

Bioinformatika



hledání významu biologických dat



Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE



Marian Novotný



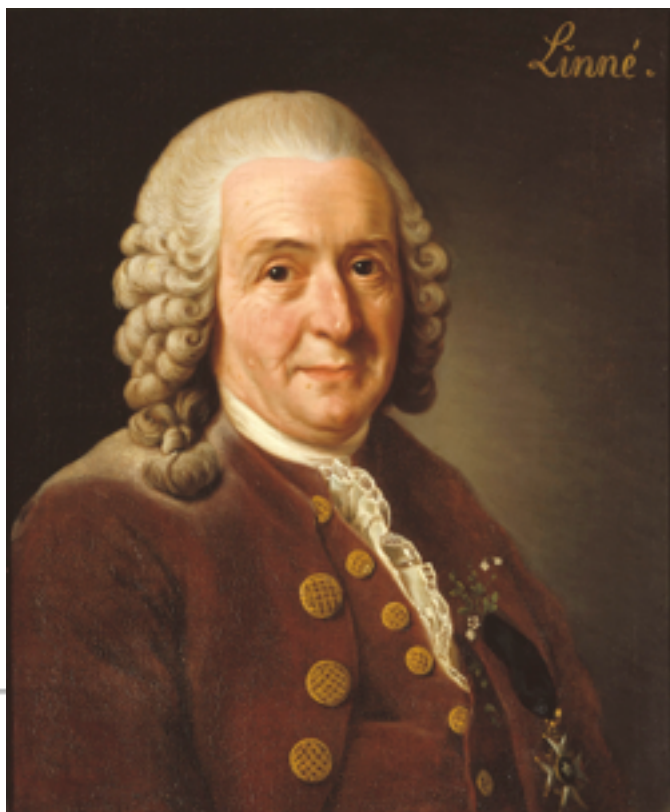
Bioinformatika

- sběr biologických dat
- archivace biologických dat
- organizace biologických dat
