

Molekulární metody ve studiích kořenových systémů

Jiří Košnar, 2016



Úvod

Řešení otázek:

- identifikace (barcoding) a kvantifikace rostlinných druhů ve společenstvu
- identifikace (barcoding) a kvantifikace rostlinných symbiontů – AMF, ECM, hlízkové bakterie apod., nebo patogenů

předpoklady:

- více druhů ve vzorku → ze všech budeme analyzovat pouze jednu homologickou část genomu (vhodně variabilní - aby rozlišila taxony) → PCR amplifikace, tj. musíme znát primery

Úvod

Schéma DNA analýzy:

Odběr vzorku – fixace vysušením silikagel, nebo sušárna (50-60 °C)



Izolace celkové DNA vzorku

- homogenizace (např. mlecí mlýnky)
- často problémy s kvantitou a kvalitou získané DNA (nutná dodatečná purifikace, časově a finančně náročné)
- nutno optimalizovat, zkoušet různé komerční kity



PCR amplifikace konkrétního úseku DNA

- prokaryota – 16S, eukaryota – LSU, SSU nrDNA, apod.



Detekce taxonů ze směsi molekul PCR produktu
= metody detekce DNA variability



Přehled metod

Detekce DNA variability:

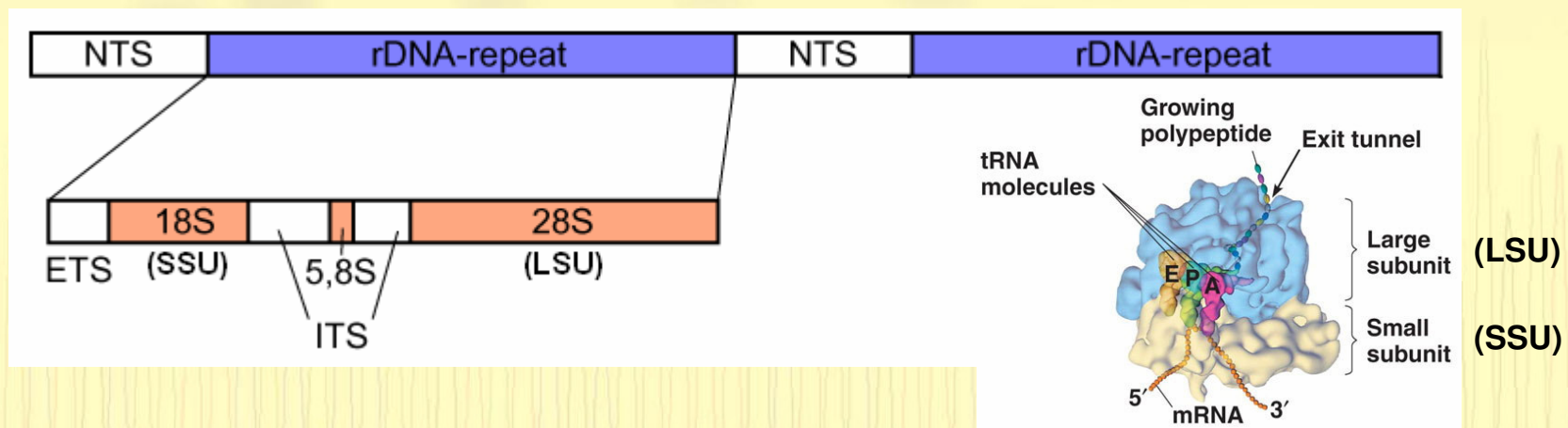
- **sekvenace DNA** – přesné stanovení sekvence nukleotidů
 - identifikace druhu podle podobnosti sekvence (k jinému známému organismu)
 - u vzorků společenstev obsahujících více druhů:
 - klonování+sekvenace
 - next-generation sequencing (NGS)
- **nepřímá detekce** variability **DNA** sekvencí (‘proužkové’ metody):
 - DGGE
 - T-RFLP
- **kvantifikace** organismu nebo skupiny taxonů, RT-PCR (qPCR)

Sekvenování DNA

Co se dá sekvenovat? To, na co jsou primery...

➤ Jaderná ribozomální DNA (nrDNA)

- primery na širší spektrum (Eukaryota, Prokaryota)
- nebo na určitou taxonomickou skupinu (Basidiomycota, Glomeromycota, rostliny)
- možno navrhnout specifické i na konkrétní druhy
- v jednom genomu ve velkém množství kopií (multi-copy), někdy kopie i navzájem odlišné



Sekvenování DNA

Co se dá sekvenovat? To, na co jsou primery...

➤ **Jaderné single-copy a low-copy geny**

- primery obvykle fungují pouze na určitou taxonomickou skupinu nebo druhy

➤ **Rostliny - cpDNA:**

- ± univerzální primery pro desítky úseků
- chloroplasty jsou v buňce ve velkém počtu
- příklady často používaných úseků:

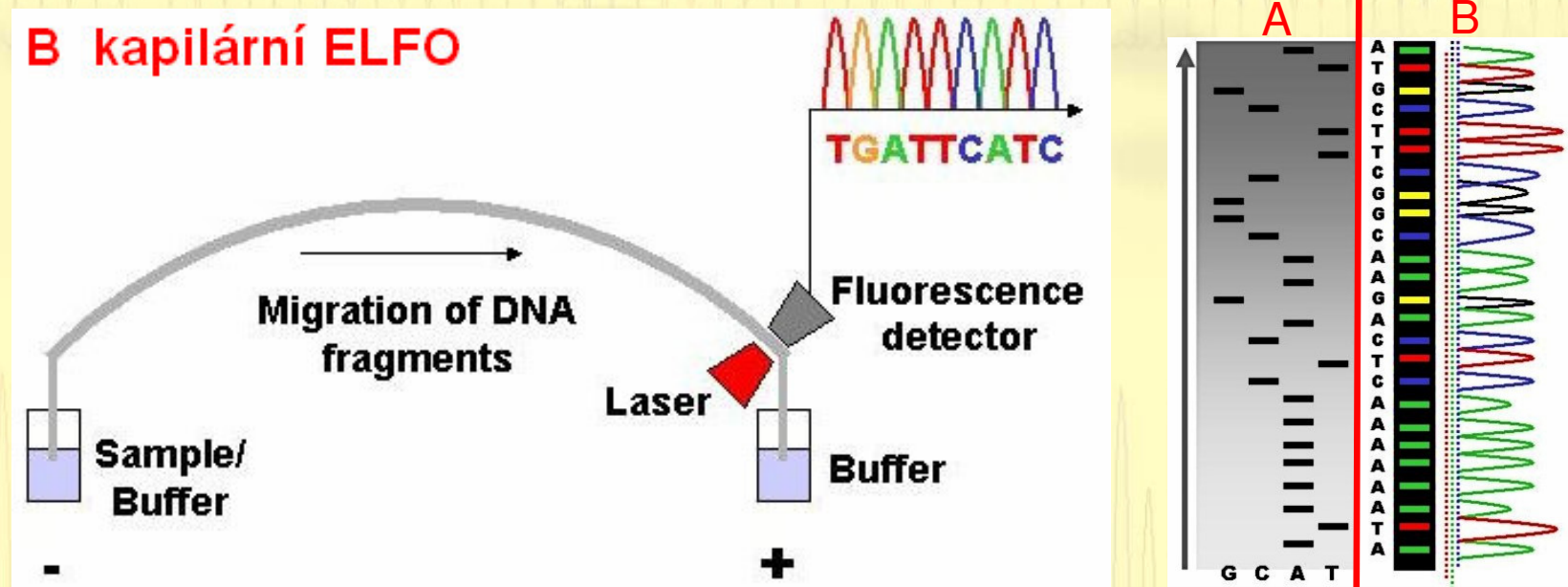
rbcL*, *matK – kódující

exony, spacery a introny oblasti ***trnT-trnL-trnF*** (Taberlet et al., 1991)

