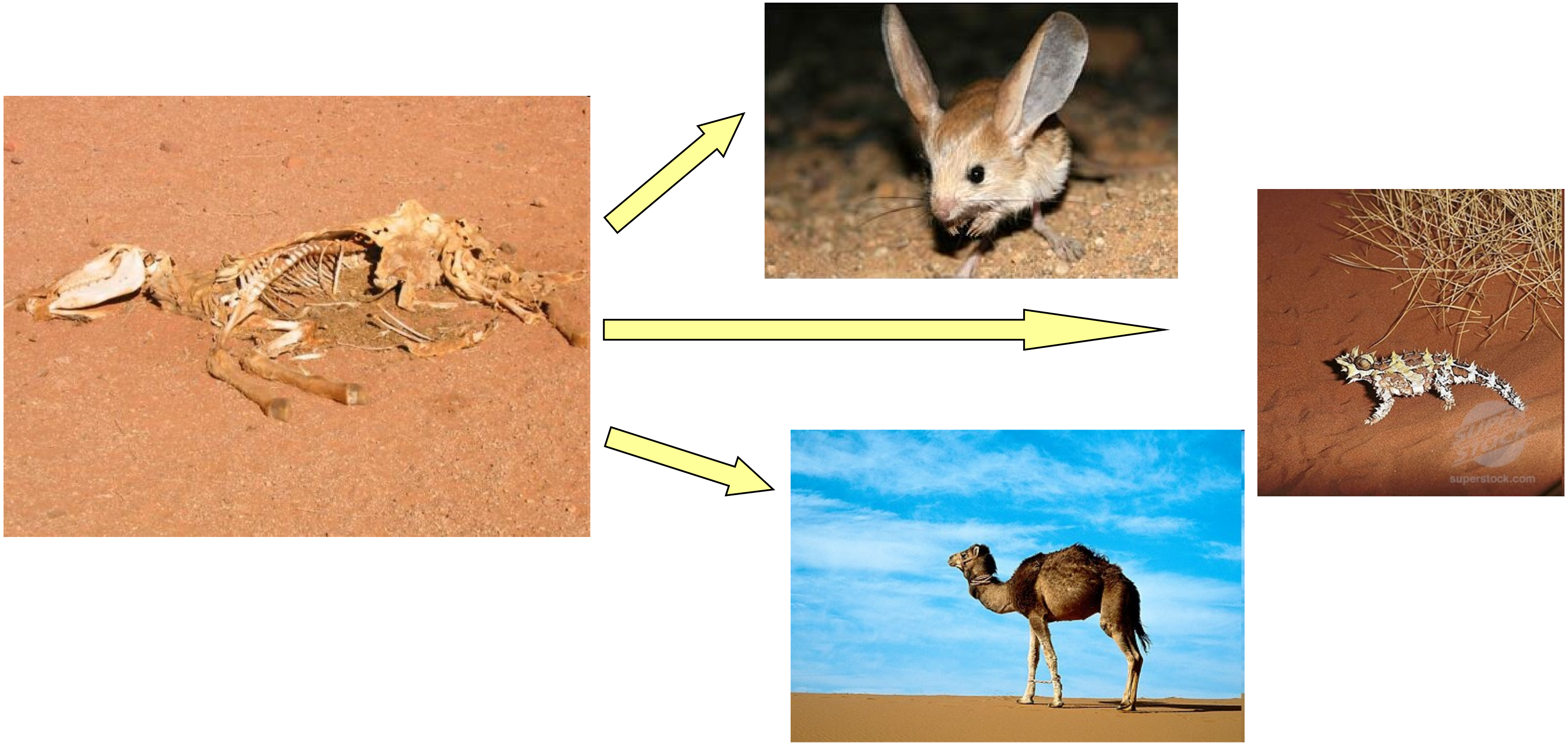
Exprimované geny a přírodní selekce

Produkty funkčních genů a jejich význam v ekologických studiích

Geny a adaptace

- studium selekčních tlaků daných prostředím a evoluční odpovědi na ně → vznik **adaptací**, tj. **geneticky** podmíněné přizpůsobení se prostředí (vs. fenotypová plasticita)
- **např. interakce s abiotickým prostředím**

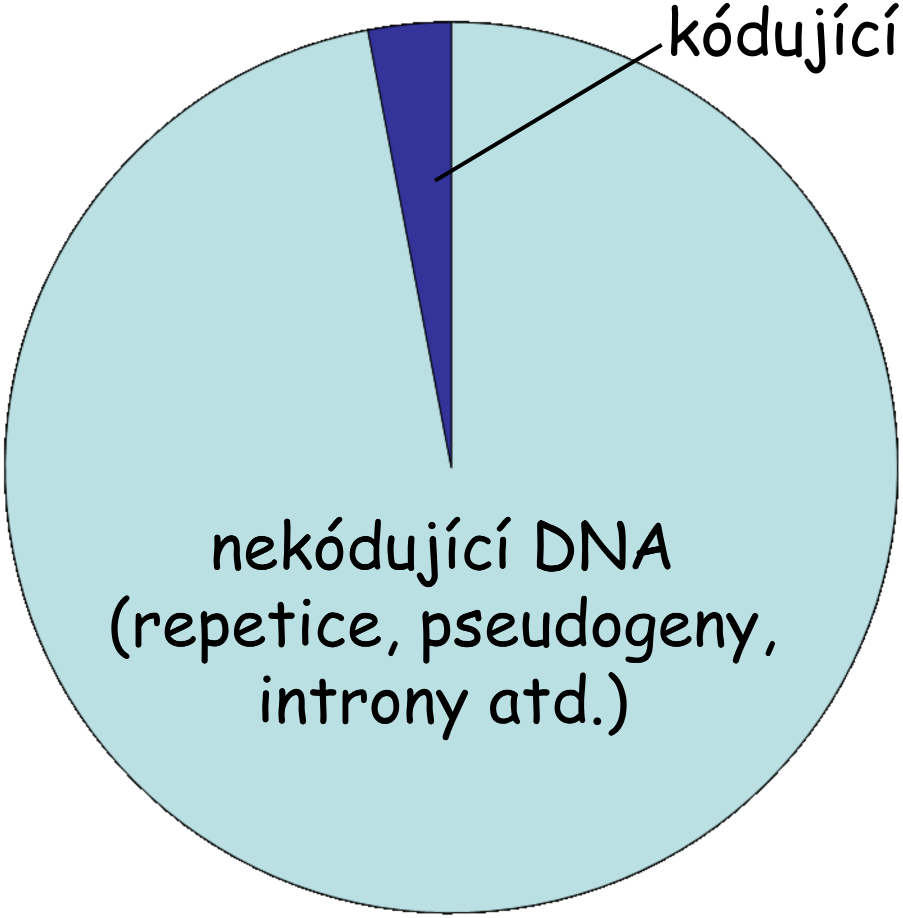


Proč geny v molekulární ekologii?

- Geny mají funkční význam – geneticky determinovaný polymorfismus
- -> studium **proximátních mechanismů**

- Př.: Proč je samec hýla rudého červeně zbarven?
- **ultimátní vysvětlení** – aby se líbil samicím a zplodil s nimi více potomků
- **proximátní vysvětlení** – protože karotenoidy získané z potravy ukládá více do peří a méně je používá v imunitní odpovědi (protože má dobré geny)

Funkční vs. neutrální genetická variabilita



kódující DNA = funkční geny

↓

fenotyp

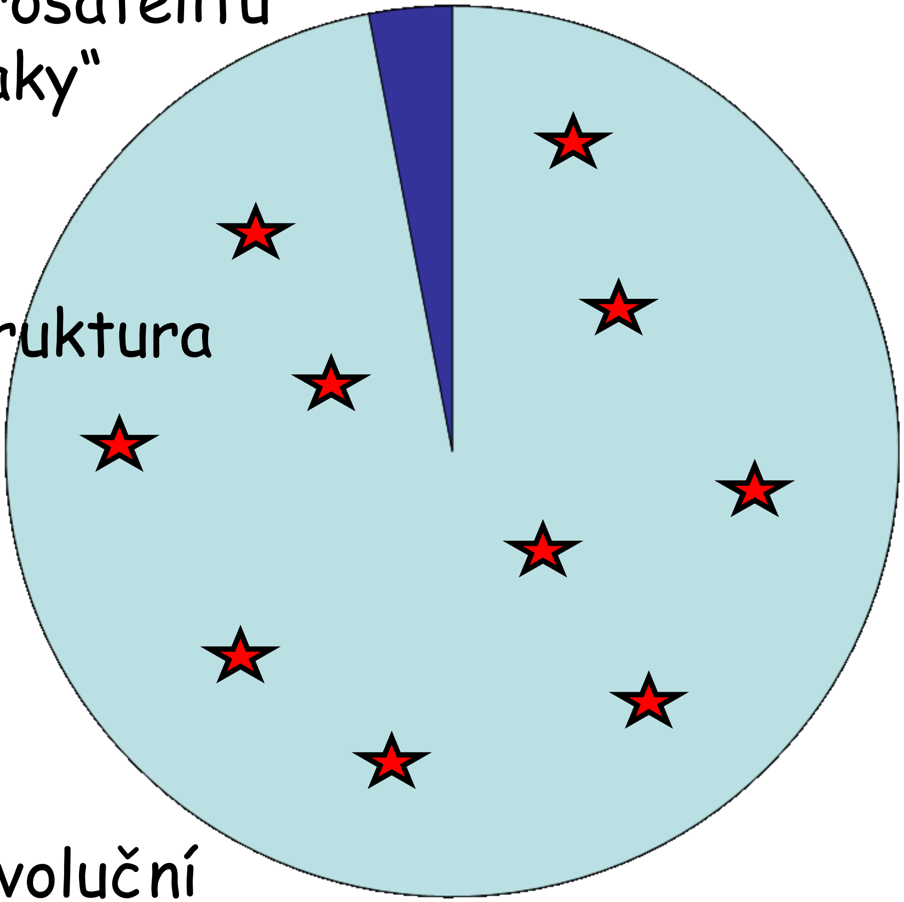
↓ přírodní výběr

adaptace ke
specifickému
prostředí

97% lidské DNA nic nekóduje!!!

Jak relevantní je informace získaná z genetických dat

Příklad: 10 microsatelitů
= „neutrální znaky"



- ✓ populačně-genetická struktura
- ✓ inbreeding
- ✓ bottleneck

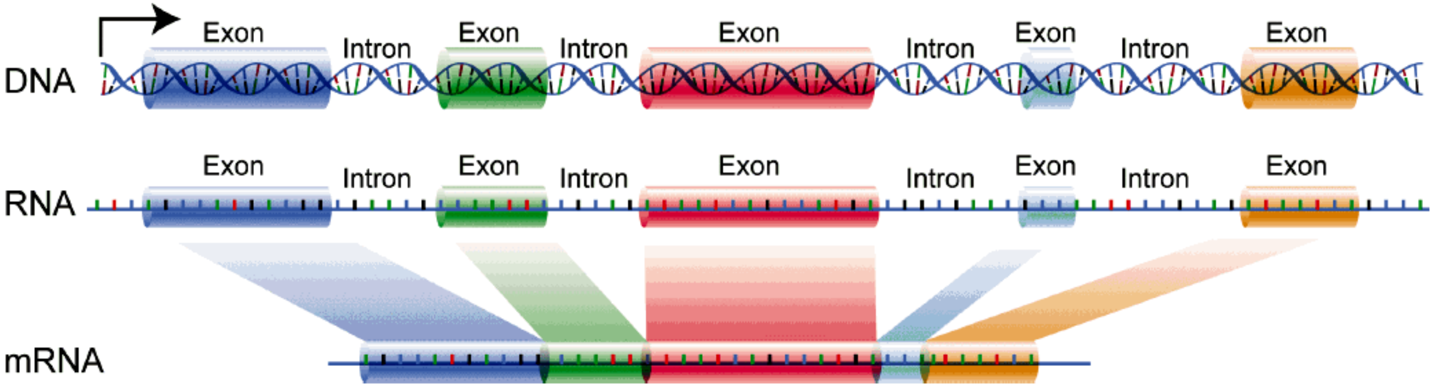
- ✗ adaptace
- ✗ proximátní evoluční mechanismy

↓

a priori neutrální k působení přírodního výběru

Struktura genů

- Exons → protein coding, under selection
- Introns → non-coding, neutral
- Intergenic regions → non-coding, neutral



Degenerovaný genetický kód

GCT → Alanine

GCA → Alanine

GCC → Alanine

GCG → Alanine

CCT → Proline

GGA → Glycine

→3rd position evolves neutrally

→ 1st and 2nd position under selection

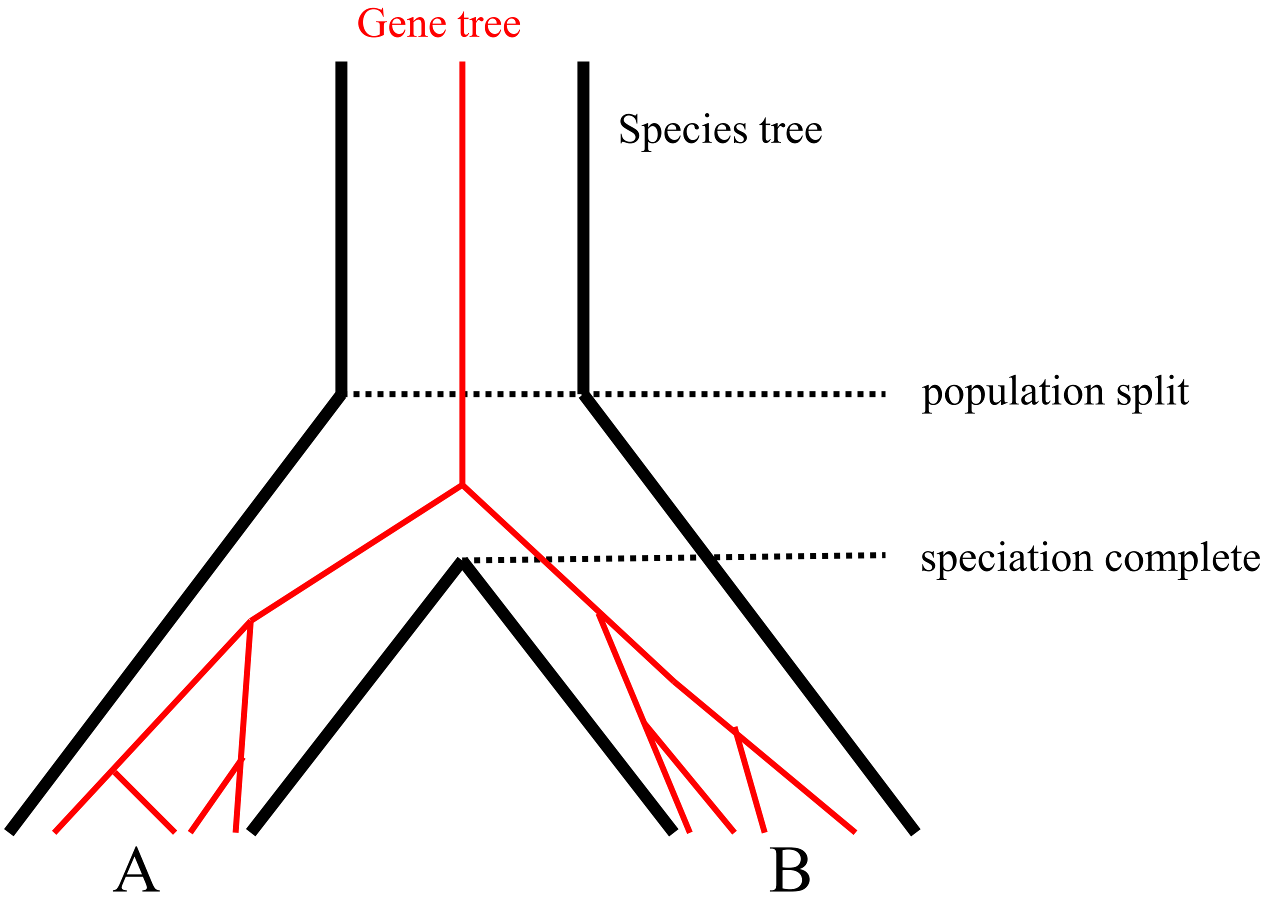
Jak poznat, že na daný gen působí selekce?

