

„Molekulárně-genetické“ metody

- analýza polymorfismu DNA (genetické markery)
- konzervativní vs. variabilní úseky (« loci »)
- **sekvenční polymorfismus:**

CGCATCTCTAGCTTC**GATTCAGGAA**

CGCATCTCTAGCTTT**GATTCAGGAA**

Sangerovo sekvenování

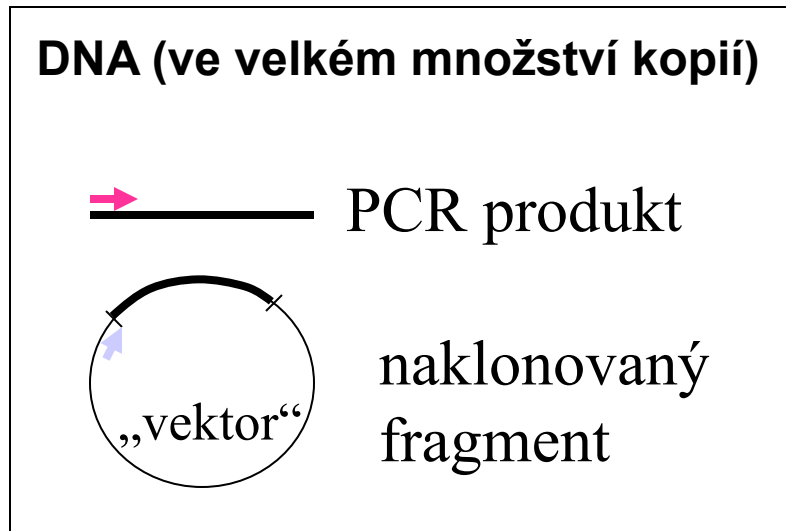
Sekvencování DNA

- ~~Maxam-Gilbertova (chemická) metoda: bázeově-specifická chem. modifikace a štěpení fragmentů DNA~~

- Sangerova (enzymatická) metoda: terminace replikace pomocí ddNTP

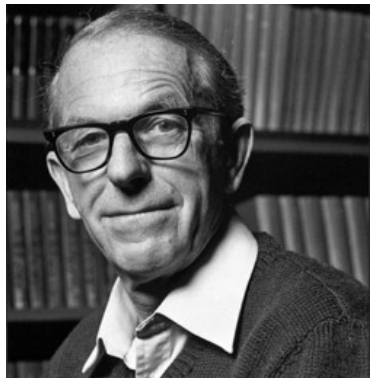
- paralelní „high-throughput“ sekvenování:
= NGS („next-generation sequencing“)

Sangerovo sekvenování DNA



Sekvenační reakce:

- směs standardních nukleotidů a značených dideoxynukleotidů
- **jeden specifický** nebo **universální** primer – poskytuje volný 3 -OH konec



„dideoxy metoda“

Frederick Sanger (1918-2013)

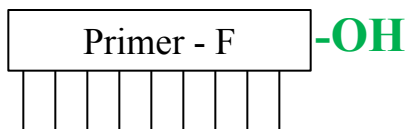
Nobel prize 1958 (struktura inzulinu) a 1980 (sekvenování bílkovin a nukleových kyselin)

- **jen jeden primer**

- vysoká koncentrace templátu (hodně kopií – buď PCR nebo klony v bakteriích)

- směs deoxynukleotidů a fluorescenčně značených dideoxynukleotidů

Primer - F	AAGTCAGTCTAAATGCGATTGGGA	Rev. Primer - R
Rev. Primer - F	TTCAGTCAGATTACGCTAACCCCT	Primer - R



1. Denaturace - 96°C

2. Nasednutí primeru - 50-60°C

Sekvenační „PCR“ - jen jeden primer

Primer - F	AAGTCAGTCTAAATGCGATTGGGA	Rev. Primer - R
------------	---------------------------------	-----------------

Přisedání deoxynukleotidů ...



1. Denaturace - 96°C

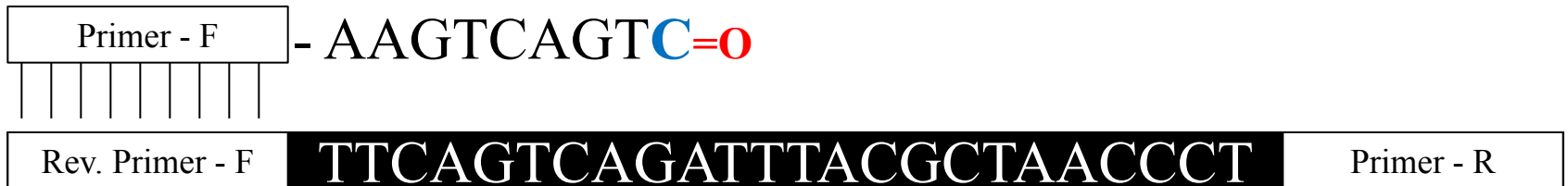
3. Polymerizace - 72°C

2. Nasednutí primeru - 50-60°C

Sekvenační „PCR“ - jen jeden primer

Primer - F **AAGTCAGTCTAAATGCGATTGGGA** Rev. Primer - R

Přisedání deoxynukleotidů ...
... až narazí na dideoxynukleotid



1. Denaturace - 96°C
2. Nasednutí primeru - 50-60°C
3. Polymerizace - 72°C

Sekvenační „PCR“ - jen jeden primer

Primer - F - AAGTCAGTC=O

Primer - F - AAGTCAGTCTAA=O

Rev. Primer - F

TTCAGTCAGATTACGCTAACCT

Primer - R

1. Denaturace - 96°C

2. Nasednutí primeru - 50-60°C

3. Polymerizace - 72°C

35 x

Sekvenační „PCR“ - jen jeden primer

Primer - F - AAGTCAGTC**C=O**

Primer - F - AAGTCAGTCTA**A=O**

Primer - F - AAG**T=O**

Rev. Primer - F

TTCAGTCAGATTACGCTAACCT

Primer - R

1. Denaturace - 96°C

2. Nasednutí primeru - 50-60°C

3. Polymerizace - 72°C

35 x

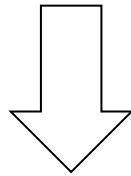
Výsledek sekvenační „PCR“

Primer - F - AAGTCAGTC**C=O**

Primer - F - AAGTCAGTCTAA**A=O**

Primer - F - AAG**T=O**

... a řada dalších fragmentů, každý z nich označený fluorescenčním dideoxynukleotidem



Kapilární elektroforéza

- seřazení fragmentů podle délky
- detekce barvy dideoxynukleotidů

Detekce fragmentů - kapilární elektroforéza

Primer - F - AAGTCAGTCTAA**A**=0

Primer - F - AAGTCAGTCT**A**=0

Primer - F - AAGTCAGTCT**T**=0

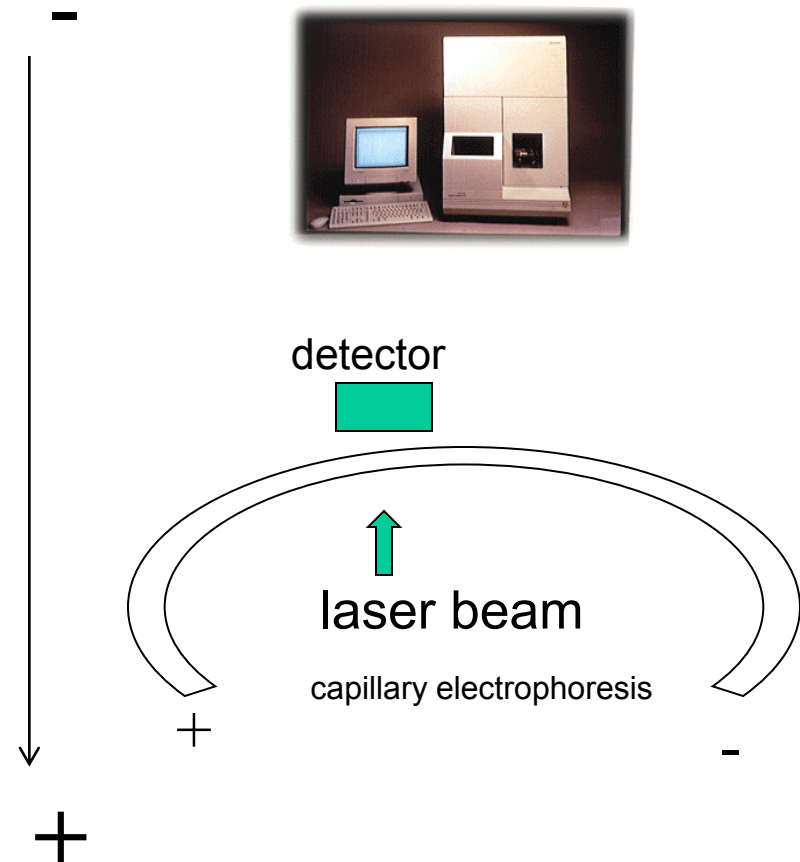
Primer - F - AAGTCAGT**C**=0

Primer - F - AAGTCAG**T**=0

Primer - F - AAGTCAG**G**=0

Primer - F - AAGTC**A**=0

Primer - F - AAGTC**C**=0



Detekce fragmentů - kapilární elektroforéza

Primer - F

- AAGTCAGTCTAA**A**=0 -

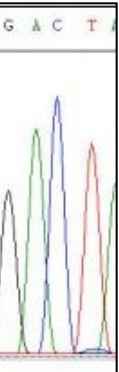
Primer - F

- AAGTCAGTCTA**A**=0

- Sekvence délky 500 – 1000 bp (cca 100 Kč za sekvenci, bez PCR a přečištění PCR produktu)

