



# Budoucnost genetických metod v ekologickém výzkumu

## 1. Nové postupy při sekvenování DNA („genomics“)

Molecular Ecology Resources (2008) 8, 3–17

doi: 10.1111/j.1471-8286.2007.02019.x

TECHNICAL REVIEW

Sequencing breakthroughs for genomic ecology and  
evolutionary biology

MATTHEW E. HUDSON

*Department of Crop Sciences, University of Illinois, Urbana, 334 NSRC, 1101 W. Peabody Blvd., IL 61801, USA*

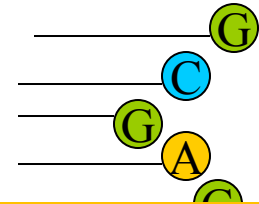
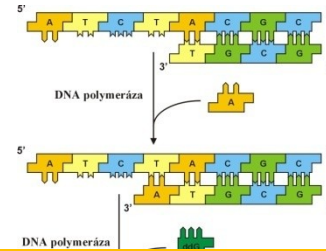


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

# Sequencing - Sangerova metoda

DNA

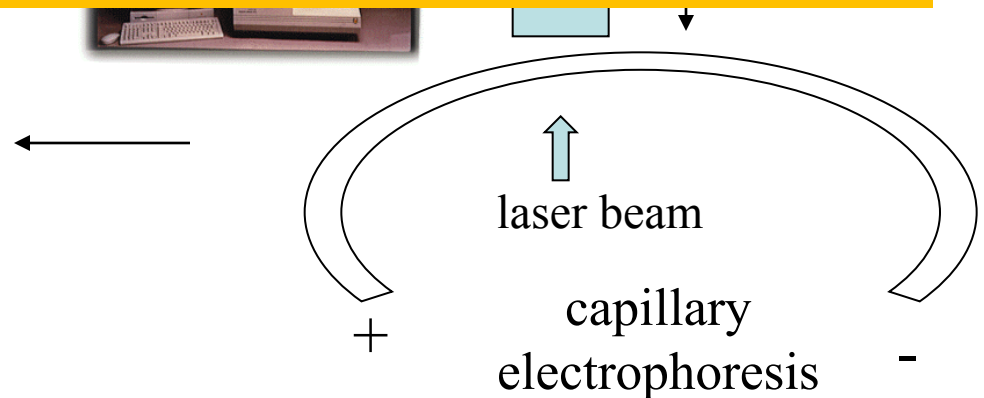
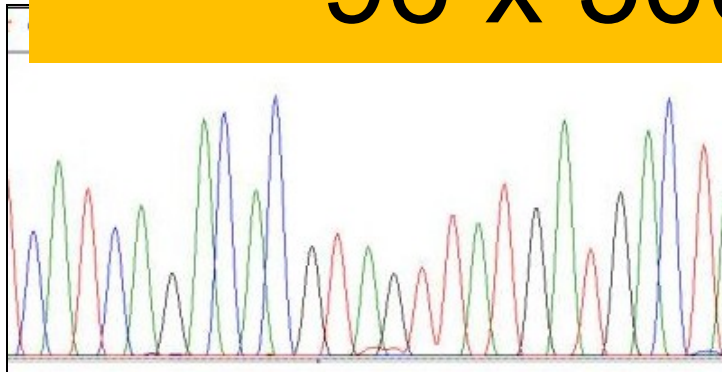
→ PCR product



4-kapilární sekvenátor

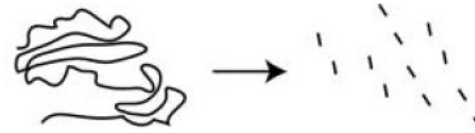
=

96 x 500 bp/12 hodin

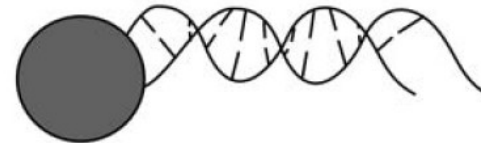


# „Next generation sequencing“

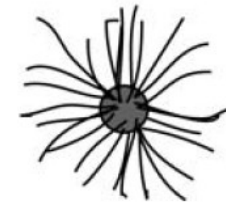
1) Randomly fragment many molecules of target DNA



2) Immobilize individual DNA molecules on solid support

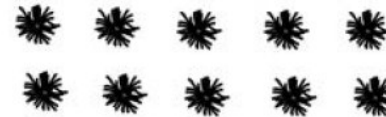


3) Amplify DNA in clonal 'polymerase colony'



„polonies“  
(polymerase colonies)

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



... commercially available since August 2007

# 454 pyrosequencing

- emulzní techniky amplifikace pikolitrové objemy
- simultánní sekvenování na destičce z optických vláken detekce pyrofosfátů uvolňovaných inkorporací bází
- První generace GS20 → 200 000 reakcí najednou (zhruba 20 milionů bp) dnes FLX → 400 000 reakcí najednou = eukaryotní genom za týden!!!
- Délka jednotlivých sekvencí 100 – 400



Molecular Ecology (2008) 17, 1629–1635

NEWS AND VIEWS

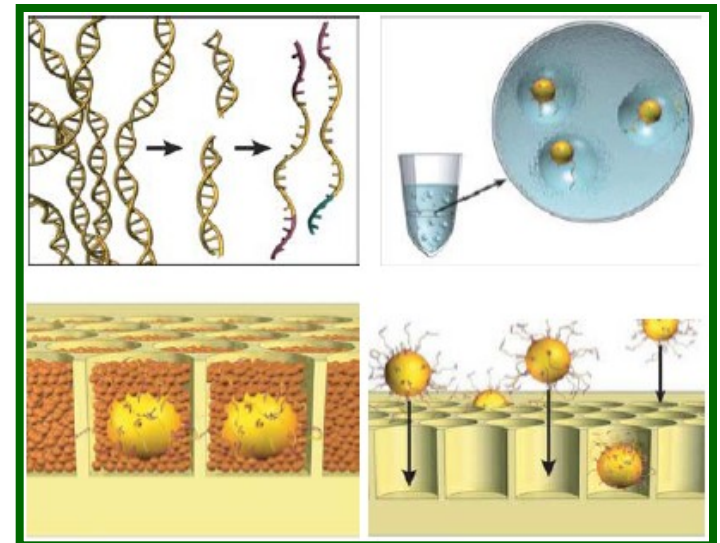
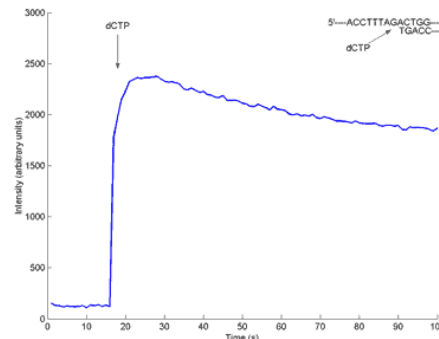
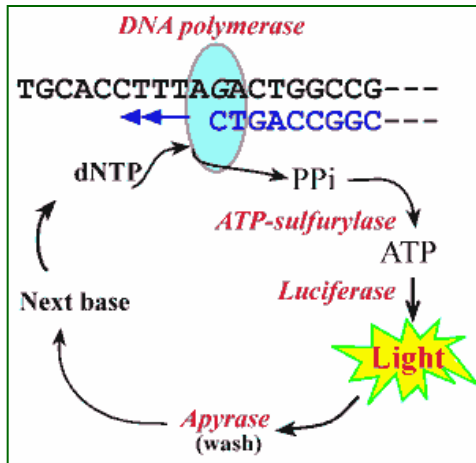
PERSPECTIVE

Sequencing goes 454 and takes large-scale genomics into the wild

HANS ELLEGREN

Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, SE-75236 Uppsala, Sweden

1 600 000 well plate



# Pracovní postup



1

## DNA Library Preparation

1. DNA Fragmentation (Nebulization)
2. DNA Fragment Size Selection
3. DNA Sample Quality Assessment (Nebulized or *LMW* DNA Sample)
4. Fragment End Polishing
5. Adaptor Ligation
6. Small Fragment Removal
7. Library Immobilization
8. Fill-In Reaction
9. Single-Stranded DNA Library Isolation
10. DNA Library Quality Assessment and Quantitation

Time: 11 - 72 h

General Laboratory 1



2

## Emulsion-Based Clonal Amplification (emPCR)

1. Preparation of the Live and Mock Amplification Mixes
2. DNA Library Capture
3. Emulsification
4. Amplification
5. Bead Recovery
6. DNA Library Bead Enrichment
7. Sequencing Primer Annealing

Time: 11 - 13 h

Controlled Room

Amplicon Room



3

## Sequencing / Genome Sequencer FLX Operation

1. The Pre-Wash
2. PicoTiterPlate Device Preparation
3. The Sequencing Run

Time: 11.5 h

General Laboratory 2



4

## Data Processing and Analysis

1. Data Processing
  - a) Image Processing
  - b) Signal Processing
2. Data Analysis
  - a) Assembly
  - b) Mapping
  - c) Amplicon Variant Analysis

Time: variable



# 1. Příprava jednořetězcové DNA knihovny (ssDNA library preparation)

## 1 DNA Fragmentation (Nebulization):



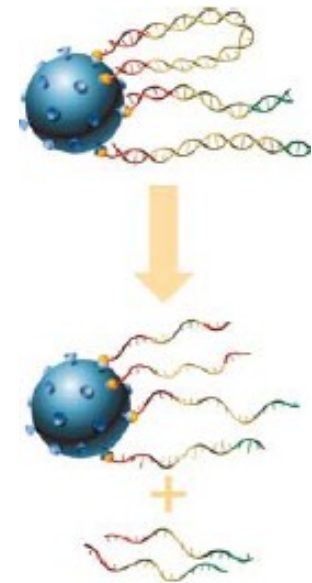
## 5 Adaptor Ligation:



