SINE, LINE, etc.

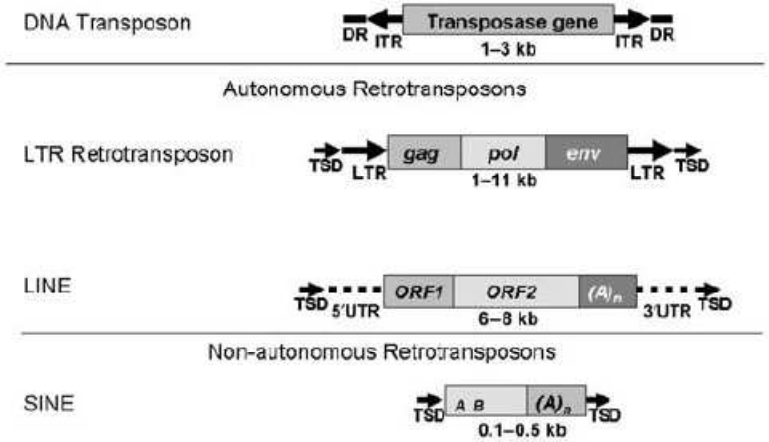
(Shedlock et al. 2004, TREE; Ray et al. 2007, MolEcol)

- **Transposable elements**
- Vytváří kopie (většinou)
- Kopie integrovány na nová místa v genomu
- Obvykle nejsou specificky odstraňovány
- Molekulární fosílie – neexistují homoplasie !!!
- Nesmírně početné
- Člověk – víc jak polovina genomu (ost. druhy – 40-90%)

Objev DNA transpozonů u kukuřice: Barbara McClintock

Typy transposabilních elementů

- **Kódující své proteiny, autonomní,** 1-10 kb
  - **DNA transposony** (cut-and-paste)
  - transposasa
  - **Retrotransposony** (copy-and-paste)
  - **LINE**
    1-2 proteiny, kopie přes RNA
  - **LTR retrotransposony**
    5-6 proteinů, také přes RNA

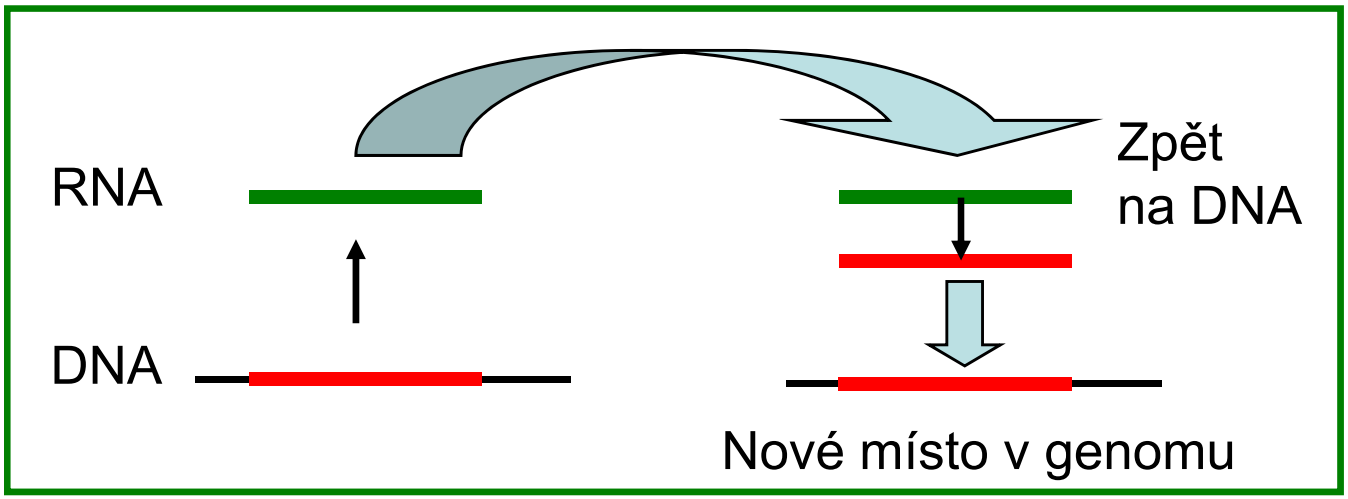


- **Nekódují proteiny, neautonomní,** 100-1000 bp
  paraziti předešlých, např. **SINE** (člověk *Alu* – více než 1 milion kopií) – nejčastěji používané v populačních a fylogenetických studiích

LINE – mechanismus transpozice

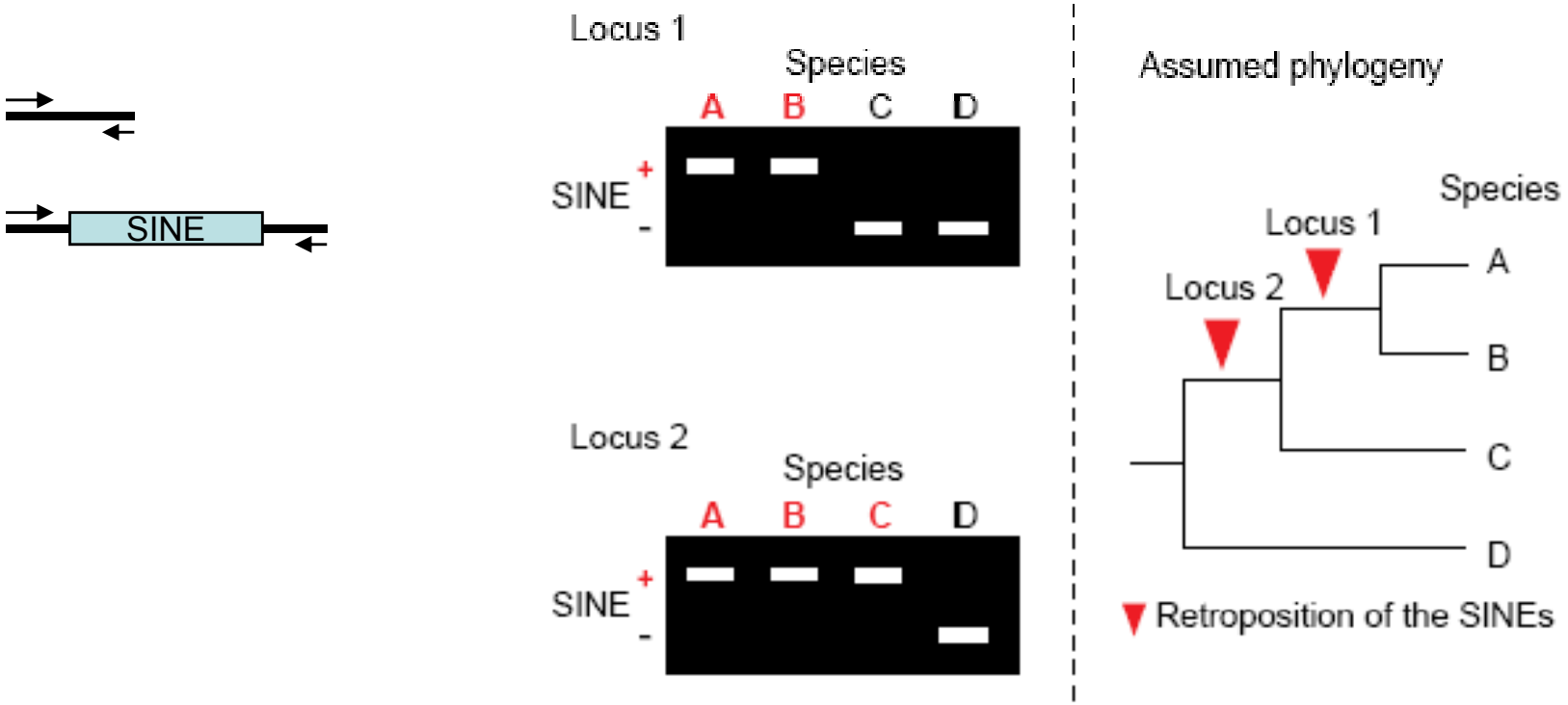
- Kopie přes RNA
- Reversní transkriptáza
- Mašinerii využívají **SINE** (jsou to „paraziti“),