- **Fylogenetická analýza** (detekce na úrovni sekvencí)
- **Populačně-genetická analýza** (detekce na úrovni frekvence alel)

Jak poznat, že na daný gen působí selekce?

- Gene-tree versus species tree

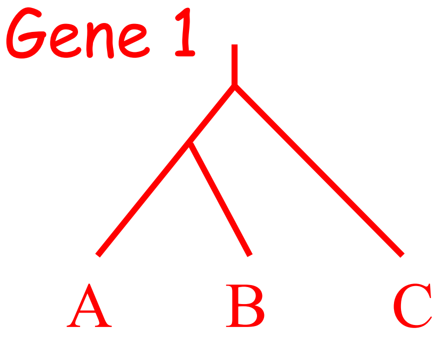
-

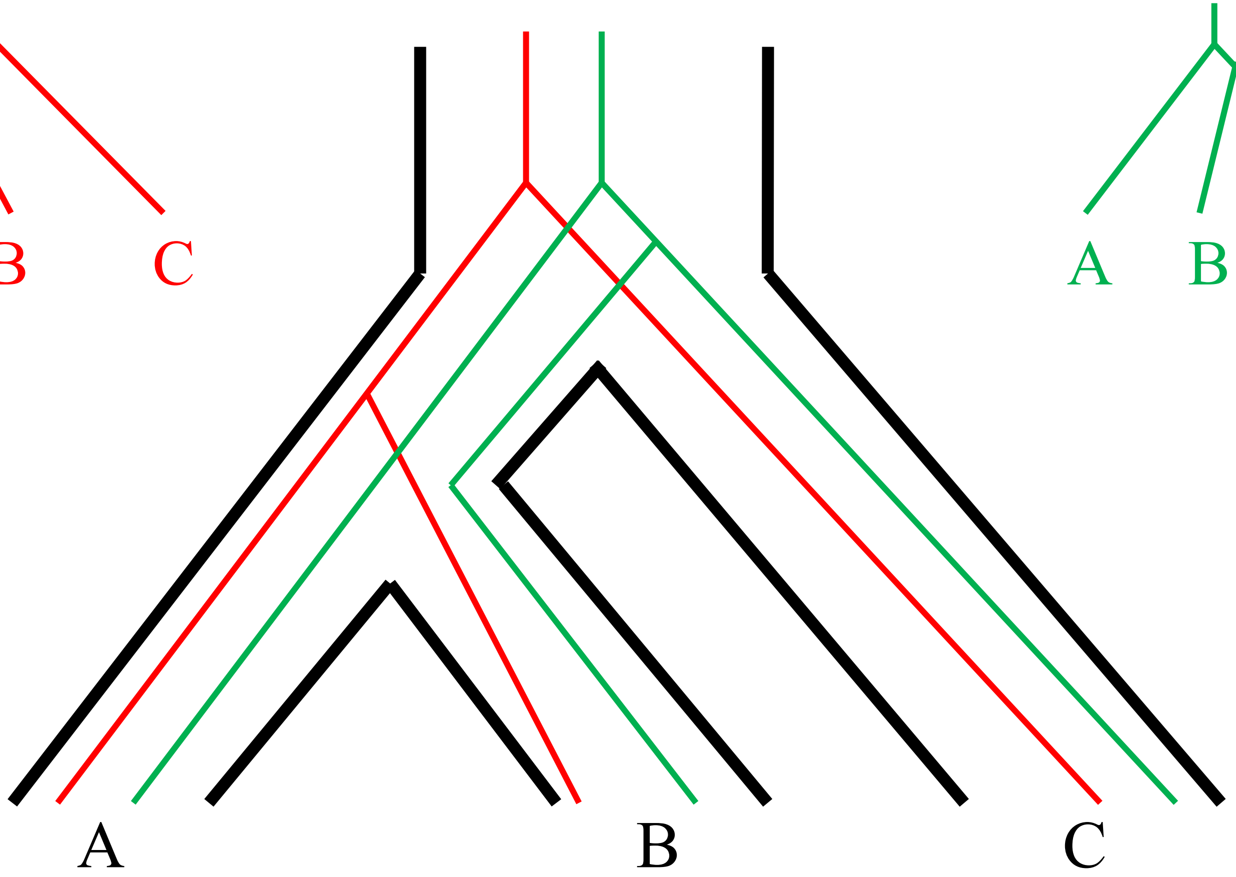


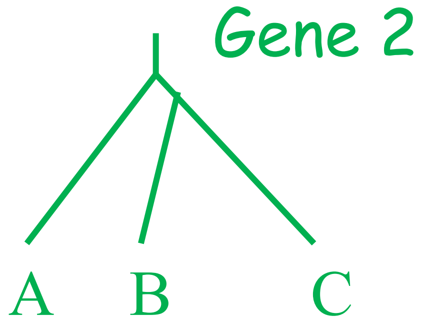
Studium selekce – fylogenetická analýza

Studium selekce - fylogenetická analýza

- Gene-tree versus species tree



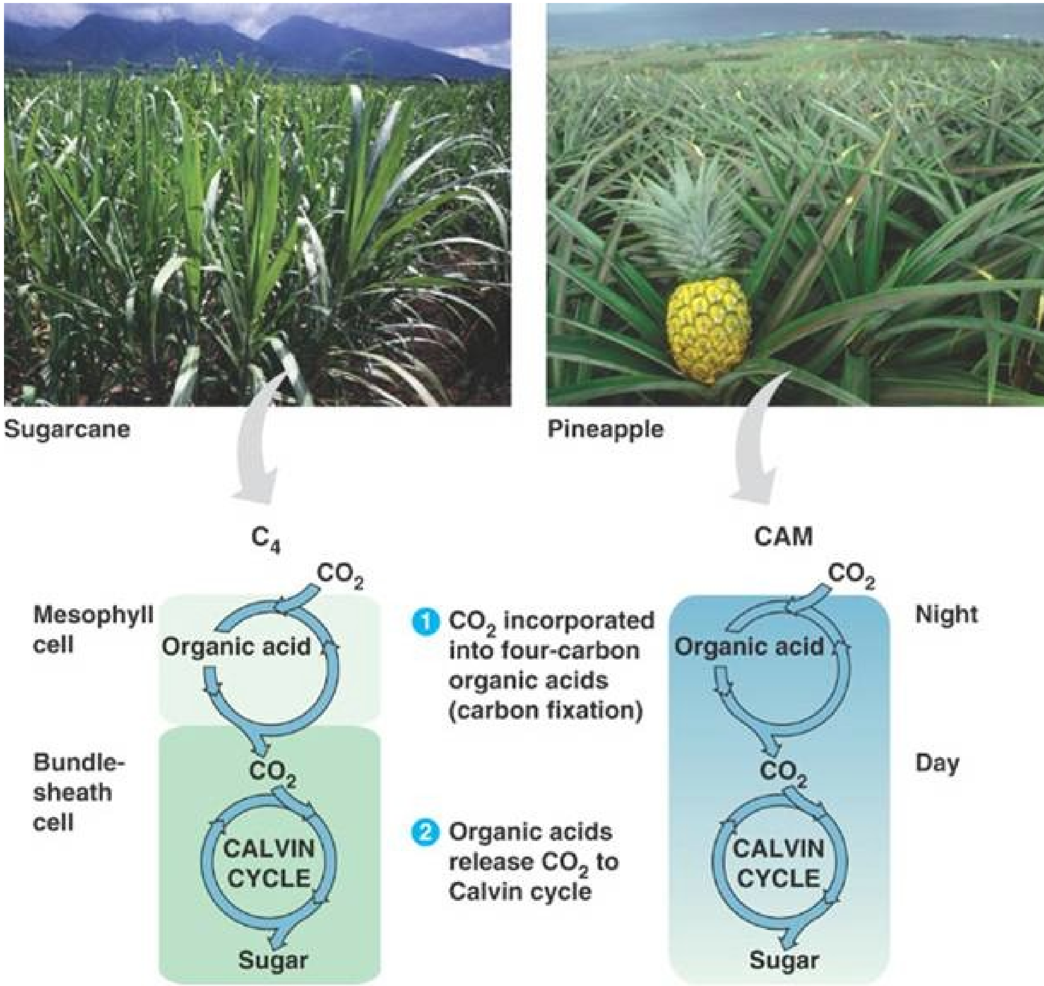




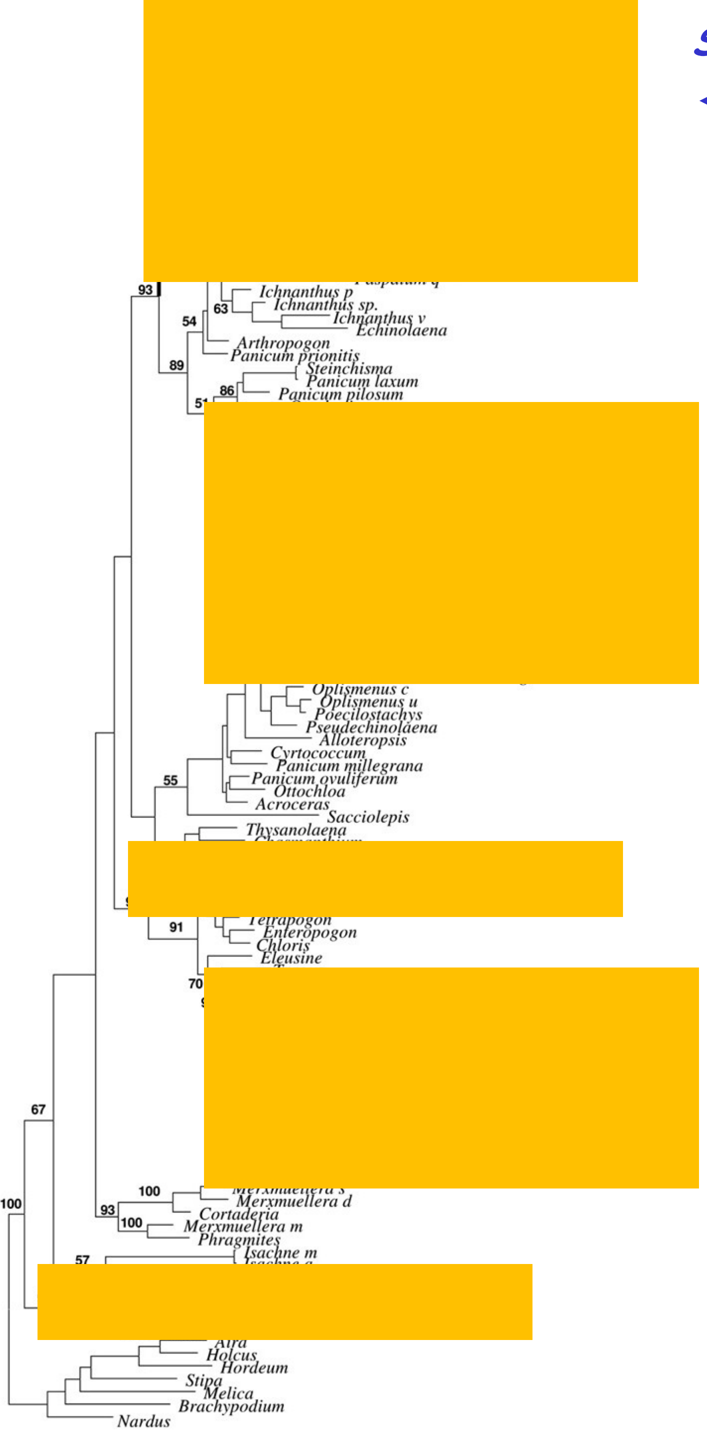
Studium selekce – fylogenetická analýza

- Gene-trees & Selection: The case of C4 photosynthesis

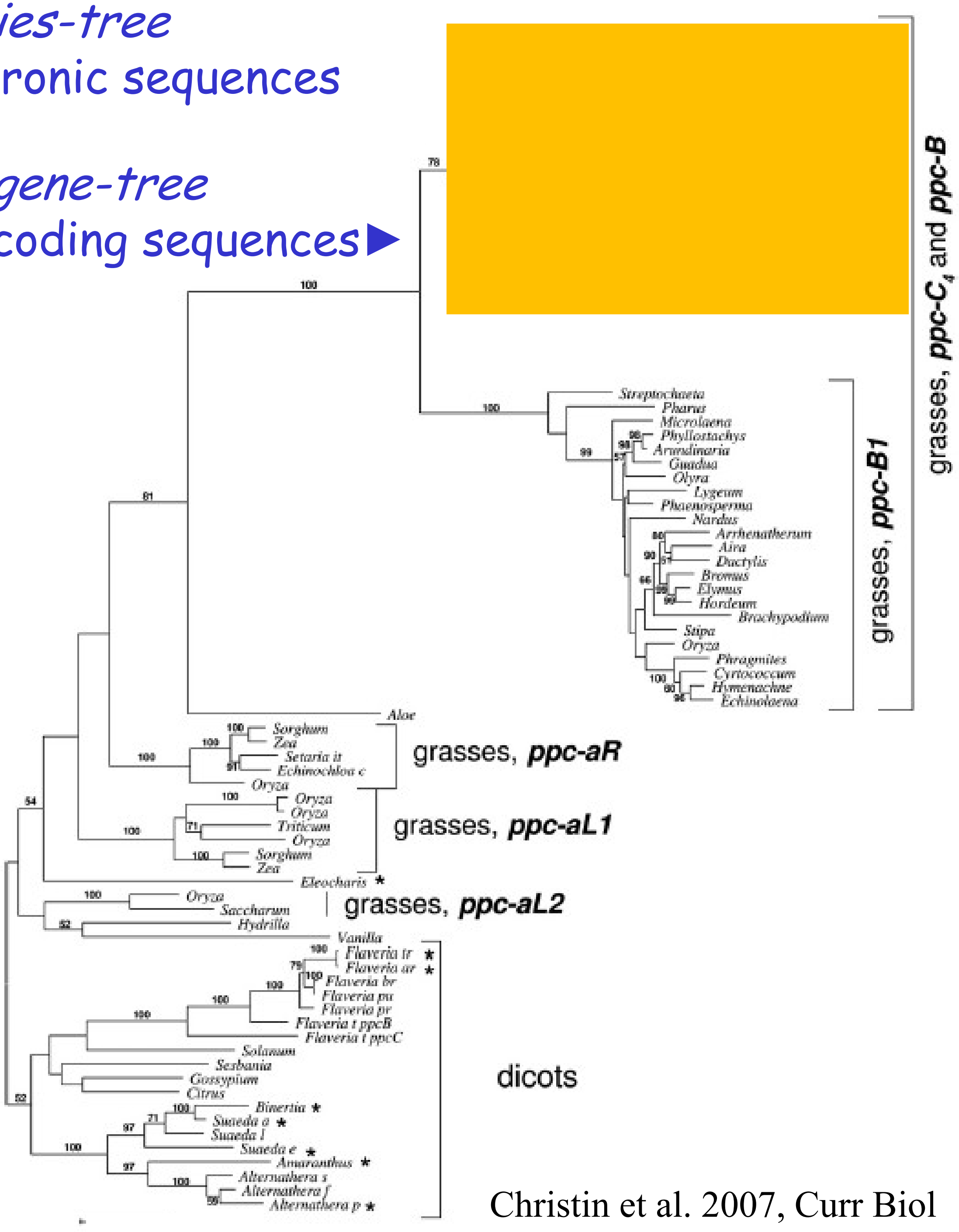
- Christin et al. 2007, Curr Biol
- *PEPC* gene
- →C4 photosynthesis