## 7 Library Immobilization:



## 9 ssDNA Library Isolation:



### Adaptor A + Adaptor B

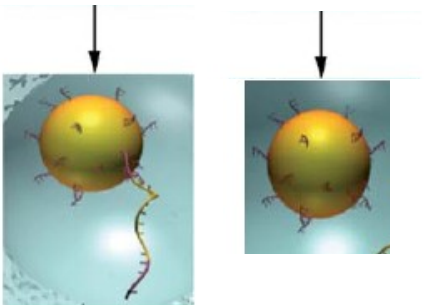
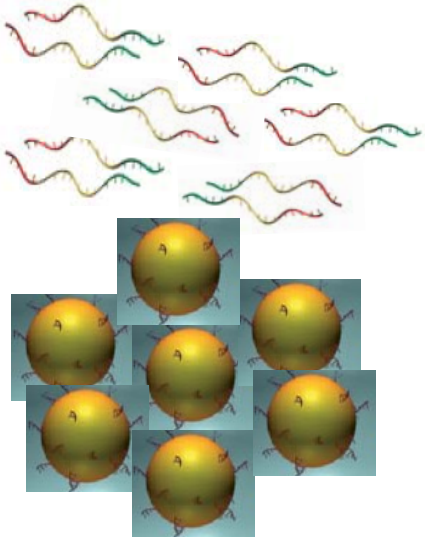
-Slouží jako vazebné místo primerů pro následnou PCR amplifikaci a sekvenování

-Slouží k uchycení na kuličky (na adaptor B je připojen **biotin**)

## 2. Namnožení každé jednotlivé molekuly pomocí emulzní PCR (emPCR)

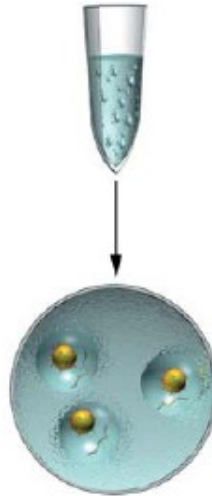
### 1 DNA Library Capture:

- poměry nastavit tak aby  
1 kulička  $\leq$  1 molekula DNA

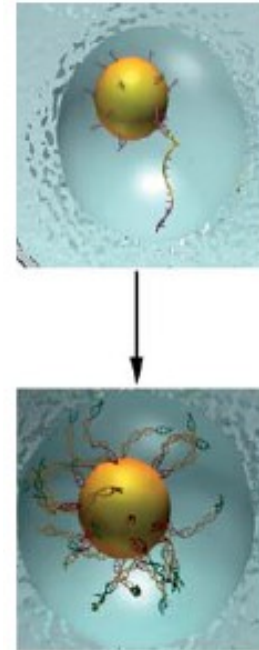


### 2 Preparation of the Amplific. Mixes

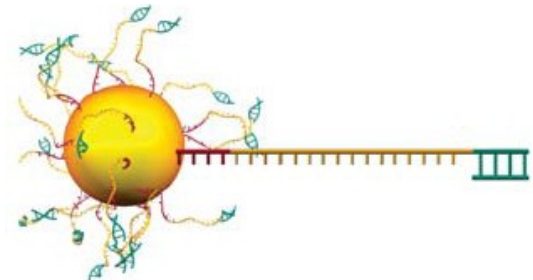
### 3 Emulsification:



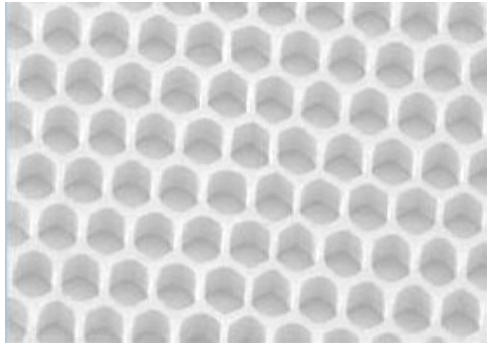
### 4 emPCR Amplification:



### 7 Sequencing Primer Annealing:

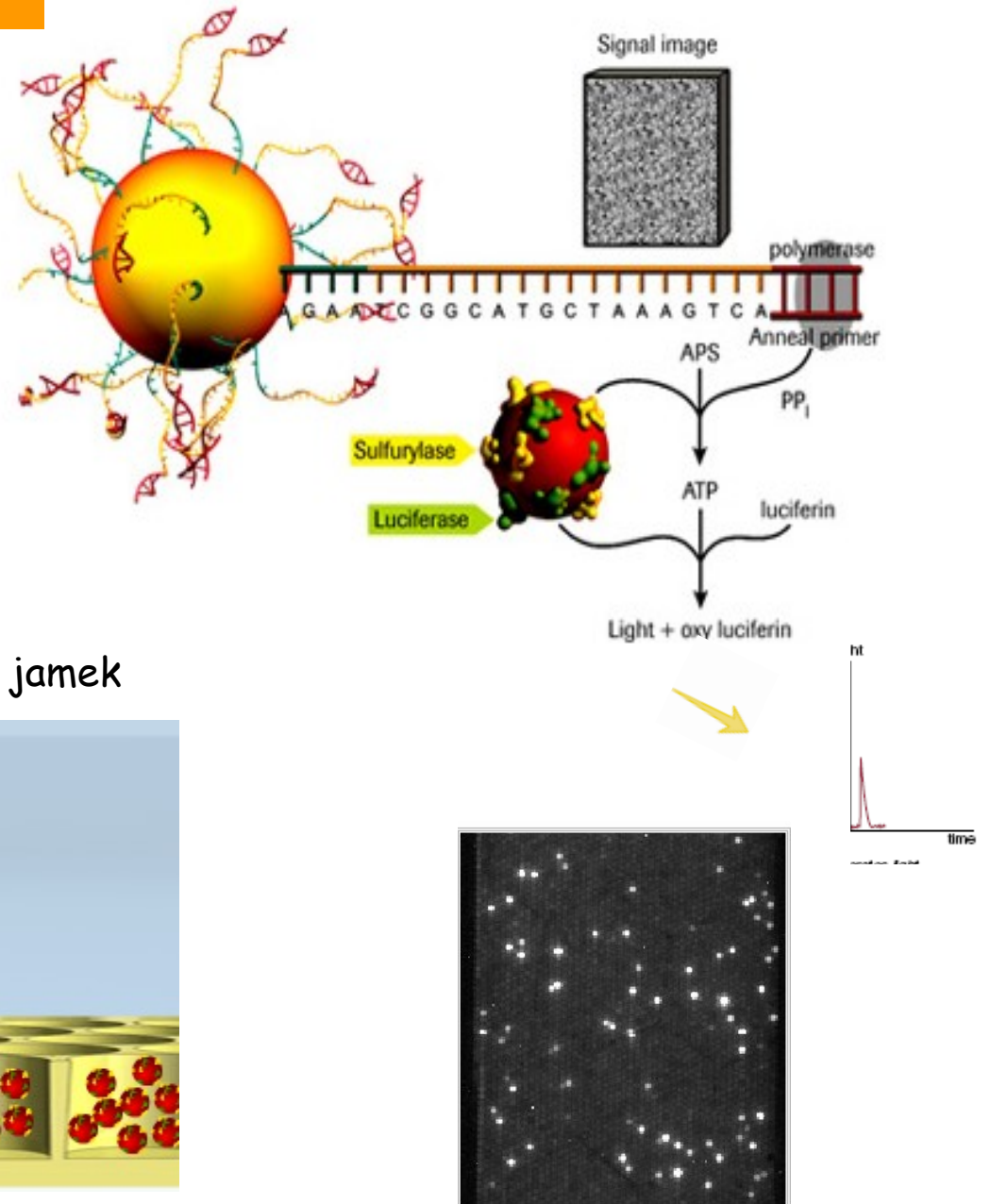
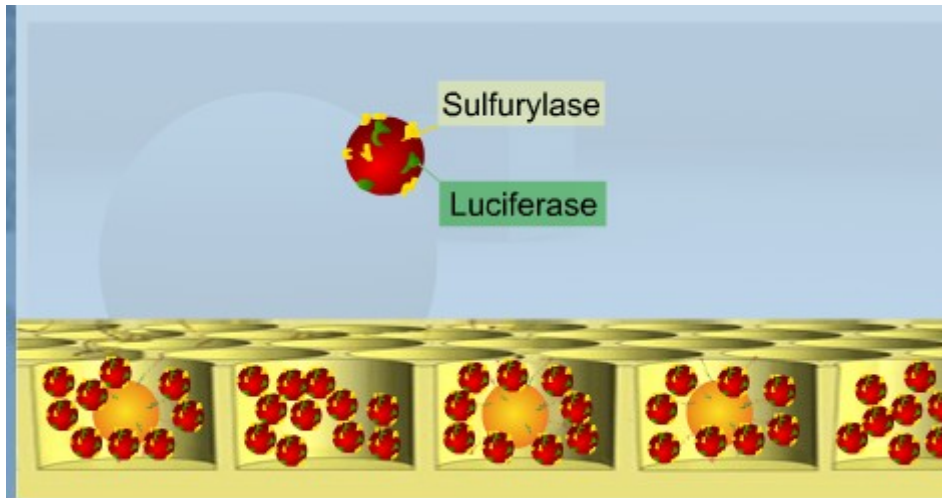


### 3. Pyrosekvenovani



pikotitrační destička

Na jedné destičce 400 000 až 1milión jamek





### 3. Pyrosekvenovani - detekce signálu

- postupně se přidávají nukleotidy v definovaném pořadí: např. TACG TACG TACG
- po přidání každého nukleotidu a detekci signálu se nukleotid odmyje a přidá se další odmyje

