

„Propojení výuky oborů Molekulární a buněčné biologie a Ochrany a tvorby životního prostředí“

Reg. č.: CZ.1.07/2.2.00/28.0032



**evropský
sociální
fond v ČR**



EVROPSKÁ UNIE



**MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY**



**OP Vzdělávání
pro konkurenceschopnost**

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Molekulární markery II.

mikrosatelity, sekvenace, NGS, metody odvozené od NGS

Miloslav Kitner
Katedra botaniky PŘF UP



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

MIKROSATELITY

MIKROSATELITY

Princip:

Detekujeme rozdíly v délce krátkých tandemově se opakujících oblastí DNA

Jedinec A ...GTTCTGTCATATATATATATATCGTACTT...

Jedinec B ...GTTCTGTCATATATATATATATATATATCGTACTT...

krátké tandemové repetice, STR (short tandem repeats)

repetice jednoduchých sekvencí, SSR (simple sequence repeats)

- krátké, tandemově se opakující jednoduché sekvenční motivy zpravidla o délce 2-6bp
 - prokaryota a eukaryota
 - kódující i nekódující oblasti
 - Vysoká mutační rychlost (u savců 10^{-3} až 10^{-4} /lokus/generaci)
- hlavní zdroj vysoké proměnlivosti - sklouznutí nukleotidového řetězce během replikace (replication slippage)
- Alely se liší svou **délkou**.....
 - Snadná separace pomocí elektroforézy

MIKROSATELITY

– **Dokonalé = perfect repeats** (jeden nepřerušný motiv):

- Jednonukleotidové: ...AAAAAAAAAAAAAAAAAAAAA...
- Dinukleotidové: ...CACACACACACACACACA... = (CA)_n
- Trinukleotidové: ...CGTCGTCGTCGTCGTCGTCGT... = (CGT)_n
- Tetranukleotidové: ...CAGACAGACAGACAGACAGA... = (CAGA)_n
- Pentanukleotidové: ...AAATTAAATTAAATTAAATT... = (AAATT)_n
- Hexanukleotidové: ...CTTTAACTTTAACTTTAACTTTAA... = (CTTTAA)_n
 - Pr: ...(AG)₃₂...; ...(TAT)₂₅...; ...(CAA)₇... *Cicer*

– **Nedokonalé = imperfect repeats** (motiv je přerušen jedním nebo několika bazemi):

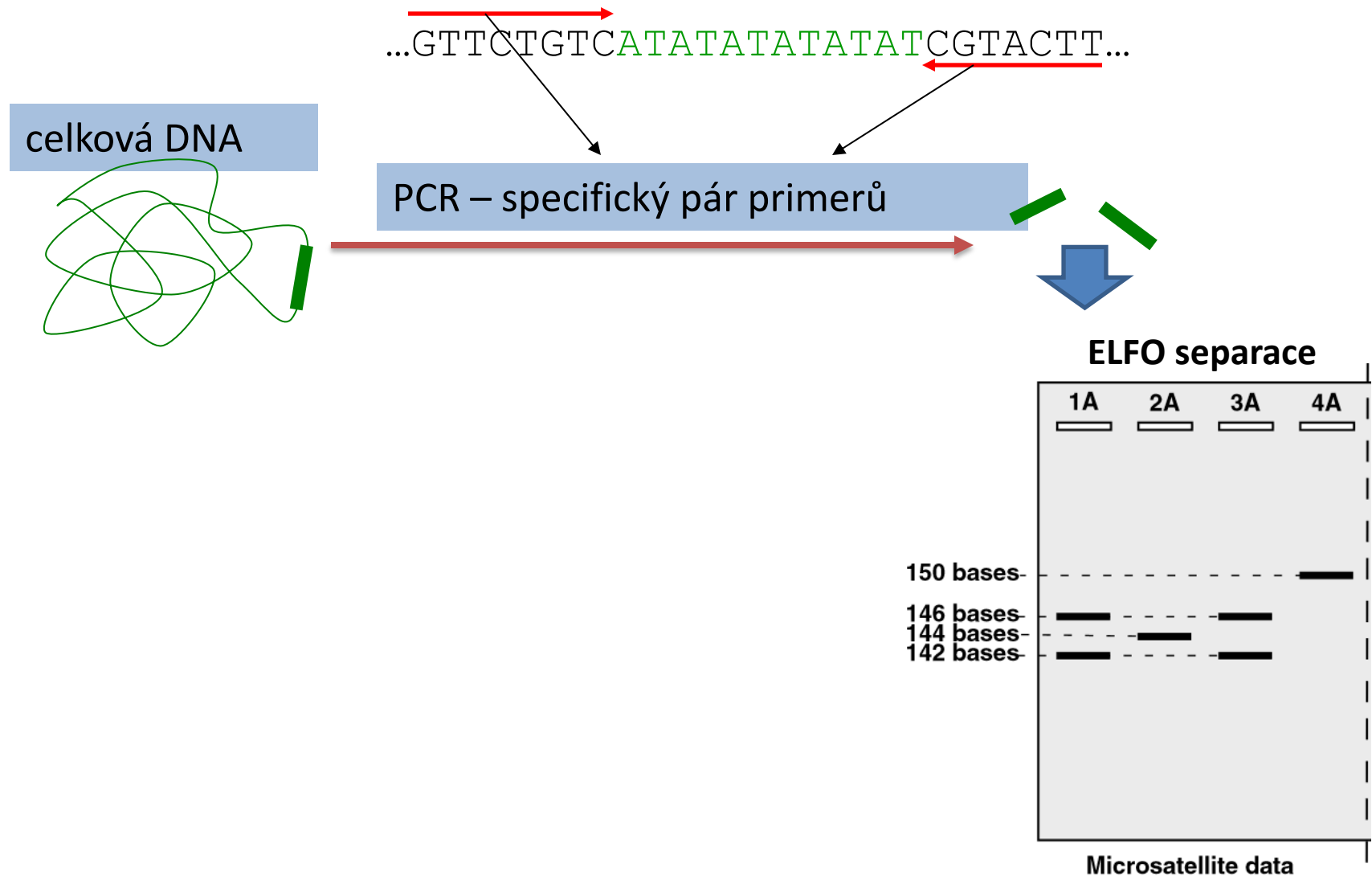
- ...(TC)₆A(TC)₁₃...; ...(AG)₁₂GG(AG)₃... *Cicer*

– **Složené = compound**: (směs dokonalých nebo nedokonalých vzorů několika motivů)

- ...(AT)₆(GT)₄₂AT(GT)₅(GT)₁₀...
- ...(AT)₁₄(AG)₈...
- ...(GAA)₂₁...(TA)₂₃

- **Nejčastější motivy:** mono a dinukleotidové; 3, 4 a 5ti méně

MIKROSATELITY



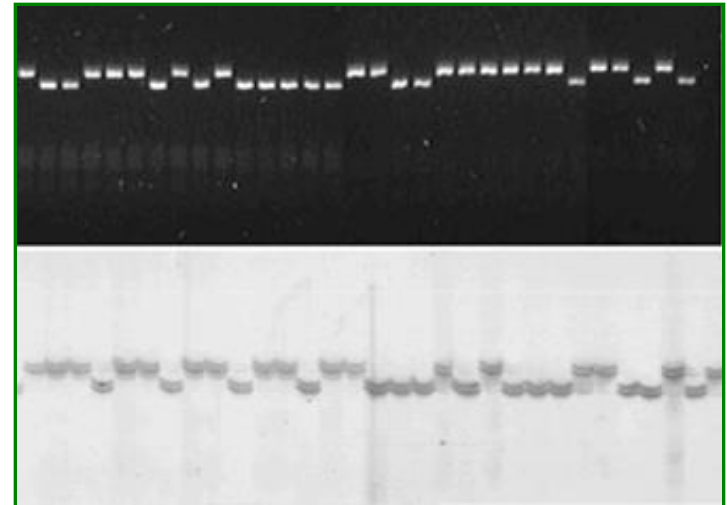
MIKROSATELITY

Výhody

- vysoká variabilita
- velká početnost a rozmístění po celém genomu
- jednoduchost analýzy (mikrosatelity lze poměrně snadno studovat pomocí PCR)
- robustní a reprodučibilní markery
- KODOMINANTNÍ MARKER, IDENTIFIKACE ALEL

Nevýhody

- Musíme znát sekvence primerů
- Pokud neznáme, tak testovat SSR z příbuzných druhů
- Optimalizace PCR
- Popřípadě vyvíjet nové



MIKROSATELITY - využití

jaderné mikrosatelity (SSRs)

- nejlepší marker pro zhodnocení variability na populační úrovni
- jsou potřeba druhově specifické primery, tj. metoda použitelná pouze v případě, že pro studovaný druh byly primery již publikovány

chloroplastové mikrosatelity (cpDNA SSRs)

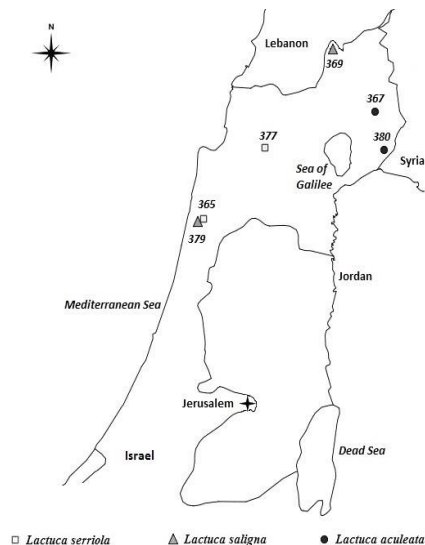
- vhodné pro hodnocení variability na úrovni příbuzných druhů, někdy i na vnitrodruhové úrovni

Obecně:

- Forezní genetika: kriminalistika, identifikace jedinců, příbuzenské vztahy
- analýza rodičovství (*parentage analysis*)
 - určení rodičů semen (semenáčků) v populacích
- identifikace klonů
- populačně-genetické studie
 - genový tok, migrace
 - historie populací



MIKROSATELITY – příkladová studie



Studium genetické variability populací tří druhů *Lactuca* spp. vyskytujících se v centru diverzity druhu *Lactuca* (Izrael)

- *L. serriola*
- *L. saligna*
- *L. aculeata*
- Samosprašné druhy – jejich populace by měly být víceméně uniformní
- Je to správný předpoklad?



L. serriola

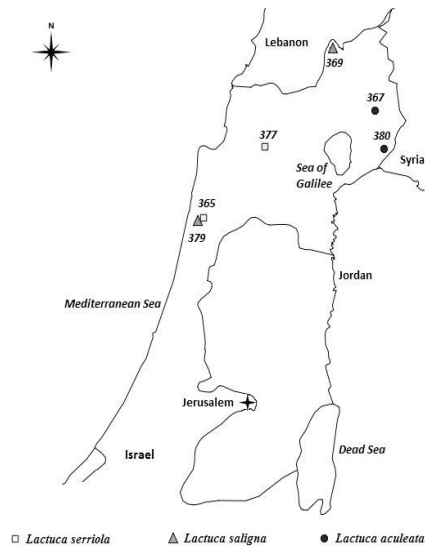


L. saligna



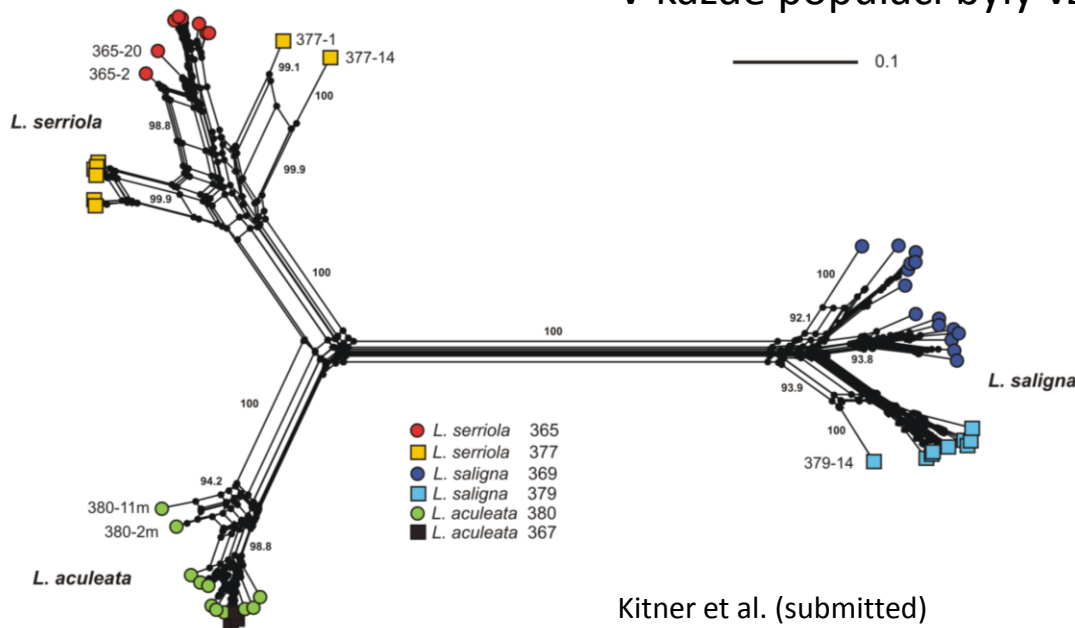
L. aculeata

MIKROSATELITY – příkladová studie



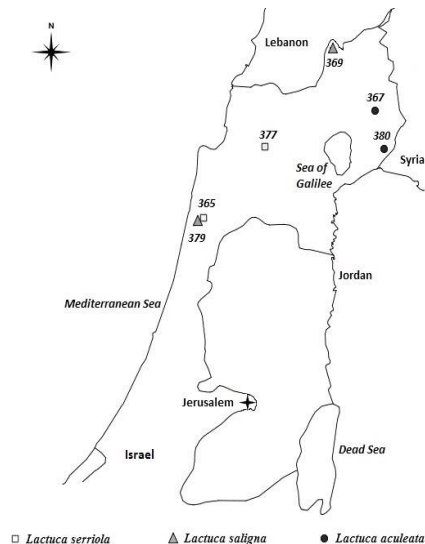
Studium genetické variability populací tří druhů *Lactuca* spp. vyskytujících se v centru diverzity druhu *Lactuca* (Izrael)