- 4 kapiláry - destička s 96 vzorky za noc

- Jsou i sekvenátory s 96 kapilárami



Primer - F

- AAGT**C**=0

+

Primer - F

AAGT**CAGTCTAA**ATGCGATTGGGA

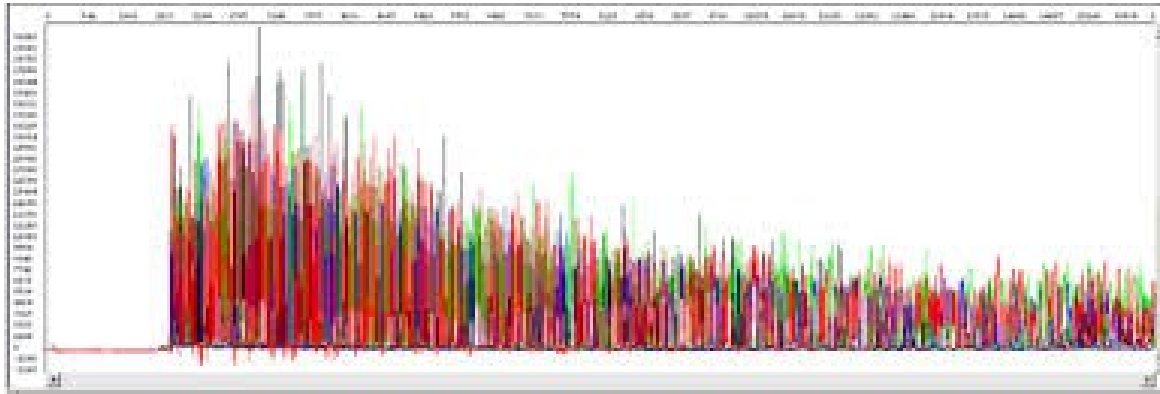
Rev. Primer - R

Rev. Primer - F

TTCAGTCAGATTACGCTAACCCCT

Primer - R

Editace sekvencí



„raw data“ (.ab1 file)

electrophoretogram



„basecalling“ (specializovaný software, např. Geneious)

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

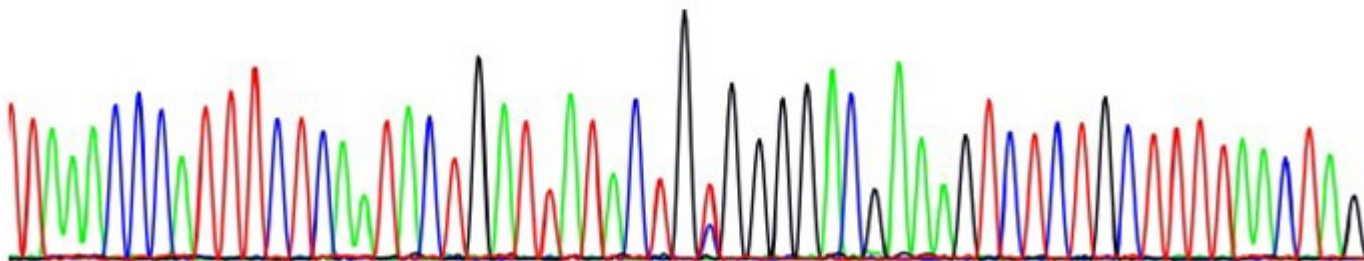
145

157

169

181

193



SNP=T/C

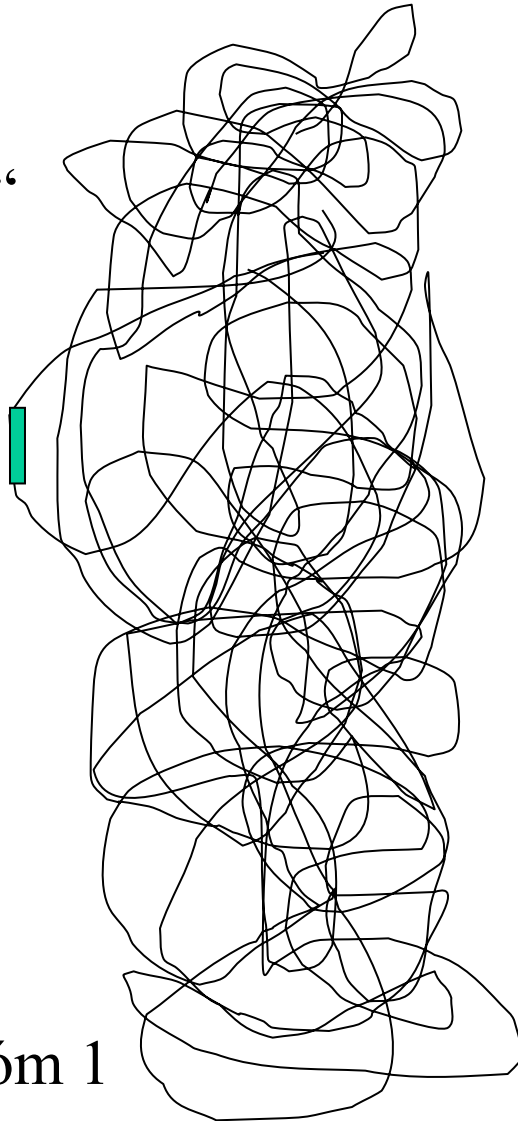
„trimming“
(ořezání konců
s nízkou kvalitou)

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)

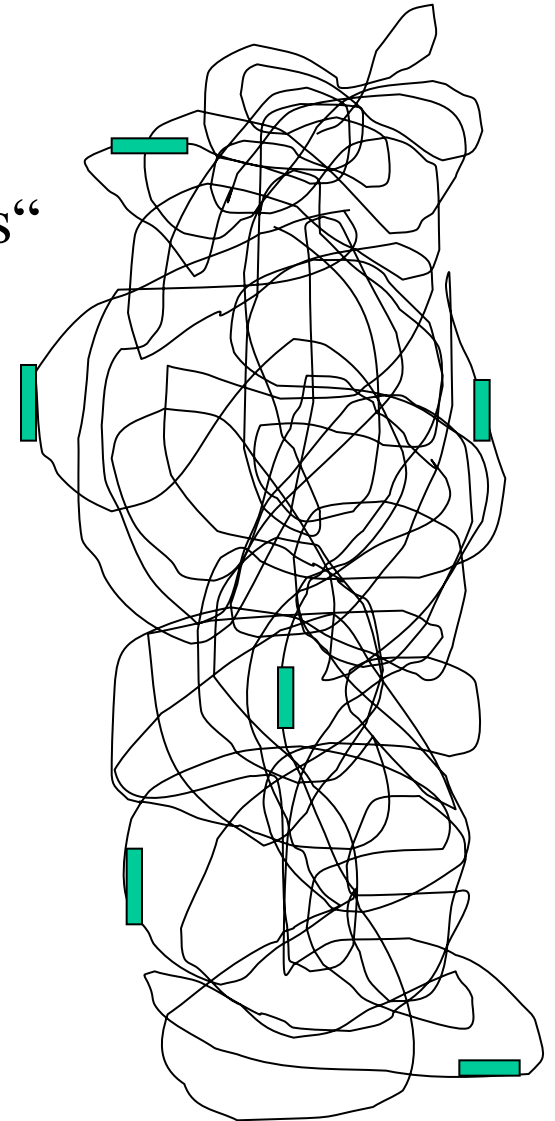
Typy genetických markerů

„single-locus“



Př.: chromozóm 1

„multi-locus“



Typy genetických markerů

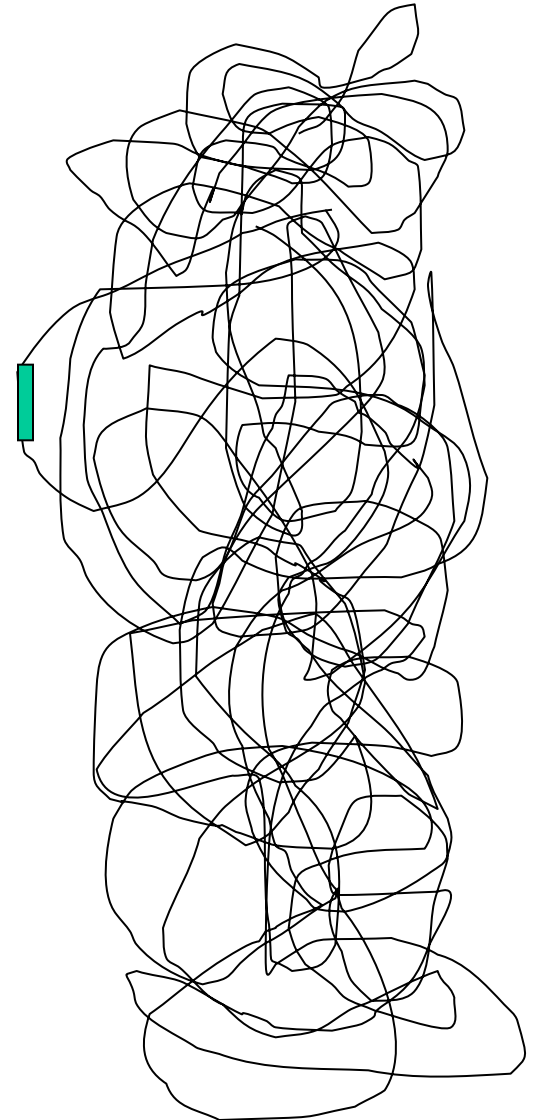
- **dominantní** markery – odliší pouze přítomnost (či nepřítomnost) daného znaku; tj. neodliší obě jeho formy na homologních chromozómech
- **kodominantní** markery – identifikace homologních alel, tj. je možno rozlišit homozygotní a heterozygotní stav (umožňují stanovit frekvenci alel)