DNA sekvence: **C T C C G**

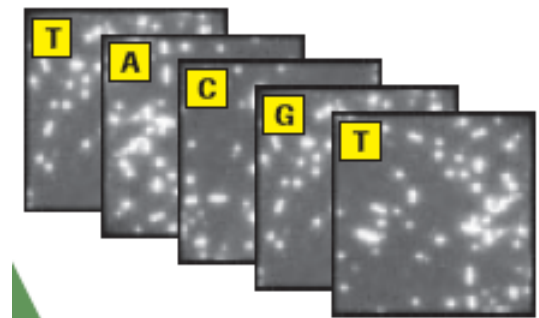
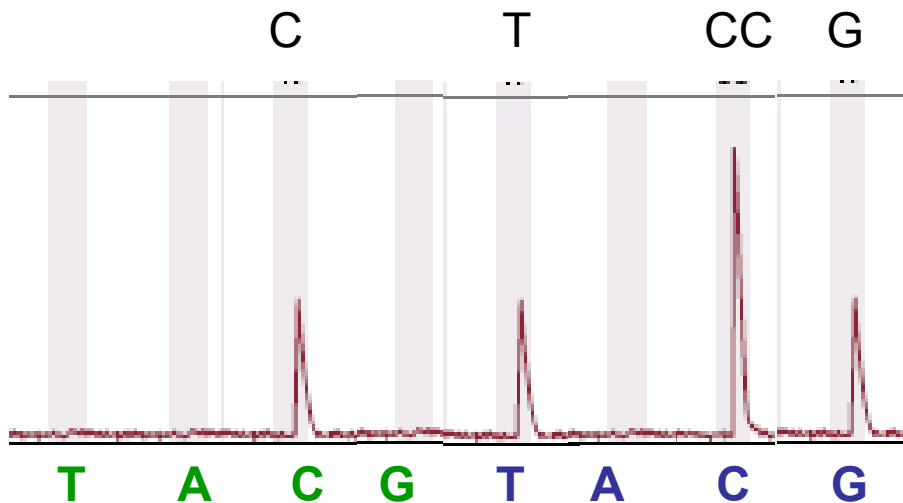


Image Files:  
12-15 gigabytes  
per run

**Problém!!!!** Homopolymery např. AAAAAAAAAA

# High-throughput - paralelní sekvenování

1 běh (run) = 1 destička:

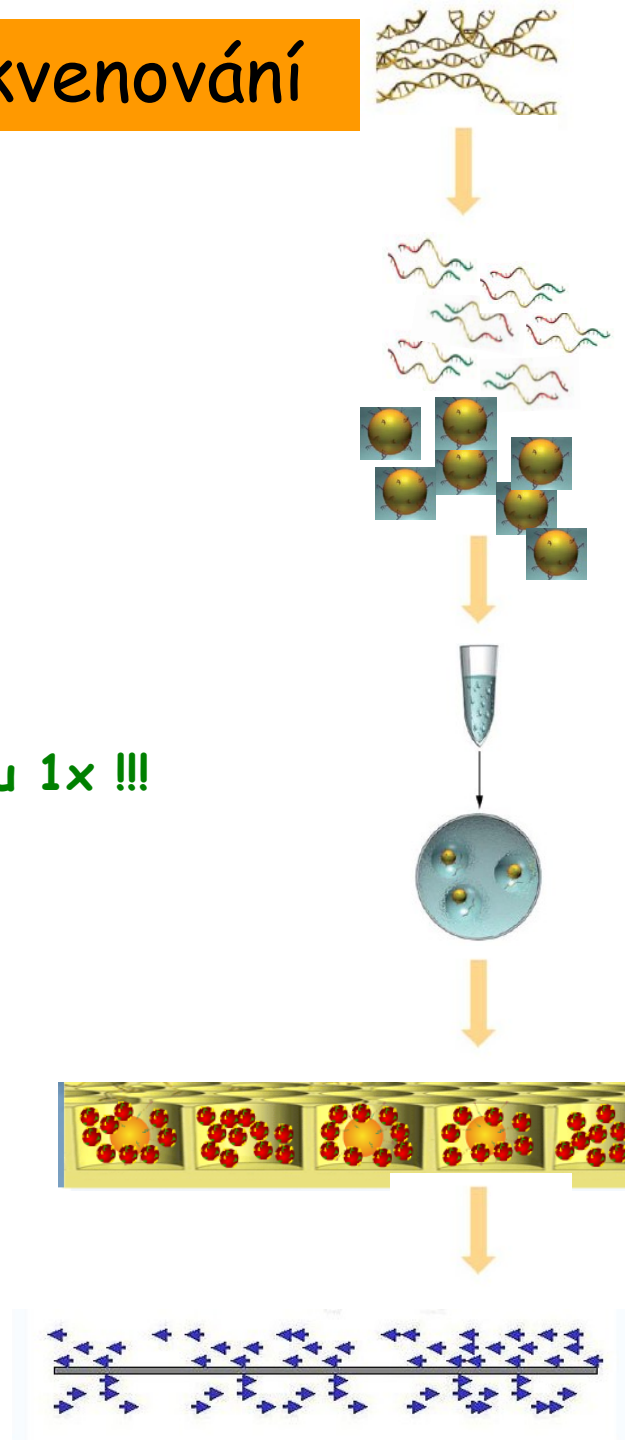
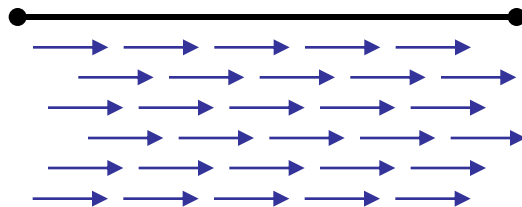
- 400 000 / 1milión jamek (reads)
- v každé 240 / 400 bp (read length)
- 7.5 / 10 hod

→ 100 Mb / 400 Mb na jednu destičku

→ cena??? 150-350 000 Kč ????

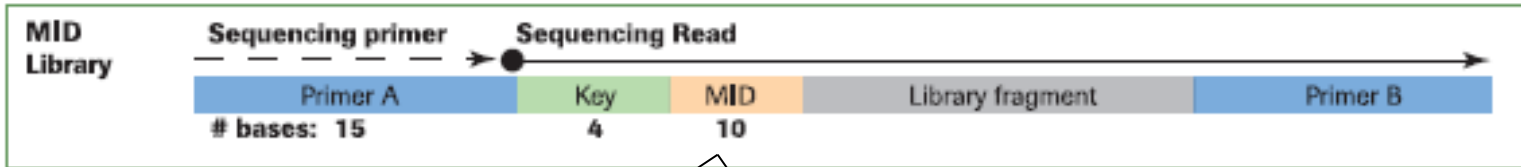
**!!! Samozřejmě nestačí mít každou bázi osekvenovanou 1x !!!**

- Pospojování (*reads assembly*) do souvislé sekvence
- Nepřesnosti - pokrytí (*coverage*)

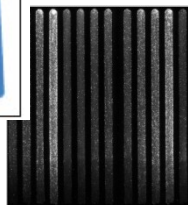
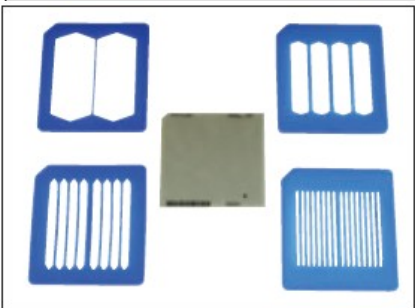
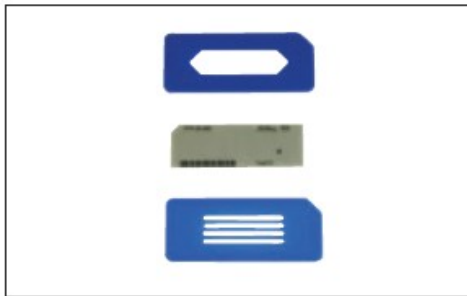


# Kapacita destičky **400 Mb**:

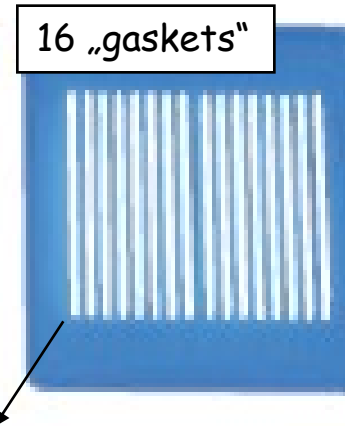
Mus:	2700 Mb	→ 7 run 1x coverage
Caenorhabditis:	100 Mb	→ 1 run 4x coverage
E. coli:	5 Mb	→ 1 run 80x coverage
mitoch. Mus:	0.016 Mb	→ 1 run 25000x coverage
HIV:	0.01 Mb	→ 1 run 40000x coverage



- k dispozici 12 odlišných MID



1. CCCCCCCCCC
2. GGGGGGGGGG
- ...
12. CCCCAAAG



$$\begin{aligned} &12 \text{ MID} \\ &\times \\ &16 \text{ gaskets} \\ &= \\ &\text{max. } 192 \text{ vzorků} \end{aligned}$$

V každém max. 12 vzorků  
(každý označen svým MID)

# Využití

1. Celogenomové sekvenování de novo
2. Celogenomové resekvenování
3. Sekvenování amplikonů (PCR produktů)  
+ to samé i s RNA (resp. cDNA)



