Budoucnost genetických metod v ekologickém výzkumu

1. Nové postupy při sekvenování DNA („genomics")

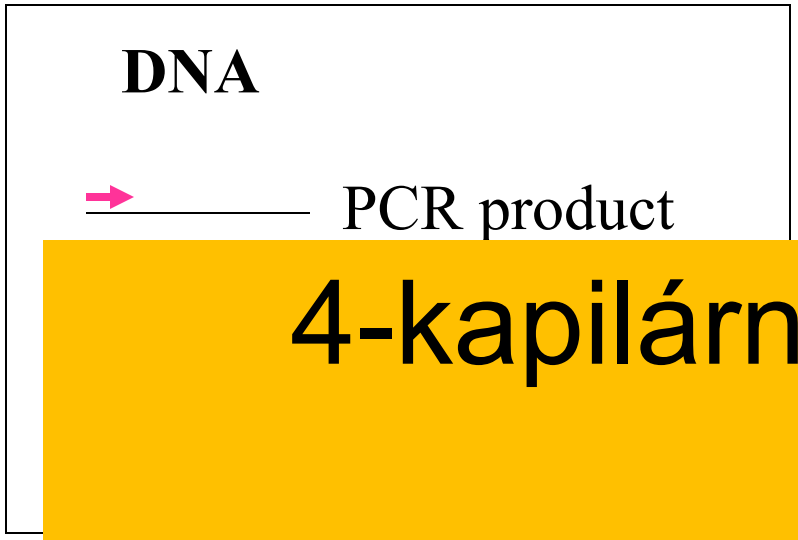
Molecular Ecology Resources (2008) 8, 3–17 doi: 10.1111/j.1471-8286.2007.02019.x

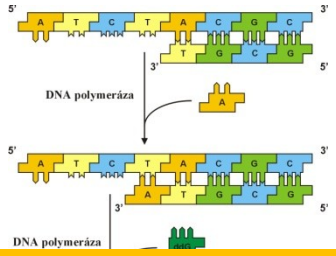
TECHNICAL REVIEW

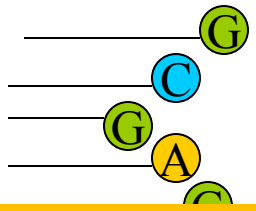
Sequencing breakthroughs for genomic ecology and evolutionary biology

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

Sequencing - Sangerova metoda



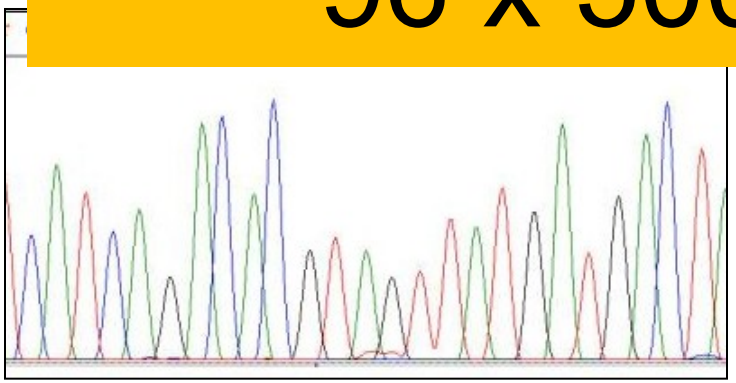


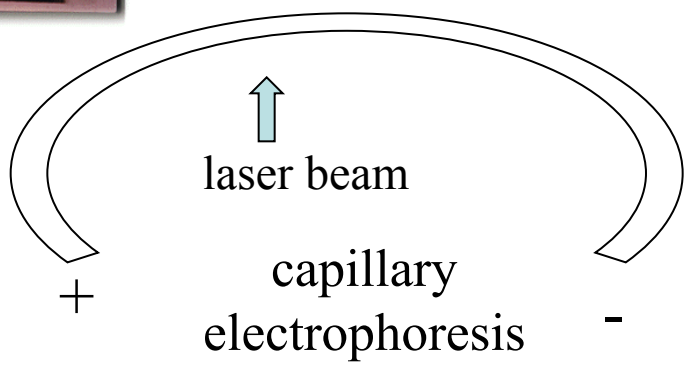


4-kapilární sekvenátor

=

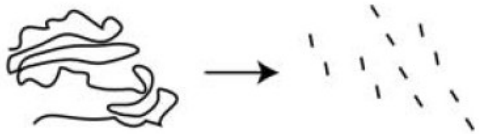
96 x 500 bp/12 hodin



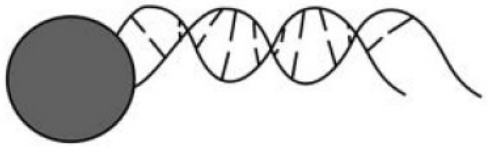


„Next generation sequencing"

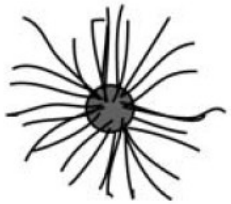
1) Randomly fragment many molecules of target DNA



2) Immoblize individual DNA molecules on solid support

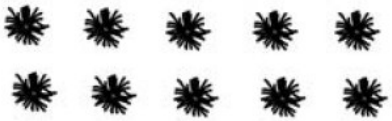


3) Amplify DNA in clonal 'polymerase colony'

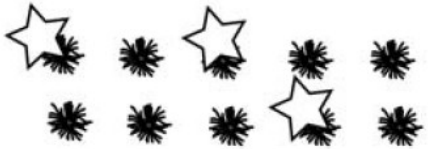


„polonies"
(polymerase colonies)

4) Sequence DNA by adding liquid reagents to immoblized 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 inkorporaci bazí
- 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

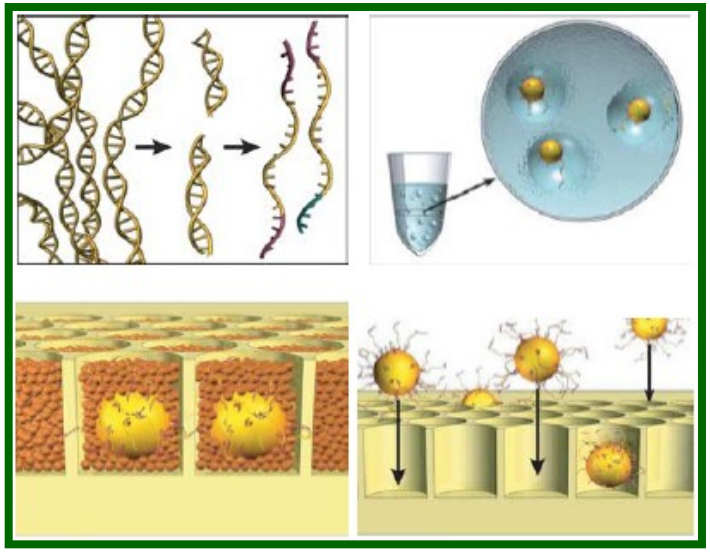
NEWS AND VIEWS

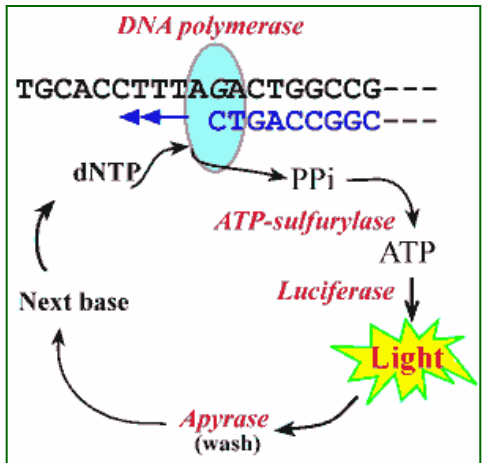
PERSPECTIVE

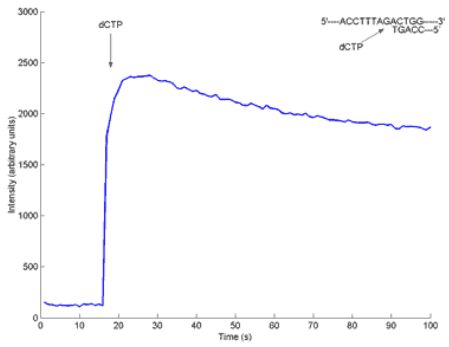
Sequencing goes 454 and takes large-scale genomics into the wild

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

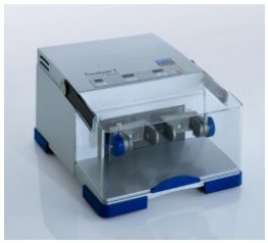
1 600 000 well plate







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

Controlled Room

4. Amplification
5. Bead Recovery
6. DNA Library Bead Enrichment
7. Sequencing Primer Annealing

Time: 11 - 13 h

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

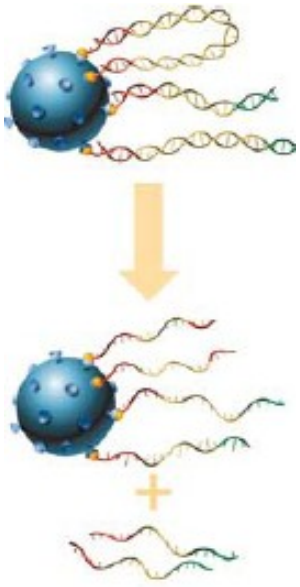
1. Příprava jednořetězcové DNA knihovny (ssDNA library preparation)

