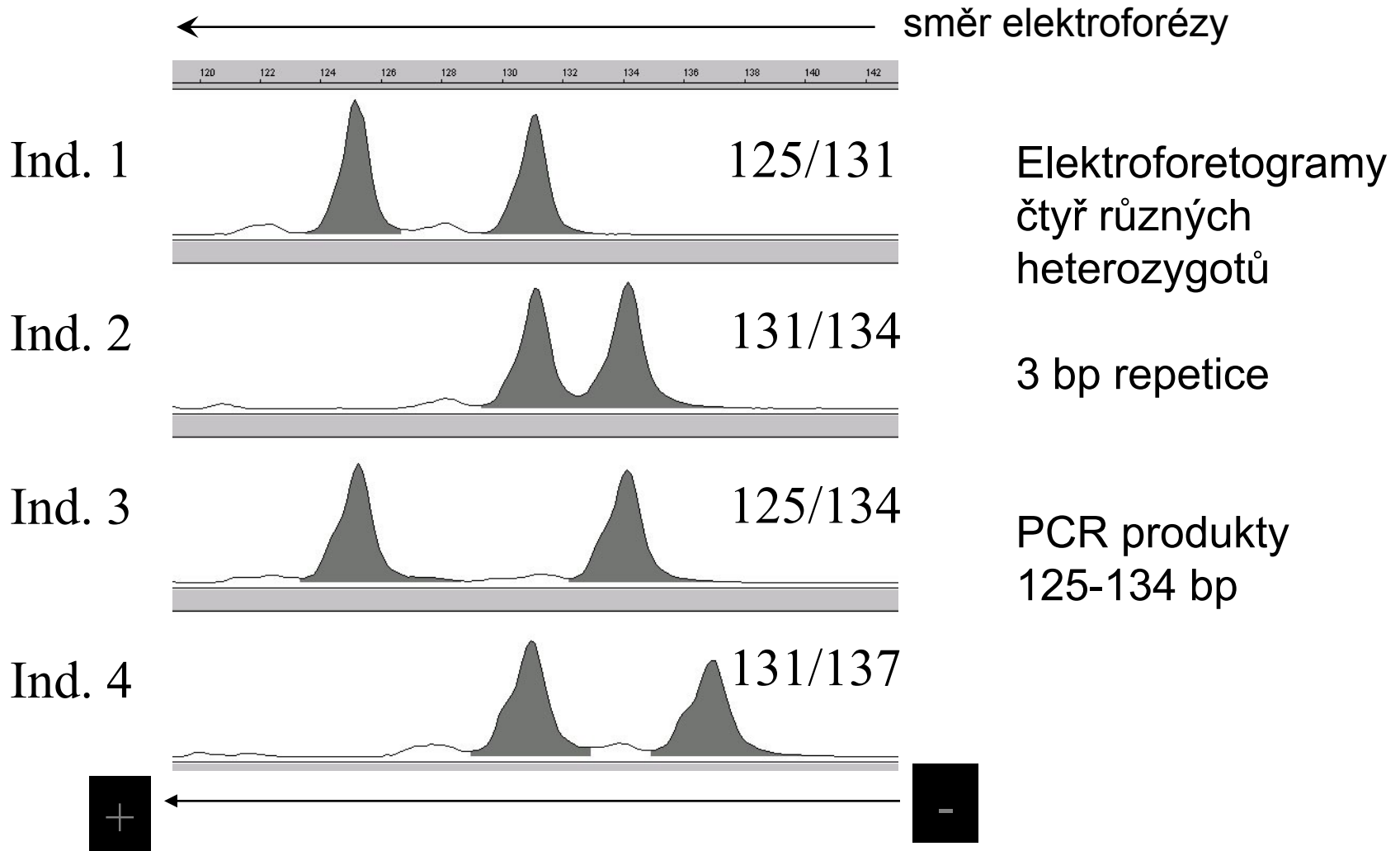
Genotyp mikrosatelitu na lokusu NED = 326/342 nebo 327/343
 Programy: GeneMapper, Genotyper, Geneious, GeneMarker, ...

Genotyp 162/174

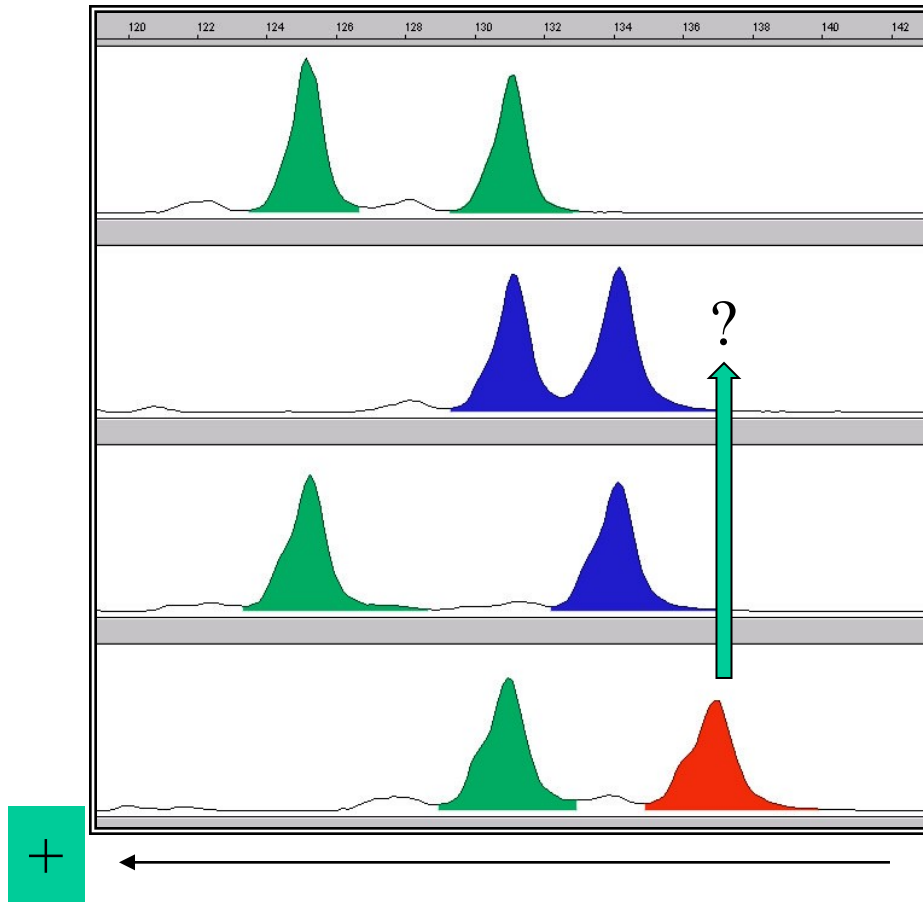


- alely a jejich stuttery jsou černě (rozdíl mezi nimi je 2 bp)
- bílé píky jsou tzv. „mínus A-alely“ a jejich stuttery = výsledek jiné chyby polymerázy, a to nepřidání koncového adeninu
- rozdíl mezi černým a sousedním bílým píkem je 1 bp (tj. chybějící adenin)
- pattern daného lokusu je vždy specifický a často záleží na PCR podmínkách