- interpretace biologických dat





Biologové sbírají data



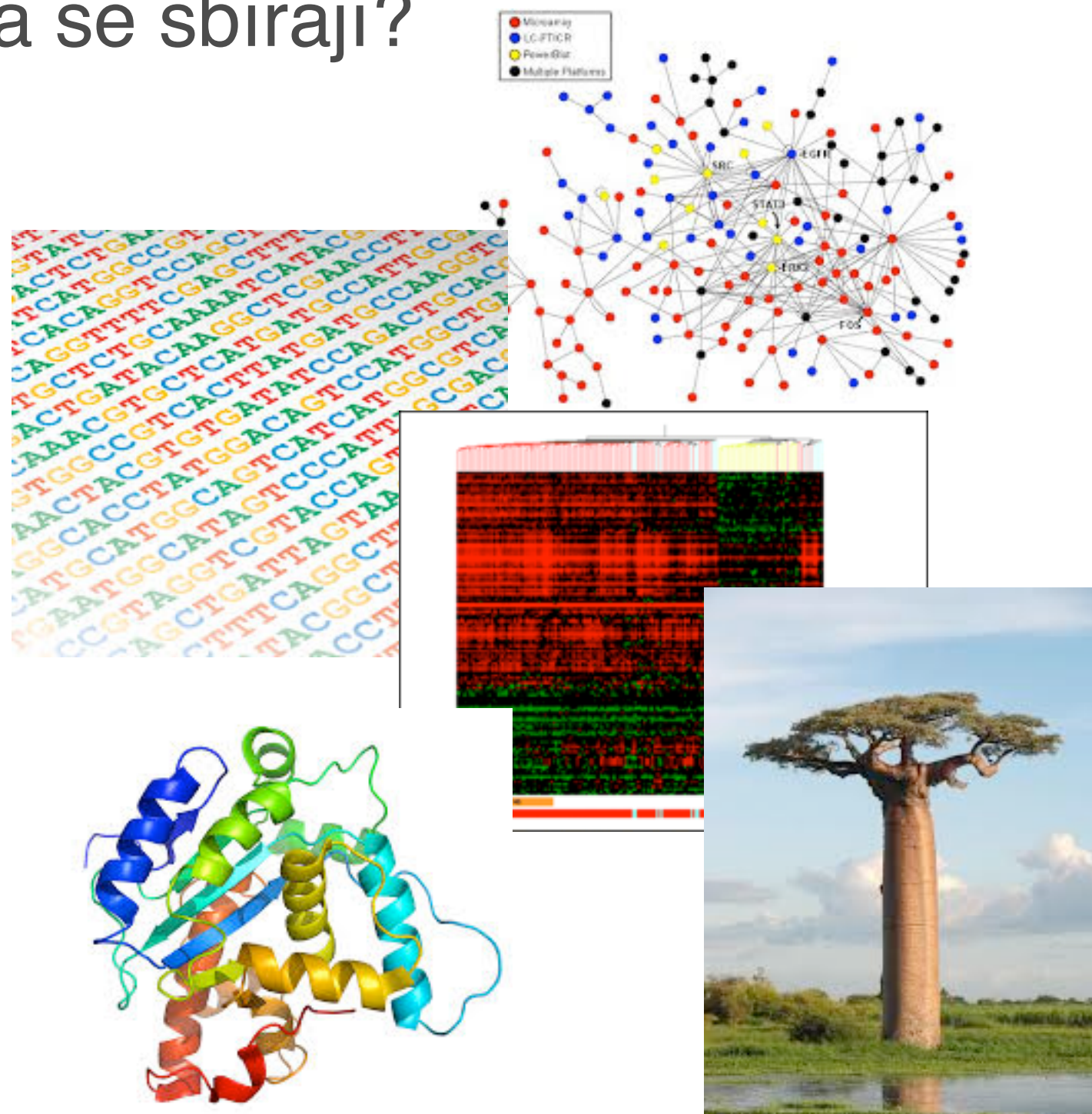
- Carl Linné (1707-1778) - sbíral, archivoval a organizoval data
- středověké botanické zahrady - již v 16. a 17. století - dovoz, produkce a výměna rostlin





