

Next-generation sequencing  
(NGS)

**High-throughput sequencing  
(HTS)**

# Sanger sequencing

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

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

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

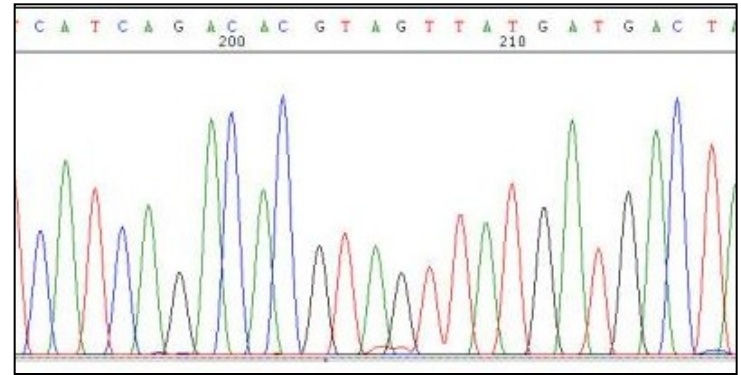
Primer - F - AAGTCAGT**C**=0

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

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

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

Primer - F - AAGT**C**=0



krátké ----- dlouhé  
(rychlé) ----- (pomalé)

+

Primer - F **AAGTCAGTCTAA**ATGCGATTGGGA Rev. Primer - R

Rev. Primer - F **TTCAGTCAGATTACGCTAACCT** Primer - R

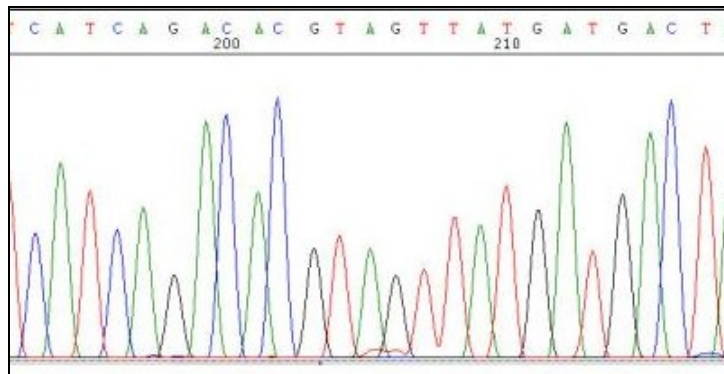
# 4-kapilární sekvenátor

=

96 x 500 bp/12 hodin

=

**cca 100 000 bp/den**



detector



laser beam



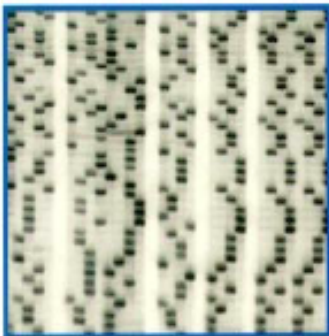
capillary  
electrophoresis

+

-

# Evolve Sangerova sekvenování

Pre-1992  
“old fashioned  
way”



S35 ddNTPs  
Gels  
Manual loading  
Manual base calling

1992-1999  
ABI 373/377



Fluorescent ddNTPs\*  
Gels  
Manual loading  
Automated base calling\*

1999  
ABI 3700



Fluorescent ddNTPs  
Capillaries\*  
Robotic loading\*  
Automated base calling  
Breaks down frequently

2003  
ABI 3730XL



Fluorescent ddNTPs  
Capillaries  
Robotic loading  
Automated base calling  
Reliable\*



96-kapilární sekvenátor

=

2304 x 500 bp/12 hodin

=

**cca 2 400 000 bp/den**

HTS (Illumina NovaSeq 6000)

=

**cca 6 000 000 000 000 bp/den**

electrophoresis

# Next-generation sequencing (NGS)

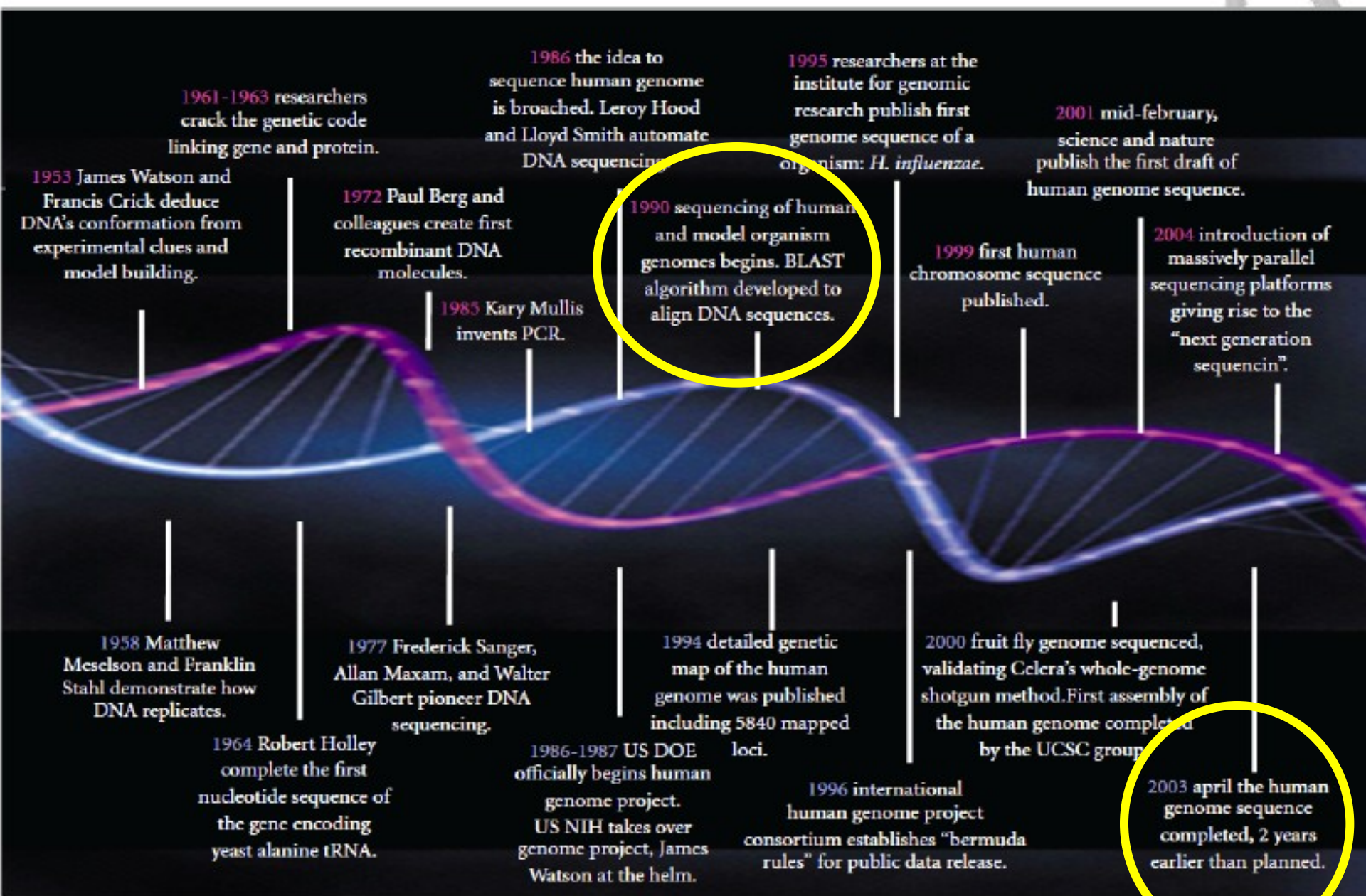
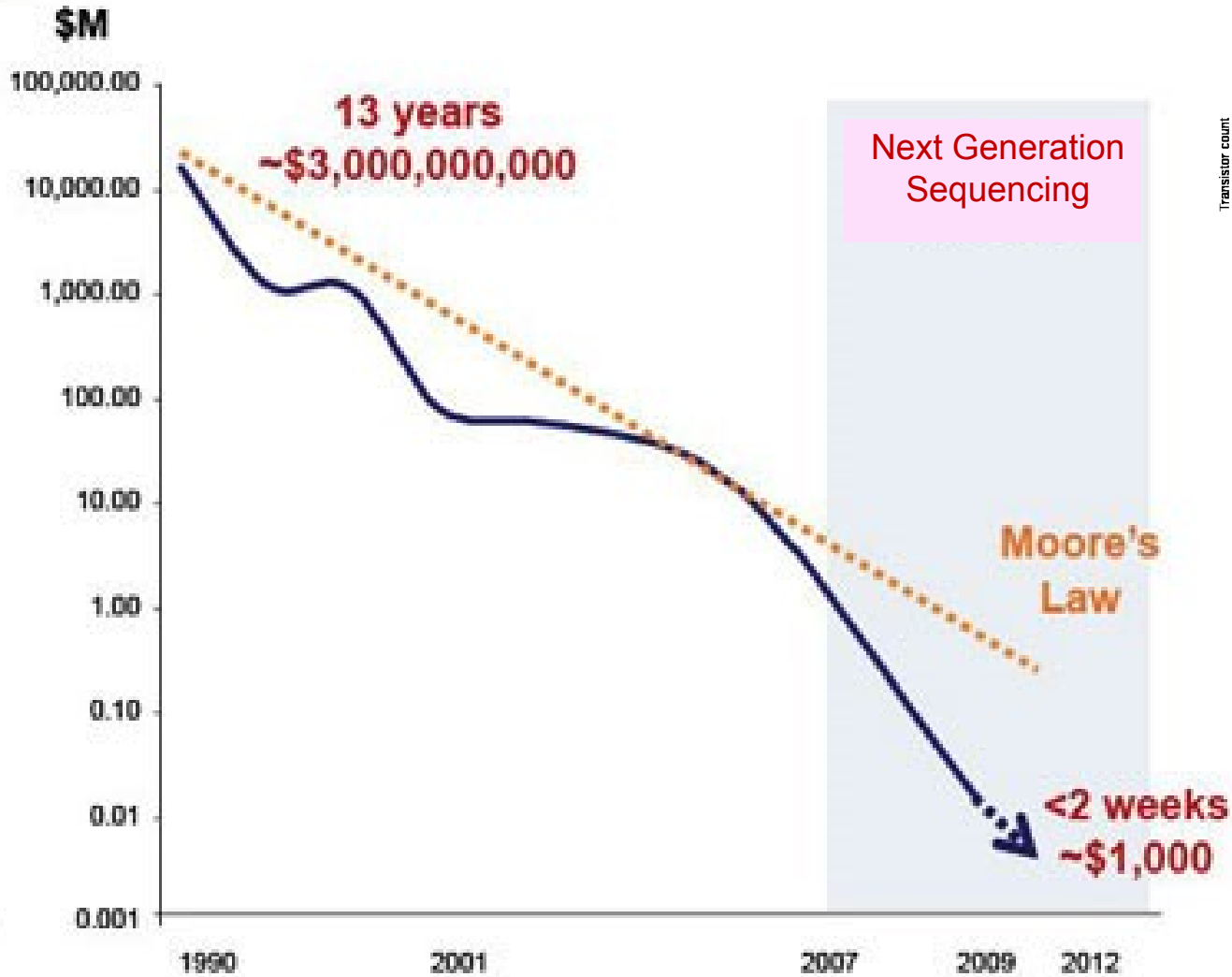
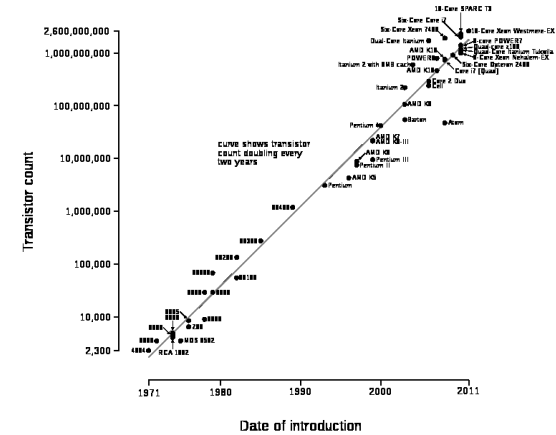


FIGURE 1: Evolution of DNA revolution.

# Cost per Human Genome



Microprocessor Transistor Counts 1971-2011 & Moore's Law



# Illumina HiSeqX10



**\$1 M** per machine

**1.8 Tbase** per machine per 3 days

**1800** human genomes per machine per year

# Welcome to immense discovery power

NovaSeq 6000 Sequencing System is by far our most powerful instrument, designed to adapt to your needs so groundbreaking discoveries are always within reach. It's flexible, scalable science on your terms.

Contact an Illumina Representative



## Illumina NovaSeq 6000



### System specifications

Output range 80 - 6000 Gb

Paired end reads per run 1.6 - 40B

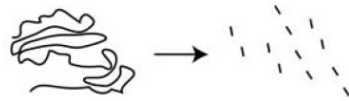
Max read length 2 × 250 bp

Run time 13 - 44 hours

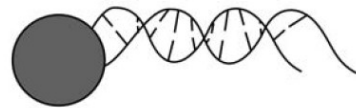
[View All NovaSeq 6000 Specifications](#)

# Historie „Next generation sequencing“

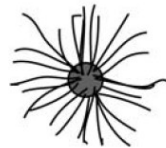
1) Randomly fragment many molecules of target DNA



2) Immobilize individual DNA molecules on solid support



3) Amplify DNA in clonal 'polymerase colony'



4) Sequence DNA by adding liquid reagents to immobilized DNA colonies



5) Interrogate sequence incorporation *in situ* after each cycle using fluorescence scanning or chemiluminescence



454 pyrosequencing ... první komerčně dostupná NGS technologie od srpna 2007

2016 – ohlášené stažení z trhu (Roche)



# Široké spektrum technologií





# Ale jen některé přežijí



# Dnes dostupné NGS platformy

- Roche 454
- **Illumina (MiSeq, NextSeq, HiSeq, NovaSeq)**
- ABI SOLiD
- IonTorrent (Life Technologies)
- **SMRT (Pacific Biosciences)**
- **Oxford Nanopore**
- ...

# Illumina HiSeq/MiSeq

- v současné době nejrozšířenější typ (cca 70%) na trhu
- v horizontu následujících let její používání spíše poroste
- NextSeq, NovaSeq, etc.

<https://www.illumina.com/techniques/sequencing/high-throughput-sequencing.html>

Illumina HiSeq

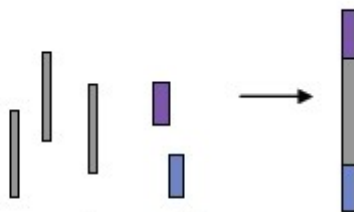


Illumina MiSeq



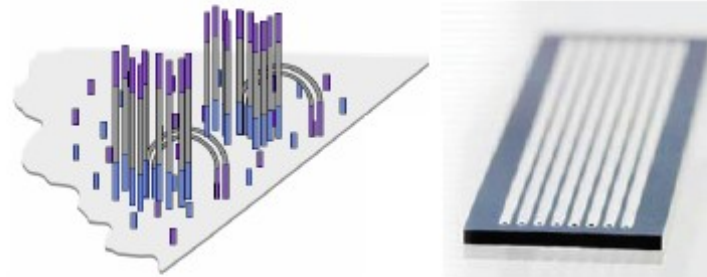
# Illumina Sequencing pipeline

## 1. Sample Prep (1-5 days)



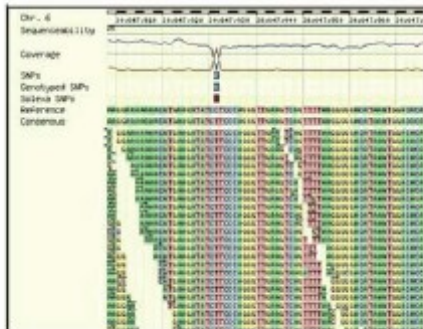
Ligate adapters

## 2. Cluster generation on flow cell (1.5 day)



Clonal Single molecular Array

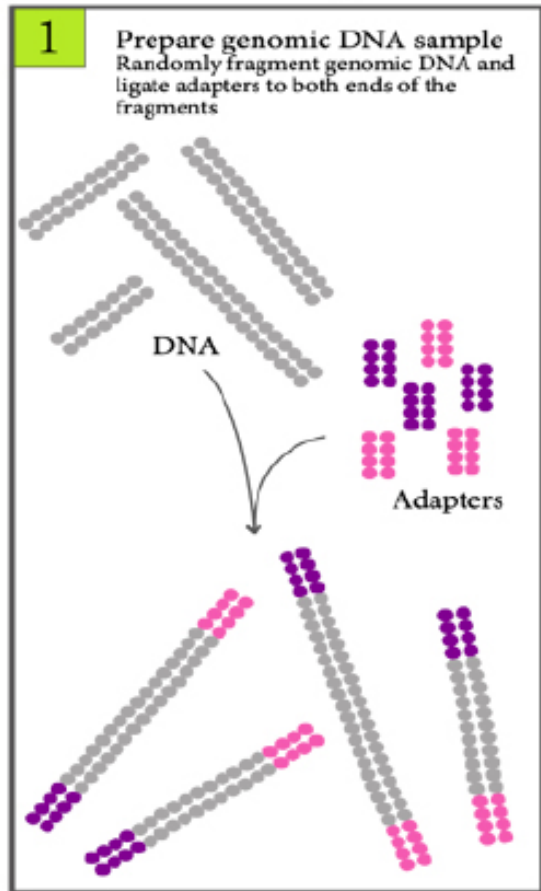
## 4. Data Analysis (days-months)



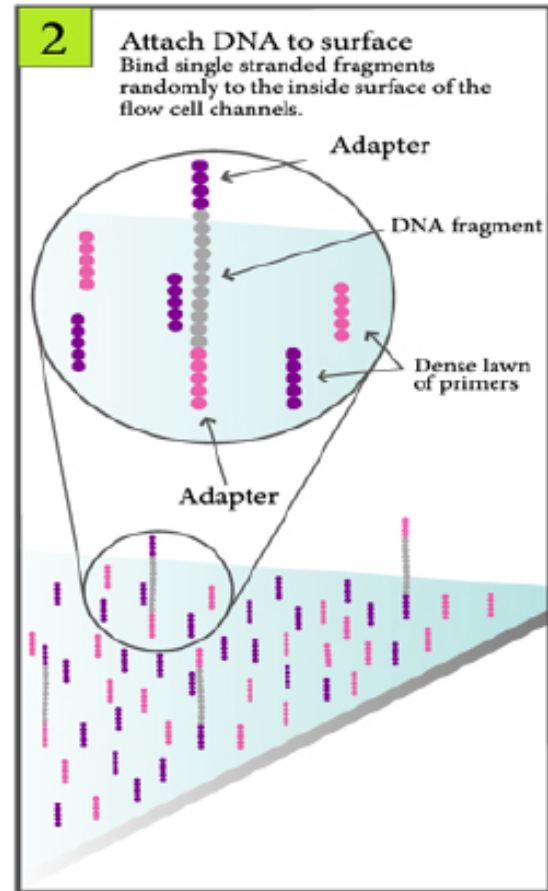
## 3. Sequencing and imaging (2-3 days)



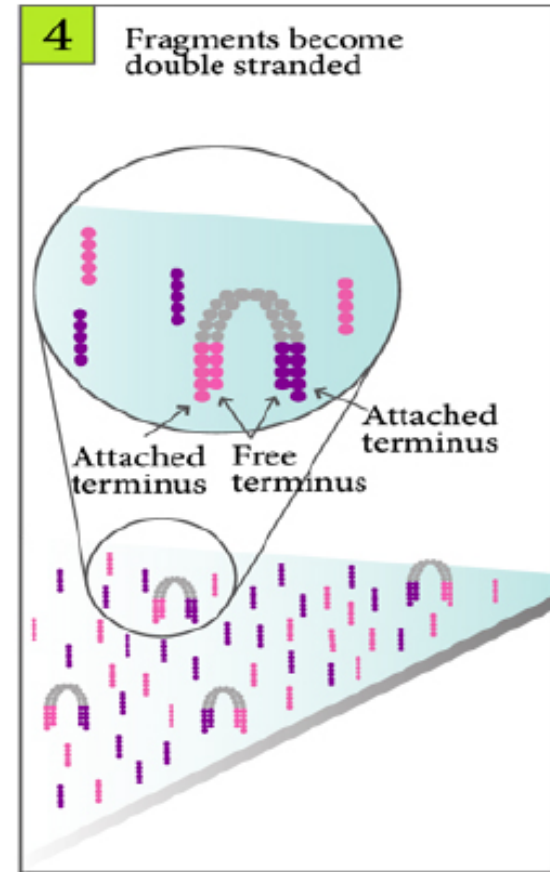
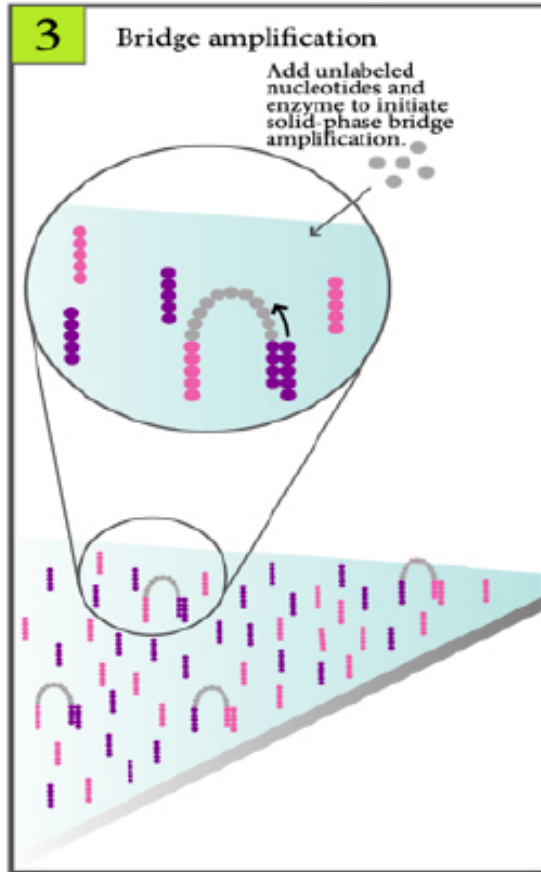
# Attach DNA to flow cell