Srovnání různých jedinců - analýzy příbuznosti



Př. Analýza příbuzenských vztahů



Genotyp (bp)

Matka: 125/131

Otec: 131/134

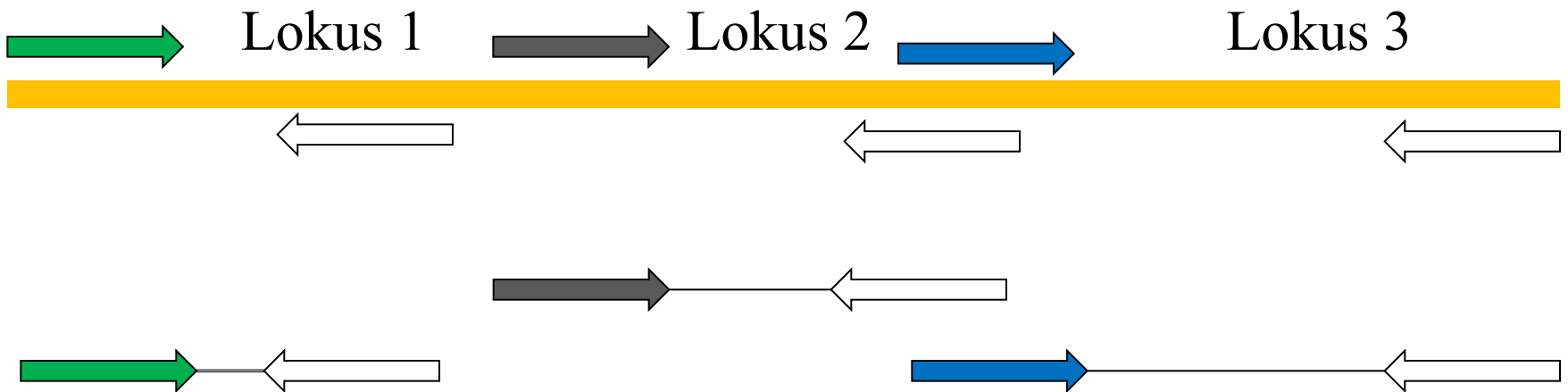
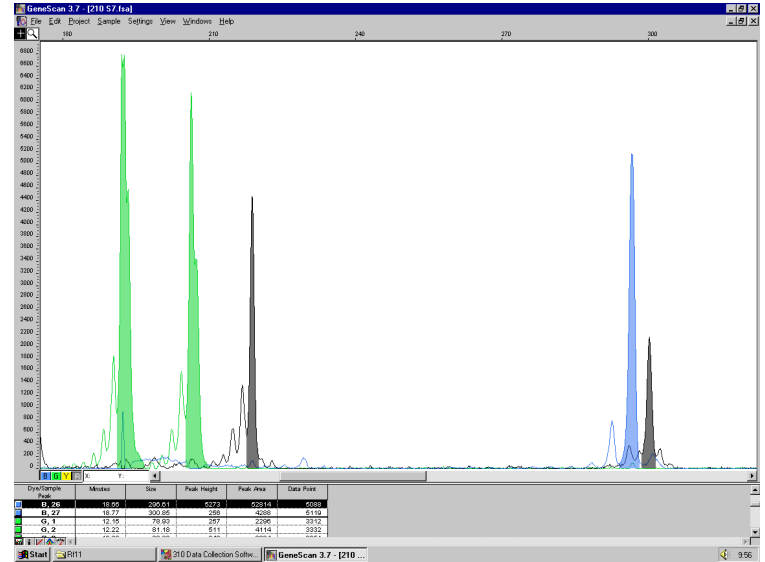
Potomek 1: 125/134

Potomek 2: 125/137

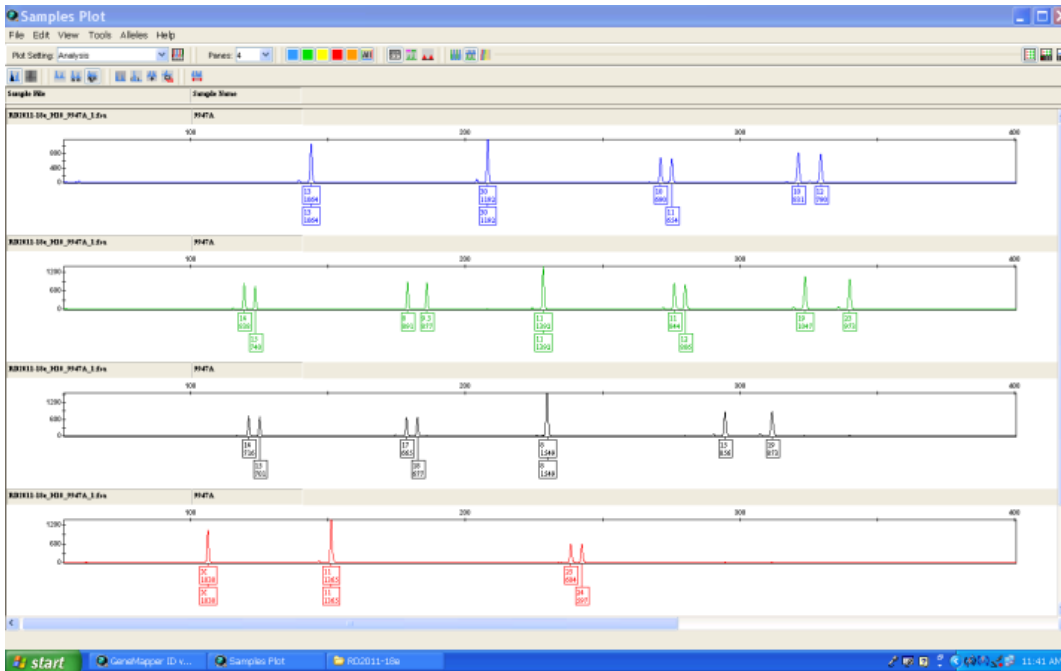
Sledovaný otec mohl zplodit potomka 1, ale zcela jistě není otcem potomka 2