- *L. serriola*
- *L. saligna*
- *L. aculeata*
- Samosprašné druhy – jejich populace by měly být víceméně uniformní.
- Je to správný předpoklad?
- Metodika: 11 SSR + 230 AFLP
- V každé populaci byly vzorky odebrány v lineárním transektu



- U některých druhů ano (*L. serriola*, *L. aculeata*).
- *L. saligna* – velice variabilní, každý jedinec představuje unikátní genotyp

MIKROSATELITY – příkladová studie



Studium genetické variability populací tří druhů *Lactuca* spp. vyskytujících se v centru diversity druhu *Lactuca* (Izrael)

- *L. serriola*
- *L. saligna*
- *L. aculeata*
- Samosprašné druhy – jejich populace by měly být víceméně uniformní.
- Je to správný předpoklad?
- Metodika: 11 SSR + 230 AFLP
- V každé populaci byly vzorky odebrány v lineárním transektu

L. serriola

Locus	SML-001		SML-019		SML-039		SML-045		SML-055		WSULs_18		WSULs_25		WSULs_45		WSULs_75		WSULs_153		WSULs_163	
377-1	195	195	170	170	238	238	229	229	228	228	244	244	370	370	189	189	184	184	194	194	188	188
377-2	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-5	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-6	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-7	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-8	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-9	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-10	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-11	195	195	163	163	232	232	229	229	230	230	240	240	360	360	184	184	187	187	194	194	194	194
377-12	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	192	196	194	194
377-13	195	195	163	163	232	232	229	229	230	230	240	240	360	360	189	189	187	187	194	194	194	194
377-14	195	195	170	170	238	238	229	229	228	228	244	244	370	370	175	175	184	184	194	194	188	188

- U některých druhů ano (*L. serriola*, *L. aculeata*).
- *L. saligna* – velice variabilní, každý jedinec představuje unikátní genotyp
- U několika populací se ukázalo, že okraje populací jsou významným zdrojem variability – přítomnost jedinců nesoucích unikátní alely, které nejsou přítomny u jedinců uvnitř populací
- Tyto alely byly přítomny u:
 - jedinců pocházejících z jiných populací
 - mezidruhových hybridů

FROM THE COVER

Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation

VERONICA NYSTRÖM,*†‡ JOANNE HUMPHREY,§ PONTUS SKOGLUND,¶ NIAL J. McKEOWN,§** SERGEY VARTANYAN,†† PAUL W. SHAW,§** KERSTIN LIDÉN,† MATTIAS JAKOBSSON,¶‡‡ IAN BARNES,§ ANDERS ANGERBJÖRN,* ADRIAN LISTER§§ and LOVE DALÉN†§

*Department of Zoology, Stockholm University, S-106 91 Stockholm, Sweden, †Archaeological Research Laboratory, Stockholm University, S-106 91 Stockholm, Sweden, ‡Department of Molecular Systematics, Swedish Museum of Natural History, S-104 05 Stockholm, Sweden, §School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK, ¶Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, S-752 36 Uppsala, Sweden, **Institute of Biological, Environmental and Rural Sciences (IBERS), Aberystwyth University, Penglais, Aberystwyth SY23 3DA, UK, ††Northeast Interdisciplinary Research Institute, Far East Branch, Russian Academy of Sciences, 16 Portovaya St., Magadan 685000, Russia, ‡‡Science for Life Laboratory, Uppsala University, S-751 23 Uppsala, Sweden, §§Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK

Table 1 List of samples used in this study. Samples are shown with their radiocarbon laboratory number and age in radiocarbon years, material used for DNA analysis, mtDNA haplotype and individual heterozygosity (values in brackets denote incomplete genotypes). Samples marked with an asterisk (*) indicate samples from Wrangel Island dated to the time period when Wrangel was part of the Beringian landmass. The number of successful genotypes for each specimen is shown for each of the four loci: EMU10 (1), EMU13 (2), FH67 (3) and FH71 (4)

¹⁴ C laboratory no.	¹⁴ C age	Material	Region	Locality	mtDNA haplotype	Loci scored	Individual heterozygosity
OxA-20032 (P23349)	>61 600	Tooth	Chukotka	Krugloe Lake	W11	1,2,3,4	1
OxA-20031 (P23348)	>55 700	Tusk	Chukotka	Ayon Island	W20	1,2,3,4	0.5
OxA-20050 (P23366)	>52 000	Tusk	Chukotka	Rakvachan River basin	W17	1,3,4	(0.33)
LU-5180	±50 500	Tusk	Chukotka	Kyttyk Peninsula	W11	2,4	(0.5)
OxA-20033 (P23350)	50 300 ± 1000	Scapula	Chukotka	Kyttyk Peninsula	W25	1,2,3,4	1



¹⁴ C laboratory no.	¹⁴ C age	Material	Region	Locality	mtDNA haplotype	Loci scored	Individual heterozygosity
Ua-13375	4210 ± 70	Tooth	Wrangel Island		W1	1,2,3,4	0.25
LU-4448	4120 ± 110	Tusk	Wrangel Island		W1	1,2,3,4	0.25
Ua-13369	4085 ± 65	Tooth	Wrangel Island		W1	1,3	(0.5)
GIN-6985	3920 ± 40	Tusk	Wrangel Island		W1	1,2,3,4	0.75
AA40665	3905 ± 47	Tooth	Wrangel Island		W1	1,2,3,4	0.25
LU-2741	3730 ± 40	Tusk	Wrangel Island		W1	1,2,3,4	0.5
Ua-13366	3685 ± 60	Tooth	Wrangel Island		W1	1,2,3,4	0.25



- Vývoj populací mamuta v S-V Sibiři v období ob 60 000 BP až do vyhynutí před 4000 BP.

- **mtDNA seq a mtDNA mikrosatelity**

- SSR odvozeny od slonů (4 z 15ti testovaných)

Mamut (*Mammuthus primigenius*)

- **700.000 BP** – první fosilní záznamy a začátek šíření po severní polokouli

- **14.000** – začátek stahování mamutů z Evropy

- **Poslední Euroasijské populace na S. Sibiři**

- **Dvě poslední populace:**

- Ostrov Sv. Pavla v Beringově moři (do cca 6.400 BP)

- Wranglerův ostrov (do 4.000 BP) – u Čukotky v Severním ledovém oceánu



FROM THE COVER

Microsatellite genotyping reveals end-Pleistocene decline in mammoth autosomal genetic variation

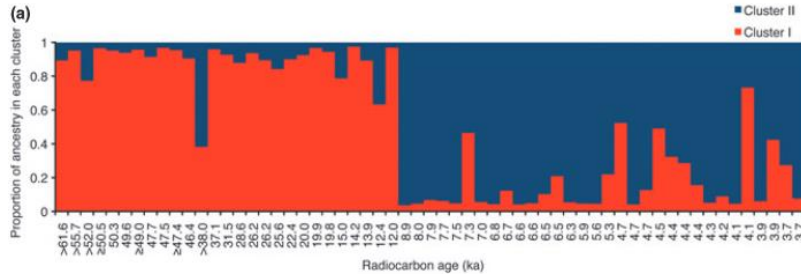


Fig. 2 Analysis of genetic differentiation. (a) Assignment of individuals to genetic clusters. The figure shown is based on the mean Q-values computed in CLUMPP from 10 clustering solutions obtained in STRUCTURE assuming two clusters. Each vertical bar represents one individual and shows its inferred cluster membership. The radiocarbon age (ka) is given below each individual. (b) Principal components analysis. Samples from before (black) and after (grey) 12 ka are differentiated along PC1. (c) Correlation between PC1 and radiocarbon age (Pearson correlation coefficient: -0.72 , $P = 2.73 \times 10^{-6}$), where underlined squares represent samples with infinite radiocarbon dates.

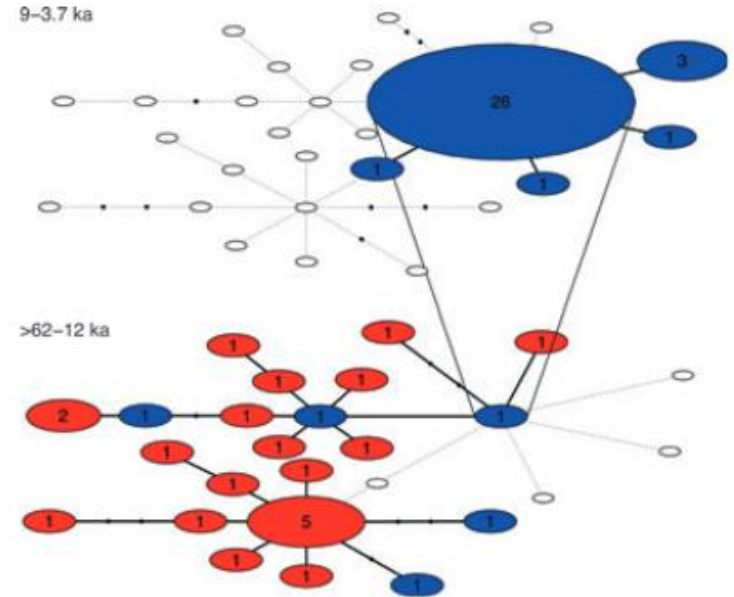


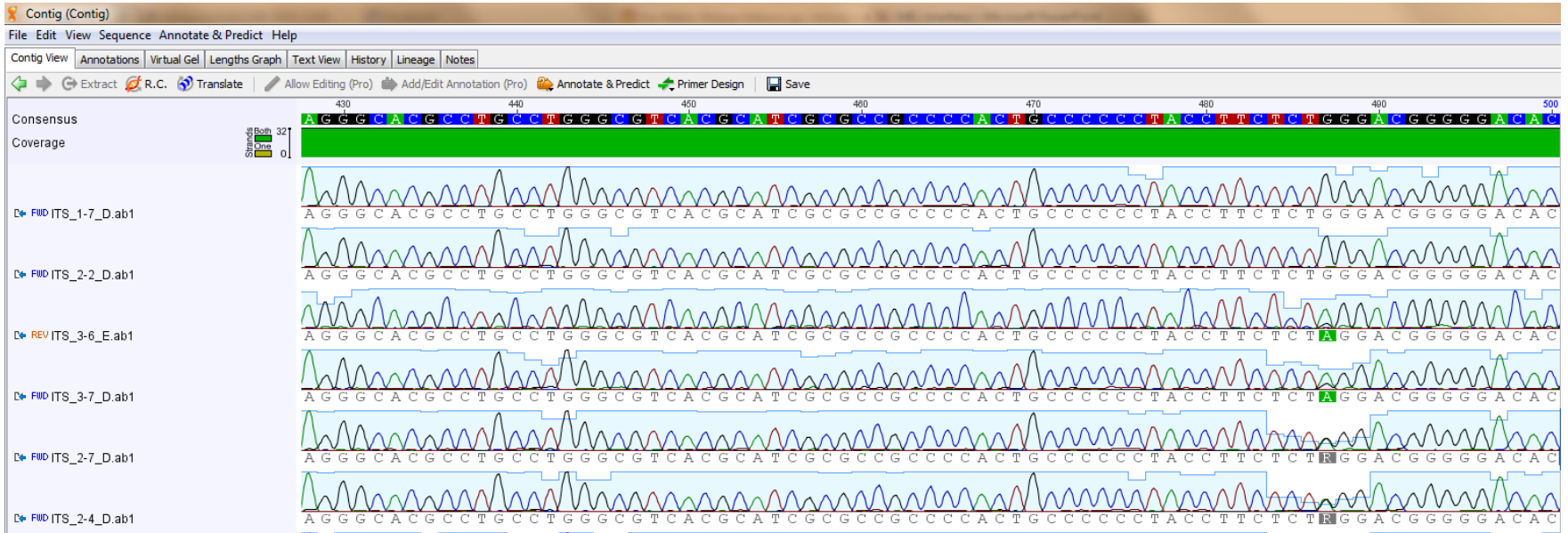
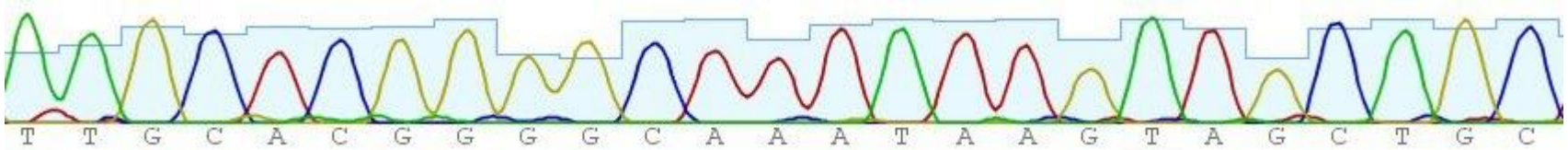
Fig. 3 A statistical parsimony network for two time periods (>62–12 ka and 9–3.7 ka). Blue haplotypes are from Wrangel and red haplotypes are from Chukotka. Haplotypes present in both time periods are connected with lines. Each branch represents one mutational step, and black circles represent missing haplotypes. Haplotypes missing only in one particular time period are shown as white ellipses. The number of individuals belonging to a particular haplotype is written within the ellipses.