1 DNA Fragmentation (Nebulization):

5 Adaptor Ligation:

7 Library Immobilization:

9 ssDNA Library Isolation:



Adaptor A + **Adaptor B**

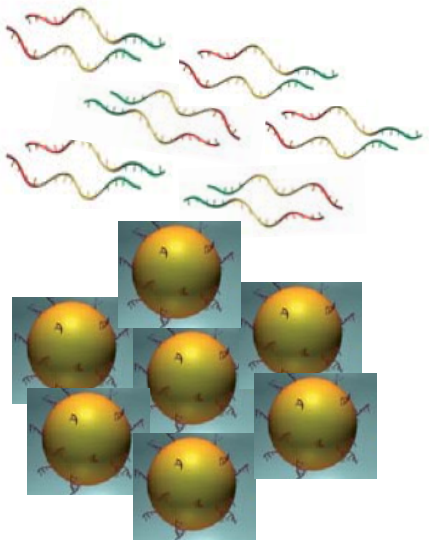
-Slouží jako vazebné místo primerů pro následnou PCR amplifikaci a sekvenování

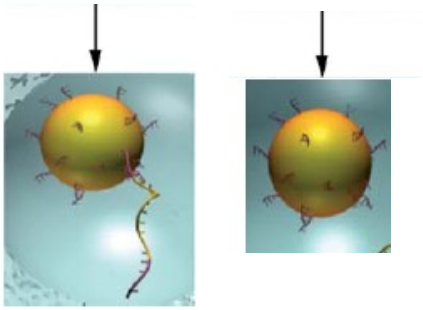
-Slouží k uchycení na kuličky (na adaptor B je připojen biotin)

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

1 DNA Library Capture:

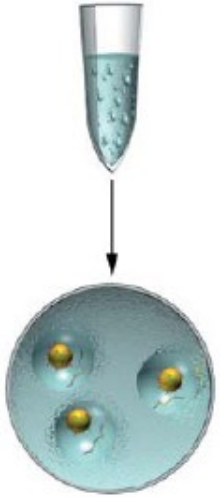
- poměry nastavit tak aby 1kulička ≤ 1 molekula DNA



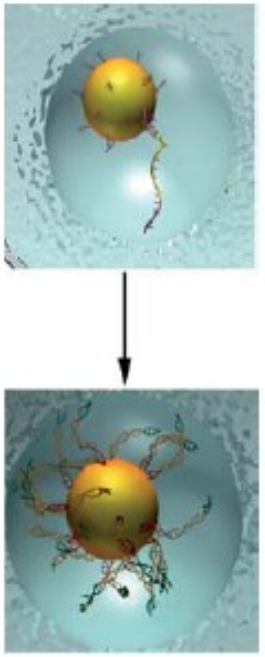


2 Preparation of the Amplific. Mixes

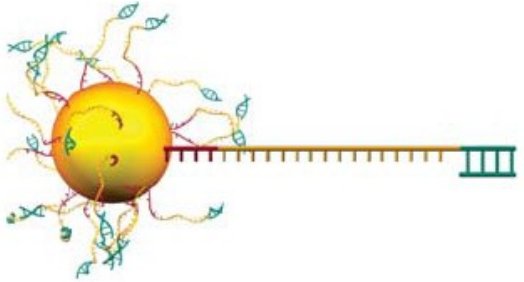
3 Emulsification:



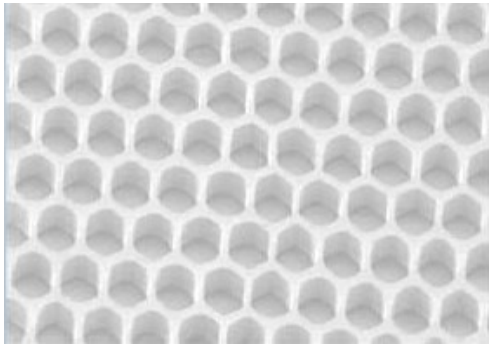
4 emPCR Amplification:



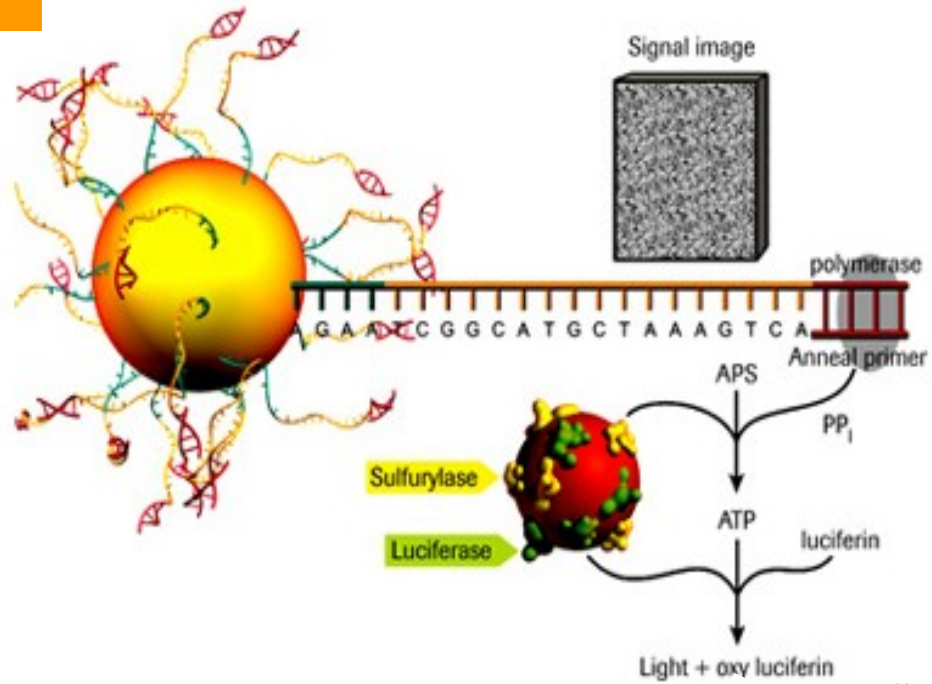
7 Sequencing Primer Annealing:



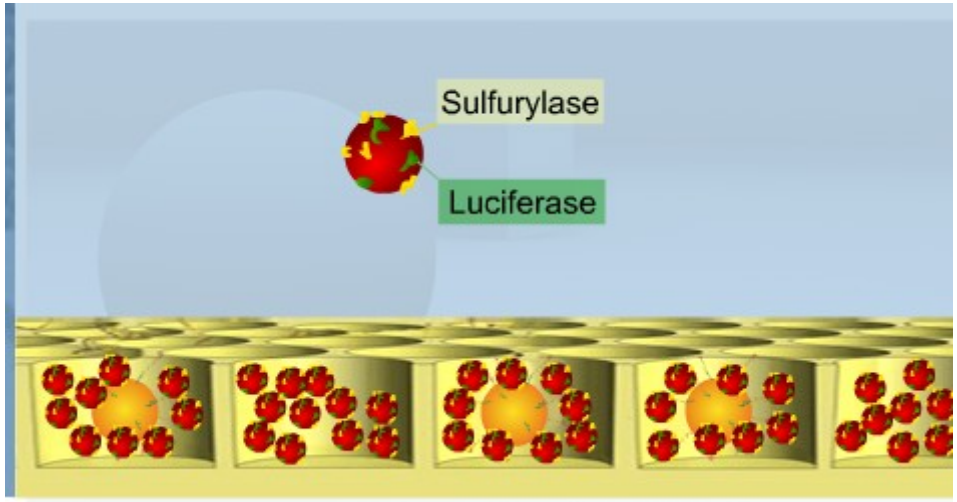
3. Pyrosekvenovani

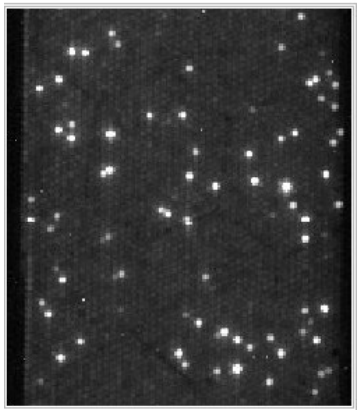


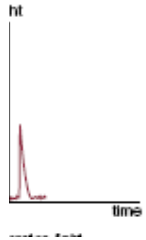
pikotitrační destička



Na jedné desticče 400 000 až 1milión jamek



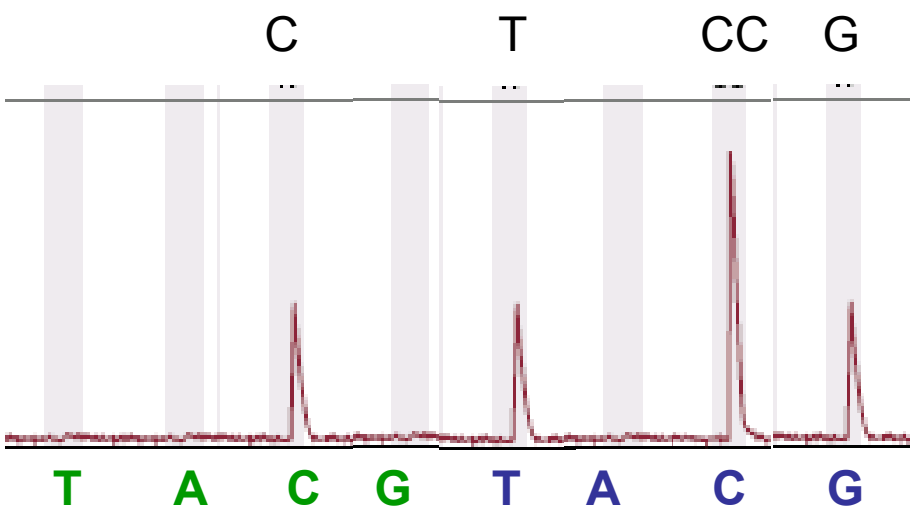




3. Pyrosekvenovani – detekce signálu

- 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**



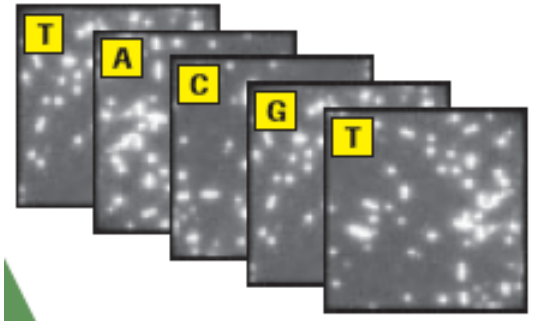


Image Files:
12-15 gigabytes
per run

Problém!!!! Homoplymery např. AAAAAAAAAA

High-throughput – paralelní sekvenování

1 běh (run) = 1 destička:

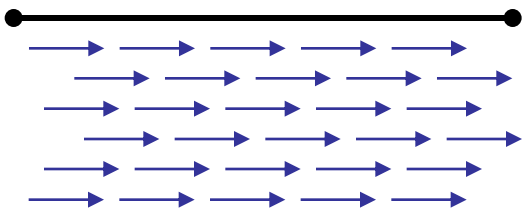
- 400 000/ 1milión jamek (reads)
- v každé 240 / 400 bp (read length)
- 7.5 / 10 hod

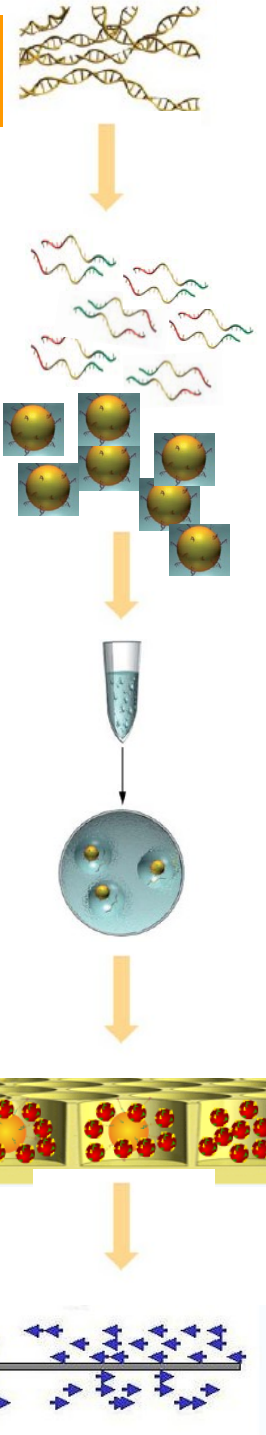
→ 100 Mb / 400 Mb na jednu destičku

→ cena??? 150-350 000 Kč ????

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

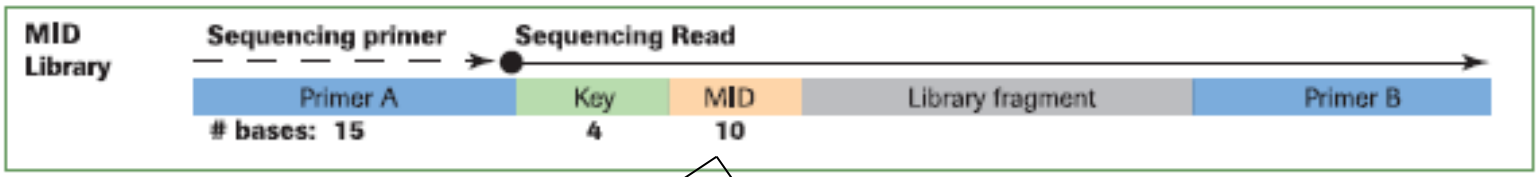
- Pospojování (reads assembly) do souvislé sekvence
- Nepřesnosti – pokrytí (coverage)





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

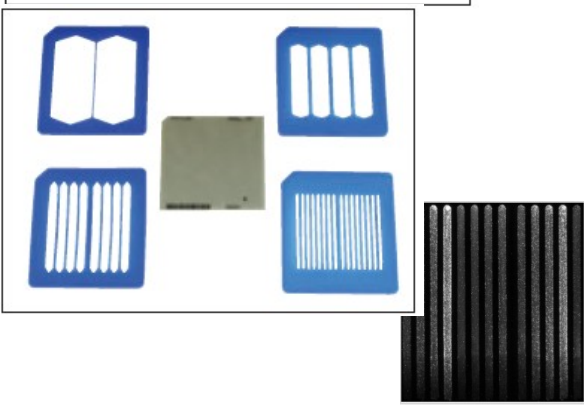


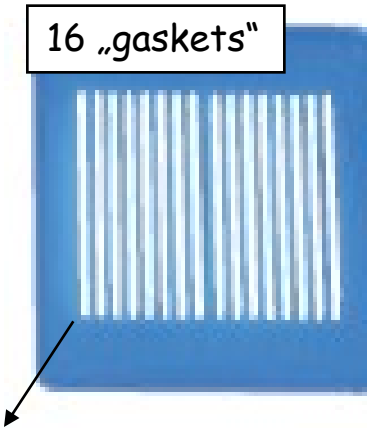
- k dispozici 12 odlišných MID

1. *CCCCCCCCCC*
2. *GGGGGGGGG*

.
.
.

12. *CCCCCAAAG*





12 MID
X
16 gaskets
=
max. 192 vzorků

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)

Chip-sequencing

Metylace

SNP

SAGE

1. Celogenomové sekvenování de novo

Problém: **KRÁTKÝ READ LENGTH**

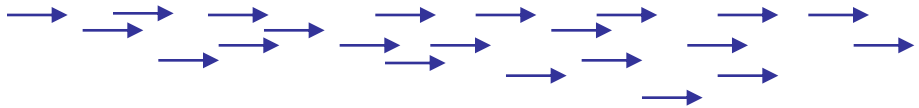
- **400bp** 454 FLX Roche
- **35-75bp** Solexa, Solid

  vs **800-1000bp** Sanger

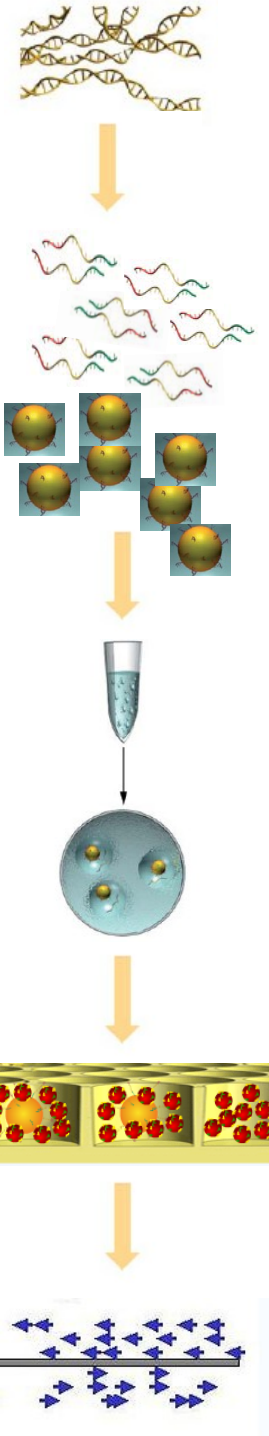
→ Uspořádání (assembly) už není problém z hlediska výpočetní kapacity

