

Využití molekulárních markerů v systematice a populační biologii rostlin

11. Next generation sequencing (NGS)

Next generation sequencing (NGS)

- první generace – Sangerovo sekvenování
- další generace – paralelní sekvenování mnoha molekul (PCR namnožených)
- ještě další generace – single molecule sequencing

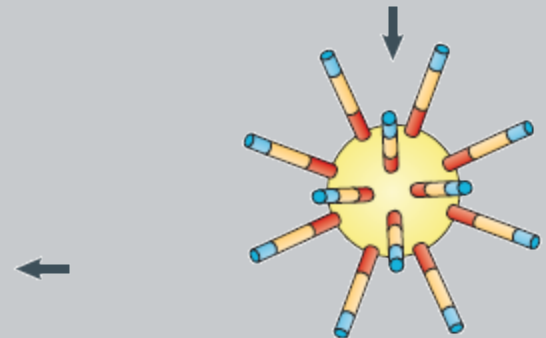
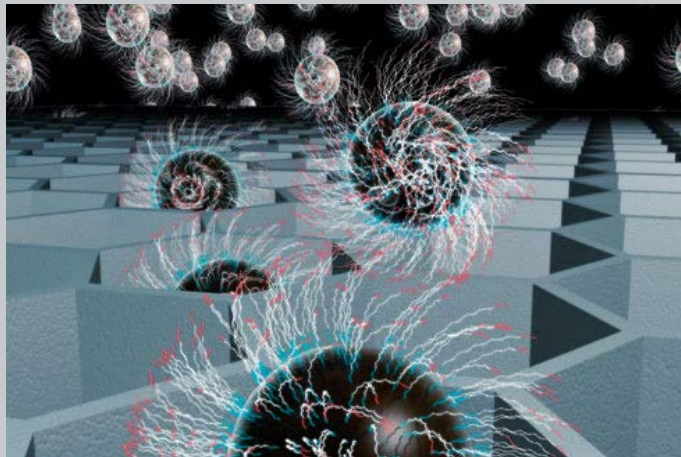
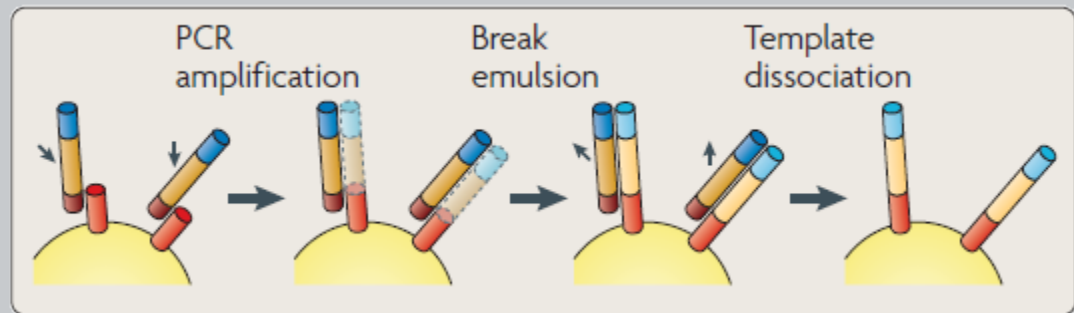
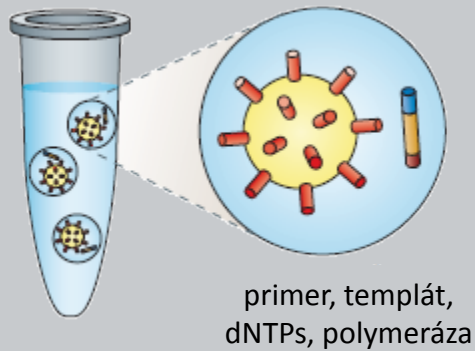
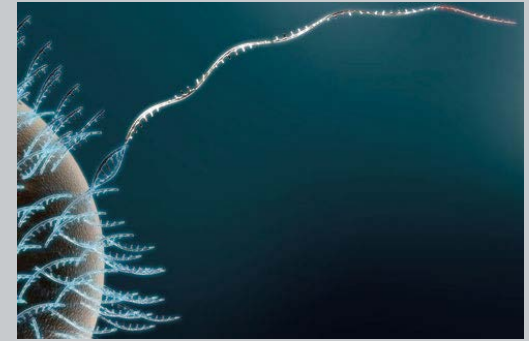
Obecný postup NGS

- příprava knihovny
 - náhodné štěpení genomové DNA na fragmenty, ligace adaptorů
- prostorová separace jednotlivých fragmentů
- dvě „základní“ možnosti sekvenování
 - sekvenování klonálně amplifikovaných templátů
 - emulsion PCR (emPCR)
 - solid-phase amplification
 - jednomolekulové sekvenování
- imobilizace k povrchu
- vlastní sekvenování a záznam dat
 - pyrosekvenování (Roche/454)
 - cyclic reversible termination (CRT) (Illumina/Solexa)
 - sequencing by ligation (SOLiD)
- analýza dat (analýza obrazových dat, kontrola kvality, ...)