species-tree
◄intronic sequences

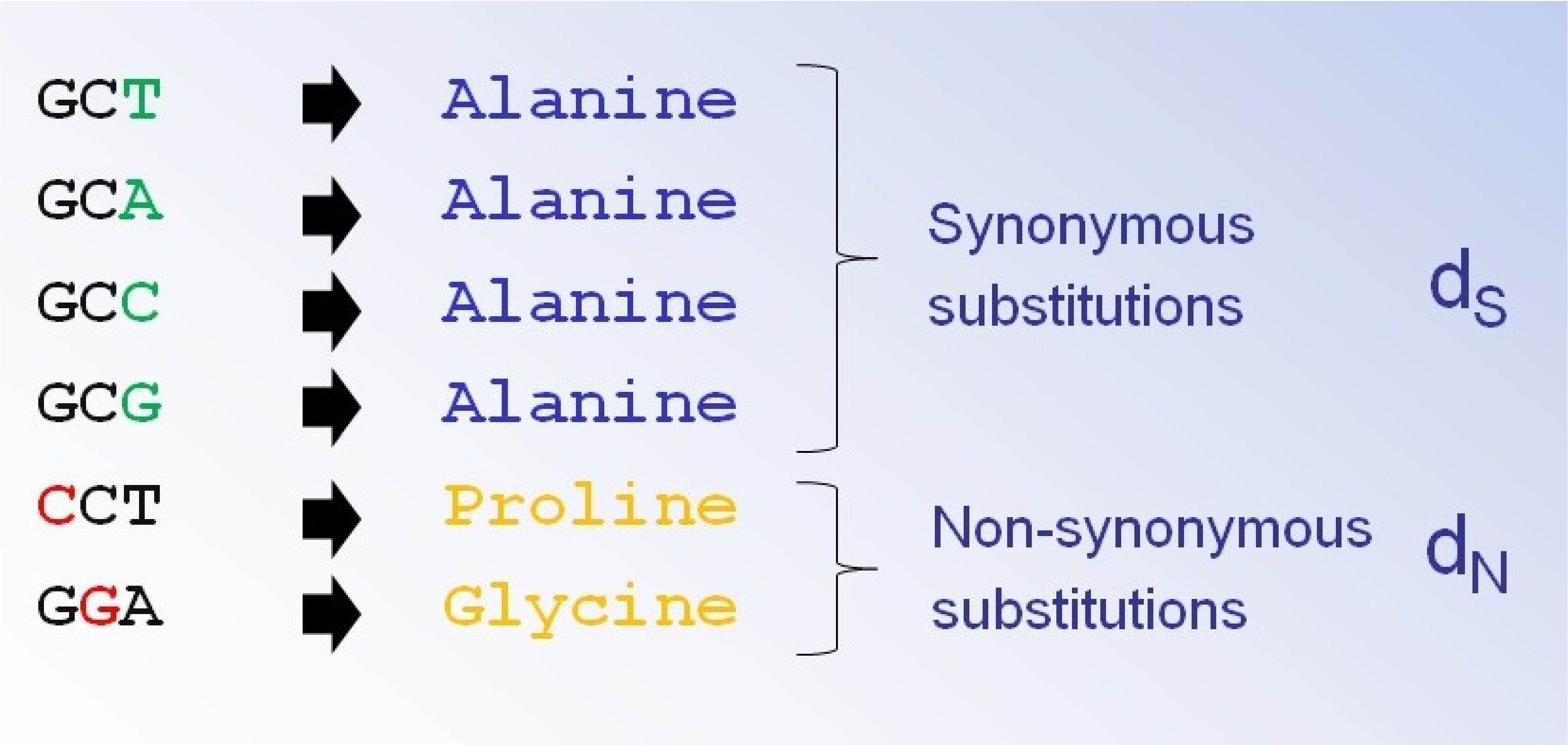


gene-tree
coding sequences►



Christin et al. 2007, Curr Biol

Selekce na úrovni sekvencí



GCT → Alanine

GCA → Alanine

GCC → Alanine

GCG → Alanine

CCT → Proline

GGA → Glycine

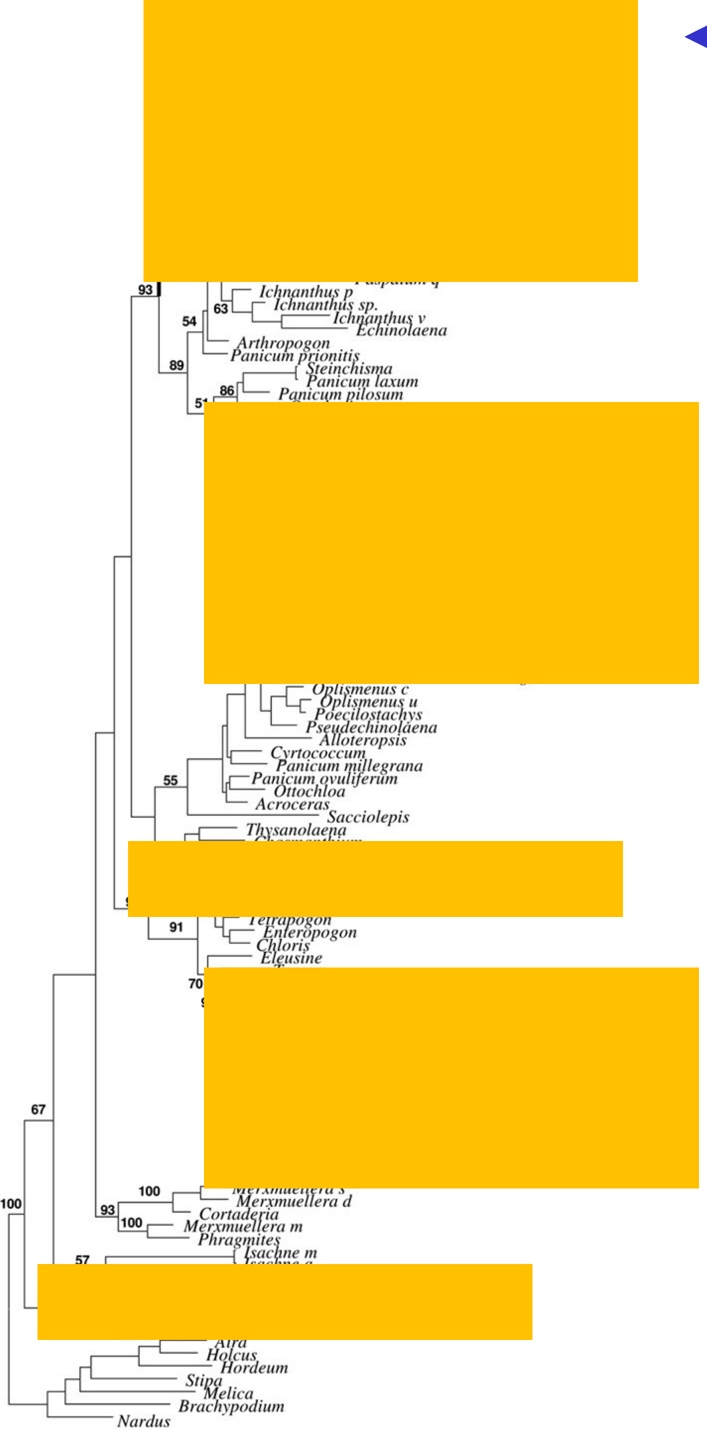
Selekce na úrovni sekvencí

- Testing for selection: d_N/d_S

d_S Rate of synonsymous substitutions, ‘neutral’ evolutionary rate

d_N Rate of non-synonsymous substitutions

- $d_N/d_S = 1$ as many syn as non-syn substitutions → neutral evolution
- $d_N/d_S < 1$ less non-syn than syn substitutions → purifying selection
- $d_N/d_S > 1$ more non-syn than syn substitutions → positive selection



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

◀intronic sequences

coding sequences▶
without positively
selected sites

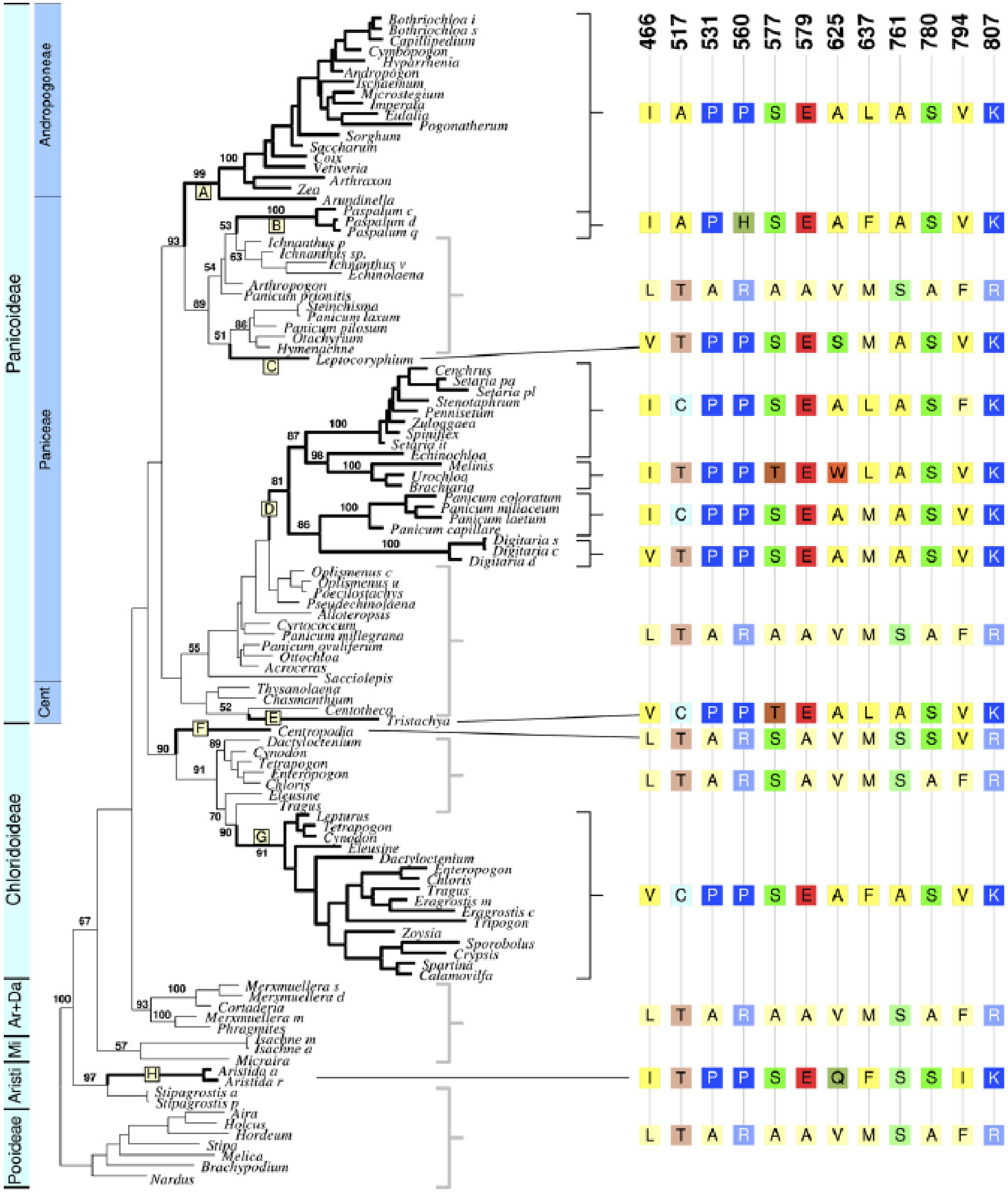
[redacted]

[redacted]

[redacted]

[redacted]

Christin et al. 2007, Curr Biol



466	517	531	560	577	579	625	637	761	780	794	807
I	A	P	P	S	E	A	L	A	S	V	K
I	A	P	H	S	E	A	F	A	S	V	K
L	T	A	R	A	A	V	M	S	A	F	R
V	T	P	P	S	E	S	M	A	S	V	K
I	C	P	P	S	E	A	L	A	S	F	K
I	T	P	P	T	E	W	L	A	S	V	K
I	C	P	P	S	E	A	M	A	S	V	K
V	T	P	P	S	E	A	M	A	S	V	K
L	T	A	R	A	A	V	M	S	A	F	R
V	C	P	P	T	E	A	L	A	S	V	K
L	T	A	R	S	A	V	M	S	S	V	R
L	T	A	R	S	A	V	M	S	A	F	R
V	C	P	P	S	E	A	F	A	S	V	K
L	T	A	R	A	A	V	M	S	A	F	R
I	T	P	P	S	E	Q	F	S	S	I	K
L	T	A	R	A	A	V	M	S	A	F	R

PEPC gene

12 codons with $d_N/d_S > 1$

Sequences of unrelated but ecologically similar species more similar at these positions than they are in related species.

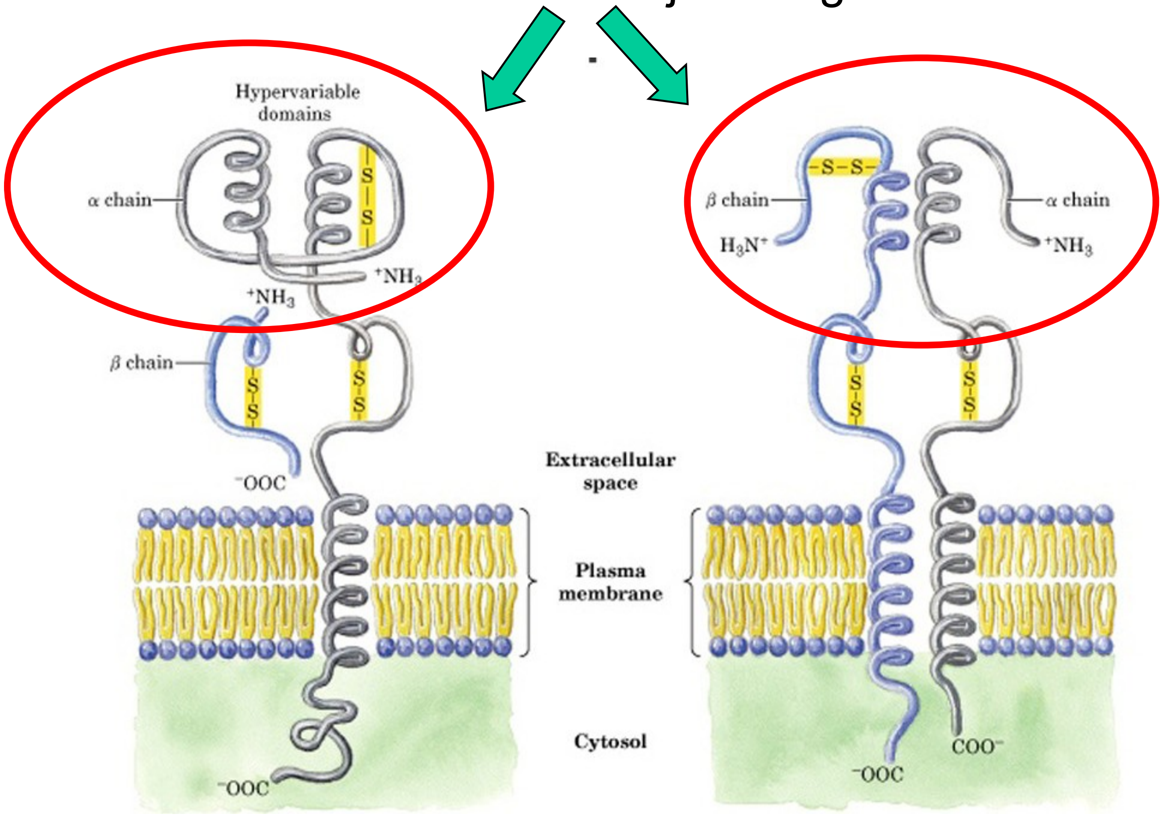
→ convergent evolution!