Typy genetických markerů

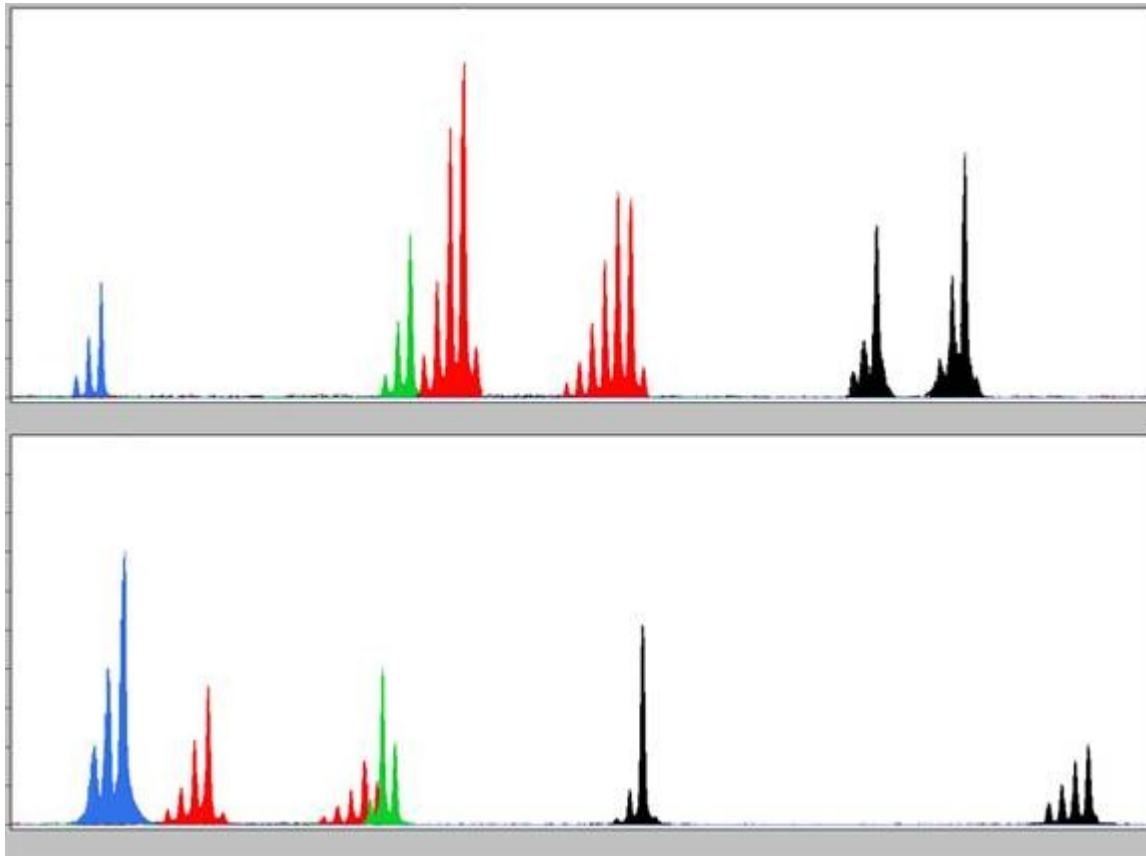
	Single locus	Codominant	PCR	Celková variabilita
Jaderné více-lokusové („nuclear multi-locus“)				
Minisatelitový DNA fingerprints	No	No	No	High
RAPD	No	No	Yes	High
AFLP	No	No	Yes	High
Jaderné jedno-lokusové („nuclear single locus“)				
Alozymy	Yes	Yes	No	Low-medium
Mikrosatelity	Yes	Yes	Yes	High
SINE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

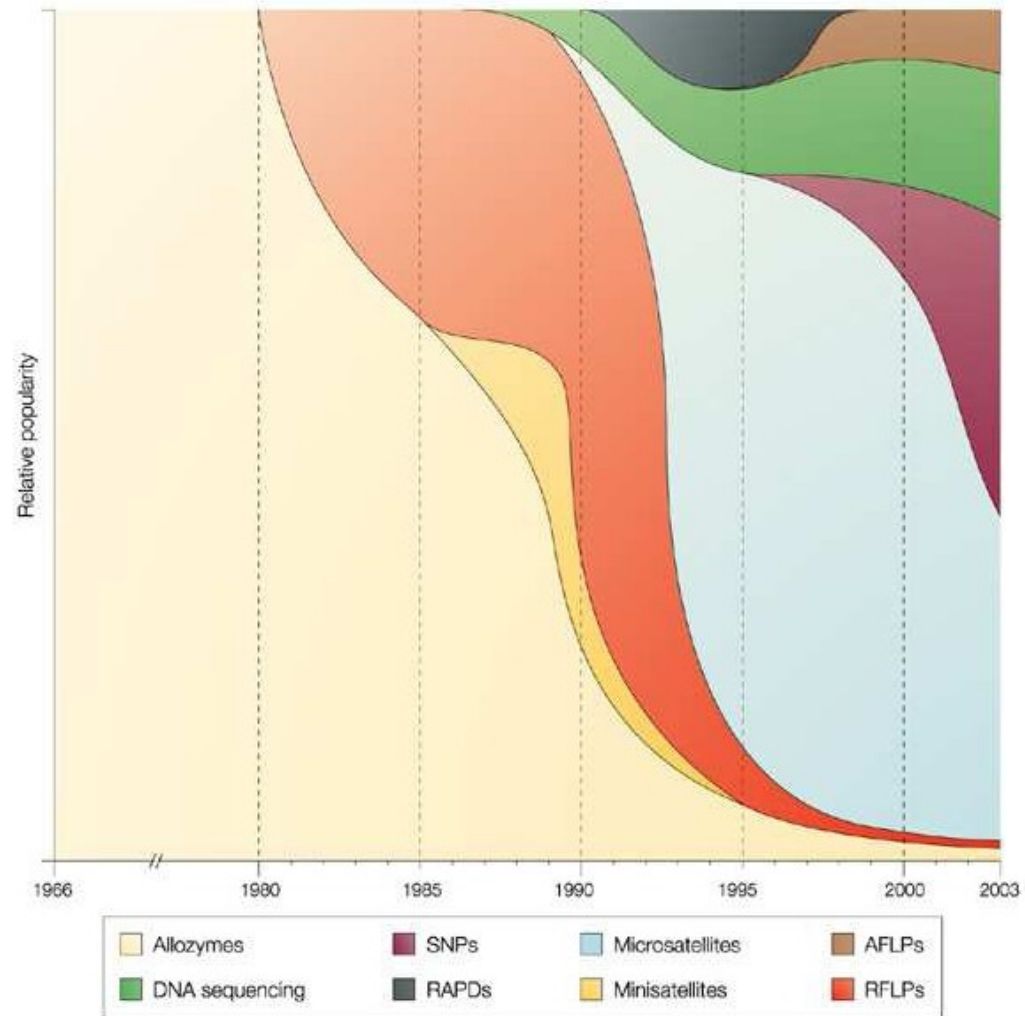


Mikrosatellity



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

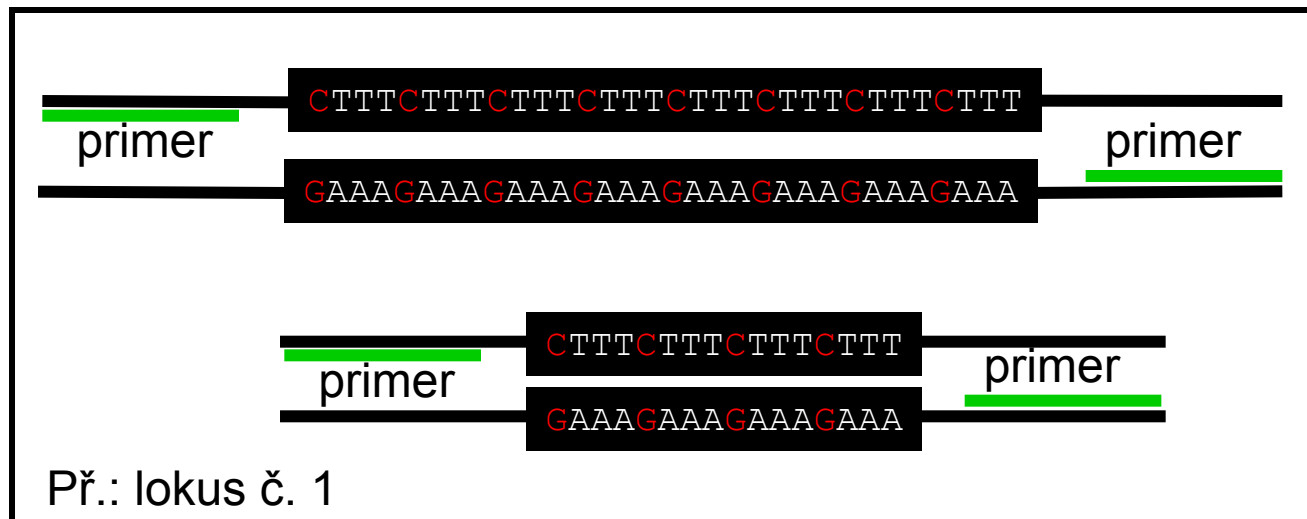
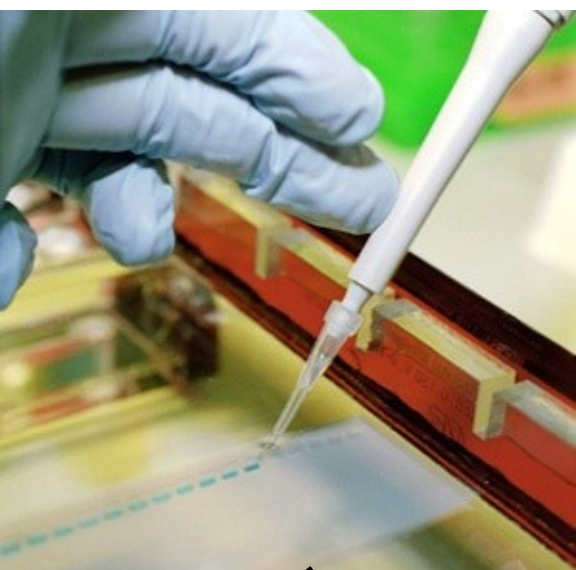
TTCAGGCACACA**ATCTCTAGCTTTGA**