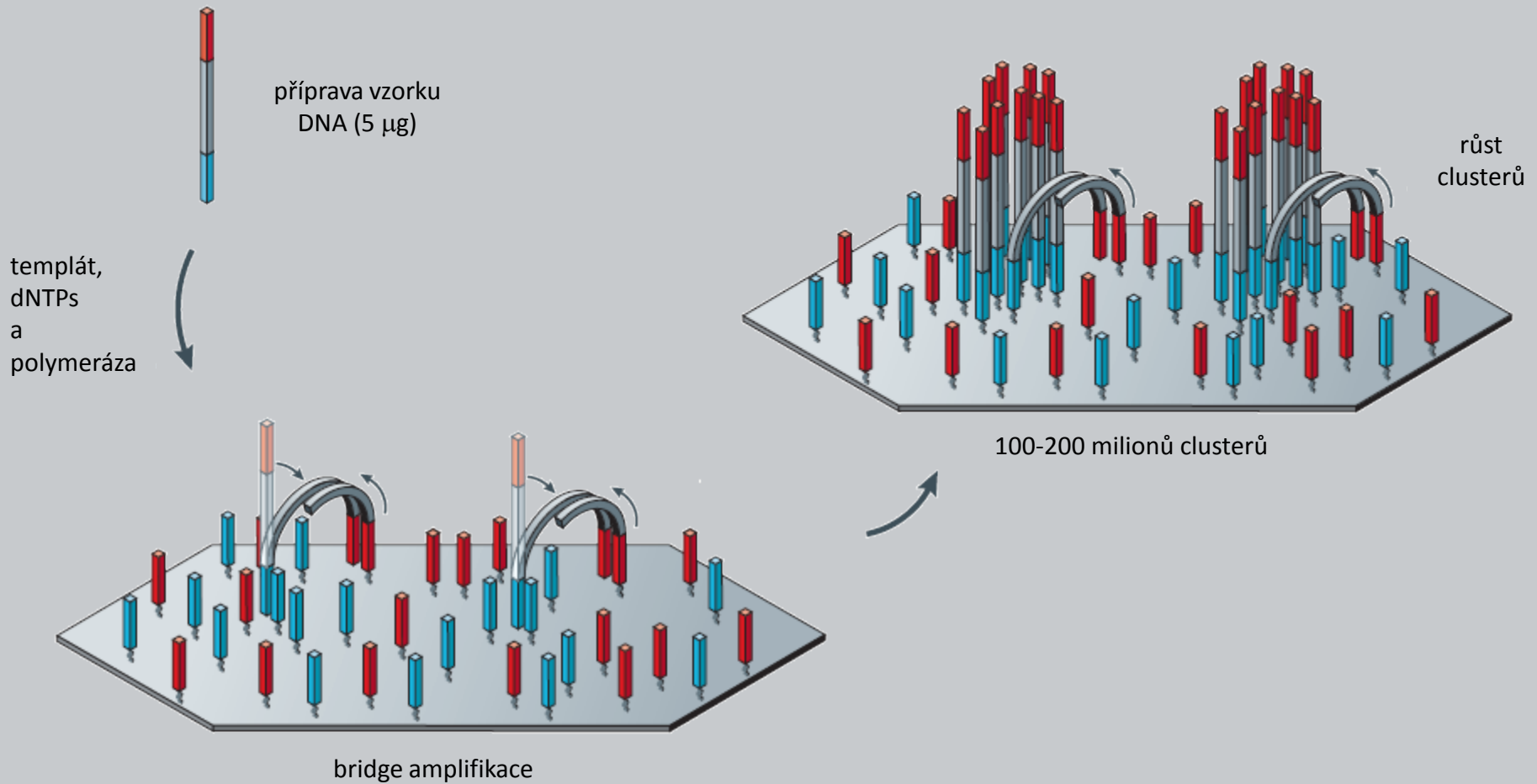
Nejrozšířenější NGS platformy

- Roche/454 – emPCR, pyrosekvenování
- Illumina/Solexa – solid phase (bridge), CRT
- Life/APG (SOLiD) – emPCR, ligation
- Pacific Biosciences – single molecule real time (SMRT)

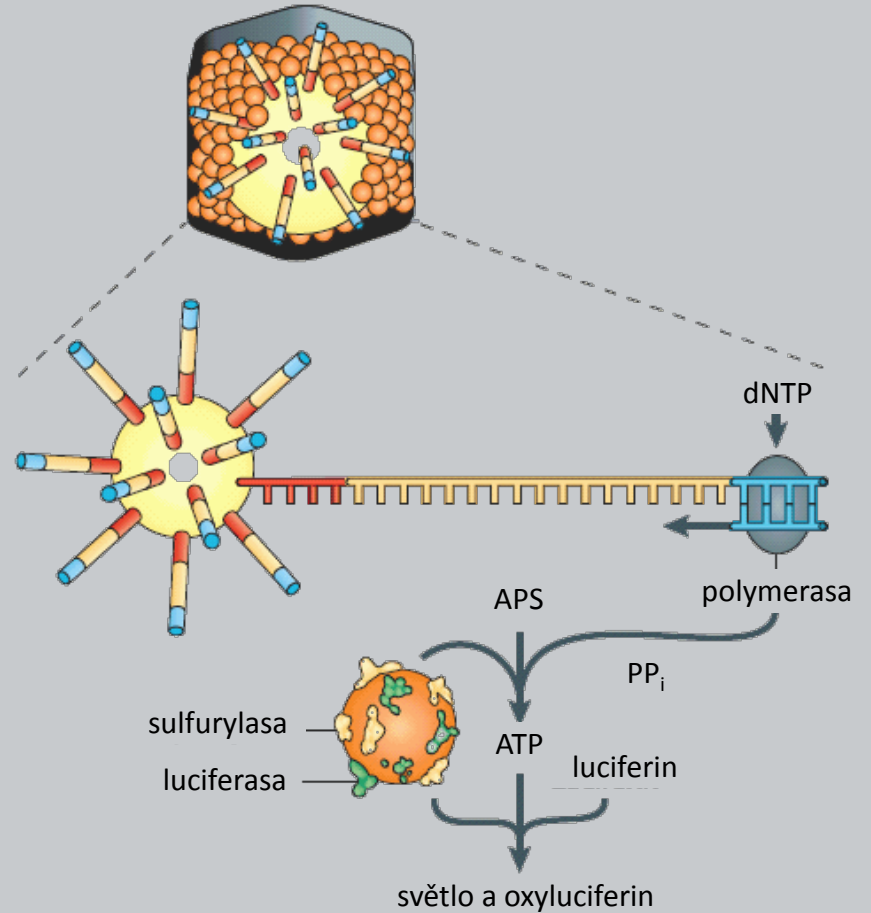
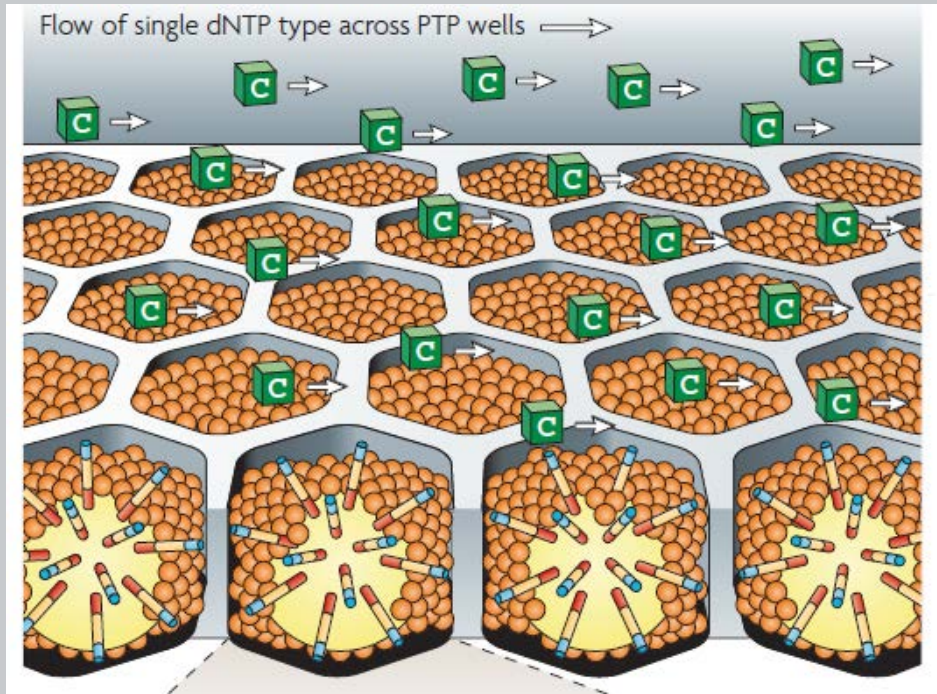
Emulsion PCR