Jaká biologická data se sbírají?

- organismy
- měření, vážení
- sekvence
- 3D struktury
- interakce
- příbuzenské vztahy
- genomické mapování
- expresní profily





Data mění biologii

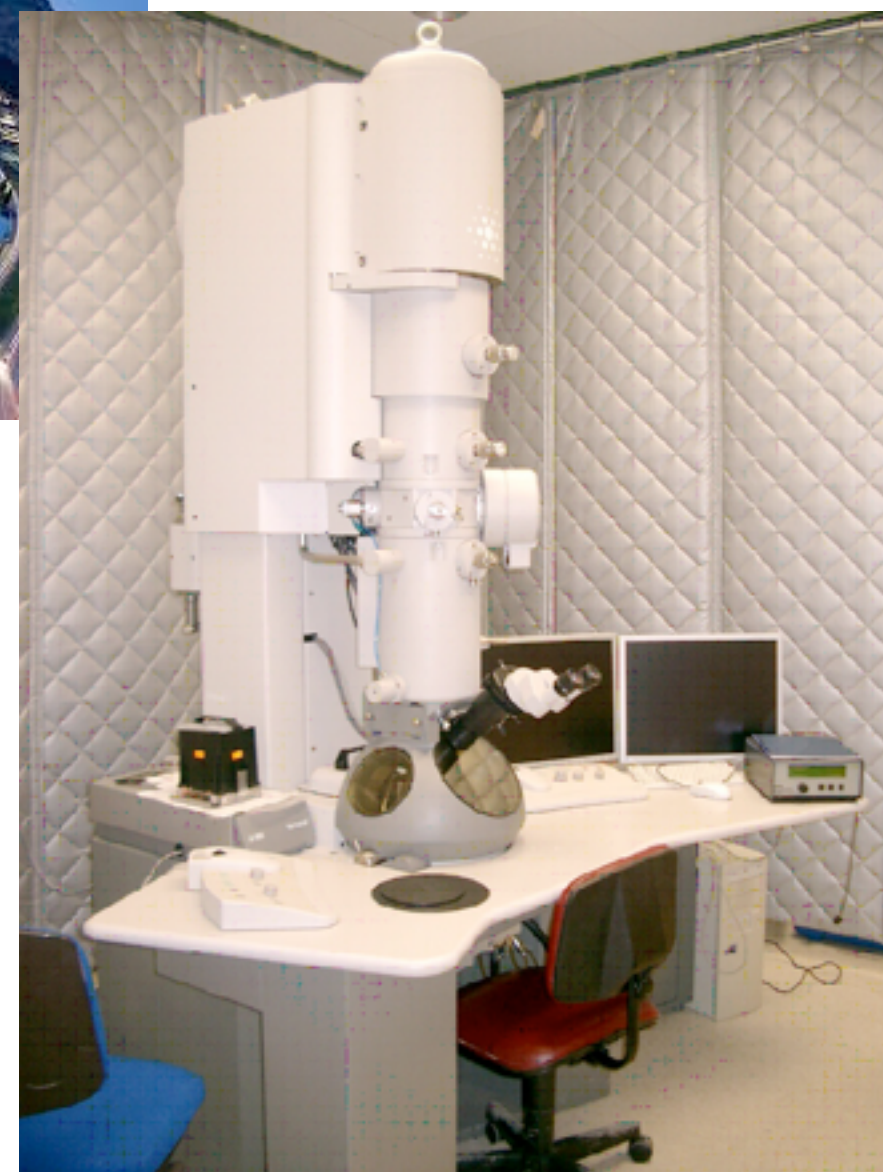




Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE



Kde se data berou?





Lidský genom

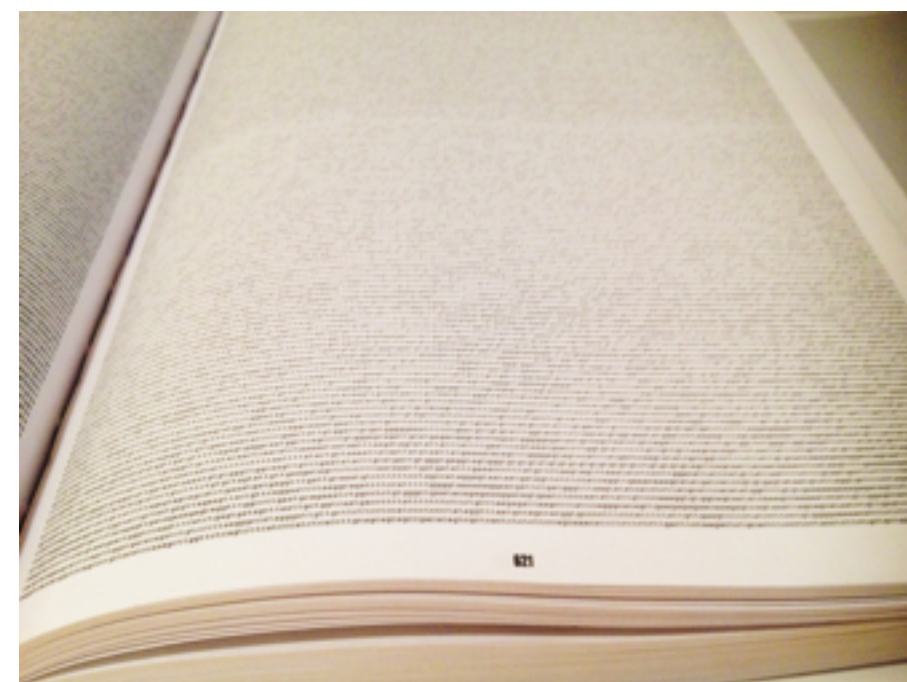
- kompletní sekvence lidské DNA
- 3.2 Gb dat
- ~20 000 genů
- 1 SNP na 1300 nukleotidů -> 3 000 000 nukleotidů rozdíl mezi dvěma jedinci