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

GTAAAAAAAAAAAAAAAAAAAC

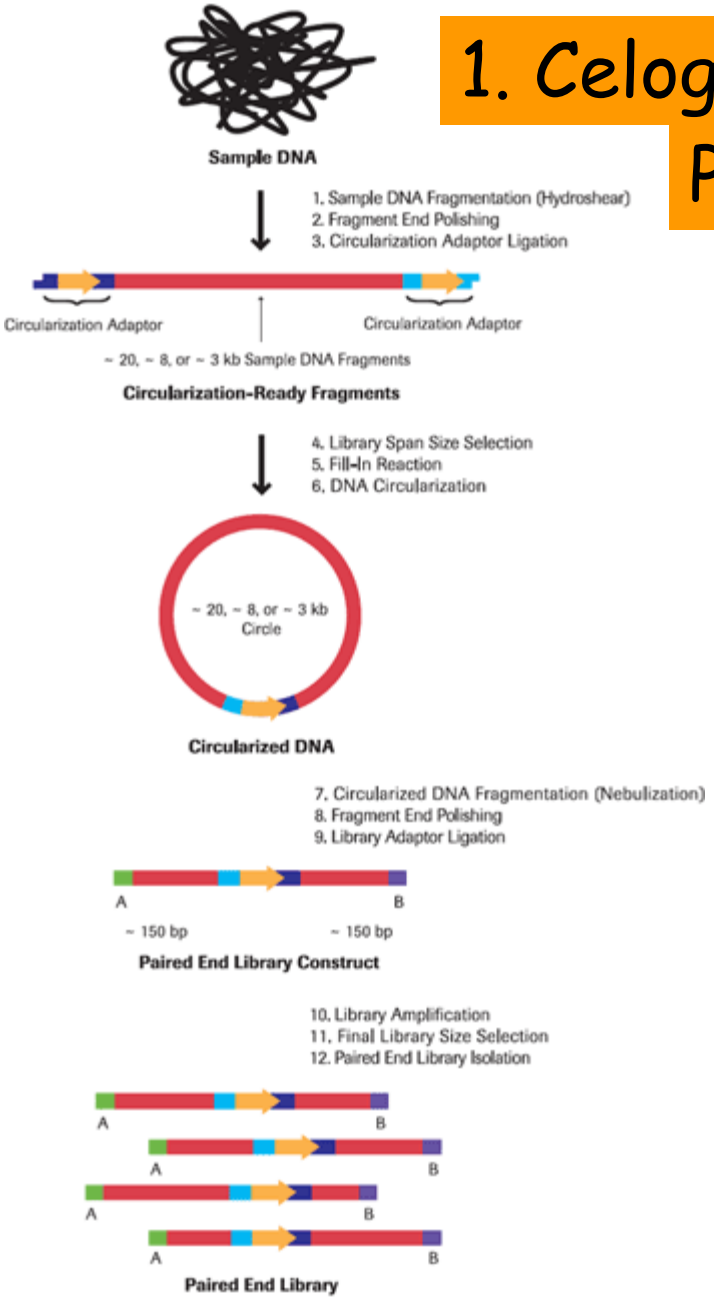


Zvláště komplexní eukaryotické genomy – úseky souvislých oblastí přerušených mezerami



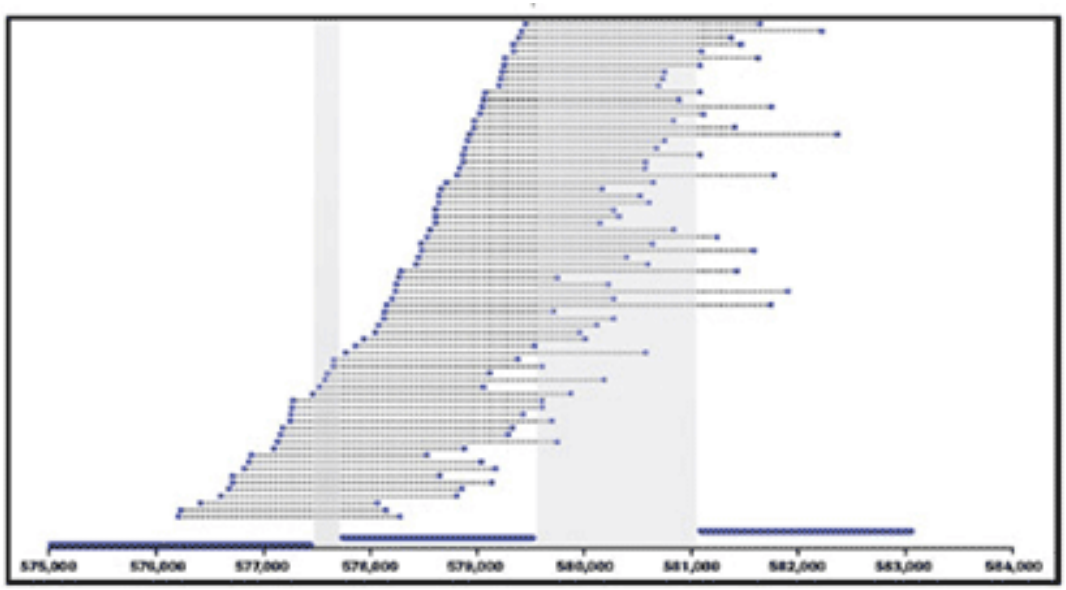
1. Celogenomové sekvenování de novo

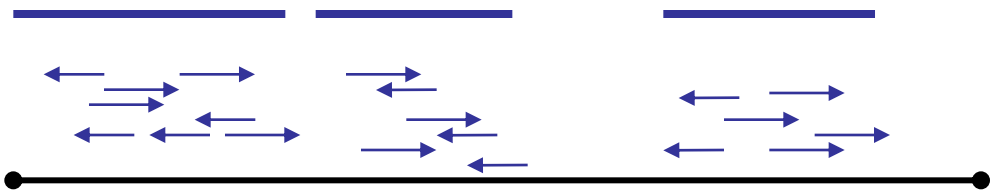
Pair end reads



???Řešení??? (asi ne úplně..)

- kombinace normálního shotgun pyrosekvenování a pair end reads





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 často nepotřebujeme kompletní a poskládanou sekvenci

Hledání (diagnostických) SNP:

smíchat a osekvenovat

10 jedinců

10 jedinců

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

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

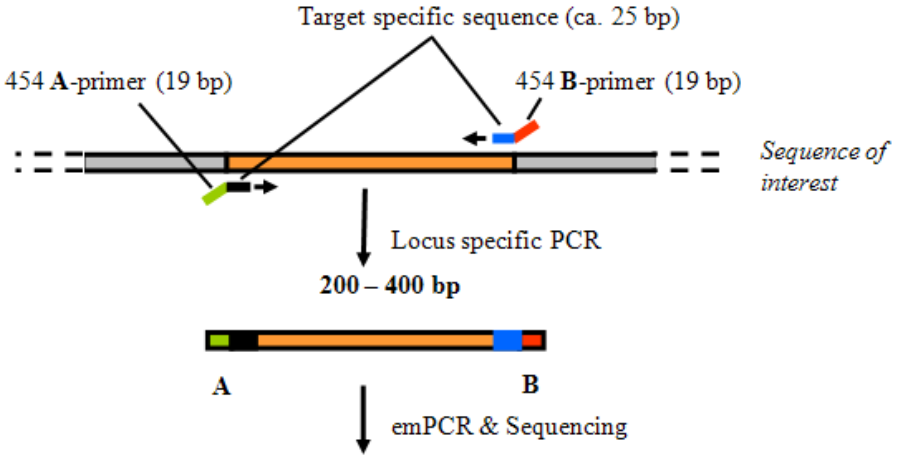
Richard E. Green,[1,*] Anna-Sapfo Malaspinas,[2] Johannes Krause,[1] Adrian W. Briggs,[1] Philip L.F. Johnson,[3] Caroline Uhler,[4] Matthias Meyer,[1] Jeffrey M. Good,[1] Tomislav Maricic,[1] Udo Stenzel,[1] Kay Prüfer,[1] Michael Siebauer,[1] Hernán A. Burbano,[1] Michael Ronan,[5] Jonathan M. Rothberg,[6]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

3.Sekvenování amplikonů (PCR produktů)

SMĚSNÉ VZORKY

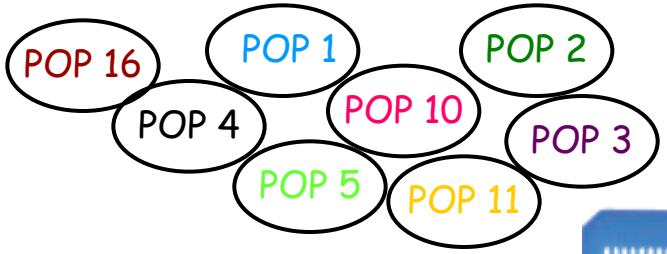
1. Metagenomika/metatranskriptomika

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



2. Složení potravy, trusu ???

3. Populační genetika



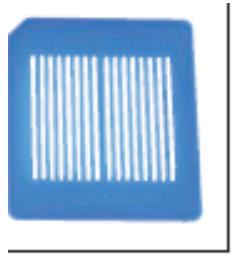
4. Studie u kandidátních genů

20x NEMOCNÉ MYŠI

20x ZDRAVÉ MYŠI

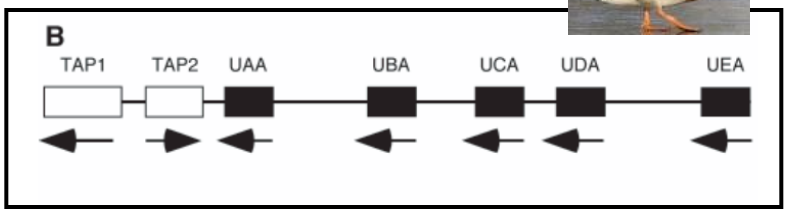
1. PCR např. **imunitního genu/genů**
2. Sekvenování
3. Které varianty jsou asociovány s chorobou??