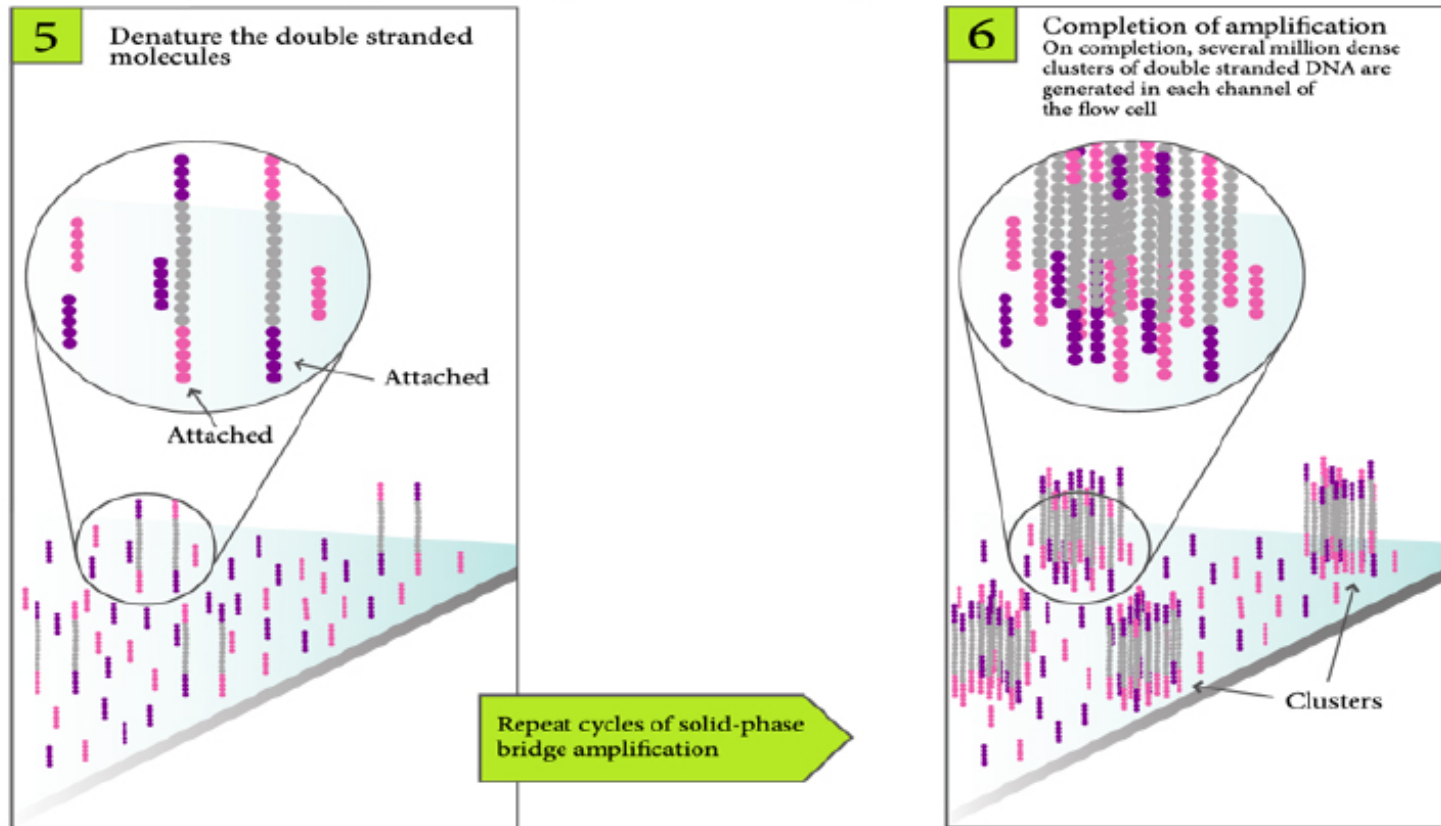
Add sample to flow cell



# Bridge Amplification



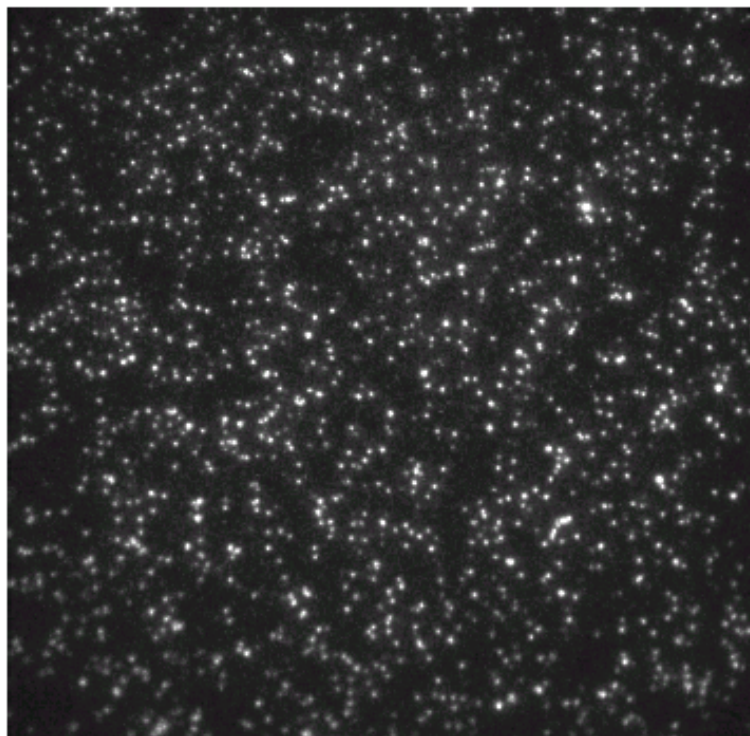
# Cluster Generation



Clonal Single molecular Array

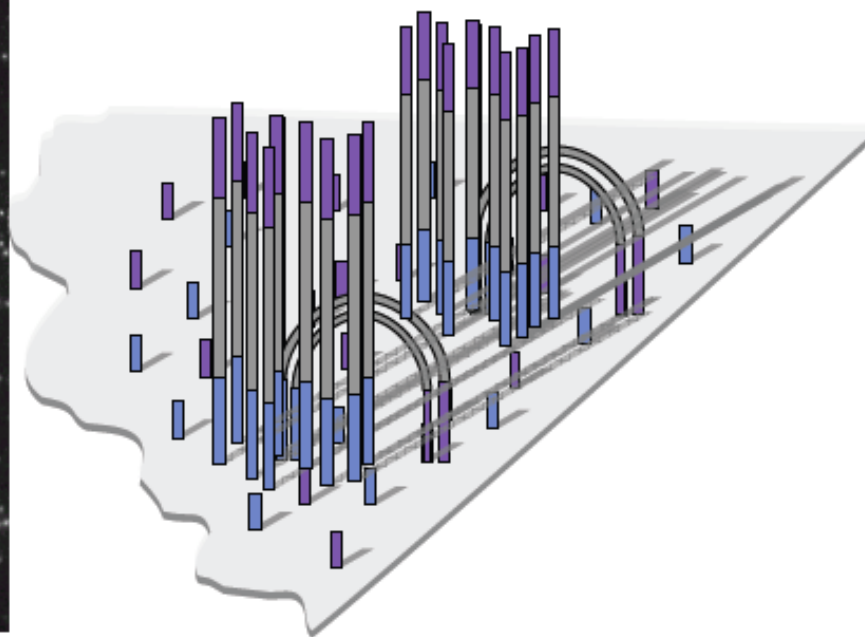


# Clonal Single molecule Array



100um

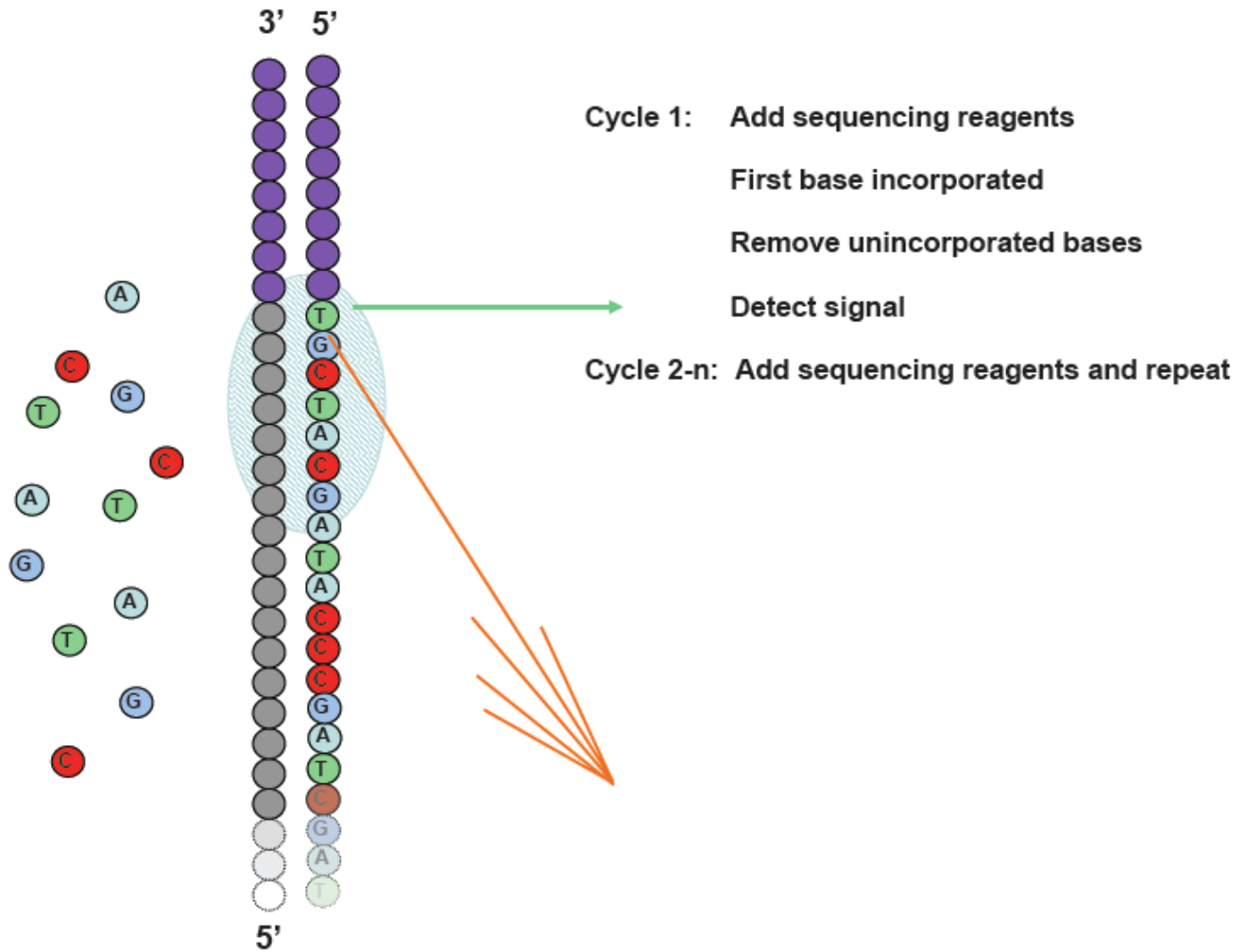
Random array of clusters



~1000 molecules per ~ 1 um cluster  
~20-30,000 clusters per tile  
~40 M clusters per flowcell



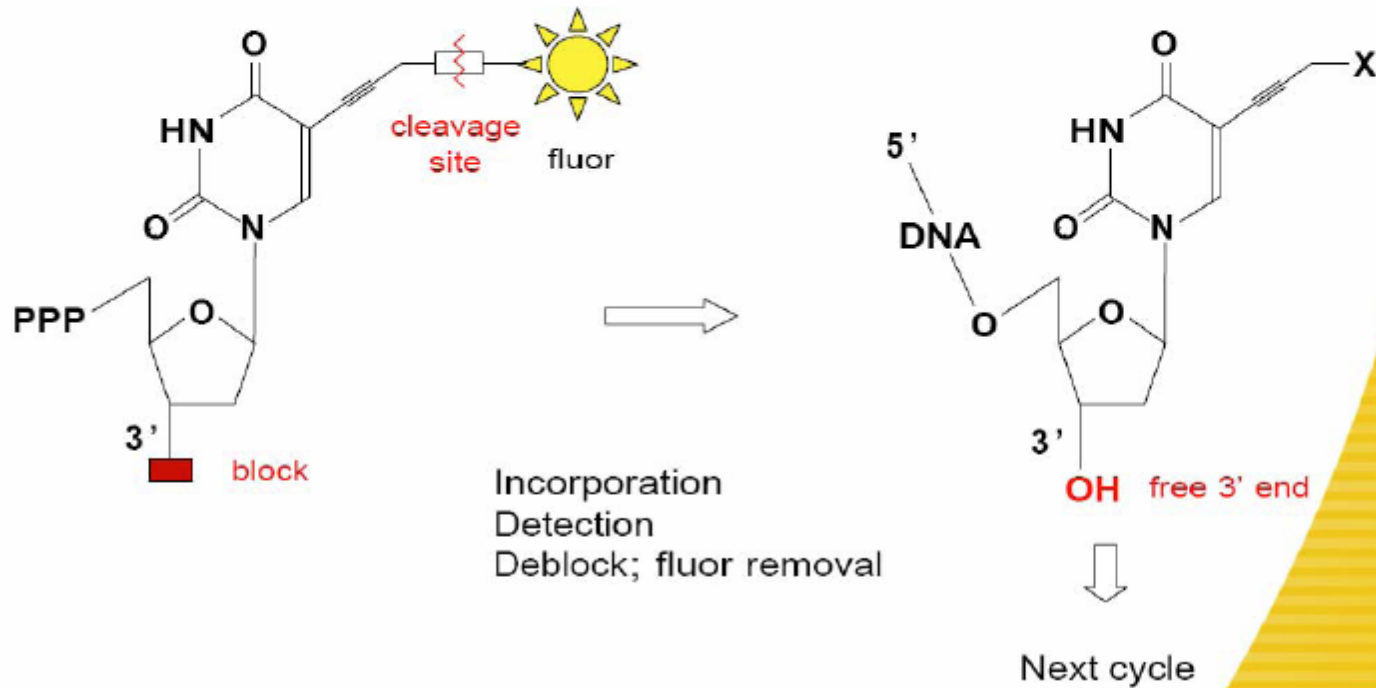
# Sequencing By Synthesis (SBS)



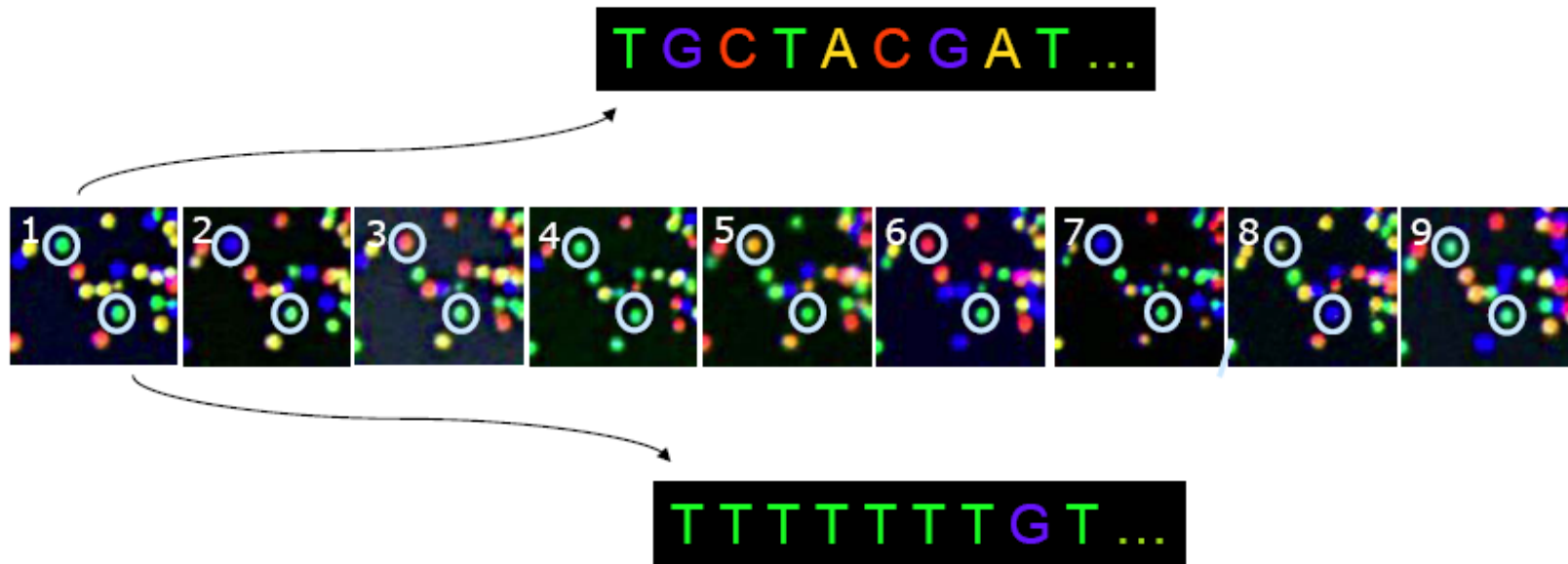
# Reversible Terminator Chemistry



- All 4 labelled nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats



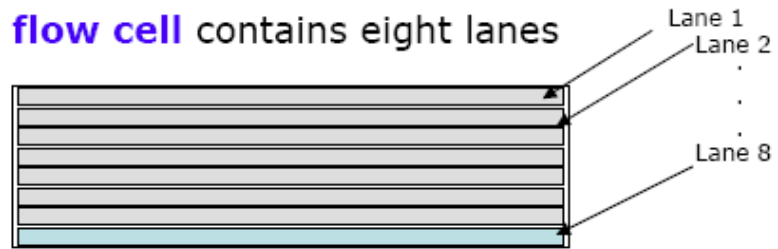
# Base Calling From Images



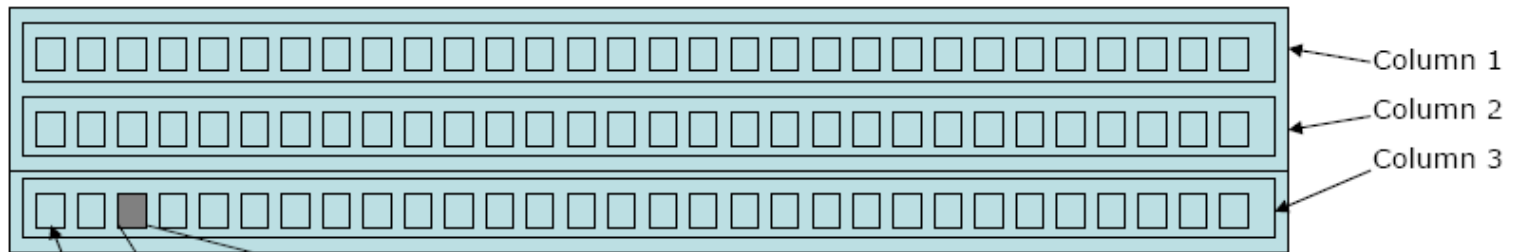
The identity of each base of a cluster is read off from sequential images



A **flow cell** contains eight lanes



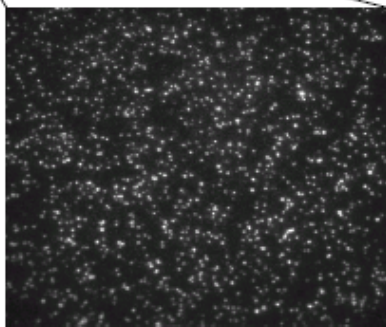
Each **lane/channel** contains **three columns** of tiles



Each **column** contains **100 tiles**

Tile

20K-30K  
Clusters

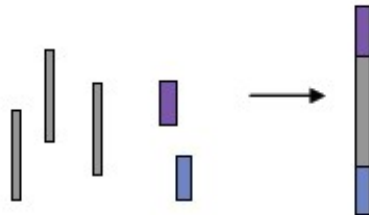


350 X 350  $\mu\text{m}$

[https://www.youtube.com/watch?annotation\\_id=annotation\\_228575861&feature=iv&src\\_vid=womKfikWlxM&v=fCd6B5HRaZ8](https://www.youtube.com/watch?annotation_id=annotation_228575861&feature=iv&src_vid=womKfikWlxM&v=fCd6B5HRaZ8)

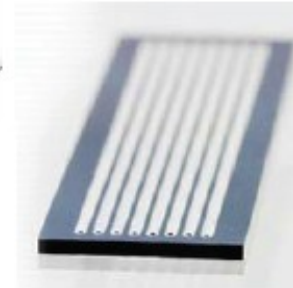
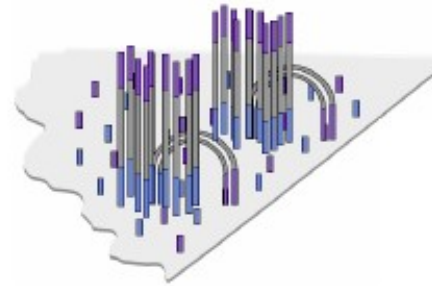
# Illumina Sequencing pipeline

## 1. Sample Prep (1-5 days)



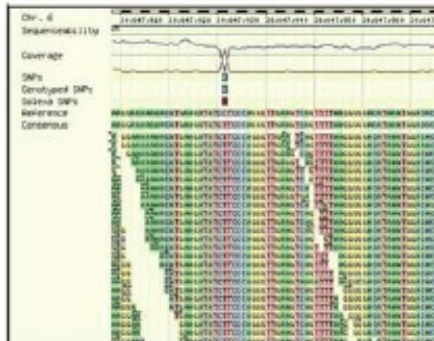
Ligate adapters

## 2. Cluster generation on flow cell (1.5 day)



Clonal Single molecular Array

## 4. Data Analysis (days-months)



## 3. Sequencing and imaging (2-3 days)





# Illumina fastq

= one „read“

```
1      2      3      4      5      6 7      8
@HWI-ST226:253:D14WFACXX:2:1101:2743:29814 1:N:0:ATCACG
TGC GGAAGGATCATTGTGGAATTCTCGGGTGCCAAGGA ACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTT
GAAAAAAAAAAAAAAAAAATTA
+
B@CFFFFFFHFFHJIIGHIHIJJJIJIIJJGDCHIIJJJJJJJGJGIHHEH@)=F@EIGHHEHFFFDCBBD:@CC@C
:<CDDDD50559<B#####
```

1. unique instrument ID and run ID
2. Flow cell ID and lane
3. tile number within the flow cell lane
4. 'x'-coordinate of the cluster within the tile
5. 'y'-coordinate of the cluster within the tile
6. the member of a pair, /1 or /2 (*paired-end or mate-pair reads only*)
7. N if the read passes filter, Y if read fails filter otherwise
8. Index sequence



# All this generates a lot of Data!

## 6 TB data/run

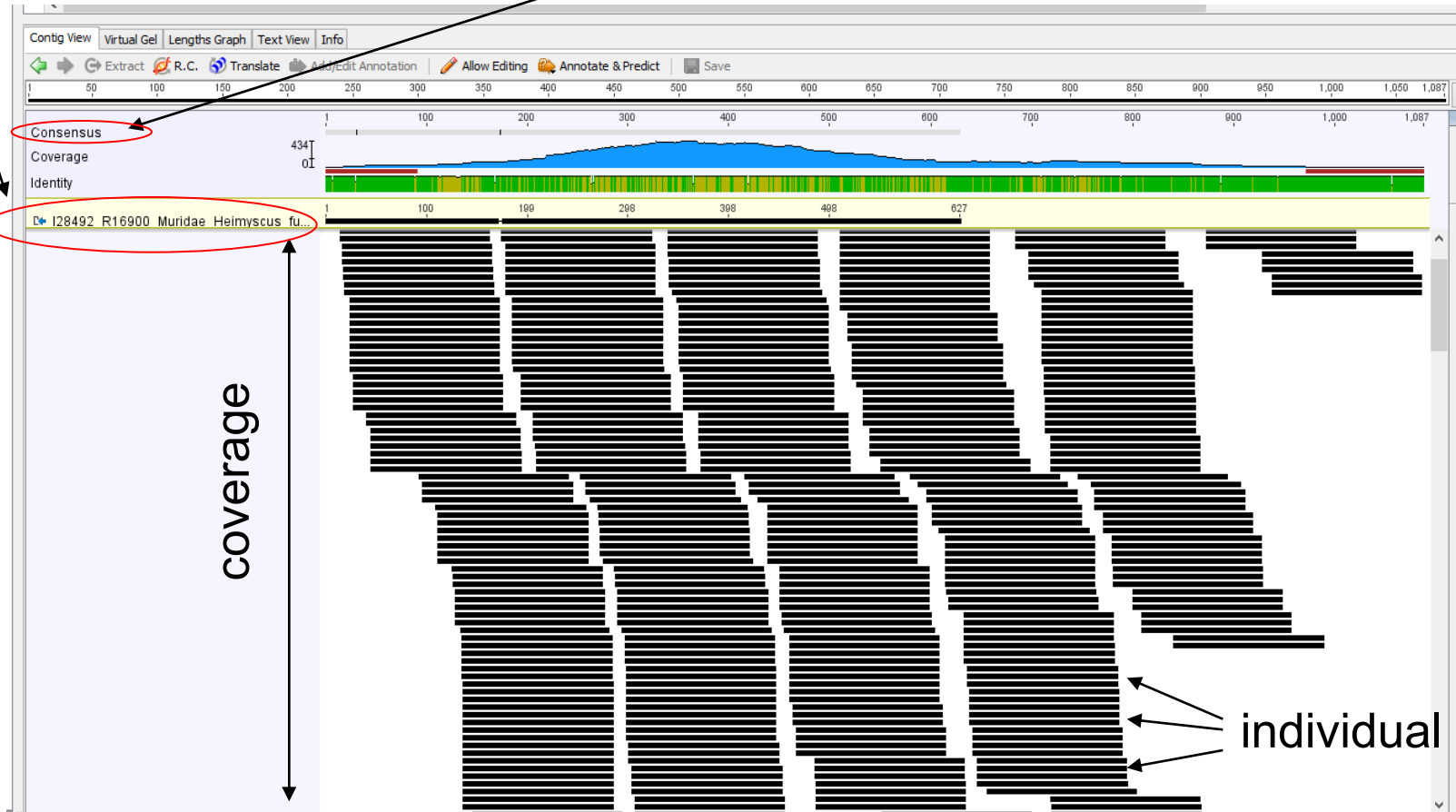
- 1 Gig of Space
  - 125,000 pages of text
  - 11 CDs of Music
  - 4000 (1024x768) JPEG images
  - 40,000 pages of PDF
- 1 TB of space
  - 220 Million pages of text
  - 300 hours of video
  - 4,000,000 JPEG images
  - 1,000 copies of the Encyclopedia Britannica
  - 1/10 of the printed Library of Congress



# Data analysis in Geneious

consensus

reference (in resequencing)

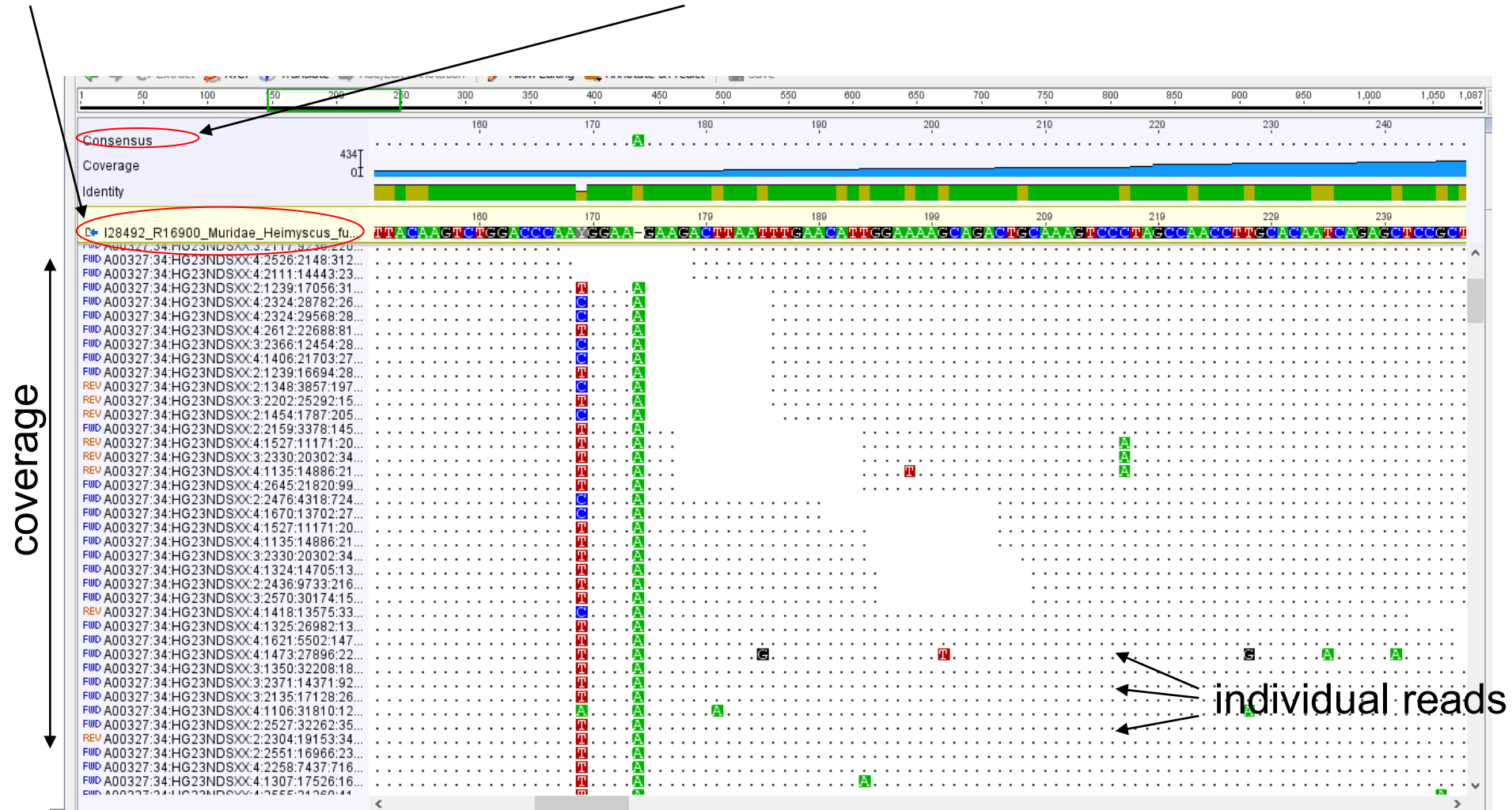


individual reads

# Data analysis in Geneious

reference (in resequencing)

consensus



## Illumina sequencers

### Illumina MiSeq

4 millions reads/run  
150 bp/read



### Illumina GAIIx

300 millions reads/run  
150 bp/read



### Illumina HighSeq

1500 – 3000 millions reads/run  
100 bp/read



## NovaSeq 6000 Sequencing System (2017)

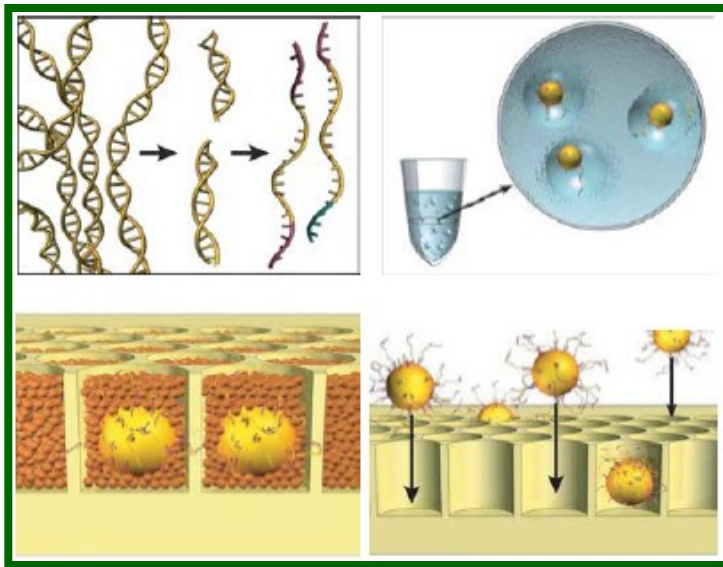
ca. 48 human genomes/run

### Sequencing Output per Flow Cell

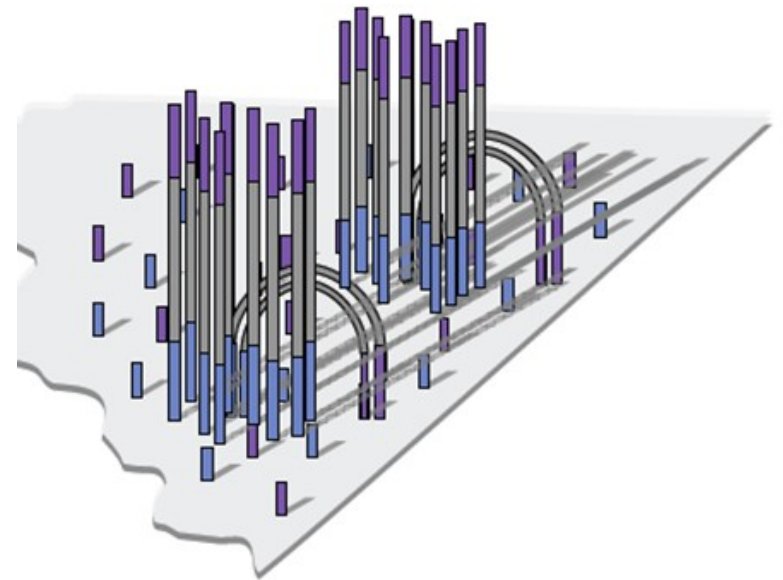
Flow Cell Type	NovaSeq 6000 System		
	S1	S2	S4
2 × 50 bp	134–167 Gb	333–417 Gb	N/A*
2 × 100 bp	266–333 Gb	667–833 Gb	N/A*
2 × 150 bp	400–500 Gb	1000–1250 Gb	2400–3000 Gb