- **Výsledky**
- Vzorky rozděleny do dvou skupin:
 - Variabilní vzorky až do 12.000 BP
 - Vzorky od 9.000 BP s omezenou variabilitou
- Snížení genetické variability (o 30% u SSR, o 65% u mtDNA) vlivem několikanásobně snížené velikosti populací
- Nebyl pozorován další pokles variability ke konci období výskytu = rychlá extinkce druhu vlivem člověka (lov) nebo změnou klimatu

SEKVENOVÁNÍ DNA

SEKVENOVÁNÍ DNA - princip

- zjištění pořadí nukleotidů v řetězci DNA



MEGA XLS CSV TA UUC PhC C V Pi S 0 2 4 Special NAME MOTF

[illegible]

Sangerova metoda SEKVENOVÁNÍ DNA - princip

- 1) Předstupeň: PCR s použitím dvojice primerů
 - namnožení studovaného úseku DNA
- 2) sekvenační reakce
 - použití pouze jednoho primeru
 - kromě dNTP jsou ve směsi přítomny i ddNTP
 - produkce fragmentů lišících se přesně o 1 bázi
- 3) Elektroforetická separace fragmentů na gelu
 - automatický sekvenátor

SEKVENOVÁNÍ DNA

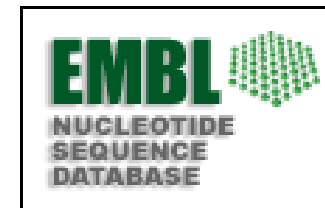
Aplikace

- Aplikační možnosti:
 1. **evoluce genů** (vznik alel, lokusů, redukce polymorfismu v důsledku selekce atd.)
 2. **vnitrodruhové (populační) studie** (geografická proměnlivost, tok genů, hybridizace, fylogeografie (př. mitochondriální geny, Y chromozom)
 3. **mezidruhové studie** (studium speciace, biogeografie)
 - volba správné sekvence
 4. **Detekce mutací (SNP, inserce, delece)** = identifikace jedinců, taxonů, populací – sekvenace pomáhá nalézt drobné populačně- i taxonově-specifické rozdíly

Výhoda

- Možnost srovnání sekvenčních dat o vašem organismu s údaji v internetových databázích

Př: <http://www.ebi.ac.uk/embl/>



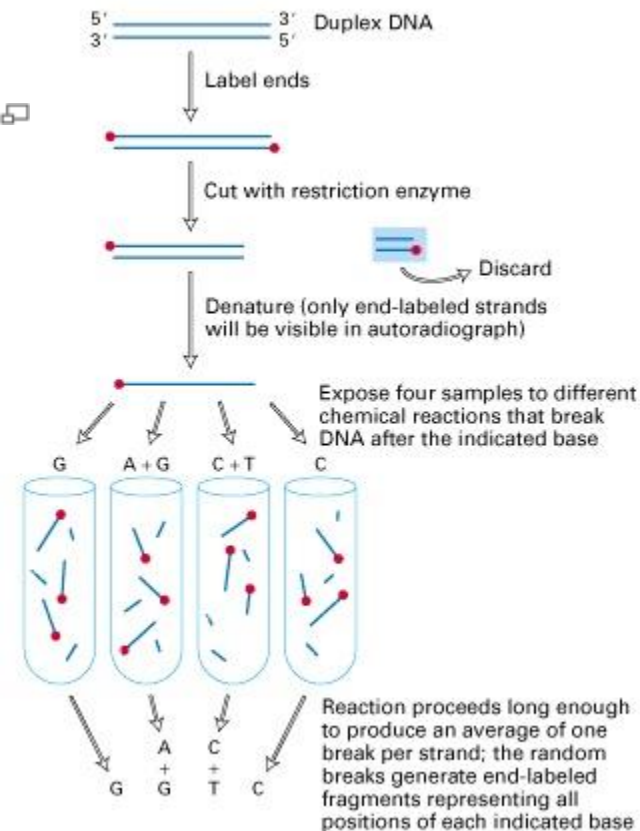
SEKVENOVÁNÍ DNA

1977: dvě metody zjištění pořadí nukleotidů v DNA

Maxamova-Gilbertova (chemická) metoda

je založena na báze-specifické chemické modifikaci a následném štěpení fragmentů

DNA - **dnes se neužívá**



Krátká sekvenční DNA (ds nebo ss)

* na 5' konci radioaktivně označena ^{32}P . * vzorek se

rozdělí na několik částí

a) Přídavek chemikáli – modifikují vždy jeden nebo dva typy nukleotidů:

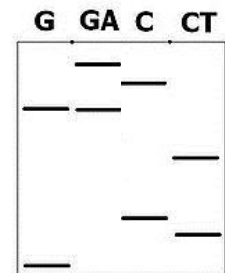
1. dimethylsulfát: G
2. kys. mravenčí: A+G
3. hydrazin: C + T
4. 1,5M NaCl: C

b) Přídavkem piperidinu: štěpení DNA v místě modifikace

c) Elfo separace

Výsledek ELFO:

sekvenční DNA: 5'-GTCTGCA-3'
(čte se zespodu)



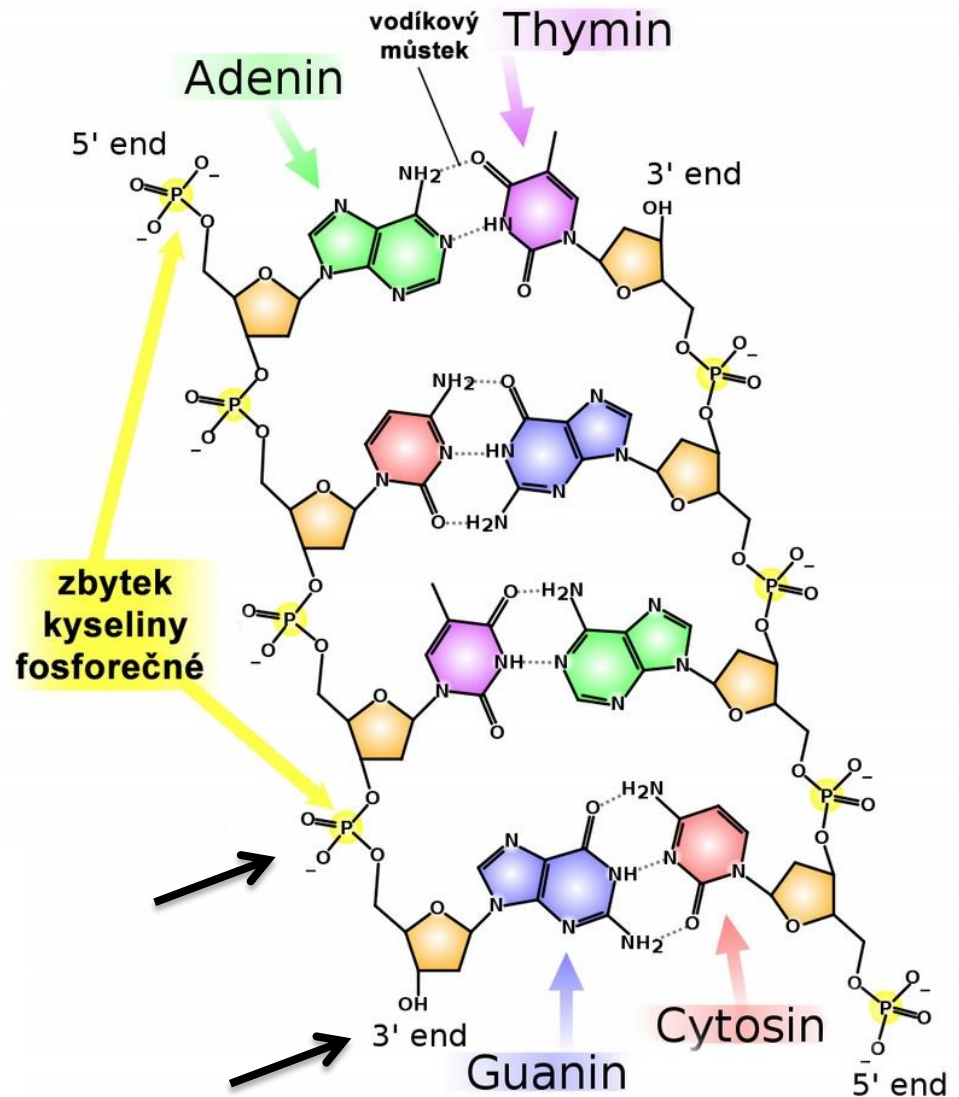
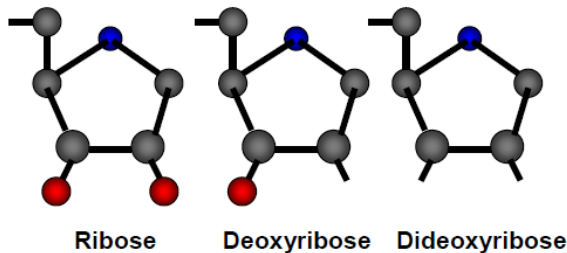
- narazila na technické problémy s vývojem standartních molekulárních kitů
- technicky příliš komplexní a příliš mnoho toxických látek

SEKVENOVÁNÍ DNA

Sangerova metoda terminace řetězce

Sangerova metoda

- založena na terminaci replikace nového řetězce podle matrice zkoumané sekvence dideoxynukleozidtrifosfátem (ddNTP)
- na 3'-uhlíku deoxyribózy chybí OH-skupina a proto k nim DNA polymeráza nemůže navázat další nukleotid
- pokud během replikace dojde k náhodné inkorporaci dideoxynukleotidu (ddA, ddC, ddG, ddT), replikace se zde zastaví