Různé značení různých znaků

- Snížení časových a finančních nákladů
- = „multiplex set“
- Až 4 různé barvy (+ 5. barva jako velikostní standard) - analýza až 4 lokusů o stejné velikosti alel



Identifikace jedinců u lidí



AmpFSTR® Identifiler™



D8S1179 D21S11 D7S820 CSF1PO

D3S1358 TH01 D13S317 D16S539 D2S1338

D19S433 vWA TPOX D18S51

A D5S818 FGA



GS500-internal lane standard

16 lokusů = spolehlivá identifikace jedinců
(v euro-americké populaci)



Mikrosatelity - omezení

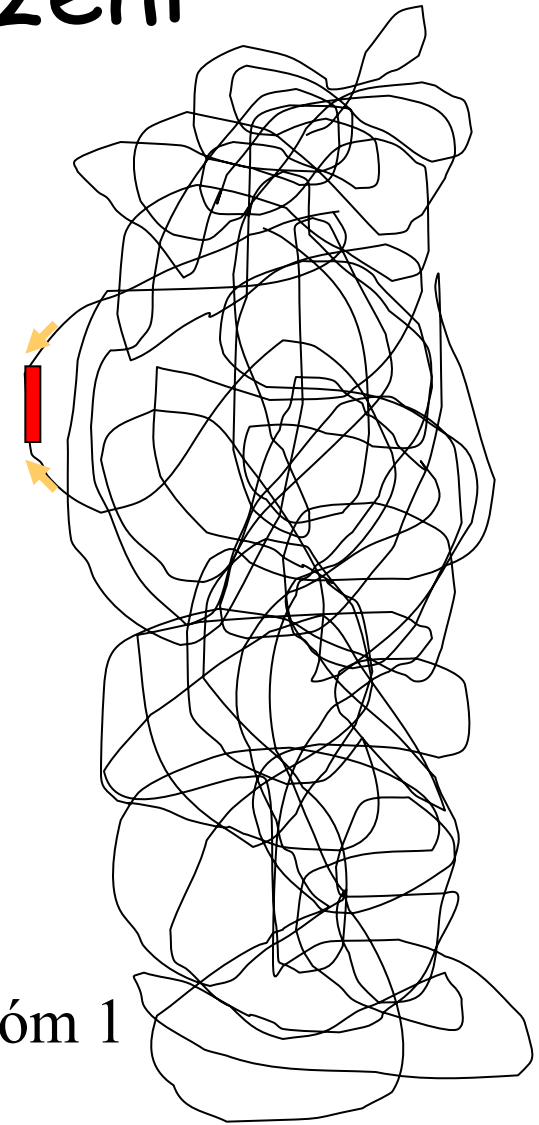
- nalezení lokusů (navržení primerů) je pracné a nákladné u volně žijících druhů (genomová knihovna, klonování, screening, sekvencování)



TTCAGG**CACACACA**TCTCTAGCTTCGA



„flanking regions“ – ohraničují repetici a zde musí být navrženy primery pro PCR

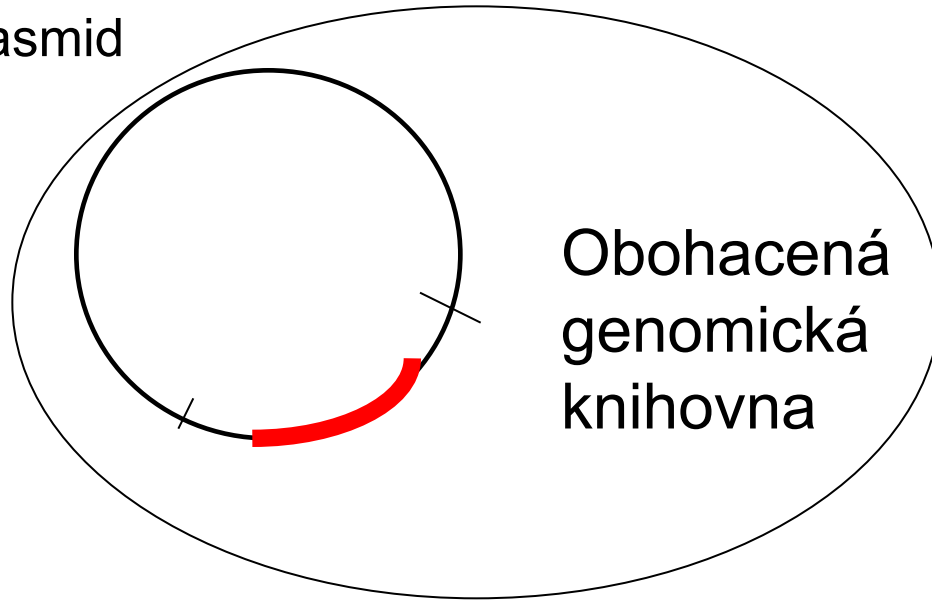


Př.: chromozóm 1

Štěpení, obohacení, klonování, screening a sekvenování

Každý klon obsahuje jednu sekvenci

vector =
plasmid



izolace vektorů s
inzertem



screening klonů obsahujících
repetice (hybridizace se sondou)



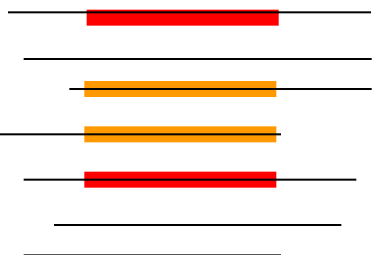
sekvenování inzertů
(repetitivní DNA + flanking regions)