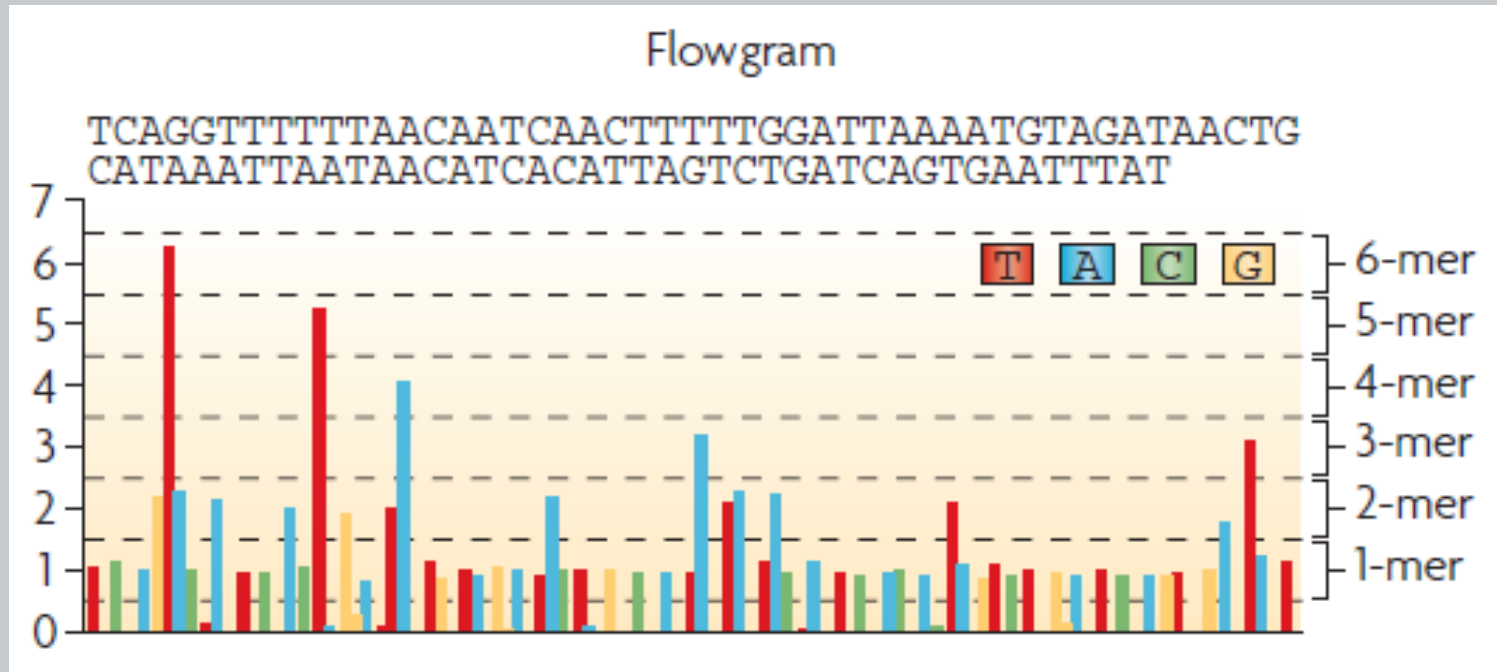
Solid-phase amplification



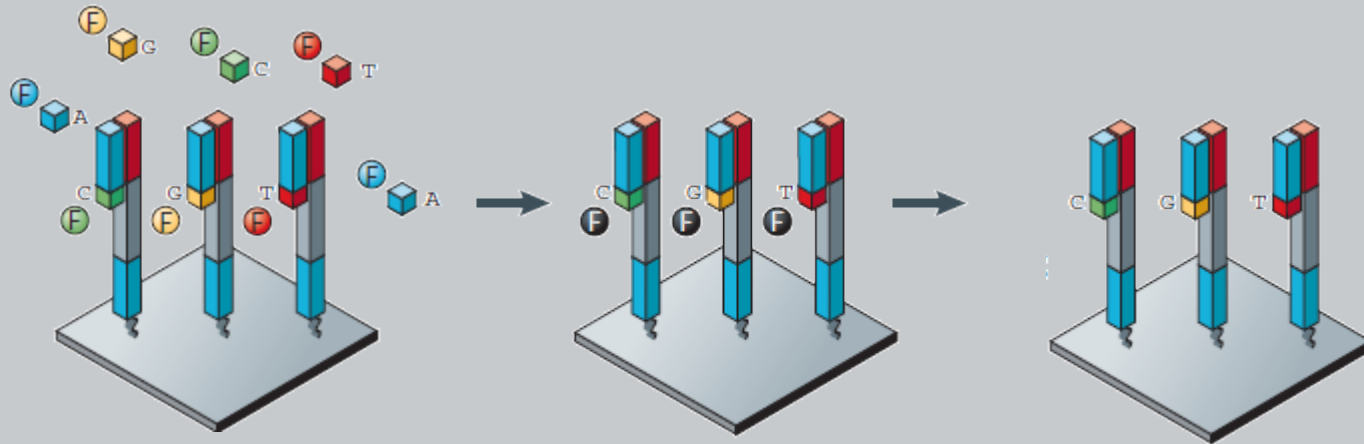
Pyrosekvenování



Pyrosekvenování



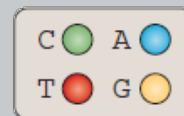
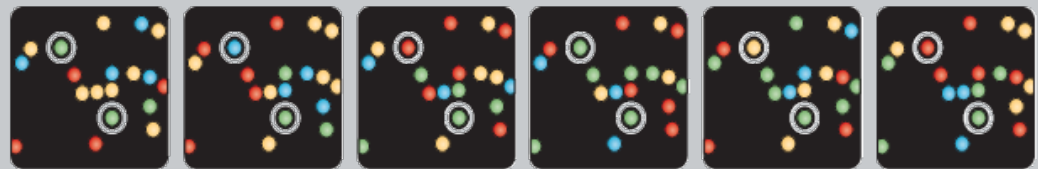
Cyclic reversible termination