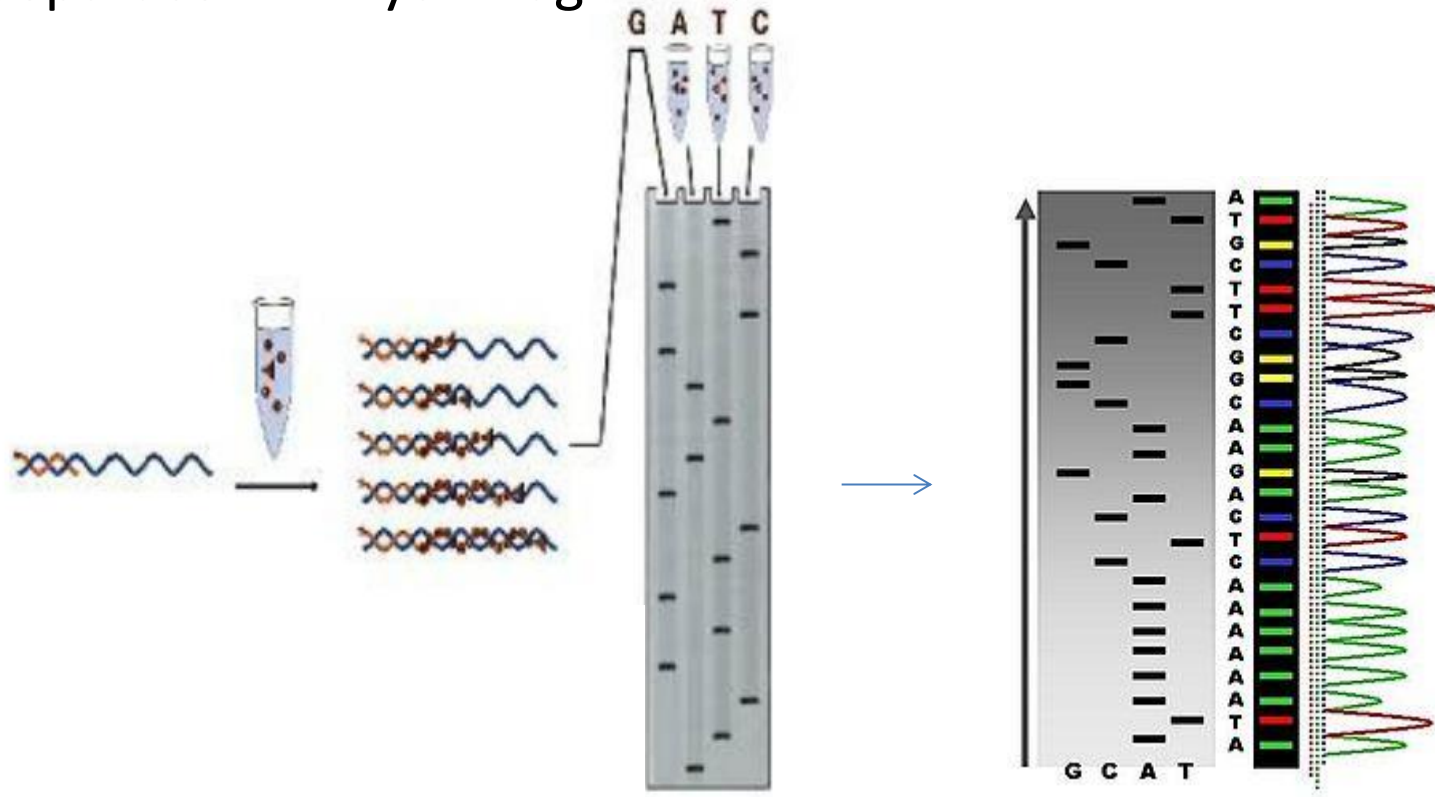


SEKVENOVÁNÍ DNA

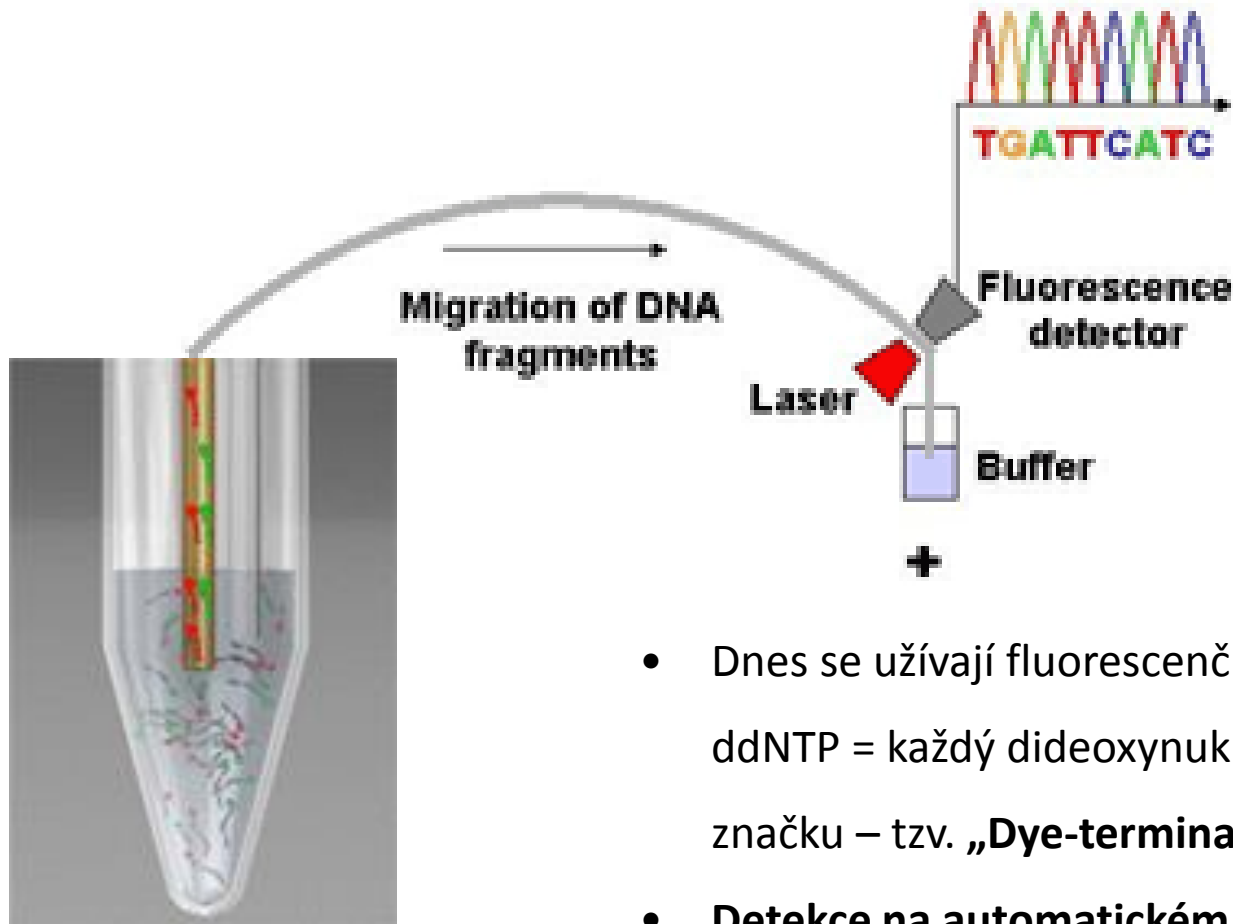
Sangerova metoda založená na terminace řetězce

Sangerovo sekvenování

- ke studované DNA (úseku) se přidá směs dNTP's a ddNTP's
- provede se amplifikace s jedním primerem
- provede se separace vzniklých fragmentů

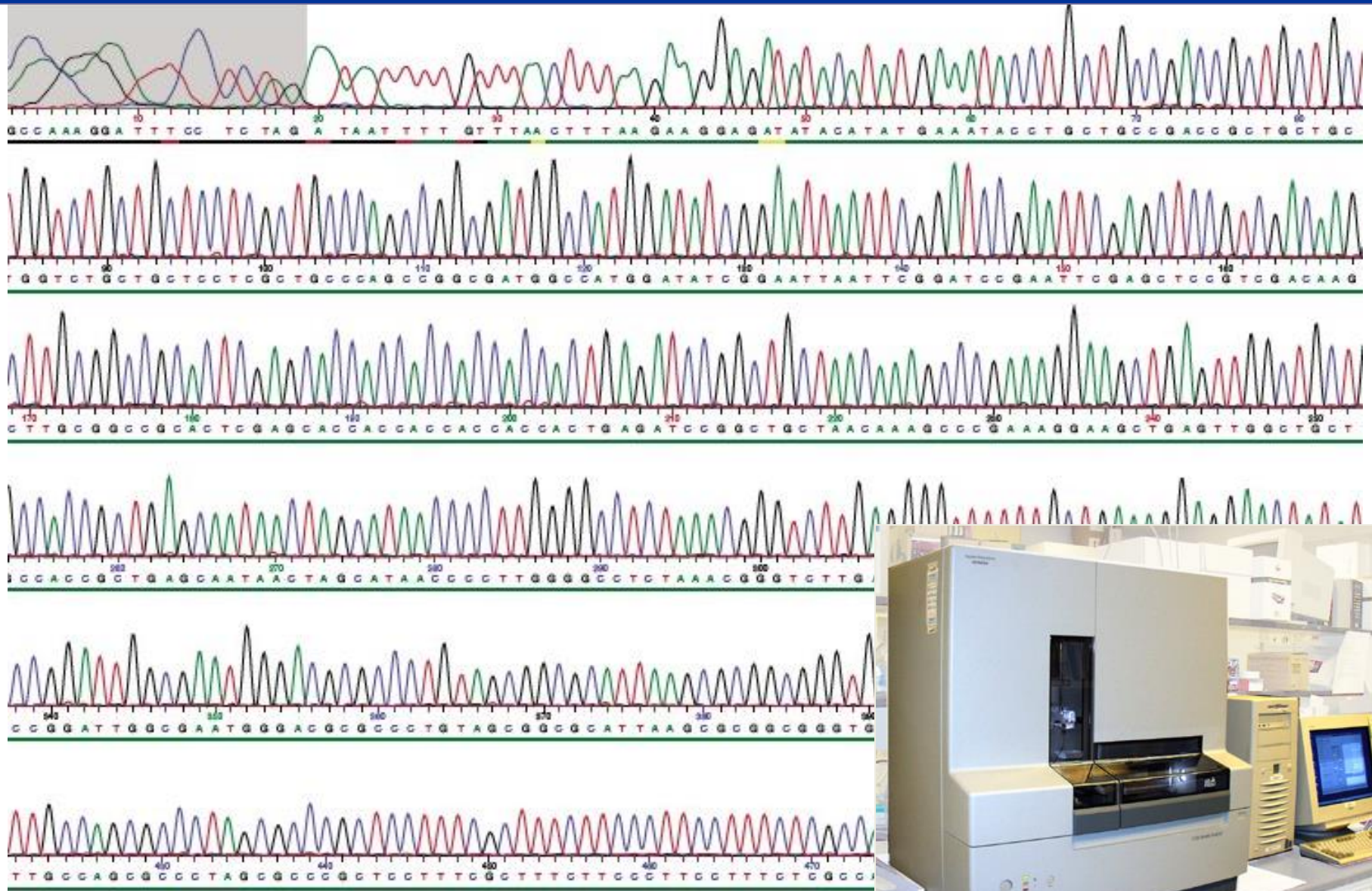


SEKVENOVÁNÍ DNA - zefektivnění Sangerovy metody vývoj kapilárních sekvenátorů



- Dnes se užívají fluorescenčně značené ddNTP = každý dideoxynukleotid má jinou značku – tzv. „**Dye-terminator sequencing**“
- **Detekce na automatickém sekvenátoru**

SEKVENOVÁNÍ DNA



- **Výhody**

- Dlouhá čtení (~900bp)
- Pro řešení malých projektů

- **Nevýhody**

- V porovnání s NGS technologiemi:
 - Relativně krátký úsek celkové výsledné sekvence DNA
 - Vysoká cena za délku analyzovaného úseku