Sangerovo sekvenování DNA

Vlastní sekvenace DNA

- provádí komerční firmy (dodáme PCR produkt)
- **sekvenování syntézou DNA:** dideoxynukleotidy (ddNTPs) nemají 3'-OH skupinu pro další prodlužování řetězce DNA, proto **terminují**

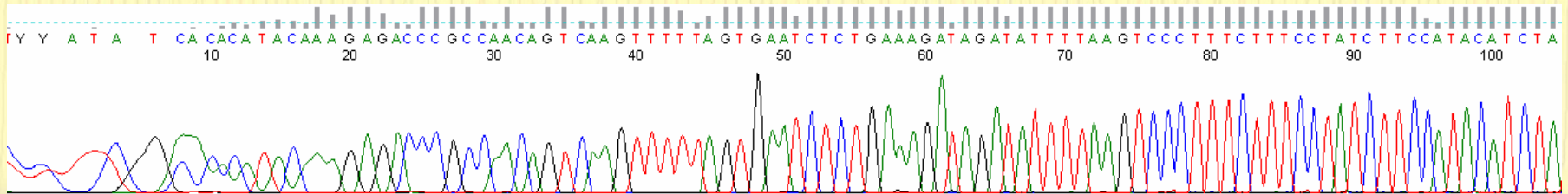


ddNTPs odlišně fluorescenčně značené – sekvenátor detekuje jejich signál

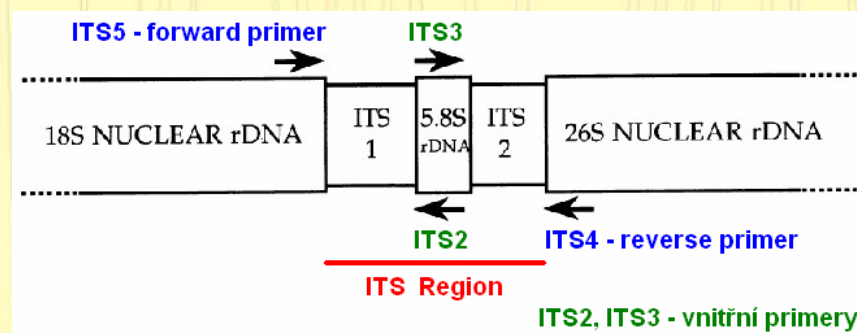
Sangerovo sekvenování DNA

Data ze sekvenátoru:

- jeden běh sekvenace pokryje 500-900 bp



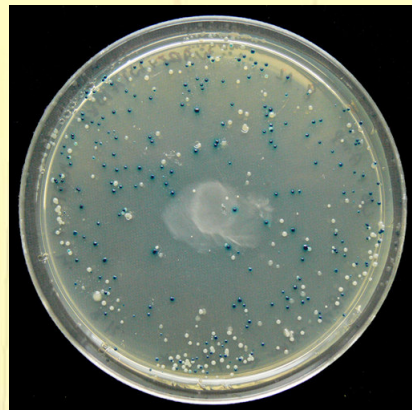
- vizualizace a editace - free programy:
FinchTV – prohlížení a editace signálu sekvenace
BioEdit (BioLign) – prohlížení a editace sekvencí
- u delších úseků nutné sekvenování z reverse směru (za použití reverse PCR primeru), případně pomocí vnitřních primerů



Sangerovo sekvenování DNA

Jak sekvenovat vzorky společenstev obsahující více druhů?

- molekuly lze separovat **klonováním**:
 - **jednotlivé** molekuly PCR produktu vloženy pomocí vektoru (plazmidu) do bakterií
 - předpoklad: **1** buňka přijme pouze **1** vektor, z každé **1** buňky bakterie vyroste jediná homogenní kolonie → **1** sekvence
 - jednotlivé bakteriální kolonie použity pro amplifikaci dané molekuly PCR produktu → sekvenace



modrobílá selekce:

bílé bakteriální kolonie obsahují inzert

Práce s databázemi DNA sekvencí

[NCBI \(National Center for Biotechnology Information\)](#)

[BLAST \(Basic Local Alignment Search Tool\)](#) <http://www.ncbi.nlm.nih.gov/>

➤ **algoritmus pro hledání homologických sekvencí**

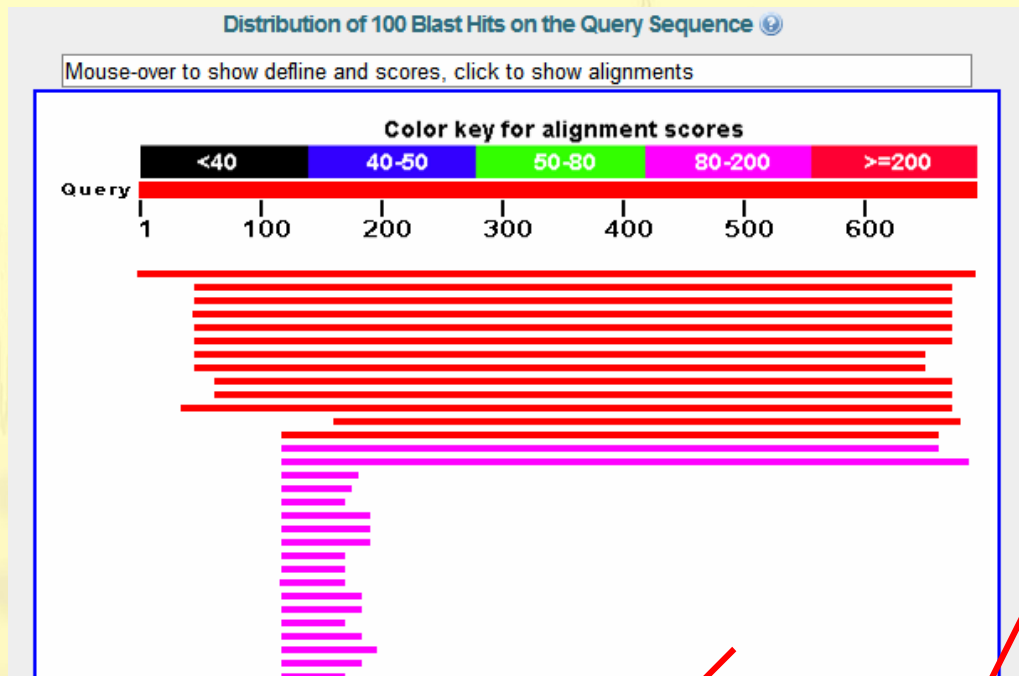
- přibližné ověření identity sekvence
- výsledek ovlivněn např. tím, zda je daný úsek / tax. skupina zastoupená v databázi

➤ **princip algoritmu:**

- naše sekvence (Query) → algoritmus z ní použije kratší motivy (words), prohledává jimi databázi
- pokud word nalezeno v sekvenci z databáze (Sbjct), dále v ní prohledává jeho okolí
- pokud celková podobnost přesáhne určitý limit, sekvenci vybere