začlenění
všech čtyř
nukleotidů,
každý je
značený jinou
barvou

odmytí zbylých
nukleotidů,
záznam
čtveřice barev

odstranění
barvičky,
odmytí



Nahoře: CATCGT
Dole: CCCCC

Srovnání platforem

	příprava templátu	chemie	délka čtení (báze)	doba běhu (dny)	Gb na jeden běh	výhody	nevýhody
Roche/454 (GS Jr., FLX)	emPCR	pyrosekve- nování	350-750	0.35	0.65	dlouhé čtení, rychlé	drahé v přepočtu na bázi, vysoká chybovost u homopolymerů
Illumina/ Solexa (GAII, MiSeq, HiSeq)	solid- phase bridge PCR	cyclic reversible termination	75-250	0.8-11	3-600	nejrozší- řenější	nízká možnost multiplexování vzorků?
Life/APG (SOLiD 3)	emPCR	sequencing by ligation	50	7-14	30-50	vysoká spolehli- vost čtení	krátké délky čtení, dlouhá doba běhu

Co dále se sekvencemi ?

- FASTA + quality scores
- assembling
 - de novo assembly
 - využití referenčního genomu (*reference-guided*)
- využití sekvencí pro
 - hledání variability (SNP)
 - hledání mikrosatelitů
 - identifikace vhodných regionů pro fylogenetické studie
 - fylogenomika – fylogeneze na základě celých genomů (např. cpDNA) a
 - ...

Assembling

generování jednotlivých sekvencí (reads)



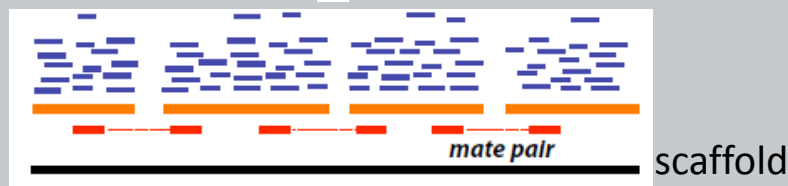
nalezení překrývajících se readů



assemblování readů do contigů



spojení contigů do scaffoldů



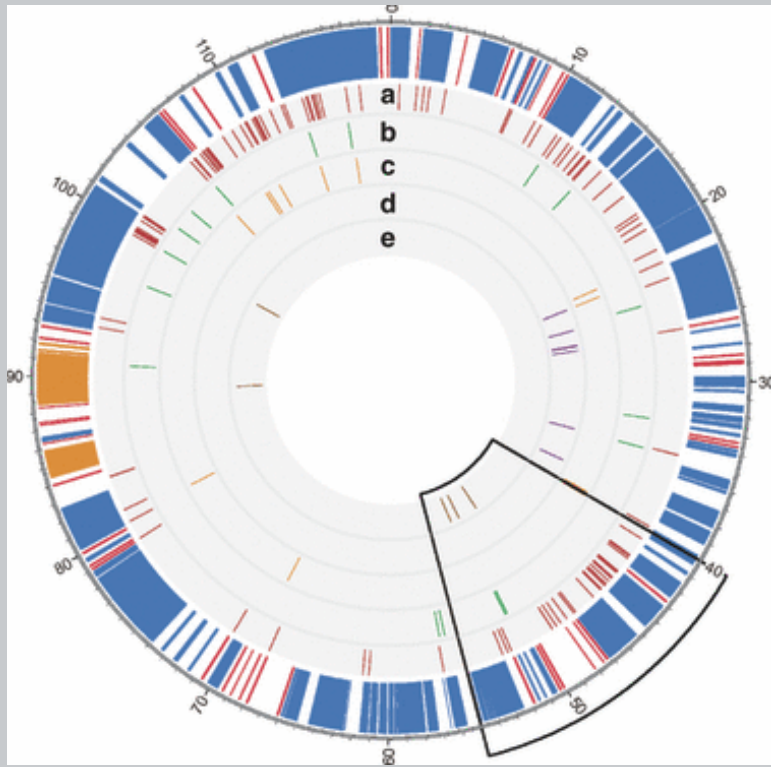
Využití NGS

- sekvenování genomu de-novo
 - cílené obohacení genomu (*targeted enrichment*), tj. sekvenování jen části genomu
- re-sekvenování genomu – *read mapping*
- sekvenování transkriptomu (RNA-Seq)
- amplikonové sekvenování
- (environmentální) metasekvenování
- ...

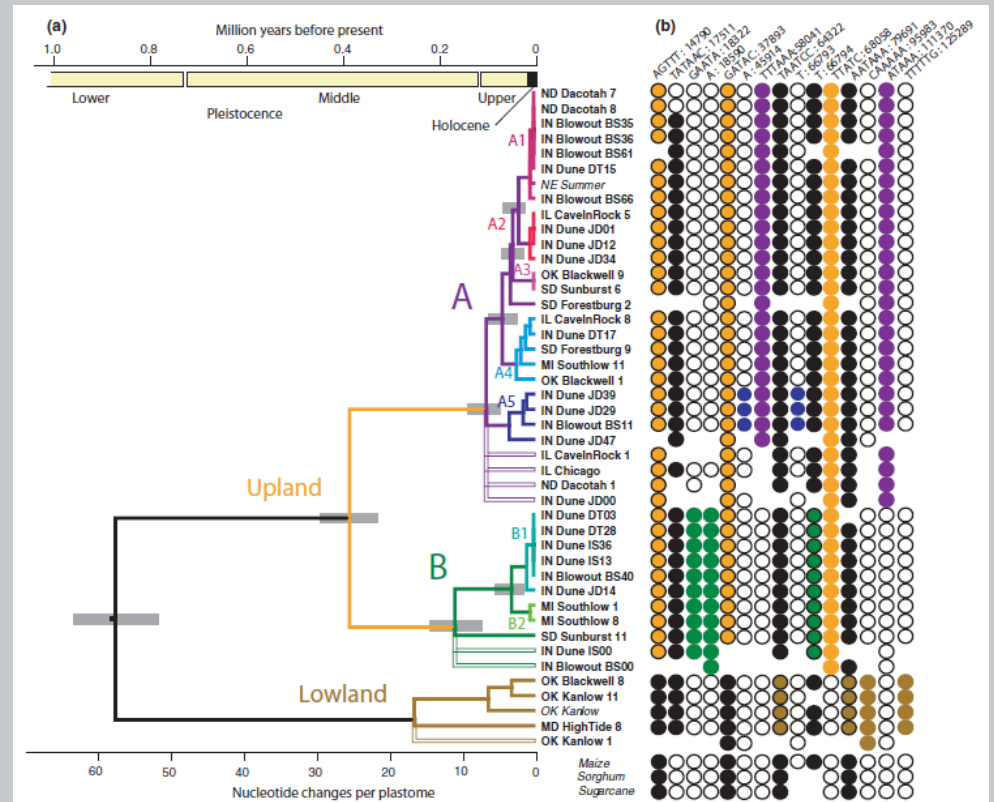
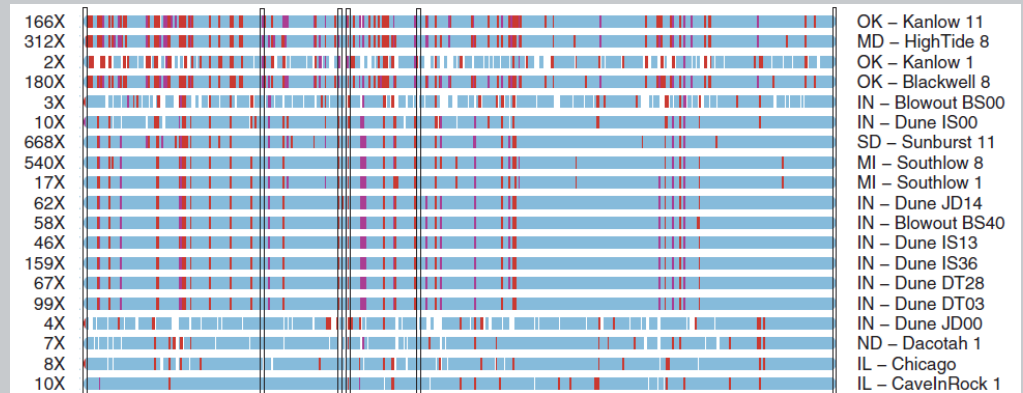
Whole genome sequencing