Next Generation Sequencing

Metody celogenomového sekvenování

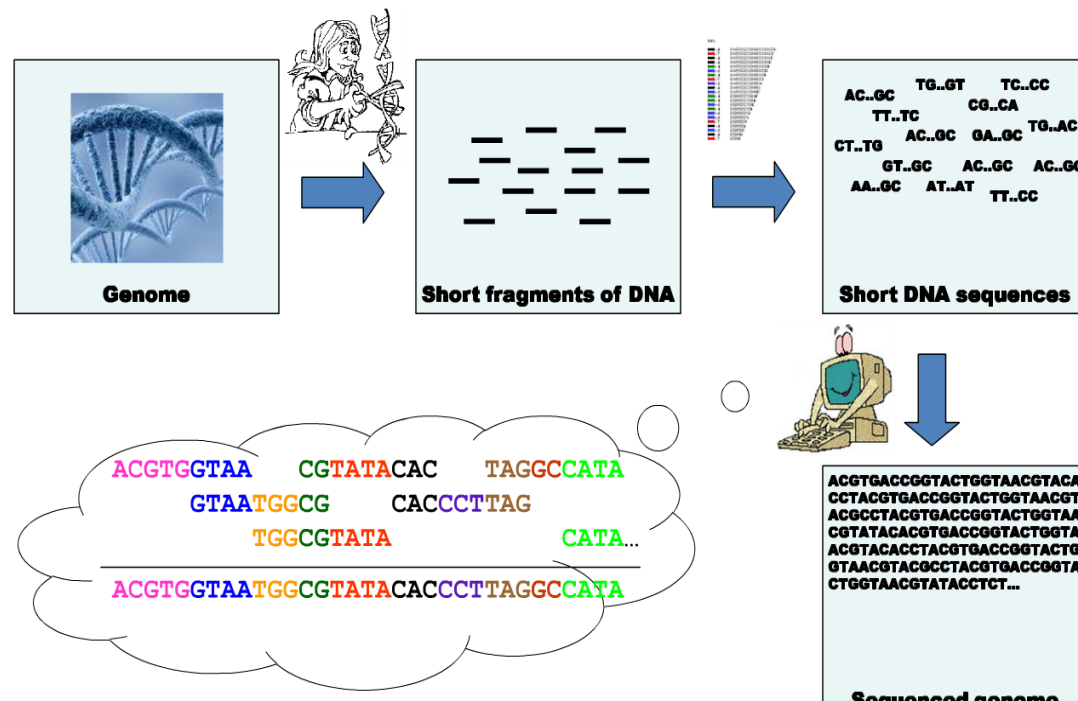
Metody celogenomového sekvenování

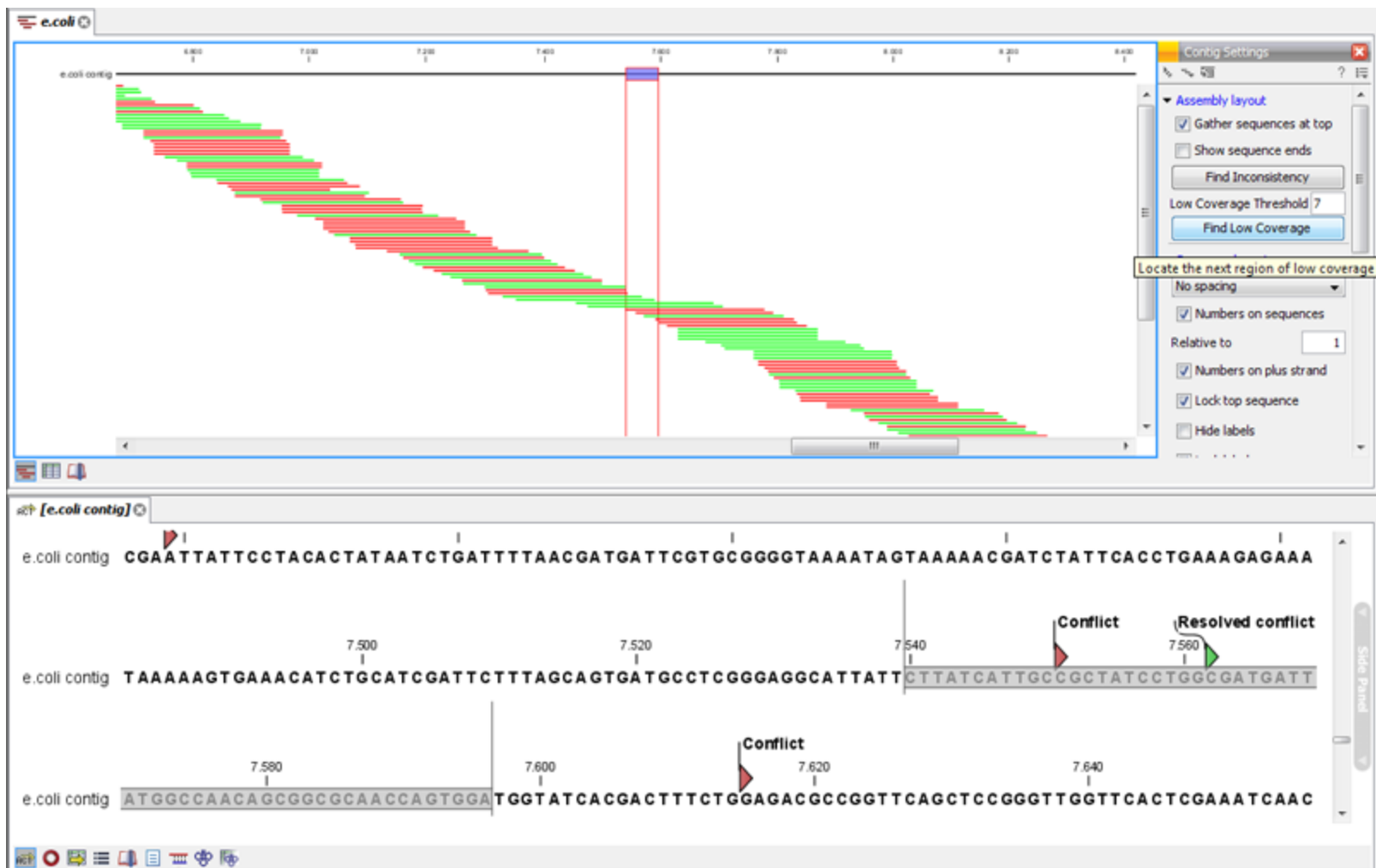
Celogenomové sekvenován

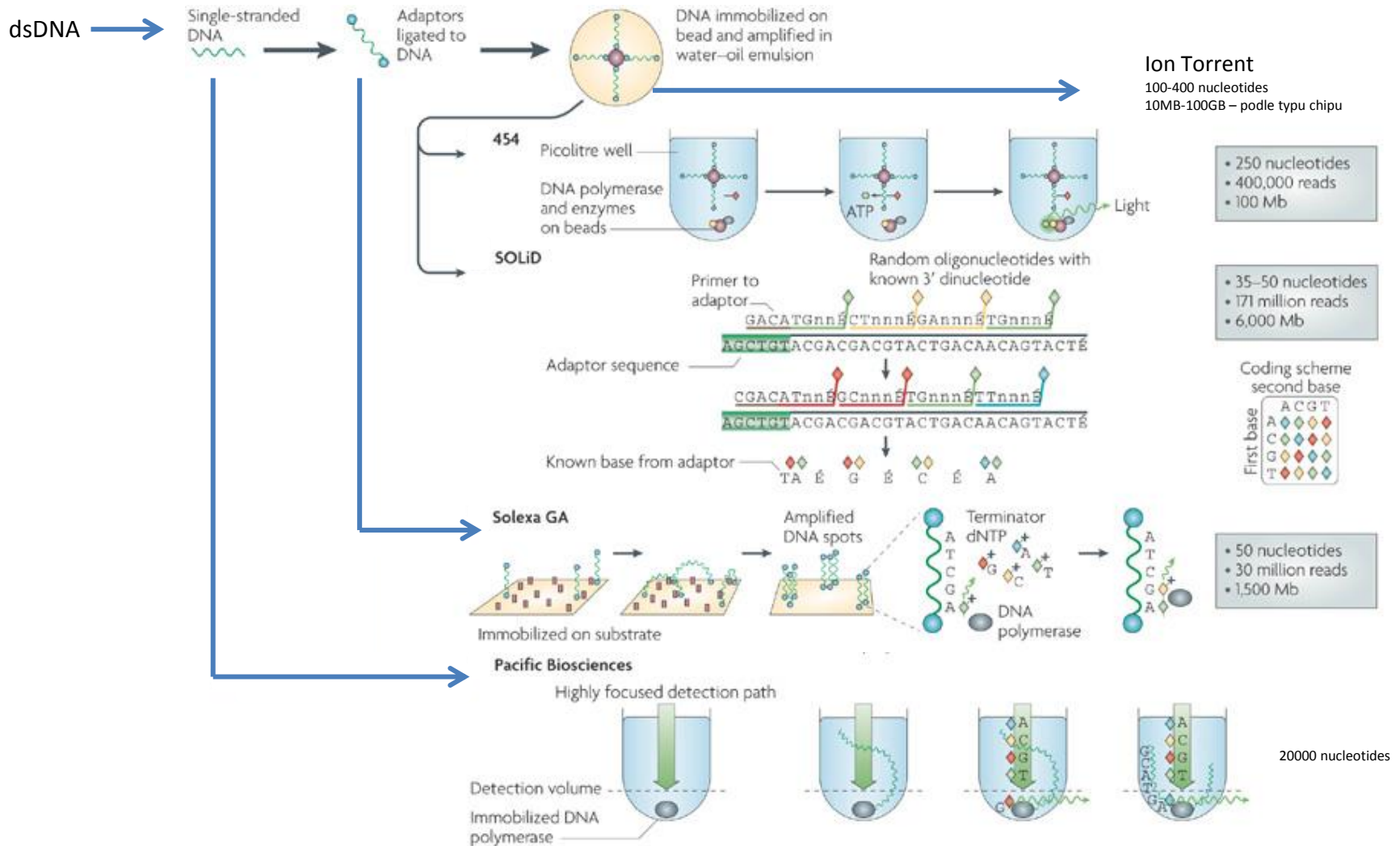
Cíl: získat informaci o pořadí nukleotidů genomické DNA nebo jejich rozsáhlých částech

Princip:

1. Příprava DNA knihovny - fragmentovat na kratší kousky a připravit pro sekvenaci (ligace adaptorů, PCR)
2. Vlastní sekvenace - paralelně probíhající čtení krátkých úseků (30-400 bp) u tisíců fragmentů DNA naráz
3. Bioinformatické zpracování dat - překryvem krátkých sekvencí získáme informaci o sekvenci gDNA







- první nová sekvenační metoda (1998)

- čteno 400 tisíc fragmentů najednou

Fragmentace DNA
- 300-800bp

- DNA kuličky do roztoku oleje a vody
- Protřepání = vznik mikroreaktoru okolo kuličky

- Na destičku se cyklicky nanáší roztok polymerázy a vždy jednoho nukleotidu... viz. pyrosekvenace

- B adaptor má biotinovou značku pro přichycení na streptavidinem obalenou magnetickou kuličku

- obsahuje 1,6 mil jamek širokých 44µm
- průměr DNA kuličky je 26 µm
- přídavek menších kuliček

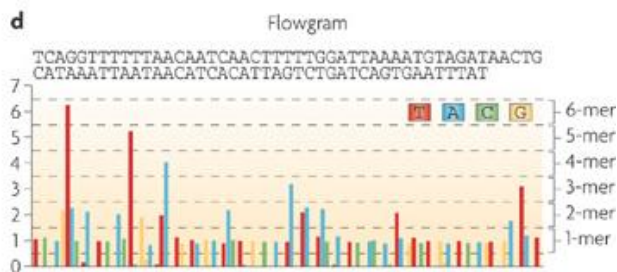
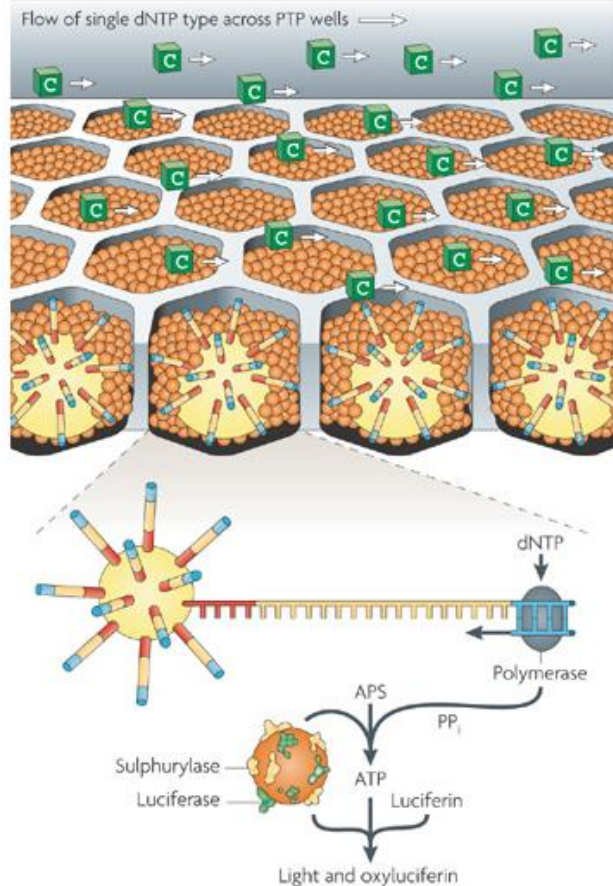
dvou typů – jeden obsahuje chemikálie pro pyrosexvenaci, druhý typ fixuje kuličky s DNA