Christin et al. 2007, Curr Biol

Major histocompatibility complex (MHC)

Buňka nabízející antigen spustí imunitní odpověď

Oblast rozeznávající antigen

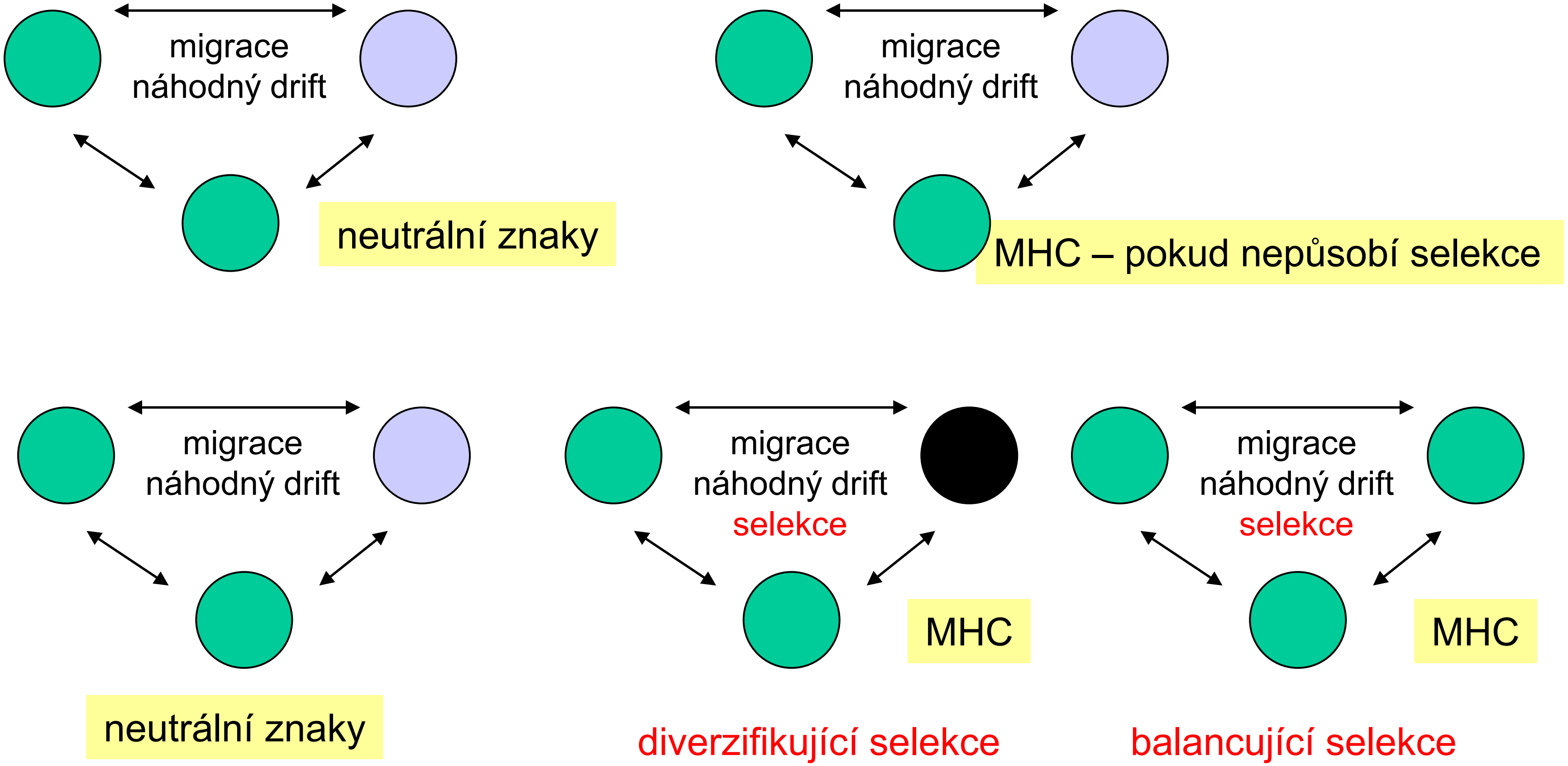


Class I

Class II

Studium selekce – populačně-genetická analýza

- Srovnání populačně-genetické struktury na MHC genech a neutrálních znacích (mikrosatelity)



migrace
náhodný drift

neutrální znaky

migrace
náhodný drift

MHC – pokud nepůsobí selekce

migrace
náhodný drift

neutrální znaky

migrace
náhodný drift
selekce

MHC

diverzifikující selekce

migrace
náhodný drift
selekce

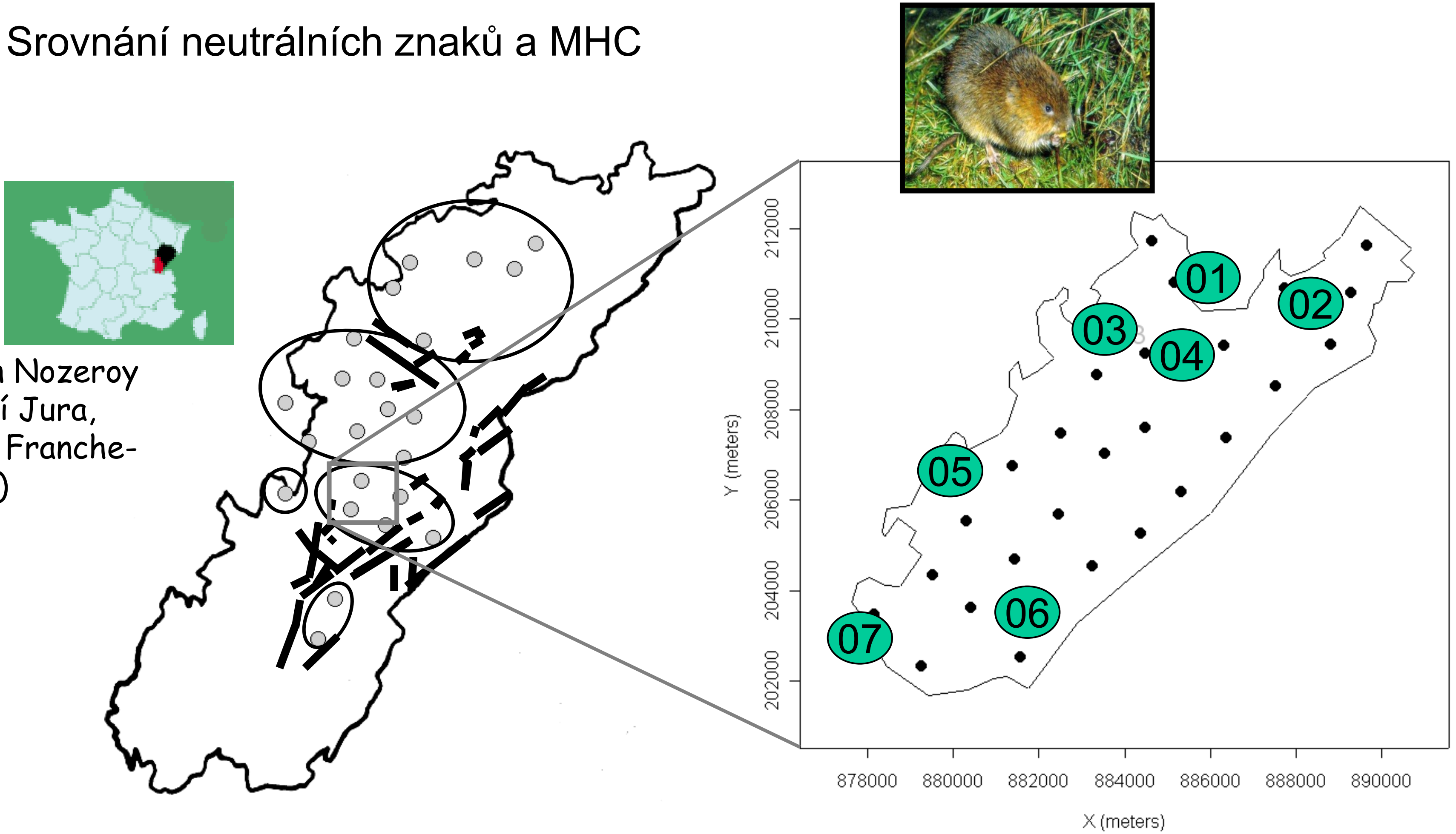
MHC

balancující selekce

Je možno kvantifikovat např. pomocí F_{ST}

Důkaz přírodního výběru v současnosti: analýza populačně-genetické struktury

- Srovnání neutrálních znaků a MHC

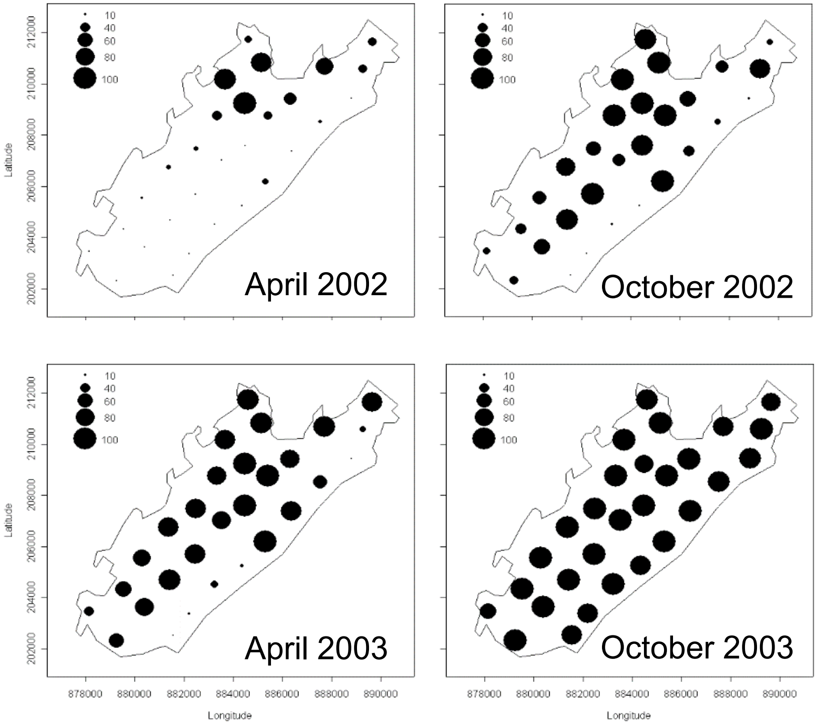


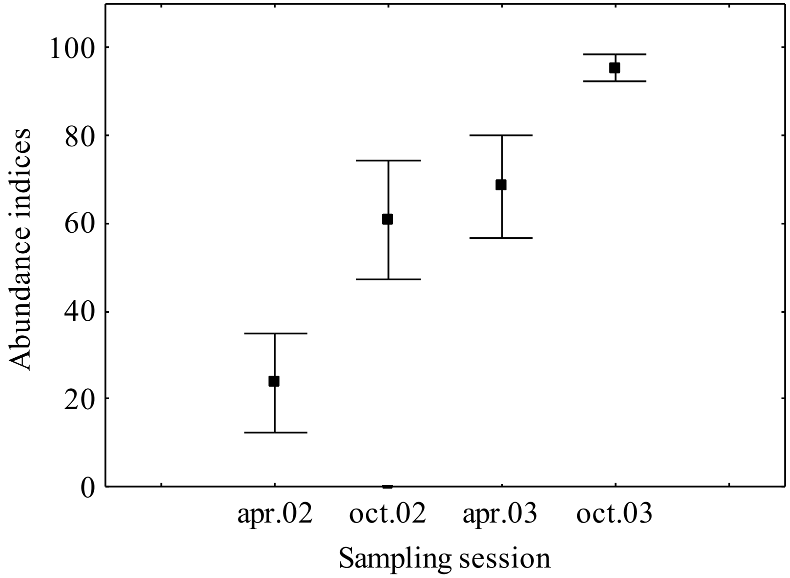
Canton Nozeroy (pohoří Jura, région Franche-Comté)

Studované lokality – 7 populací ve stejné fázi populačního cyklu

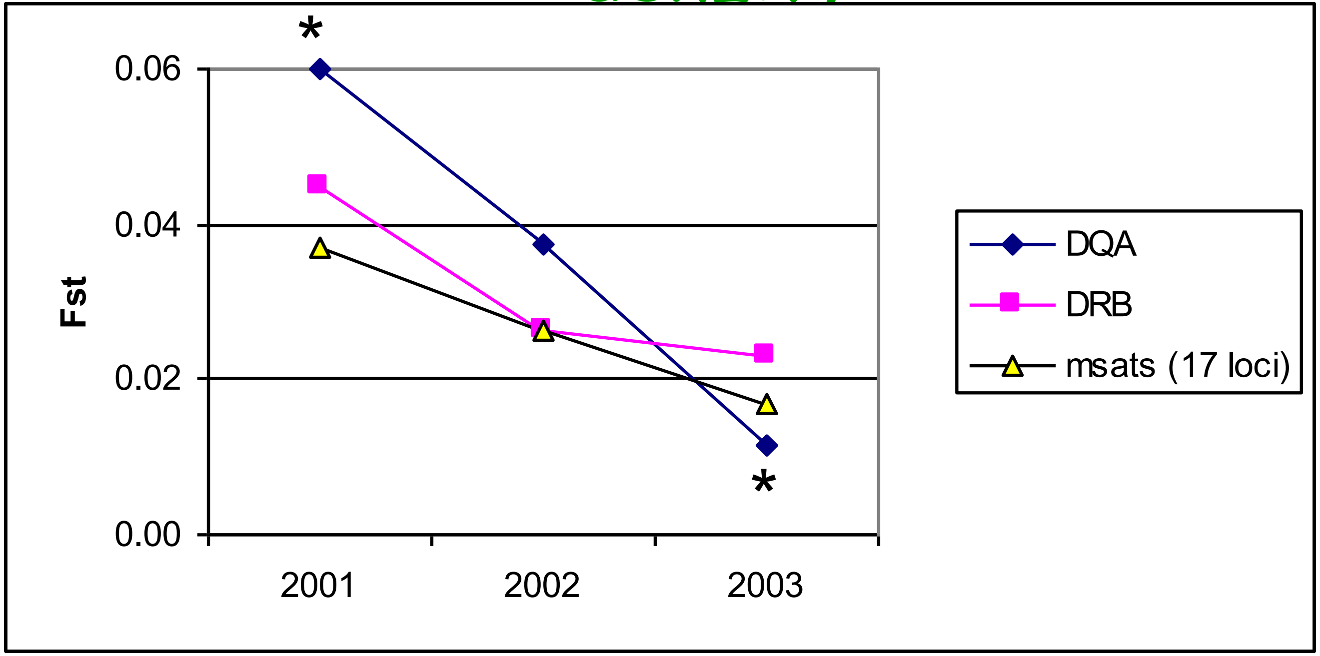
Bryja et al. 2007, Molecular Ecology

→ 2001-2003: fáze růstu populační hustoty





Diferenciace populací v průběhu růstu denzity



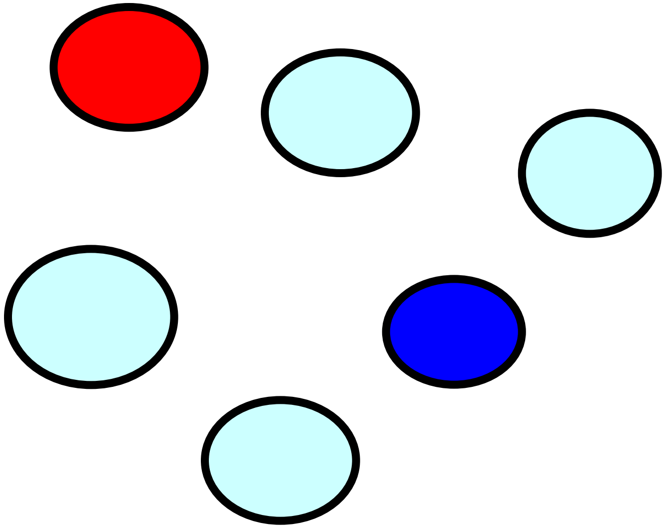
* Signifikantní rozdíl DQA1 vs. mikrosatelity