design primerů a testování
polymorfismu



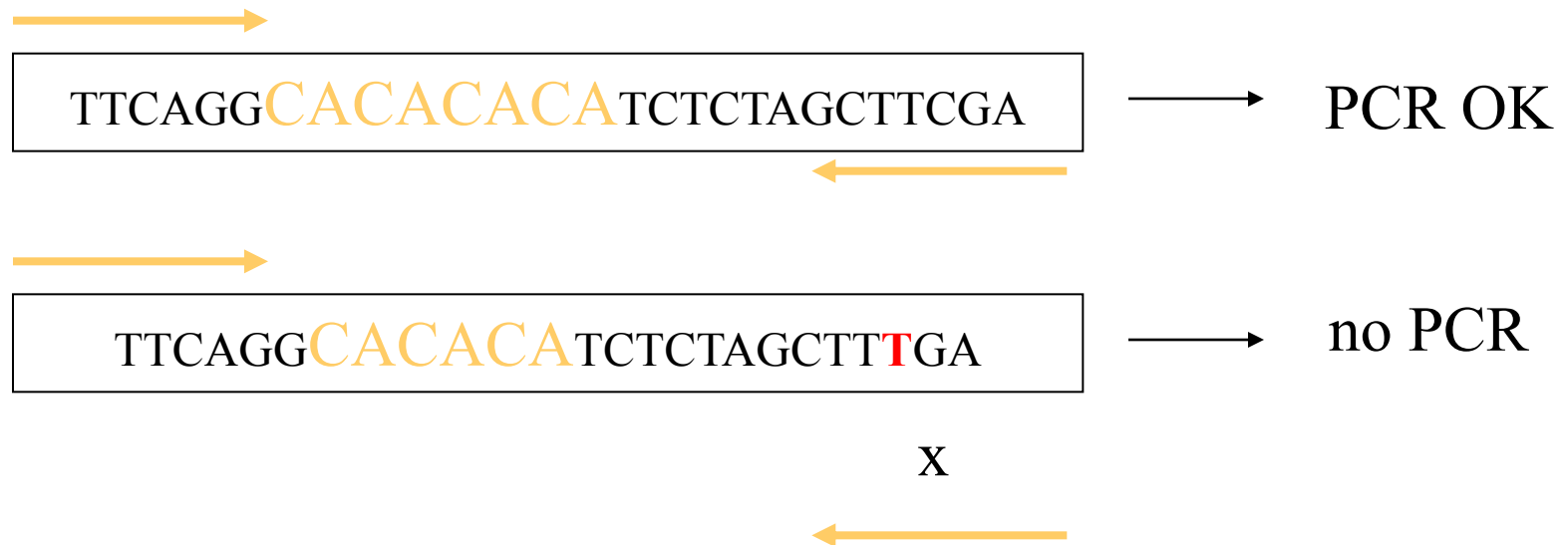
ligace do plasmidu, transformace



Genomická DNA po rozštěpení
a obohacení na repetice

Alternativa: cross-species amplification

- „**cross-amplification**“ – úspěšnost klesá s fylogenetickou vzdáleností
- **nulové alely** (mutace v primerových sekvencích)
→ vyšší proporce „homozygotů“



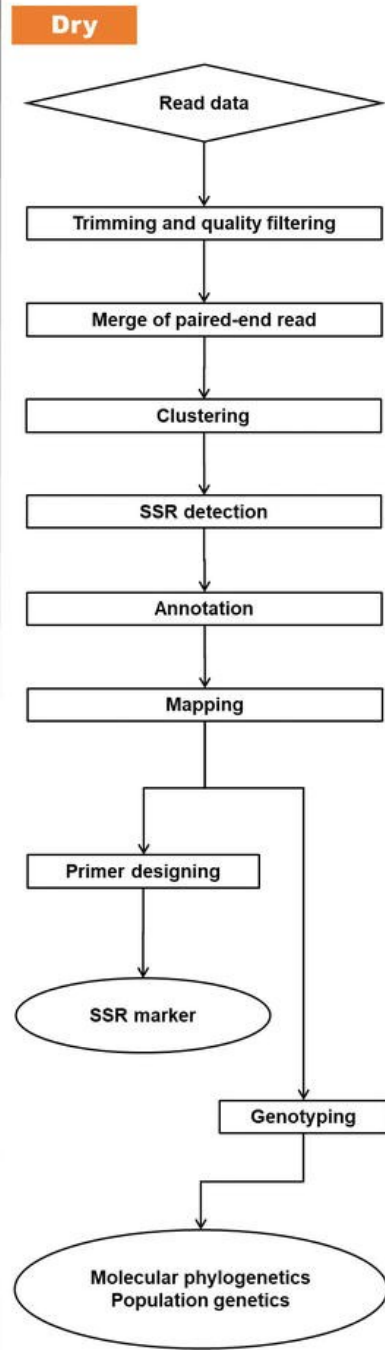
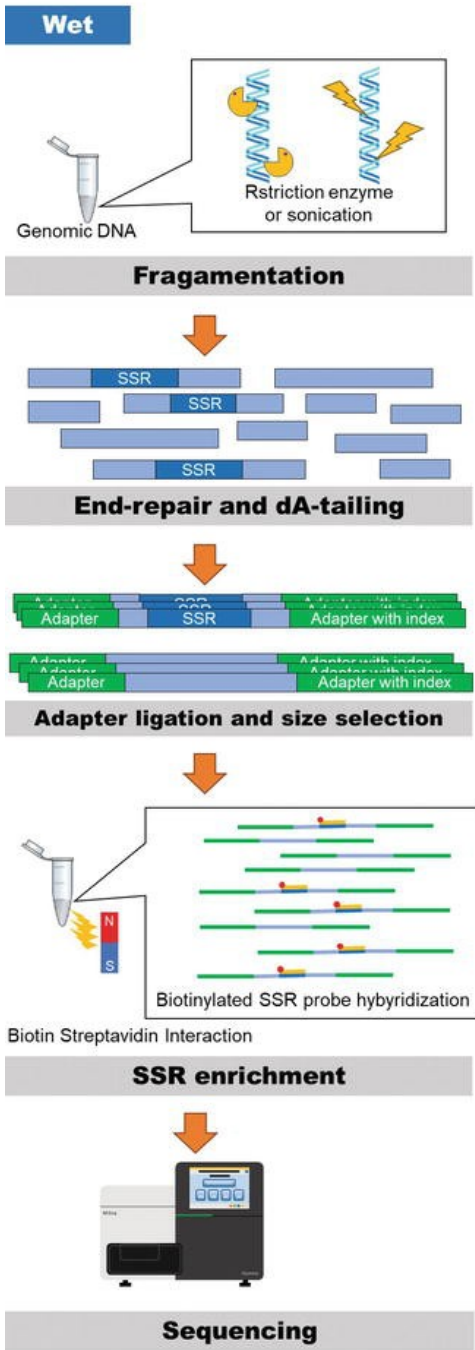
Nulové alely a genotypizační chyby komplikují analýzy příbuznosti

	lokus 1 null alleles		lokus 2 genotyping error	
Matka	100	150	300	350
Samec 1	100	100	300	367
Mládě	150	150	350	365