  Alu (SINE) a *L1* (LINE) se stejně rychle množí



- **LTR retrotransposony – opět přes RNA, složitější proces**

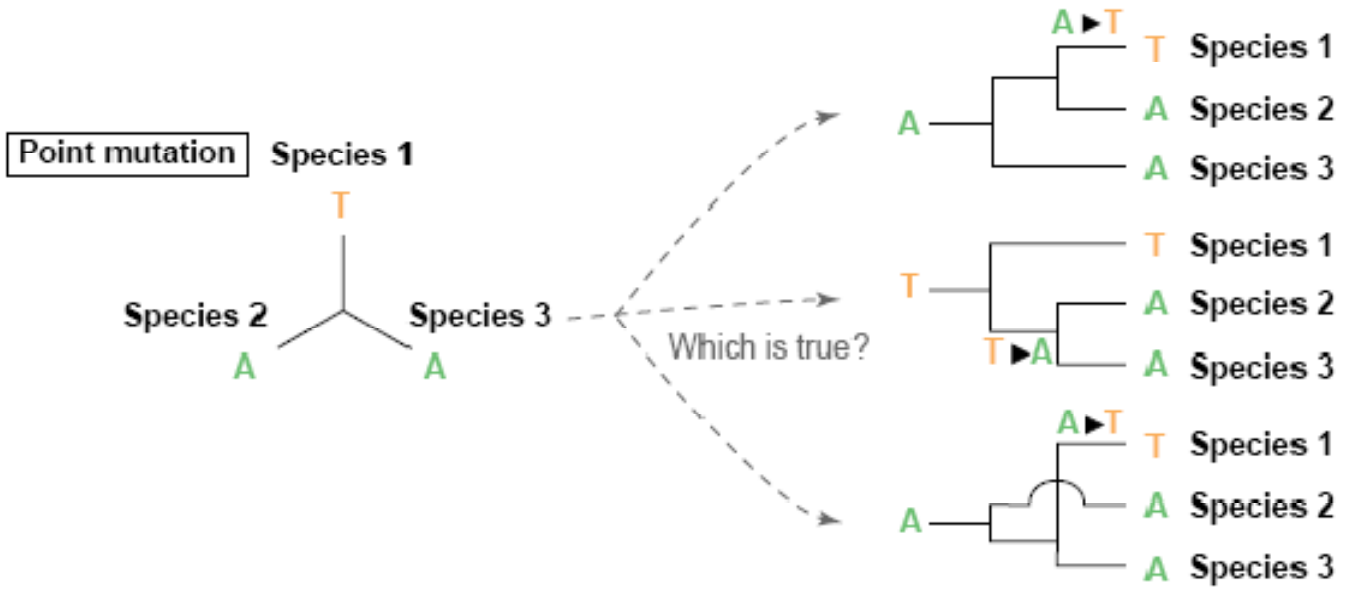
Velmi nízké riziko homoplázií → SINE = ideální fylogenetické markery



„single-locus marker“

- PCR amplifikace daného úseku a elektroforéza

Neexistují zpětné mutace = výhoda oproti sekvenačním datům



Příklad aplikace: kytovci vs. sudokopytníci (hroch je bratr velryby)

Budoucnost genetických metod v ekologickém výzkumu

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

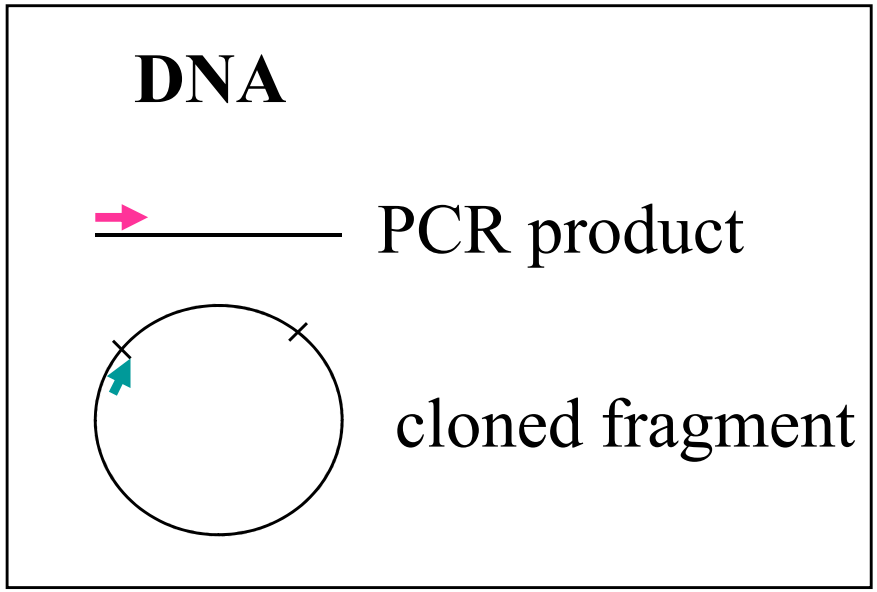
Molecular Ecology Resources (2008) 8, 3–17 doi: 10.1111/j.1471-8286.2007.02019.x

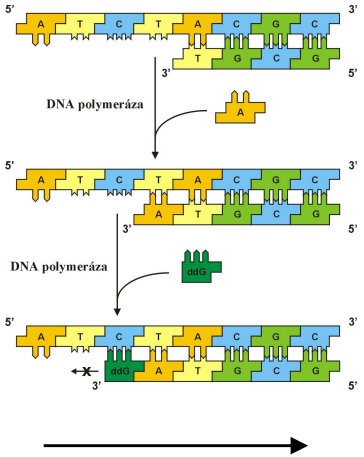
TECHNICAL REVIEW

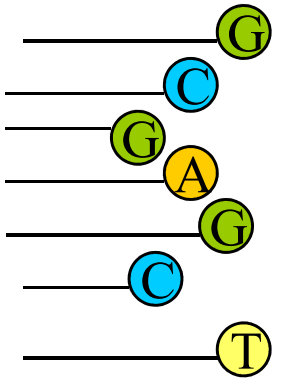
Sequencing breakthroughs for genomic ecology and evolutionary biology