# 1. Celogenomové sekvenování de novo

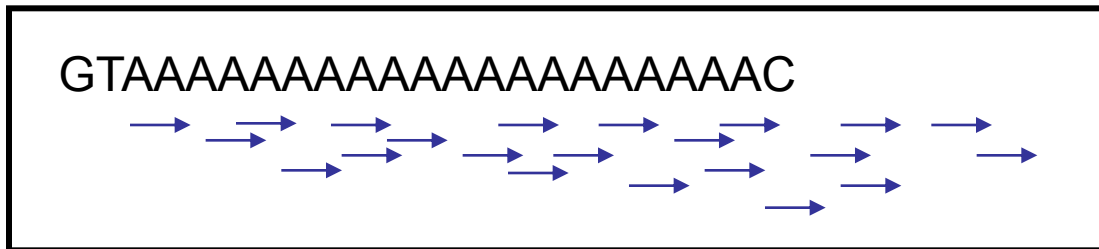
Problém: **KRÁTKÝ READ LENGTH**

- **400bp** 454 FLX Roche
- **35-75bp** Solexa, Solid
- vs **800-1000bp** Sanger

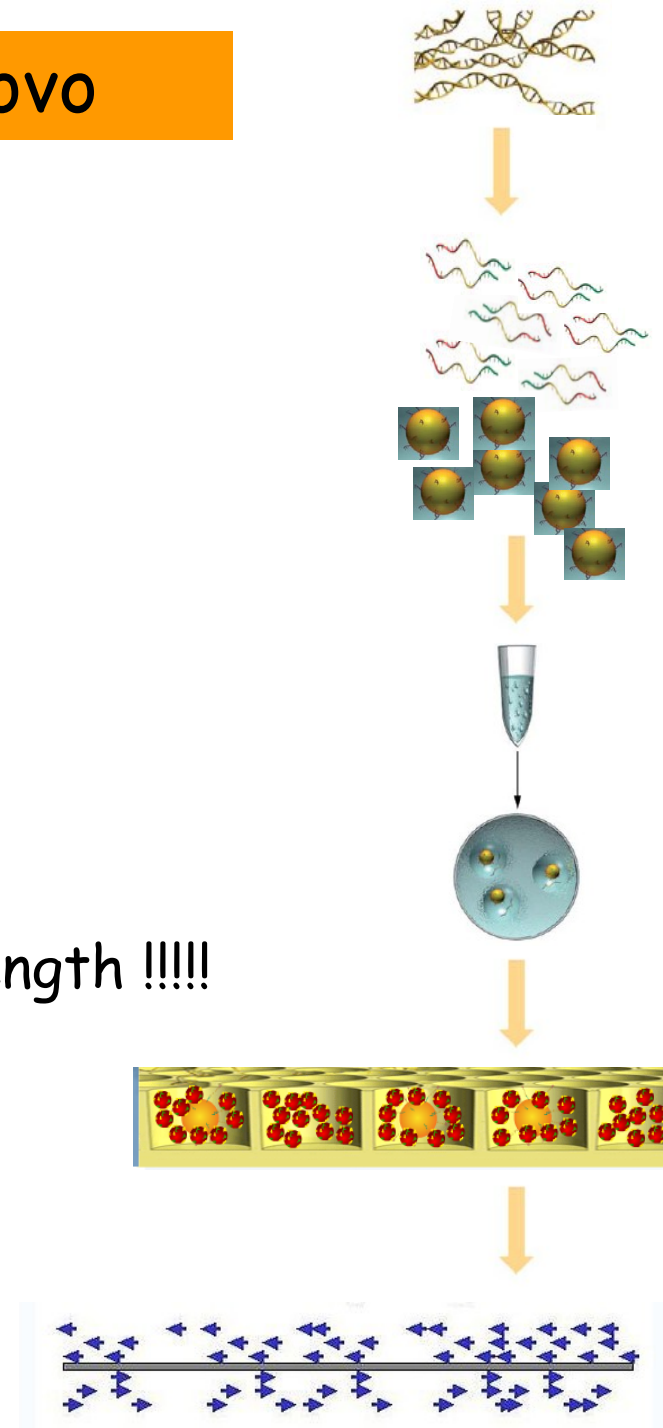


→ Uspořádání (assembly) už není problém z hlediska výpočetní kapacity

!!!! **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 Sangera)
- viry, prokaryota, malá eukaryota, mitochondrie/plastidy/plasmidy

× Ale č...



## Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever–Associated Arenavirus from Southern Africa

Thomas Briese<sup>1,3\*</sup>, Janusz T. Paweska<sup>2,3</sup>, Laura K. McMullan<sup>3</sup>, Stephen K. Hutchison<sup>4</sup>, Craig Street<sup>1</sup>, Gustavo Palacios<sup>1</sup>, Marina L. Khristova<sup>5</sup>, Jacqueline Weyer<sup>2</sup>, Robert Swanepoel<sup>2</sup>, Michael Egholm<sup>4</sup>, Stuart T. Nichol<sup>3</sup>, W. Ian Lipkin<sup>1\*</sup>

<sup>1</sup>Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York, United States of America, <sup>2</sup>Special Pathogens Unit, National Institute for Communicable Diseases of the National Health Laboratory Service, Sandringham, South Africa, <sup>3</sup>Special Pathogens Branch, Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, <sup>4</sup>454 Life Sciences, Branford, Connecticut, United States of America, <sup>5</sup>Biotechnology Core Facility Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

### Abstract

Lujo virus (LUJV), a new member of the family *Arenaviridae* and the first hemorrhagic fever–associated arenavirus from the Old World discovered in three decades, was isolated in South Africa during an outbreak of human disease characterized by nosocomial transmission and an unprecedented high case fatality rate of 80% (4/5 cases). Unbiased pyrosequencing of RNA extracts from serum and tissues of outbreak victims enabled identification and detailed phylogenetic characterization within 72 hours of sample receipt. Full genome analyses of LUJV showed it to be unique and branching off the ancestral node of the Old World arenaviruses. The virus G1 glycoprotein sequence was highly diverse and almost equidistant from that of other Old World and New World arenaviruses, consistent with a potential distinctive receptor tropism. LUJV is a novel, genetically distinct, highly pathogenic arenavirus.



ců

s  
G  
G  
T  
T  
G  
G  
T  
G

G  
G

T  
T

## 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>

# 3. Sekvenování ampliconů (PCR produktů)

## SMĚSNÉ VZORKY

### 1. Metagenomika/metatranskriptomika

- Celé společenstvo půdních, vodních mikroorganismů, střevní mikroflóra
- PCR genu 16S (18S) rRNA
- lze i kvantifikovat

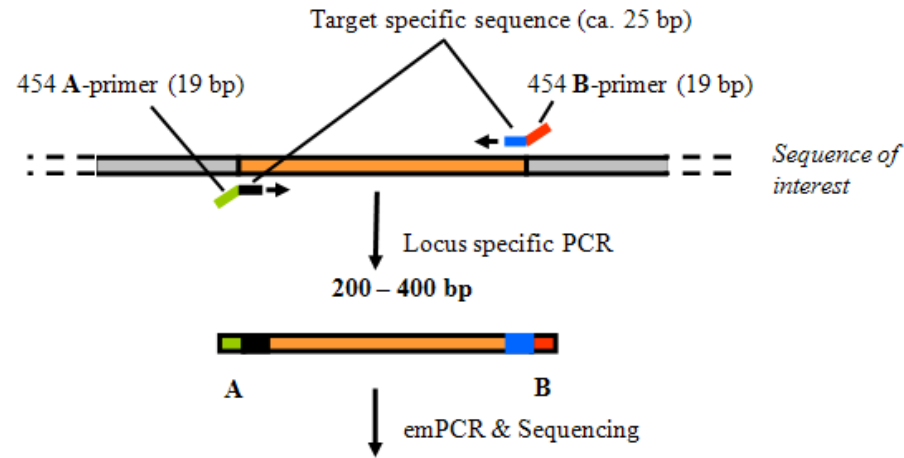
### 2. Složení potravy, trusu ???

### 4. Studie u kandidátních genů

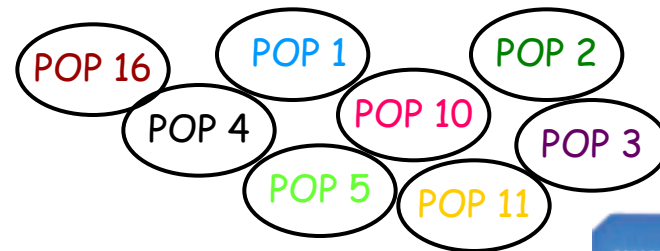
20x  
NEMOCNÉ MYŠI

20x  
ZDRAVÉ MYŠI

1. PCR např. imunitního genu/genů
2. Sekvenování
3. Které varianty jsou asociovány s chorobou??



### 3. Populační genetika



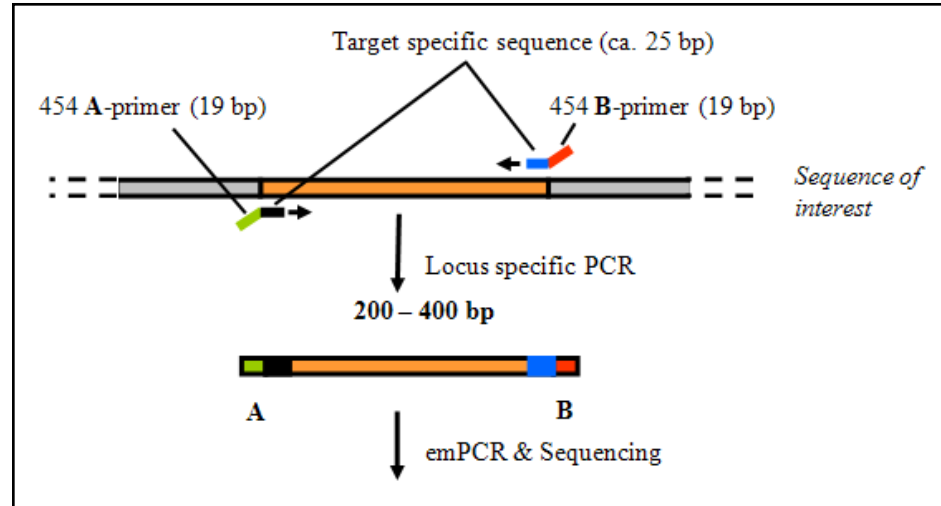
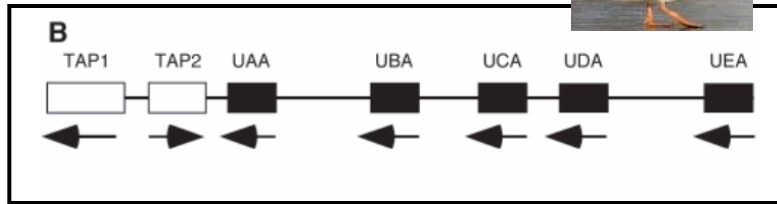
1. PCR genu/genů
2. Sekvenování
3. Zjištění sekvencí variant a frekvencí variant v každé populaci (záleží na pokrytí)