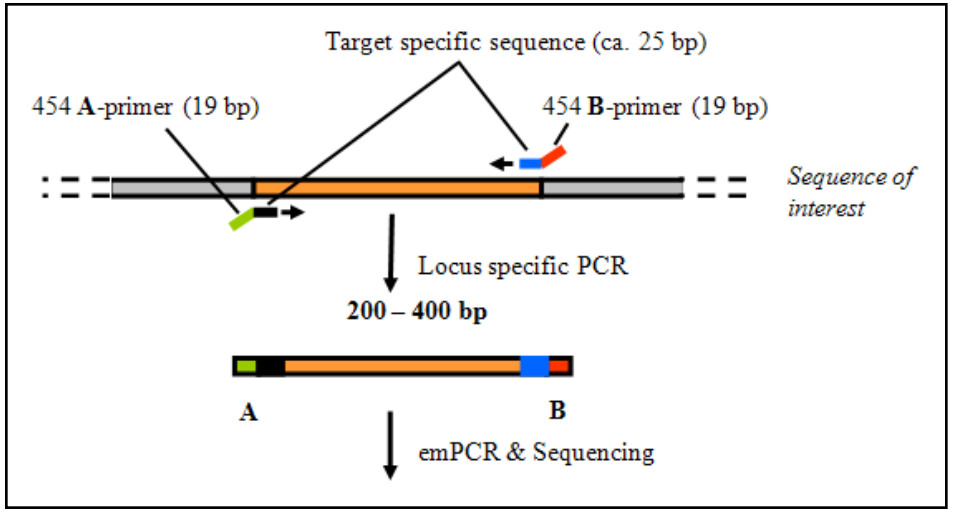
1. **PCR genu/genů**
2. Sekvenování
3. Zjištění sekvencí variant a frekvencí variant v každé populaci (záleží na pokrytí)



3.Sekvenování amplikonů (PCR produktů)

5. Genové duplikace



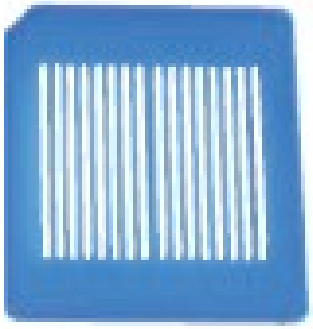


A-adaptor MID Target specific

Označí jedince

Amplifikuje všechny kopie MHC genů

Potřeba k emPCR, sekvenování..

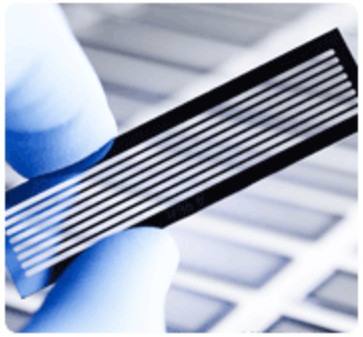


192 jedinců

Solexa/Illumina 1G SBS technology

(SBS = sequencing by synthesis)

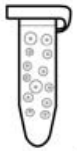
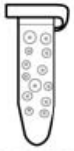
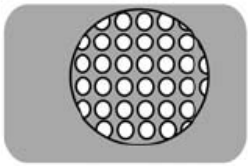
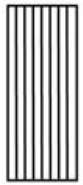
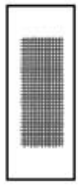
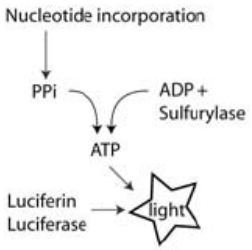
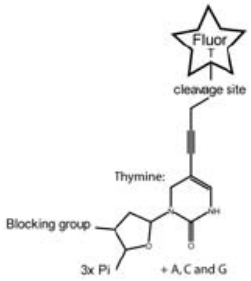
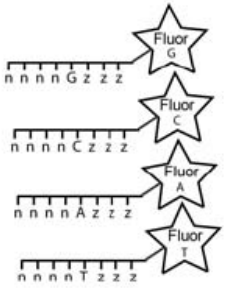
- 1 Gb (šestinásobek genomu *Drosophily*)
- Výrazně levnější
- Sekvence délky 35 bp
- Flourescence, reversibilní terminátory
- Spíš pro resequencing



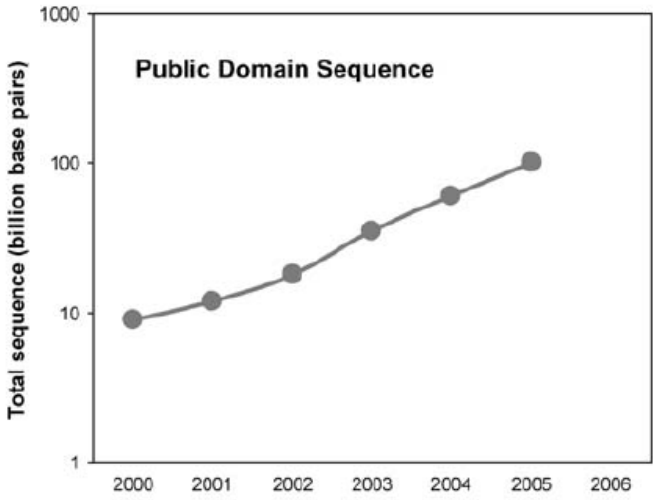
SOLiD

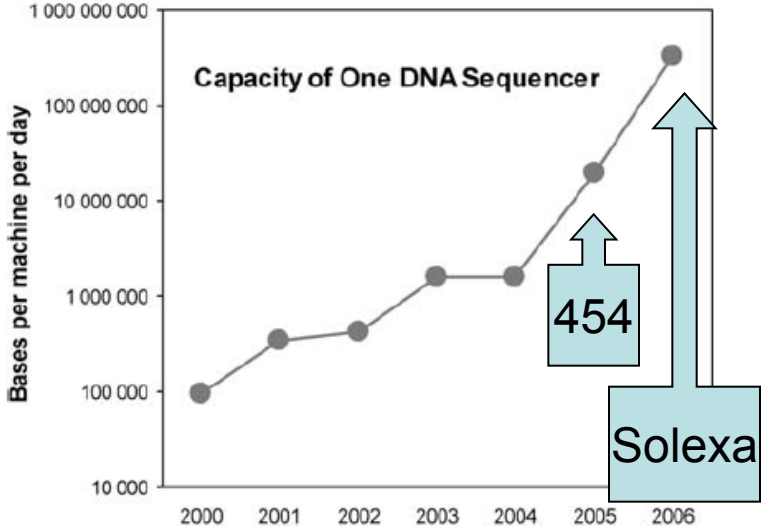
(sequencing by Oligonucleotide Ligation and Detection)

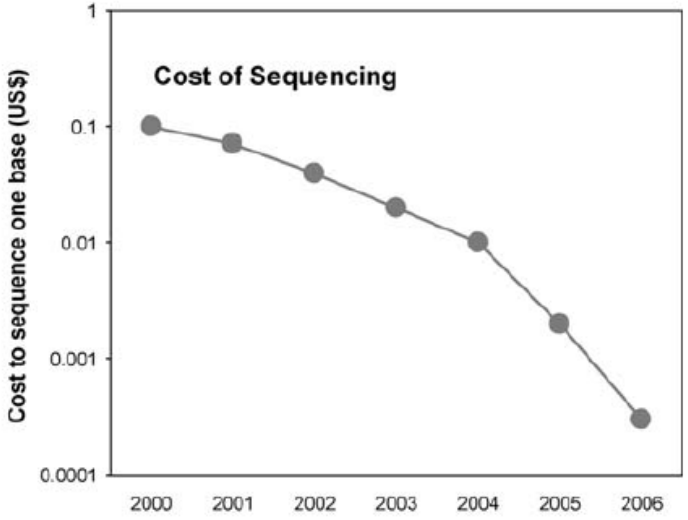
… a další (každého půlroku nová technologie – bouřlivý rozvoj !!!