Práce s databázemi DNA sekvencí

BLAST



vizualizace Query coverage
(~ jak dlouhý je homolog. úsek srovnávaných sekvencí)

pravděpodobnost, že podobnost sekvencí je náhodná

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
FJ546412.1	Syntrichia ruralis chloroplast, complete genome	1077	1077	100%	0.0	94%	
EU854273.1	Fissidens luisieri isolate 2155 trnM-trnV intergenic spacer and tRNA-Val	754	754	90%	0.0	88%	
EU854274.1	Fissidens luisieri isolate 2156 trnM-trnV intergenic spacer and tRNA-Val	754	754	90%	0.0	88%	
EU854283.1	Fissidens serrulatus isolate 2167 trnM-trnV intergenic spacer and tRNA-Val	750	750	90%	0.0	88%	
EU854282.1	Fissidens serrulatus isolate 2145 trnM-trnV intergenic spacer and tRNA-Val	749	749	90%	0.0	88%	
EU854284.1	Fissidens serrulatus isolate 2168 trnM-trnV intergenic spacer and tRNA-Val	749	749	90%	0.0	88%	
EU854271.1	Fissidens luisieri isolate 2139 trnM-trnV intergenic spacer and tRNA-Val	730	730	87%	0.0	88%	
EU854272.1	Fissidens luisieri isolate 2160 trnM-trnV intergenic spacer and tRNA-Val	730	730	87%	0.0	88%	
EU854277.1	Fissidens osmundoides isolate 2164 trnM-trnV intergenic spacer and tRNA-Val	669	669	88%	0.0	86%	
EU854276.1	Fissidens osmundoides isolate 2165 trnM-trnV intergenic spacer and tRNA-Val	664	664	88%	0.0	86%	
AP005672.1	Physcomitrella patens subsp. patens chloroplast DNA, complete genome	628	628	92%	1e-176	84%	
JN006037.1	Funaria hygrometrica isolate FH32 tRNA-Val (trnV) gene, partial sequence	407	407	74%	6e-110	81%	
AB299142.1	Takakia lepidiozoides chloroplast DNA, partial sequence	261	261	78%	5e-66	76%	
DQ821119.1	Angiopteris evecta chloroplast, complete genome	143	143	78%	2e-30	72%	
HM222519.1	Ptilidium pulcherrimum chloroplast, complete genome	119	119	81%	3e-23	72%	
AP004638.1	Psilotum nudum chloroplast DNA, complete genome	104	104	9%	8e-19	95%	
EF380353.1	Dioscorea elephantipes chloroplast, complete genome	93.5	93.5	8%	2e-15	95%	
AJ627251.1	Nymphaea alba chloroplast, complete genome	93.5	93.5	7%	2e-15	98%	

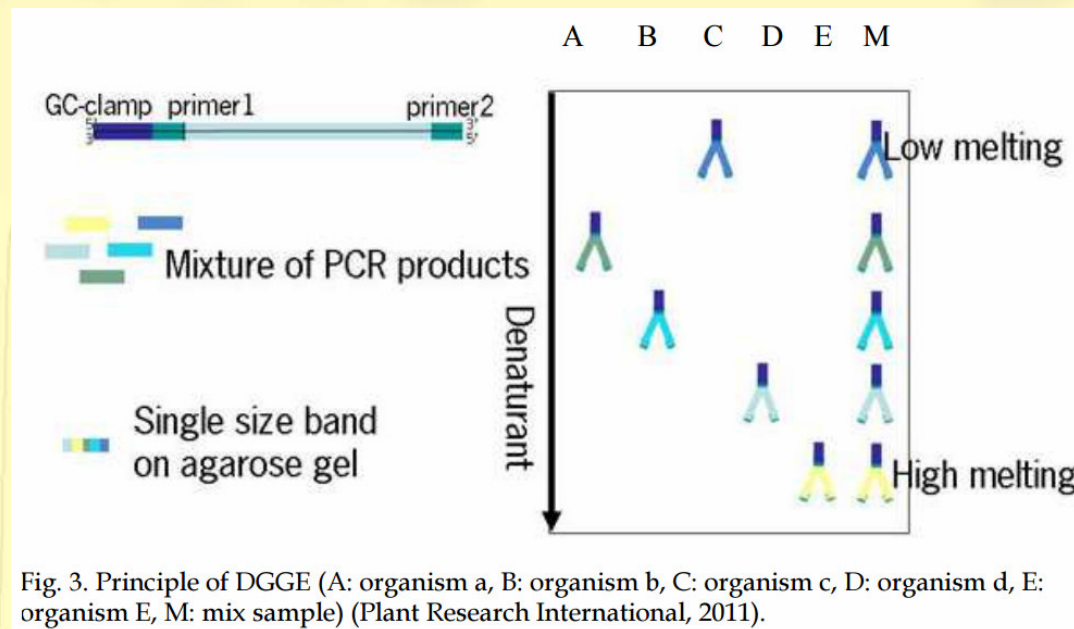
je dobré brát v potaz:

někdy může např. Max. Ident. 95%, ale jen na krátkém úseku Query

Nepřímá detekce DNA variability

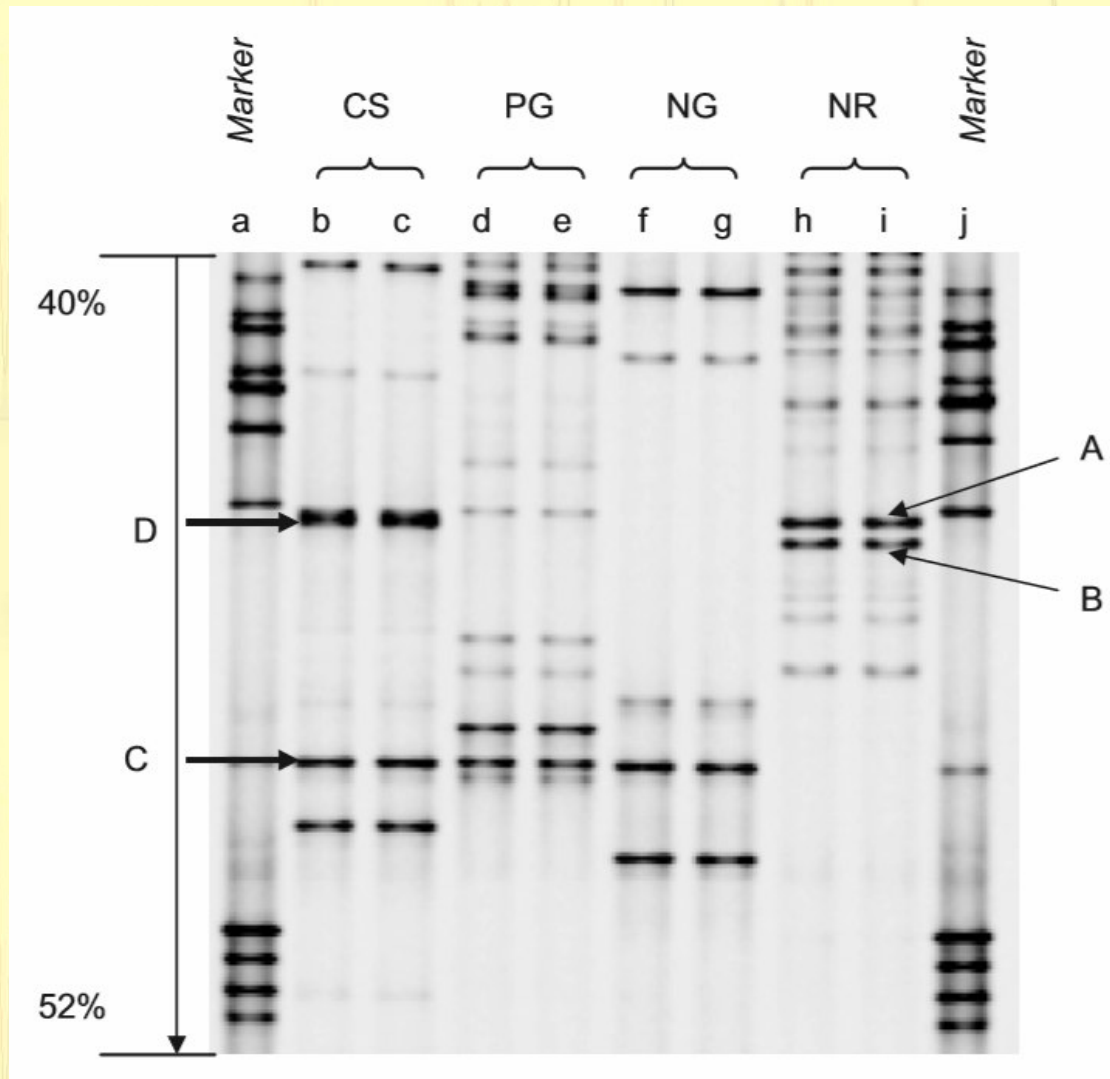
DGGE (denaturing gradient gel electrophoresis)