Samec 1 je vždy opravdovým otcem, ale jednoduchá „exclusion“ metoda ho vždy vyloučí

Optimalizace mikrosatelitů v současnosti = NGS

- „next-generation sequencing“ (= HTS, „high-throughput sequencing“) – velice rychlá sekvenace velkého množství DNA fragmentů z jakéhokoliv genomu
- vyhledání repetitivních sekvencí vhodným softwarem a navržení primerů
- identifikace nových mikrosatelitů v genomu rychle, elegantně a relativně levně



Target species	Library style	NGS platform	Number of reads	Sequences including SSR
<i>Agkistrodon contortrix</i> (Copperhead snake)	WG	GS-FLX	128,773	14,612
<i>Anisogramma anomala</i>	WG	GAIIX	26,036,313	44,247
<i>Aristeus antennatus</i> (Red shrimp)	WG	GS-FLX	165,507	247
<i>Aristotelia chilensis</i> (Maqui)	WG	GS-FLX	165,043	24,494
<i>Artocarpus altilis</i>	WG	MiSeq	2,341,465	47,607
<i>Aspidistra saxicola</i>	cDNA	HiSeq2000	13,133,336	4764
<i>Brachiaria ruziziensis</i> (Ruzigrass)	WG	GAIIX	186,764,108	139,098
<i>Callosobruchus chinensis</i> (Adzuki bean weevil)	WG	HiSeq2500	106,888,024	6593
<i>Camelina sativa</i>	cDNA	GAIIX	10,830,000	14,140
<i>Camellia sinensis</i> (Tea plant)	cDNA	HiSeq2000	26,874,116	5649
<i>Carthamus tinctorius</i> (Safflower)	WG	HiSeq2000	48,502,680	23,067
<i>Catha edulis</i> (Khat)	WG	GS-FLX	65,401	11,678
<i>Catla catla</i> (Catla)	WG	PGM	29,794	21,477

Open access peer-reviewed chapter

Microsatellite Capture Sequencing

By Keisuke Tanaka, Rumi Ohtake, Saki Yoshida and Takashi Shinohara

Submitted: July 31st 2017 Reviewed: November 22nd 2017 Published: June 20th 2018

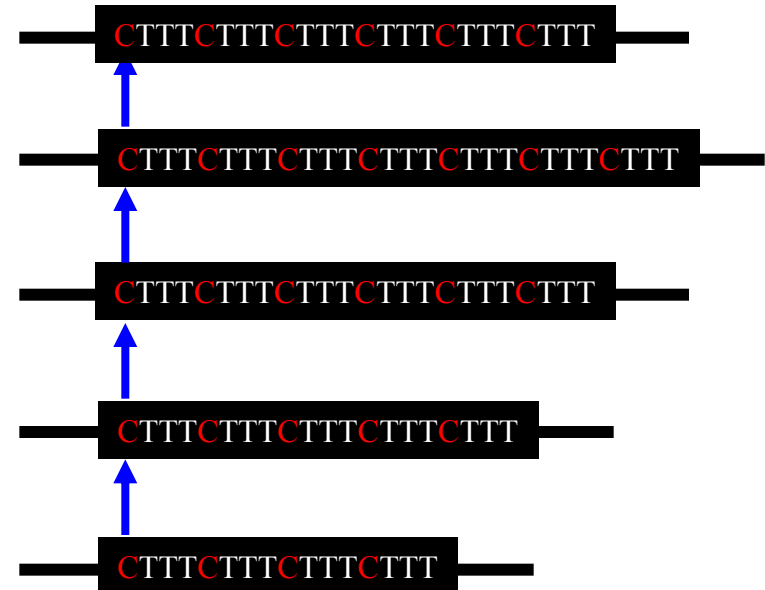
DOI: 10.5772/intechopen.72629

Využití mikrosatelitů

- identifikace jedinců a analýzy příbuznosti (zejména rodičovství)
- populační genetika (conservation genetics, landscape genetics, etc.)
- fylogeografie a analýzy historické demografie (omezeně – nutno znát **mutační model**)
- fylogenetika, tj. vzdálenější příbuzní – téměř vůbec, vysoké riziko **homoplázií**

Teoretické mutační modely (nutno definovat pro analýzy vyžadující údaj o podobnosti alel, např. při analýzách historické demografie)

- **IAM – infinite allele model**
Při mutaci ztráta nebo získání libovolného počtu opakování. Vzniká nová alela, která doposud v populaci nebyla - každá alela vznikne pouze jednou a pak už se nemění. **Není možno** určit podobnost (similarity) alel, ale pouze **identitu**.
- **SMM – stepwise mutation model**
(Mutace způsobeny pouze ztrátou nebo získáním jediného opakování motivu. Mutací může vzniknout alela, která je již v populaci přítomna – tzv. **homoplázie**. Je možno odhadnout podobnost (**similarity**) alel.



Pravda bude někde mezi ...
= Two-phase model (TPM)

Indels

- inzerce nebo delece 1bp či delších úseků – použití pouze pro populačně-genetické analýzy vyžadující „**identity**“ (nepoužitelné pro modely vyžadující „**similarity**“)