	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	 *in situ* 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	 pyrosequencing	 reversible-terminator sequencing by synthesis	 Ligation of sequence-specific labeled oligos
Sequence detection	Chemiluminescence (one channel)	Fluorescence (four channel)	Fluorescence (four channel)
Read length and number	100–400 bp $> 2 \times 10^5$ reads	35 bp $\sim 4 \times 10^7$ reads	25 bp (paired) $> 10^7$ reads

„genomics era“



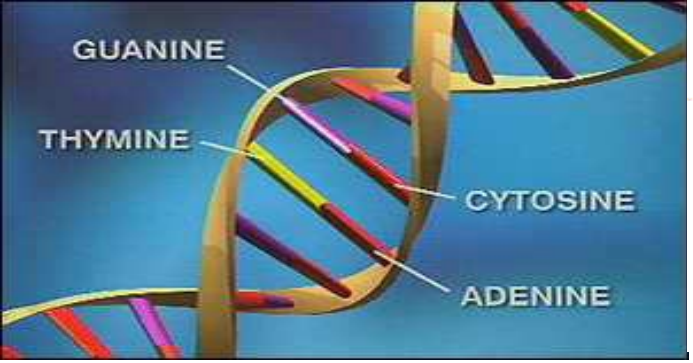


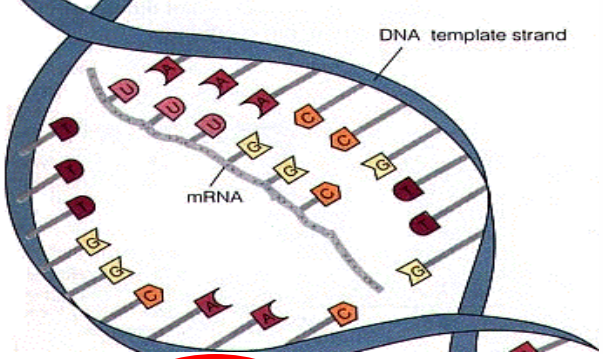


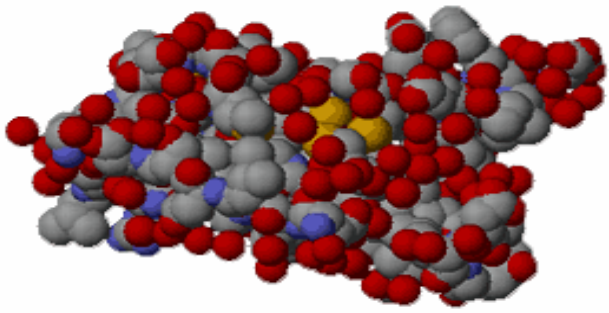
Budoucnost genetických metod
v ekologickém výzkumu

2. Analysis of expression by microarrays („transcriptomics")

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



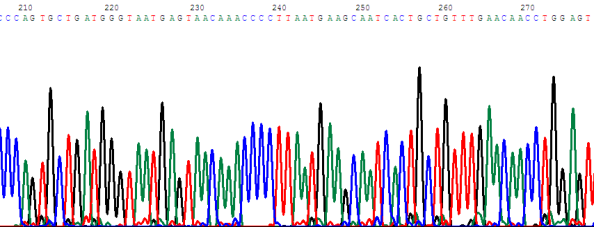


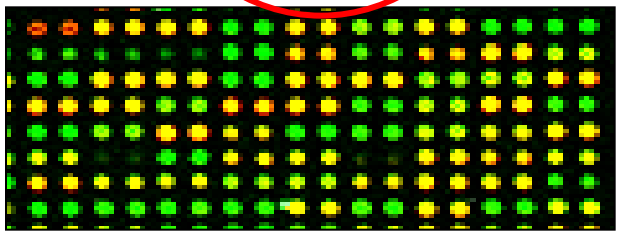


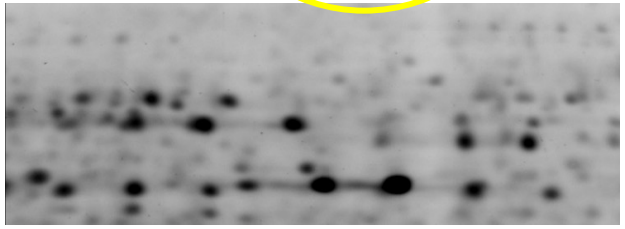
DNA → Transcription → mRNA → Translation → Proteins

Genome	Transcriptome	Proteome
3 billion bases	20-30,000 genes	~100,000 proteins
Genomics	Transcriptomics	Proteomics

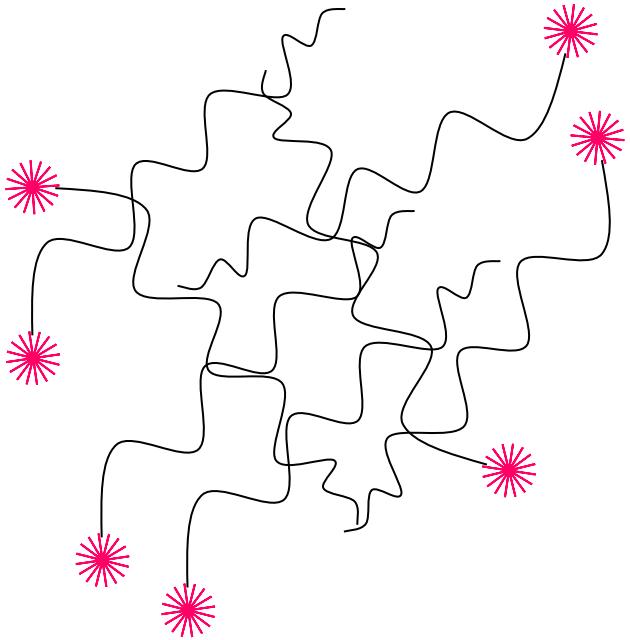
Functional Genomics

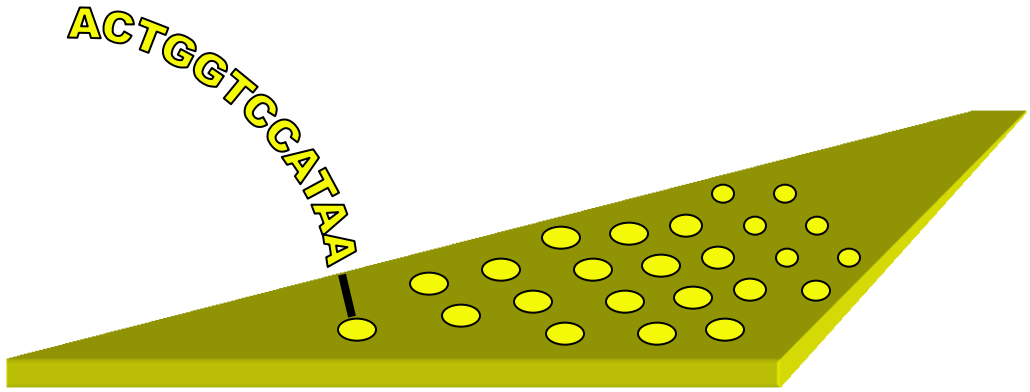






Microarray analysis of transcriptome
(~ specific DNA hybridization)





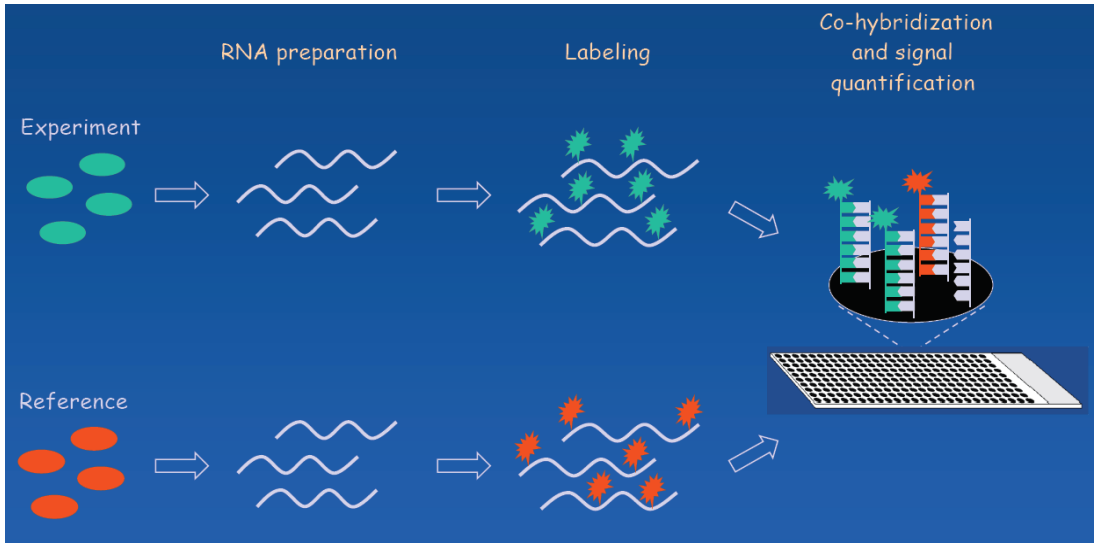
Target (i.e. mix of transcripts in a form of cDNA)