6

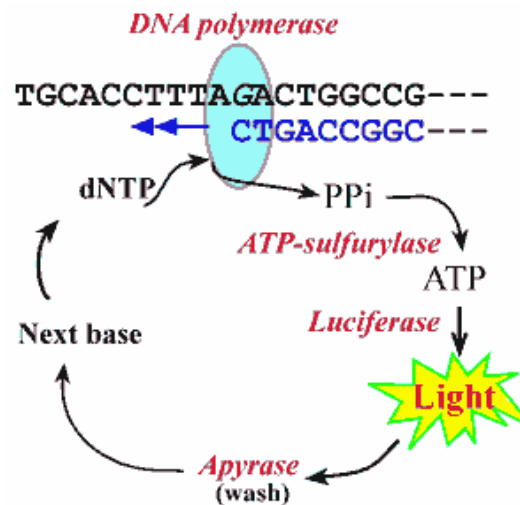
Záznam a zpracování dat



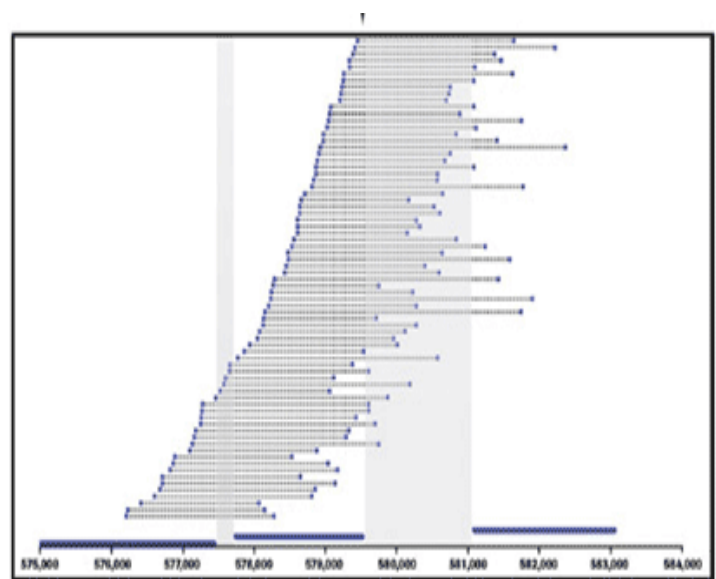
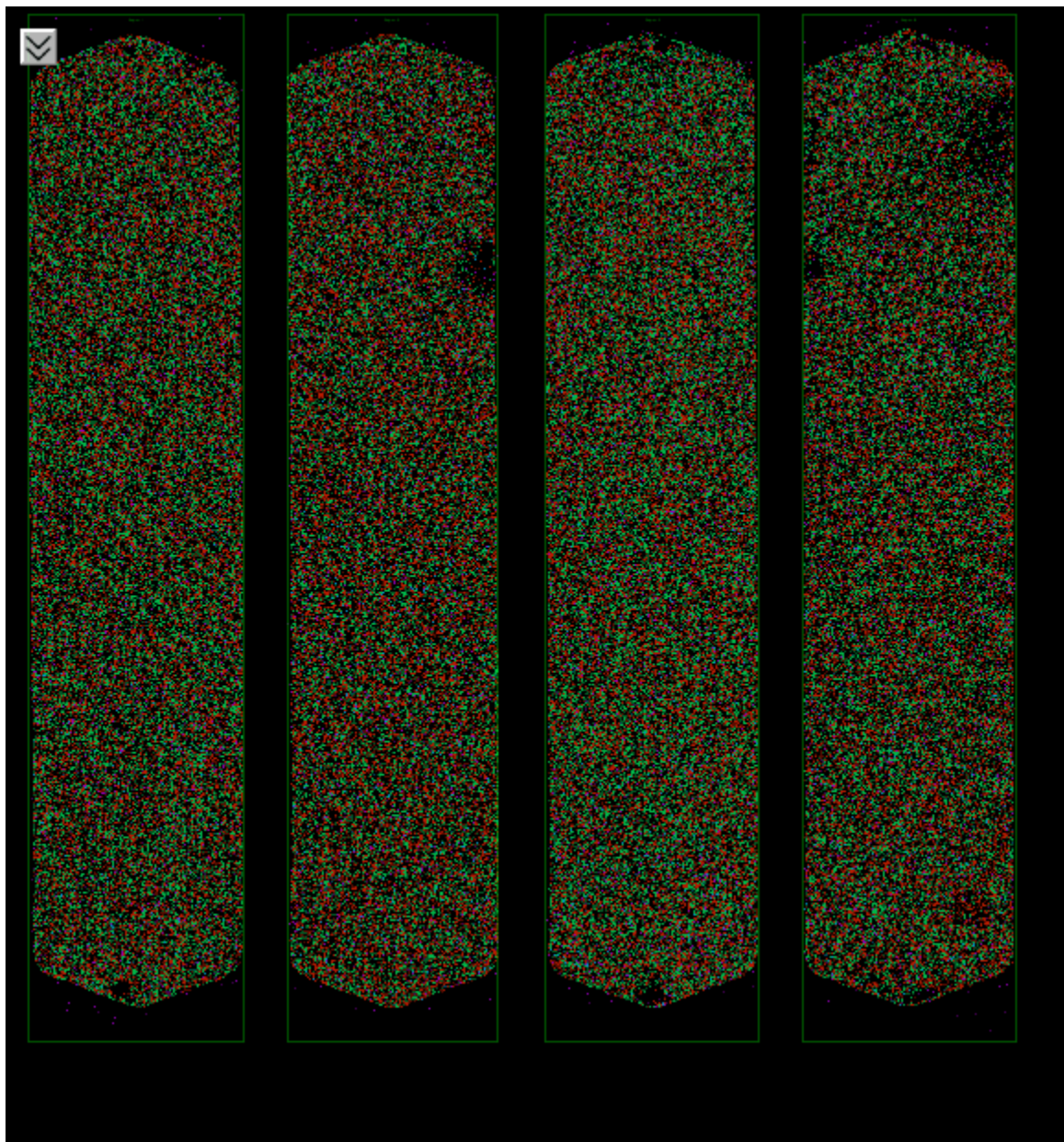
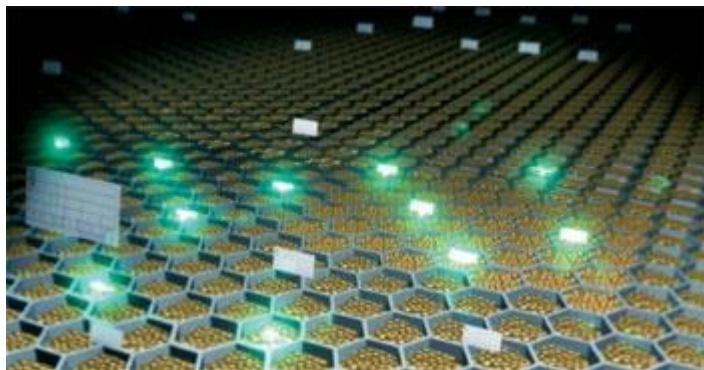
1–2 million template beads loaded into PTP wells



Pyrosekvenování



1. začlenění nukleotidu (konkrétní typ, ne směs)
2. uvolnění pyrofosfátu (PP_i)
3. Vznik ATP
 - Enzym ATP sulfuryláza katalyzuje vznik ATP reakcí PP_i s adenosinfosfosulfátem (APS)
4. Emise záření:
 - enzym luciferáza spotřebuje vzniklý ATP na oxidaci luciferinu – záblesk
5. Detekce záblesku
6. Enzym apyráza odstraní nespotřebované nukleotidy a ATP
7. Promytí systému
8. Změna typu nukleotidu ve směsi – krok č 1.



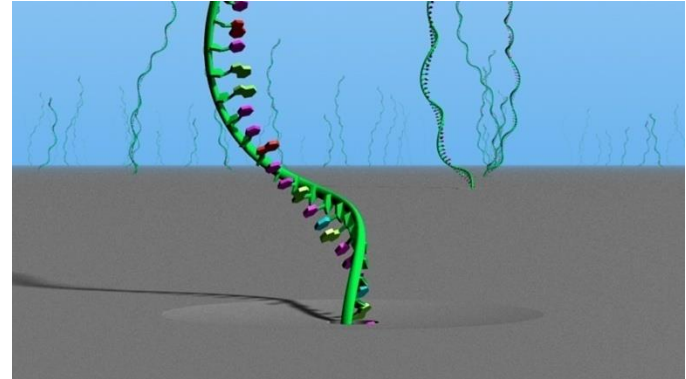
pico-titer plate from Roche's FLX

Illumina Genome Analyzer

- metoda sekvenace syntézou (SBS-sequencing by synthesis)
- r. 2006; úseky 35-50bp

- 1) Příprava knihovny

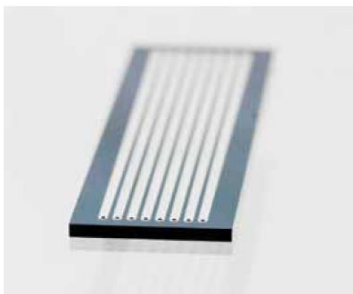
- fragmentace DNA
- na konce dva adaptéry
- ELFO: vyříznou se fragmenty 150-200bp
- přečištění



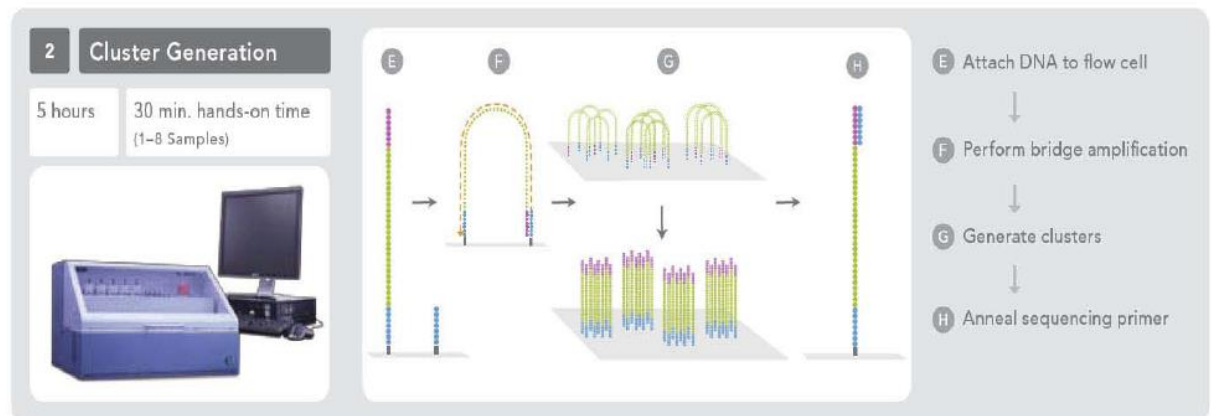
- 1a) Amplifikace fragmentů (bridge amplification)

- na sklíčku jsou oligonukleotidy komplementární k oběma adaptorům
- jednovláknové fragmenty se navážou na sklíčko – vytvoří „můstek“
- během PCR - dosyntetizování komplementárního vlákna,
- při každé denaturaci – dsDNA se rozdělí na dvě ssDNA, které vytvoří „můstky“ ležící na sklíčku blízko sebe
- opakování = vznik tisíců kopií těsně vedle sebe v tzv. „clusterech“ = knihovna je hotová

Figure 1: Illumina Flow Cell



Several samples can be loaded onto the eight-lane flow cell for simultaneous analysis on an Illumina Sequencing System.



Illumina Genome Analyzer

2) sekvenační reakce

- v každém cyklu NARÁZ všechny nukleotidy
 - mají fluorescenční značku a modifikovaný 3' konec (brání zařazení více nukleotidů v jednom kroku)
- zařazení nukleotidu provázené emisí světla, charakteristického pro daný nukleotid
- detekce
- odstranění fluoresc. značky a odblokování 3' konce
- opakování postupu
- x cyklů

3) zpracování dat

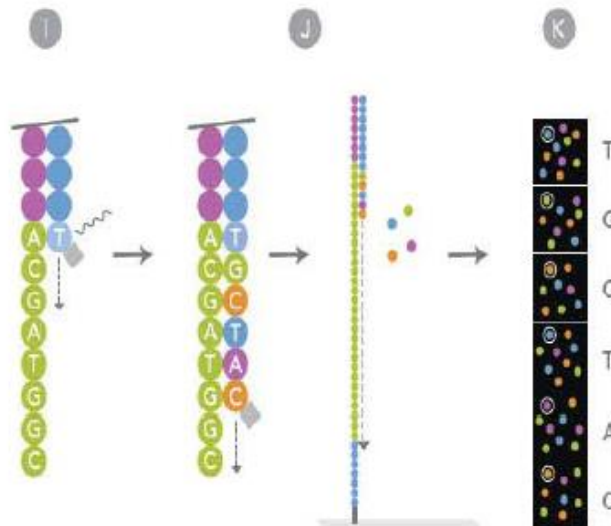
3

Sequencing

2-3 days (single-read)