TTCAGG CACACACA TCTCTAGCTTCGA

27 bp

SMM model – možno kvantifikovat podobnost („similarity“) alel

TTCAGG CACACACA CA TCTCTAGCTTCGA

27 → 29 bp

TTCAGG CACACACA TCTC G TAGCTTCGA

27 → 28 bp

TTCAGG CACAC GACA TCTCTAGCTTCGA

27 → 28 bp

TTCAGG CACACCA TCTCTAGCTTCGA

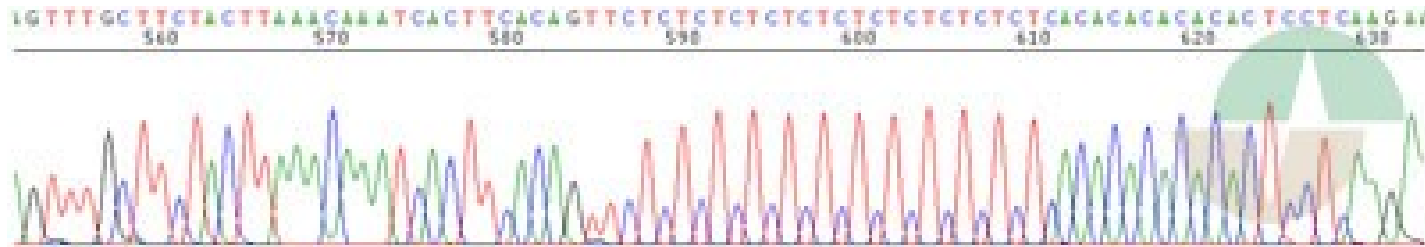
27 → 26 bp

TTCAGG CACACACA TCTCTAGTTCGA

27 → 26 bp

„Indels“ – pouze pro analýzy, kde je vyžadována „identity“ a nikoliv podobnost

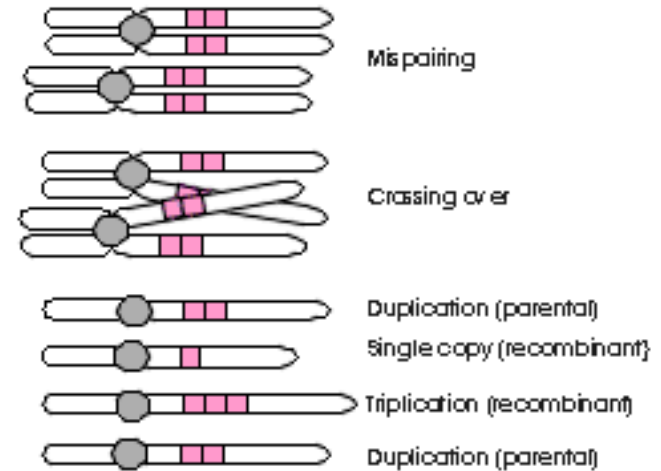
Další nepravidelnosti (tj. možné komplikace při analýzách)



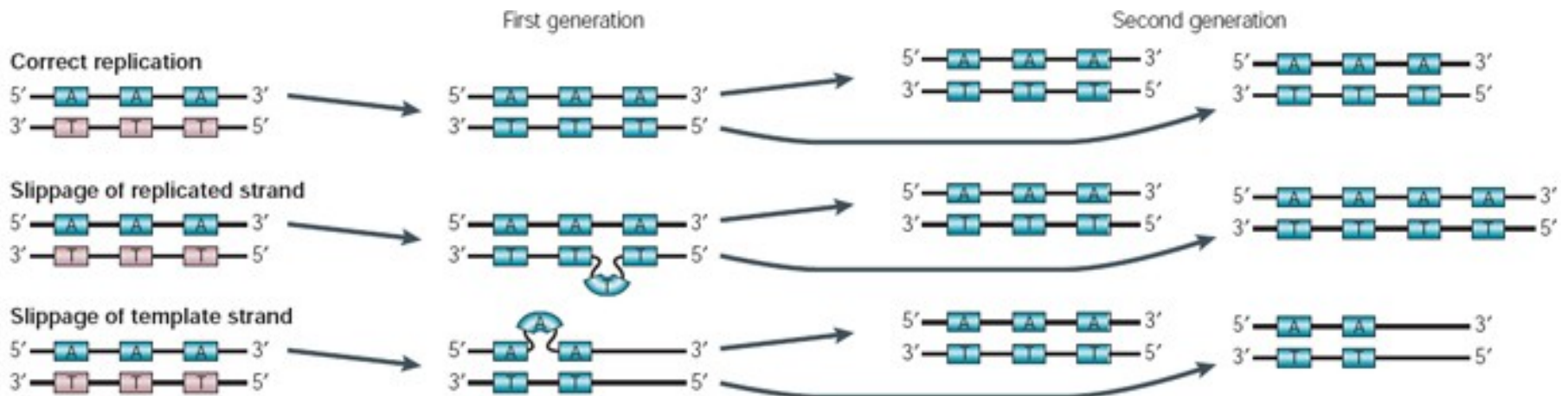
„složený mikrosatelit“ – rovněž není možné aplikovat jednoduchý mutační model

Proč je tolik alel? (microsatellite instability)

- **Nerovnoměrný (Unequal) crossing-over**
(díky špatnému alignmentu)



- **Sklouznutí polymerázy při replikaci**
Slip-strand mispairing
(při replikaci nejprve polymeráza sklouzne a vyrobí odlišný počet opakujícího se motivu mikrosatelitu, při alignmentu je pak část opakování vykloněna mimo dvoušroubovici, flanking regions tedy párují)



Bias (skutečná data)

- **Kratší mikrosatelity** (s malým počtem opakování motivu) **mají zřejmě tendenci se spíše prodlužovat** (slabě převládají adice nad delecemi)
- **Delší mikrosatelity se spíše zkracují** (náchylnější k velkým delecím)
- **Delší mikrosatelity rychleji mutují** (díky více opakováním je vyšší pravděpodobnost pro sklouznutí polymerázy (SSM) – mají více alel)

Mikrosatelity - závěry

- Homoplasie – nevhodné pro fylogenetické analýzy
- Stepwise mutation model (SMM) platí jen omezeně – obtížně aplikovatelné v analýzách evoluční historie (pokud se uplatňují mutace)
- Potřeba neustálé standardizace při genotypizacích - i v populační genetice jsou rychle nahrazovány jinými markery, např. SNPs
- Tolik to nevádí při identifikaci jedinců a pro analýzy příbuznosti (paternity), kde je možno zanedbat mutace – zde se budou používat ještě hodně dlouho

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%)