# 3. Sekvenování amplikonů (PCR produktů)

## 5. Genové duplikace



A-adaptor MID Target specific

Označí jedince

Amplifikuje všechny kopie MHC genů

Potřeba k emPCR, sekvenování..

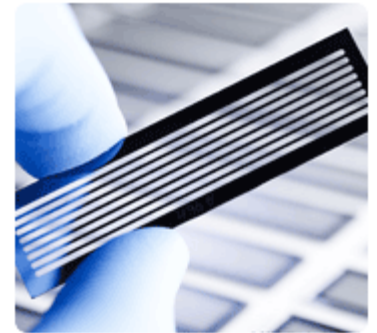


192 jedinců

# Solexa/Illumina 1G SBS technology

(SBS = sequencing by synthesis)

- 1 Gb (šestinásobek genomu *Drosophily*)
- Výrazně levnější
- Sekvence délky 35 bp
- Fluorescence, reversibilní terminátory
- Spíš pro resequencing


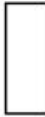

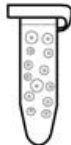


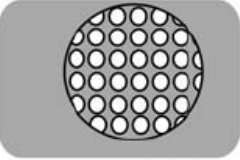
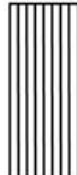
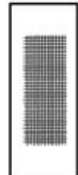
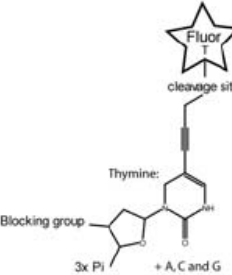


# SOLiD

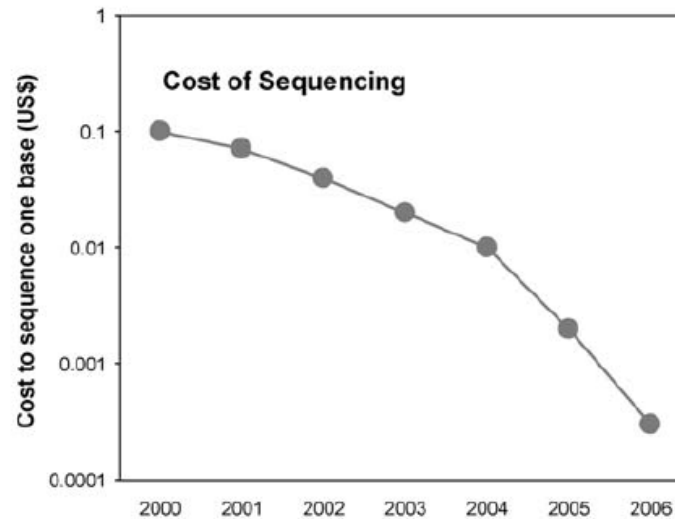
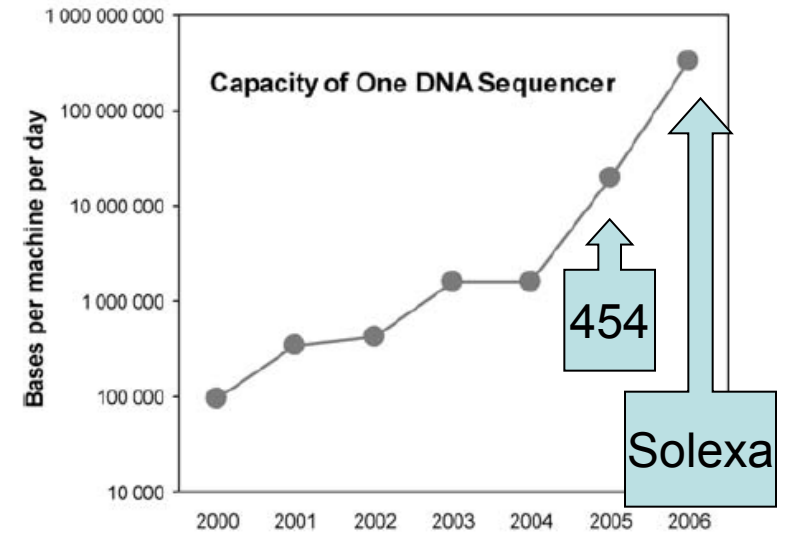
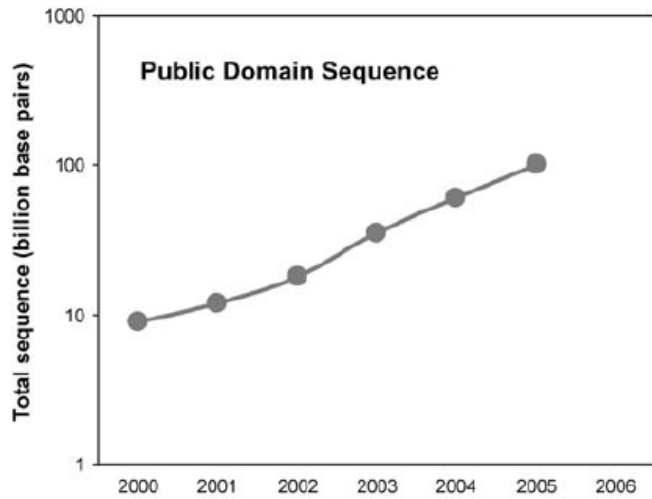
(sequencing by Oligonucleotide Ligation and Detection)



... a další (každého půlroku nová technologie – bouřlivý rozvoj !!!)

	454 pyrosequencing	Solexa SBS sequencing	Agencourt / ABI SOLiD polony sequencing
All methods ligate single, randomly sheared DNA molecules to support			
DNA support	 25–36 µm bead	surface of flow cell 	 ~1 µm bead
Amplification	 emulsion-phase PCR	 <i>in situ</i> PCR on solid surface	 emulsion-phase PCR
Sequencing surface	 1 600 000 well plate one bead per well	 8-channel flow cell clusters of DNA randomly located	 Single slide imaged in panels beads random
Sequencing chemistry	<p>Nucleotide incorporation</p> <p>↓</p> <p>PPi      ADP + Sulfurylase</p> <p>    ↓      ↓</p> <p>    ATP      ↓</p> <p>Luciferin Luciferase → light</p> <p>pyrosequencing</p>	<p>Fluor T</p> <p>↓</p> <p>cleavage site</p> <p>Thymine: </p> <p>Blocking group      3x Pi      + A, C and G</p> <p>reversible-terminator sequencing by synthesis</p>	<p>Fluor G</p> <p>Fluor C</p> <p>Fluor A</p> <p>Fluor T</p> <p>nnnnGzzz</p> <p>nnnnCzzz</p> <p>nnnnAzzz</p> <p>nnnnTzzz</p> <p>Ligation of sequence- specific labeled oligos</p>
Sequence detection	Chemiluminescence (one channel)	Fluorescence (four channel)	Fluorescence (four channel)
Read length and number	100–400 bp > 2 x 10 <sup>5</sup> reads	35 bp ~ 4 x 10 <sup>7</sup> reads	25 bp (paired) > 10 <sup>7</sup> reads