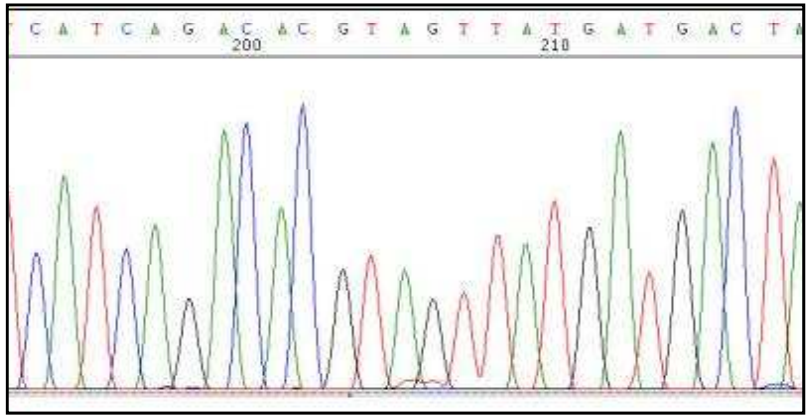
MATTHEW E. HUDSON
Department of Crop Sciences, University of Illinois, Urbana, 334 NSRC, 1101 W. Peabody Blvd., IL 61801, USA

Sequencing - Sangerova metoda









detector

laser beam

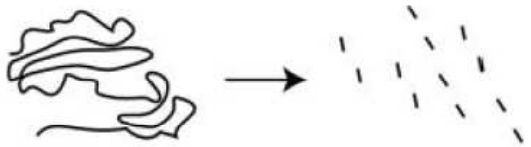
capillary electrophoresis

+

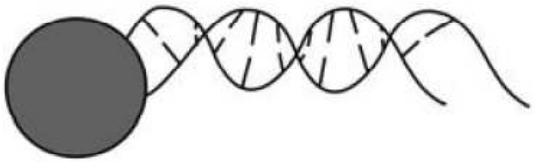
-

„Next generation sequencing"

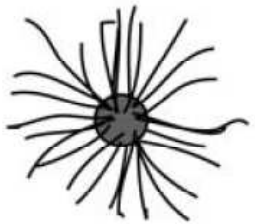
1) Randomly fragment many molecules of target DNA



2) Immoblize individual DNA molecules on solid support

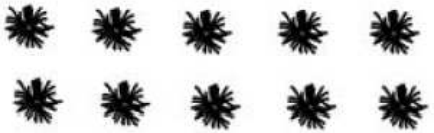


3) Amplify DNA in clonal 'polymerase colony'

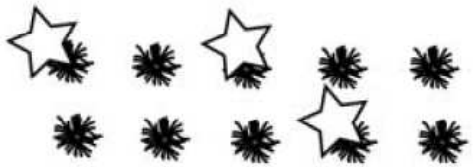


„polonies"
(polymerase colonies)

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 při inkorporaci bazí
- První generace GS20
  → 200 000 reakcí najednou
  (zhruba 20 milionů bp)
  dnes FLX → 400 000 reakcí najednou
- Problémy s homopolymery
- Délka jednotlivých sekvencí 100 – 400

Molecular Ecology (2008) 17, 1629–1635

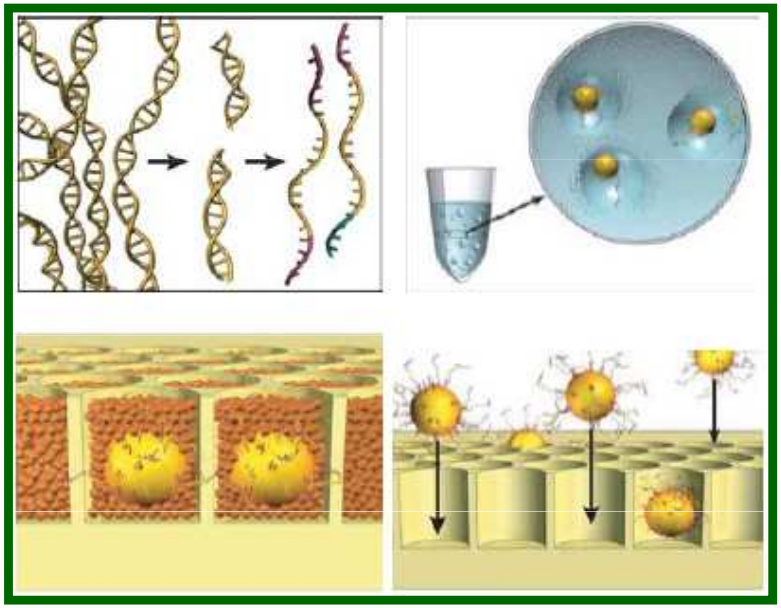
NEWS AND VIEWS

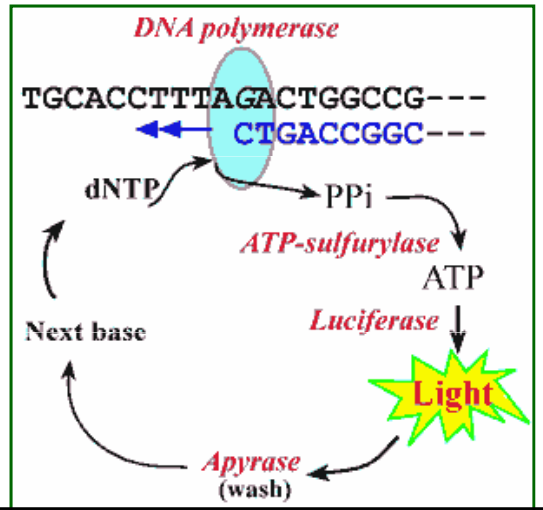
PERSPECTIVE

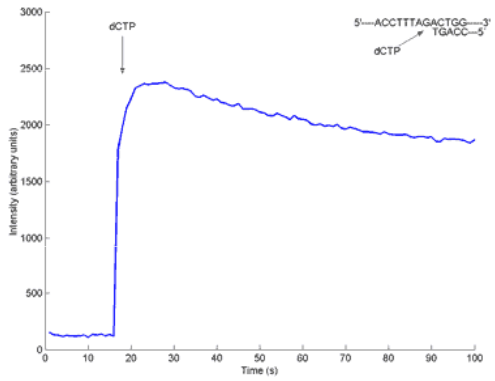
Sequencing goes 454 and takes large-scale genomics into the wild

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

1 600 000 well plate



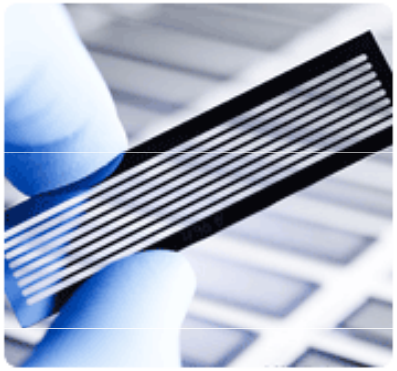




Solexa/Illumina 1G SBS technology

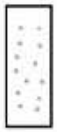
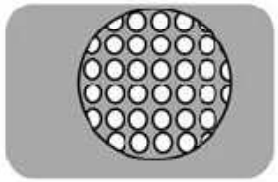
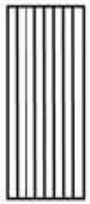
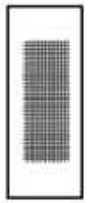
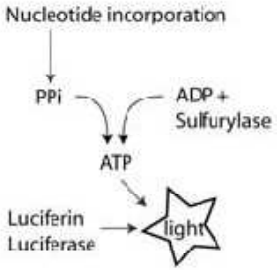
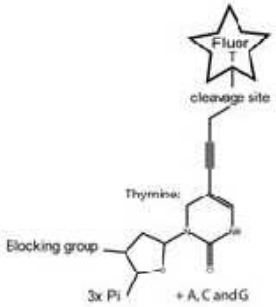
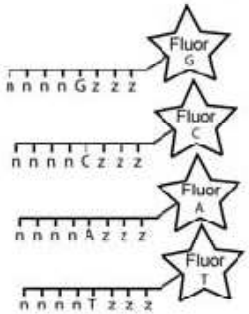
(SBS = sequencing by synthesis)