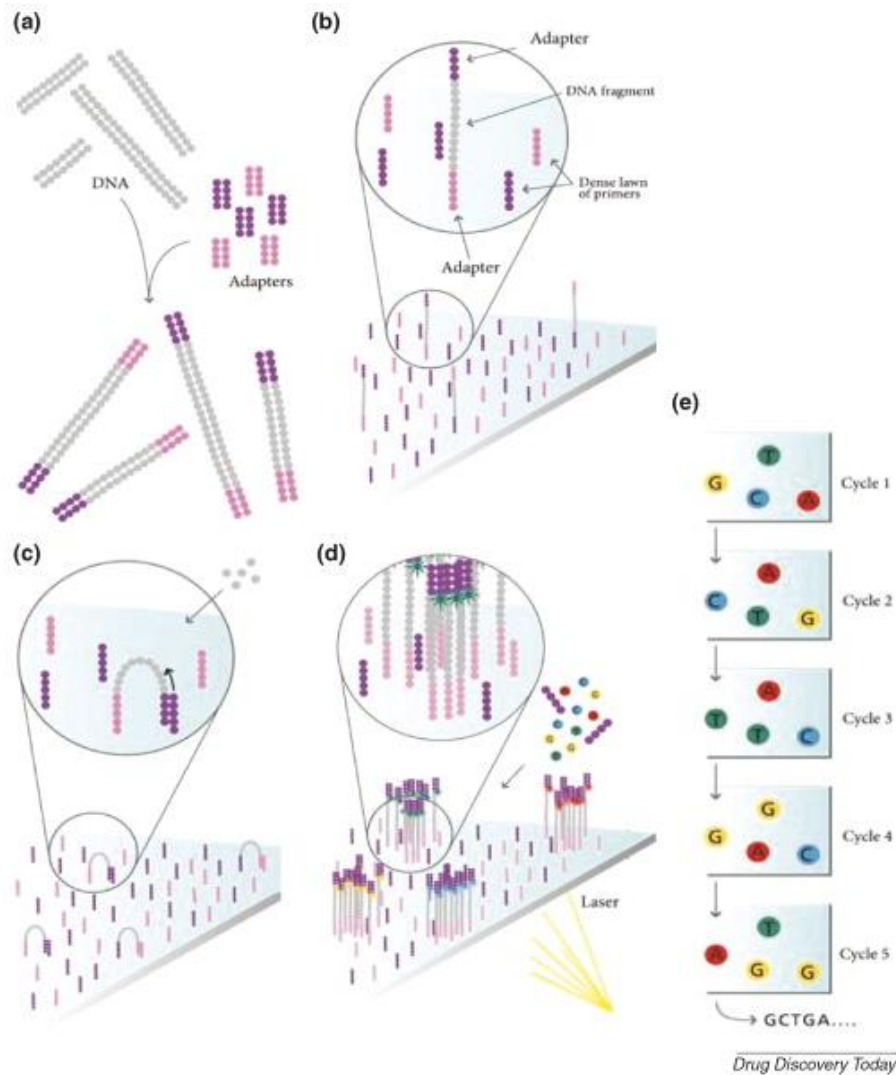
4-6 days (paired-end)

30 min. hands-on time (1-8 Samples)



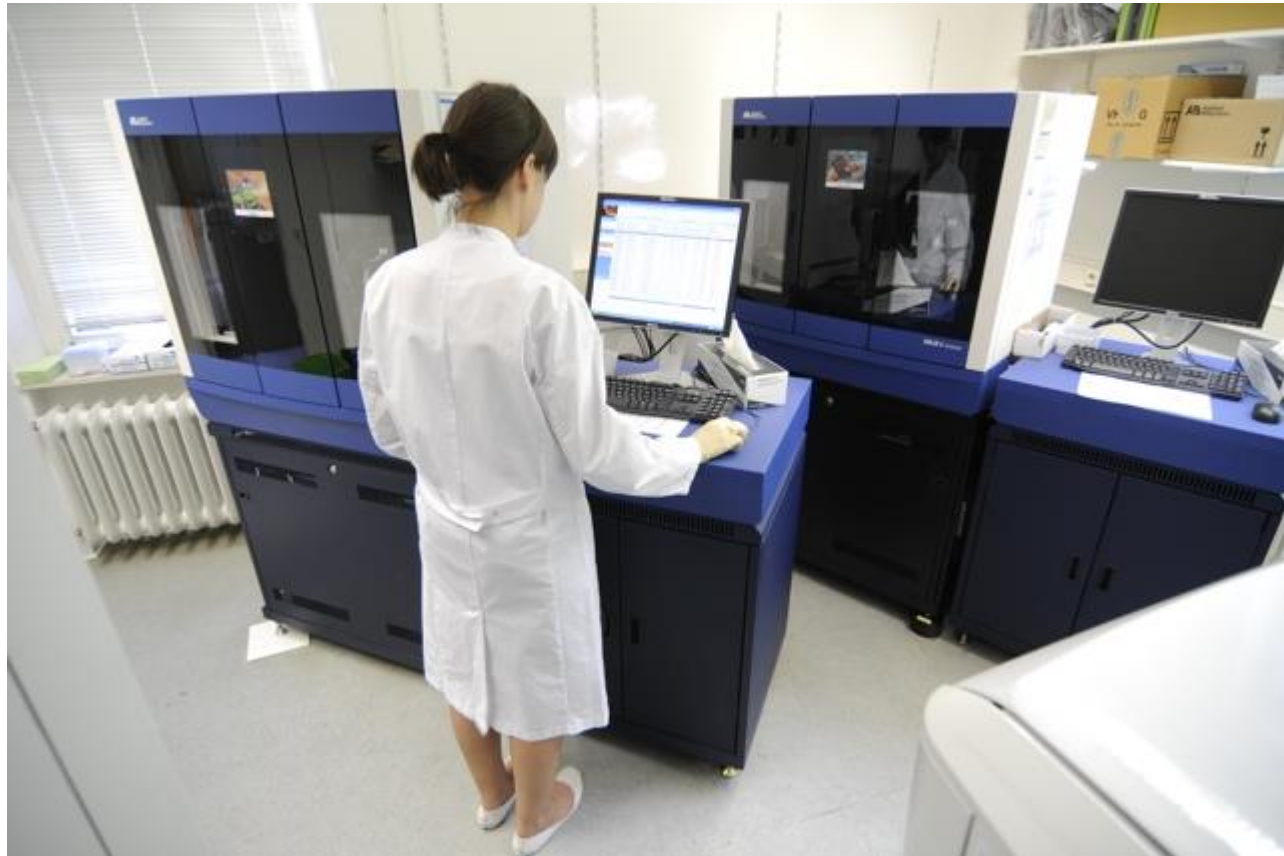
- I Extend first base, read, and deblock
- J Repeat step above to extend strand
- K Generate base calls

Illumina/Solexa seq.



[Drug Discovery Today](#)
[Volume 13, Issues 13–14](#), July 2008, Pages 569–577

Illumina/Solexa process overview. Genomic DNA is randomly fragmented and adapters are added to each fragment end **(a)**. Single-stranded fragments are then attached randomly inside flow cells **(b)**. Bridge amplification is performed to generate double-stranded fragments **(c)**. Following repeated cycles of denaturation and bridge amplification, millions of DNA copies are present in each flow cell **(d)**. DNA sequence is determined with four-labeled reversible terminators primers and polymerase **(e)**. Unincorporated terminators are washed, and the sequence is determined in each flow cell following laser excitation. The blocked 3' terminus and fluorophore are removed from the incorporated base and the cycle is repeated (figure adapted from <http://www.illumina.com>).



Third Generation Sequencing

Pacific Biosciences — Real-time sequencing

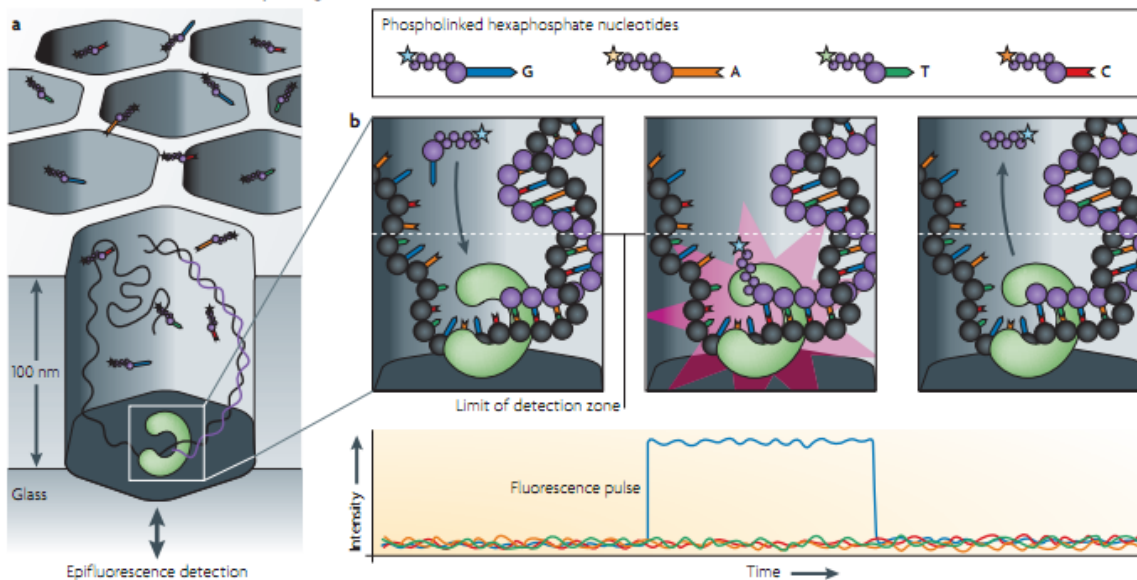


Figure 4 | Real-time sequencing. Pacific Biosciences' four-colour real-time sequencing method is shown. **a** | The zero-mode waveguide (ZMW) design reduces the observation volume, therefore reducing the number of stray fluorescently labelled molecules that enter the detection layer for a given period. These ZMW detectors address the dilemma that DNA polymerases perform optimally when fluorescently labelled nucleotides are present in the micromolar concentration range, whereas most single-molecule detection methods perform optimally when fluorescent species are in the pico- to nanomolar concentration range⁴⁴. **b** | The residence time of phospholinked nucleotides in the active site is governed by the rate of catalysis and is usually on the millisecond scale. This corresponds to a recorded fluorescence pulse, because only the bound, dye-labelled nucleotide occupies the ZMW detection zone on this timescale. The released, dye-labelled pentaphosphate by-product quickly diffuses away, dropping the fluorescence signal to background levels. Translocation of the template marks the interphase period before binding and incorporation of the next incoming phospholinked nucleotide.

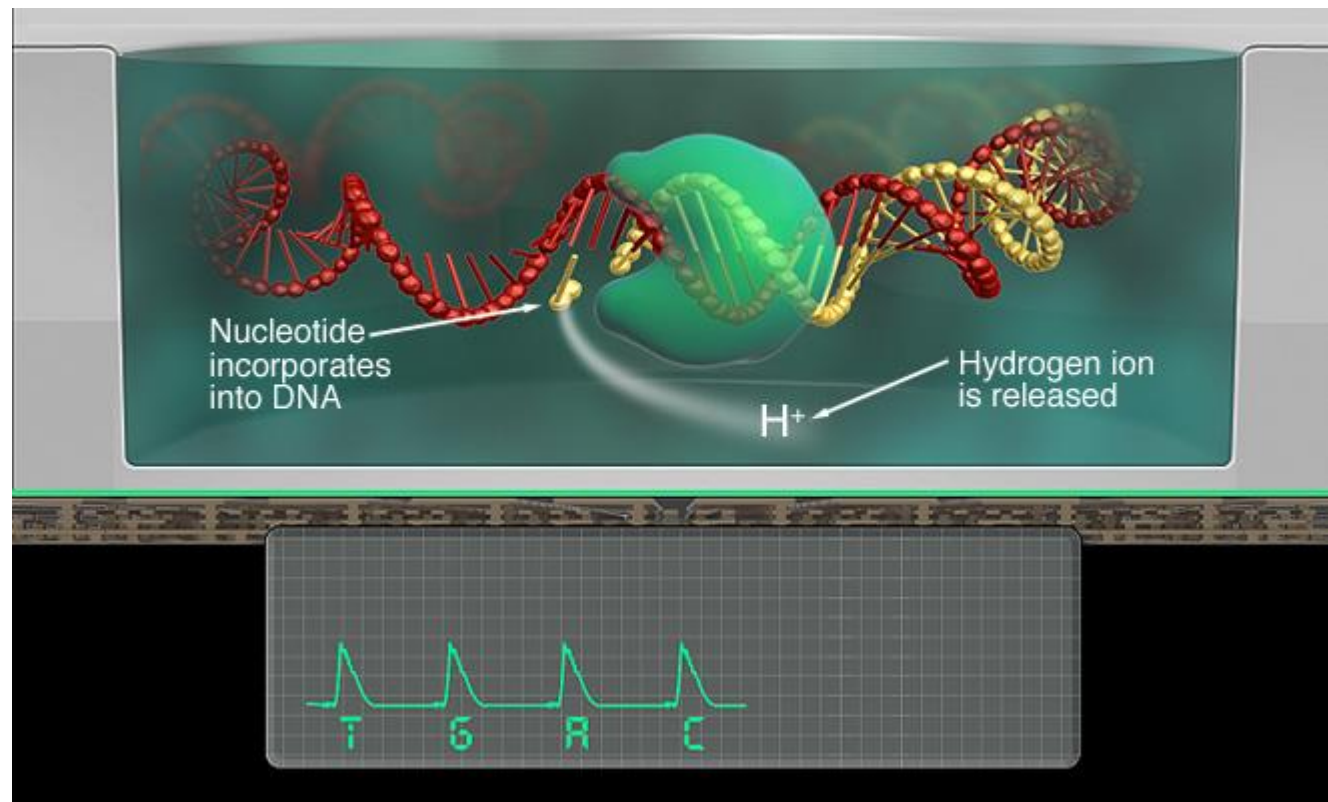
reads as long as 20,000 bases.