25 bp

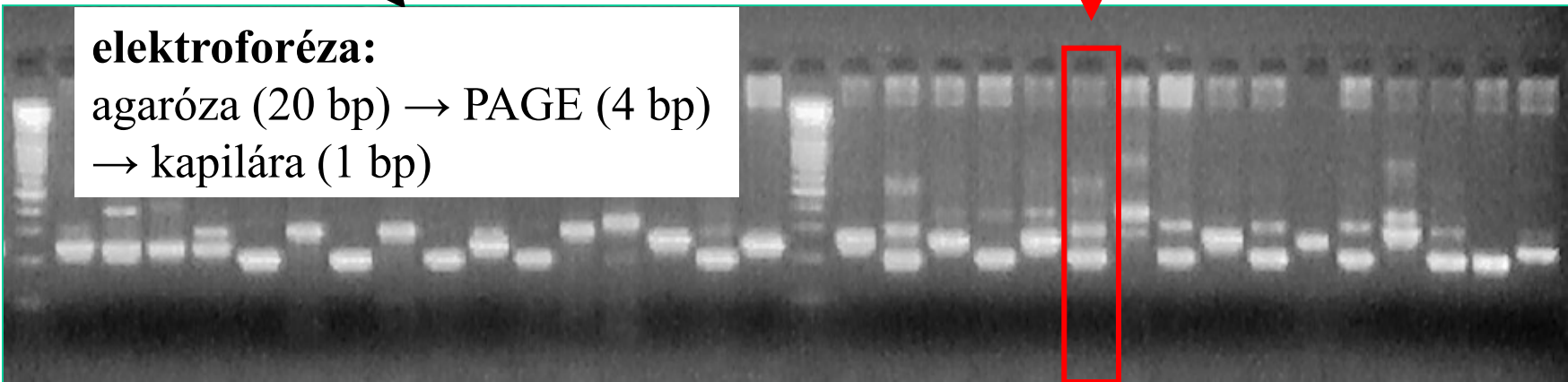
genotyp diploidního jedince: **25/27**

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

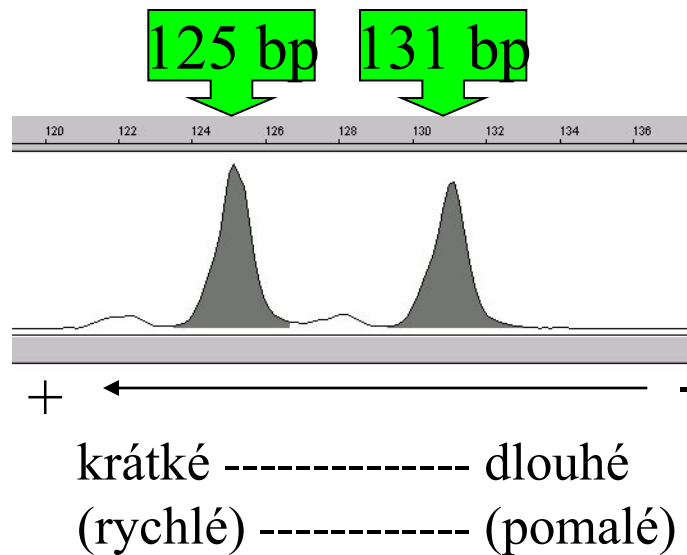


elektroforéza:
agaróza (20 bp) → PAGE (4 bp)
→ kapilára (1 bp)

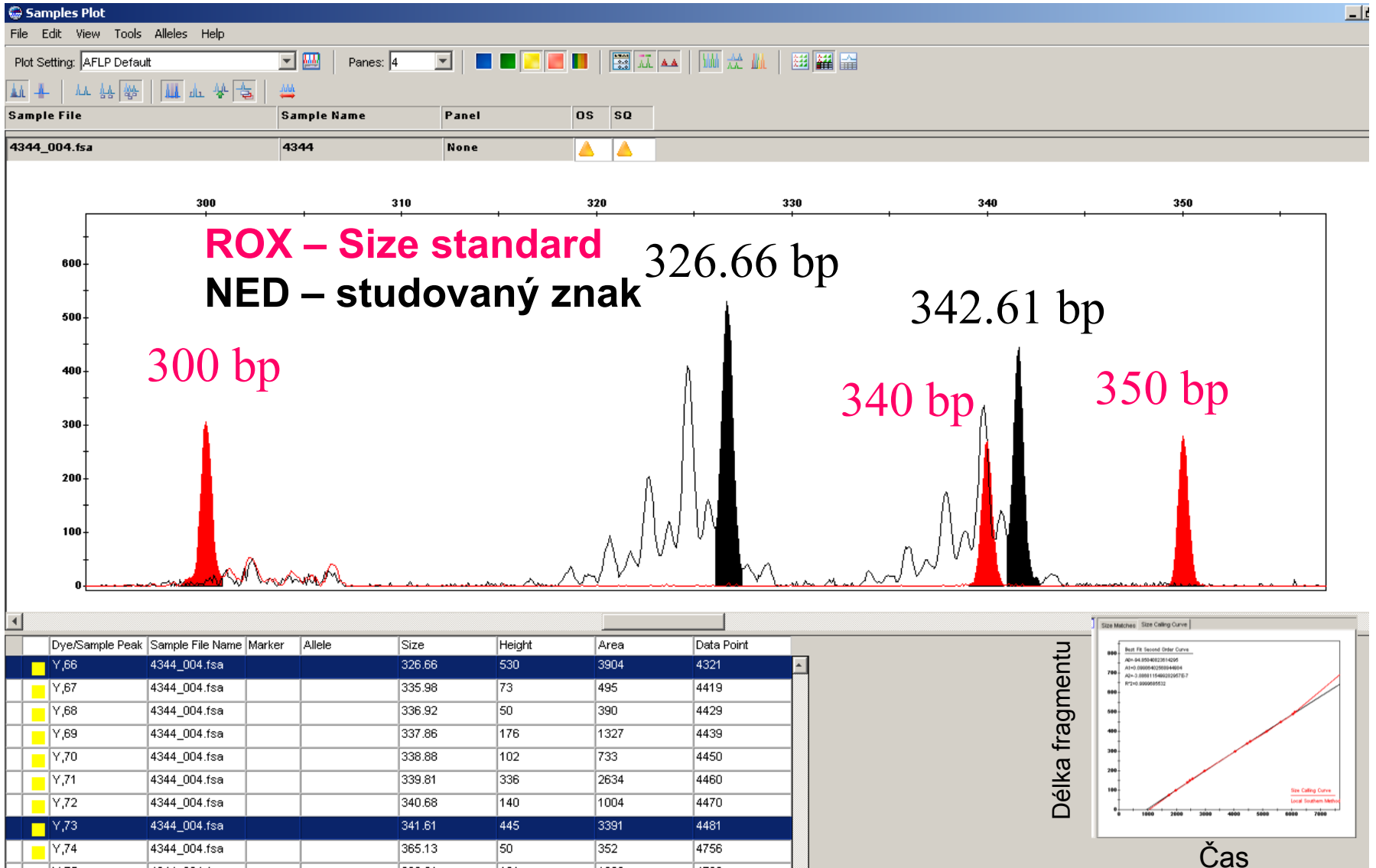


Kapilární elektroforéza ~ Fragmentační analýza

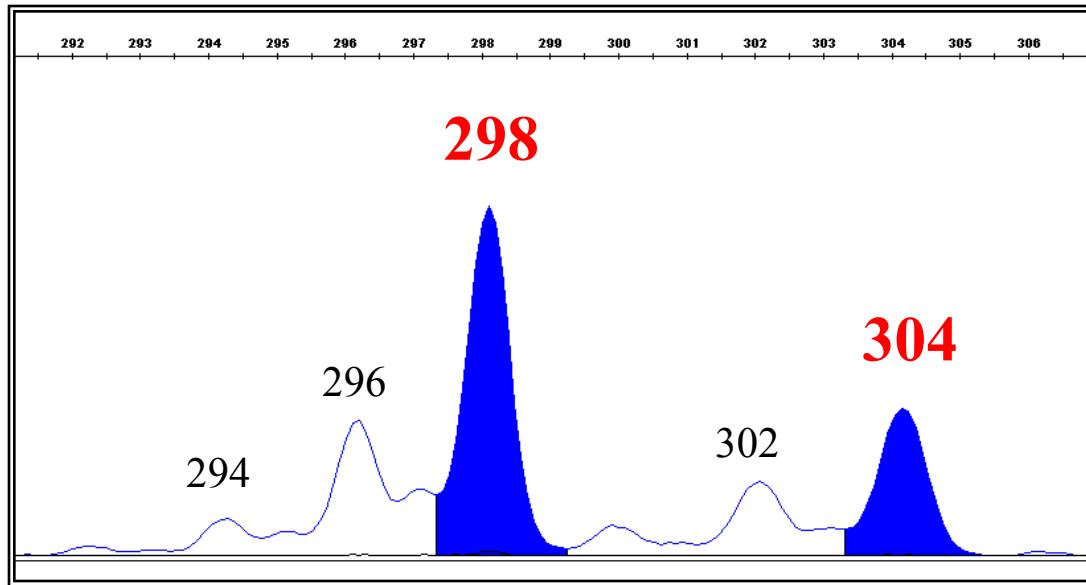
(denaturující polymer POP7 - ssDNA, jeden značený primer)