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

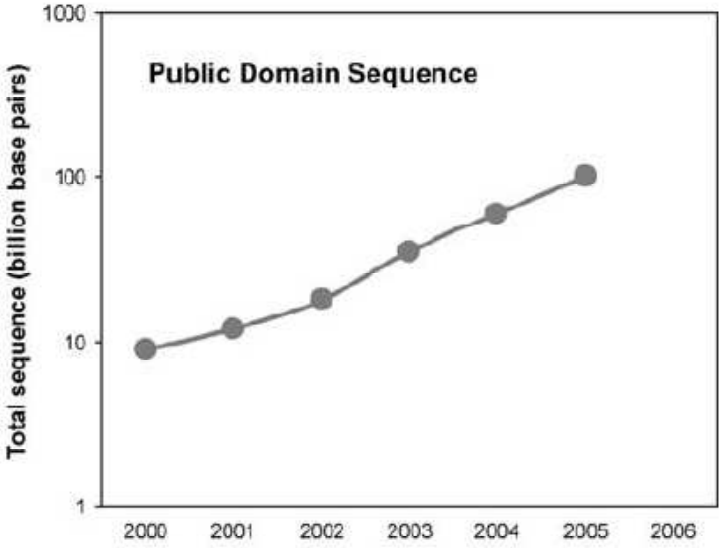


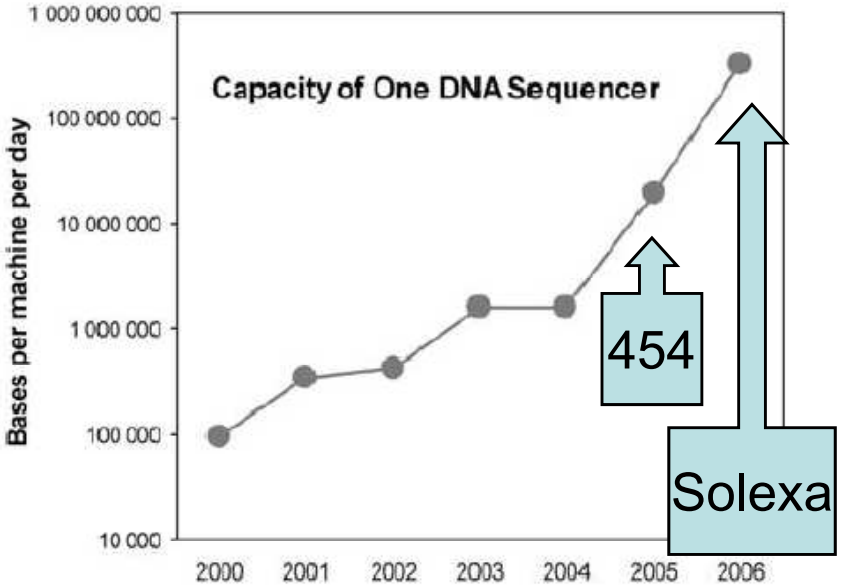
SOLiD

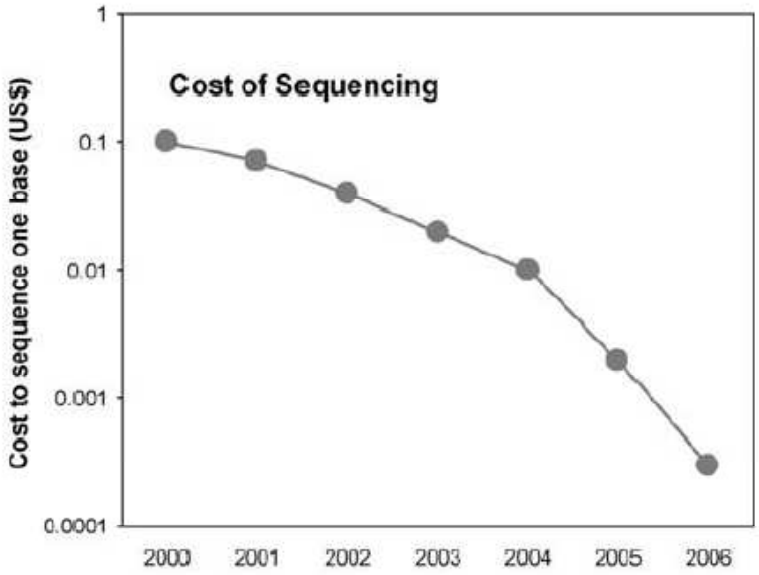
(sequencing by Oligonucleotide Ligation and Detection)

	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“



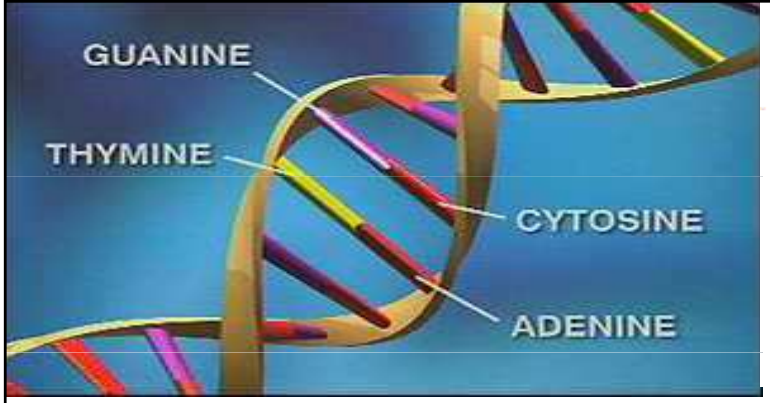


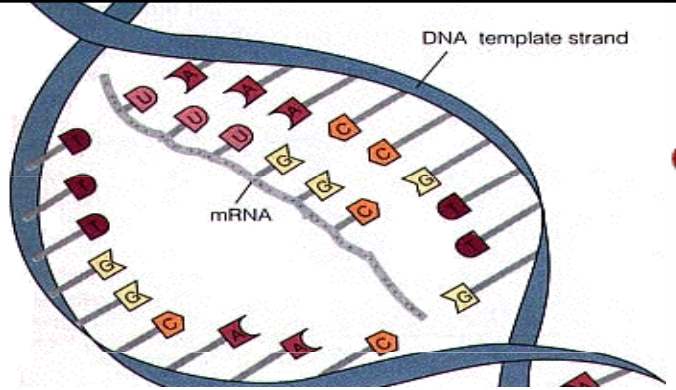


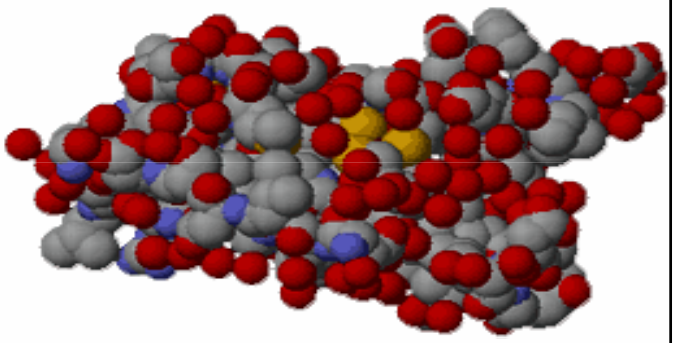
Budoucnost genetických metod
v ekologickém výzkumu