Specifications based on Illumina PhiX control library at supported cluster densities.  
\* N/A: not applicable

# Další NGS technologie



454 pyrosequencing  
(Roche)



Illumina

# Ion Torrent technology



Microbial  
sequencing



Targeted  
sequencing



Transcriptome  
sequencing



Exome  
sequencing



Ion PGM™ Sequencer

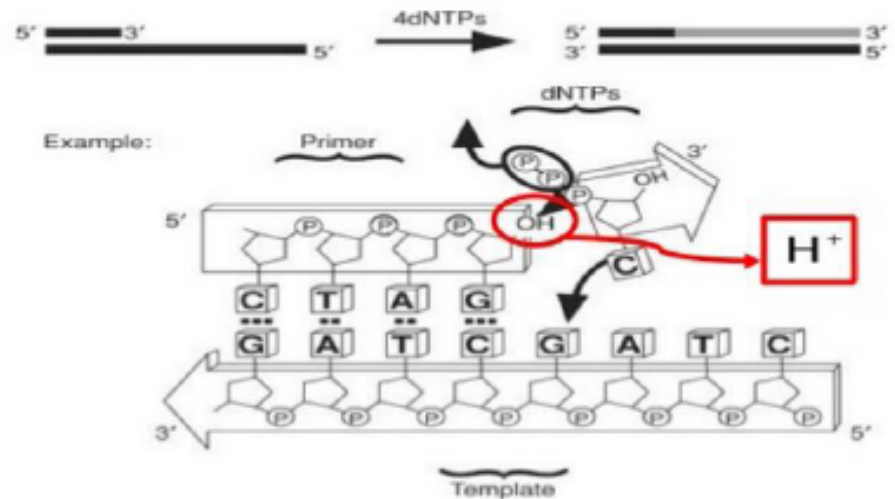


Ion Proton™ Sequencer

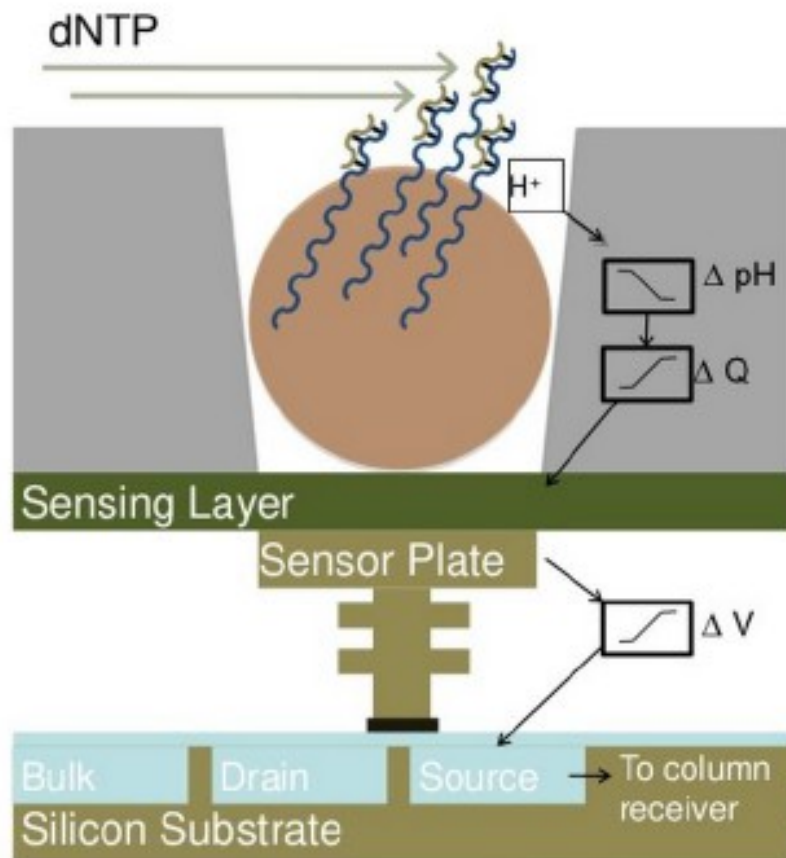
## Ion sequencing: Life Technologies

# Využívá změny pH při syntéze DNA

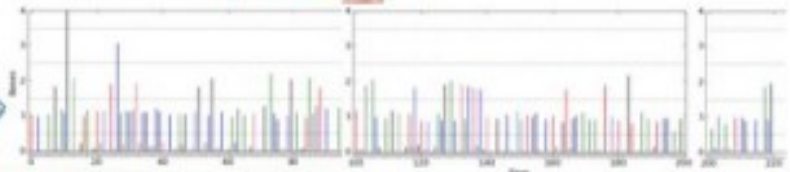
- Ion Semiconductor Sequencing
- Detection of hydrogen ions during the polymerization DNA
- Sequencing occurs in microwells with ion sensors
- No modified nucleotides
- No optics



# Ion Torrent



- DNA → Ions → Sequence
  - Nucleotides flow sequentially over Ion semiconductor chip
  - One sensor per well per sequencing reaction
  - Direct detection of natural DNA extension
  - Millions of sequencing reactions per chip
  - Fast cycle time, real time detection





# Ion Torrent: System Updates

## 314 Chip

- 100bp reads ~10 Mb/run (1.5 hrs)

## 316 Chip

- 100 bp reads ~100 Mbp / run (2 hrs)
- 200 bp reads ~200 Mbp/run (3 hrs)

## 318 Chip

- 200 bp reads ~1 Gbp / run (4.5 hrs)

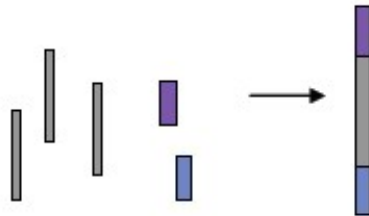
400 bp reads





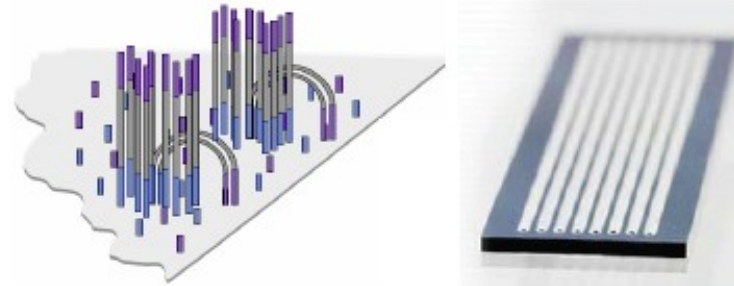
# Illumina Sequencing pipeline

## 1. Sample Prep (1-5 days)



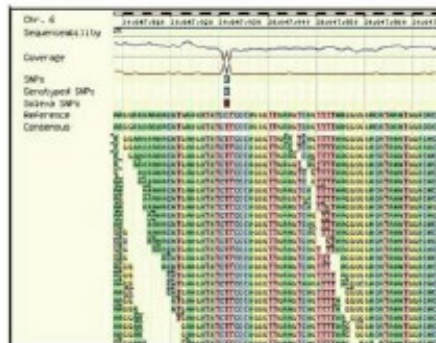
Ligate adapters

## 2. Cluster generation on flow cell (1.5 day)



Clonal Single molecular Array

## 4. Data Analysis (days-months)



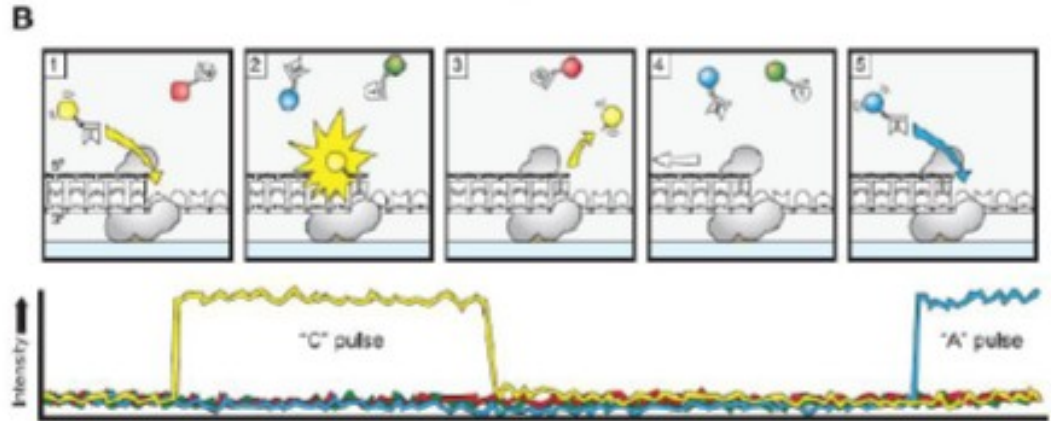
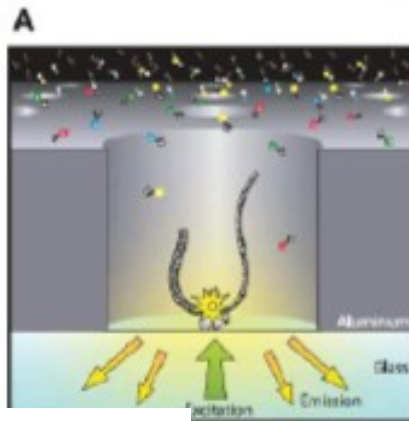
## 3. Sequencing and imaging (2-3 days)



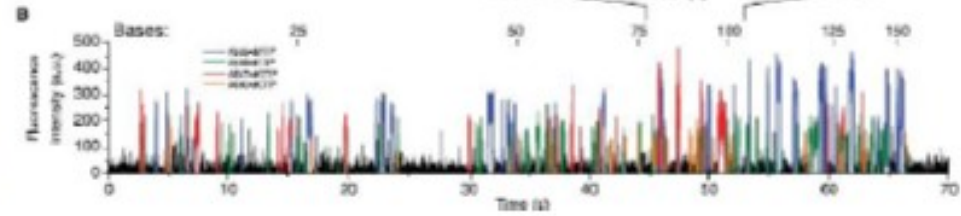
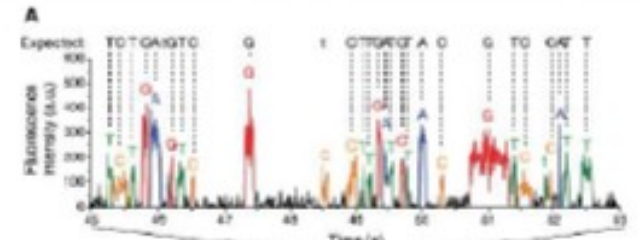
# 3rd generation of sequencing (TGS)

- Long reads with no amplification
- Direct detection of epigenetic modifications on native DNA
- Direct sequencing through regions of the genome inaccessible or difficult to analyze by short-read platforms
- Uniform coverage of the genome as they are not as sensitive to GC content as short-read platforms

# SMRT („single molecule real-time sequencing”) – Pacific Biosciences

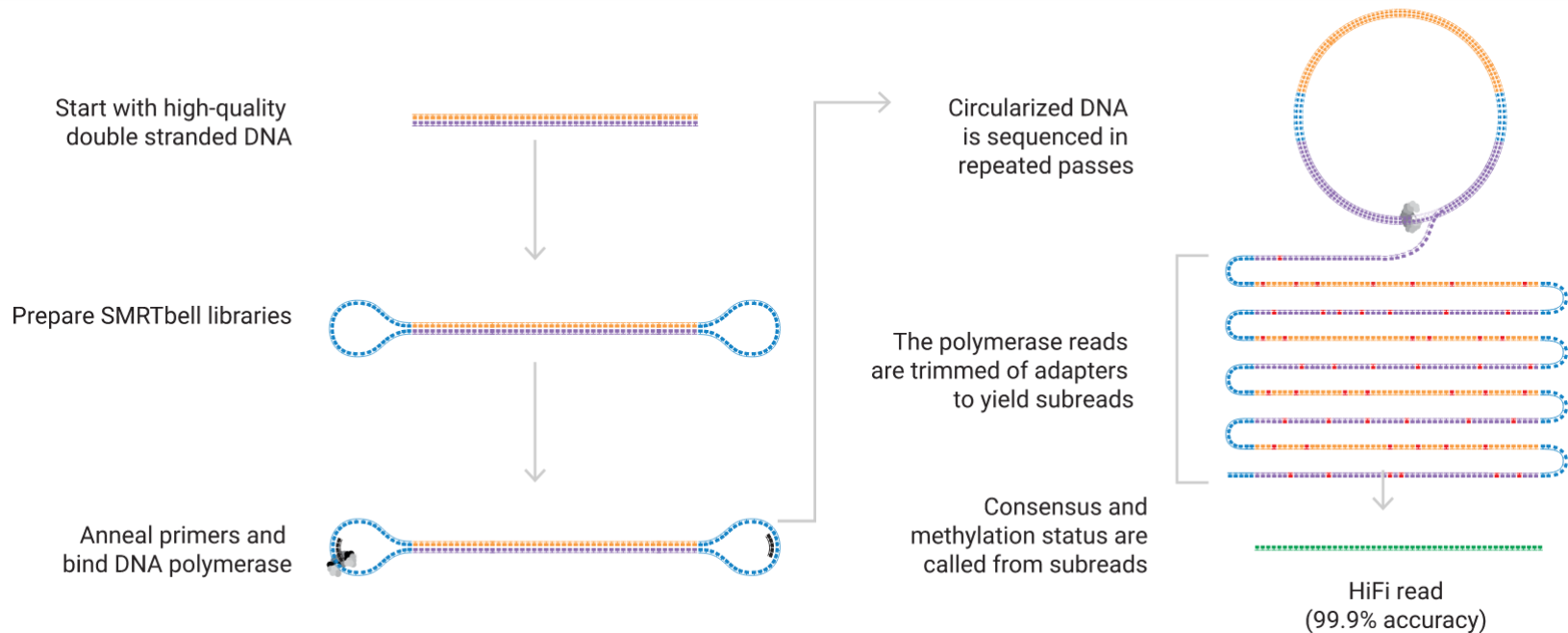


Pacbio RS – raw data



dlouhé čtení (15 kb), hodně chyb

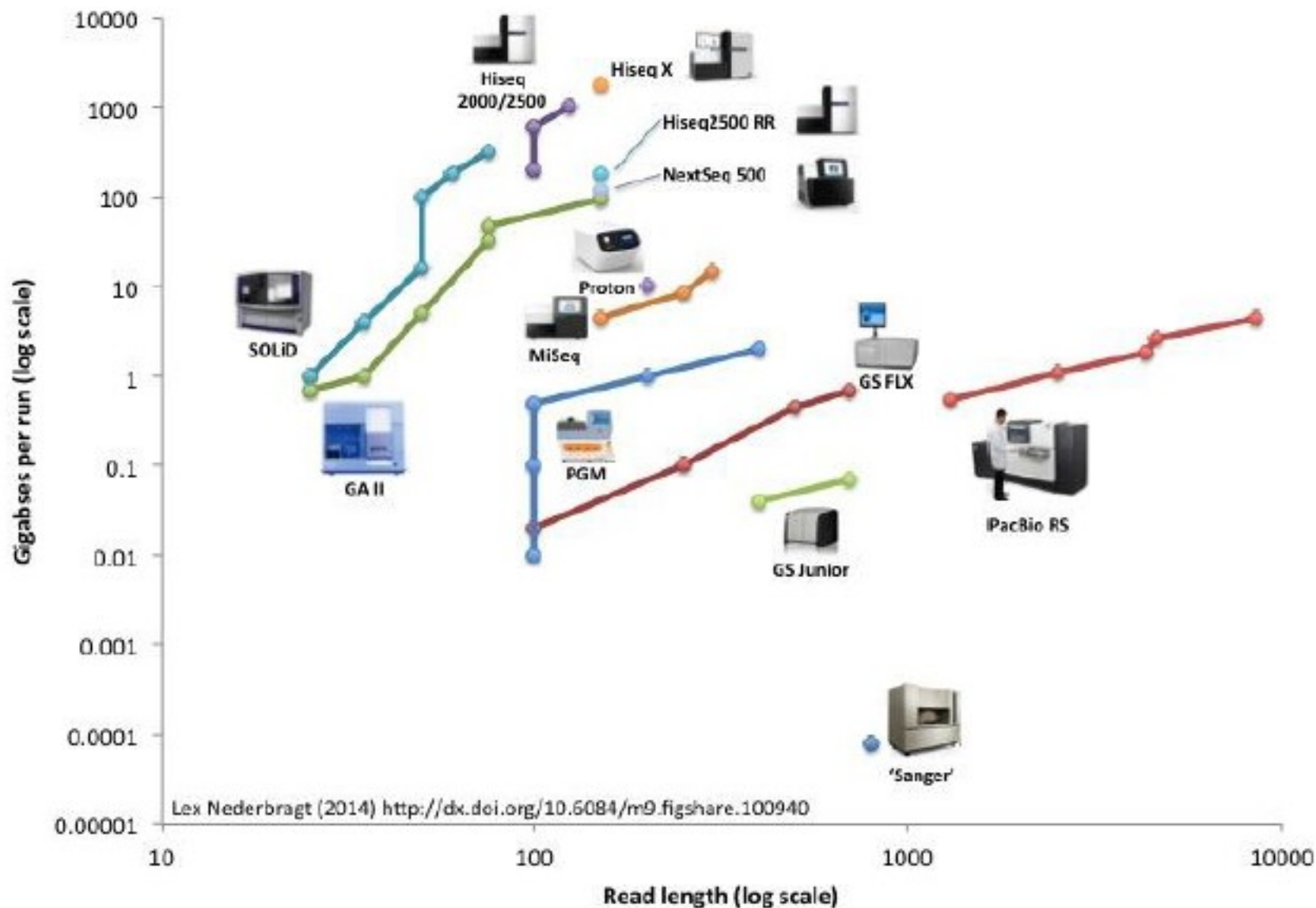
# HiFi long-read sequencing



**PacBio**

dlouhé čtení, velmi přesné

### Developments in High Throughput Sequencing



Lex Nederbragt (2014) <http://dx.doi.org/10.6084/m9.figshare.100940>

# Oxford Nanopore



**MinION**  
512 pores



**GridION**  
5 000 pores



# Future Sequencing Technologies

## Oxford Nanopore

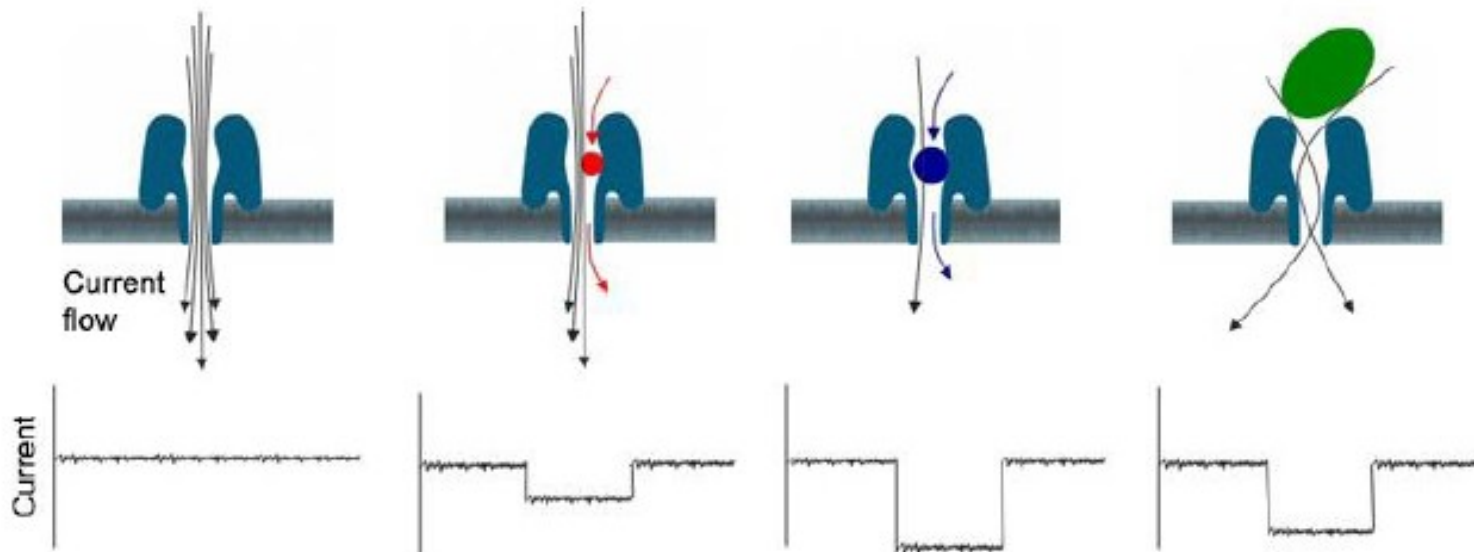
Nanopore sequencing  
up to 50 kb

„Run until sequencing ...“





# Princip technologie



<https://www.youtube.com/watch?v=CGWZvHli3i0>

# Sekvenování přímo v terénu (?)



Ebola outbreak

*Quick et al., Nature 2016*



# Traditional Sequencing vs. Next Generation Sequencing: Data Throughput

1 x Illumina GAI



200+ of 3730xl



Vs.

Days vs. Years

**The Sequencing Landscape is Changing**



# Bioinformatika - největší brzda dalšího rozvoje

Basically, analyzing genomes in interaction with their environment is now feasible and accessible to anyone



# Sekvenační strategie

- nutno velmi dobře počítat než se začne sekvenovat
- celkový výtěžek sekvenování = **počet „reads“ \* délka „reads“ \* coverage**
- zásadně závisí na konkrétním cíli výzkumu a použité technologii

Benchtop Sequencers

Production-Scale Sequencers



iSeq 100



MiniSeq



MiSeq Series +



NextSeq 550 Series +



NextSeq 1000 & 2000

Popular Applications & Methods	Key Application <span style="color: gray;">■</span>	Key Application <span style="color: green;">■</span>	Key Application <span style="color: orange;">■</span>	Key Application <span style="color: cyan;">■</span>	Key Application <span style="color: pink;">■</span>
Large Whole-Genome Sequencing (human, plant, animal)					
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●	●
Exome & Large Panel Sequencing (enrichment-based)				●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)				●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)				●	●
Targeted Gene Expression Profiling	●	●	●	●	●

Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours	11–48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	330 Gb*
Maximum Reads Per Run	4 million	25 million	25 million †	400 million	1.1 billion*
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp

[Explore iSeq 100](#)

[Explore MiniSeq](#)

[Compare MiSeq](#)

[Compare NextSeq 550](#)

[Explore NextSeq 1000 & 2000](#)





NextSeq 550 Series



NextSeq 1000 & 2000



NovaSeq 6000

Popular Applications & Methods	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)			●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●
Exome & Large Panel Sequencing (enrichment-based)	●	●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)	●	●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)	●	●	●
Chromatin Analysis (ATAC-Seq, ChIP-Seq)	●	●	●
Methylation Sequencing	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)	●	●	●
Cell-Free Sequencing & Liquid Biopsy Analysis	●	●	●

Run Time	12-30 hours	11-48 hours	~13 - 38 hours (dual SP flow cells) ~13-25 hours (dual S1 flow cells) ~16-36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	360 Gb*	6000 Gb
Maximum Reads Per Run	400 million	1.2 billion*	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 250**

[Compare NextSeq 550](#)

[Request Pricing](#)

[Explore NextSeq 1000 & 2000](#)

[Request Pricing](#)

[Explore NovaSeq 6000](#)

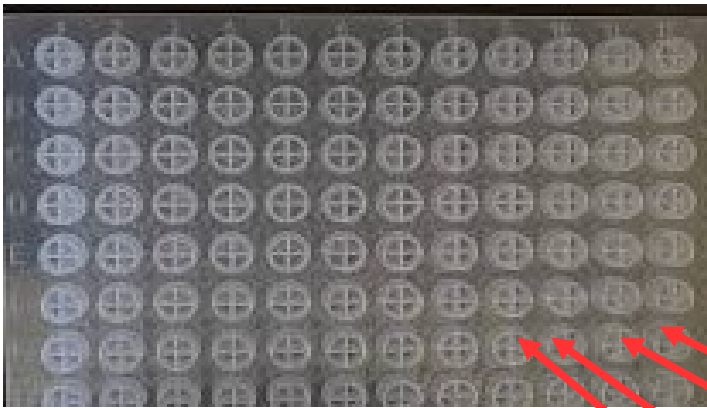
[Request Pricing](#)

# Sekvenační strategie

...JEDEN VZOREK NA RUN JE MÁLO

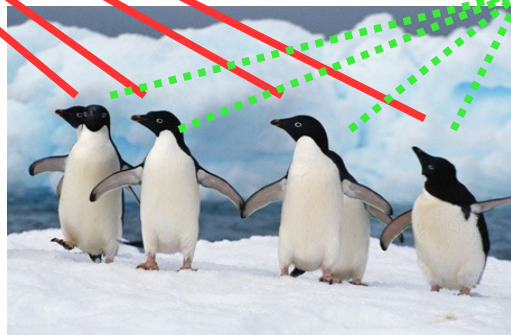
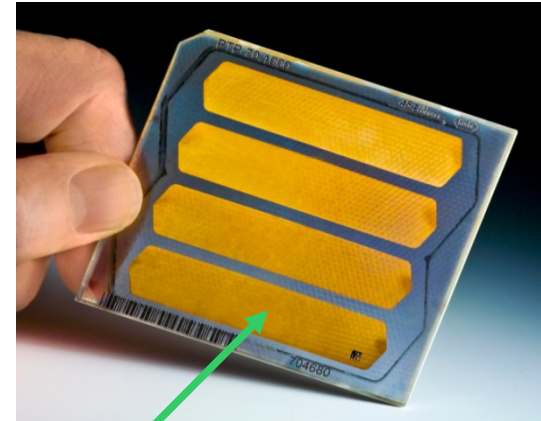
## Kapilární sekvenátor

U kapilárních sekvenátorů není problém přiřadit sekvenci k jednotlivým vzorkům na základě pozice na platíčku



## Sekvenátor druhé generace

U sekvenátorů druhé generace se najednou sekvenuje pool desítek až stovek vzorků



# Sekvenační strategie

...JEDEN VZOREK NA RUN JE MÁLO

Jednotlivé vzorky pro sekvenátory druhé generace se značí tzv. barcodes (midy, tagy)

Krátká (obvykle 6-12bp) oligonukleotidová sekvence před primerem (pokud sekvenujeme PCR amplikon), která je specifická pro daný vzorek

Přiřazení identity jednotlivých sekvencí k vzorkům probíhá bioinformaticky

BARCODE

PRIMER

SEQUENCE

```
AGCGTAGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....  
TTCGTAGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....  
TGGGTAGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....  
TGCCTAGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....  
TGCGCAGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....  
TGCGTIGGTCATTTCGATGCGGTCATGCCTGGATTAAAGCT.....
```