- elektroforéza v gradientovém denaturačním polyakrylamidovém gelu
- primery vytvoří stabilizační část produktu, GC-clump
- detekce variability - separací fragmentů podle rozdílů v délce fragmentů + rozdílech v sekvenci, které ovlivní denaturaci (melting, tání) fragmentu



Nepřímá detekce DNA variability

DGGE



Nepřímá detekce DNA variability

DGGE

Z. Liang et al. / Soil Biology & Biochemistry 40 (2008) 956–966

963

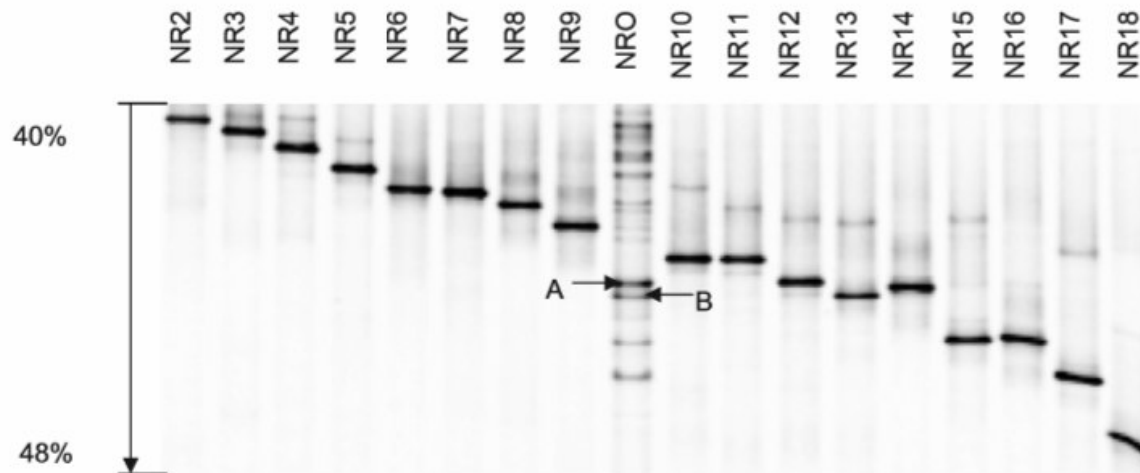


Fig. 4. DGGE profile of AMF 18S rDNA clone library from NR (red cedar forest ecosystem at the Nebraska National Forest, Halsey, Nebraska) soil sample. NR2–NR18: coded single clone (clone NR1, which is chimeric is not shown on the gel). Lane NRO: original fingerprint of NR sample. Arrows marked with A and B represent the two dominant bands shown in Fig. 3 that were excised and sequenced. Vertical arrow indicates that the denaturant gradient increases from 40% on the top to 48% at the bottom of the gel with 100% denaturant defined as containing 7 M urea and 40% (v/v) formamide.

- možnost přímé identifikace druhů – pokud je možné daný organismus získat v čisté formě (kultuře), nebo máme jeho sekvenci v plazmidu z klonování, zpracujeme izolát na samostatný DGGE profil

Nepřímá detekce DNA variability

Výhody DGGE

- nízká cena analýzy – možno zpracovat víc vzorků než klonováním+sekvenací
- možnost sekvenovat DNA fragmenty

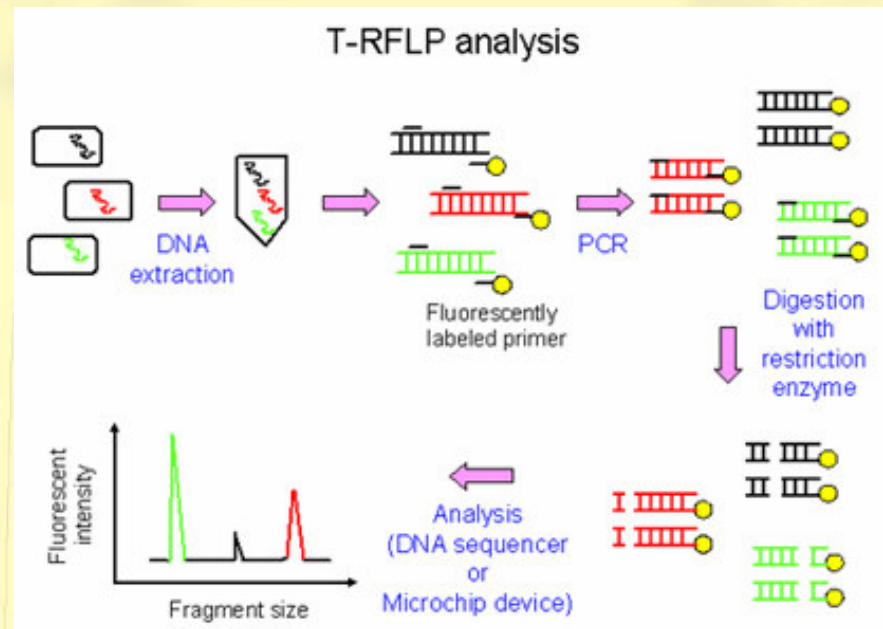
Nevýhody DGGE

- speciální vybavení - elektroforéza
- relativně dost manuální práce
- nižší rozlišení druhů než ostatní metody
- pokud primery amplifikují nespecificky i necílové organismy, a zároveň nemáme databázi předpokládaných profilů, tak výsledek zkreslen signálem z necílových organismů

Nepřímá detekce DNA variability

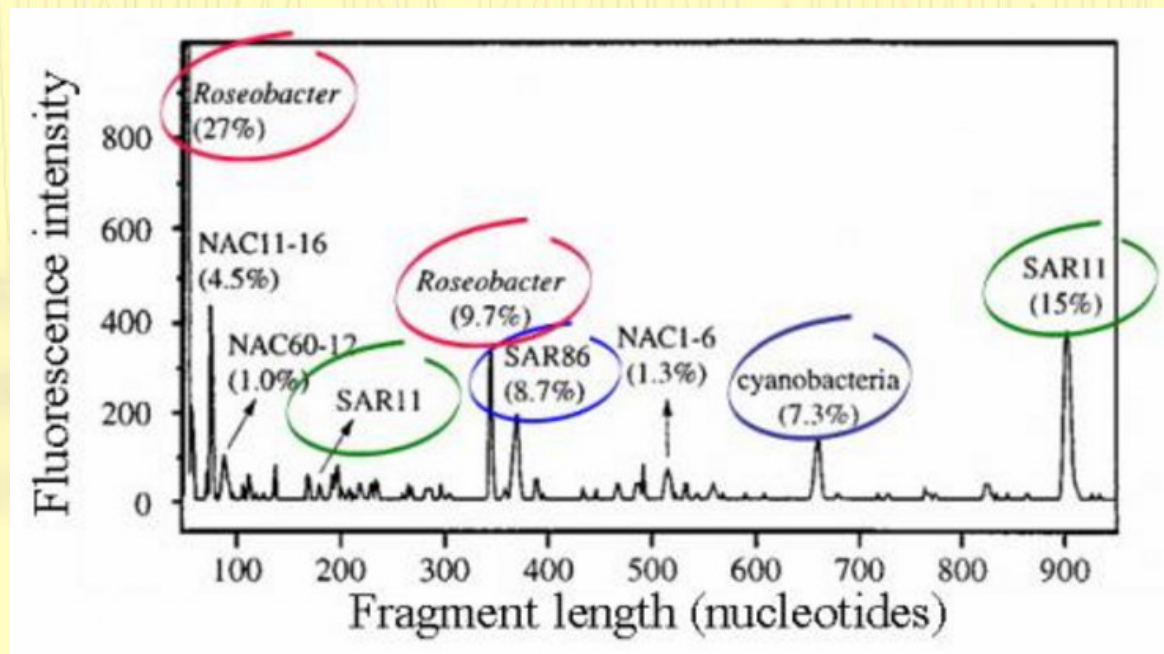
T-RFLP (Terminal Restriction Fragment Length Polymorphism)