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

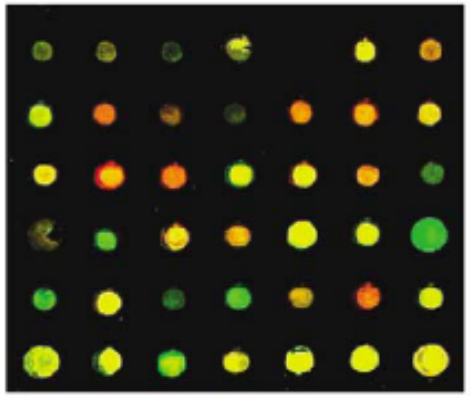
How to get a transcription profile

- vždy srovnání kontroly a „treatement“

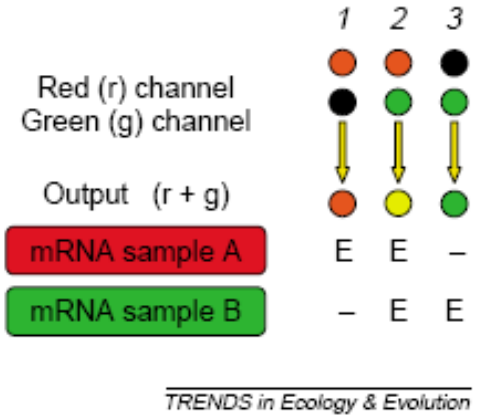
(a)



(b)



(c)



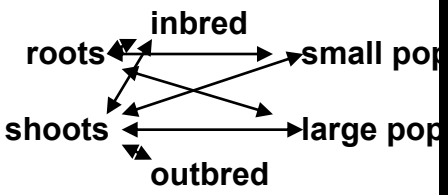
Analysis of expression level

Case study: Joop Ouborg et al.

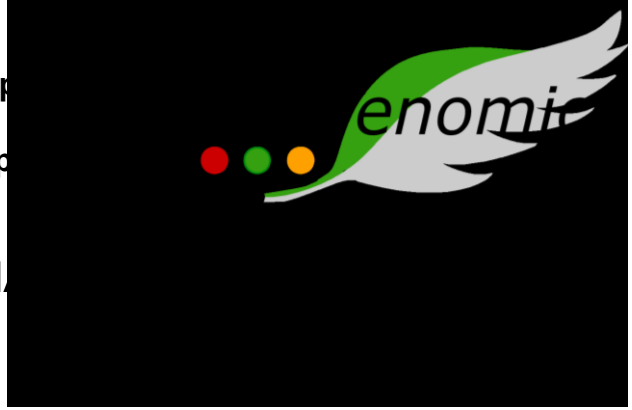
Transcriptional profiling of inbreeding depression and genetic erosion in *Scabiosa columbaria*: the balance between genetic drift and selection in the genetic erosion process.

Example:

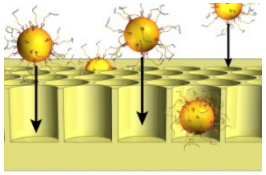
Scabiosa columbaria



cDN



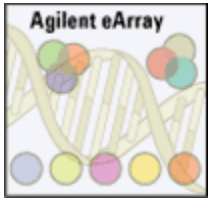
454 LIFE SCIENCES

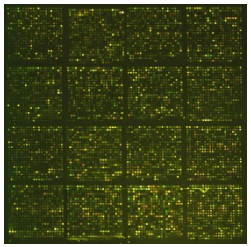


NCBI

530.000 sequences in one run, leading to ~ 40.000 ESTs

Agilent Technologies



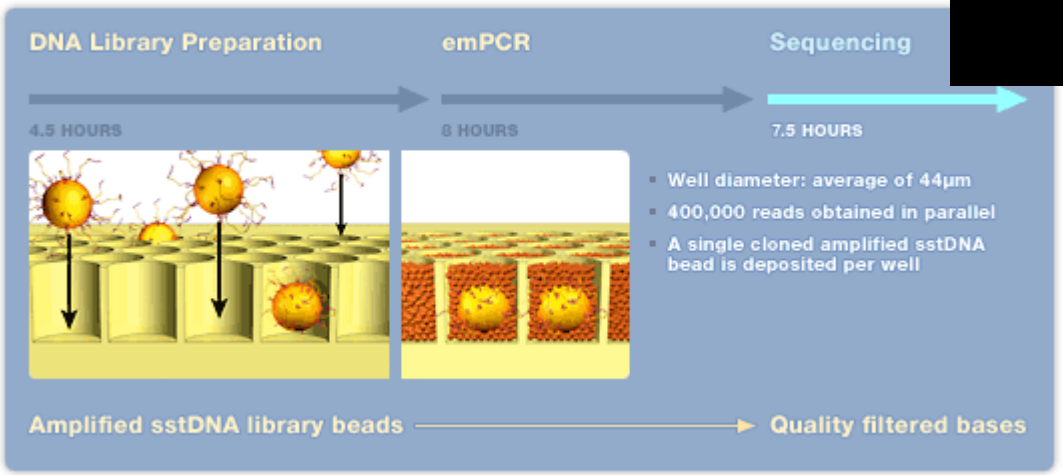


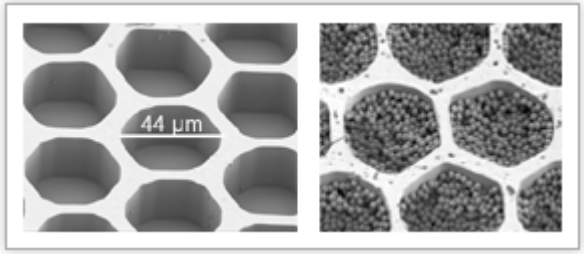
15k – 30k 60-mer microarrays

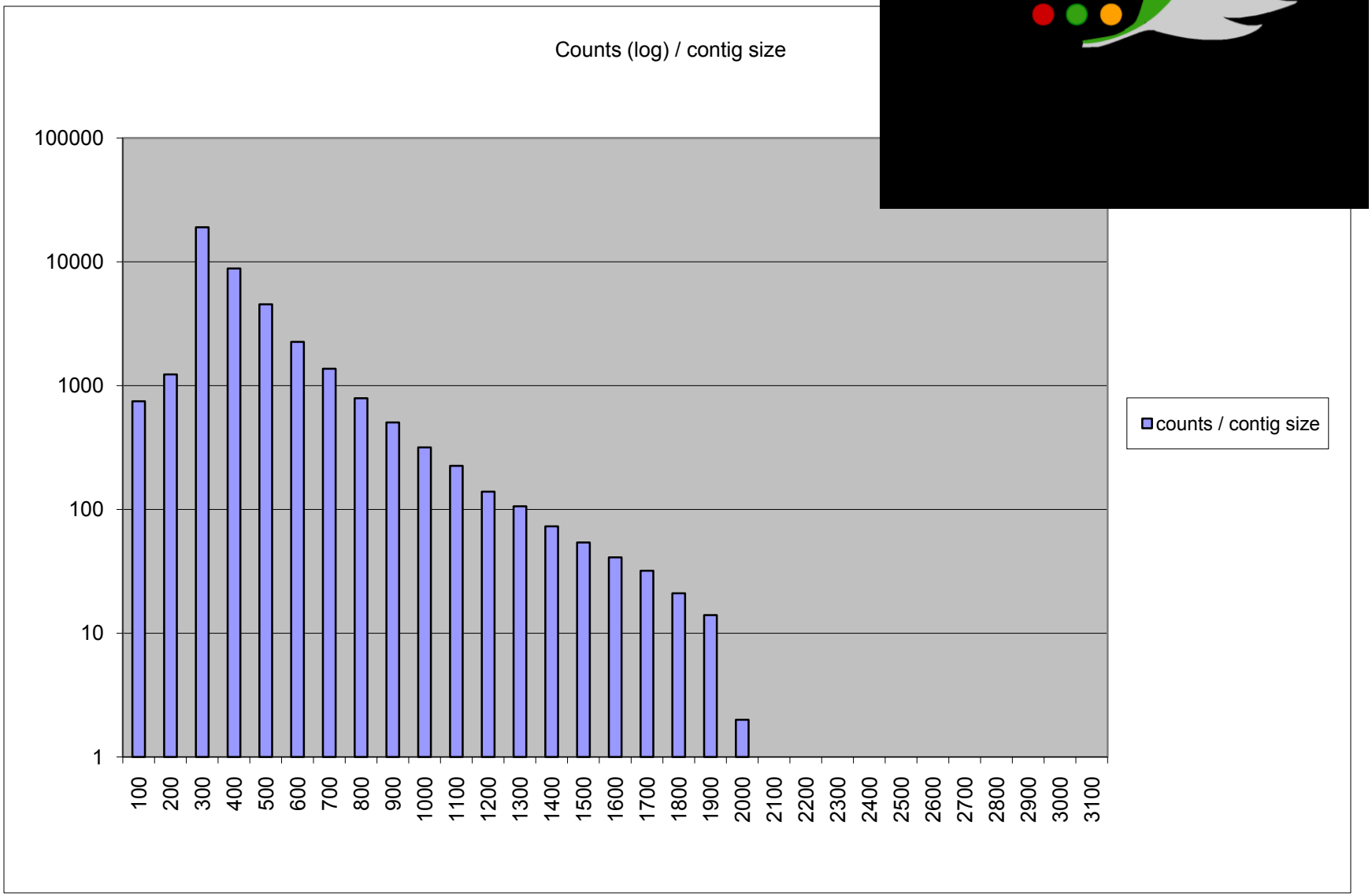
Experiment: transcriptional profiling of inbreeding depression