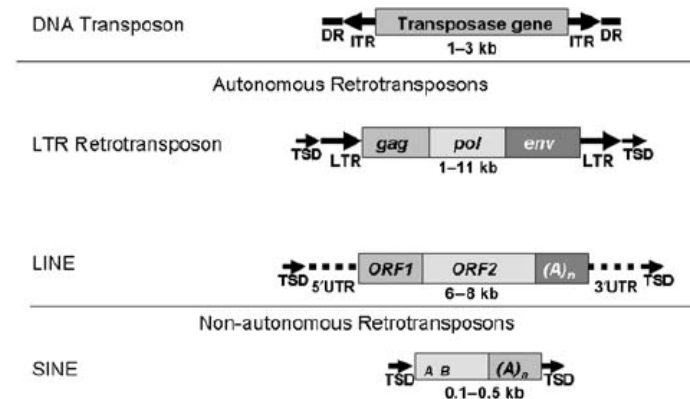


*Objev DNA transpozonů u kukuřice:
Barbara McClintock*

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

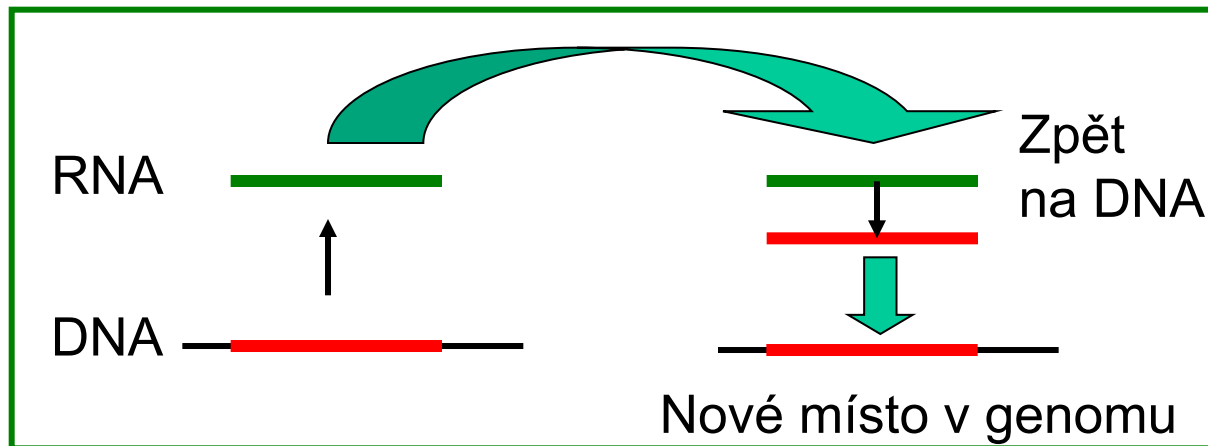


- **Nekódují proteiny, neautonomní, 100-1000 bp**

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

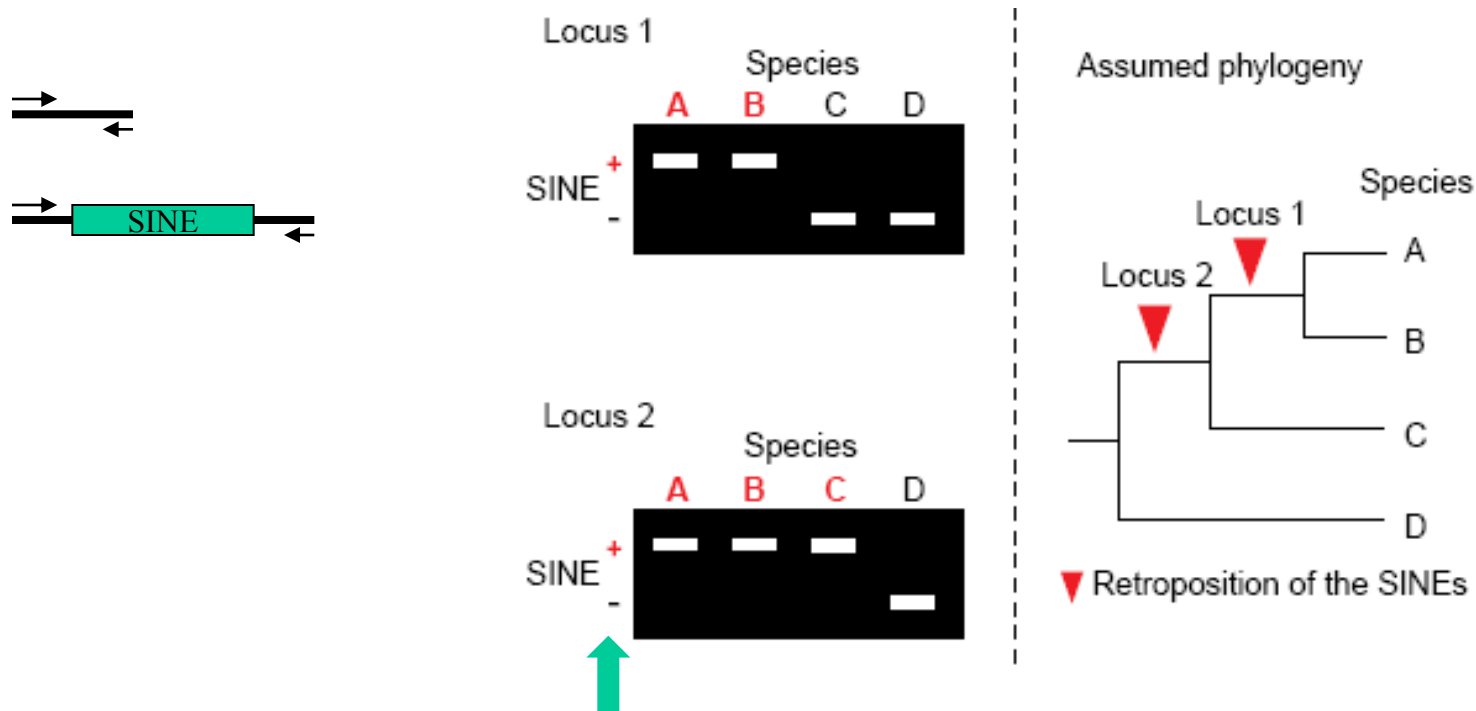
LINE - mechanismus transpozice

- 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)