- sekvenování + assembling
- jednoduché pro malé genomy
 - bakterie
 - cpDNA
- pro velké eukaryotické genomy stále složité a náročné – kombinace dat z více platforem

Sekvenování celých chloroplastů

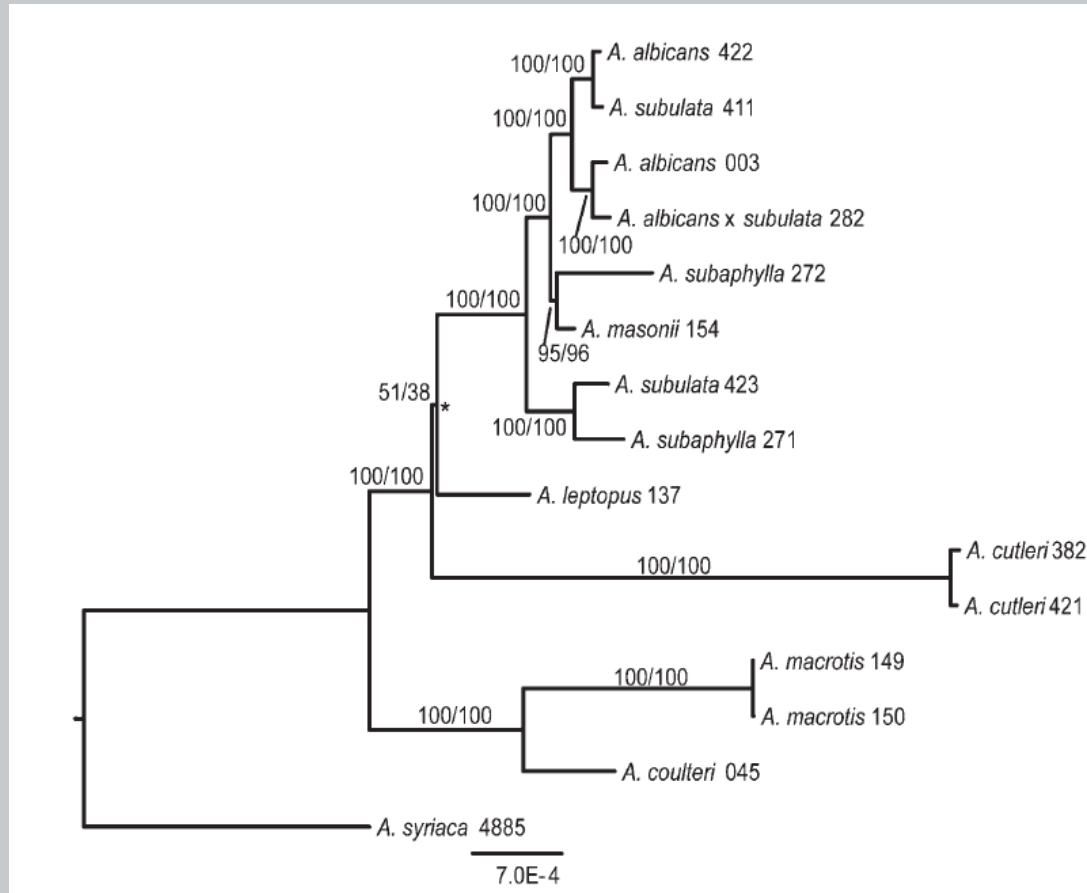


Whittall et al. (2010): Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19:100-114.



Morris et al. (2011): Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology* 20: 4938-4952