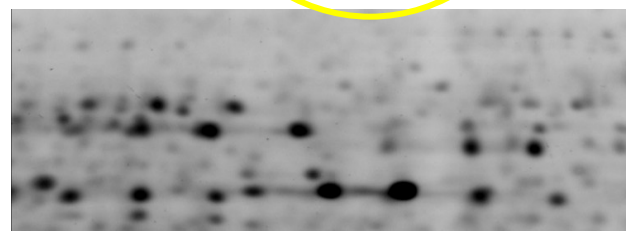
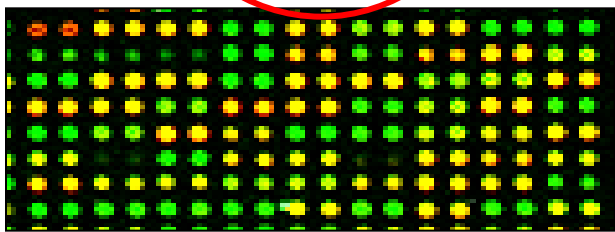
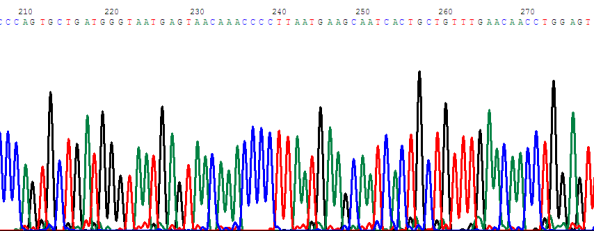
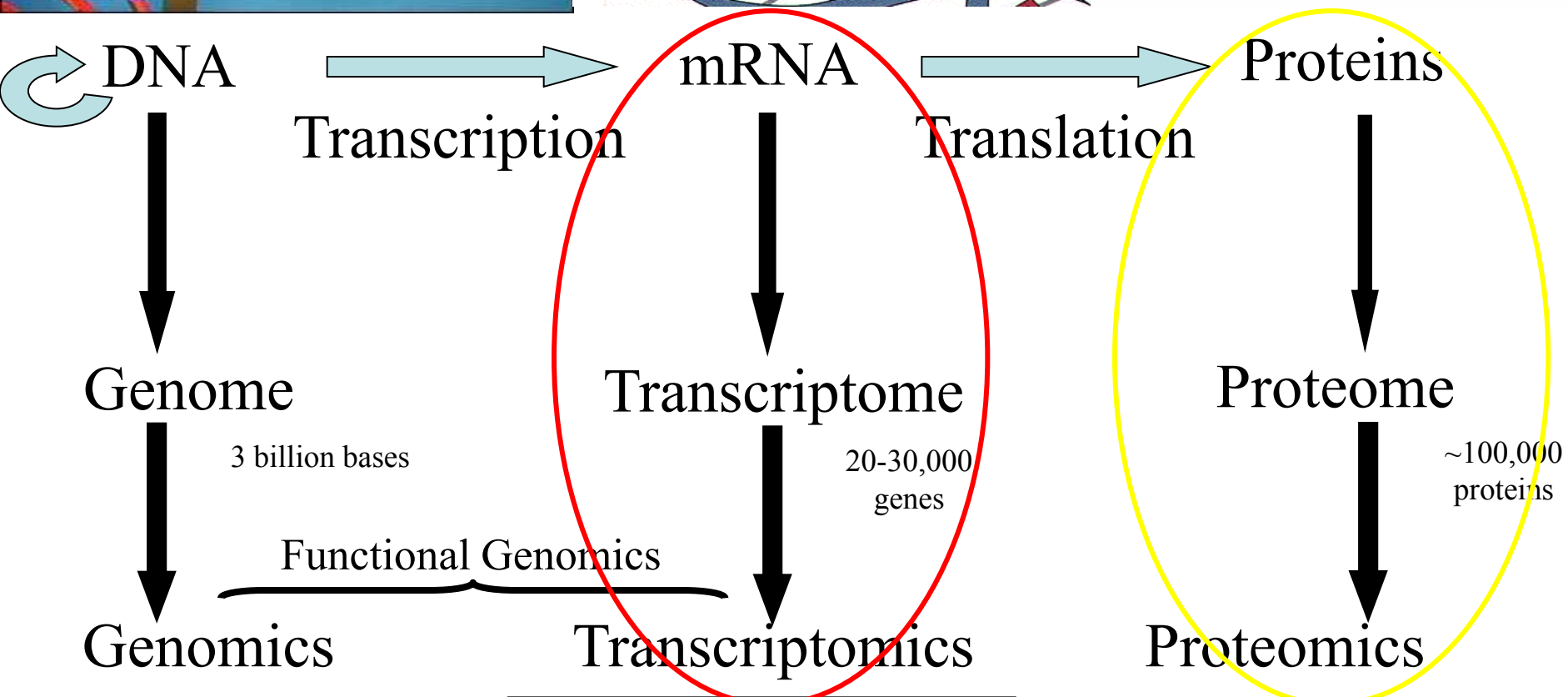
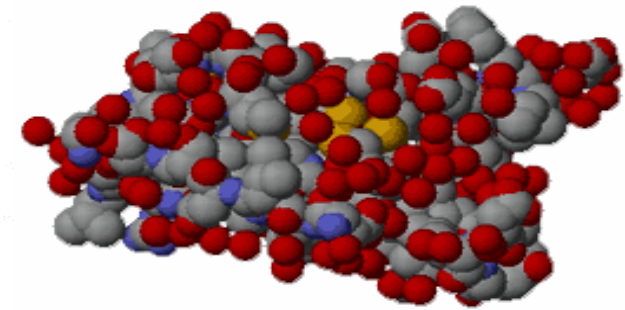
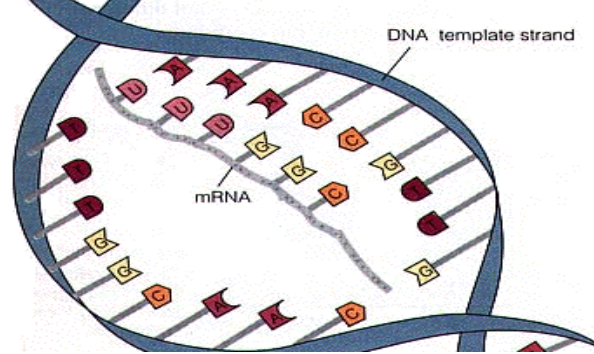
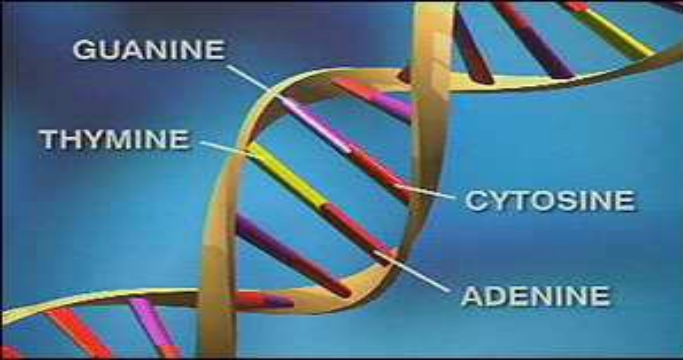
# „genomics era“



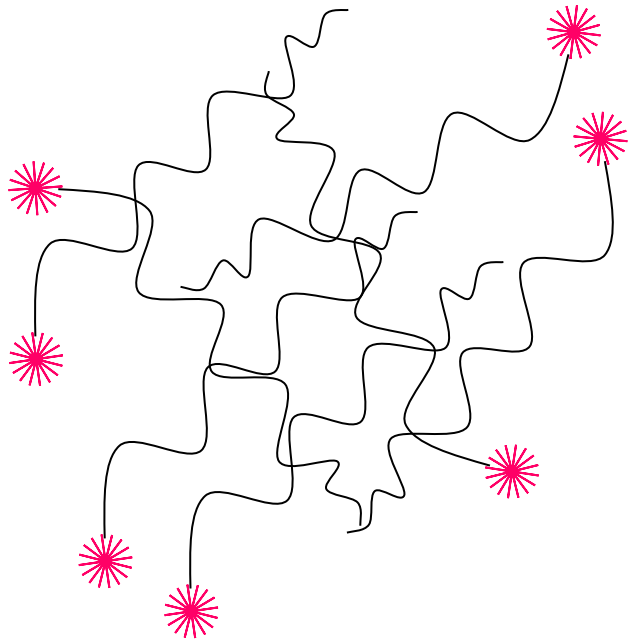
# Budoucnost genetických metod v ekologickém výzkumu

## 2. Analysis of expression by microarrays („transcriptomics“)

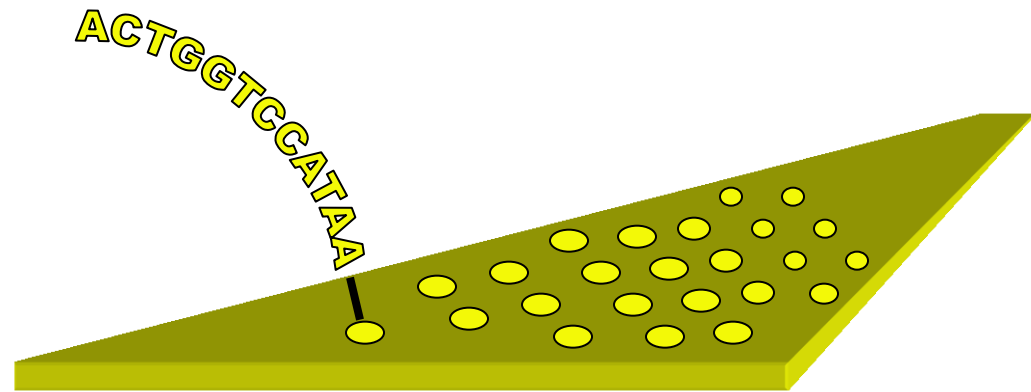
Ranz JM, Machado CA: Uncovering evolutionary patterns of gene expression using microarrays. TREE, 21(1): 29-37