# Sekvenační strategie

AMPLIKONOVÉ SEKVENOVÁNÍ (amplikony kratší než délka readů)

SHOT GUN SEKVENOVÁNÍ

LONG-RANGE PCR + SHOT GUN (amplikony delší než délka readů)

---

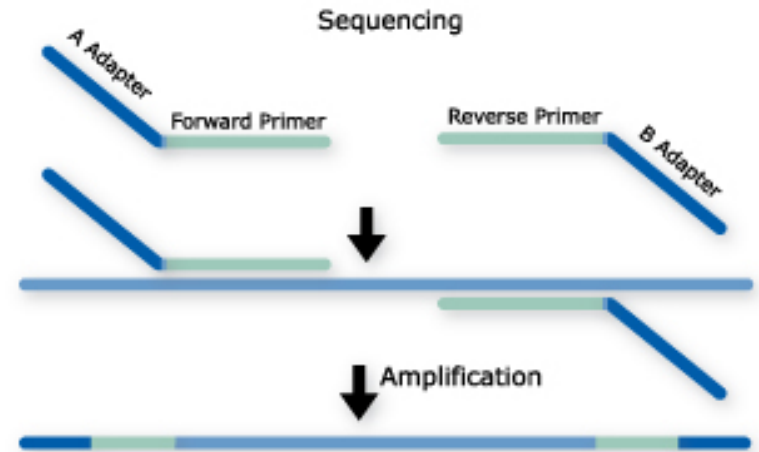
# Sekvenační strategie

## AMPLIKONOVÉ SEKVENOVÁNÍ

PCR Amplifikace konkrétního úseku daného genomu pomocí specifických primerů (se sekvenačními adaptory)

Následná sekvenace

*Taxonomické složení daného vzorku („metabarcoding“), variabilita konkrétních genů apod.*



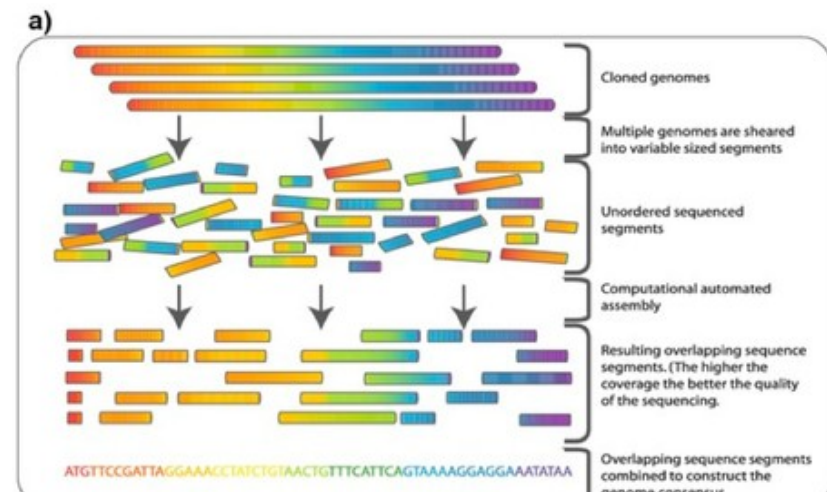
## SHOT GUN SEKVENOVÁNÍ

Fragmetace celogenomové DNA

Ligace sekvenačních adaptorů

Následná sekvenace náhodných fragmentů

*De novo assembly, resekvenování, transkriptomika, funkční složení daného společenstva*



# Sekvenační strategie

## LONG RANGE PCR + SHOT GUN

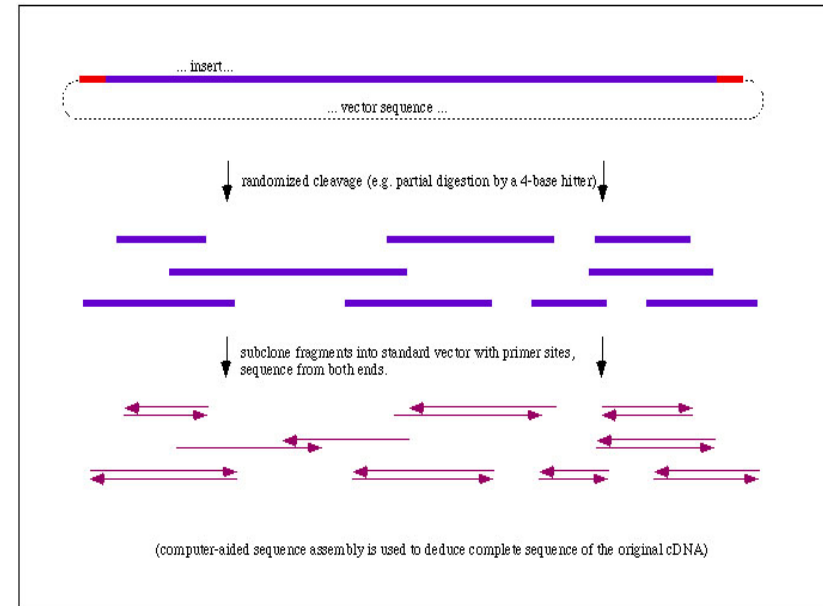
*Dlouhé PCR produkty, které nejdou vcelku osekvenovat*

*Jejich fragmentace*

*Sekvenování fragmetů*

*Zpětná rekonstrukce původní sekvence („assembly“)*

*Použitelné pokud nás zajímá variabilita v jednolitém úseku DNA. Např. sekvenace kompletní mitochondriální DNA (3 různé PCR produkty).*



# Sekvenační strategie

AMPLIKONOVÉ SEKVENOVÁNÍ (amplikony kratší než délka readů)

SHOT GUN SEKVENOVÁNÍ

LONG-RANGE PCR + SHOT GUN (amplikony delší než délka readů)

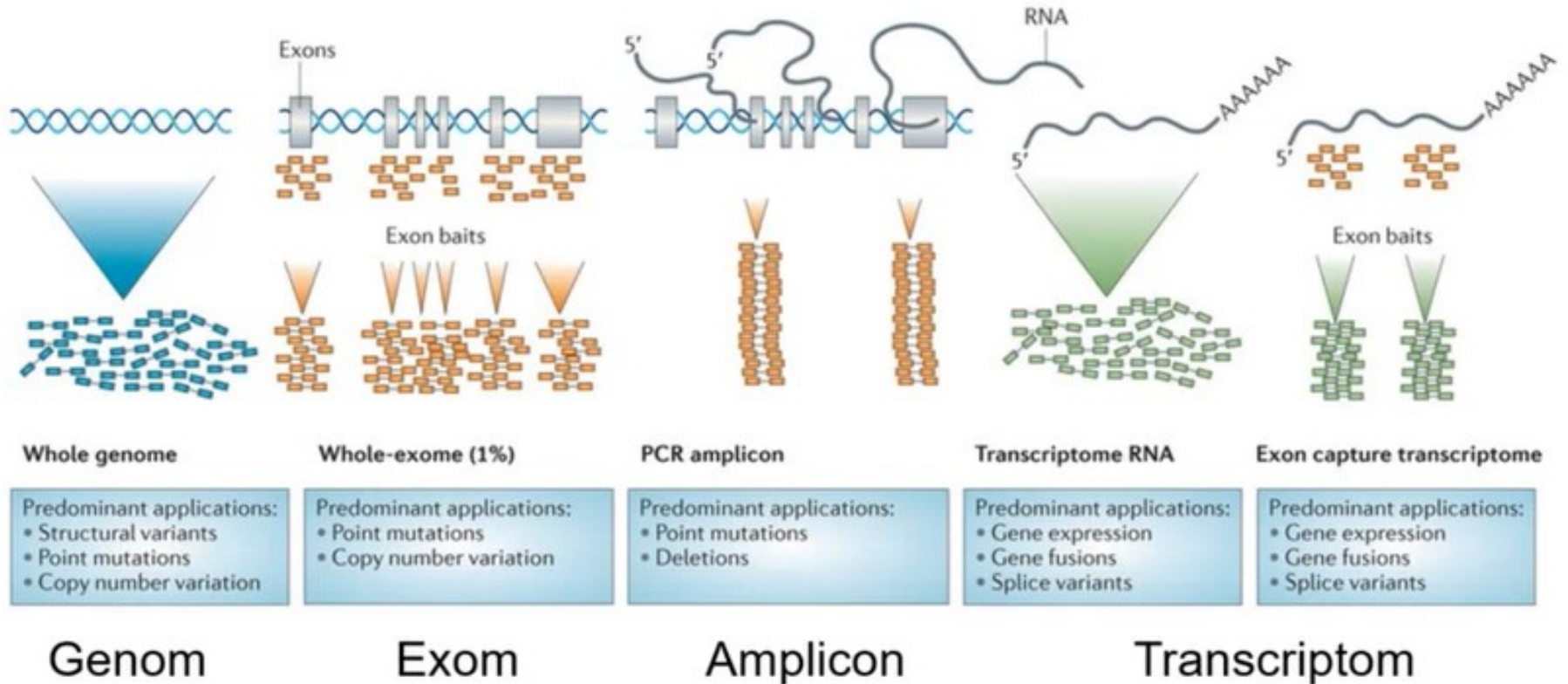
COMPLETE GENOME (e.g. viral genome from enriched samples)

REDUCED GENOME

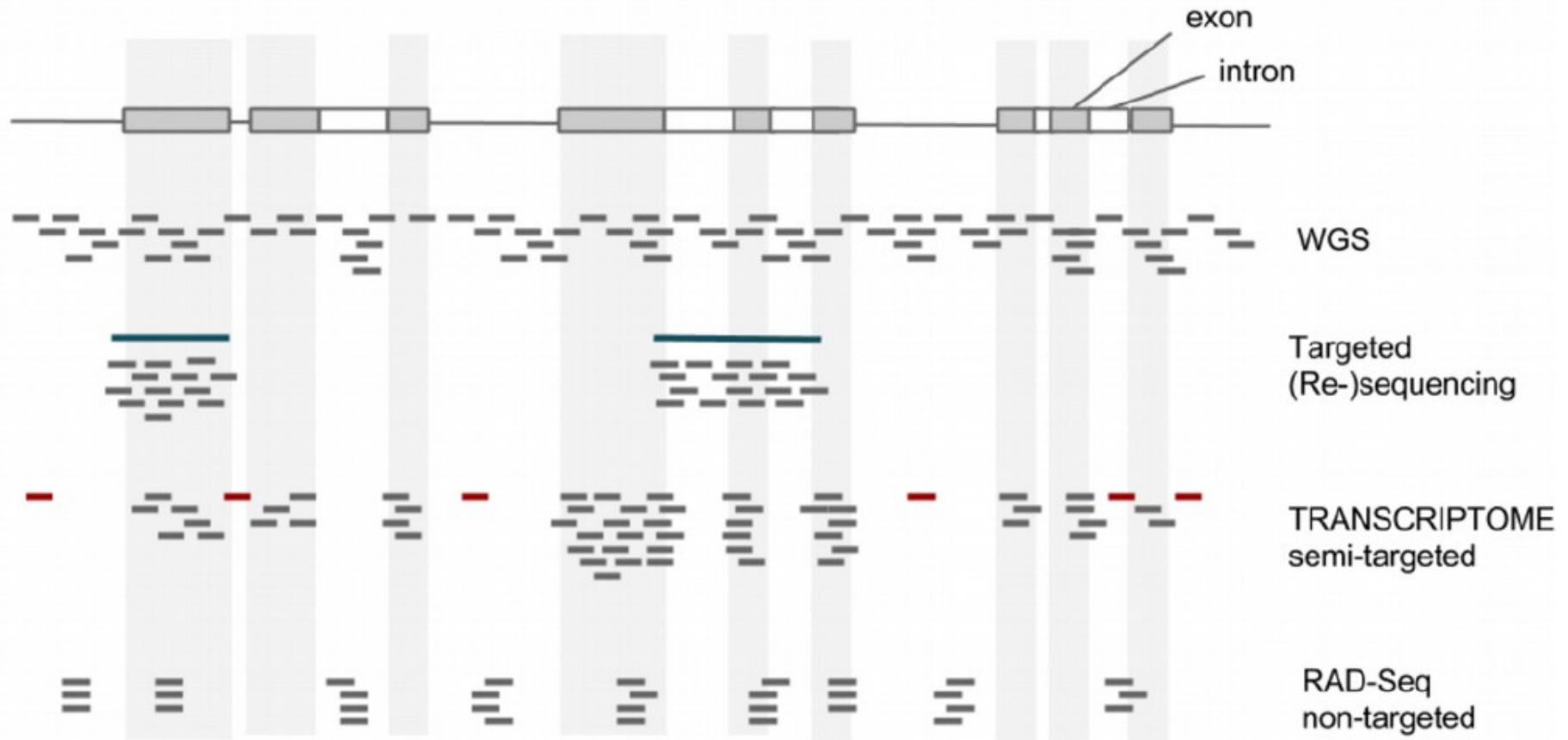
- **PCR** amplicons
- Enriched libraries by **hybridization** (development of microsatellite markers, exom, anchored phylogenomics, UCE = ultraconserved elements, etc.)
- Enriched libraries by **restriction enzymes** (RAD sequencing)
- RNAseq (**transcriptomics**)



# Sekvenační strategie



# Sekvenační strategie



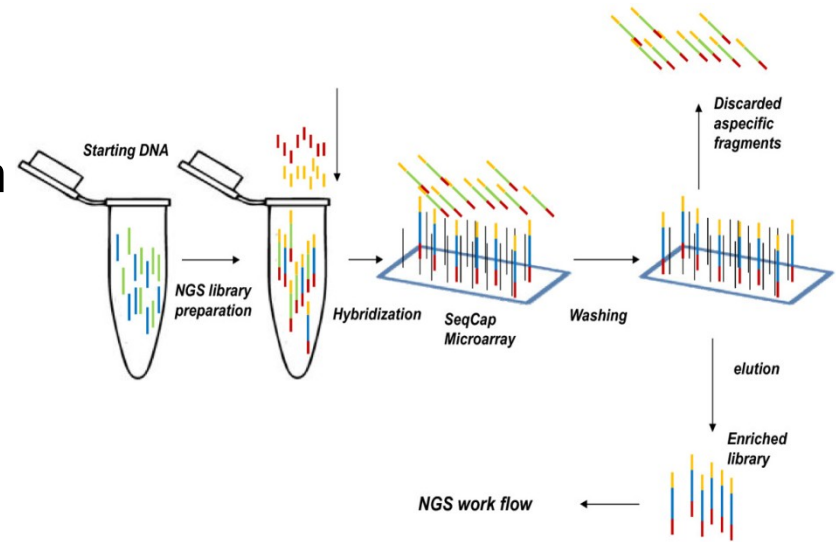
# Sekvenační strategie

## Enrichment by hybridization + shot gun

Separace úseků genomu které nás zajímají na základě jejich hybridizace

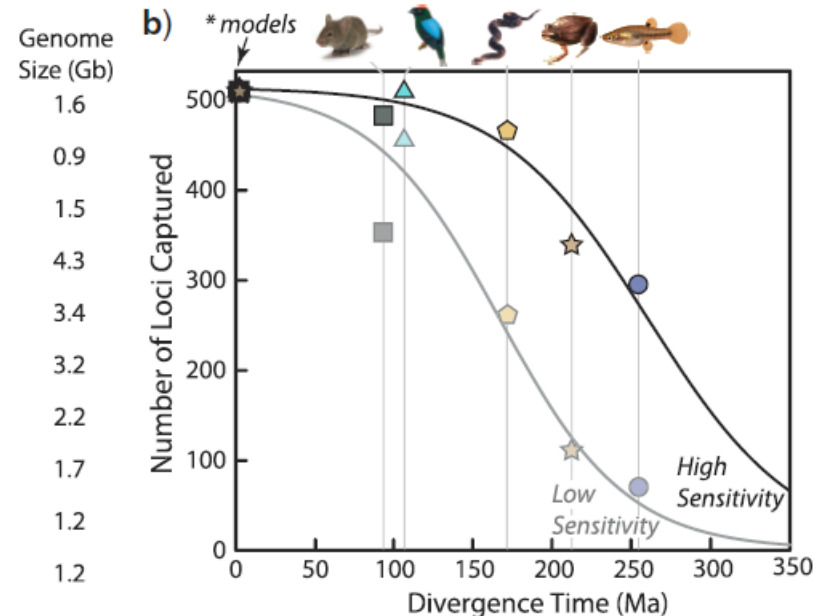
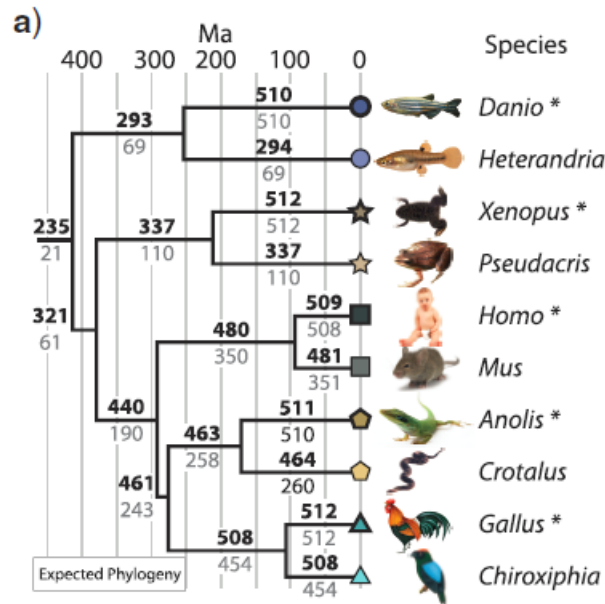
Následná sekvenace obohacených knihoven („enrichment by baits“)

Nové markery (mikrosatelity apod.), kódující oblasti genomu („exom“), „anchored phylogenomics“ apod.



## Anchored phylogenomics

- hundreds of conserved loci
- hybridization enrichment
- u velmi příbuzných taxonů bude málo variability





# CENTER FOR ANCHORED PHYLOGENOMICS

*ACCELERATING THE RESOLUTION OF LIFE™*



## A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing

Richard O. Prum<sup>1,2\*</sup>, Jacob S. Berv<sup>3\*</sup>, Alex Dornburg<sup>1,2,4</sup>, Daniel J. Field<sup>2,5</sup>, Jeffrey P. Townsend<sup>1,6</sup>, Emily Moriarty Lemmon<sup>7</sup> & Alan R. Lemmon<sup>8</sup>



**Nature Paper Resolves Bird Tree of Life**

October 2015

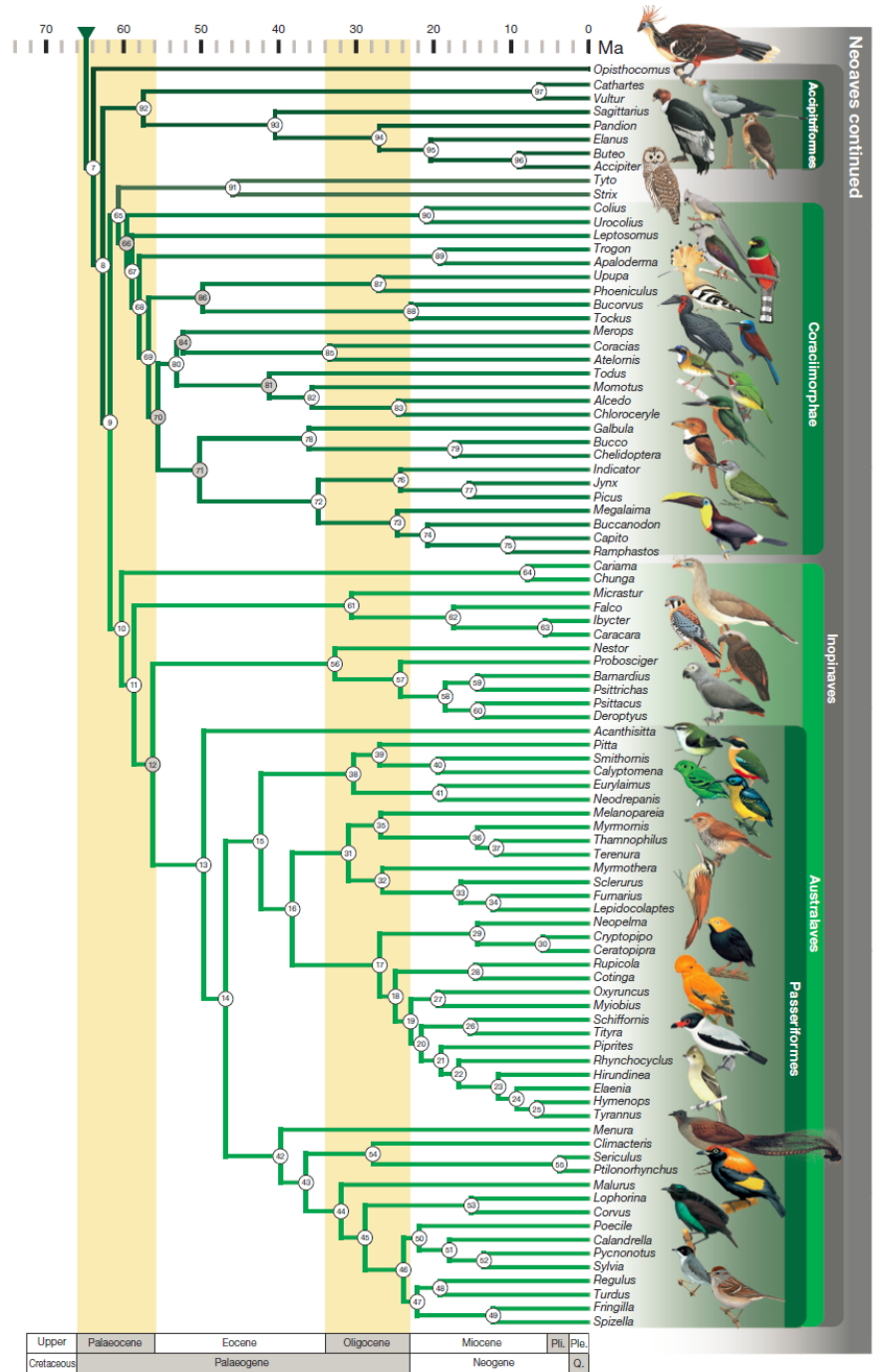
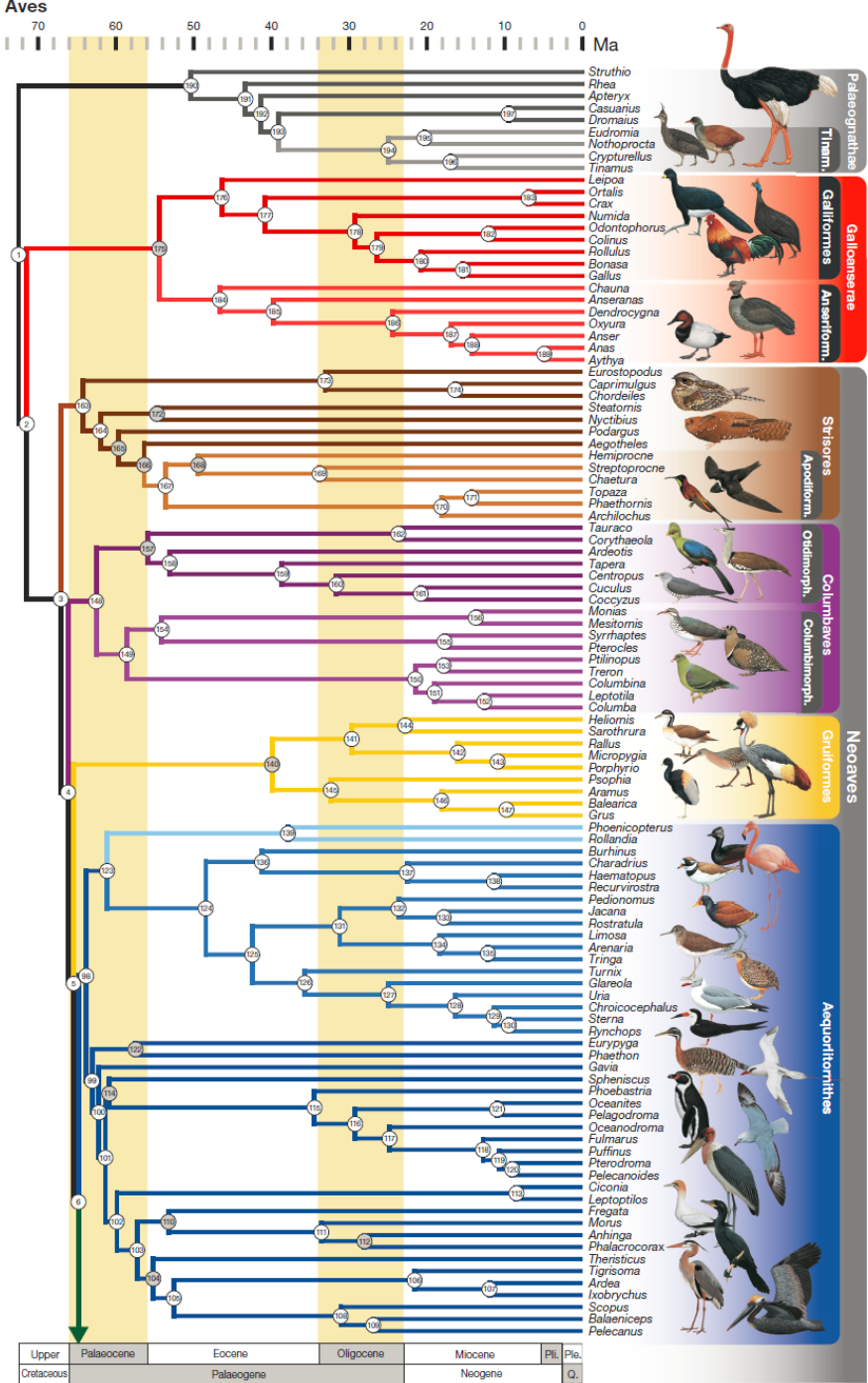
Posted on [October 6, 2015](#) by [ameer](#)

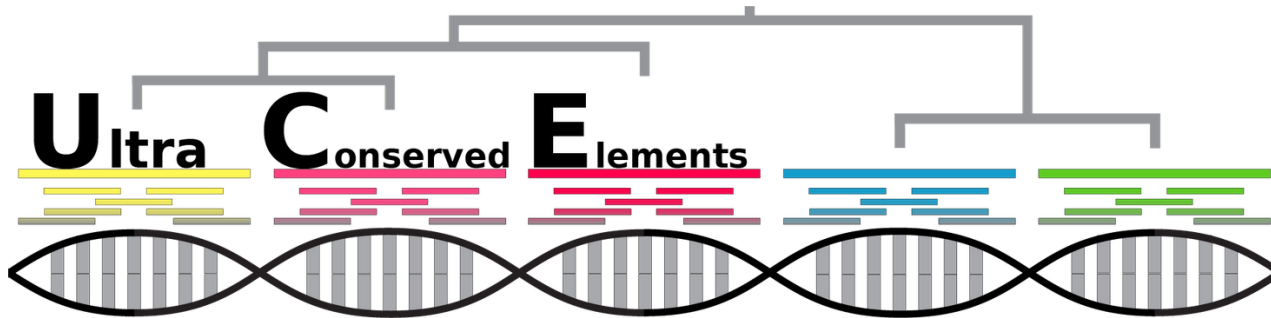
198 species

259 nuclear loci (ca 1500 bp each)

> 390 000 bp







### What are UCEs?

As their name implies, ultraconserved elements (UCEs) are highly conserved regions of organismal genomes shared among evolutionary distant taxa - for instance, birds share many UCEs with humans. UCEs were first described in a wonderful manuscript by Gil Bejerano et al. (2004) from David Haussler's group and subsequently identified in several classes of organisms outside the group of original taxa (Siepel et al. 2005) used to identify these genomic elements. The 27-way vertebrate genome alignment (Miller et al. 2007) identified additional regions of high conservation.