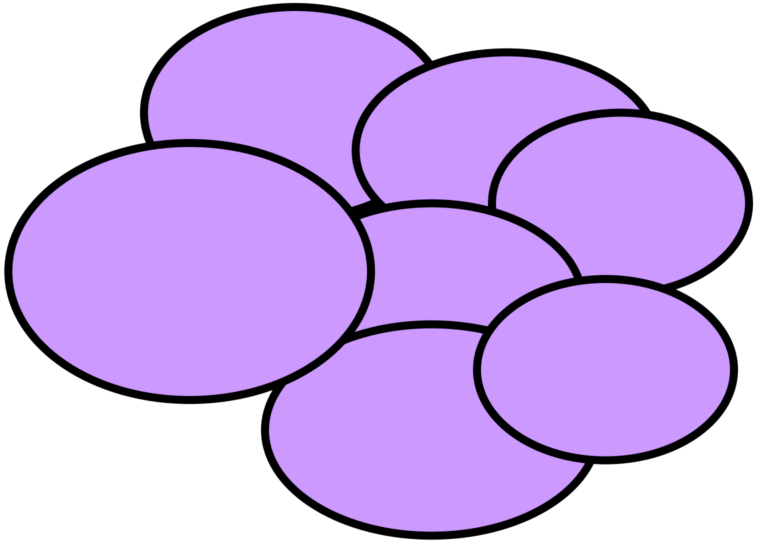
Pokles diferenciace s nárůstem denzity (nárůst disperze, tj. toku genů)

MHC (zejména DQA1) – signifikantně odlišné od mikrosatelitů

Bryja et al. 2007, Molecular Ecology

Závěr: Typ selekce na MHC závisí na početnosti populace





Nízká denzita

Lokální rozdíly ve společenstvech patogenů

Lokální diverzifikující selekce

Vysoká denzita

Nárůst diverzity parazitů v důsledku disperze

Balancující selekce

Bryja et al. 2007, Molecular Ecology

Metody studia funkční variability

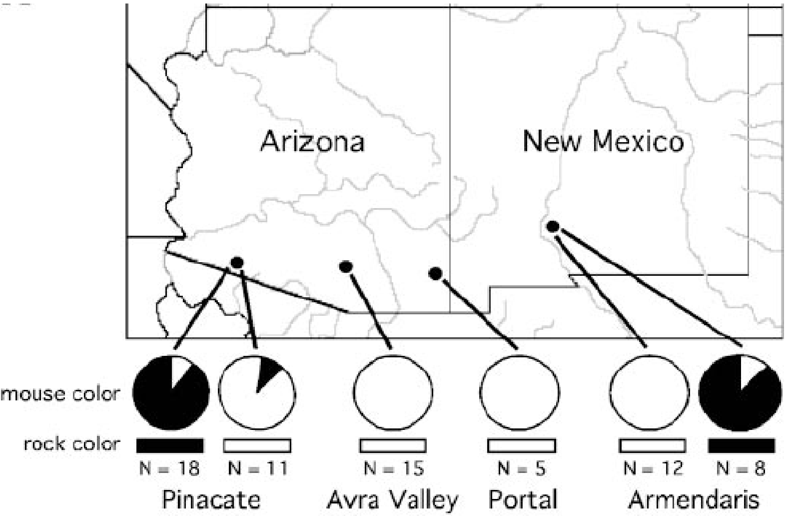
1. Sledování kandidátních genů

2. Genomické přístupy (mnoho genů najednou)

pytlouš

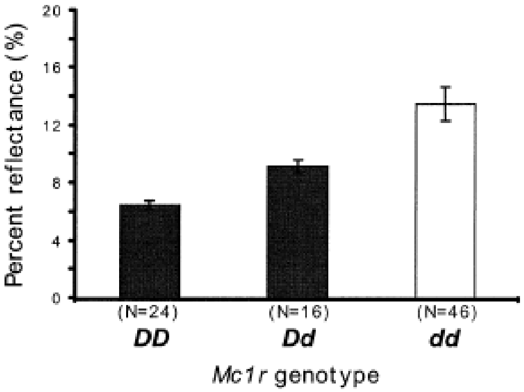
Chaetodipus intermedius

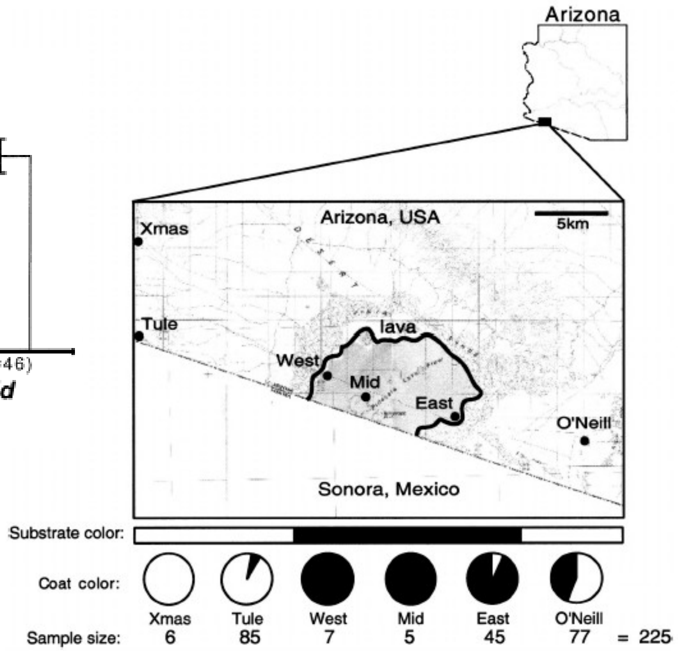
Hoekstra, Nachman et al.

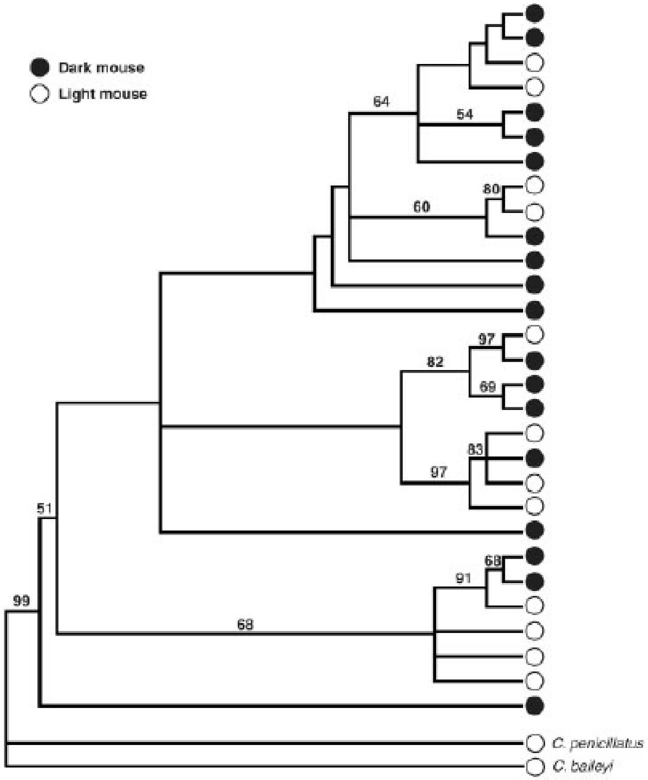


- Tmavé a světlé zbarvení
- Odpovídá barvě prostředí (tmavé zbarvení na lávě)

- **Arizona**
- Korelace zbarvení s prostředím i na malé škále
- mtDNA nekoreluje se zbarvením
- Sekvenování kandidátních genů (známých z inbredních myší)
- melanocortin-1 receptor MC1R
- Záměna 4 aminokyselin
- Jednoduchá dědičnost alel a zbarvení

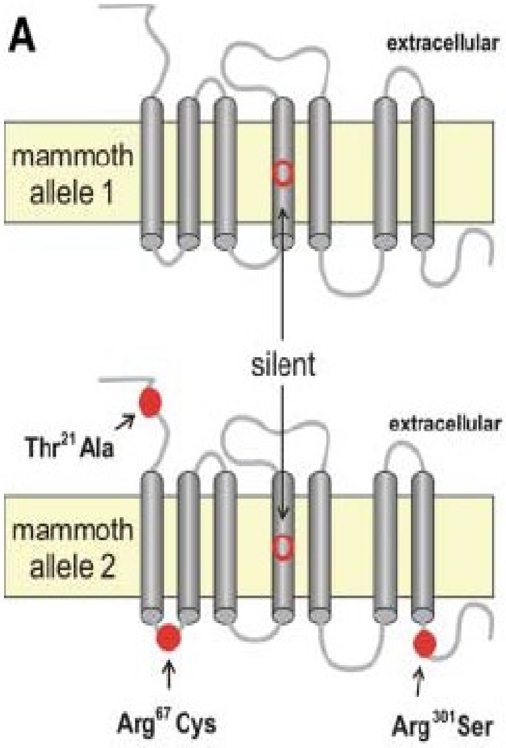






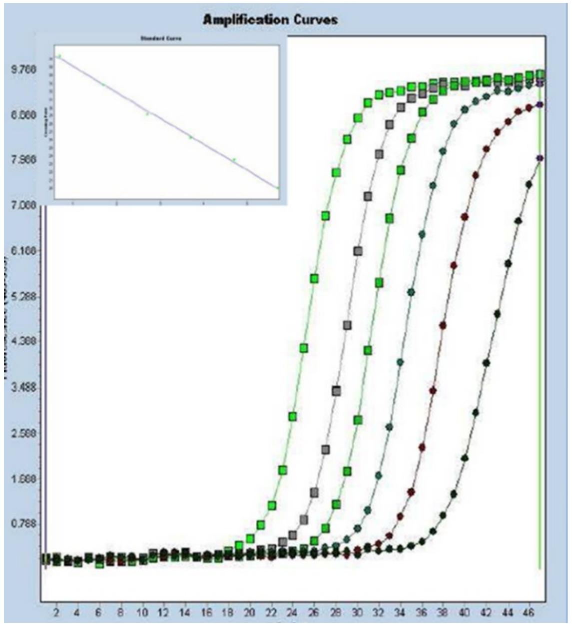
MC1R u člověka, mamuta a dalších

- U člověka zrzavé vlasy a neschopnost se opálit
- Zbarvení krav, koňů a psů
- Výskyt dvou odlišných variant u mamutů

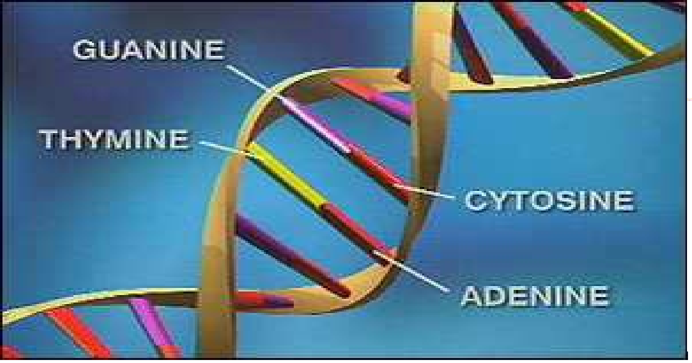


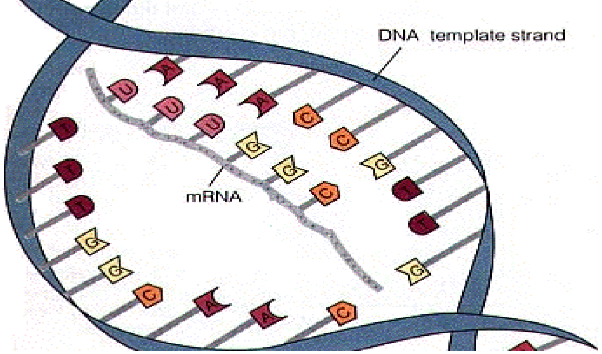
Measuring of expression - qPCR

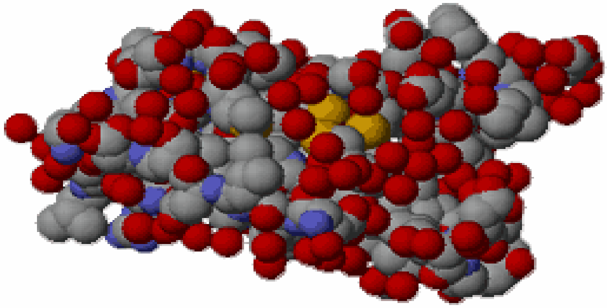
- Relative comparison of quantity of particular DNA, e.g. level of specific gene expression (i.e. particular RNA = cDNA) (e.g. comparison of different tissue types, elevations, treatement vs. non-treatement etc.)
- **housekeeping genes** – use as standard for quantification
- same number of copies in all cells (e.g. genes encoding proteins of cytoskeleton)
- constitutive genes - expressed in all cells, independent on experimental treatment
- validation of housekeeping genes should be performed before their use in gene expression experiments



Sledování mnoha genů najednou







DNA → mRNA → Proteins

Transcription

Translation

Exome

Genome

Transcriptome

Proteome

3 billion bases

20-30,000 genes

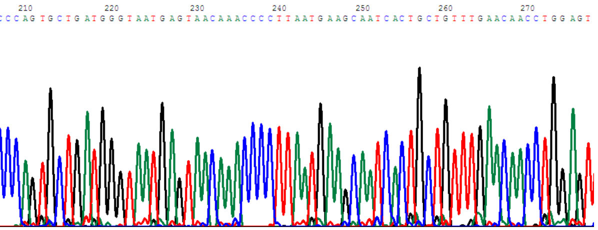
~100,000 proteins

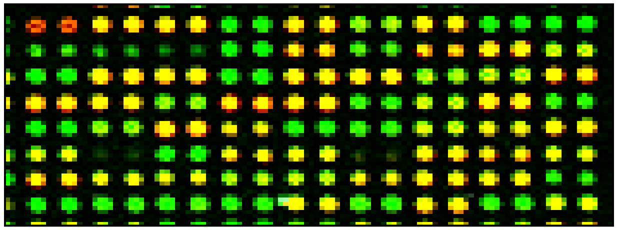
Functional Genomics

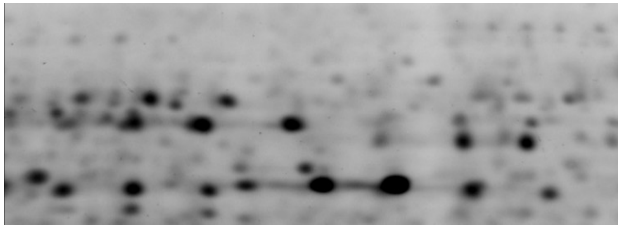
Genomics

Transcriptomics

Proteomics



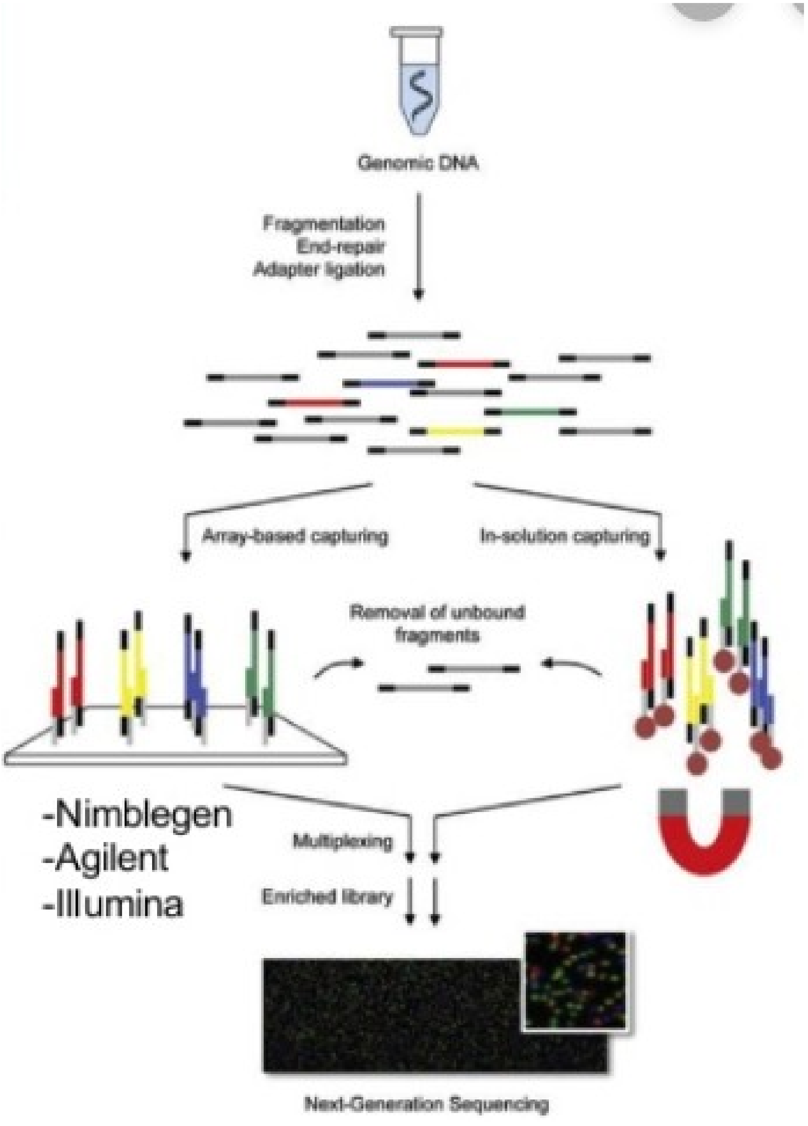




Exome-Seq

Targeted exome capture

- targets ca. 20,000 coding sequences
- high depth of coverage for more accurate variant calling