Well controlled
electrophoresis parameters,
high sensitivity



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

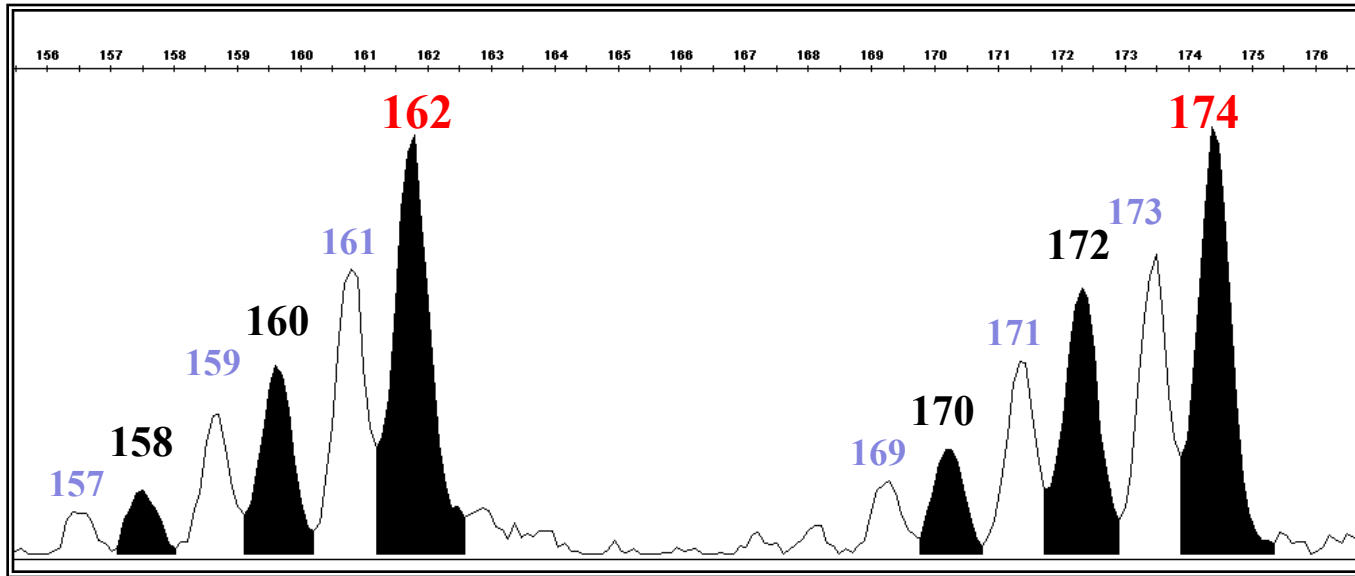


Genotyp 298/304

- „stutters“ – chyby v důsledku „sklouznutí“ polymerázy při PCR
- často odlišují mikrosatelity od nespecifických PCR produktů
 - rozdíl mezi alelou a „stutter“ je délka repetice (zde 2 bp)