### Why are UCEs useful?

We have discovered (see Citations) that we can collect data from UCEs and the DNA adjacent to UCE locations (flanking DNA), and that these data are useful for reconstructing the evolutionary history and population-level relationships of many organisms. Because UCEs are conserved across disparate taxa, UCEs are also universal genetic markers in the sense that the locations (or loci) that we can target in humans are identical, in many cases, to the loci that we can target in ducks or snakes or lizards.

### What do UCEs do?

That's an extremely good question, and one to which we do not entirely know the answer (Dermitzakis et al. 2005). UCEs have been associated with gene regulation (Pennachio et al. 2006) and development (Sandelin et al. 2004, Woolfe et al. 2004) and we generally assume that UCEs must be important by the very nature of their near-universal conservation across extremely divergent taxa. However, gene knockouts of UCE loci in mice resulted in viable, fertile offspring (Ahituv et al. 2007), suggesting that their role in the biology of the genome may be cryptic.

#### Arachnida

📍 14,799 baits for 1,120 UCEs  
(Arachnida 1.1Kv1)

Described as part of Faircloth 2017. First use as part of Starret et al. 2017.

[Get 1.9Kv1 bait design for Arachnida »](#)

#### Diptera

📍 31,328 baits for 2,711 UCEs  
(Diptera 2.7Kv1)

Described as part of Faircloth 2017. First use has not been published, yet.

[Get 2.7Kv1 bait design for Diptera »](#)

#### Hymenoptera (ver. 1)

📍 2,749 baits for 1,510 UCEs  
(Hymenoptera 1.5Kv1)

Described as part of Faircloth et al. 2015. First used as part of Faircloth et al. 2015.

[Get 1.9Kv1 bait design for Hymenoptera »](#)

#### Anthozoa

📍 16,306 baits for 720 UCEs and 1,071 exons  
(Anthozoa 1.7Kv1)

Described as part of Quattrini et al. 2017. First use as part of Quattrini et al. 2017.

[Get 1.9Kv1 bait design for Anthozoa »](#)

#### Coleoptera

📍 13,674 baits for 1,172 UCEs  
(Coleoptera 1.1Kv1)

Described as part of Faircloth 2017. First use as part of Baca et al. 2017.

[Get 1.1Kv1 bait design for Coleoptera »](#)

#### Hemiptera

📍 40,207 baits for 2,731 UCEs  
(Hemiptera 2.7Kv1)

Described as part of Faircloth 2017. First use has not been published, yet.

[Get 2.7Kv1 bait design for Hemiptera »](#)

#### Hymenoptera (ver. 2)

📍 31,829 baits for 2,590 UCEs  
(Hymenoptera 2.5Kv2)

Described as part of Branstetter et al. 2017. First use as part of Branstetter et al. 2017.

[Get 2.9Kv2 bait design for Hymenoptera »](#)

#### Tetrapod probe sets.

Below are several probe designs that we have used to study relationships among amniotes/tetrapods (e.g. Crawford et al. 2012, McCormack et al. 2013). We are constantly evaluating the utility of given probe sets and probe designs, in addition to expanding the number of UCE loci we are targeting. We have several larger bait sets in the works, and we are also working on optimizing probe sets based on their capture success, phylogenetic utility, etc. Please check back for updates.  
You can now buy each of these probe sets direct from Arbor Biosciences in the form of a capture kit. Arbor Biosciences has even made a **discounted "pilot" sized kit** available for labs who want to do some test enrichments.

[Order enrichment kits from Arbor Biosciences »](#)

📍 2,560 baits for 2,386 UCEs  
(Tetrapods-UCE-2.5Kv1)

Described as part of Faircloth et al. 2012. First use as part of Faircloth et al. 2012.

[Get 2.5Kv1 bait design for Tetrapods »](#)

📍 5,472 baits for 5,060 UCEs  
(Tetrapods-UCE-5Kv1)

Described in Faircloth et al. 2012 and first use as part of Kepez et al. 2014.

[Get 5Kv1 bait design for Tetrapods »](#)

#### Fish probe sets.

Below are two bait set designs that we have used (1) to understand relationships among the early diverging teleosts (Faircloth et al. 2013) and (2) to study the diversification of Acanthomorphs (Aflaro et al. 2018). We are currently working on several other bait set designs, as well as optimizing existing bait sets based on their capture success, phylogenetic utility, etc. Please check back for updates.

You can now buy both probe sets directly from Arbor Biosciences in the form of a capture kit. Arbor Biosciences has even made a **discounted "pilot" sized kit** available for labs who want to do some test enrichments.

[Order enrichment kits from Arbor Biosciences »](#)

📍 Actinopterygians  
2,001 baits for 500 UCEs  
(Actinopterygians 0.5Kv1)

Described as part of Faircloth et al. 2013. First use as part of Faircloth et al. 2013.

[Get 0.5Kv1 bait design for Actinopterygians »](#)

📍 Acanthomorphs  
2,628 baits for 1,314 UCEs  
(Acanthomorphs 1Kv1)

Described as part of Aflaro et al. 2013. First used as part of McCoe et al. 2016.

[Get 1Kv1 bait design for Acanthomorphs »](#)



# Sekvenační strategie

## Sekvenování podél restričních míst (Enriched libraries by restriction enzymes)

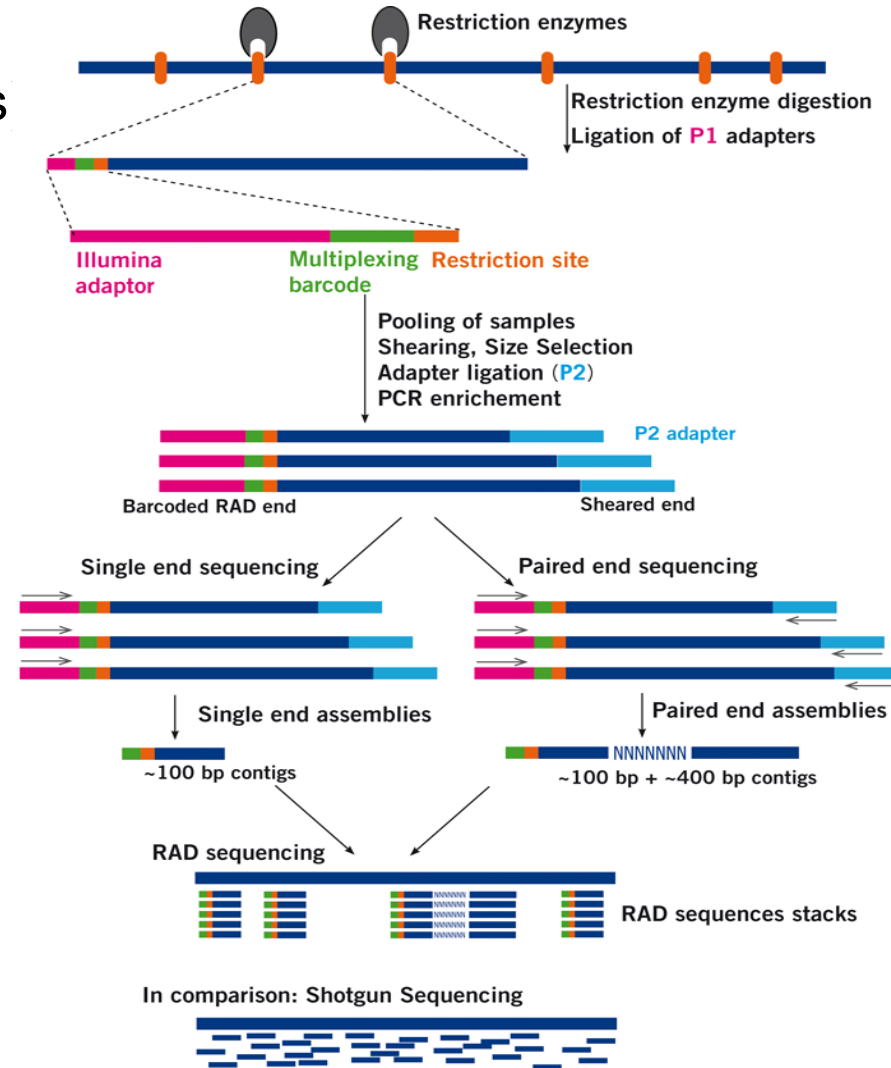
Fragmetace gelogenomové DNA pomocí  
restričních enzymů

Ligace sekvenačních adaptorů na výsledné  
fragmety

Následná sekvenace podél restričních míst

Celogenomové scany genetické variability

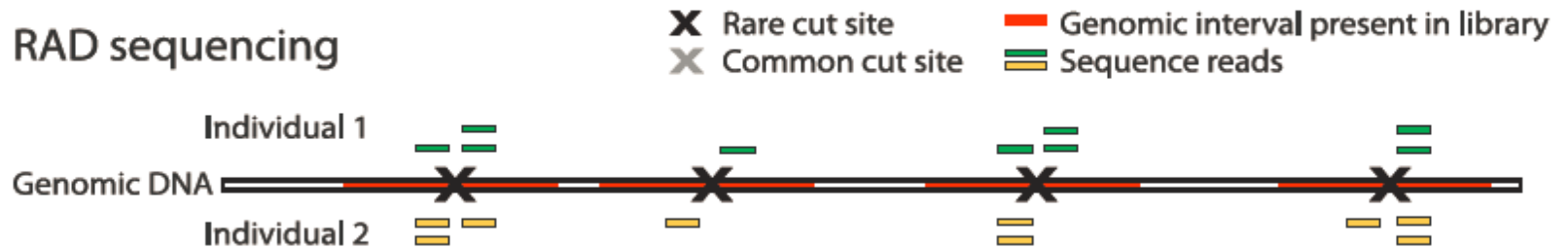
*Hledání SNPs, populační genomika (např. RAD-  
SEQ) apod.*



# RAD vs. ddRAD

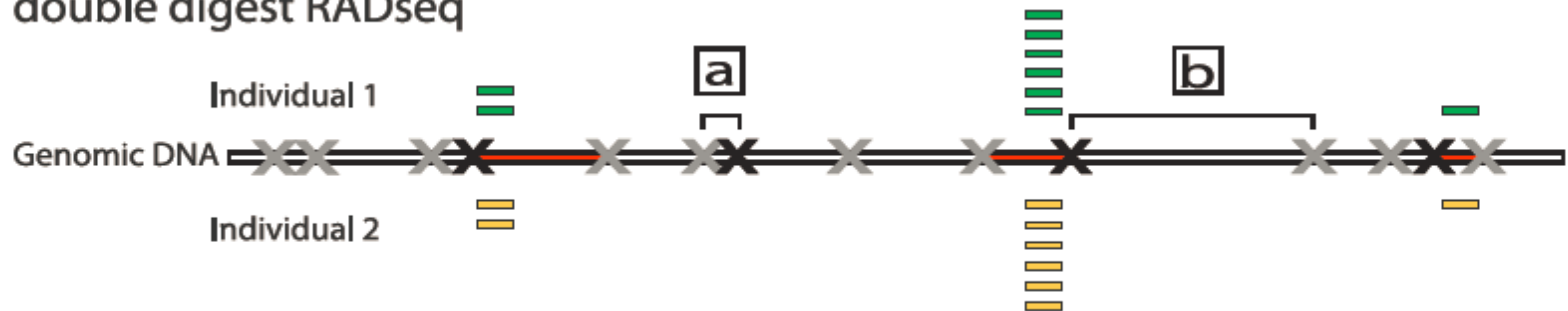
A

RAD sequencing



B

double digest RADseq



# Process of ddRAD

## 1-Digestion

2-size-selection with beads

## 3-Ligation

4-washing and cleaning

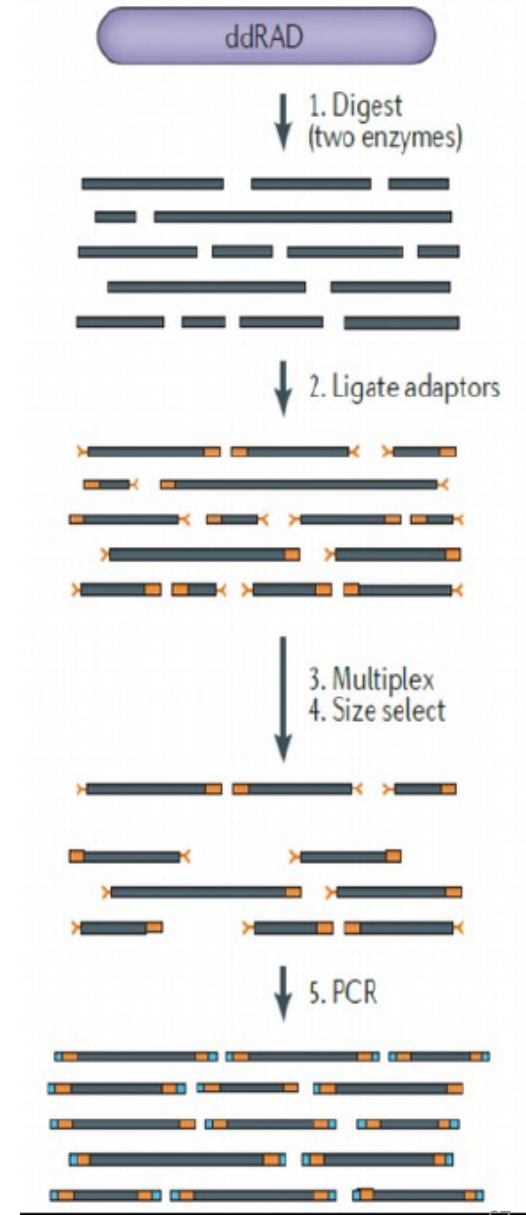
## 5-PCR

6-Pooling

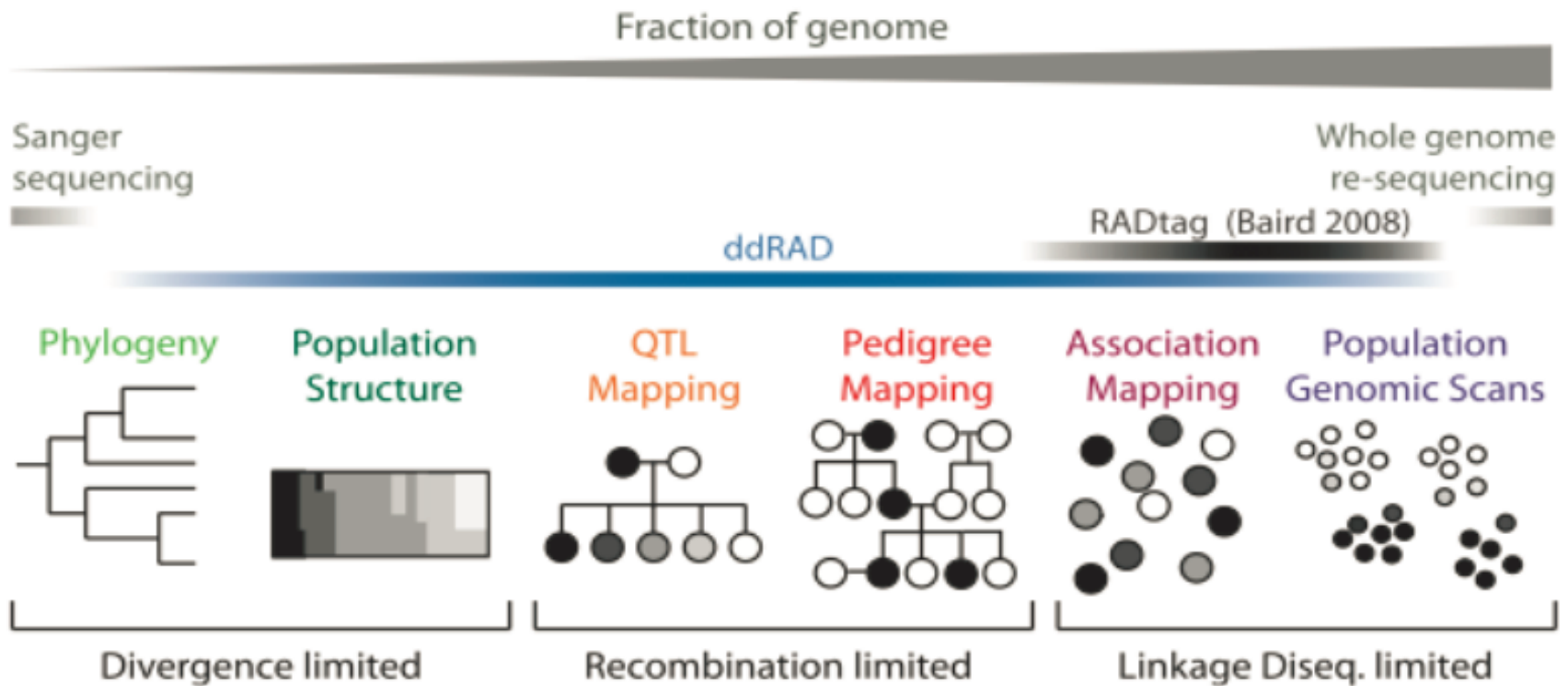
7-Size-selection, pippin prep and bioanalyser

8-qPCR

9-Sequencing



# Sekvenování podél restričních míst



# ddRAD library

o tuto sekvenci nám jde!

## DATA ANALYSIS

Complete adapter+insert(EcoRI and MspI) = SEQUENCING LIBRARY :

5' - AATGATACGGCGACCACCAGATCTACACACCGACAACACTCTTTCCCTACACGACGCTCTTCCGATC CATCCAAAT CGAGATCGGAAGAGCACACGTC TGAAC TCCAGT CACAGGTC ACTATCTCGTATGCCGCTTCTTGCTTG-3'  
 3' - TTACTATGCCGCTGGTGGCTCTAGATGTGTGGCTGTTGTGAGAAAGGGATGTGCTGCGAGAAGGCTAG GTAGGTTTA GCTCTAGCCTTC TCGTGTGCAGACTTGAGGTCAGTGT CCAAGTATAGAGCATACGGCAGAAGACGAAC-5'

Illumina adapter

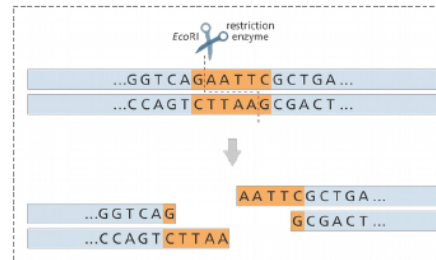
sekvenační primery

Illumina adapter

ACCGACAA P5 Index example  
 TCCAGTGA P7 Index example  
 CATCCA Inline barcode

(can be preceded by 1-2 bp to increase complexity)

AATT RE overhang  
 CG RE overhang  
 -----> R1 sequence  
 <----- R2 sequence  
 -----> I1 sequence  
 <----- I2 sequence  
 XXXXXX insert - the RE fragment



identifikace vzorku (jedince)

# Data analysis in ddRADseq

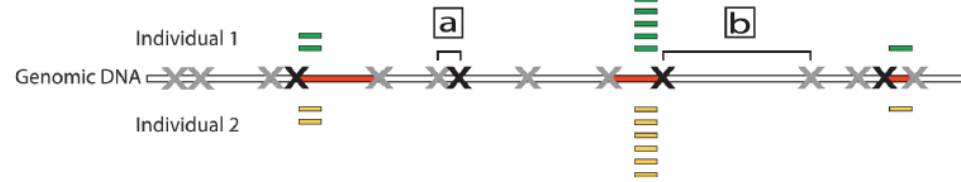
## DATA ANALYSIS

### B) DEFINE LOCI AND FIND VARIABILITY

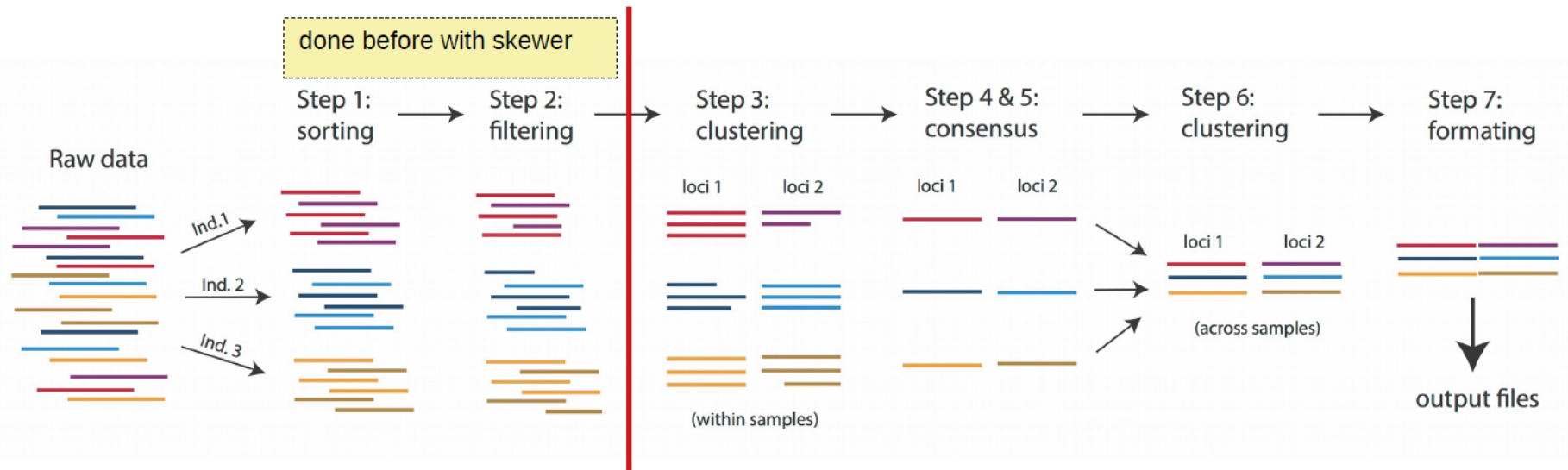
SOFTWARE AVAILABLE:

Stacks, dDocent, **iPyrad**....

double digest RADseq



### iPyrad denovo assembly workflow (no reference genome)



program Skewer

program iPyrad

- stovky až desítky tisíc lokusů



# Aplikace

1. Celogenomové sekvenování de novo
2. Celogenomové resekvenování
3. Sekvenování amplikonů (PCR produktů)
4. Další aplikace - např. hledání klasických DNA markerů (mikrosatelity, SNPs)



# 1. Celogenomové sekvenování de novo

Problém: **KRÁTKÝ READ LENGTH**

- max **300bp** u Illumina, **35-75bp** Solid

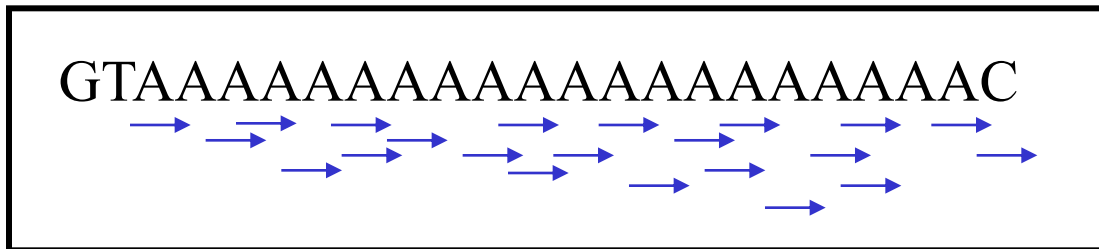
vs **800-1000bp** Sanger

- nové technologie (PacBio, Nanopore) už s tím takový problém nemají



→ Uspořádání (assembly) ještě stále může být problém z hlediska výpočetní kapacity

!!!! **REPETITIVNÍ OBLASTI** delší než read length !!!!



Zvláště komplexní eukaryotické genomy - úseky souvislých oblastí přerušovaných mezerami

# 1. Celogenomové sekvenování de novo

- získání kompletní uspořádané sekvence celých velkých eukaryotních genomů pomocí next-generation sequencing de novo je problém (ale to je nakonec i u Sangeru)
- viry, prokaryota, malá eukaryota, mitochondrie/plastidy/plasmidy

**Genetic Det**  
**New Hemor**  
**Southern Af**

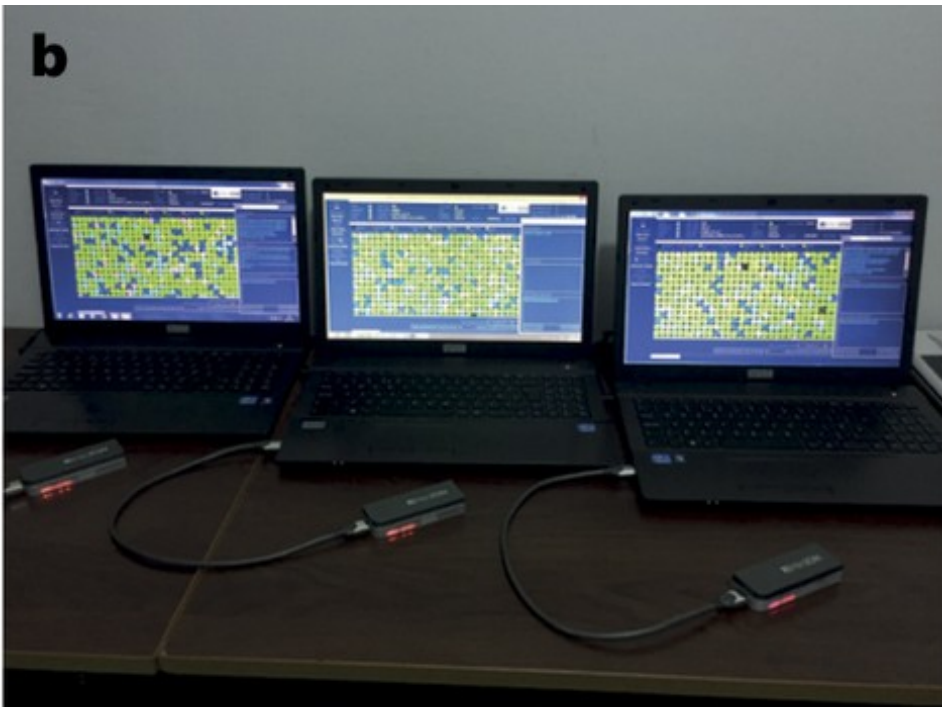
**Thomas Briese<sup>1,3\*</sup>, Jan**  
**Gustavo Palacios<sup>1</sup>, Ma**  
**Stuart T. Nichol<sup>3</sup>, W. I**

**1** Center for Infection and Immunity,  
National Institute for Communicable  
Rickettsial Diseases, Centers for Disease  
America, **5** Biotechnology Core Facil

**Abstract**

Lujo virus (LUJV), a new  
Old World discovered in  
nosocomial transmission  
extracts from serum ar  
within 72 hours of sam  
node of the Old World  
that of other Old World  
novel, genetically distinct, highly pathogenic arenavirus.

**b**



**2015**

**2009**

## 2. Celogenomové resekvenování

- podobné problémy jako u de novo, ale méně (větší strukturální přestavby..)