Transcriptomics - analysis of mRNAs

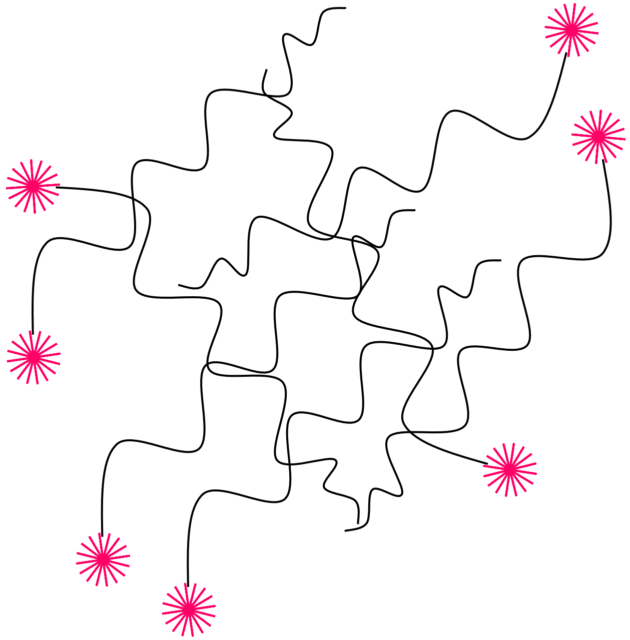
1. microarrays
2. RNA seq (NGS)

1. Analysis of gene expression by microarrays

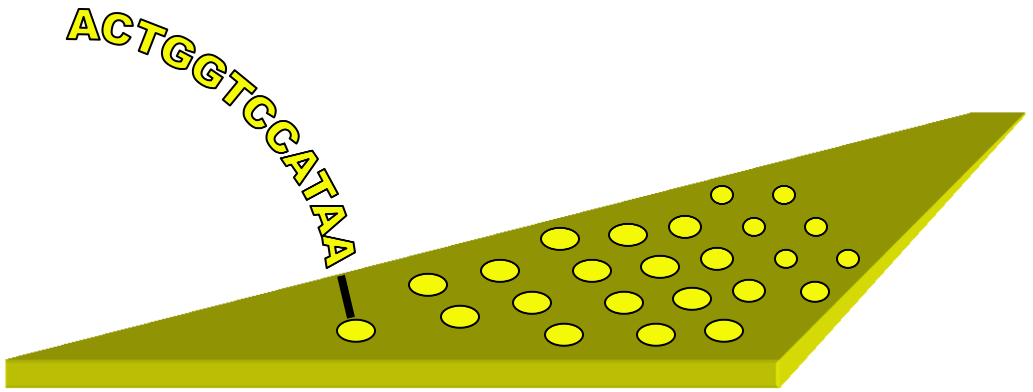
Ranz JM, Machado CA: Uncovering evoutionary patterns of gene expression using microarrays. TREE, 21(1): 29-37

Microarray analysis of transcriptome

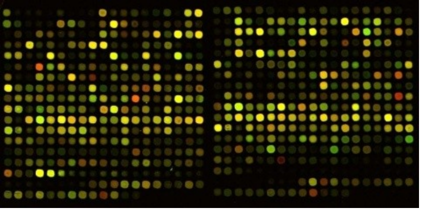
(~ specific DNA hybridization)



Target (i.e. mix of transcripts in a form of cDNA = mRNA přepsaná do DNA reverzní transkriptázou, tj. neobsahuje introny)



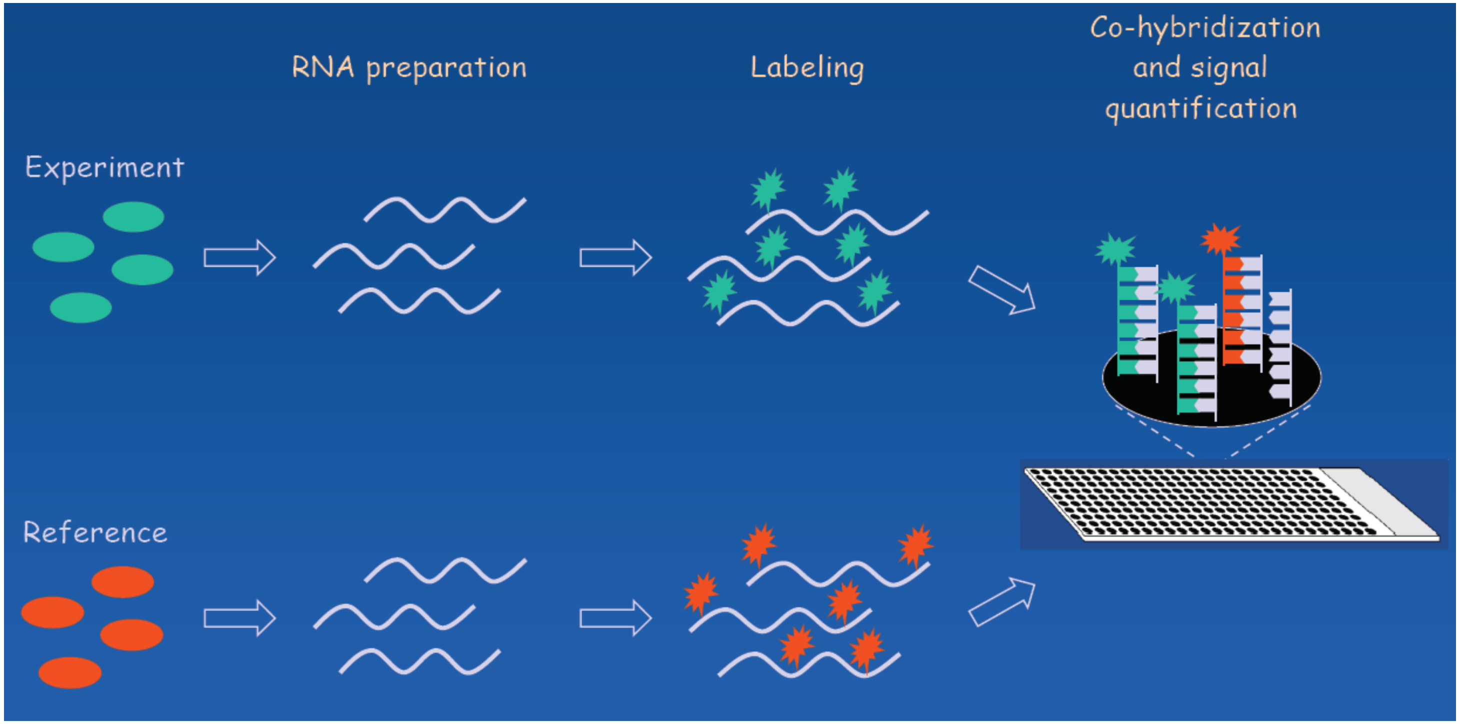
Probe (i.e. synthesized oligonucleotides complementary to particular genes)



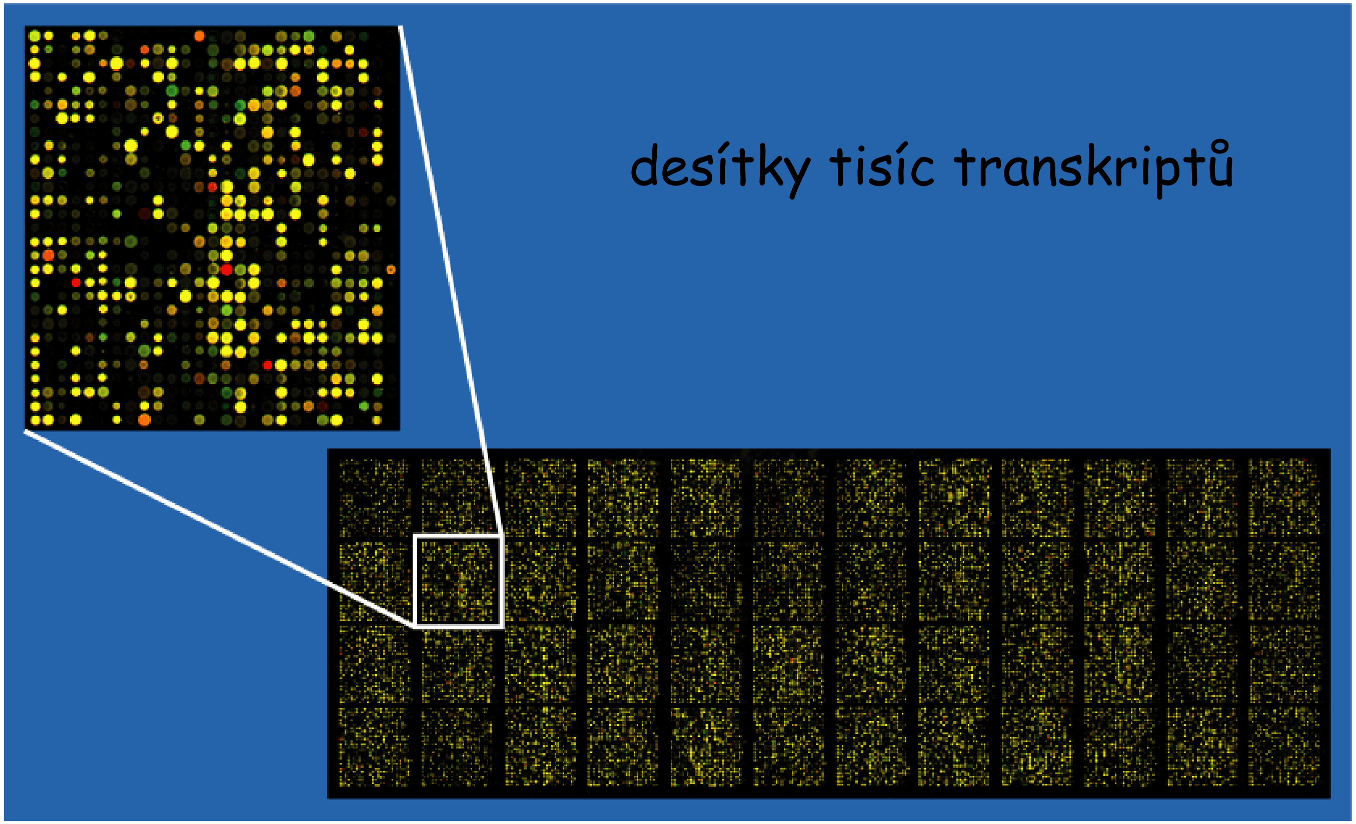
Sledování exprese genů microarrays

- Sledování exprese mnoha (tisíce) genů najednou
- Založeno na hybridizaci
- Sleduje se rozdíl vůči kontrole ("heterologous hybridization") = dvoukanálový experiment

Affymetrix
Agilent Technologies

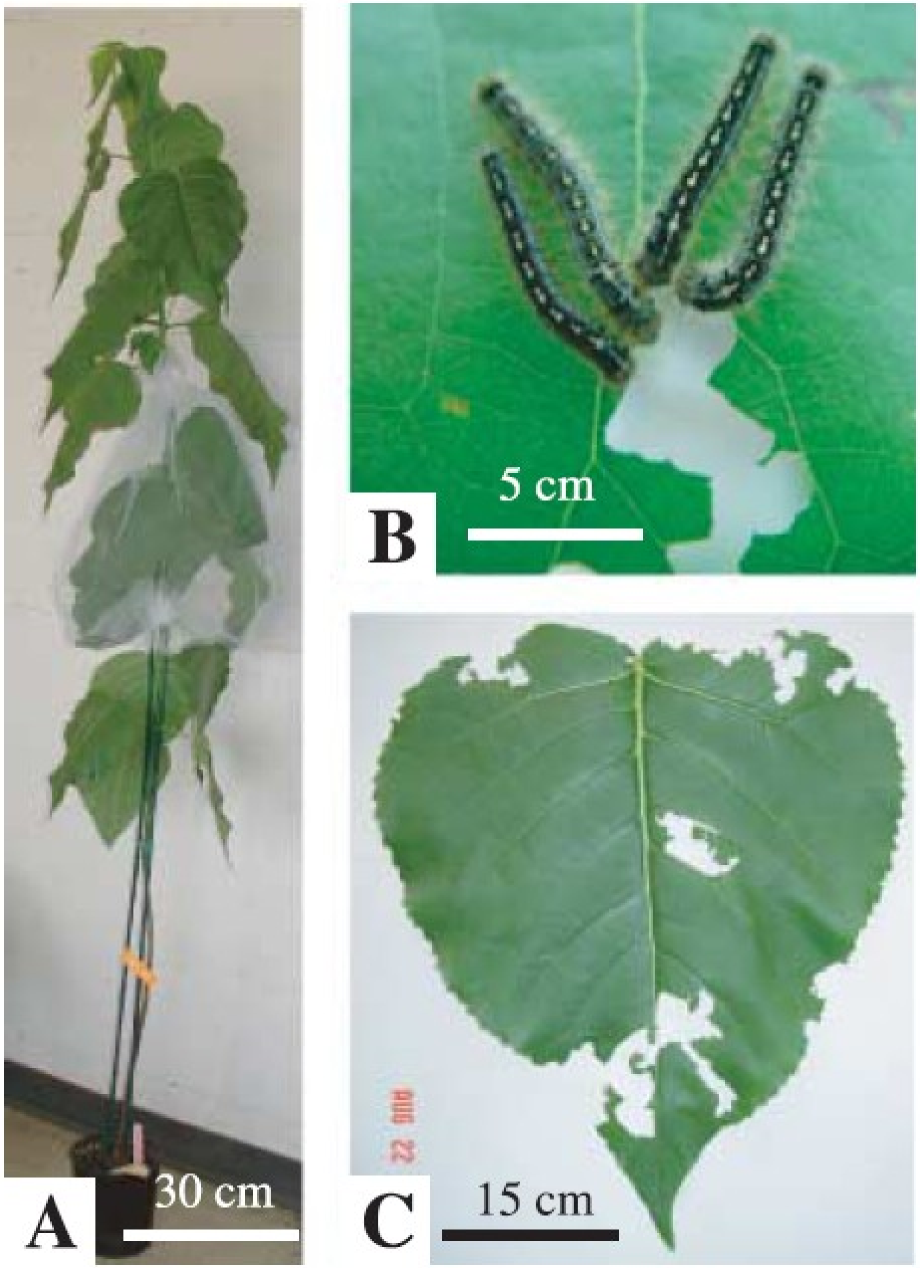


Vyhodnocení chipu – analýza obrazu (srovnání úrovně exprese mezi kontrolou a experimentem)



desítky tisíc transkriptů

- komerčně dostupné pro kompletní transkriptom cca 25 druhů (Affymetrix)
(další jsou rychle vyvíjeny, i na zakázku)
- celkově ale microarrays ustupují před RNAseq