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

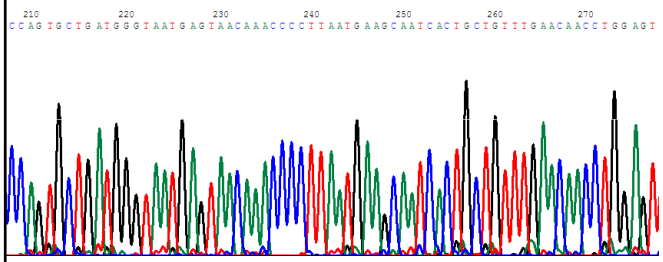
Genome — 3 billion bases

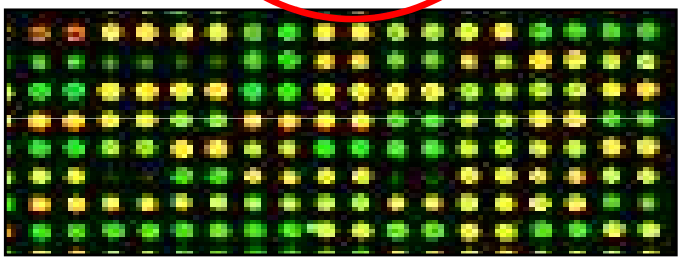
Transcriptome — 20-30,000 genes

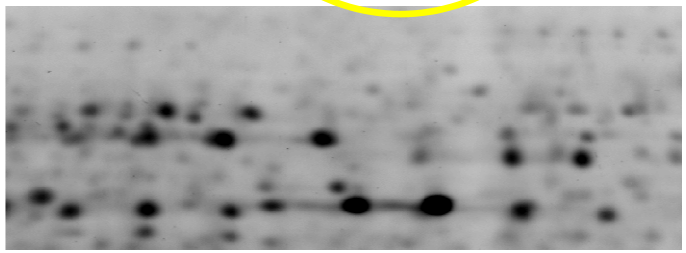
Proteome — ~100,000 proteins

Functional Genomics

Genomics — Transcriptomics — Proteomics

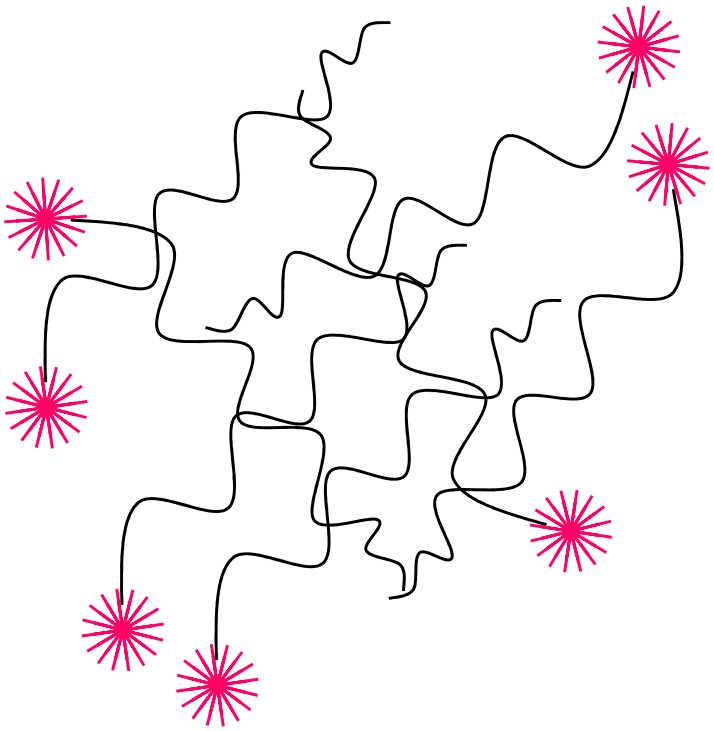




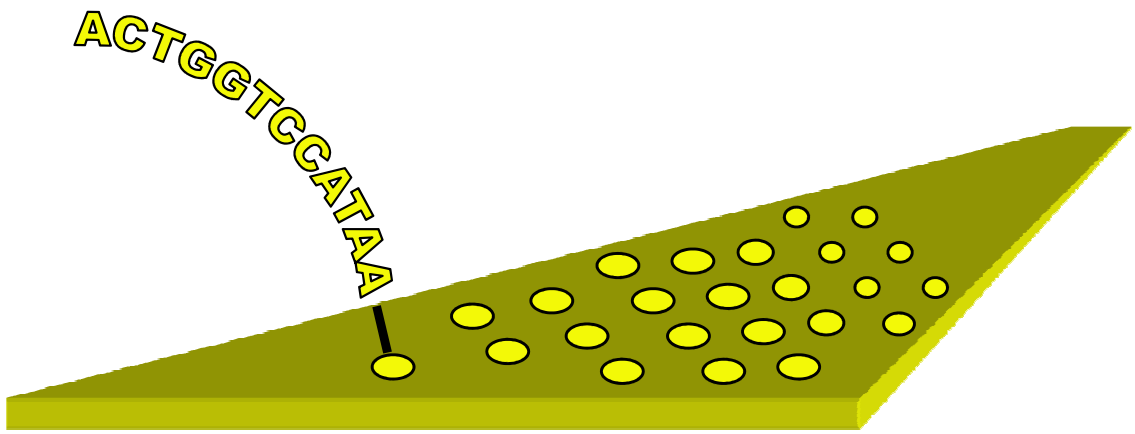


Microarray analysis of transcriptome

(~ specific DNA hybridization)