cDNA library preparation – 454 sequencing of transc

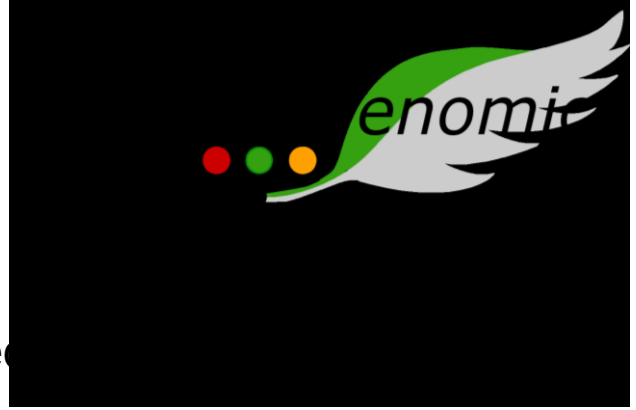
FIGURE 9







Total number of reads: **528557** Number of contigs: **40302**



In the next phase:

Annotation of these 40.000^{+} ESTs („expresse

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/)

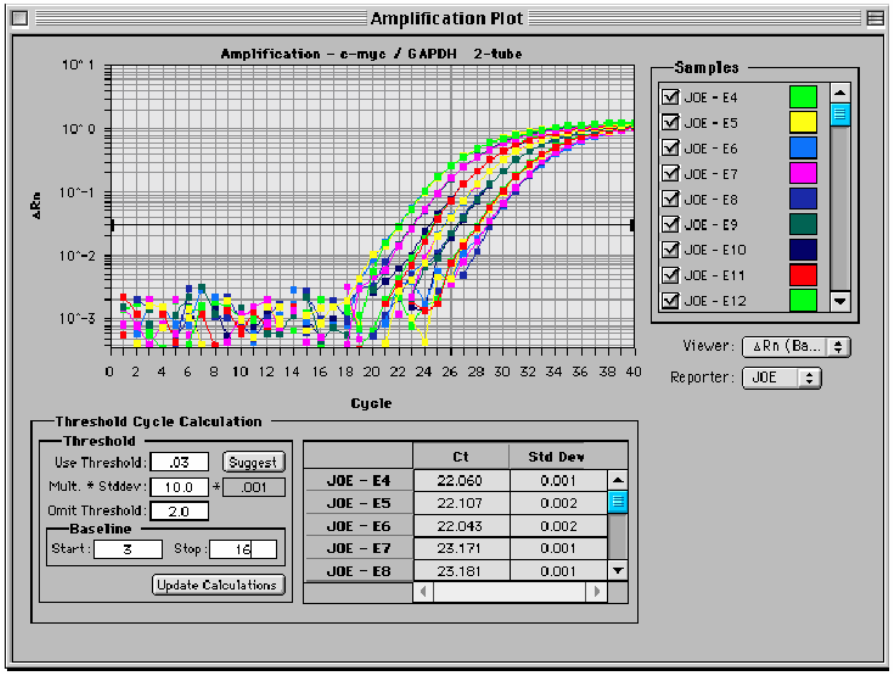
SNP markers: SNP-mining software like PolyBayes (http://genome.wustl.edu/tools/software/polybayes.cgi)

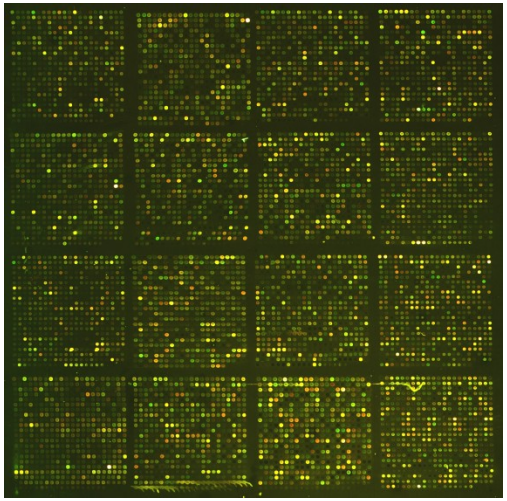
Again by using search software, freeware

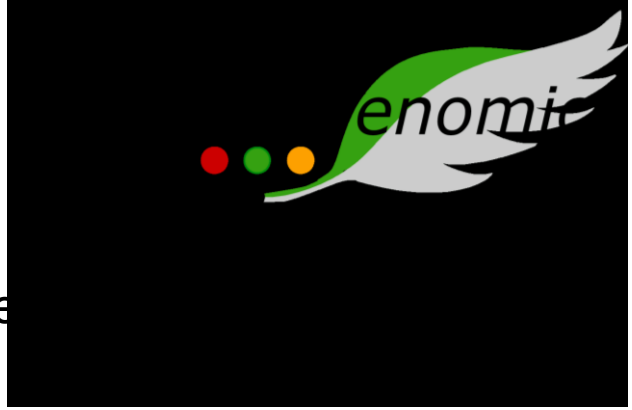
ALMOST HALF OF GENES (ESTs) ARE UNKNOWN !!!

1. Design of quantitative RealTime-PCR methods, bas
sequences

2. Design of a Scabiosa specific microarray

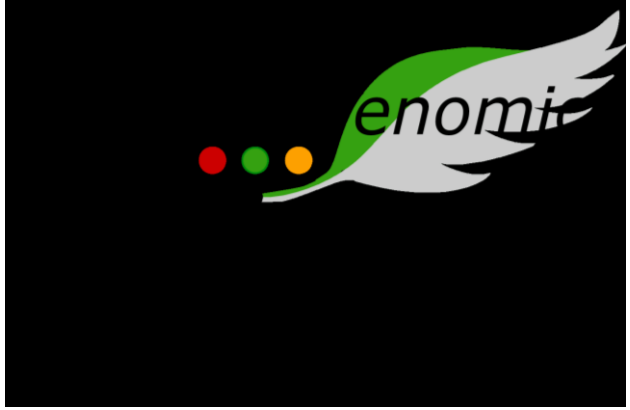






Expected pay-off:

- Ecogenomic approach to conservation gene
  effects of genetic erosion on functional genetic
- 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:

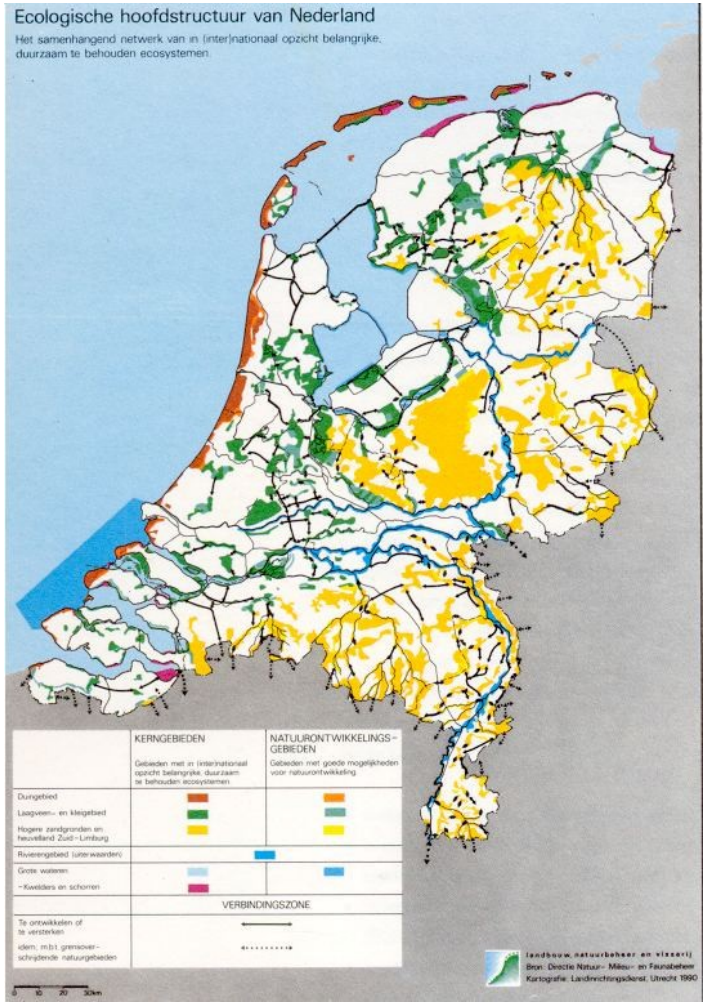
Costs are diminishing continuously

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 !!!