- primery s fluorescenčním barvivem
- variabilitu generuje štěpením restrikním enzymem (cílené na variabilní oblasti úseku DNA)
- detekce variability – elektroforetickou separací fragmentů na sekvenátoru (fragmentační analýza, komerční firmy)



Nepřímá detekce DNA variability

T-RFLP



- podobně jako u DGGE možnost přímé identifikace druhů – pokud si vytvoříme databázi předpokládaných velikostí píků jednotlivých druhů

Nepřímá detekce DNA variability

Výhody T-RFLP

- nízká cena analýzy – možno zpracovat víc vzorků než klonováním+sekvenací
- stačí obvyklé přístrojové vybavení
- patrně větší rozlišení druhů než DGGE
- z výšky píků možno kvantifikovat(?) abundanci druhů

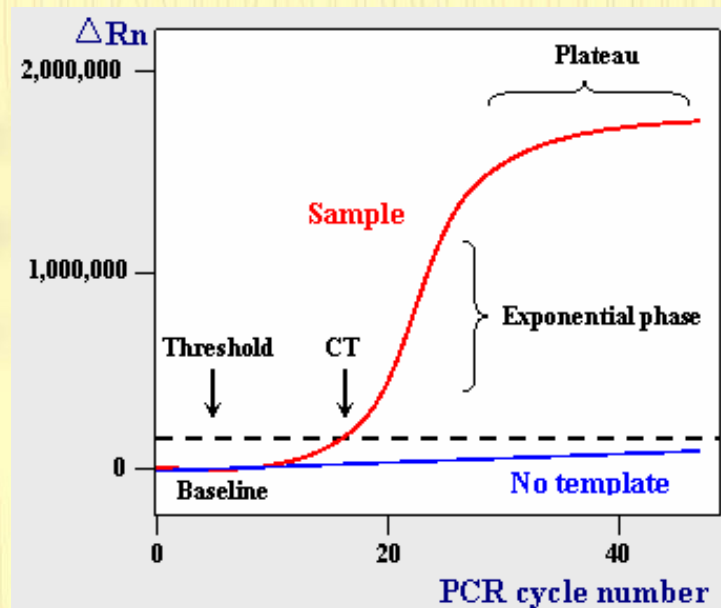
Nevýhody T-RFLP

- nižší rozlišení druhů než klonování+sekvenace
- pokud primery amplifikují nespecificky i necílové organismy, a zároveň nemáme databázi předpokládaných profilů, tak výsledek zkreslen signálem z necílových organismů

Kvantifikace organismů pomocí DNA

RT-PCR (qPCR)

- musíme mít primery na studovaný úsek DNA o délce 100-150(-500) bp
- detekce / kvantifikace amplifikované molekuly z parametrů kinytiky PCR amplifikace, pomocí specializovaných thermocyklerů

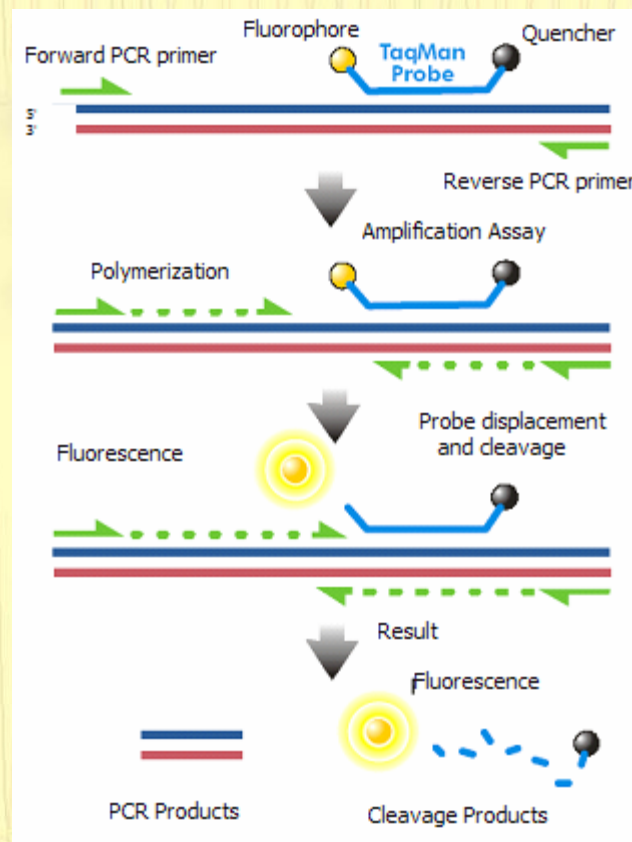


- umožňuje i vysoce specifickou detekci – primery mohou být cílené např. na konkrétní druh organismu

Kvantifikace organismů pomocí DNA

RT-PCR (qPCR)

- kvantifikace syntetizovaného dsDNA produktu barvivem SybrGreen
- nebo častěji přesnější kvantifikace pomocí **hybridizačních sond** – opět mohou být vysoce (druhově) specifické



Kvantifikace organismů pomocí DNA

Výhody RT-PCR (qPCR)

- jediná opravdu přesná metoda kvantifikace

Nevýhody RT-PCR (qPCR)

- speciální thermocykler
- nutnost designu primerů, popř. sond (pokud nejsou k dispozici)
- nutná verifikace:
 - zda je design specifický (= kvantifikujeme cílový organismus)
 - u rDNA ideálně i zjištění počtu kopií analyzovaného úseku (organismy se navzájem mohou lišit v počtu kopií, to ovlivní výsledky)

Next generation sequencing

Next Generation Sequencing (NGS)

High-throughput Sequencing

Massive Parallel Sequencing

- v jednom runu nezávisle sekvenuje tisíce až miliony molekul
= není nutné klonování
- **velké množství dat za nižší cenu** než klasické (např. Sangerovo) sekvenování
- **cena – desítky až stovky tisíc Kč za 1 běh (run)**
- méně manuální práce - vlastní sekvenování obvykle provádí komerční firma
- Illumina, Ion Torrent, 454 pyrosequencing a několik dalších nových platform

Next generation sequencing

Příprava templátu (DNA library) pro NGS sekvenaci:

připravíme PCR produkt (amplicon sequencing)

- **úprava produktu přidáním NGS adaptorů:**

pro navázání fragmentů na sekv. destičku a nasedání sekv. primerů
sekvence adaptorů vždy specifická pro danou platformu

a) pomocí naší **PCR s tzv. fúzními primery** (= oligonukleotid se specifickou NGS sekvencí + vlastní sekvencí specifickou pro daný typ templátu)

b) nebo **ligací** adaptorů (obvykle už provádí sekvenační firma)

- **clonal amplification:** DNA fragmenty se pomocí adaptorů vážou na specifický typ nosiče a proběhne jejich PCR amplifikace