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

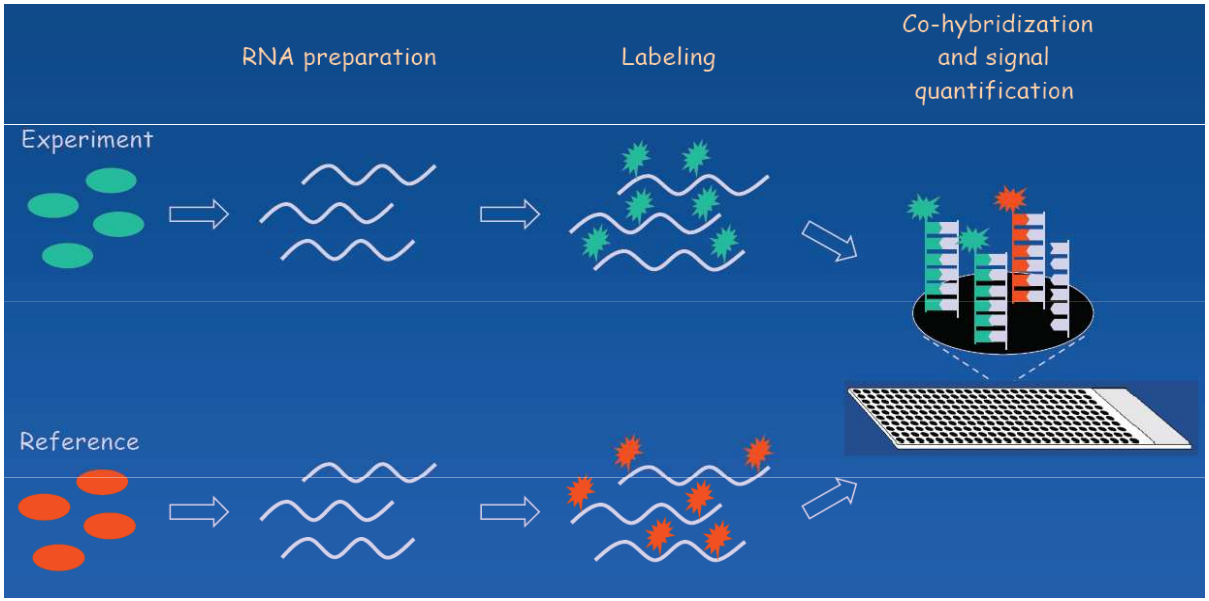


Probe (i.e. synthesized oligonucleotides complementary to particular genes)

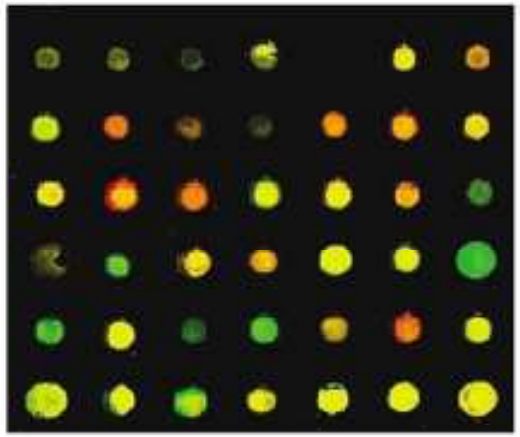
How to get a transcription profile

- vždy srovnání kontroly a „treatement“

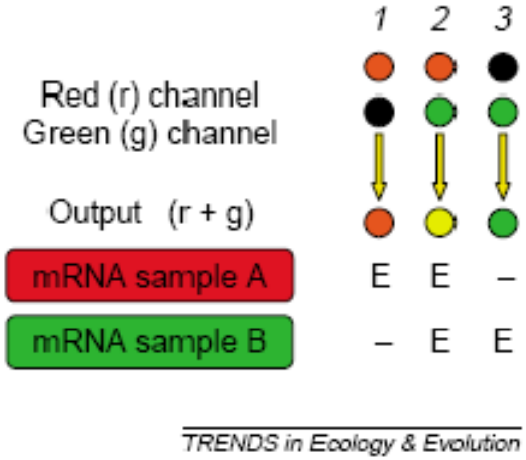
(a)



(b)



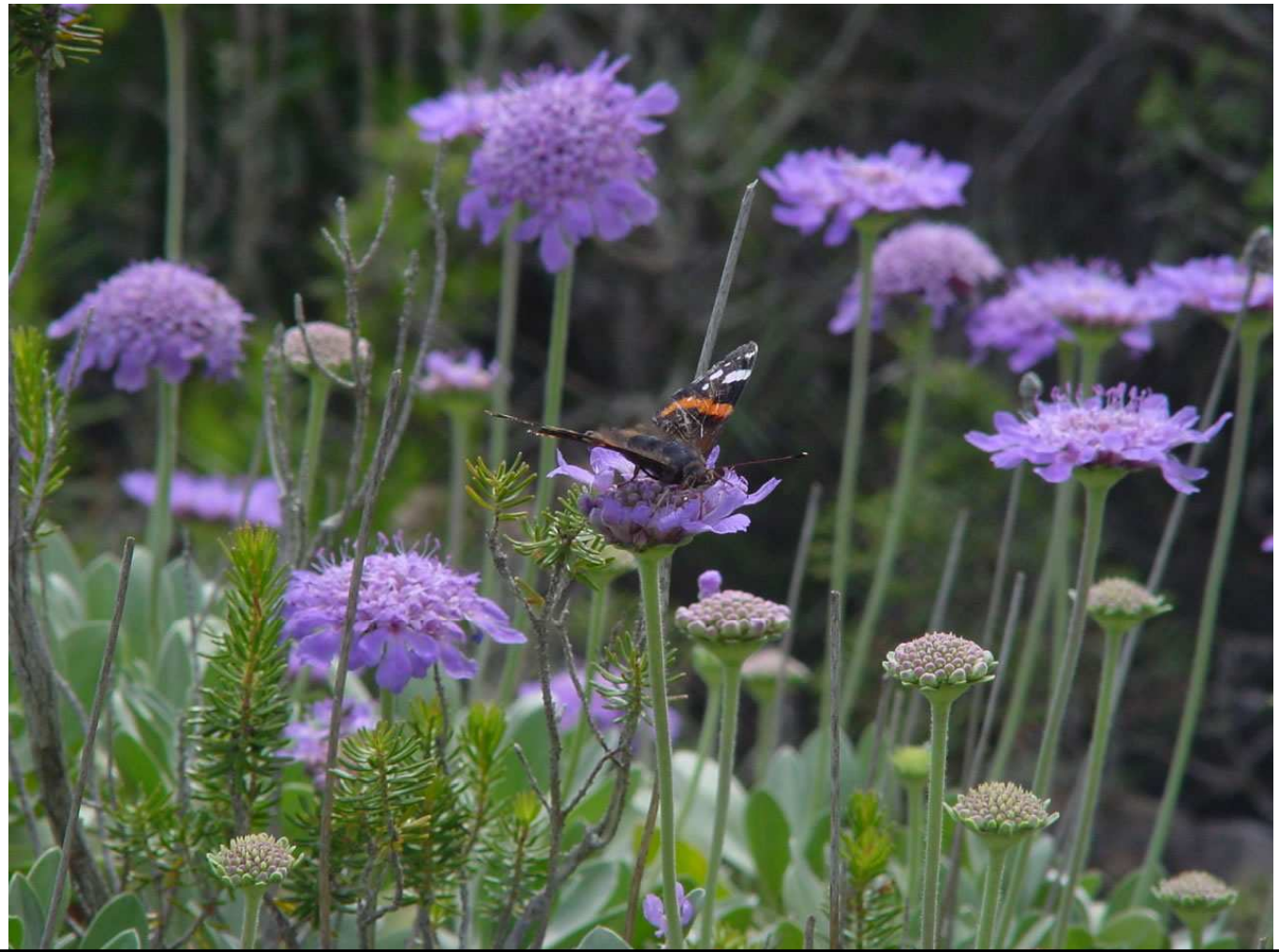
(c)



Analysis of expression level

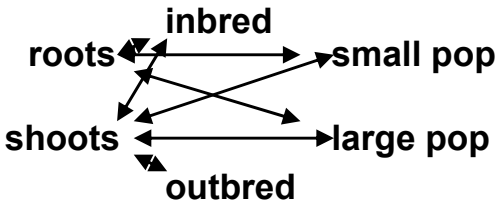
Case study: Joop Ouborg et al.