### KOMPARATIVNÍ GENOMIKA

- viry, prokaryota, malá eukaryota
- mitochondrie/plastidy/plasmidy

### ANCIENT (mt) DNA

- různé směsné, degradované vzorky, např. fosilie

---

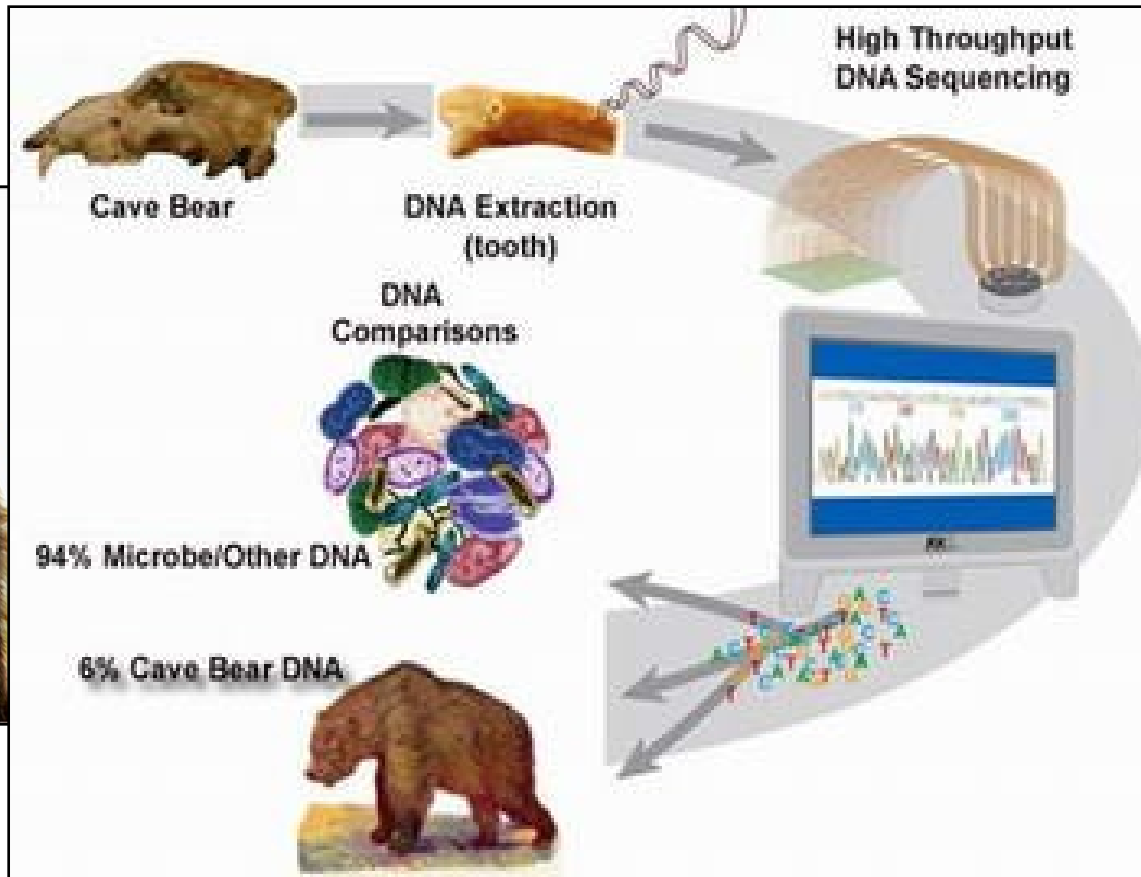
Cell

### A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing

Richard E. Green,<sup>1,\*</sup> Anna-Sapfo Malaspinas,<sup>2</sup> Johannes Krause,<sup>1</sup> Adrian W. Briggs,<sup>1</sup> Philip L.F. Johnson,<sup>3</sup> Caroline Uhler,<sup>4</sup> Matthias Meyer,<sup>1</sup> Jeffrey M. Good,<sup>1</sup> Tomislav Maricic,<sup>1</sup> Udo Stenzel,<sup>1</sup> Kay Prüfer,<sup>1</sup> Michael Siebauer,<sup>1</sup> Hernán A. Burbano,<sup>1</sup> Michael Ronan,<sup>5</sup> Jonathan M. Rothberg,<sup>6</sup> Michael Egholm,<sup>5</sup> Pavao Rudan,<sup>7</sup> Dejana Brajković,<sup>8</sup> Željko Kučan,<sup>7</sup> Ivan Gušić,<sup>7</sup> Märten Wikström,<sup>9</sup> Liisa Laakkonen,<sup>10</sup> Janet Kelso,<sup>1</sup> Montgomery Slatkin,<sup>2</sup> and Svante Pääbo<sup>1</sup>

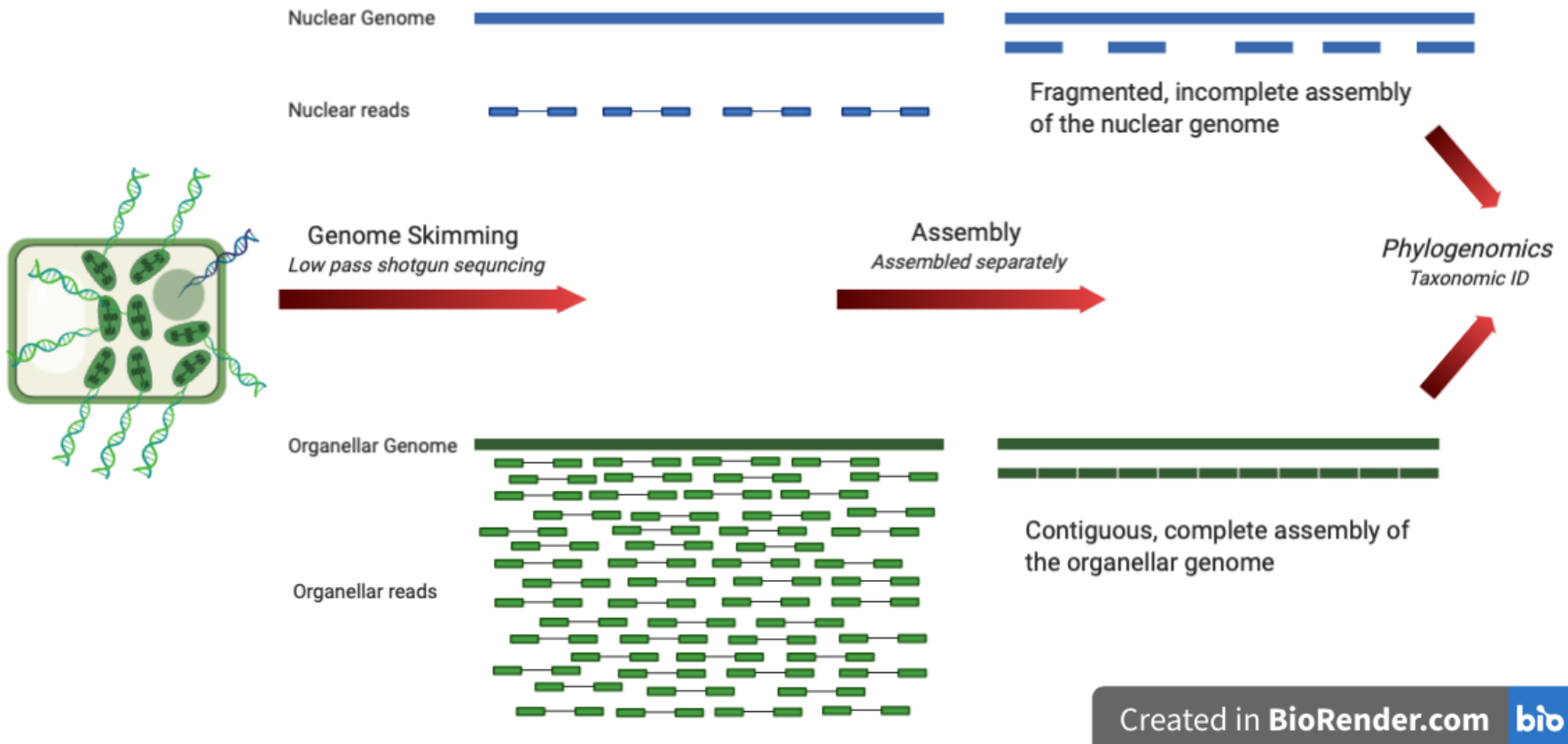
# Ancient Genomes Resurrected

- Degraded state of the sample → mtDNA sequencing
- Nuclear genomes of ancient remains: cave bear, mammoth, Neanderthal ( $10^6$  bp )



**Problems: contamination modern humans and coisolation bacterial DNA**

# Genome skimming



Analysis of museum material (e.g. holotypes, etc.)

### 3. Sekvenování amplikonů (PCR produktů)

SMĚSNÉ VZORKY - paralelní sekvenování nahrazuje klonování

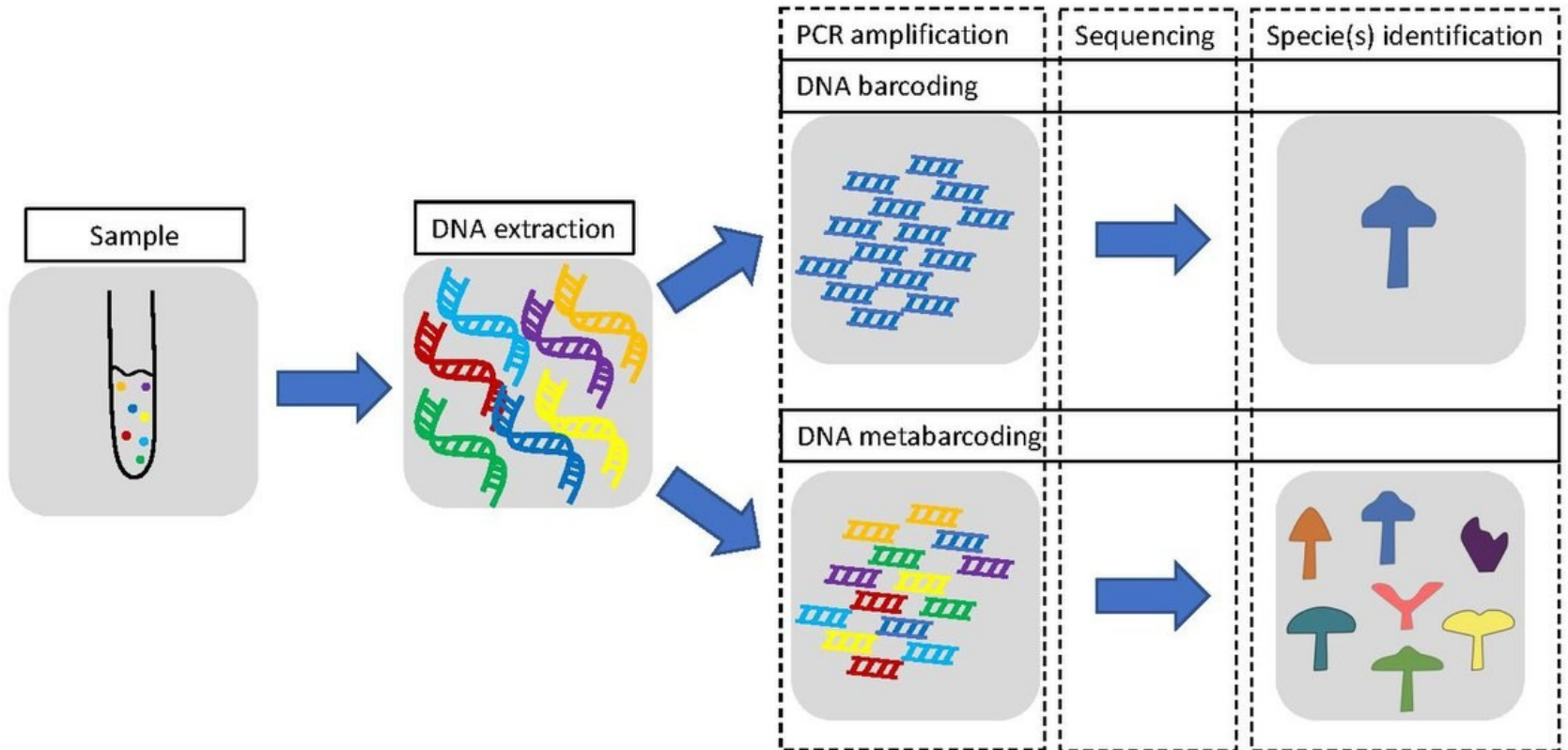
#### Metagenomika (= hlavně prokaryota)

- Celé společenstvo půdních, vodních mikroorganismů, střevní mikroflóra - **mikrobiom**
- PCR genu 16S rRNA
- lze i kvantifikovat

#### Metabarcoding (= hlavně eukaryota, ale dnes používáno jako obecný termín)

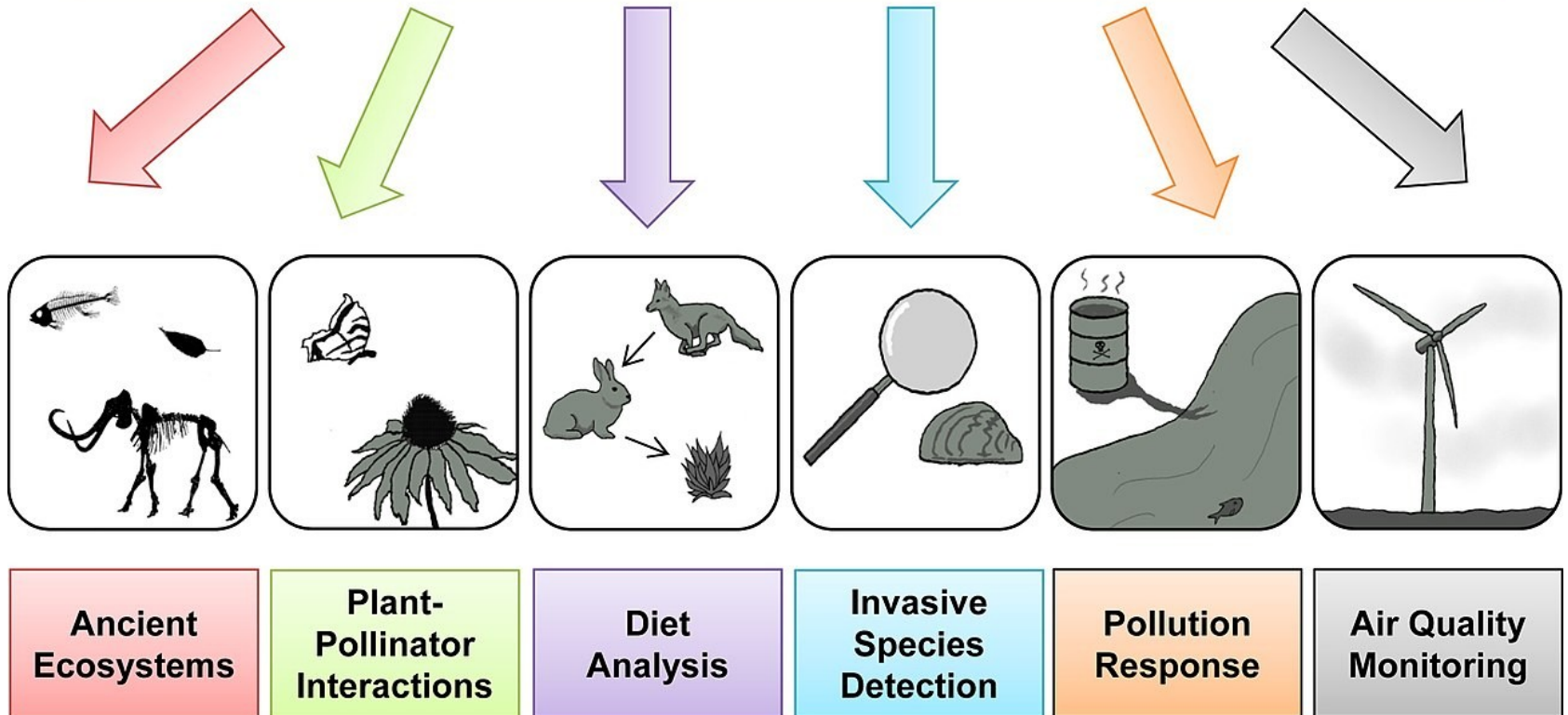
- COI gen, příp. jiný barcodingový marker

# Metabarcoding





# eDNA Metabarcoding Applications



eDNA = environmental DNA

# **Metabarcoding: Taxonomické složení společenstva v environmentální DNA na základě taxonomicky informativního úseku DNA (cyt b, COI, ITS, rRNA...)**

- Směsný vzorek environmentální DNA
- Amplifikace pomocí primerů specifických pro cílovou skupinu, pokrývajících taxonomicky informativní úsek (COI, 16s/18s RNA...)
- Paralelní sekvenování
- Filtrování nekvalitních sekvencí
- Klastrování na základě sekvenční podobnosti do OTUs („operational taxonomic units“)
- Jejich taxonomické zařazení na základě referenčních databází

## **Využití: Analýza druhového složení vzorků kde lze makroskopicky jednotlivé druhy obtížně odlišit**

- Potravní analýza z trusu
- Vzorky půdy
- Mikrobiální společenstva
- Permafrost
- Exotická/špatně probádaná společenstva
- Druhově bohatá společenstva („insect traps“ v tropech)
- Rutinní analýza velkého množství vzorků

# Metabarcoding - příklady využití

Monitoring vzácných, nedávno popsáných druhů savců na základě sekvenování krve pijavic

Výrazně větší úspěšnost prokázání přítomnosti než za použití klasických technik – fotopasti, terénní pozorování apod.

## Correspondences

### Screening mammal biodiversity using DNA from leeches

Ida Bærholm Schnell<sup>1,2,†</sup>,  
Philip Francis Thomsen<sup>2,†</sup>,  
Nicholas Wilkinson<sup>3</sup>,  
Morten Rasmussen<sup>2</sup>,  
Lars R.D. Jensen<sup>1</sup>, Eske Willerslev<sup>2</sup>  
Mads F. Bertelsen<sup>1</sup>,  
and M. Thomas P. Gilbert<sup>2,\*</sup>

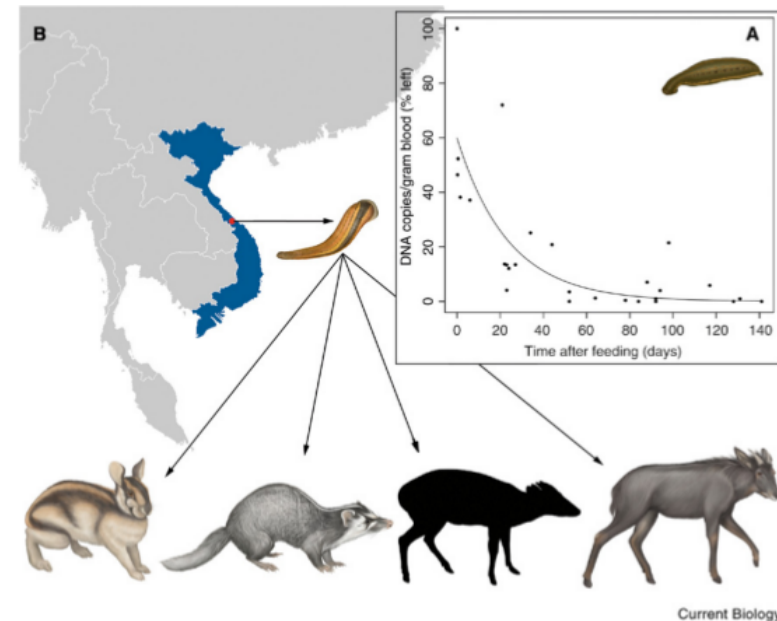
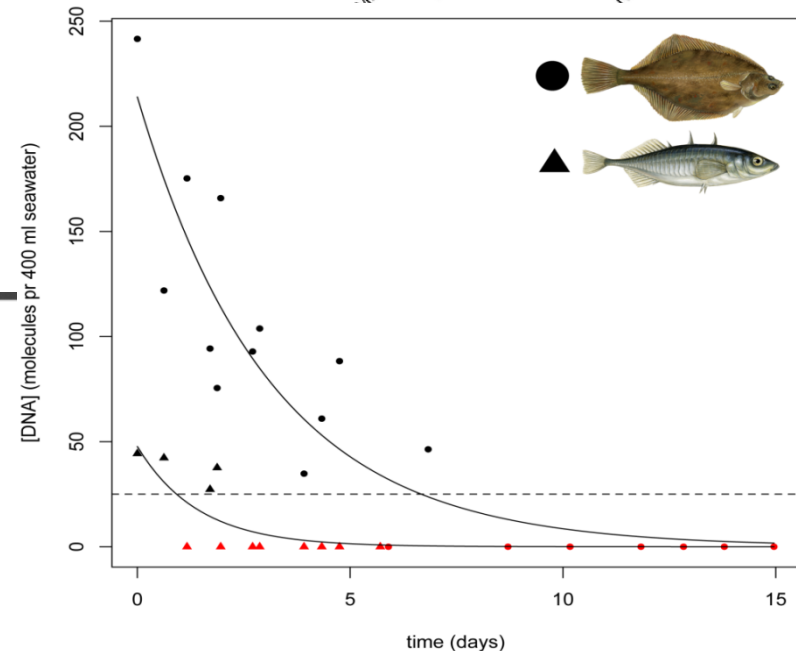
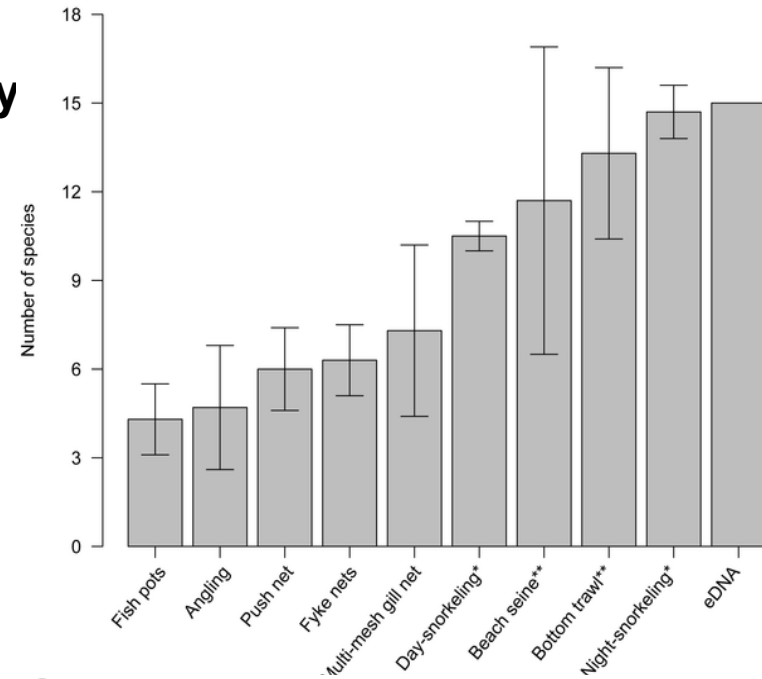
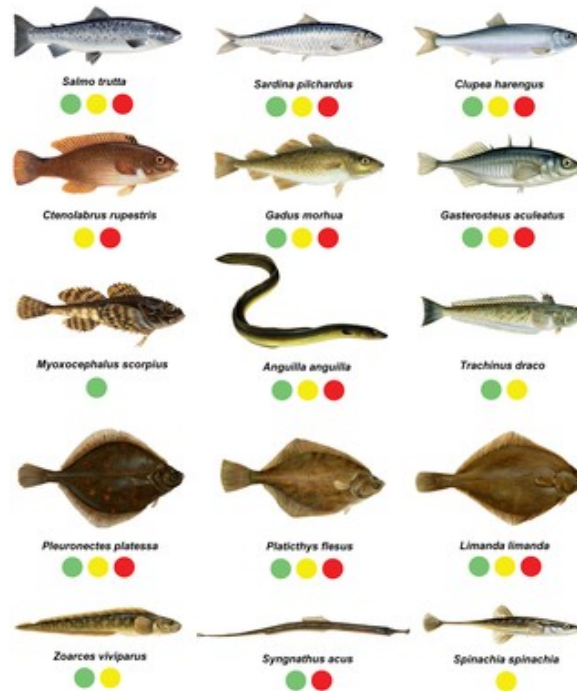


Figure 1. Monitoring mammals with leeches. (A) Survival of mtDNA in goat blood ingested by *Hirudo medicinalis* over time, relative to freshly drawn sample (100%, ca.  $2.4E+09$  mtDNA copies/gram blood). Mitochondrial DNA remained detectable in all fed leeches, with a minimum observed level at  $1.6E+04$  mtDNA/gram blood ingested. The line shows a simple exponential decay model,  $p < 0.001$ ,  $R^2 = 0.43$  (Supplemental information). (B) Vietnamese field site location and examples of mammals identified in *Hae madipsa* spp. leeches. From left to right: Annamite striped rabbit, small-toothed ferret-badger Truong Son muntjac (coat coloration and markings remain unknown), serow. Pictures do not reflect true size proportions. See also Supplemental information.