Sekvenování celých chloroplastů



Asclepias

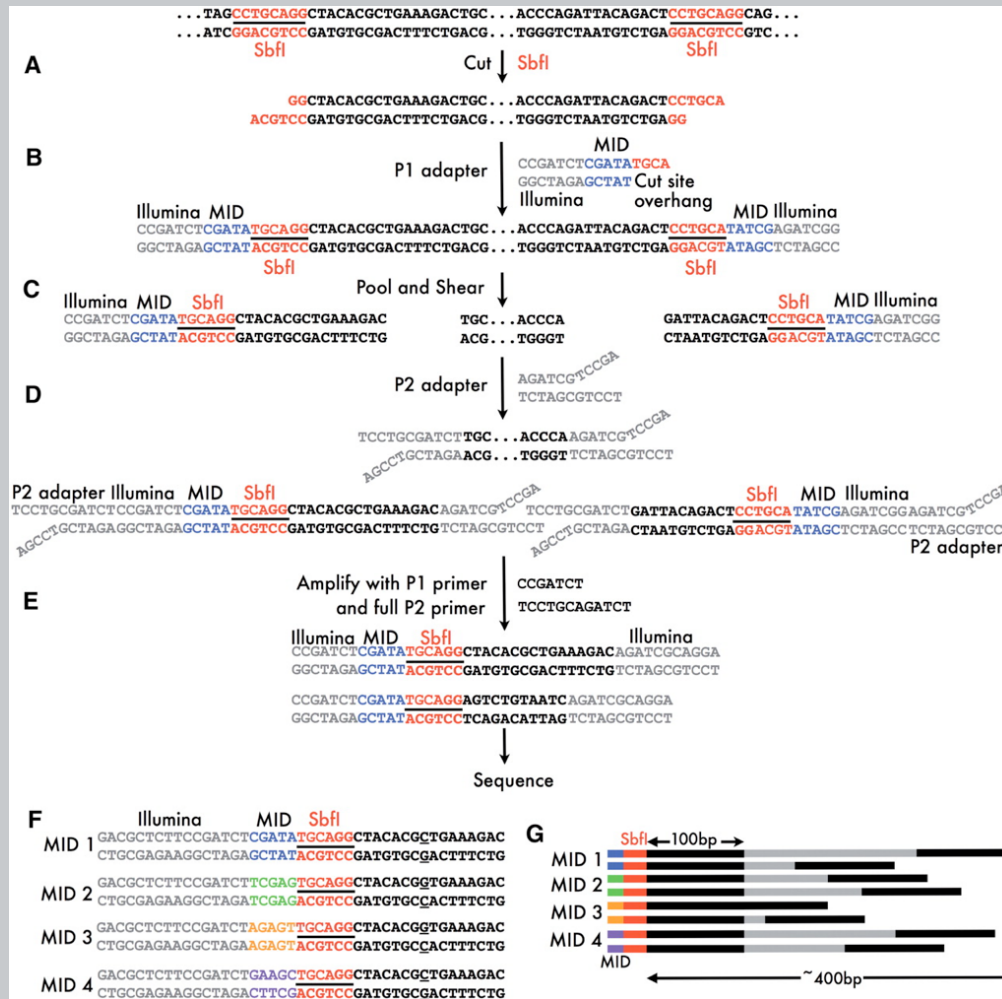
Straub et al. (2012): *Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics*. *American Journal of Botany* 99: 349–364.

Targeted enrichment

- pro snížení complexity
- restriční štěpení genomu
 - sekvenování jen části genomu za štěpnými místy
 - hledání SNP -> binární data
 - RAD-sequencing
 - GBS (genotyping-by-sequencing)
 - ...
- Hyb-Seq
 - hybridization based enrichment
 - obohacení o specifické (předem dané) sekvence

RAD-sequencing

Restriction-site-associated DNA sequencing

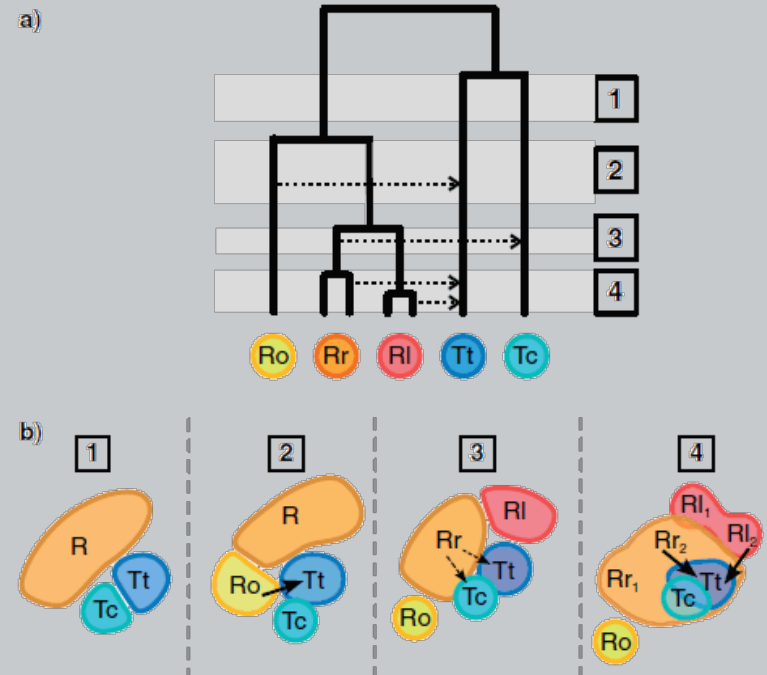
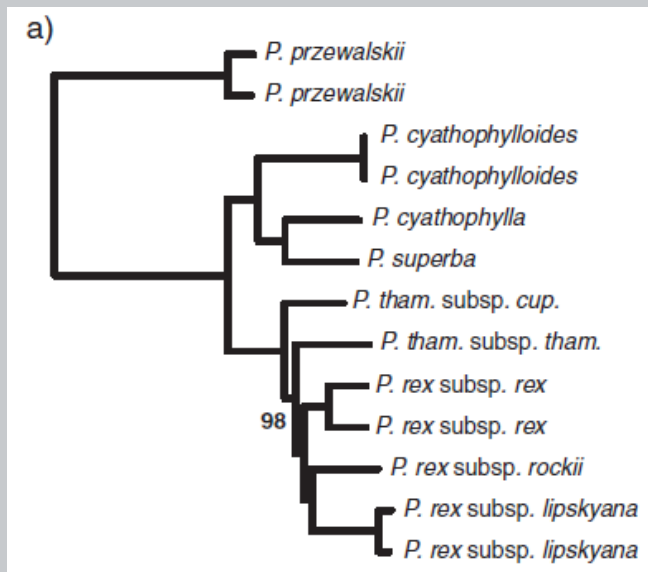
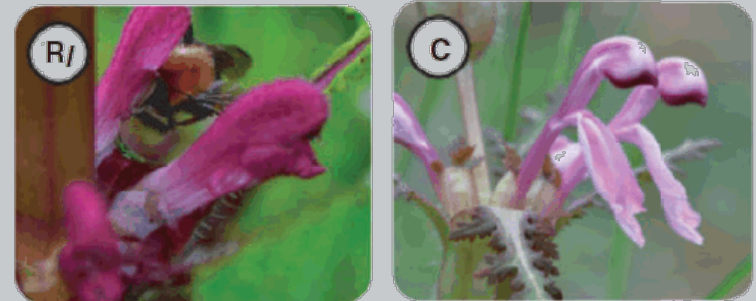


Davey J.W. & Blaxter M.L. (2011): *RADSeq: next-generation population genetics*. Briefings in Functional Genomics 9: 416-423.

Davey J.W. et al. (2011): *Genome-wide genetic marker discovery and genotyping using next-generation sequencing*. Nature Reviews 12: 499-510.