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

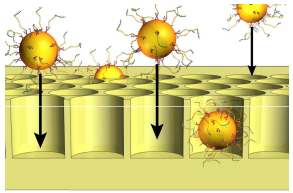


Ecogenomics

cDNA library

530.000 sequences in one run, leading to ~ 40.000 ESTs

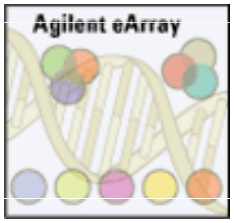
454 LIFE SCIENCES

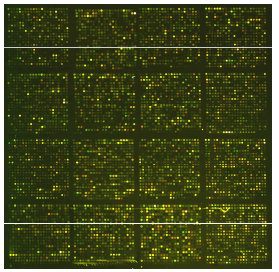


NCBI

15k – 30k 60-mer microarrays

Agilent Technologies

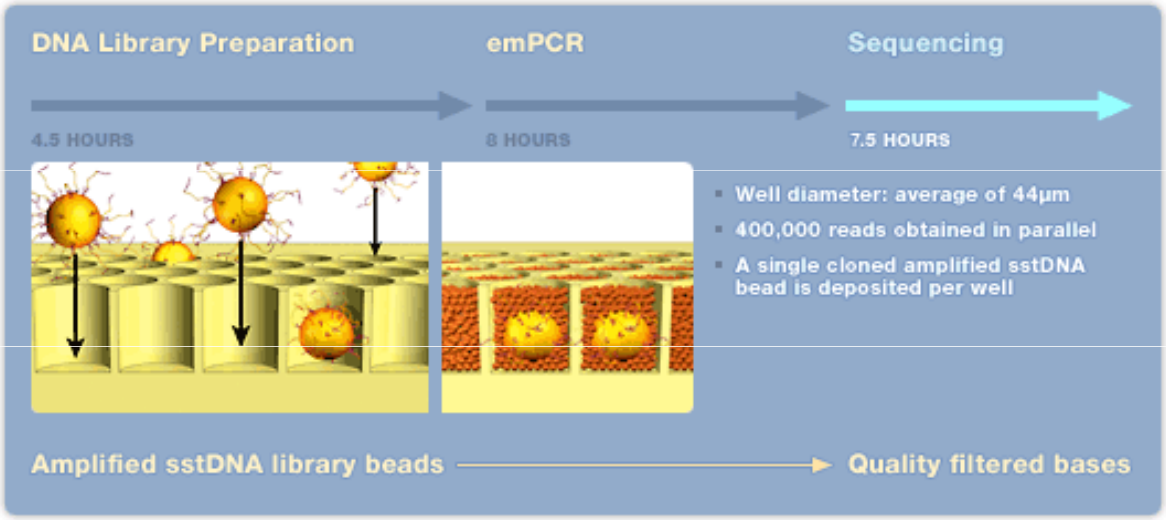


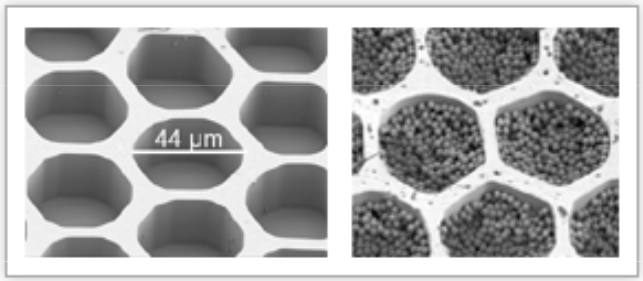


Experiment: transcriptional profiling of inbreeding depression

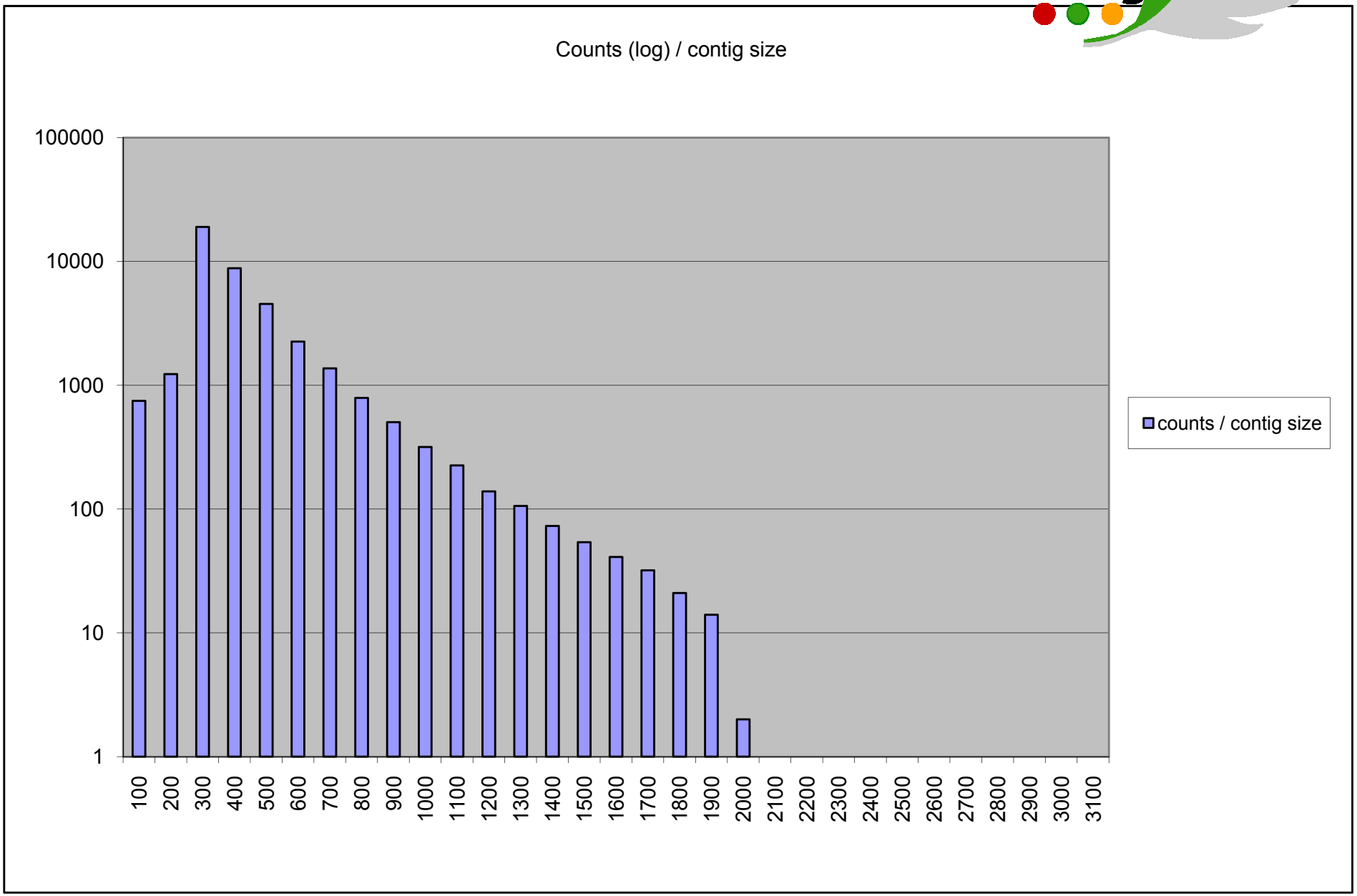
cDNA library preparation – 454 sequencing of transcriptome