# Microarray analysis of transcriptome (~ specific DNA hybridization)



**Target** (i.e. mix of transcripts  
in a form of cDNA)



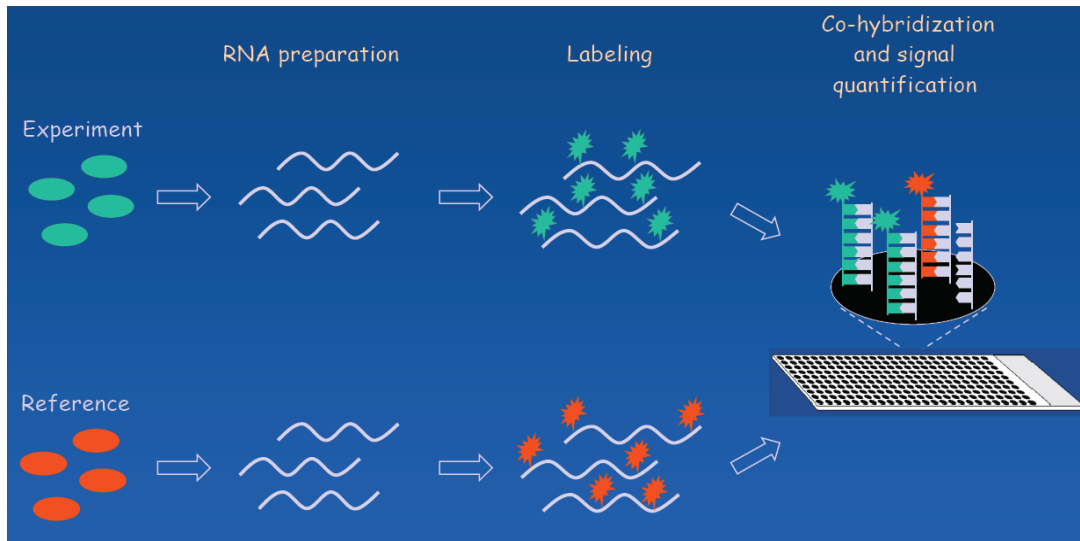
**Probe** (i.e. synthesized  
oligonucleotides  
complementary to particular  
genes)



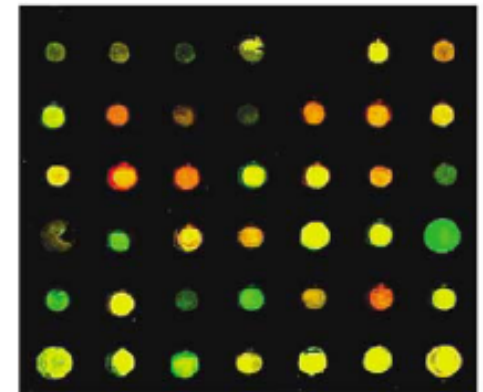
# How to get a transcription profile

- vždy srovnání kontroly a „treatment“

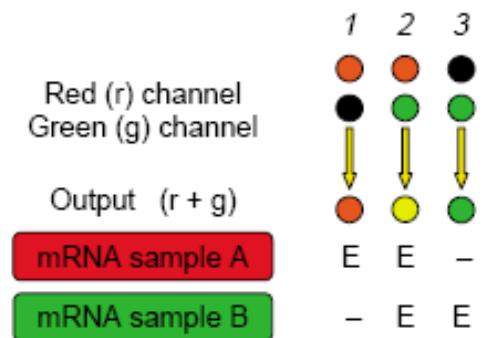
(a)



(b)



(c)



TRENDS in Ecology & Evolution

Analysis of expression level

# Case study: Joop Ouborg et al.

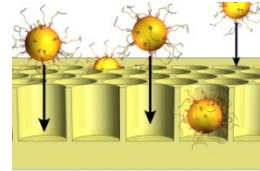
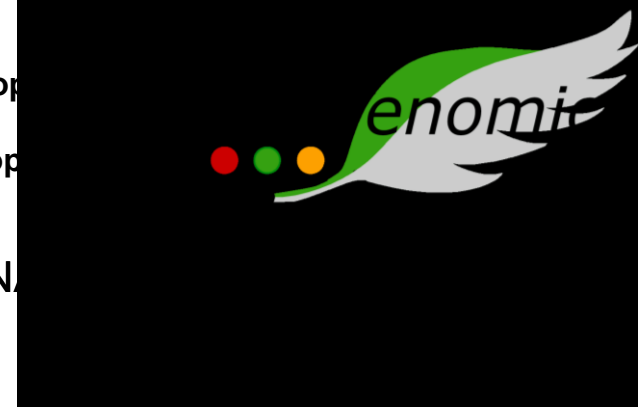
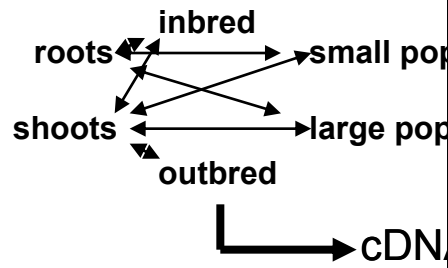
Transcriptional profiling of inbreeding depression and genetic erosion in *Scabiosa columbaria*: the balance between genetic drift and selection in the genetic erosion process.





Example:

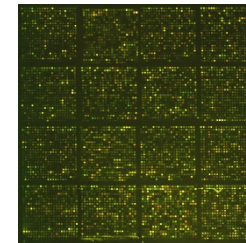
*Scabiosa columbaria*



530.000 sequences in one run, leading to ~ 40.000 ESTs



**Agilent Technologies**

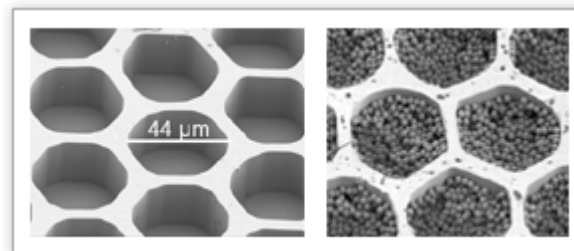
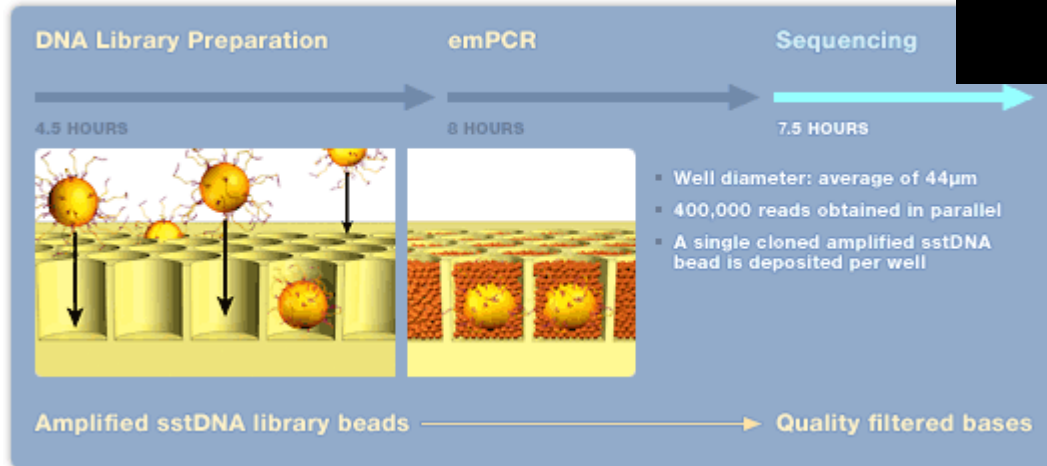


15k – 30k  
60-mer  
microarrays

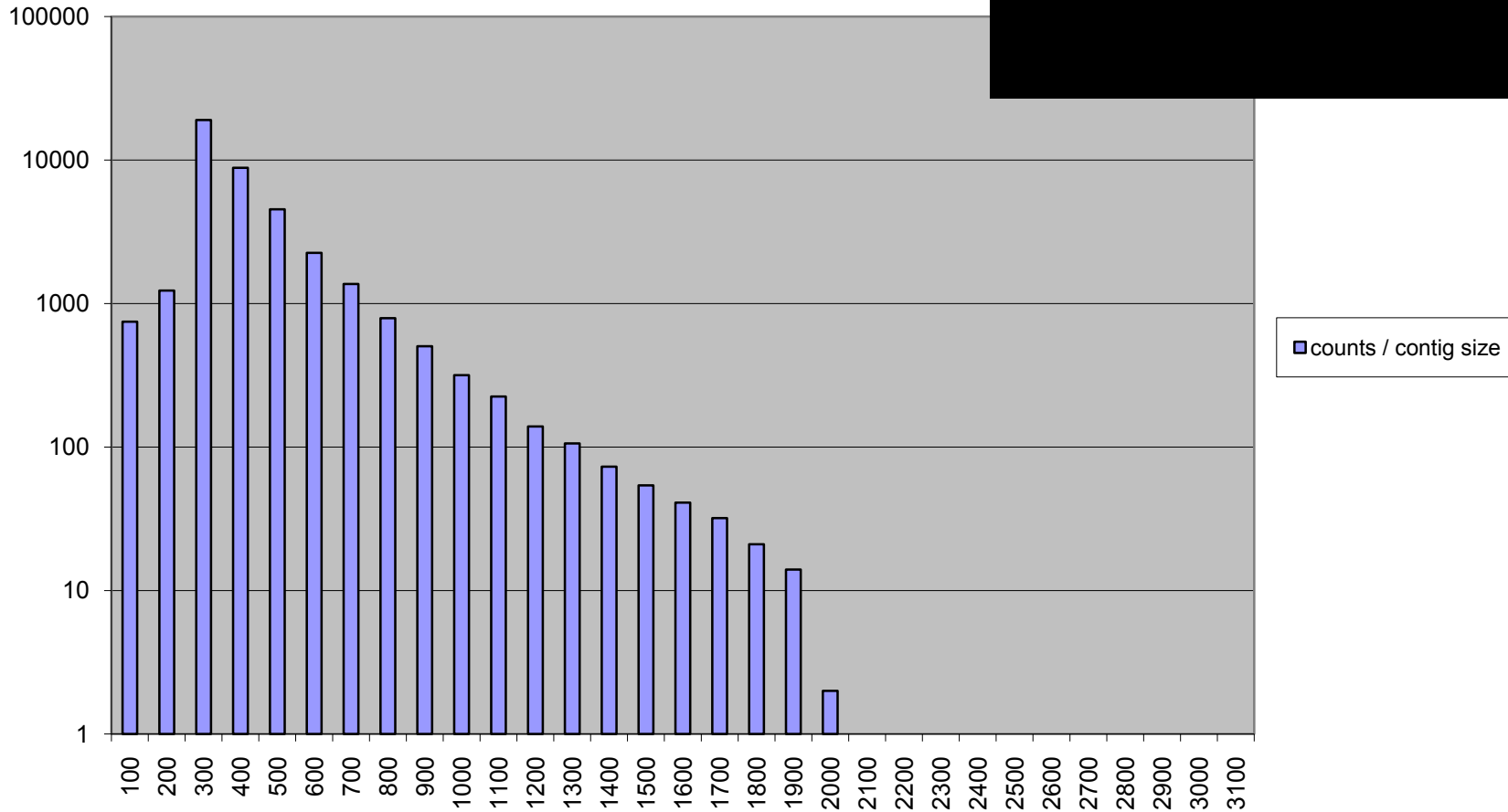
Experiment: transcriptional profiling of inbreeding depression