# Metabarcoding - příklady využití

Detekce ryb pomocí izolace eDNA z mořské vody  
-taky jedna z nejefektivnějších metod



OPEN ACCESS Freely available online

PLOS ONE

## Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples

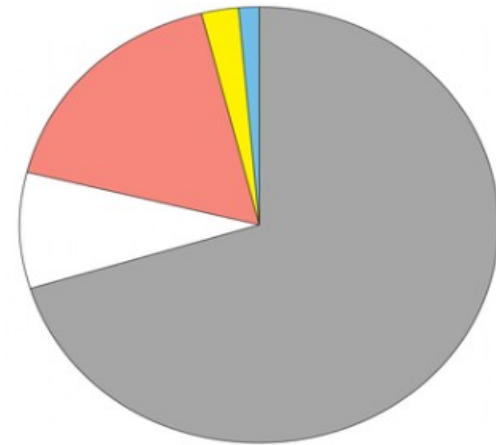
Philip Francis Thomsen<sup>1\*</sup>, Jos Kielgast<sup>1,3</sup>, Lars Lønsmann Iversen<sup>2</sup>, Peter Rask Møller<sup>3</sup>, Morten Rasmussen<sup>1</sup>, Eske Willerslev<sup>1\*</sup>

<sup>1</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade, Copenhagen, Denmark, <sup>2</sup>Freshwater Biology Section, Department of Biology, University of Copenhagen, Helsingørgade, Hillerød, Denmark, <sup>3</sup>Vertebrate Department, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken, Copenhagen, Denmark

# Metabarcoding - příklady využití

## Analýza potravy

Podíl hospodářských zvířat v potravě irbise je minimální



OPEN ACCESS Freely available online

PLoS one

## Prey Preference of Snow Leopard (*Panthera uncia*) in South Gobi, Mongolia

Wasim Shehzad<sup>1</sup>, Thomas Michael McCarthy<sup>2</sup>, Francois Pompanon<sup>1</sup>, Lkhagvajav Purevjav<sup>3</sup>, Eric Coissac<sup>1</sup>, Tiayyba Riaz<sup>1</sup>, Pierre Taberlet<sup>1\*</sup>

<sup>1</sup>Laboratoire d'Ecologie Alpine, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5553, Université Joseph Fourier, Grenoble, France, <sup>2</sup>Snow Leopard Program, Panthera, New York, New York, United States of America, <sup>3</sup>Snow Leopard Conservation Fund, Ulaanbaatar, Mongolia

Siberian ibex  
(*Capra sibirica*)

Domestic sheep  
(*Ovis aries*)

Argali sheep  
(*Ovis ammon*)

Chukar partridge  
(*Alectoris chukar*)

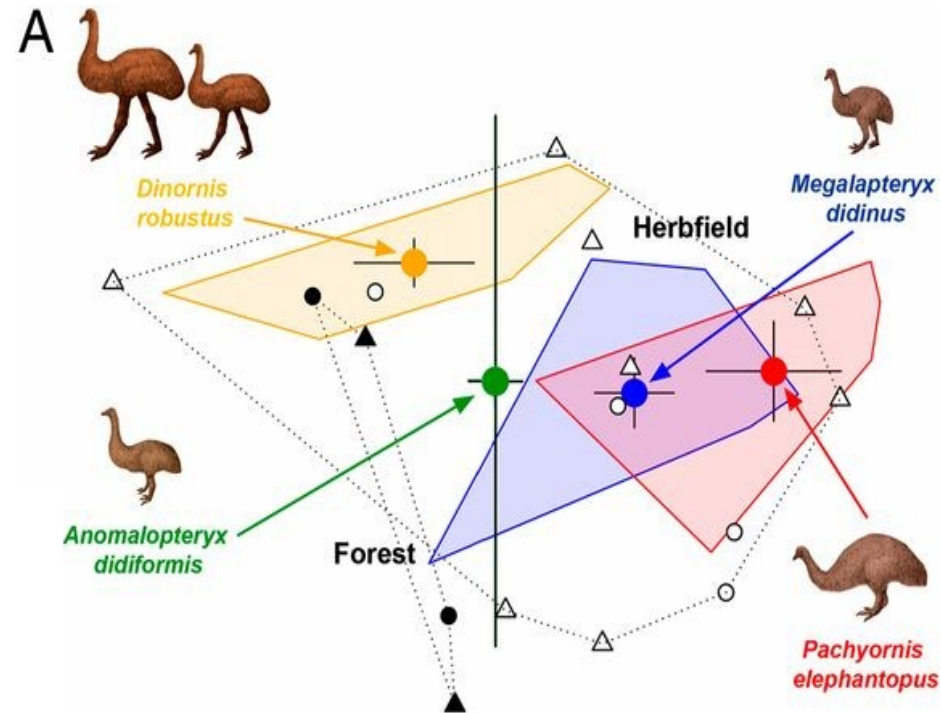
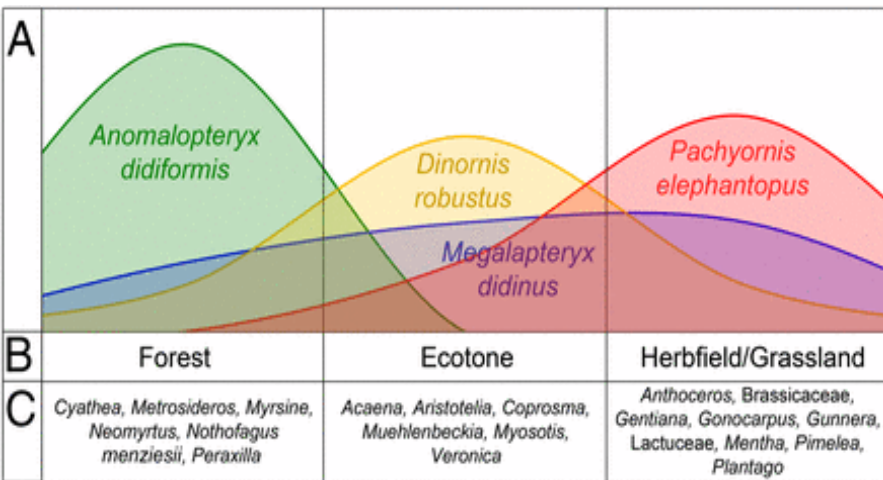
Domestic goat  
(*Capra hircus*)



# Metabarcoding - příklady využití

## Analýza složení společenstva na základě ancient DNA z koprolitů moa (Nový Zéland)

Umožňuje odhadnout typ prostředí které jednotlivé druhy obývaly a separaci ekologických nik



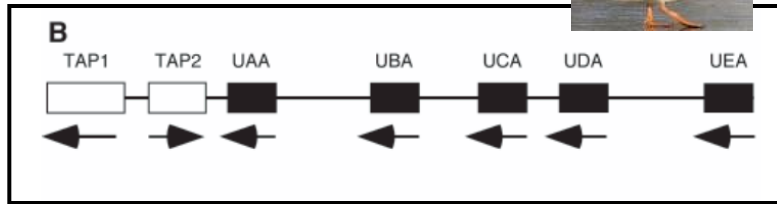
## Resolving lost herbivore community structure using coprolites of four sympatric moa species (Aves: Dinornithiformes)

Jamie R. Wood<sup>a,1</sup>, Janet M. Wilmshurst<sup>a</sup>, Sarah J. Richardson<sup>a</sup>, Nicolas J. Rawlence<sup>b,2</sup>, Steven J. Wagstaff<sup>a</sup>, Trevor H. Worthy<sup>a,3</sup>, and Alan Cooper<sup>b</sup>

<sup>a</sup>Landcare Research, Lincoln, Canterbury 7640, New Zealand; <sup>b</sup>Australian Centre for Ancient DNA, University of Adelaide, Adelaide, SA 5005, Australia;

# 3. Sekvenování amplikonů (PCR produktů)

## Genové duplikace

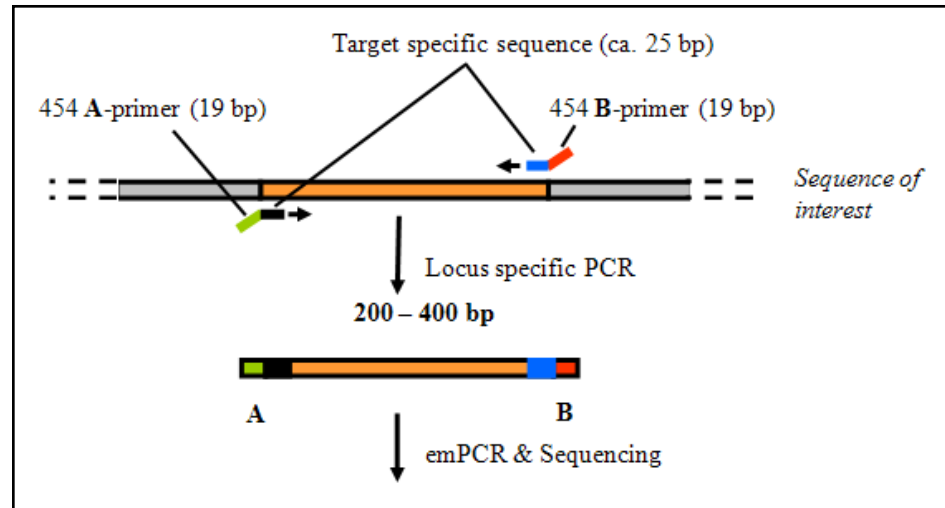


A-adaptor MID Target specific

Označí jedince

Amplifikuje všechny kopie MHC genů

Potřeba k HTS sekvenování

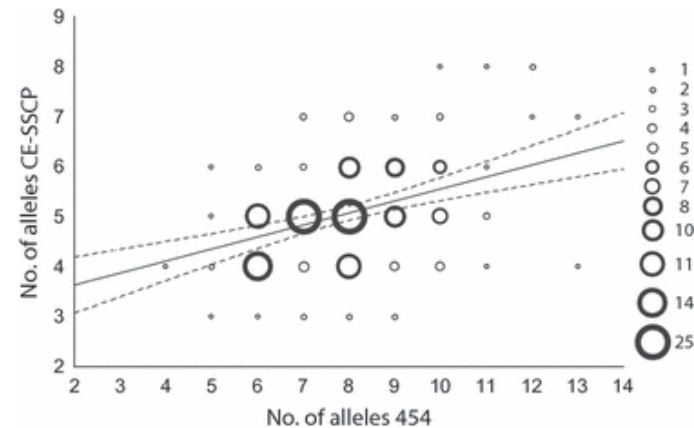
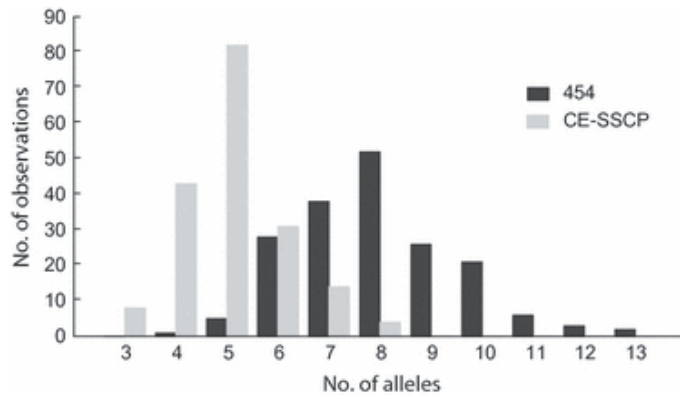




# Amplikonové sekvenování

## MHC u hýla rudého

- HTS má větší rozlišovací schopnost než SSCP + klonování



### MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2012) 12, 285–292

doi: 10.1111/j.1755-0998.2011.03082.x

## Evaluation of two approaches to genotyping major histocompatibility complex class I in a passerine—CE-SSCP and 454 pyrosequencing

MARTA PROMEROVÁ,\* WIESŁAW BABIK,† JOSEF BRYJA,\* TOMÁŠ ALBRECHT,\*‡ MICHAŁ STUGLIK† and JACEK RADWAŃŚ

## 4. Další aplikace - hledání nových genetických markerů

### Mikrosatelity

- sekvenování obohacených knihoven

### SNPs

- kompletní genomické sekvence pro hledání diagnostických SNPs
- např. RAD-sequencing

# Hledání nových genetických markerů - mikrosatelity

## Obvyklý postup:

- Obohacení genomické knihovny o mikrosatelitové motivy – sequence capture
- Sekvenování obohacených knihoven
- Detekce mikrosatelitů a návržení vhodných primerů

### MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2011) 11, 638–644

doi: 10.1111/j.1755-0998.2011.0295

## High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries

THIBAUT MALAUSA,\* ANDRÉ GILLES,† EMESE MEGLÉCZ,† HÉLÈNE BLANQUART,‡ STÉPHANIE DUTHOY,‡ CAROLINE COSTEDOAT,† VINCENT DUBUT,† NICOLAS PECH,† PHILIPPE CASTAGNONE-SERENO,\* CHRISTOPHE DÉLYE,§ NICOLAS FEAU,¶ PASCAL FREY,\*\* PHILIPPE GAUTHIER,†† THOMAS GUILLEMAUD,\* LAURENT HAZARD,\*‡ VALÉRIE LE CORRE,§ BRIGITTE LUNG-ESCARDANT,¶ PIERRE-JEAN G. MALÉ,§§ STÉPHANIE FERREIRA‡ and JEAN-FRANÇOIS MARTIN††

\*INRA, UMR 1301 IBSV INRA/INSA/CNRS, 400 Route des Chappes, BP 167, 06903 Sophia-Antipolis Cedex, France, †Aix-Marseille Université, CNRS, IRD, UMR 6116 – IMEP, Equipe Evolution Génome Environnement, Centre Saint-Charles, Case 31 3 Place Victor Hugo, 13331 Marseille Cedex 3, France, ‡Genoscreen, Genomic Platform and R&D, Campus de l'Institut Pasteur, rue du Professeur Calmette, Bâtiment Guérin, 59000 Lille, France, §INRA, UMR 1210 Biologie et Gestion des Adventices, 17 rue Sully, 21000 Dijon, France, ¶INRA, UMR 1202 BIOGECO, Equipe de Pathologie Forestière, Domaine de Pierroton, 69 route d'Arcachon, 33612 Cestas Cedex, France, \*\*INRA, Nancy-Université, UMR 1136, Interactions Arbres – Microorganismes, IFR 1: 54280 Champenoux, France, ††UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, C: 30016, 34988 Montpellier-sur-Lez Cedex, France, ‡‡INRA – UMR 1248 AGIR, BP 52627, 31326 Castanet-Tolosan Cedex, France §§UMR Evolution et Diversité Biologique (Université Toulouse III; CNRS), 118 Route de Narbonne, 31062 Toulouse, France



allgenetics

HOME

COMPANY

SERVICES ▾

HOME » SERVICES » Microsatellite Development

### Experts in Microsatellite Development

Microsatellites (also known as short tandem repeats) are repetitive DNA elements usually found in non-coding regions of the genome. They have high mutation rates, and therefore are frequently highly polymorphic. Variations in the number of repetitions generate different alleles. This makes them appropriate molecular markers for population genetics and molecular ecology projects.

## We develop microsatellite markers for your study species

At AllGenetics, we use next-generation sequencing to obtain primer pairs which amplify polymorphic microsatellite loci in your study species. Genomic DNA is used to generate genomic libraries. We usually enrich these libraries with 4 to 6 different microsatellite motifs. However, we can customise the number of motifs to your needs. We obtain thousands of microsatellite-containing reads by using high-throughput sequencing. Our bioinformaticians then filter these reads for primer design. The primers obtained are multiplexed and tested for polymorphism in a number of individuals from different populations.

## How we work

High quality DNA at a concentration of 100 ng/μL in a minimum volume of 50 μL from a number of individuals is required. Alternatively, we can isolate DNA from your samples. These samples should be adequately preserved to ensure DNA integrity. We will deliver tested primer pairs which amplify polymorphic loci for your study species. A detailed methodological report and all sequencing reads generated will also be provided.

Our microsatellite development projects are divided into four steps. For your convenience, we can carry out the entire project or only the parts you need.

OPEN ACCESS Freely available online

PLOS ONE

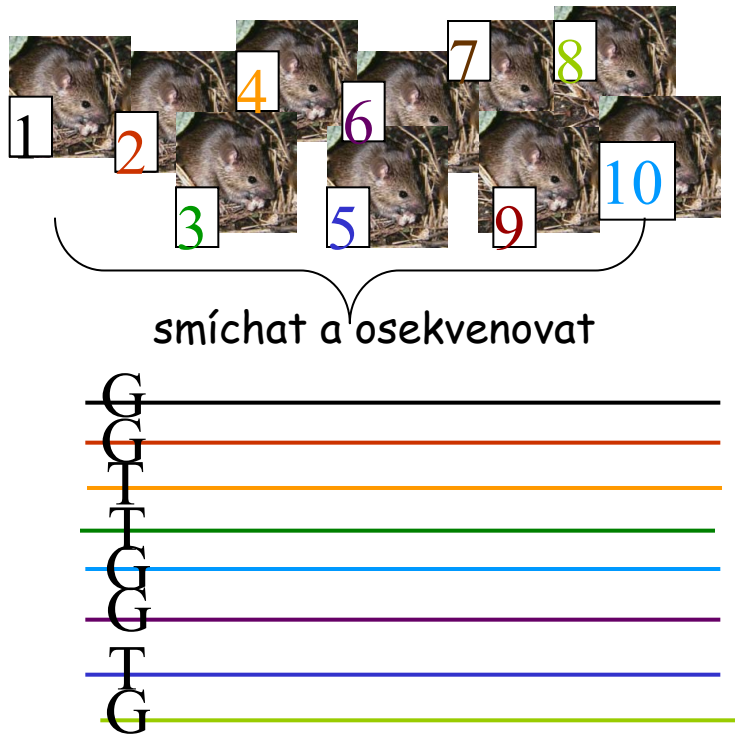
## 32 species validation of a new Illumina paired-end approach for the development of microsatellites

Stacey L. Lance<sup>1</sup>, Cara N. Love<sup>1</sup>, Schyler O. Nunziata<sup>1</sup>, Jason R. O'Bryhim<sup>1</sup>, David E. Scott<sup>1</sup>, R. Wesley Flynn<sup>1</sup>, Kenneth L. Jones<sup>2</sup>

<sup>1</sup> Savannah River Ecology Laboratory, University of Georgia, Aiken, South Carolina, United States of America, <sup>2</sup> Department of Biochemistry and Molecular Genetics, University of Colorado, Fort Collins, Aurora, Colorado, United States of America



# Hledání diagnostických SNP (např. pro studium hybridizace)



10 jedinců



10 jedinců



# Hledání nových SNPs - RAD-sequencing

Sekvenování podél restričních míst

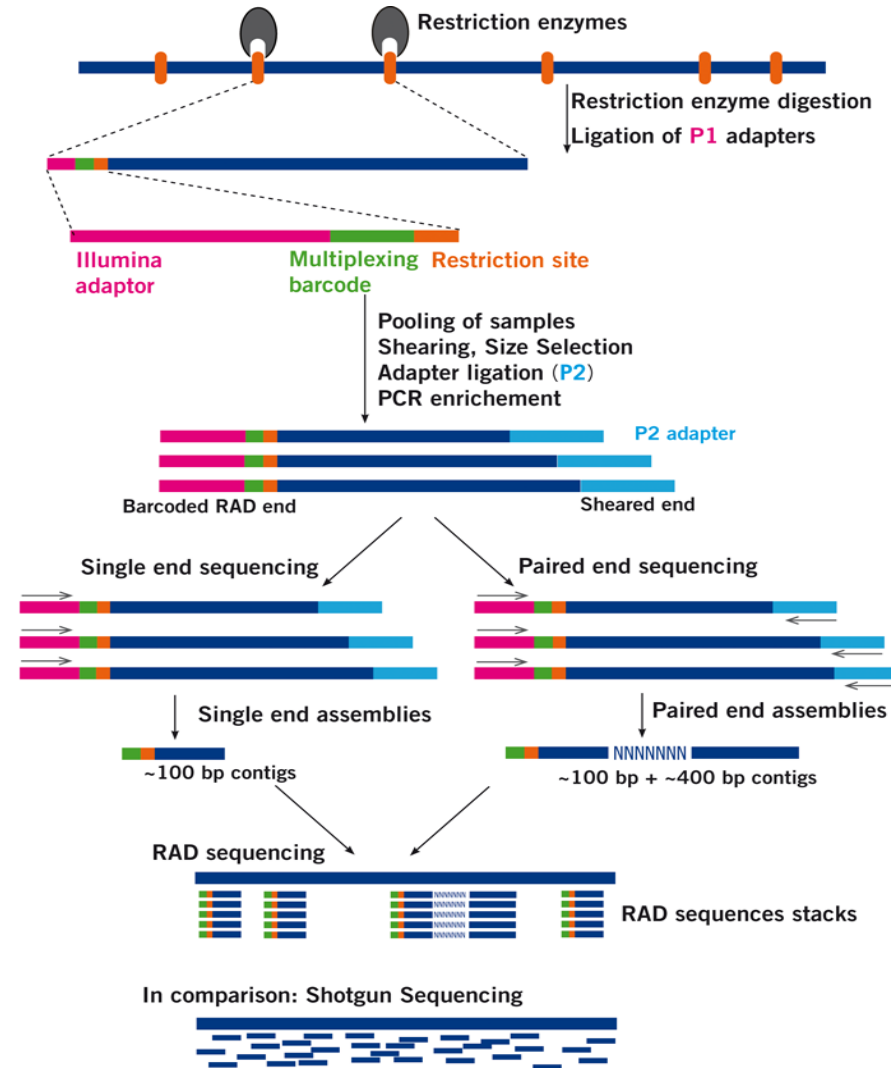
Fragmetace gelogenomové DNA po mocí restričních enzymů

Ligace sekvenačních adaptorů na výsledné fragmenty

Následná sekvenace podél restričních míst

Celogenomové scany genetické variability

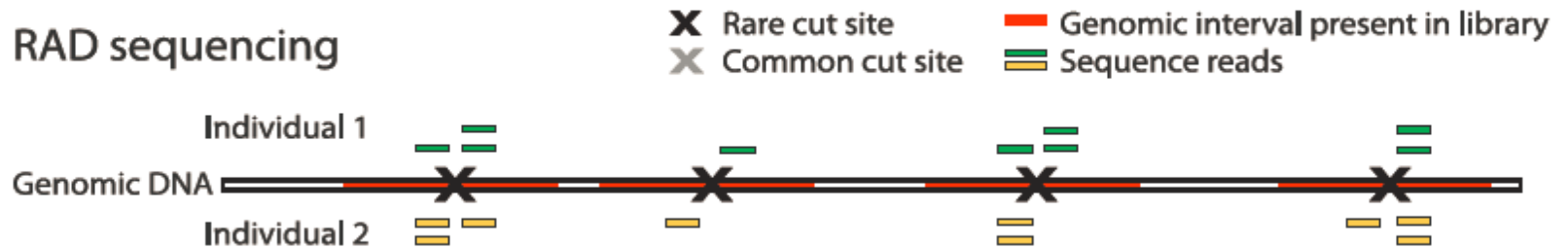
*Hledání SNPs, populační genomika (např. RAD-SEQ) apod.*



# RAD vs. ddRAD

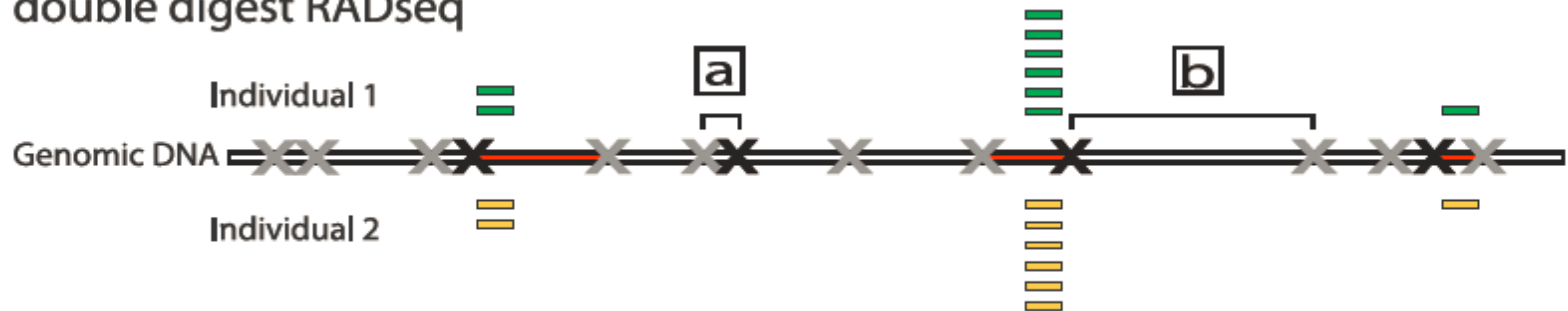
A

RAD sequencing



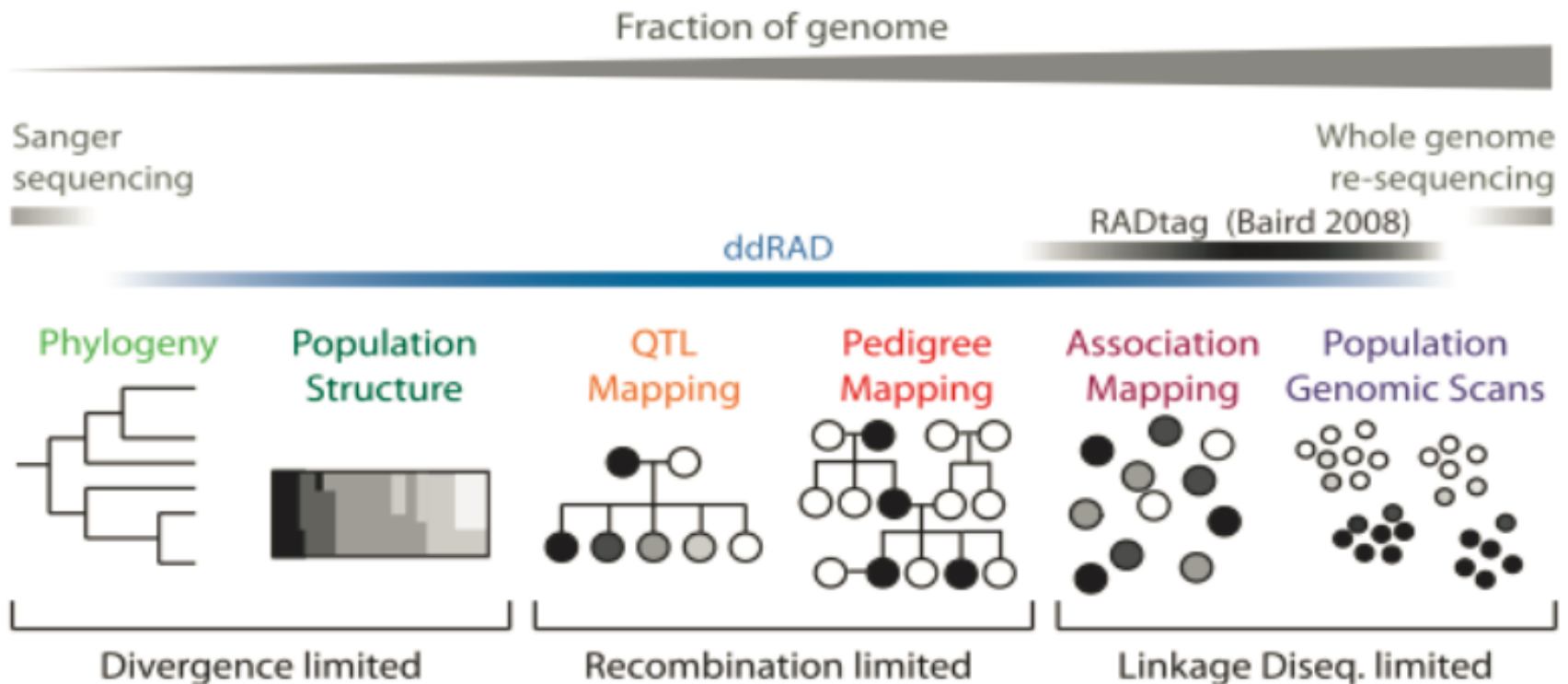
B

double digest RADseq

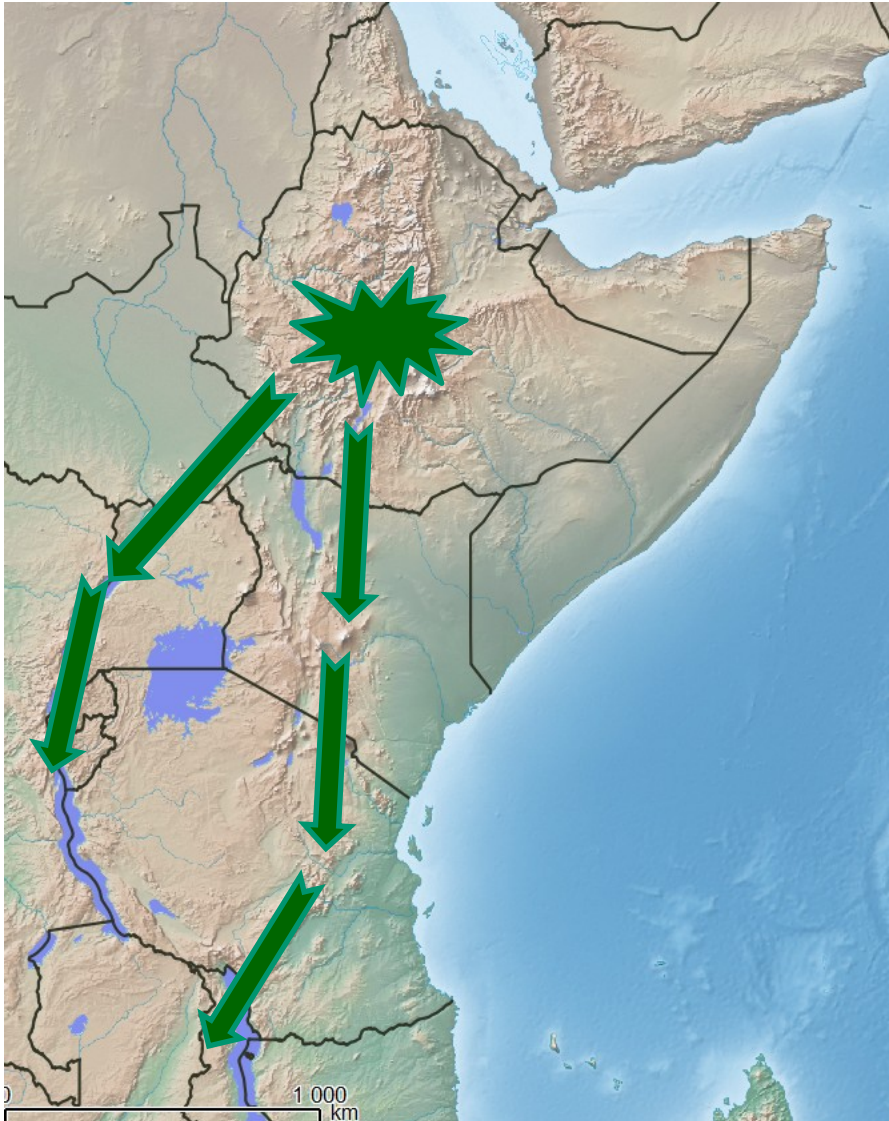




# Sekvenování podél restričních míst



# Phylogenomics of *Lophuromys*



- ancestral lineage „trapped“ in Ethiopian highlands, where diversified and sourced the colonization of other mountains (mostly in Pleistocene)
- *Lophuromys flavopunctatus* complex (9 Ethiopian species)