FIGURE 9





Ecogenomics



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

In the next phase:

Annotation of these 40.000+ ESTs („expressed sequence tags“)

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)

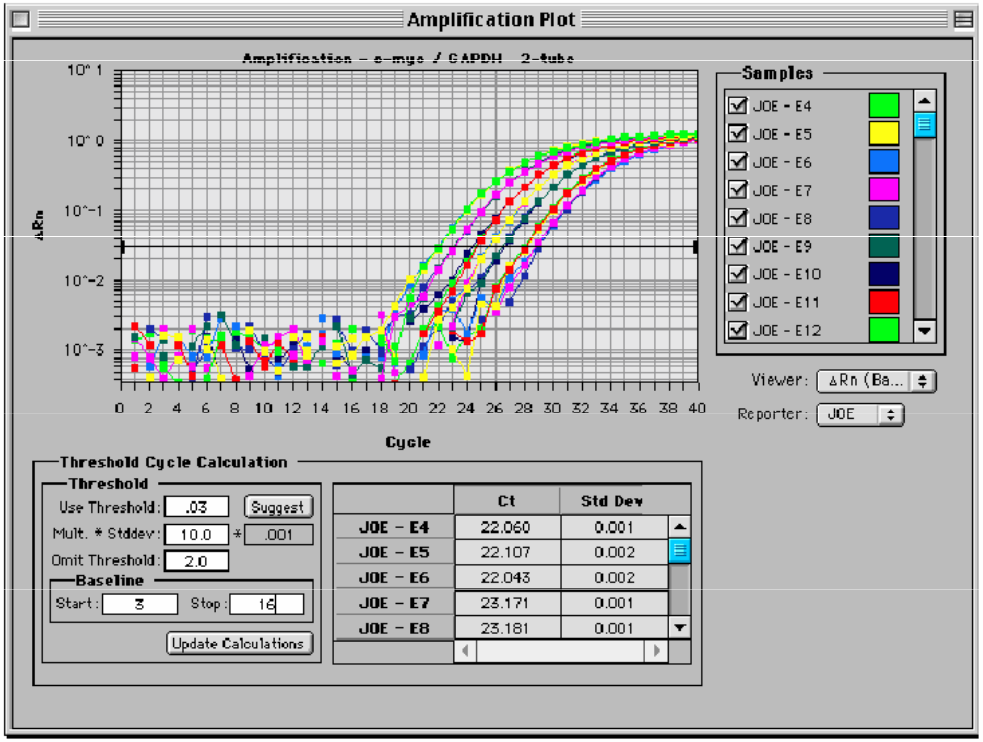
Again by using search software, freeware

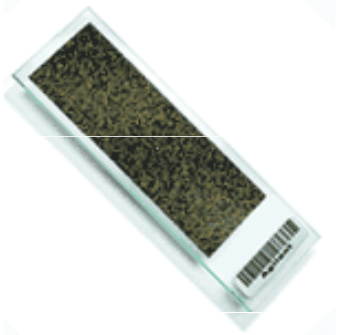
ALMOST HALF OF GENES (ESTs) ARE UNKNOWN !!!

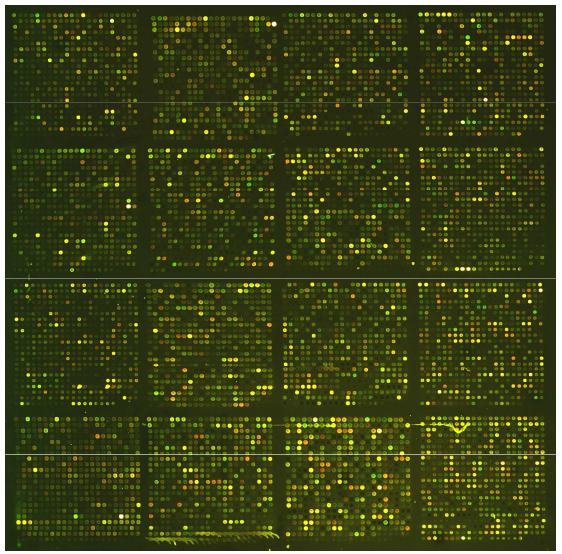
Ecogenomics

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

2. Design of a Scabiosa specific microarray







Ecogenomics

Expected pay-off:

- Ecogenomic approach to conservation genetics leads to insight in effects of genetic erosion on functional genetic variation
- 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**?

Ecogenomics

Costs/requirements:

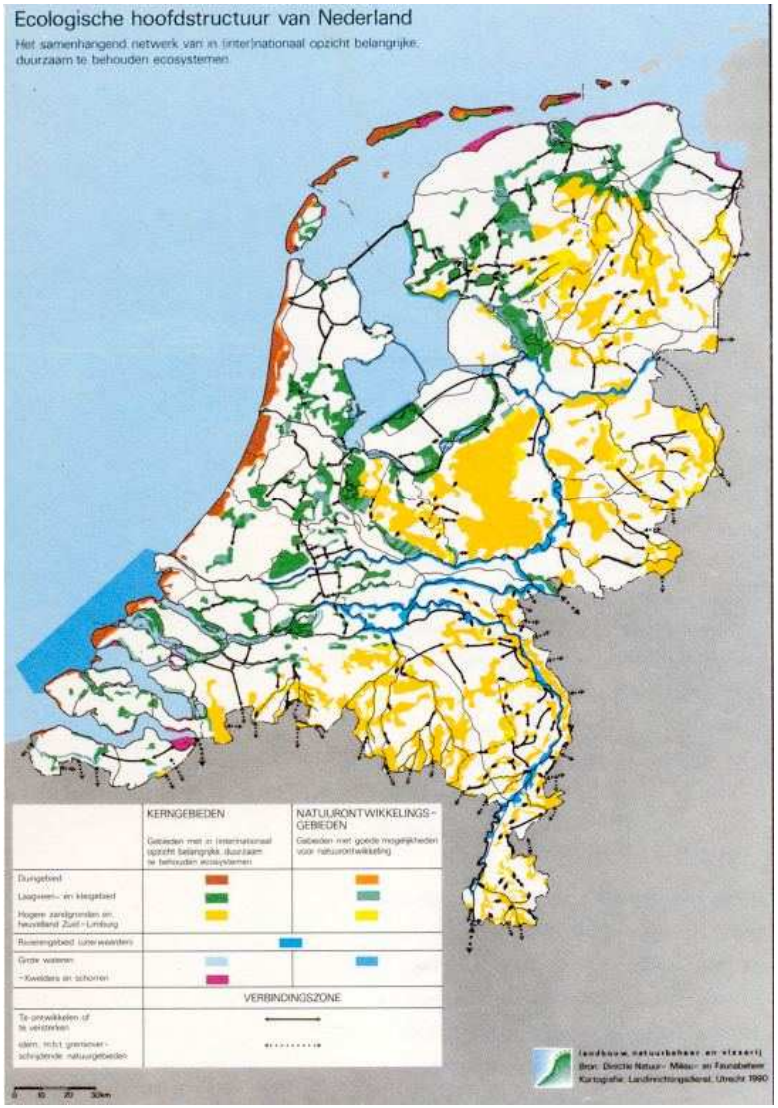
Costs are diminishing continuously

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 !!!