RAD u blízce příbuzných druhů

- recentně divergující skupina – blízce příbuzné druhy
- reduced representation sequencing (RAD Seq)
- fylogeneze a detekce ancestrální hybridizace
- 40 000 lokusů



Hyb-Seq

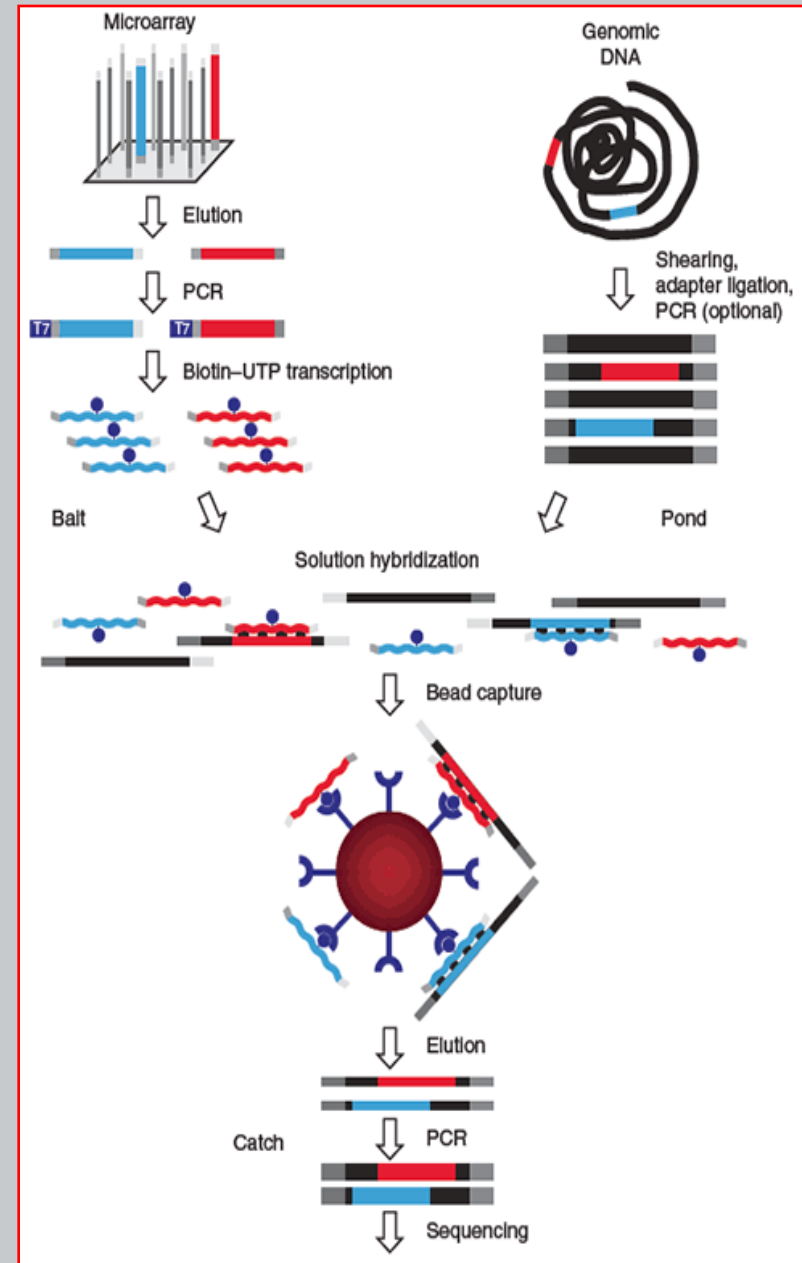
- solution phase hybridization
- ‘baits’ (krátké úseky RNA) syntetizované na array
- hybridizace v roztoku
- immobilizace via biotin-streptavidin
- obohacení o cílové sekvence

Cronn et al. (2012) Amer. J. Bot 99: 291-311

Lemmon et al. (2012) Syst. Biol.

McCormack et al. (2012) Syst. Biol.

Bi et al. (2012) BMC Genomics



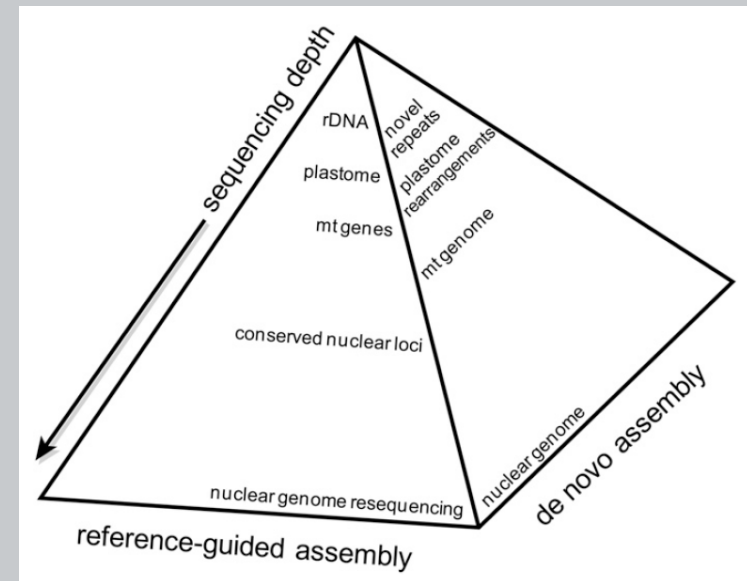
Mycroarray

<http://www.onekp.com>

- sekvenování transkriptomu pro 1300 různých druhů rostlin (z toho cca 750 krytosemenných)
- cílem je shromáždit informace pro robustní fylogenetické studie a pro biotechnologie
- vhodné pro selekci vhodných genů pro fylogenezi, např. pro design baits pro enrichment

Genome-skimming

- sekvenování genomické DNA s velmi nízkým celkovým pokrytím
- získání dostatečného pokrytí k assemblingu
 - celého plastomu
 - velké části mtDNA
 - rDNA cistronu
 - řady kandidátních single-copy genů



Straub et al. (2012): *Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics*. American Journal of Botany 99: 349–364.