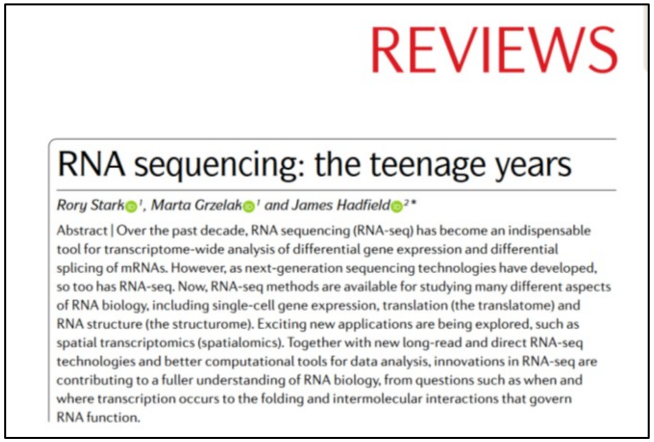
Populus trichocarpa x *deltoides* a *Malacosoma disstria* bourovec

Ralph et al. *2006*

- cDNA microarray
- 15496 genů > ¾ genomu
- Po 24 hodinách
  1191 genů up-regulated
  537 down-regulated
- Obrana: endochitinázy, inhibitory proteáz
- Signální funkce
- Transport, metabolismus, regulace transkripce

2. RNAseq

(„high-throughput sequencing of cDNAs“)



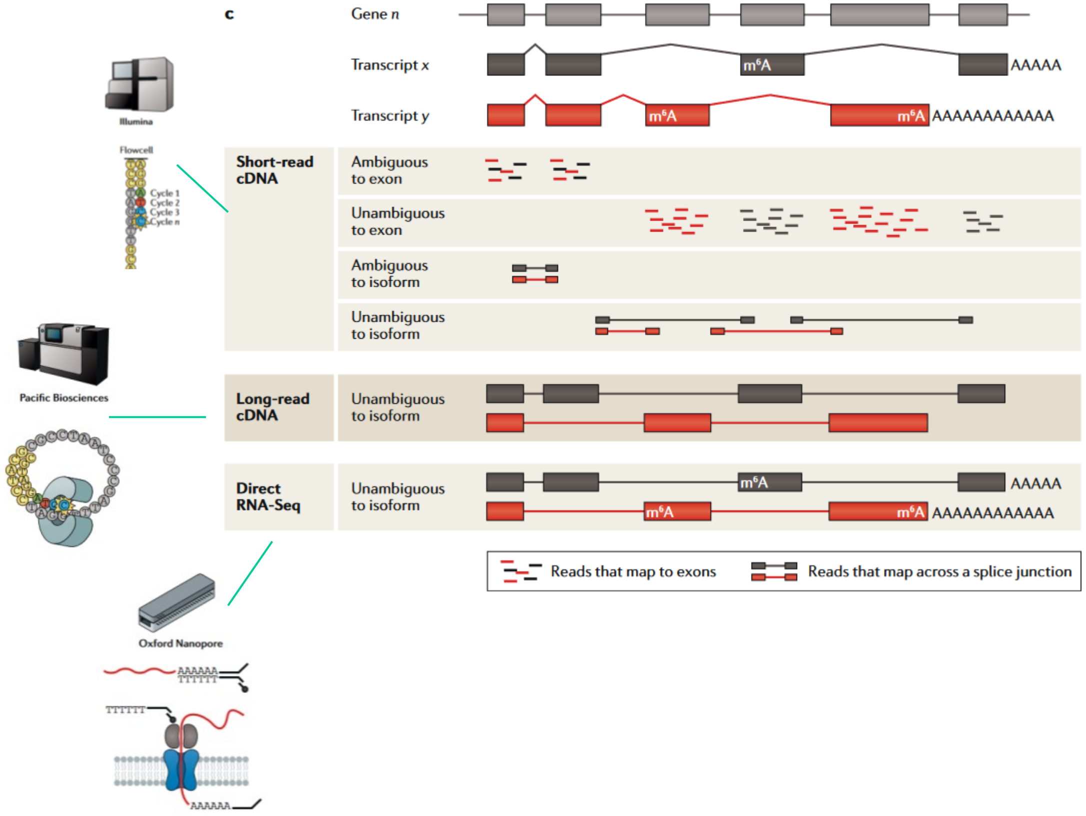
REVIEWS

RNA sequencing: the teenage years

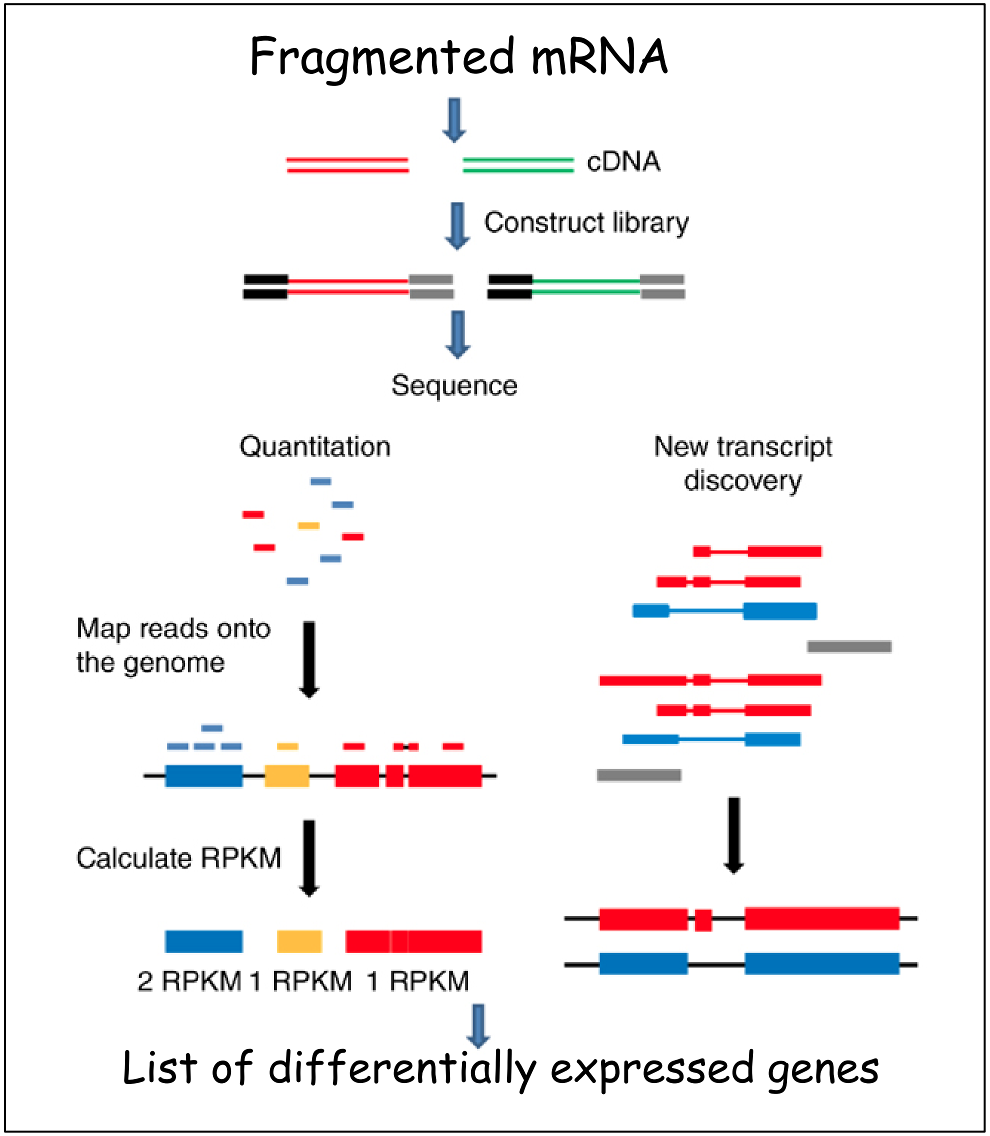
Rory Stark[1], Marta Grzelak[1] and James Hadfield[2]*

Abstract | Over the past decade, RNA sequencing (RNA-seq) has become an indispensable tool for transcriptome-wide analysis of differential gene expression and differential splicing of mRNAs. However, as next-generation sequencing technologies have developed, so too has RNA-seq. Now, RNA-seq methods are available for studying many different aspects of RNA biology, including single-cell gene expression, translation (the translatome) and RNA structure (the structurome). Exciting new applications are being explored, such as spatial transcriptomics (spatialomics). Together with new long-read and direct RNA-seq technologies and better computational tools for data analysis, innovations in RNA-seq are contributing to a fuller understanding of RNA biology, from questions such as when and where transcription occurs to the folding and intermolecular interactions that govern RNA function.

Nature Reviews Genetics 2019



RNA-Seq workflow for gene expression analysis



Fragmented mRNA

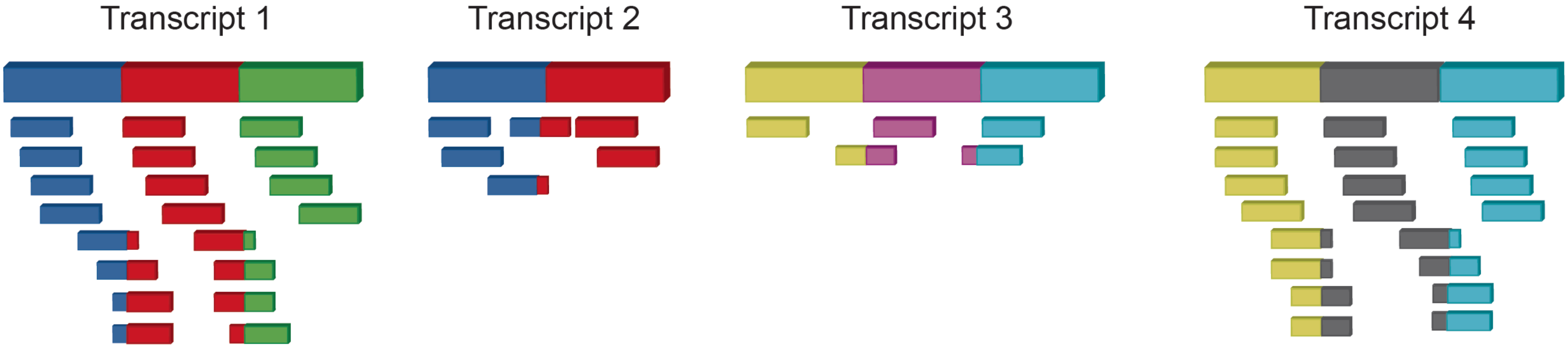
List of differentially expressed genes

Adapted from Zeng and Mortazavi, *Nature Immunology* **13** (2012)

RNA-Seq quantification (short reads Illumina)

(RPKM = reads per kilobase per million reads)

Actual Read Alignment to Alternative Splice Variants

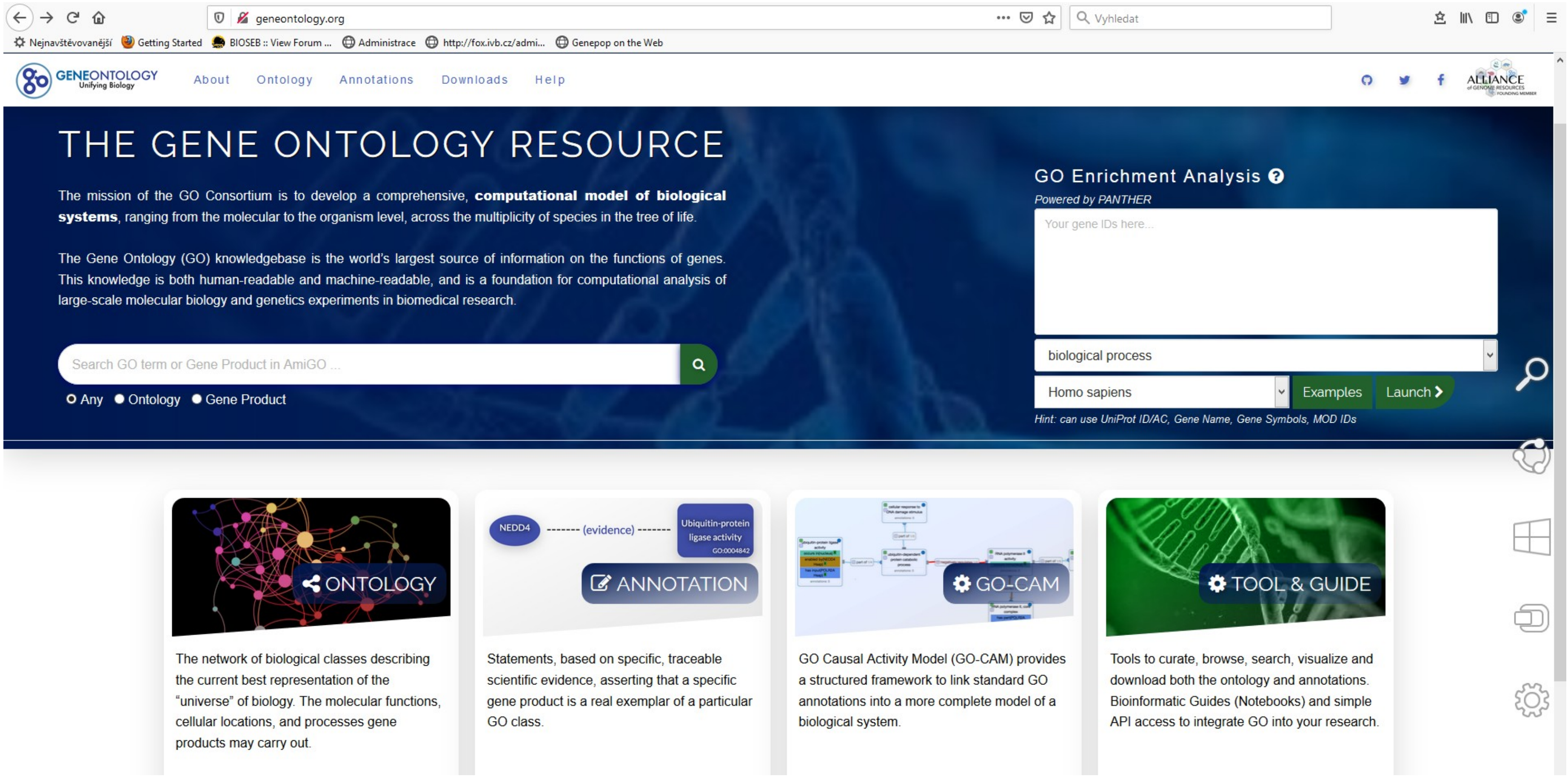


The reference transcriptome is required

Gene ontology (http://geneontology.org/)

= functional annotation analysis

- založena na databázích dostupných anotovaných genů u modelových organismů
- **Cellular Component** - the parts of a cell or its extracellular environment
- **Molecular Function** - the elemental activities of a gene product at the molecular level, such as binding or catalysis
- **Biological Process** - operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms.