Semiconductor Sequencing for Life™

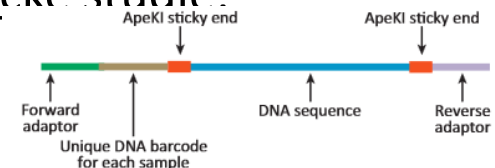
ion torrent
△ ★ △ ○ × □ + ≈



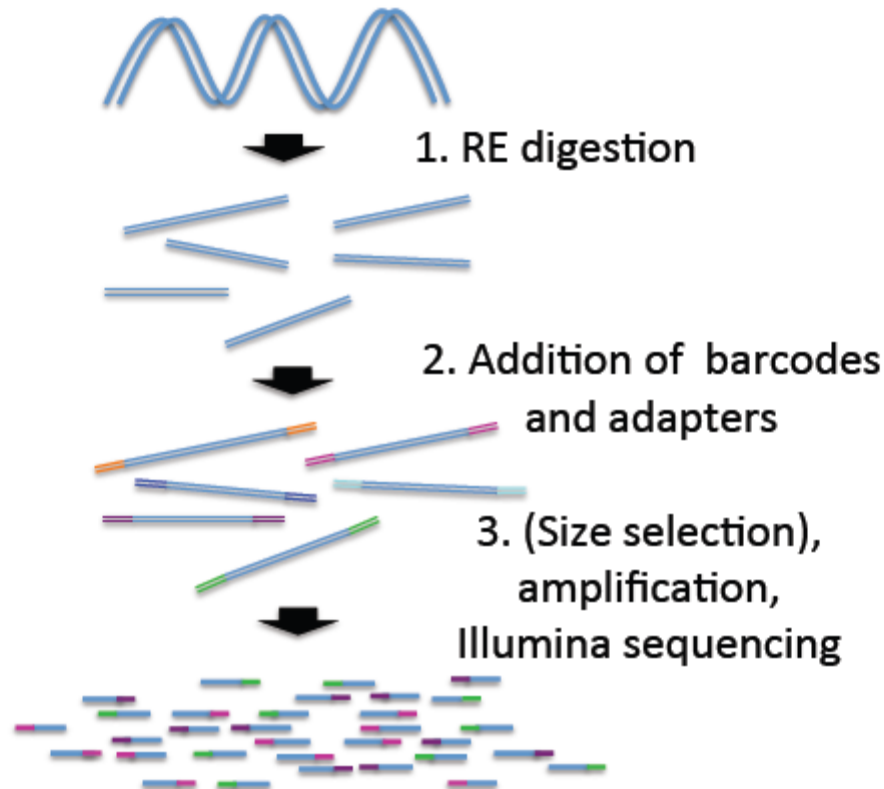
Metody odvozené od NGS

Odvozené metody využívající NGS sekvenace při hledání SNP polymorfismů

- Využití NGS pro detekci populačně genetické studie
- Detekce počtu SNP v řádech 1.000-10.000
- Předpokladem použití NGS je:
 - Sekvenujeme pouze část genomické DNA = snižujeme komplexnost genomu
 - Různé přístupy
 - RNA sequencing, PCR amplicon seq,
 - Pro populačně genetické studie, fylogeografické studie:
 - Metody založené na restrikci DNA
 - (Restriction Digest-based Methods)
- Barcoding
 - Použití adaptorů se specifickou sekvencí pro daný vzorek
 - Umožňuje spojit velký počet různých vzorků a osekvenovat je naráz



Restriction Digest-based Methods

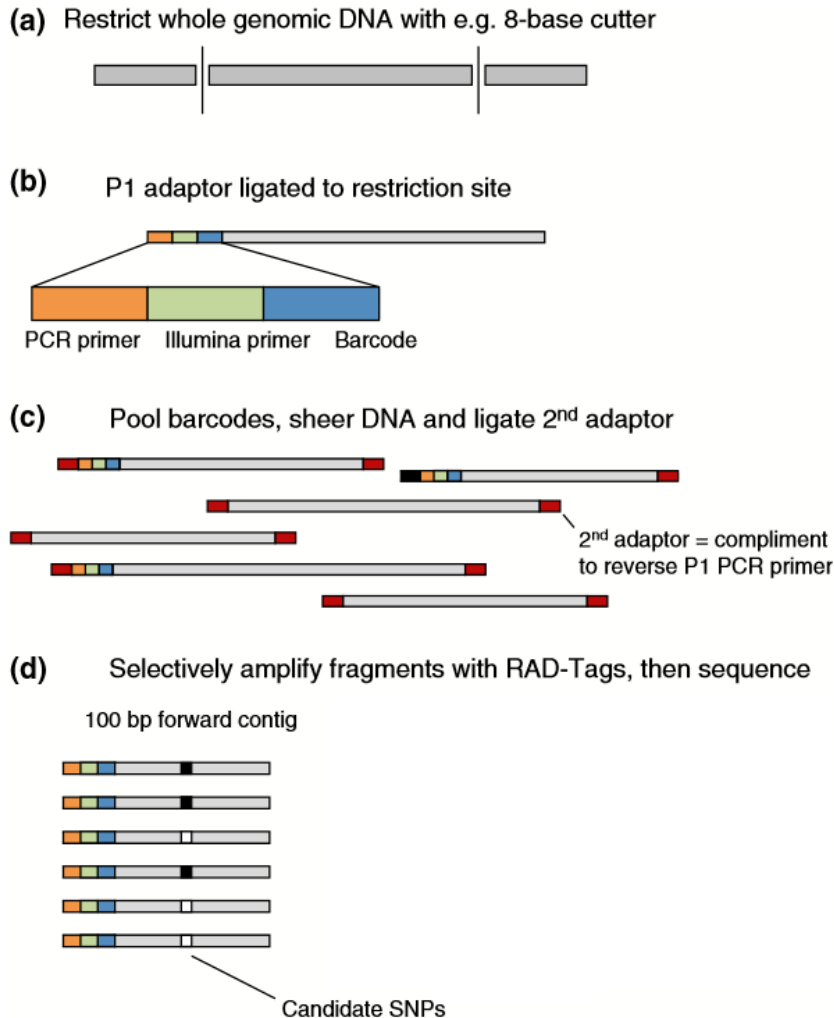


- restrikce: DNA je naštípána restriktázami,
- ligace: na vzniklé fragmenty je navázán adaptor,
- vysokokapacitní NGS sekvenování
(tagged fragment amplification)
- analýza dat
- Hledání SNP – a) anotace k referenčnímu genomu
b) pooling vzorků a hledání SNP lokusů
- **Kombinace snížení komplexnosti genomu
restrikčním štěpením s NGS sekvenováním**

Restriction Digest-based Methods

- Methods:
 - CROPs (Van Orsouw et al. 2007)
 - RAD-seq (Baird et al. 2008)
 - Modified CROPs (Gompert et al. 2010)
 - MSG (Andolfatto et al. 2011)
 - GBS (Elshire et al. 2011)
 - Modified GBS (Parchman et al. 2012)
 - ddRAD-seq (Hohenlohe et al. 2012, Peterson et al. 2012)
 - Modified GBS (Poland et al. 2012)
 - SBG (Truong et al. 2012)
 - 2b-RAD-seq (Wang et al. 2012)
 - SLAF-seq (Sun et al. 2013)

RAD-Taq sequencing



- restrikce: DNA je naštípána dvojicí restriktáz
- ligace: na vzniklé fragmenty jsou navázány adaptory
- vysokokapacitní NGS sekvenování
(tagged fragment amplification)
- analýza dat
- Hledání SNP – a) anotace k referenčnímu genomu
b) pooling vzorků a hledání SNP lokusů
(*putative RADtag locus*)
- Kombinace snížení komplexnosti genomu
restrikčním štěpením s NGS sekvenováním.

Fig. 1 The RAD-Tag SNP discovery process (after Baird *et al.* 2008). Following restriction with an appropriate enzyme (a), the P1 adaptors are ligated to the restriction sites (b). DNA from individuals (or pooled individuals) has different barcoded adaptors ligated in separate reactions. Adaptor-ligated fragments are pooled, sheared and selectively amplified under PCR (c). Products are sequenced by single or paired-end sequencing and aligned to identify SNPs (d). The paired-end contig will reduce sequence depth at candidate SNPs, but increase the length of SNP flanking region for assay design.

Wyeomyia smithi (the purple pitcher-plant mosquito)

Emerson K J et al. PNAS 2010;107:16196-16200

Outline: vysledovat směr šíření a genetickou variabilitu populací *W. smithii* na východním pobřeží USA

Materiál: 21 pops á 6 individuals

Metoda: restriction site-associated DNA tags (RAD tags)

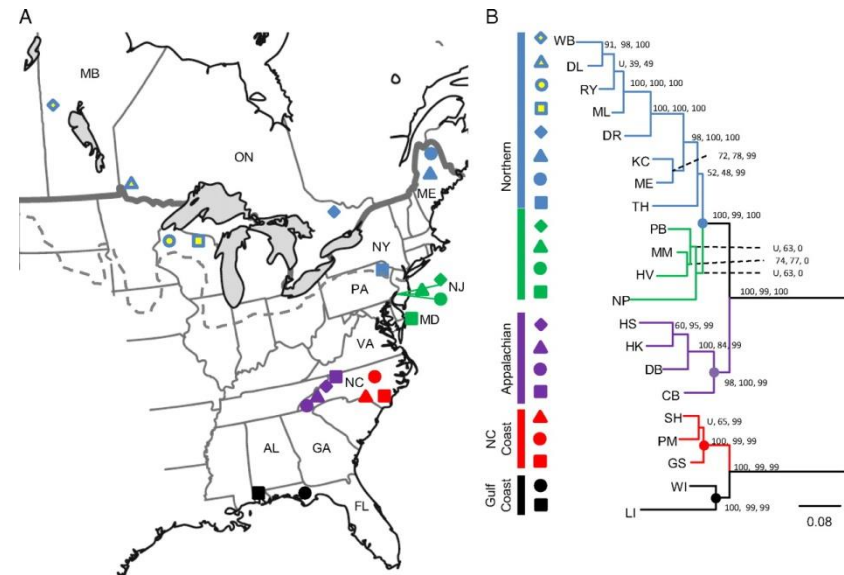
Výsledky: crude: 27.5 milRAD tag seq.

- filter pass: 14.9 mil
- 3,741 SNPs within the RAD tag that were fixed within at least two populations and were variable among populations.

Results:

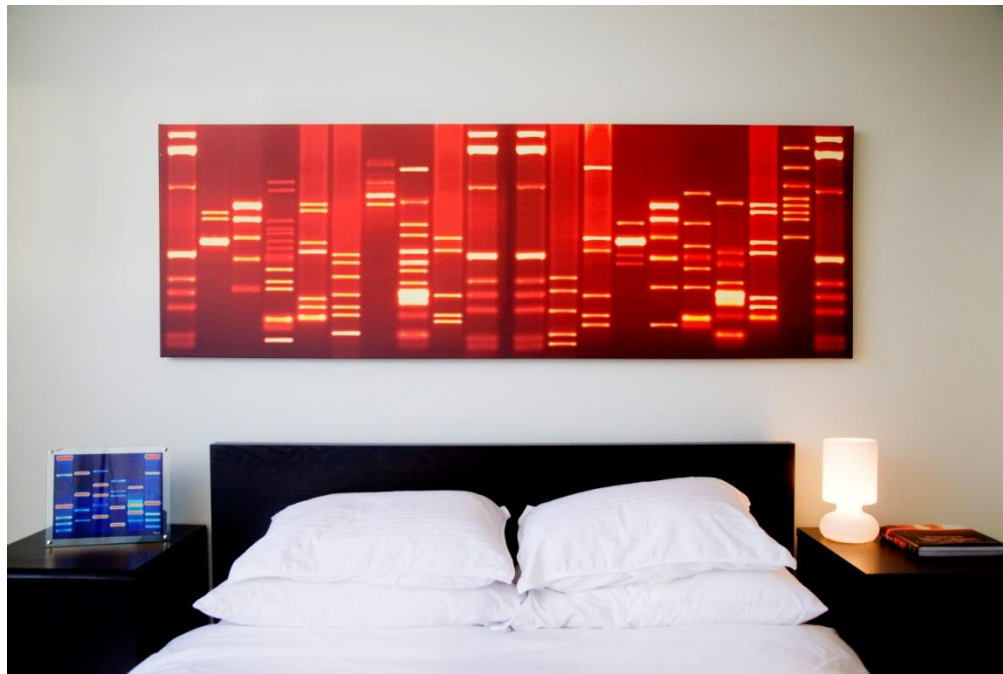
- confirms the **phylogeographic separation of *W. smithii* into southern and northern groups**
- reveals previously unresolved genetic structure and direction of evolution
- from a southern Appalachian Mountain refugium following recession of the Laurentide Ice Sheet at 22,000–19,000 B.P.

- **The distribution of host plant, *S. purpurea*, and prevailing winds provide clues to the origin of populations within the northern group.**



(A) Map of eastern North America showing the geographic range of *W. smithii* used in this study. Dashed line shows the maximum extent of the Laurentide Ice Sheet at ~20,000 B.P.

(B) Maximum likelihood phylogenetic tree of *W. smithii* using the RAD tag SNPs



That's all folks!