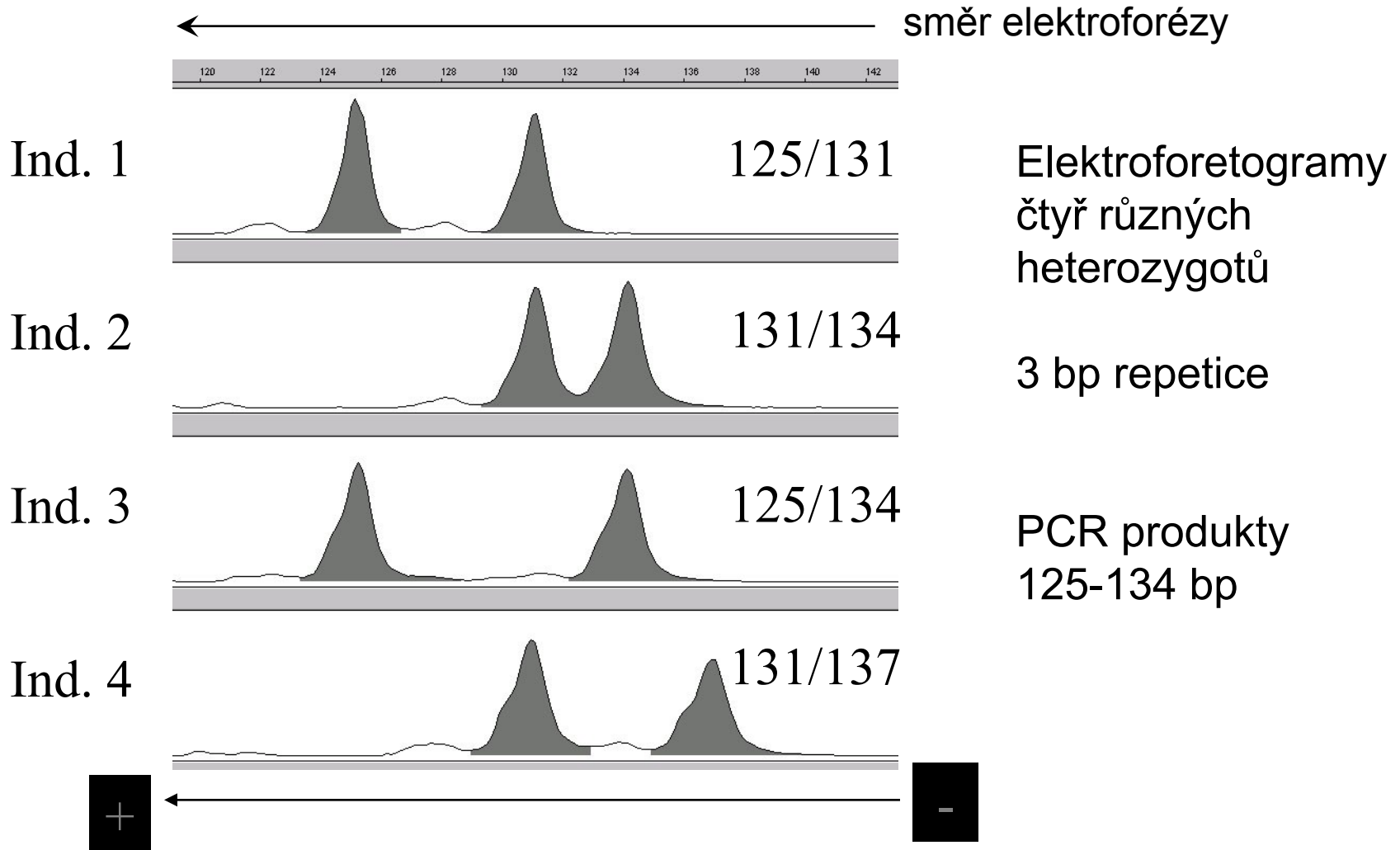
CT

Genotyp 162/174

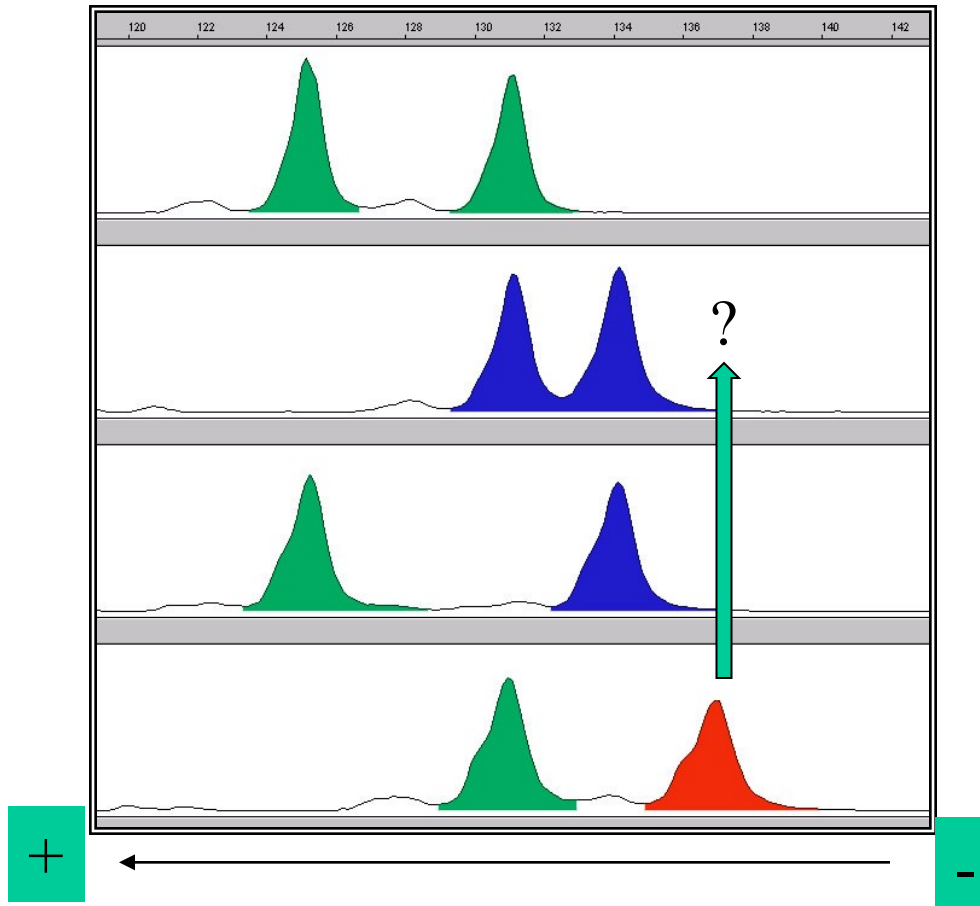


- alely a jejich stuttery jsou černě (rozdíl mezi nimi je 2 bp)
- bílé píky jsou tzv. „minus A-alely“ a jejich stuttery = výsledek jiné chyby polymerázy, a to nepřidání koncového adeninu
- rozdíl mezi černým a sousedním bílým píkem je 1 bp (tj. chybějící adenin)
- pattern daného lokusu je vždy specifický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

Srovnání různých jedinců - analýzy příbuznosti



Př. Analýza příbuzenských vztahů



Genotyp (bp)

Matka: 125/131

Otec: 131/134

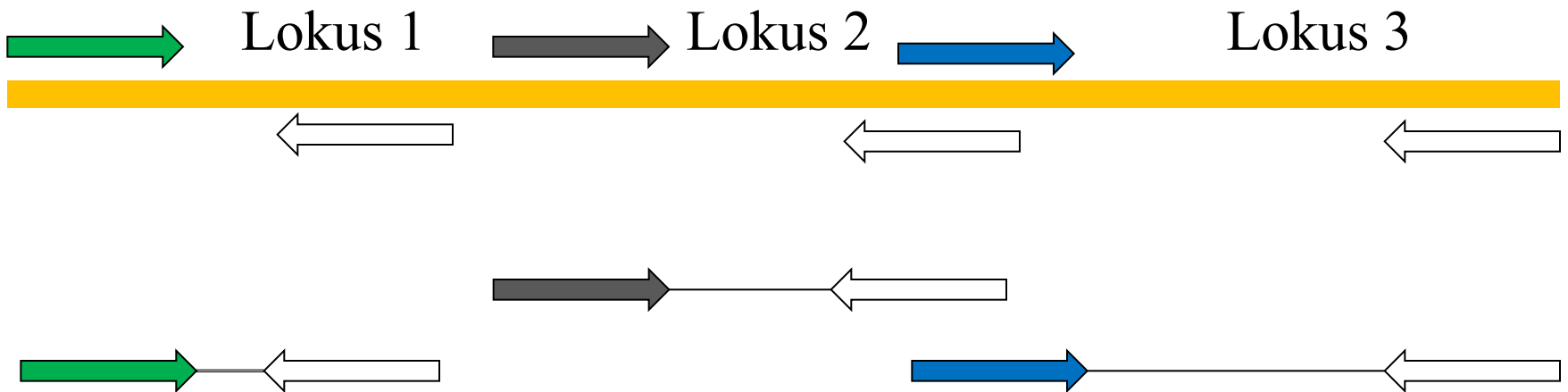
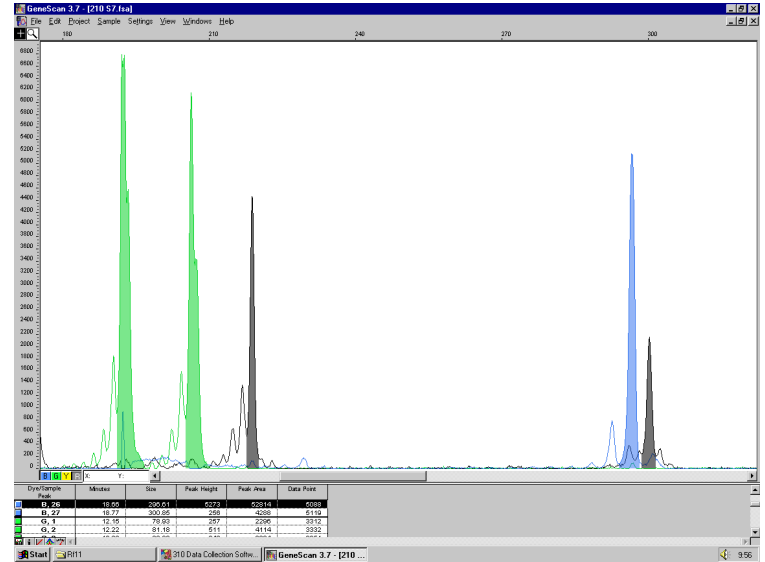
Potomek 1: 125/134

Potomek 2: 131/137

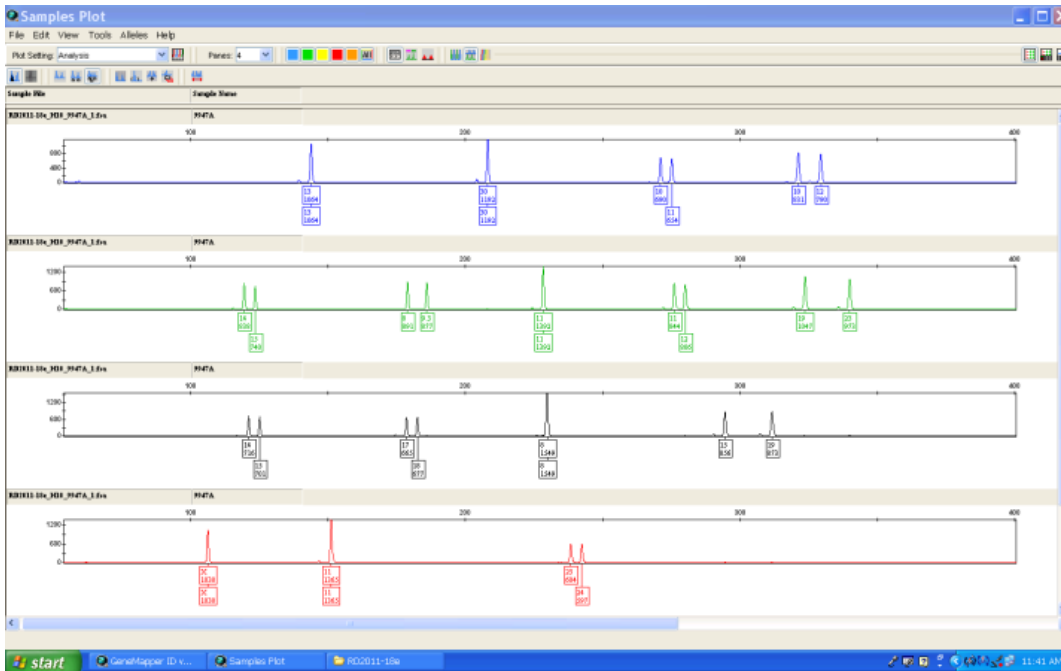
Sledovaný otec mohl zplodit potomka 1, ale zcela jistě není otcem potomka 2

Různé značení různých znaků

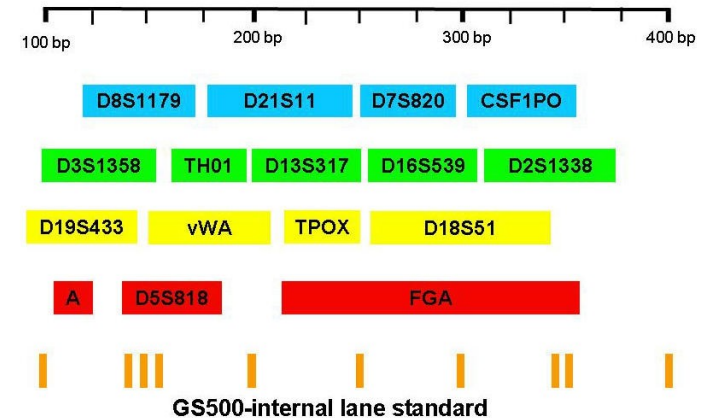
- Snížení časových a finančních nákladů
- = „multiplex set“
- Až 4 různé barvy (+ 5. barva jako velikostní standard) - analýza až 4 lokusů o stejné velikosti alel