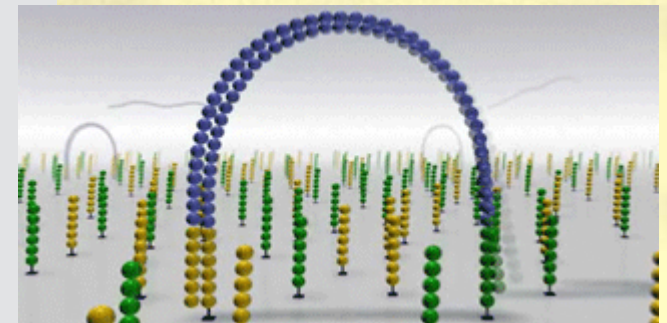
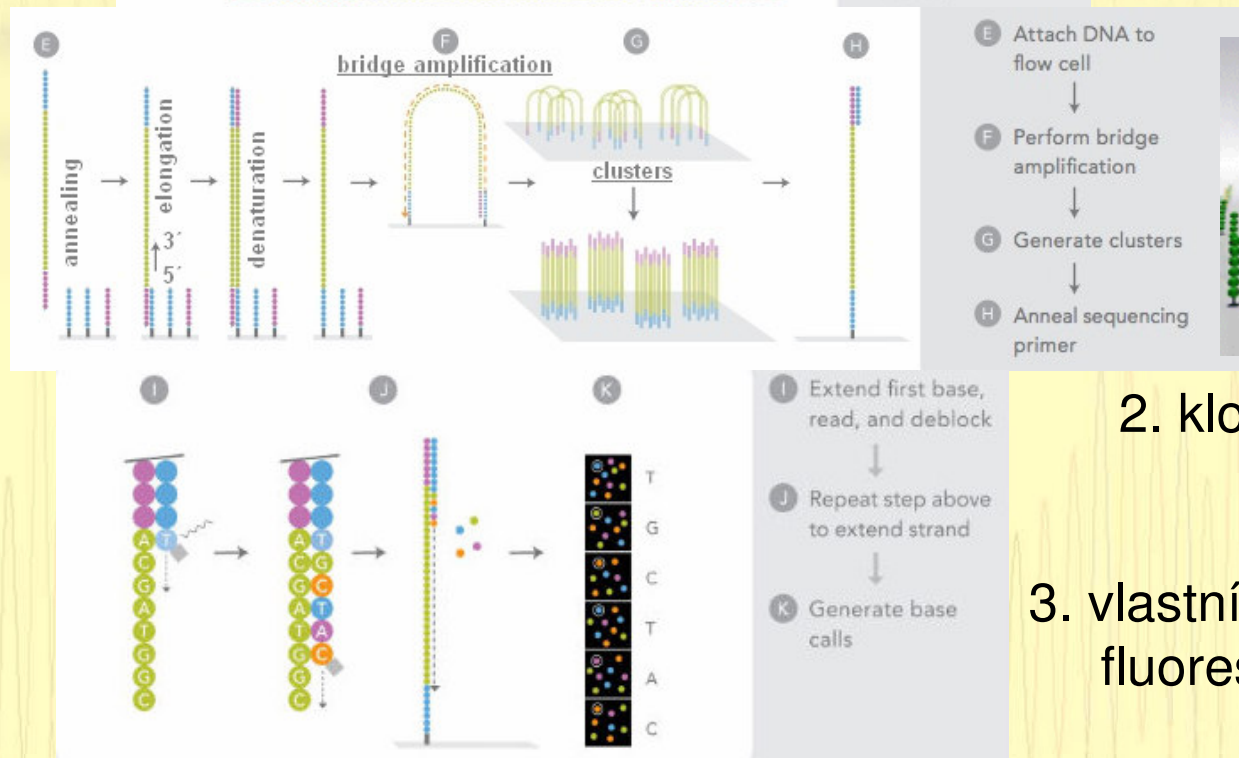
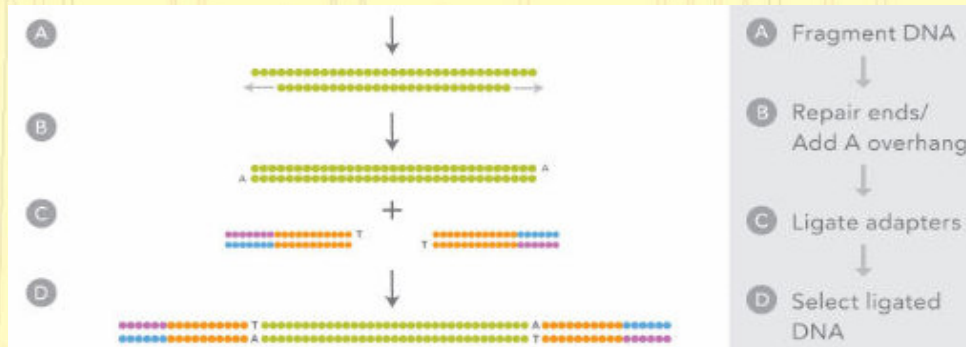
- **vlastní sekvenace**

Next generation sequencing - Illumina

Illumina Workflow



1. příprava templátu



2. klonální 'bridge amplification'

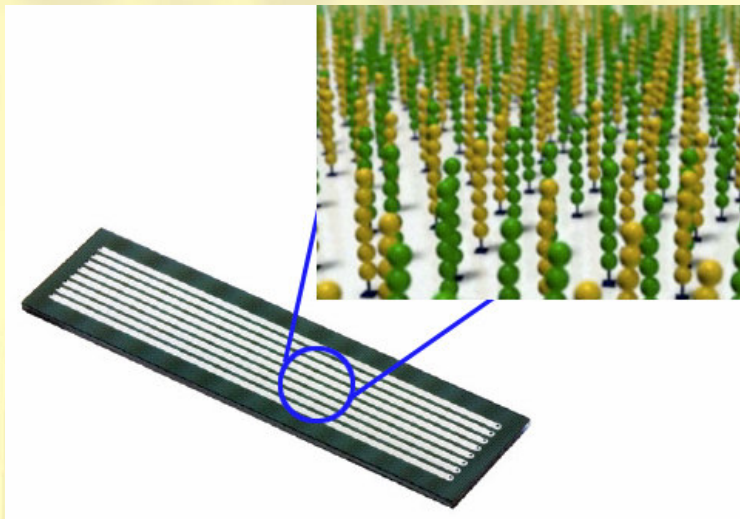
3. vlastní sekvenace = detekce fluorescenčního signálu

Next generation sequencing - Illumina

- ▶ **výstup:** stovky milionů sekvencí o délce 50-300 bp (záleží na konkrétním typu přístroje)

paired end sekvenace – oboustranné čtení

Illumina MiSeq – ready max. 300 bp, tj. až 600 bp při sekvenaci z obou konců dané molekuly DNA (2 x 300 bp)

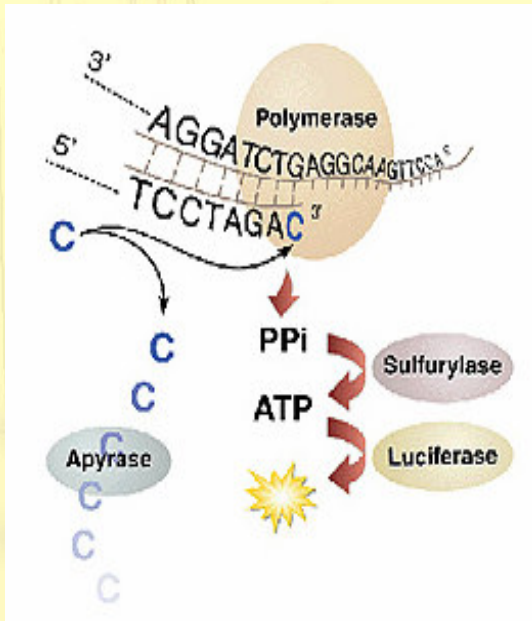


nosič pro bridge amplification

- ▶ Illumina HiSeq – ready max. 2 x 250 bp, možnost analýzy více vzorků najednou (8 lanes)

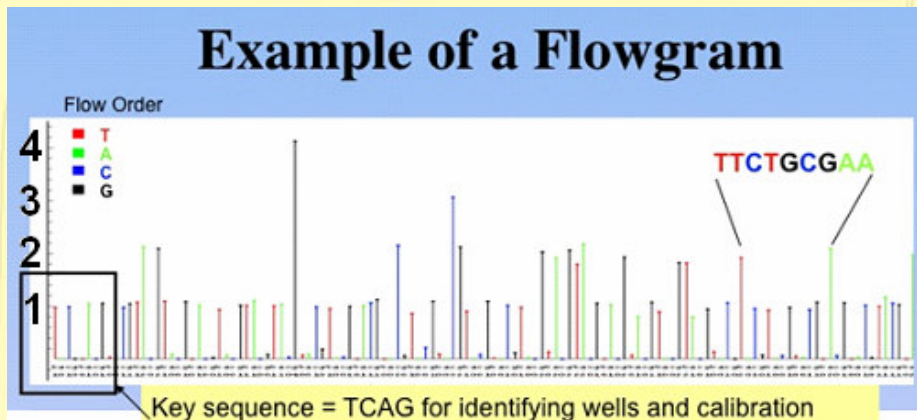
- ▶ **přesnost:** 99.9% (~ jako Sangerovo sekvenování)