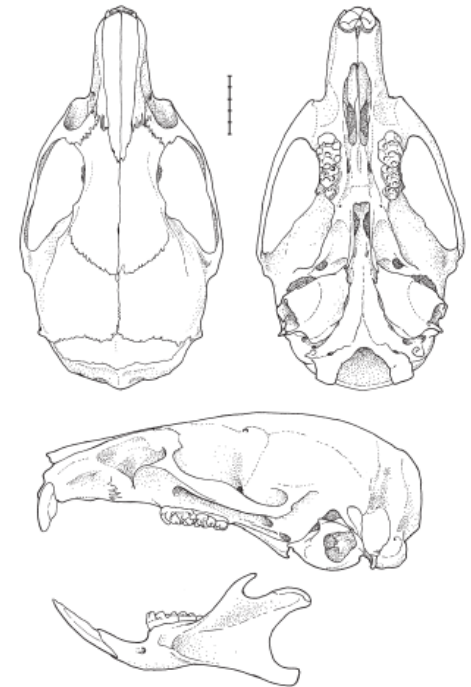
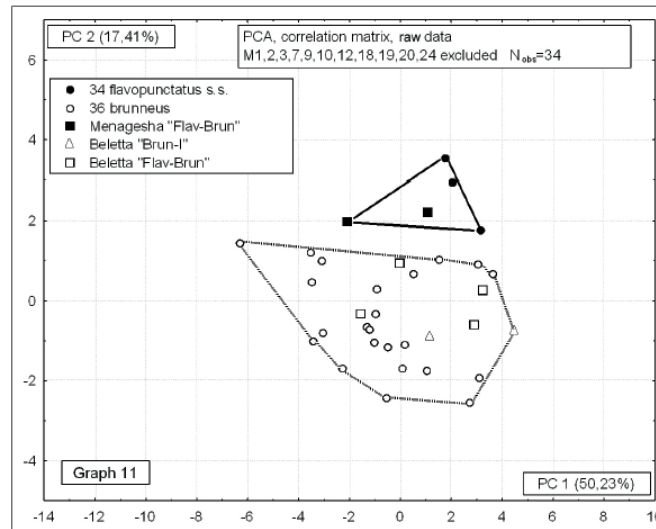
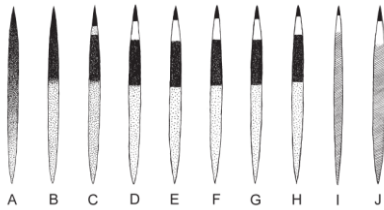
# 9 endemic species in Ethiopia

BULLETIN DE L'INSTITUT ROYAL DES SCIENCES NATURELLES DE BELGIQUE  
BULLETIN VAN HET KONINKLIJK BELGISCH INSTITUUT VOOR NATUURWETENSCHAPPEN

BIOLOGIE, 77: 77-117, 2007  
BIOLOGIE, 77: 77-117, 2007

Morphometric and genetic study of Ethiopian *Lophuromys flavopunctatus* THOMAS, 1888 species complex with description of three new 70-chromosomal species (Muridae, Rodentia)

by Leonid A. LAVRENTCHENKO, Walter N. VERHEYEN, Erik VERHEYEN, Jan HULSELMANS & Herwig LEIRS



3.2. Views of skull and mandible of *Lophuromys menageshae* n.sp. (ZMMU S-165969, holotype). Scale bar = 5 mm.

# *Lophuromys* - questions

- Are there really 9 well delimited species?
- Are they easily (genetically) recognizable? (e.g. mtDNA-barcoding)
- What is their distribution and ecological requirements? -> IUCN assessment, etc.



ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

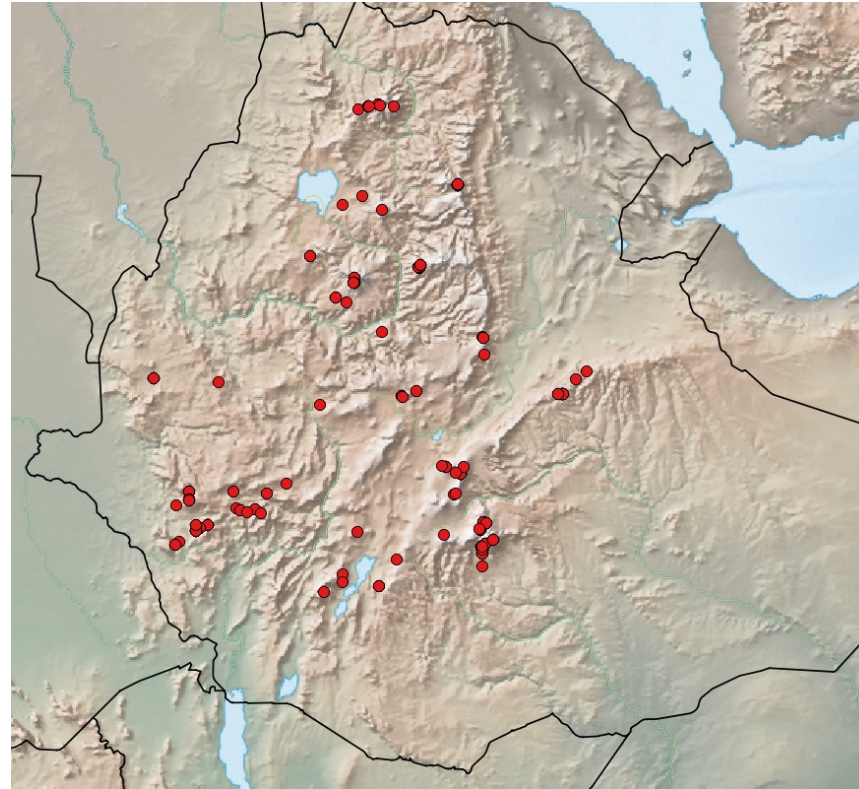
Complex reticulate evolution of speckled brush-furred rats (*Lophuromys*) in the Ethiopian centre of endemism

Valeria A. Komarova<sup>1</sup> | Danila S. Kostin<sup>1</sup> | Josef Bryja<sup>2,3</sup> | Ondřej Mikula<sup>2</sup> | Anna Bryjová<sup>2</sup> | Dagmar Čížková<sup>2</sup> | Radim Šumbera<sup>4</sup> | Yonas Meheretu<sup>5</sup> | Leonid A. Lavrenchenko<sup>1</sup>



# Material and Methods

- cca 500 specimens from all major mountain ranges
- mtDNA marker (CYTB)
- 4 nuclear markers (2 introny + 2 exony)
- **genomic approach - ddRAD sequencing**



# Retaining well-covered & informative loci

## All loci

No. of individuals:	213
No. of loci:	80570
No. of informative loci:	69724
No. of SNPs / PISs per informative locus:	
Min:	1 / 1
25%:	5 / 4
50%:	10 / 9
75%:	20 / 17
Max:	60 / 57
Loci per individual:	
Min:	5178
25%:	9719
50%:	12000
75%:	14607
Max:	23205
Individuals per locus:	
Min:	4
25%:	6
50%:	13
75%:	37
Max:	208
Proportion of missing data:	0.85

## HQ loci

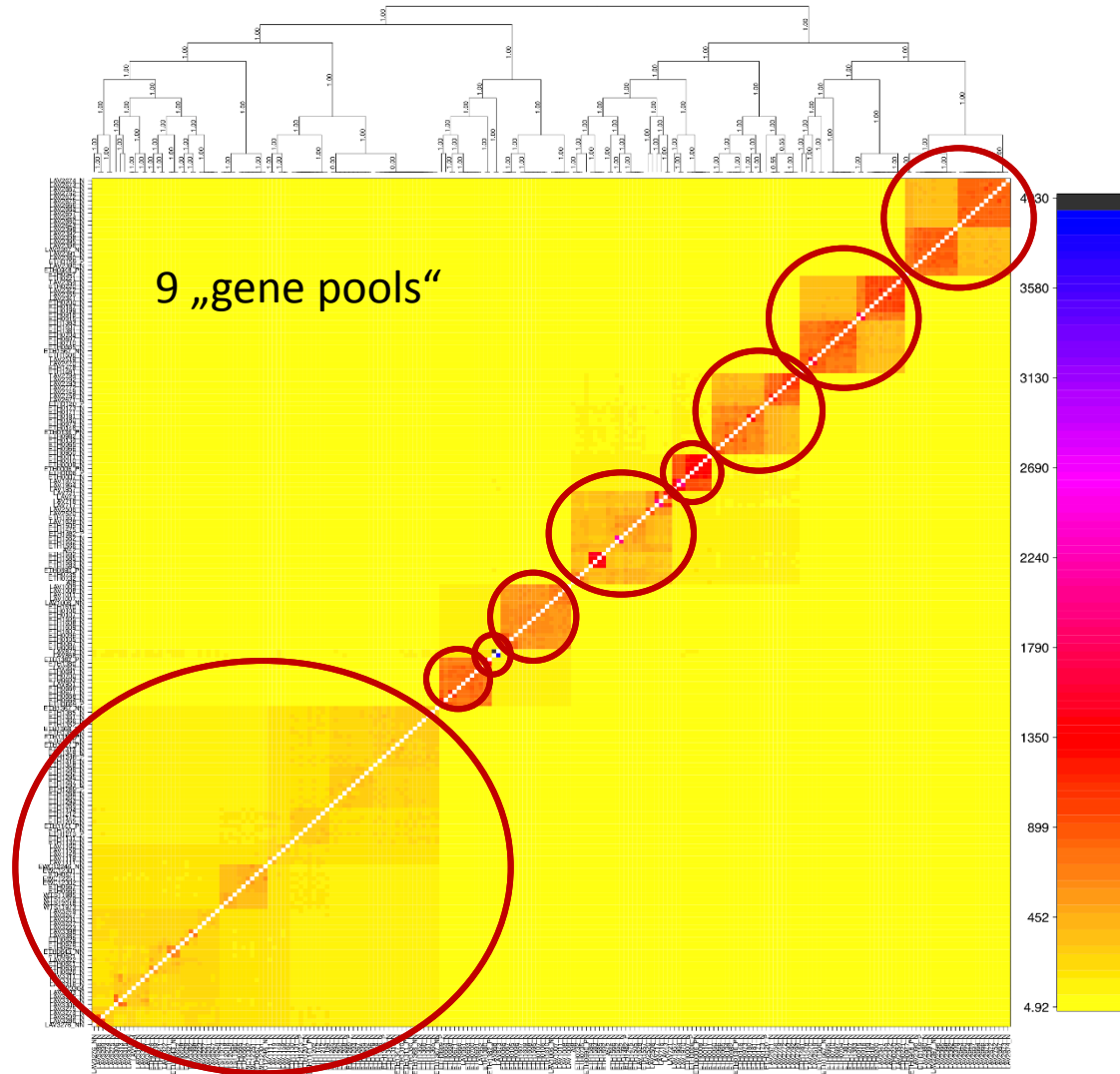
No. of individuals:	213
No. of loci:	15164
No. of informative loci:	15164
No. of SNPs / PISs per informative locus:	
Min:	1 / 1
25%:	17 / 14
50%:	25 / 21
75%:	32 / 28
Max:	57 / 54
Loci per individual:	
Min:	3393
25%:	6912
50%:	8074
75%:	9297
Max:	11912
Individuals per locus:	
Min:	54
25%:	74
50%:	103 ✓
75%:	149
Max:	208
Proportion of missing data:	0.47 ✓

80 570 loci → filtering → 15 164 loci

# ddRADseq: co-ancestry matrix

209  
individuals

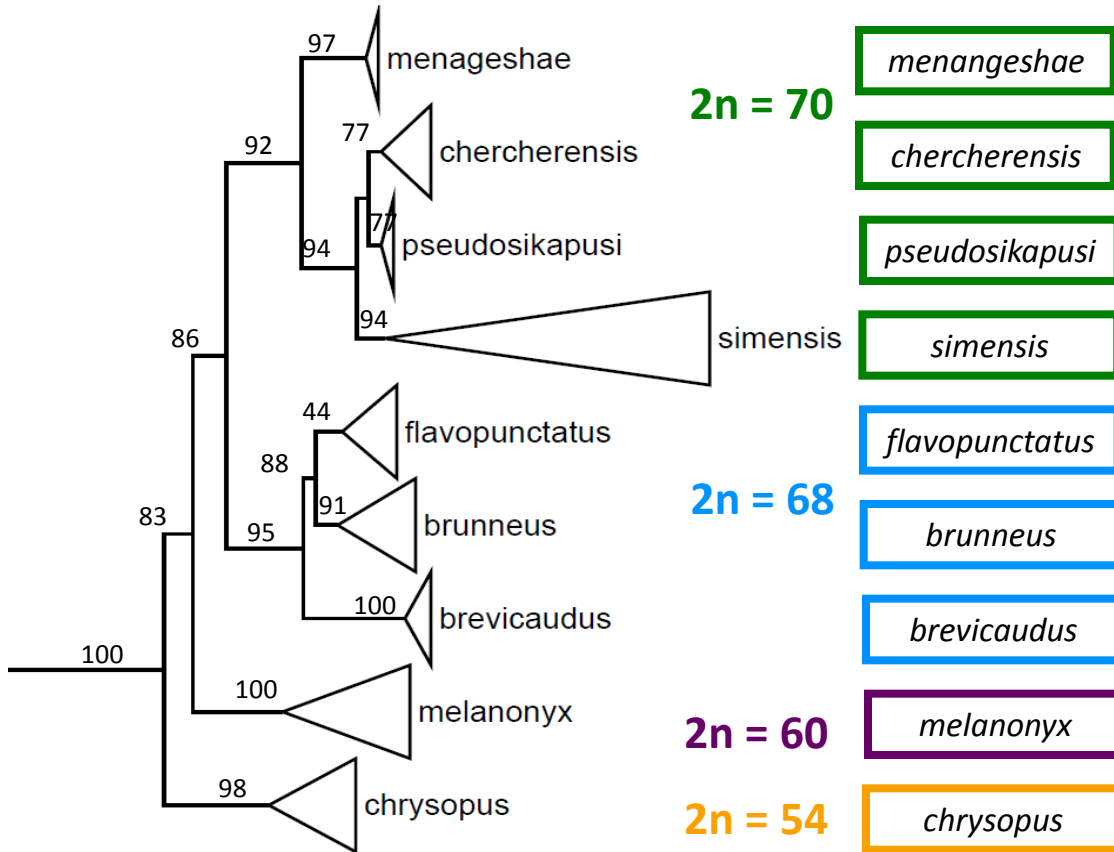
15 623  
informative  
loci





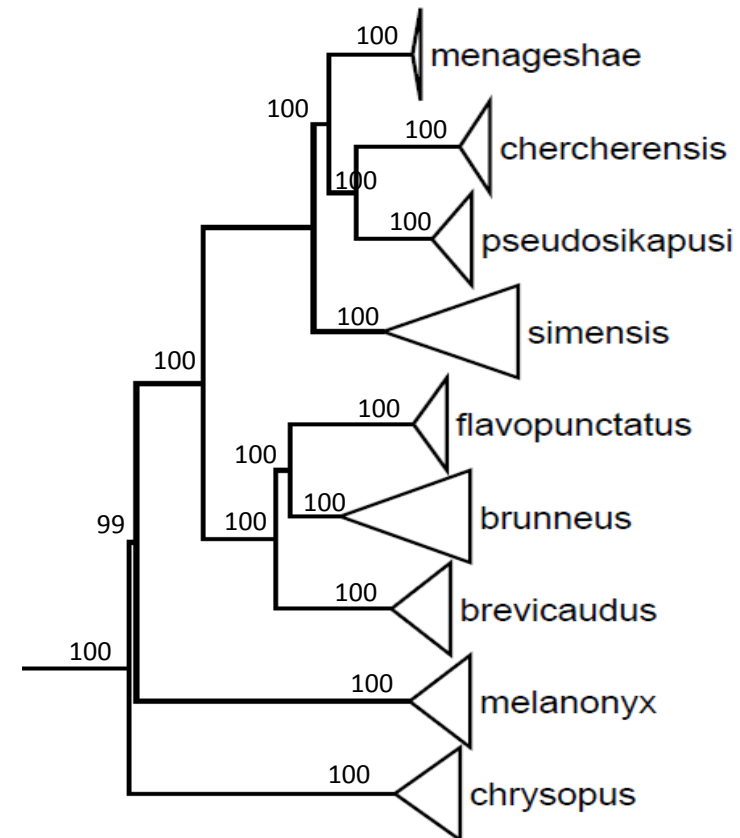
# Maximum likelihood analysis of concatenated nuclear dataset

## Sanger sequencing



4 nuclear markers (V. Komarova et al.)  
(2 604 bp concatenated dataset)

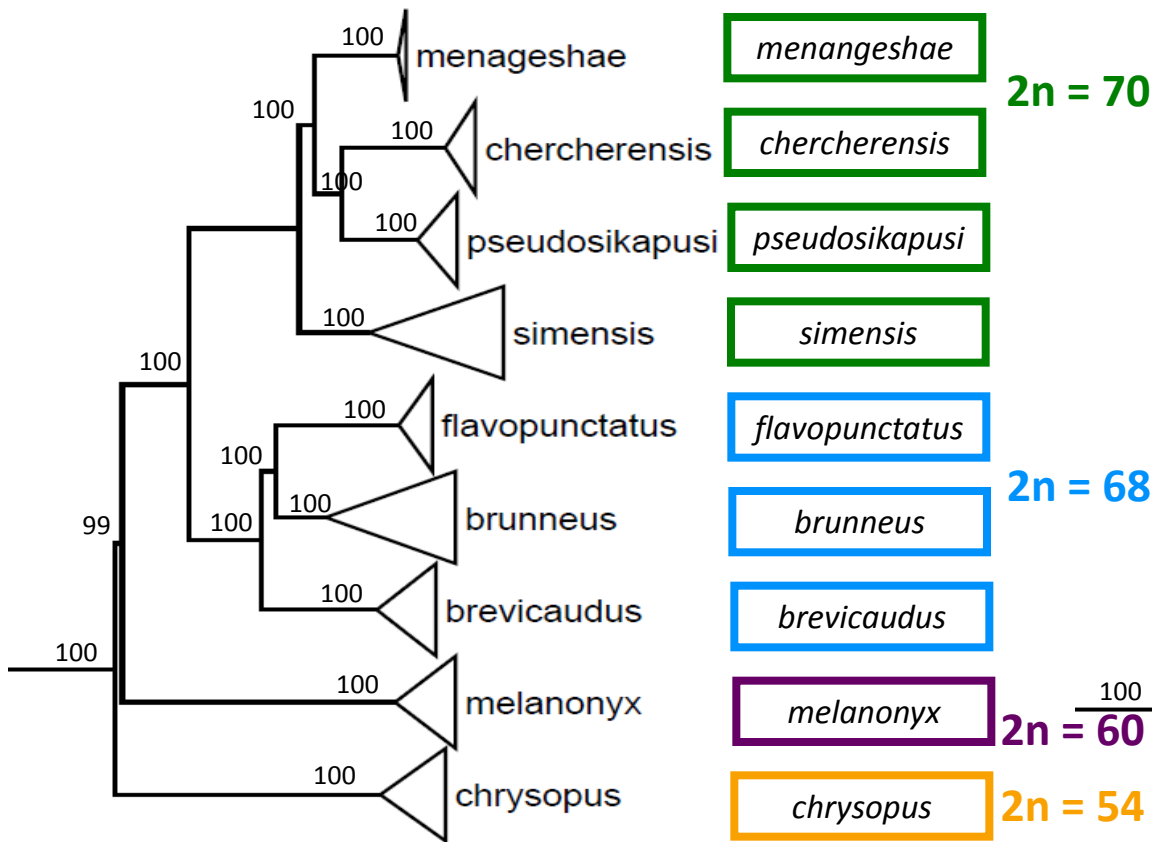
## ddRADseq



15 623 informative loci

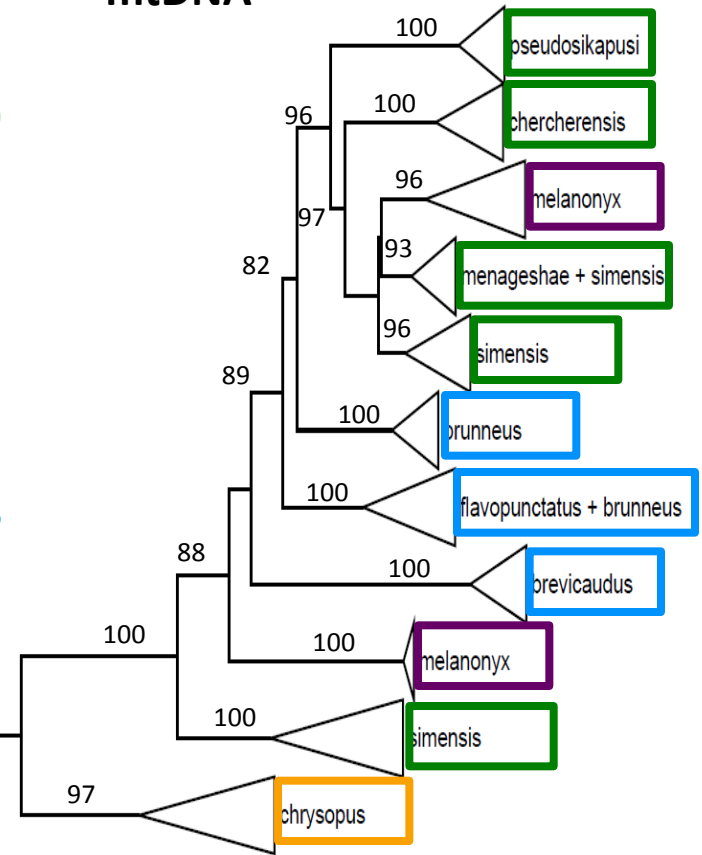
# And what about mtDNA?

ddRADseq



15 623 informative loci

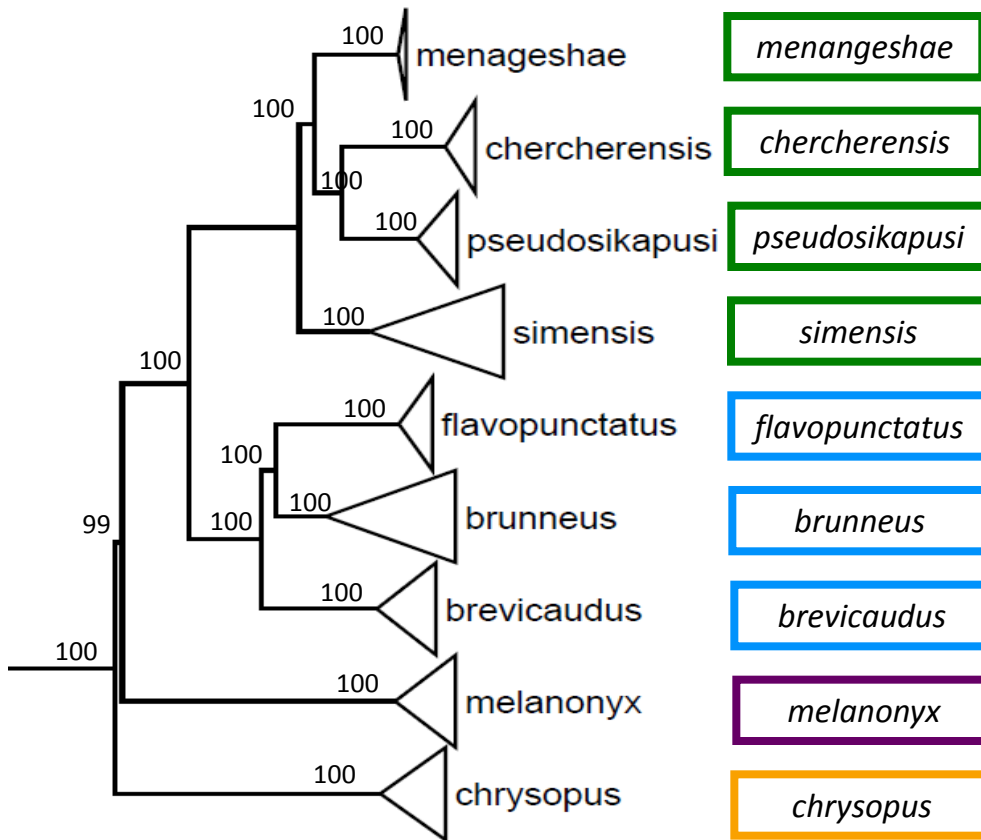
mtDNA



cytochrome *b* (1140 bp)

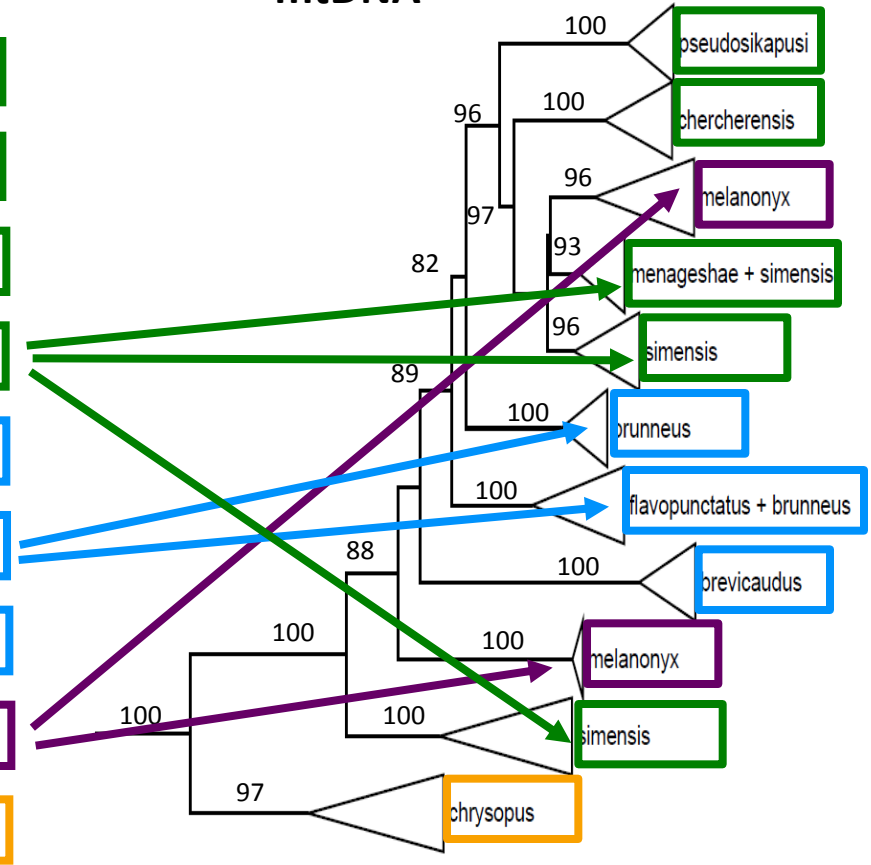
# And what about mtDNA?

ddRADseq



15 623 informative loci

mtDNA



cytochrome *b* (1140 bp)

„reticulate evolution“