Identifikace jedinců u lidí



AmpFSTR® Identifier™



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



Mikrosatelity - omezení

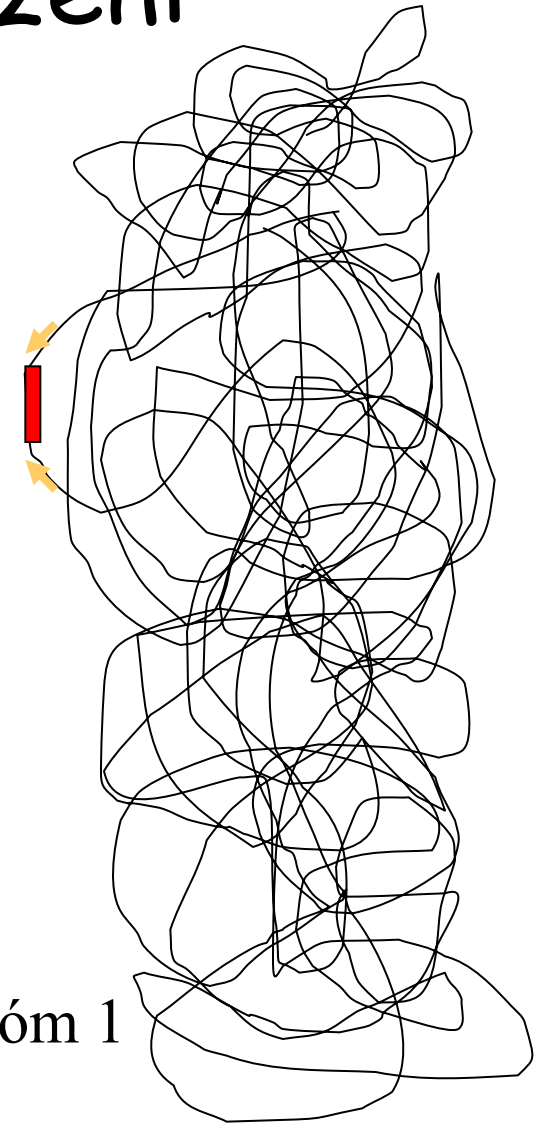
- nalezení lokusů (navržení primerů) je pracné a nákladné u volně žijících druhů (genomová knihovna, klonování, screening, sekvencování)



TTCAGG**CACACACA**TCTCTAGCTTCGA



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

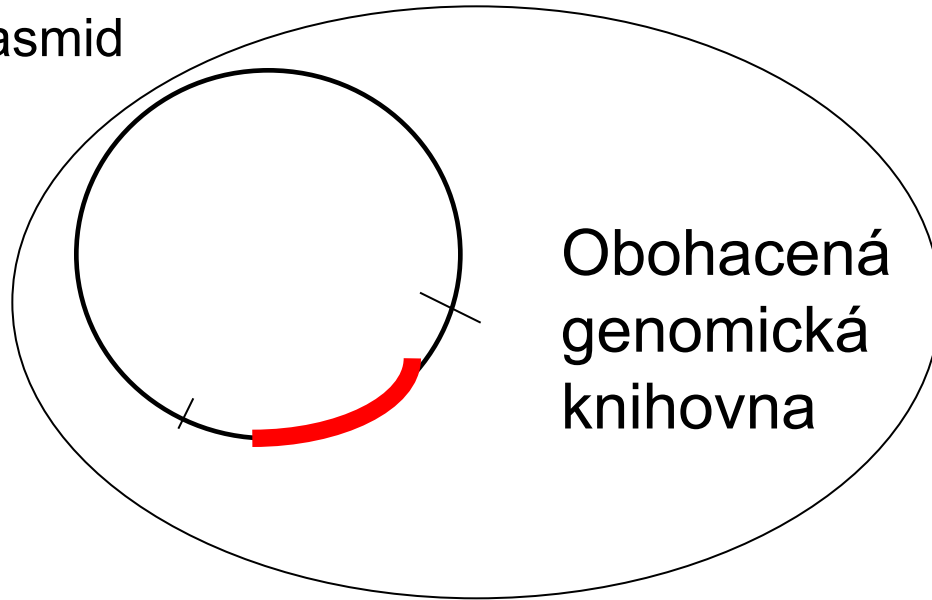


Př.: chromozóm 1

Štěpení, obohacení, klonování, screening a sekvenování

Každý klon obsahuje jednu sekvenci

vector =
plasmid



izolace vektorů s
inzertem



screening klonů obsahujících
repetice (hybridizace se sondou)



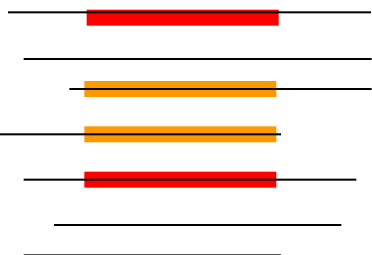
sekvenování inzertů
(repetitivní DNA + flanking regions)



design primerů a testování
polymorfismu



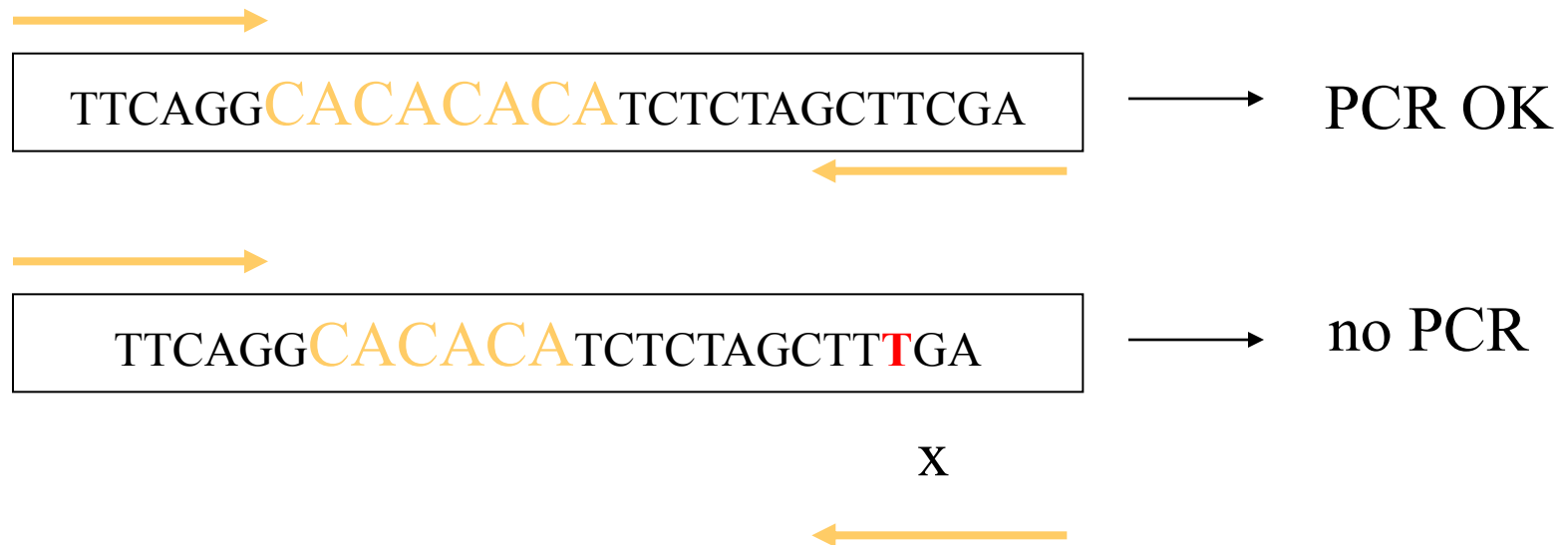
ligace do plasmidu, transformace



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“



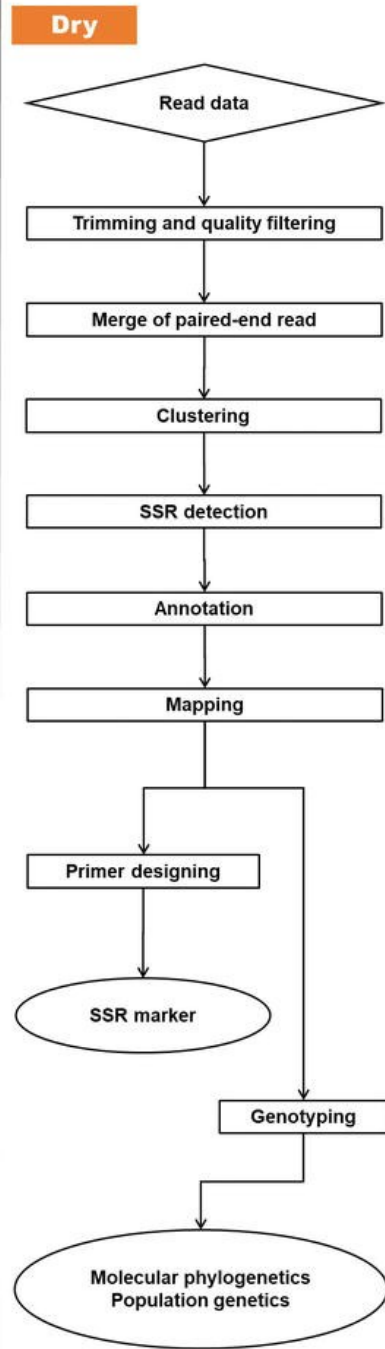
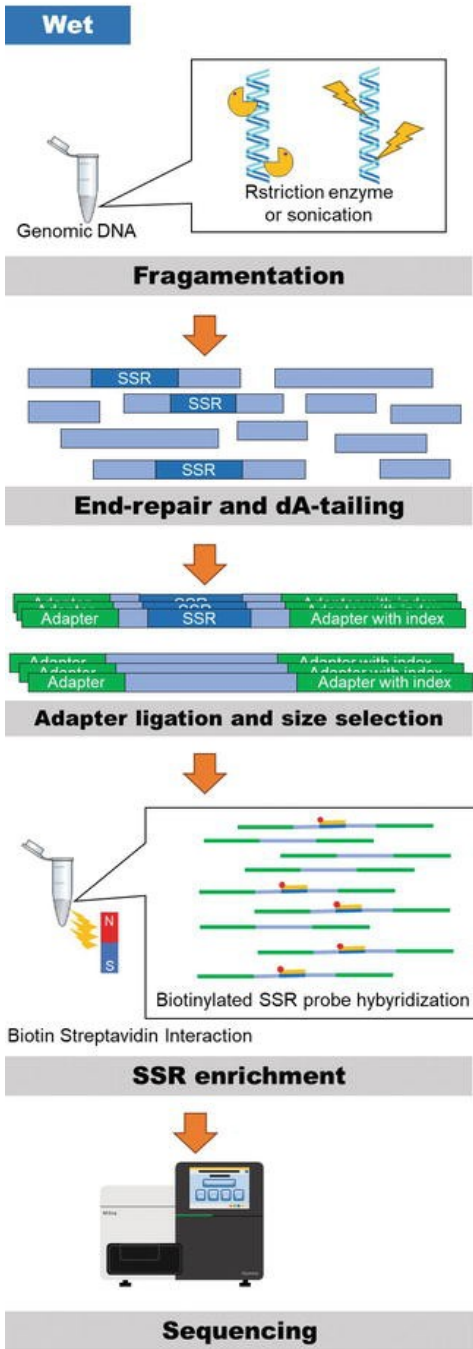
Nulové alely a genotypizační chyby komplikují analýzy příbuznosti