Next generation sequencing – 454 pyrosequencing



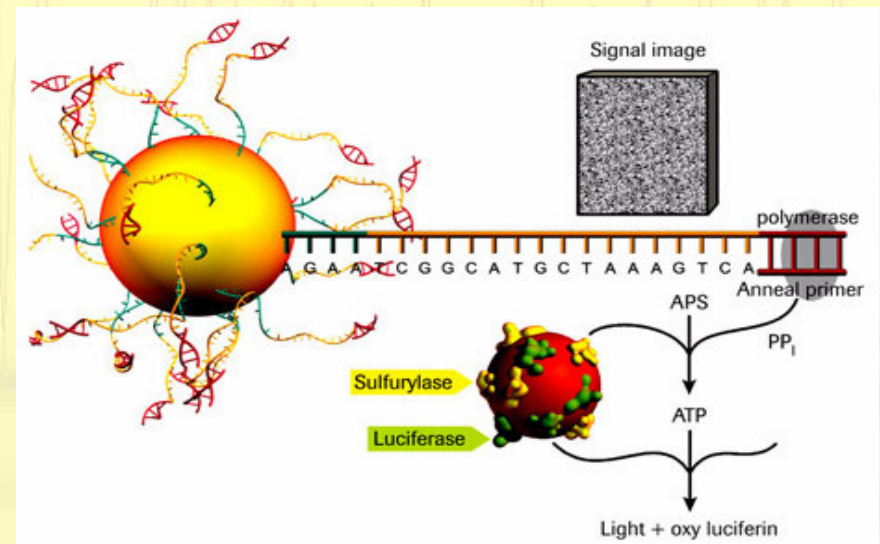
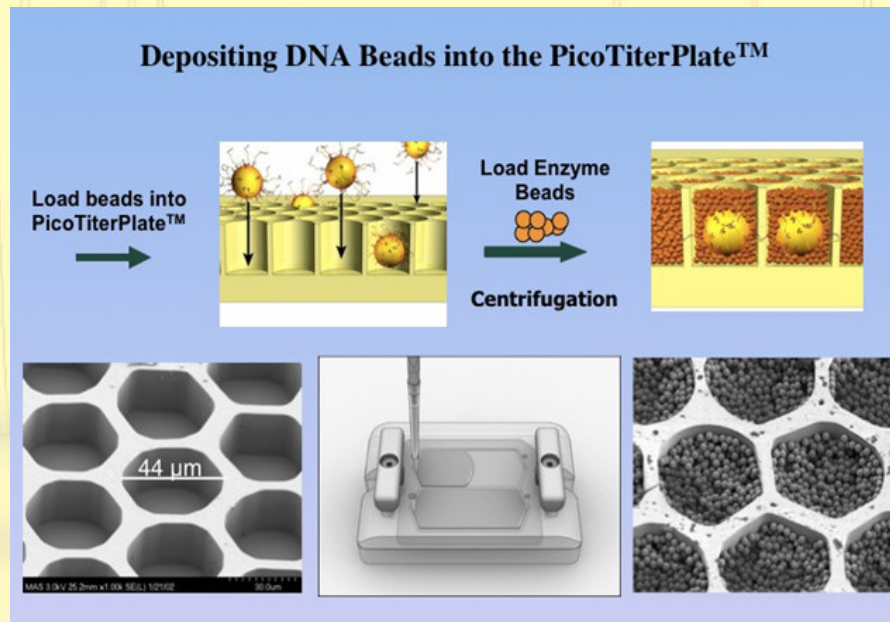
princip 454 pyrosekvenování

- po inkorporaci dNTP se odštěpí **pyrofosfát** (PPi)
 - ATP sulfuryláza jej přemění na **ATP**
 - luciferáza za přítomnosti ATP přemění **luciferin** na **oxyluciferin** → **záblesk**
 - (→ apyráza – degraduje neinkorp. dNTP a ATP)
- v každém cyklu na templát pouštěn jen **jeden typ dNTP**
 - např. když se přidá **T** a vznikne záblesk, tak na dané pozici je opravdu **T**



- ▶ osa y: výška píku záblesku odráží počet inkorporovaných nukleotidů daného typu
- ▶ delší homopolymerní sekvence → hlavní zdroj chyb 454 sekvenace

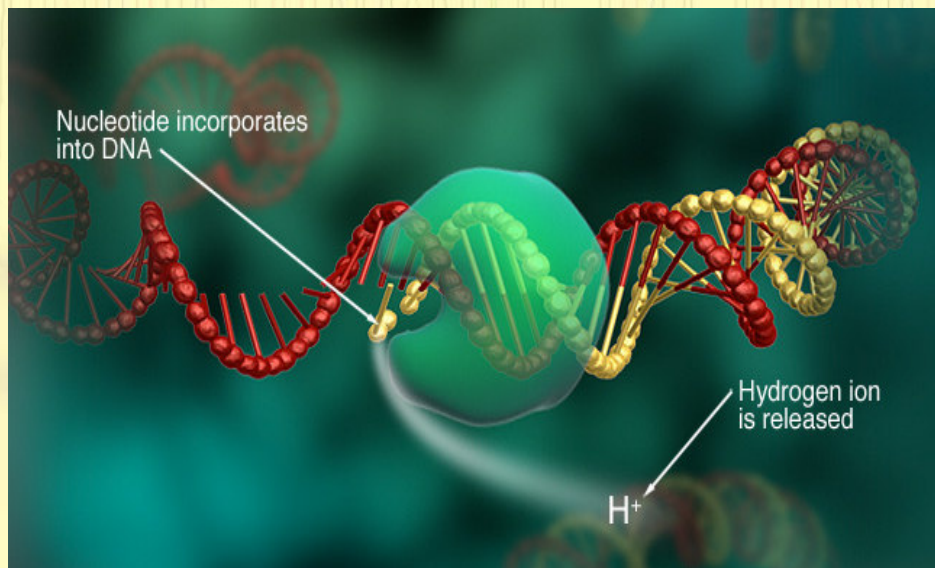
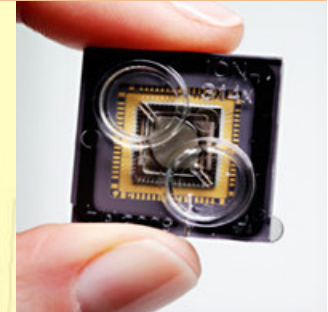
Next generation sequencing – 454 pyrosequencing



- ▶ **výstup:** desítky až stovky tisíc sekvencí o délce < 1000 bp, (nejčastěji max. 400-700 bp, podle typu použité sequencing chemistry)
- ▶ **přesnost:** 99% (nižší než u Sangerova sekvenování!)
- ▶ po r. 2016 přestává výroba kitů (metoda končí)

Next generation sequencing – Ion Torrent

princip Ion Torrent: navázání dNTP při prodlužování řetězce DNA vede k odštěpení H^+ → **detekována změna pH**



- ▶ podobně jako při 454 postupně pouštny jednotlivé typy dNTP
- ▶ 100 tis. až miliony 200-400 bp readů
- ▶ nejlevnější run (< 10 tis. Kč) i přístroj



homopolymerní sekvence (TT): jako u 454 – např. 2x vyšší pík



Next generation sequencing

Jak analyzovat víc vzorků najednou?