THE GENE ONTOLOGY RESOURCE

The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

GO Enrichment Analysis

Powered by PANTHER

Hint: can use UniProt ID/AC, Gene Name, Gene Symbols, MOD IDs

The network of biological classes describing the current best representation of the "universe" of biology. The molecular functions, cellular locations, and processes gene products may carry out.

Statements, based on specific, traceable scientific evidence, asserting that a specific gene product is a real exemplar of a particular GO class.

GO Causal Activity Model (GO-CAM) provides a structured framework to link standard GO annotations into a more complete model of a biological system.

Tools to curate, browse, search, visualize and download both the ontology and annotations. Bioinformatic Guides (Notebooks) and simple API access to integrate GO into your research.

Example of GO annotation: **cytochrome c**

- *molecular function*: oxidoreductase activity
- *biological proces*: oxidative phosphorylation
- *cellular component*: mitochondrial matrix

Examples

Jaká je úroveň exprese v různých tkáních?

- srdeční, červená a bílá svalovina
- rozdíly v expresi 113, 81 a 196 genů
- rozdíly jsou způsobeny tkáňově specifickým stupněm endotermie

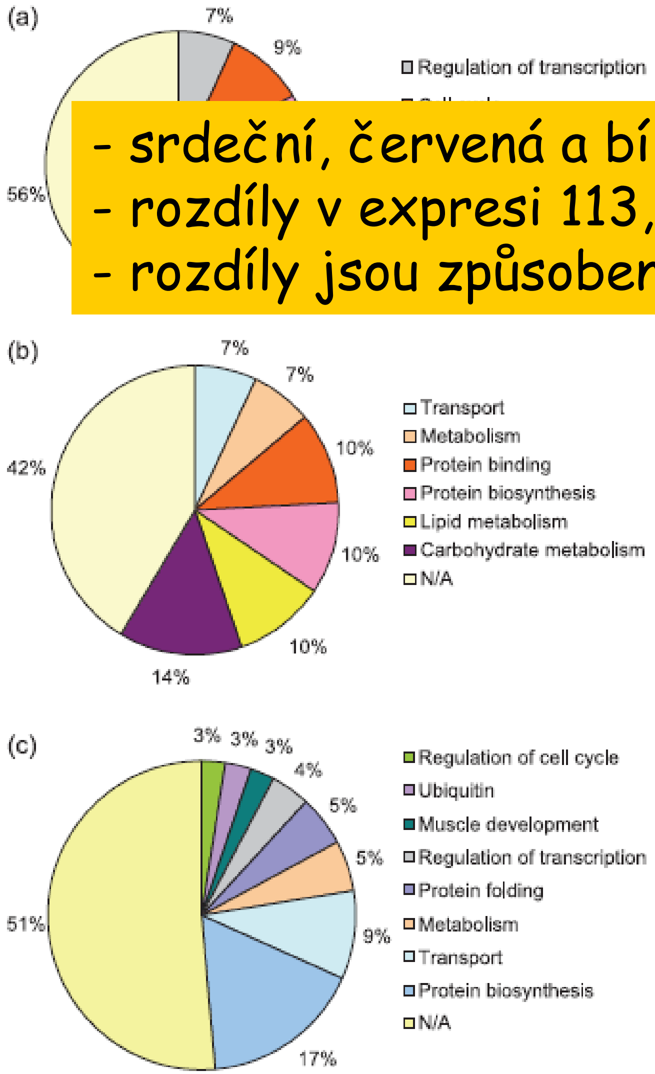


Fig. 1 Percentage of total up-regulated genes identified in the Gene Ontologies (Biological Process) of ventricle (a), red muscle (b) and white muscle (c).

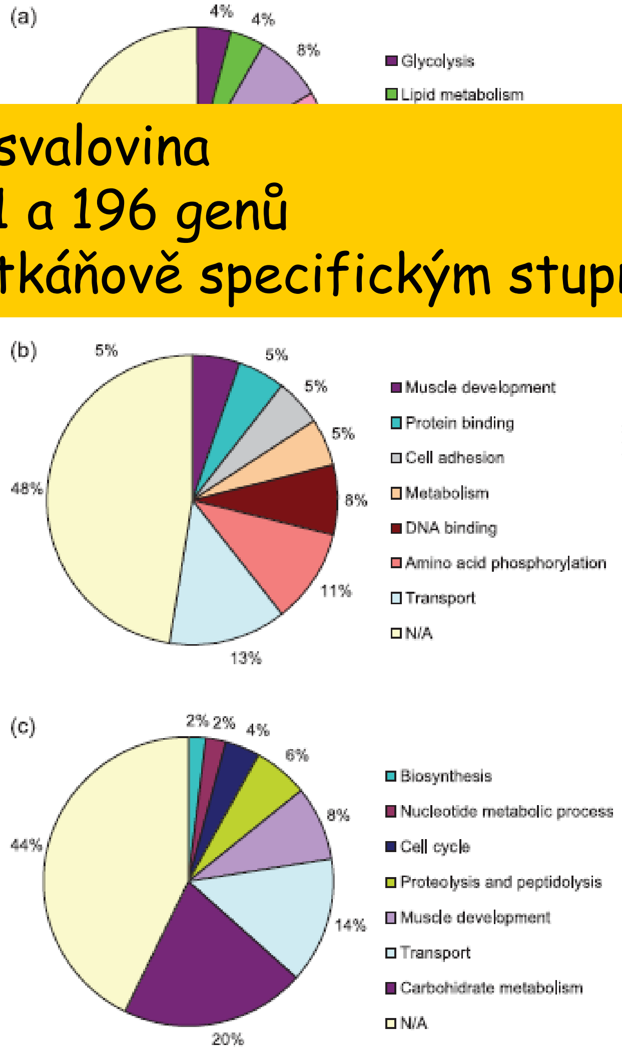


Fig. 2 Percentage of total down-regulated genes identified in the Gene Ontologies (Biological Process) of ventricle (a), red muscle (b) and white muscle (c).

quantitative real-time PCR

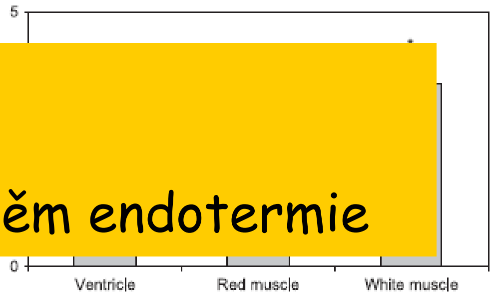
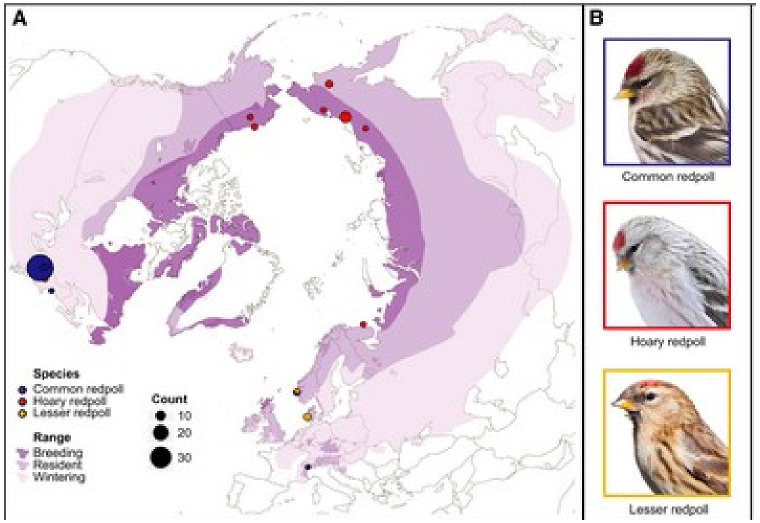


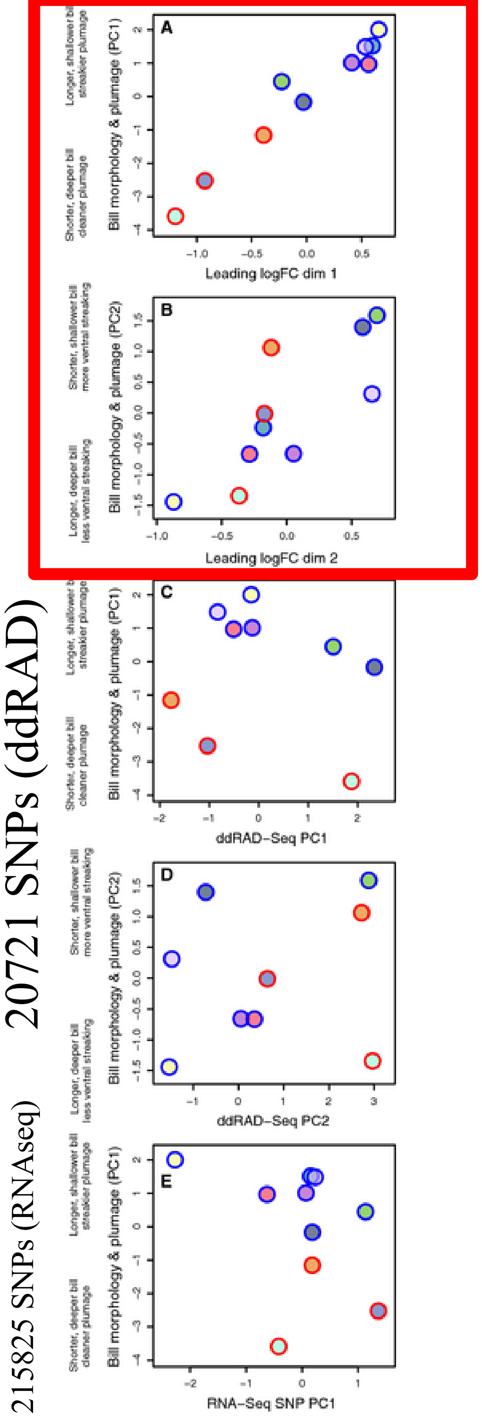
Fig. 3 Relative gene expression level of candidate gene *Hsp70* in cold-stressed Thunnus orientalis (*n* = 7) determined by quantitative real-time PCR. Expression levels were normalized to the howsekeeping gene β-actin. (*) Significant values between means.

vliv aklimatizace k chladu (20° vs 15°C)

Mason and Taylor 2015

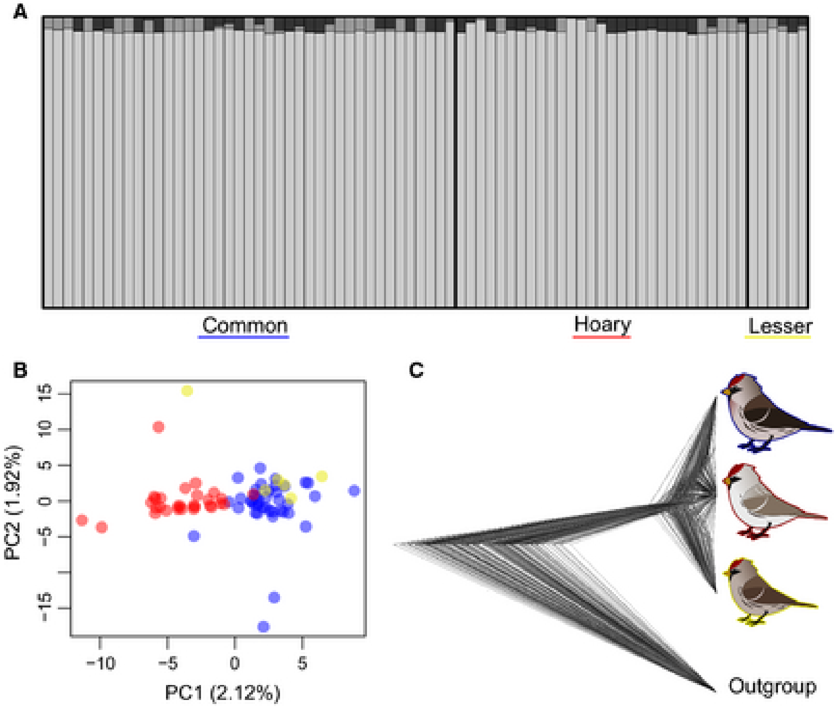


Only differentially expressed genes are responsible for morphological changes (zobák, zbarvení)



20721 SNPs (ddRAD)

215825 SNPs (RNAseq)



20721 SNPs (ddRAD) – no genetic difference at neutral loci

Sequencing of 73 genomes of all three „species"

nature COMMUNICATIONS

ARTICLE

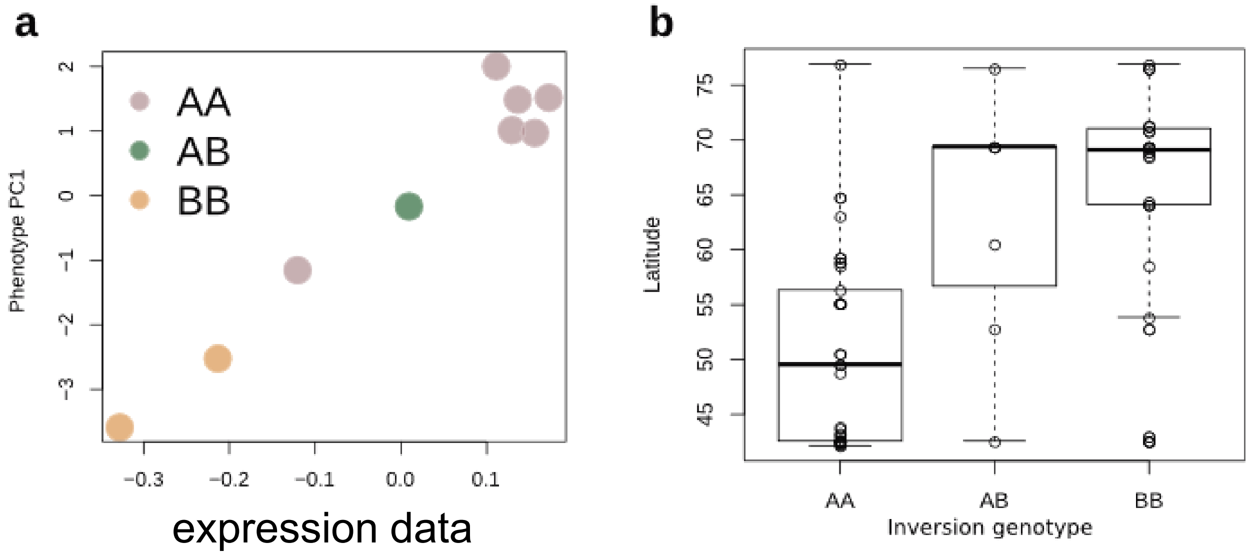
Check for updates

https://doi.org/10.1038/s41467-021-27173-z OPEN

A supergene underlies linked variation in color and morphology in a Holarctic songbird

Erik R. Funk [1✉], Nicholas A. Mason[2], Snæbjörn Pálsson [3], Tomáš Albrecht [4,5], Jeff A. Johnson [6] & Scott A. Taylor [1]

Funk et al. 2021



- ~55-Mb inversion of chromosome 1 („supergene")
- multiple candidate genes related to melanogenesis, carotenoid coloration, and bill shape
- latitudinal gradient in ecotype distribution – balanced polymorphism of supergene haplotypes

Závěr

- Molekulární ekologie se rychle vyvíjí
- Metody se zásadně vylepšují a mění
- Co platilo dnes, nemusí platit zítra – těšme se tedy na zítřek