	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

Experts in microsatellite development.

Microsatellite markers for your species with our microsatellite development service

AllGenetics' microsatellite development service uses high-throughput sequencing to obtain primer pairs which amplify polymorphic microsatellite loci in your study species. The primers obtained are multiplexed and tested for polymorphism in a number of individuals from different populations.

1

>

2

>

3

aatgtc**cgccgcccggcgggcgggcggg**taaaggagt
 ccag**ttcattcattcattcattcattcattc**atgtcagggta
 agtct**gaggaggaggaggaggaggaggaggaggaggagg**ataaatt
 atal**aacaacaacaacaacaacaacaacaacaac**gtacga
 tagtg**gatcgatcgatcgatcgatcgatcgatc**ttagagt
 atcgaagt**tcttcttcttcttcttcttcttcttctt**ctgtagttat



Pricing information

Please refer to the tables below for the prices of the different steps.

Step 1	1 species	2 species	3 species	4 species	5 species
	€ 1,565	€ 1,465	€ 1,370	€ 1,276	€ 1,181

Step 2	Primer pairs	7 individuals	11 individuals	15 individuals
	36	€ 2,833	€ 3,735	€ 4,635
	48	€ 3,675	€ 4,875	€ 6,073
	72	€ 5,355	€ 7,152	€ 8,947
	108	€ 7,872	€ 10,565	€ 13,282
144	€ 10,388	€ 13,995	€ 17,710	

Step 3	Primer pairs	3 individuals
	9	€ 399
	12	€ 475
	18	€ 610
	27	€ 795
	36	€ 973

Další firmy:
 GenoScreen
 Ecogenics
 StarSeq

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)

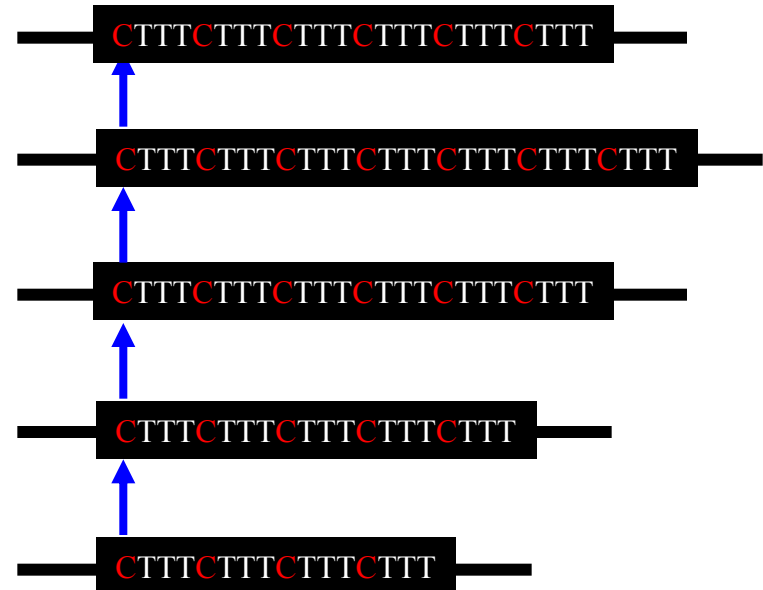
- **IAM – infinitive allele model**

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

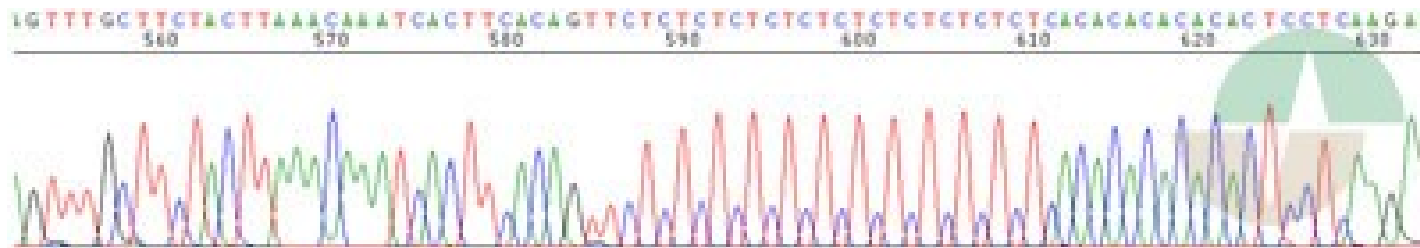
27 → 26 bp

TTCAGG CACACACA TCTCTAGTTCGA

27 → 26 bp

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

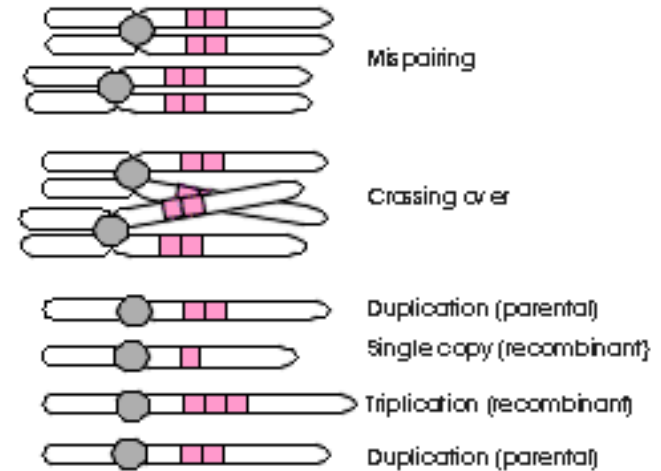
Další nepravidelnosti (tj. možné komplikace při analýzách)



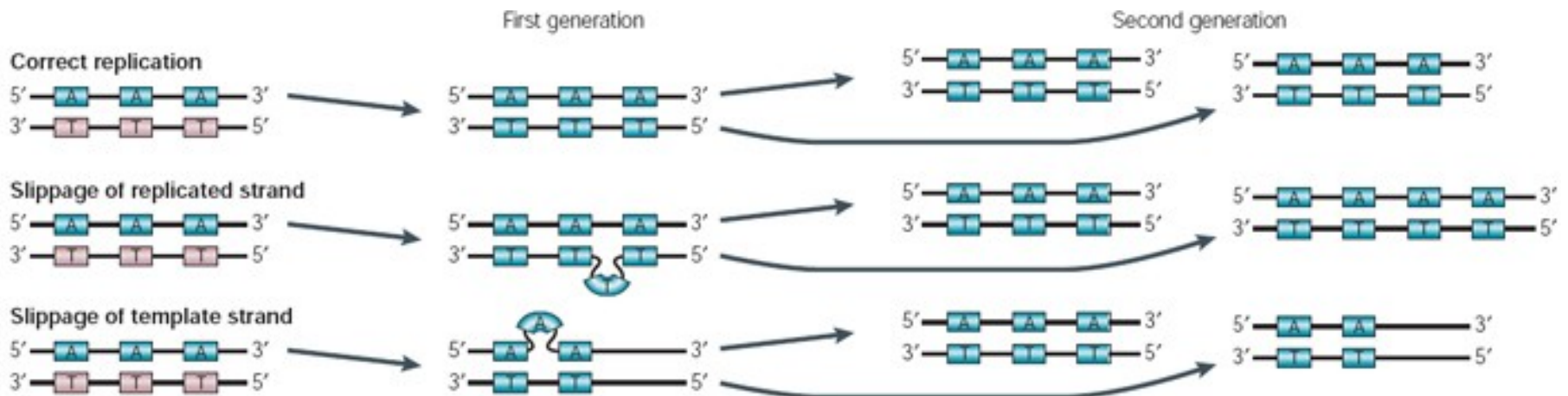
„složený mikrosatelit“ – rovněž není možné aplikovat jednoduchý mutační model

Proč je tolik alel? (microsatellite instability)

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



- **Skouznutí 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** (náchýlnější 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