- ▶ předpokladem je odlišit jednotlivé sekvence = určit, kterému ze vzorků patří
 - a) **fyzická separace** jednotlivých vzorků - možné pouze u určitých typů dané platformy (např. Illumina HiSeq, 454 GS FLX)

454 gaskets



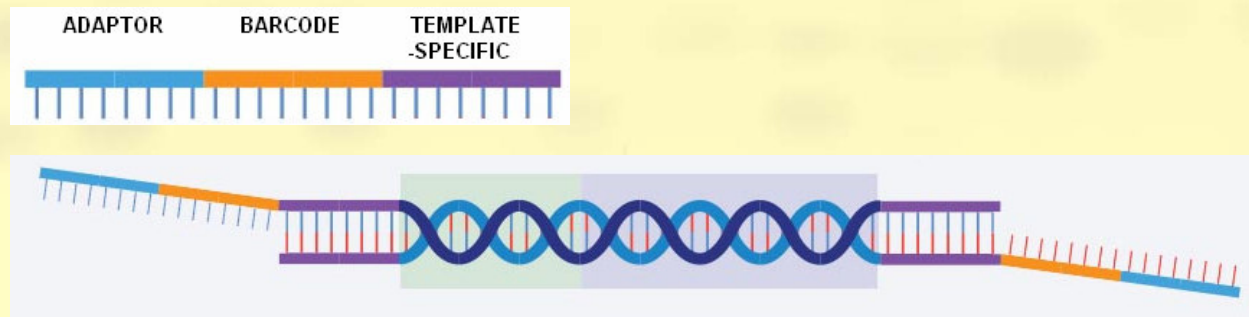
Illumina lanes

Next generation sequencing

Jak analyzovat víc vzorků najednou?

- předpokladem je odlišit jednotlivé sekvence = určit, kterému ze vzorků patří

b) multiplexování, pooled samples: každý dílčí vzorek identifikován připojením několik bp dlouhé 'značkovací' sekvence (barcode; index; tag; MID = Multiplex Identifier): součástí adaptorů ligovaných na DNA fragmenty, nebo součástí PCR primerů



barcodey možné i kombinovat: např. 10 různých forward barcodeů + 10 různých reverse barcodeů umožní celkem $10 \times 10 = 100$ kombinací = možné poolovat 100 vzorků

- vede ke snížení počtu sekvencí na 1 vzorek, ale šetří peníze

Next generation sequencing

Amplicon sequencing – NGS analýza společenstev

- ▶ **počet readů** – zásadní u studií společenstev
(př.: očekáváme 1000 druhů / vz. – 500 readů / vz. bude určitě málo
10 500 dostačující)
- ▶ **při porovnávání diverzity vzorků je nutný stejný počet readů / vz.**
- ▶ **Quality Check (QC)** – NGS data mají vyšší chybovost než Sangerovo sekvenování, z technických důvodů není možné editovat raw signál
FASTQ (Illumina), .sff (454) - datový výstup obsahuje sekvence bází + kvalitu jejich čtení
např. $Q > 20$ – daná báze byla přečtena s 99% přesností
quality trimming podle minimální průměrné Q value daného readu
dílčí báze s $Q < \text{threshold}$ vyhodnocené jako N (nebo brána za trimming point)

Next generation sequencing

Amplicon sequencing – analýza dat:

Software:

Mothur – třídění vzorků podle indexů, úprava hrubých dat i vlastní analýzy

SEED – český software s implementací nejdůležitějších programů (Mothur, Mafft, Uclust aj.)

Identifikace MOTU (molecular operational taxonomic unit, 'molekulární druh'):

- ▶ clusterování readů se sekvenční identitou – obvykle >97%
- ▶ pro každý cluster vytvořená idealizovaná sekvence, consensus
- ▶ identifikace konsensů se známými taxony z databáze pomocí BLAST search
- ▶ nejčastěji se hodnotí prezence / absence MOTU
- ▶ kvantifikace spíš nejistá: PCR bias – určité molekuly (MOTU) se mohou amplifikovat s vyšší efektivitou než jiné

Next generation sequencing

Amplicon sequencing – analýza dat:

Databáze sekvencí:

A) Vytvoříme si vlastní, např. pomocí klasického (Sanger) sekvenování:

- ▶ vhodné např. u lokálních studií, kdy sledujeme omezený počet druhů dané lokality (např. identifikace rostlin)

B) Veřejně dostupné databáze:

- ▶ optimálně takové, u kterých je garantovaná správnost tax. určení (neplatí např. pro GenBank záznamy v NCBI vyhledávání)

Ribosomal Database Project – bakterie, houby <http://rdp.cme.msu.edu/>

MaarjAM – arbuskulární mykorhizy <http://maarjam.botany.ut.ee/>

Next generation sequencing

Amplicon sequencing – analýza dat:

Ribosomal Database Project – příklad klasifikace sekvencí (Classifier)

Classifier: RDP Naive Bayesian rRNA Classifier Version 2.10, October 2014
Taxonomical Hierarchy: UNITE Fungal ITS trainset 07/04/2014
Query File: Tulasnella_GB.fasta
Query Submit Date: Thu Apr 21 04:56:18 EDT 2016



Display depth: Auto ▾ **Confidence threshold:** 95% ▾ [Refresh](#)

domain	%	Library
Fungi	100.0	

Hierarchy View (click a node to make it the root -- only show sequences assigned to that node with confidence above the threshold):

[null Root \(58 sequences\) \[show assignment detail for Root only\]](#) [download entire hierarchy as text file](#)

- » » domain Fungi (58)
 - » » » phylum Basidiomycota (58)
 - » » » » class Agaricomycetes (58)
 - » » » » » order Cantharellales (58)
 - » » » » » » family Tulasnellaceae (57)
 - » » » » » » » genus Epulorhiza (5)
 - » » » » » » » » species Epulorhiza_sp_FB1a|SH218365.06FU (2)
 - » » » » » » » » species Epulorhiza_sp_Eb3c|SH199356.06FU (3)
 - » » » » » » » » genus Tulasnella (6)
 - » » » » » » » » » species Tulasnella_sp_Sno4_4|SH199344.06FU (2)
 - » » » » » » » » » species Tulasnella_sp|SH199346.06FU (4)
 - » » » » » » » » genus Tulasnellaceae_unidentified (35)
 - » » » » » » » » » species Tulasnellaceae_sp|SH203305.06FU (4)
 - » » » » » » » » » species Tulasnellaceae_sp|SH199360.06FU (1)
 - » » » » » » » » » species Tulasnellaceae_sp|SH199345.06FU (2)
 - » » » » » » » » » species Tulasnellaceae_sp|SH205501.06FU (10)
 - » » » » » » » » » species Tulasnellaceae_sp|SH218358.06FU (4)
 - » » » » » » » » » species Tulasnellaceae_sp|SH199343.06FU (4)
 - » » » » » » » » » species Tulasnellaceae_sp|SH218360.06FU (5)
 - » » » » » » » » » species Tulasnellaceae_sp|SH199359.06FU (1)
 - » » » » » » » » » species Tulasnellaceae_sp|SH199353.06FU (1)
 - » » » » » » » » » unclassified_Tulasnellaceae_unidentified (3)
 - » » » » » » » » » unclassified_Tulasnellaceae (11)

Next generation sequencing

Amplicon sequencing – NGS analýza společenstev

Na co dávat pozor:

- ▶ kontroly:
 - ▶ negativní – neměla by v ní proběhnout PCR amplifikace; je vhodné negativní kontroly i osekvenovat
 - ▶ pozitivní – ke vzorku před zpracováním přidáme známé množství DNA která se v experimentu nevyskytuje; sledujeme, s jakou efektivitou 'přežije' celý proces zpracování
 - ▶ vnitřní standard – relativní kvantifikace pomocí pozitivní kontroly: pro porovnávání vzorků vyčíslíme četnost daného MOTU vzhledem k pozitivní kontrole
- ▶ vyhazovat MOTU které se ve vzorku vyskytují s malou četností (singletony, nebo do urč. %) – potenciální kontaminace, artefakty