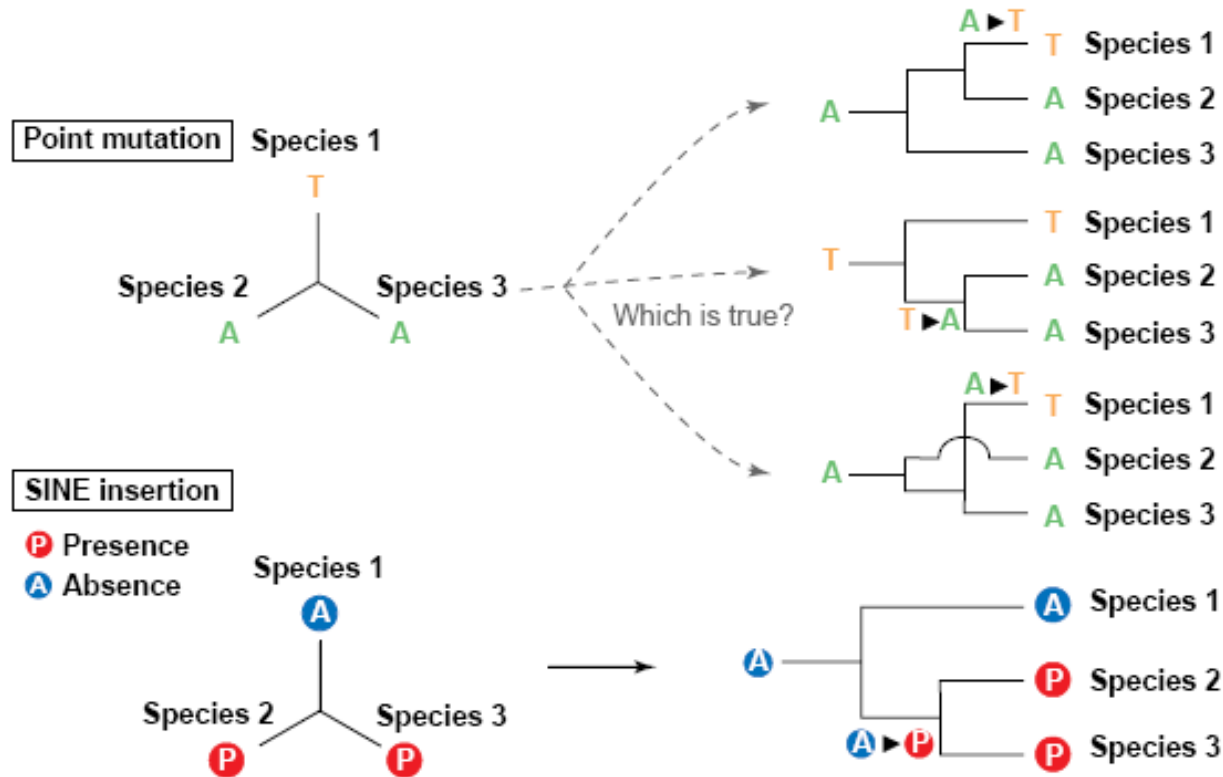


prezence/absence SINE (nikoliv póly elektroforézy)

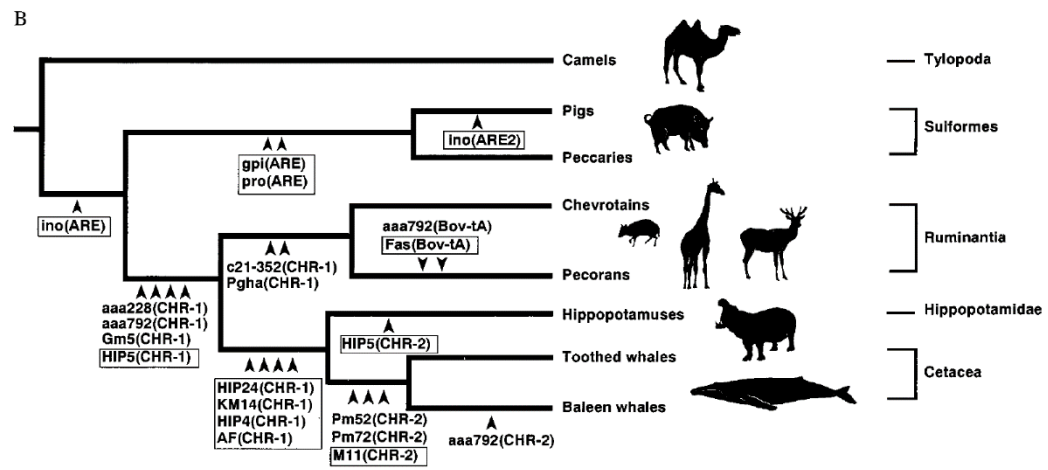
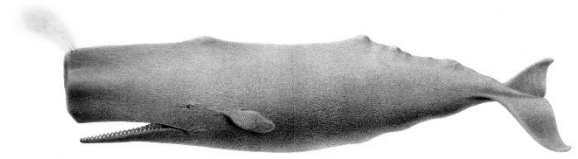
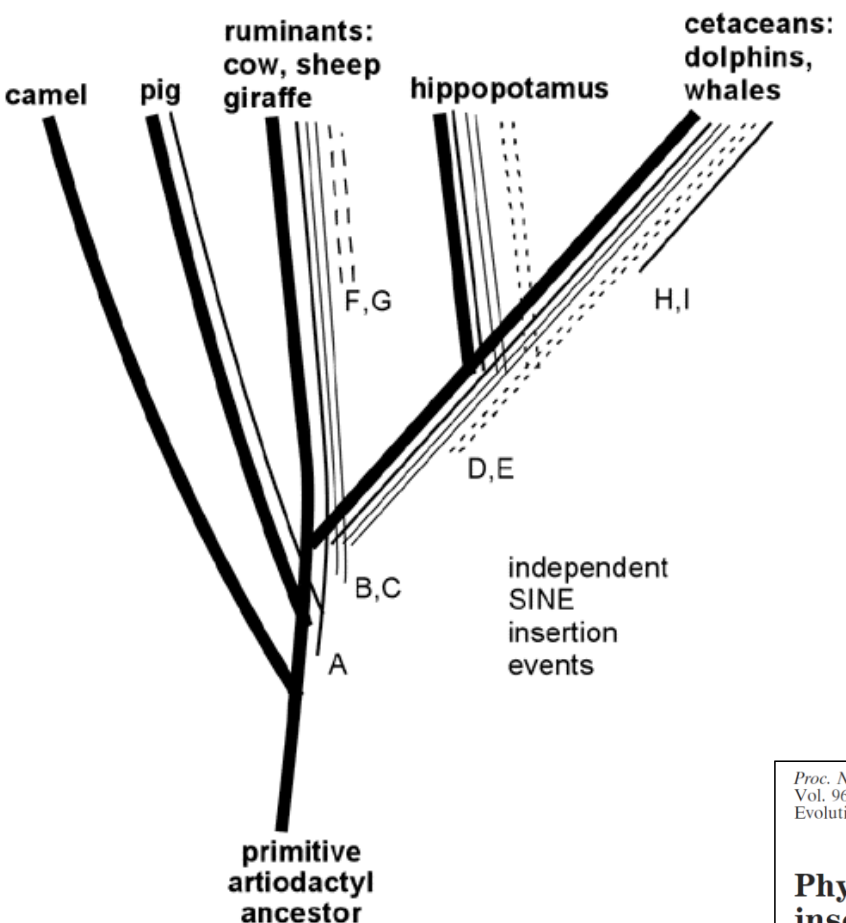
„single-locus marker“

- PCR amplifikace daného úseku a elektroforéza

Neexistují zpětné mutace = výhoda oproti sekvenačním datům



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



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[†], ALEJANDRO P. ROONEY[‡], AND NORIHIRO OKADA^{†§}

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