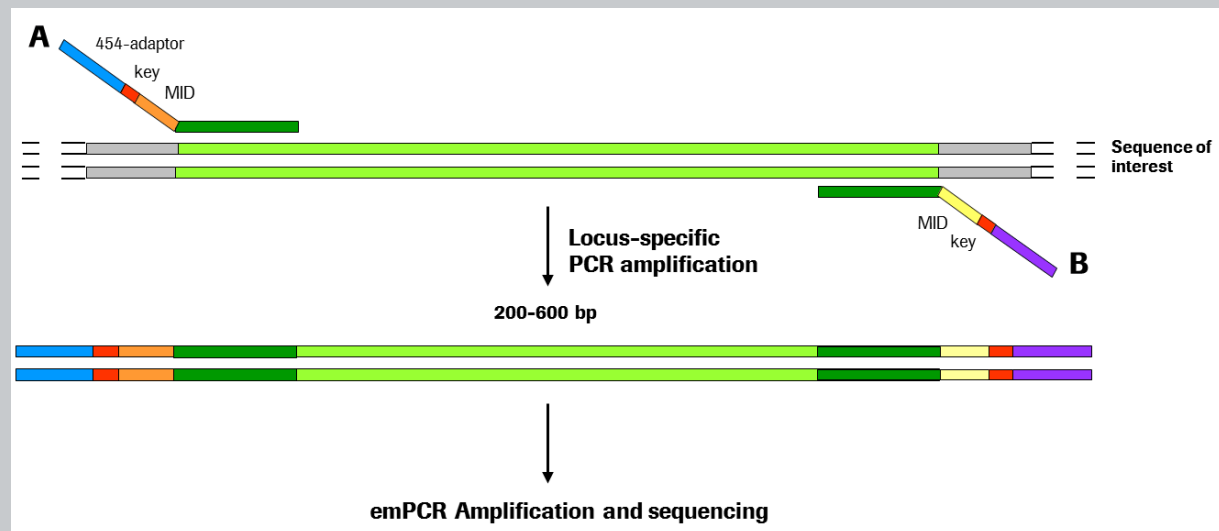
Steel et al. (2012): *Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae*. American Journal of Botany 99: 330-348.

Sekvenování transkriptomu

- sekvenování cDNA (získané reverzní transkripcí mRNA)
- transkriptom mnohem menší než genom
- vhodné pro nemodelové organismy
- využití
 - hledání vhodných genů pro fylogenetické studie (variabilní úseky při porovnání informace z více jedinců)
 - identifikace mikrosatelitů
 - ...

Amplikonové sekvenování

- PCR konkrétního genu (intergenické oblasti)
- označení jednotlivých vzorků specifickou sekvencí (MID)
- paralelní sekvenování všech PCR reakcí
- oddělení sekvencí v počítači na základě MID identifikace



Metasekvenování

- PCR amplifikace konkrétního genu z environmentálního vzorku (voda, půda atd.)
- sekvenování všech produktů
- srovnání výsledných sekvencí s databází
- identifikace druhů a jejich frekvence

- použití – zjištění složení společenstva
 - bakteriální nebo houbové společenstvo
 - historické – např. z DNA z permafrostu
 - potravní preference živočichů

Historické složení arktické vegetace

	%
22 960 ± 120 years BP	
<i>Bistorta vivipara</i>	47.25
<i>Equisetum arvense</i> / <i>E. fluviatile</i> / <i>E. sylvaticum</i>	24.31
<i>Salix</i> sp./ <i>Chosenia arbutifolia</i> / <i>Populus balsamifera</i>	4.74
<i>Armeria scabra</i>	3.03
<i>Thymus oxyodontus</i>	2.77
<i>Lagotis glauca</i>	2.17
Asteraceae 1*	1.87
<i>Avenella flexuosa</i>	1.77
<i>Aconogonon alaskanum</i> / <i>A. ocreatum</i> / <i>A. tripterospermum</i>	1.36
<i>Rumex</i> sp.	1.31
<i>Packera</i> sp./ <i>Senecio</i> sp.	0.96
Poaceae 1†	0.96
<i>Ranunculus acris</i> / <i>R. subborealis</i> / <i>R. turneri</i>	0.81
<i>Festuca</i> sp.	0.76
<i>Hulteniella integrifolia</i>	0.66
<i>Saxifraga hirculus</i>	0.55
<i>Trientalis europaea</i>	0.45
Asteraceae 2‡	0.40
<i>Valeriana capitata</i> / <i>V. officinalis</i> agg.	0.35

Sønstebo et al. (2010): *Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate*. *Molecular Ecology Resources* 10: 1009-1018.

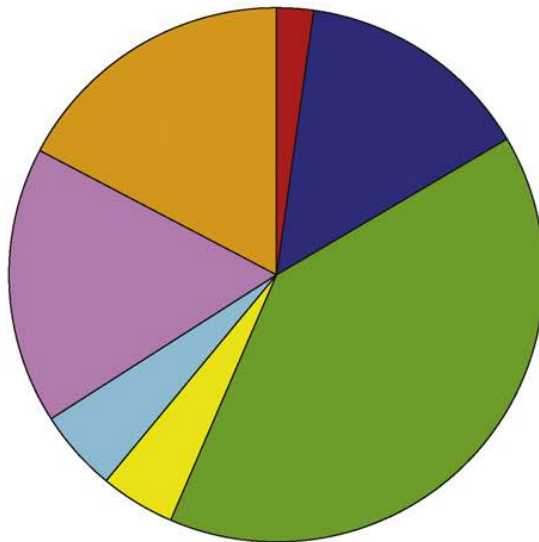
Potravní preference živočichů



Golden marmot

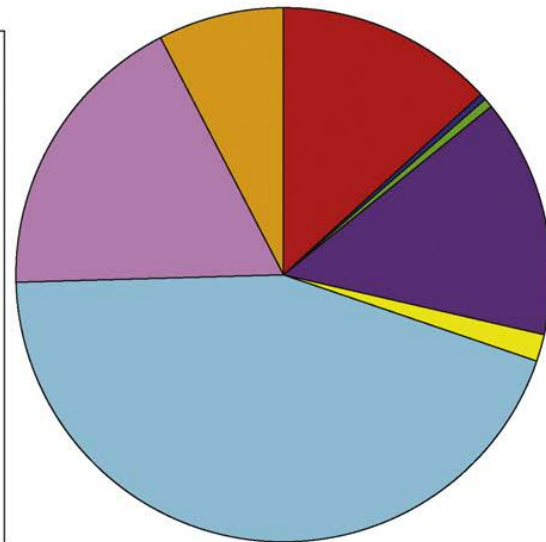


Brown bear



Key:

- Apiaceae
- Asteraceae
- Caryophyllaceae
- Cyperaceae
- Fabaceae
- Poaceae
- Polygonaceae
- Others



TRENDS in Ecology & Evolution

Literatura

- Metzker M.L. (2010) Sequencing technologies – the next generation. *Nature Reviews Genetics*, 11, 31–46.
- Bräutigam A. & Gowik U. (2010): What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology*, 12, 831–841.
- Ansorge W.J. (2009): Next-generation DNA sequencing techniques. *New Biotechnology*, 25, 195–203.
- Glenn T.C. (2011): Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759–769.
- McCormack J.E. et al. (2011): *Applications of next-generation sequencing to phylogeography and phylogenetics*. Mol. Phylogenet.Evol.
- Straub et al. (2012): *Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics*. American Journal of Botany 99: 349–364.