# cDNA library preparation – 454 sequencing of transcr

FIGURE 9



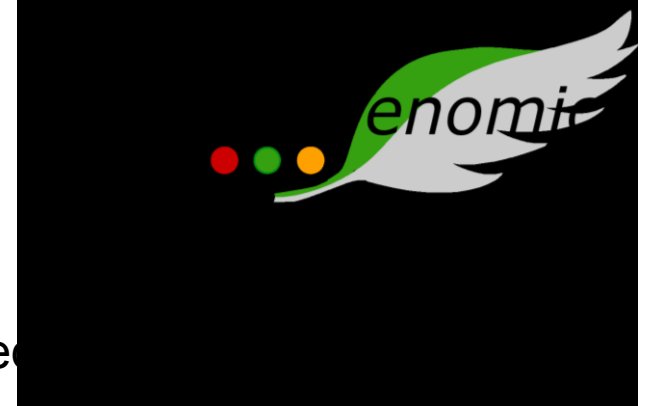
Counts (log) / contig size



Total number of reads: **528557**

Number of contigs: **40302**

In the next phase:



*Annotation* of these 40.000+ ESTs („expressed

Automated programs available, like **BLAST2GO** (<http://www.blast2go.de/>):  
just feed a file with the ESTs into the program, and turn it on.....

1 week later you will have the results, being:

- Homology with known sequences
- Known function

The sequences may also be searched for:

EST-associated SSR markers: MISA (<http://pgrc.ipk-gatersleben.de/misa/>)

SNP markers: SNP-mining software like PolyBayes

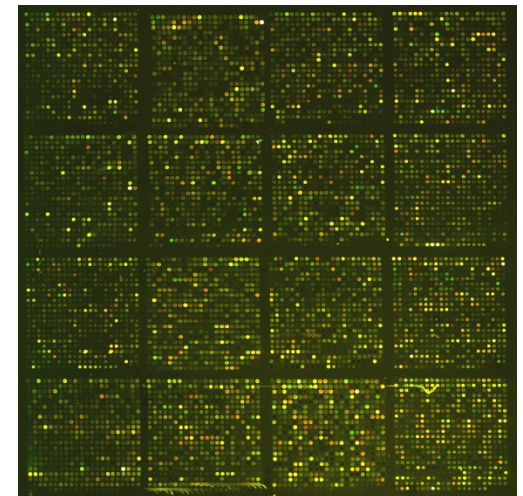
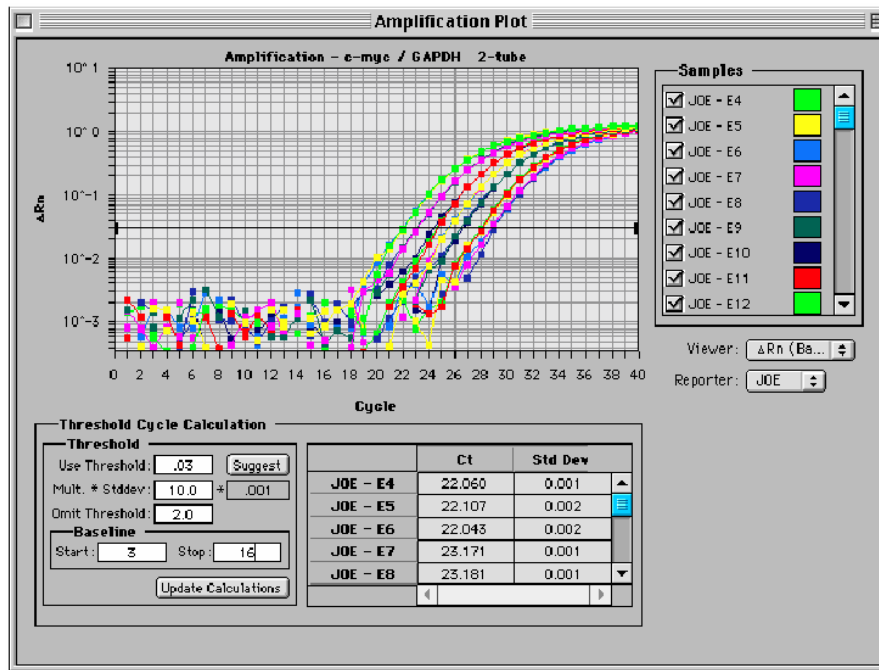
(<http://genome.wustl.edu/tools/software/polybayes.cgi>)

Again by using search software, freeware

**ALMOST HALF OF GENES (ESTs) ARE UNKNOWN !!!**

1. Design of quantitative RealTime-PCR methods, based on primer sequences

2. Design of a Scabiosa specific microarray

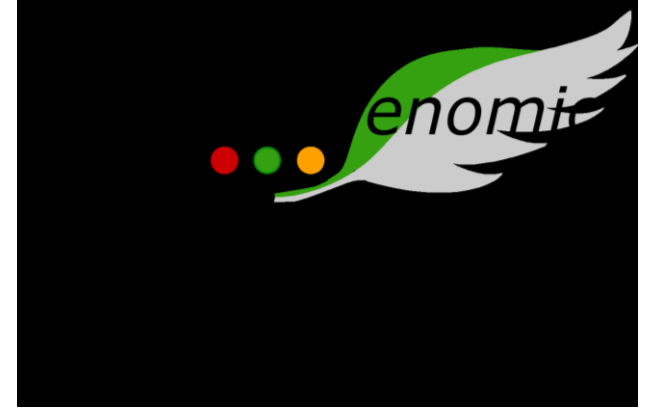




## Expected pay-off:

- Ecogenomic approach to conservation genetics: effects of genetic erosion on functional genetic diversity?
- How does genetic erosion affect evolutionary potential?
- What is the **balance between genetic drift and natural selection** in effects of habitat fragmentation?
- Are there general **inbreeding depression genes**, or is inbreeding depression a random phenomenon?
- **Which genes are involved in inbreeding depression in different life history stages**, and can this explain the non-correlation of IBD between these stages?
- What are the **footprints of selection** in the genomes of individuals from small and large populations?
- What is the **selective value of variation in gene expression**?





## Costs/requirements:

Costs are diminishing continuously

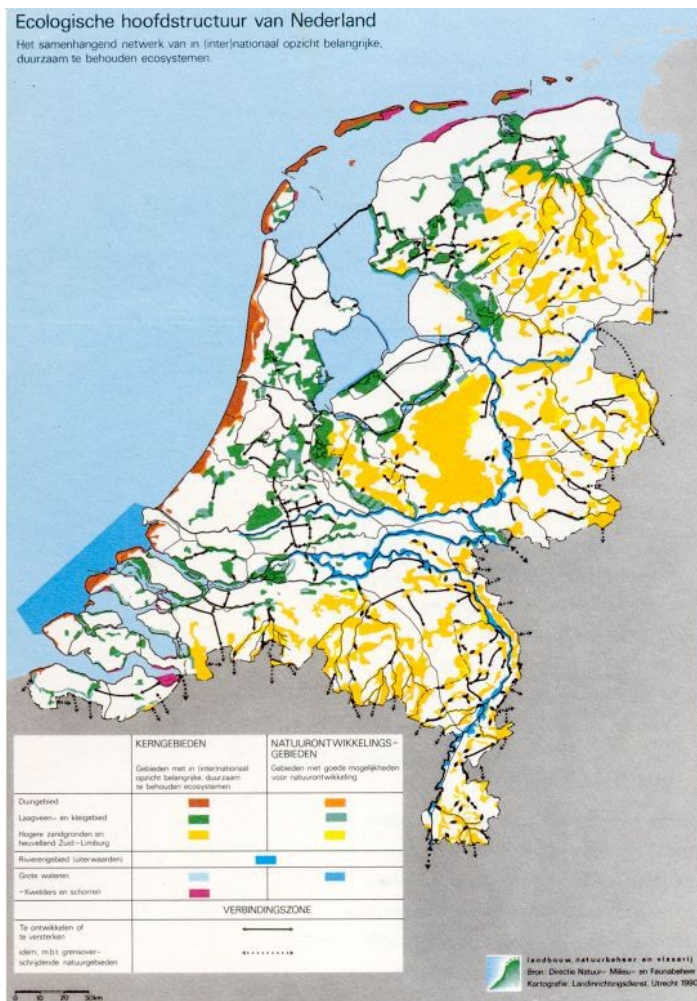
454 FLX-cDNA sequencing : 1 month, 15.000 € (used to be 200.000 € with Sanger technology)

microarray production: 100 € per array

microarray screening: 150 € per array

cheaper options (like SOLEXA technology) are becoming available, at much lower costs

## Relative costs of conservation genomics:



Projected costs (but this is almost certain a severe underestimation):

20 billion Euro

That is:

20.000.000.000 Euro

That is equivalent to 40.000.000 microarray runs.....

**We live in exciting times !!!**