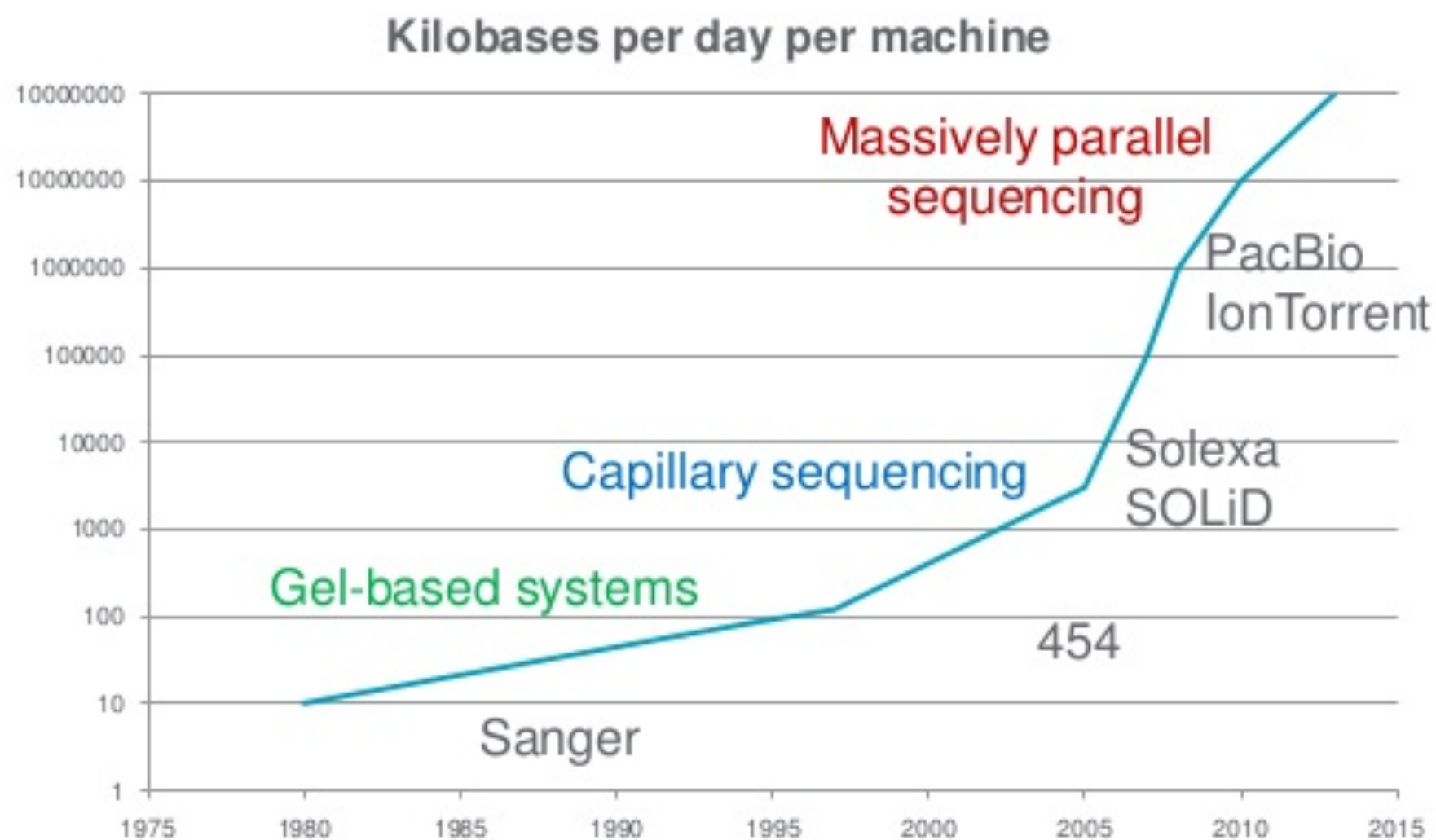
Lidský genom

- kompletní sekvence lidské DNA
- 3.2 Gb dat
- ~20 000 genů
- 1 SNP na 1300 nukleotidů -> 3 000 000 nukleotidů rozdíl mezi dvěma jedinci





Rychlost sekvenování





Cena sekvenování lidského genomu



2001



2007



2015



Sekvenování a data

osekvenované lidské genomy v roce 2014 $228k \times 3 \times 10^9$ bazí
 $= 7 \times 10^{14}$ bazí

genom všech občanů ČR $10m \times 3 \times 10^9 = 3 \times 10^{16}$ bazí

100 000 genomes - UK

celá populace - Faorské ostrovy

velké skupiny obyvatel - Finsko, Island, Německo, Španělsko

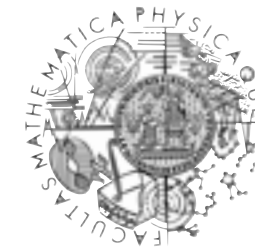


1 pb dat = 1 000 000 000 000 000

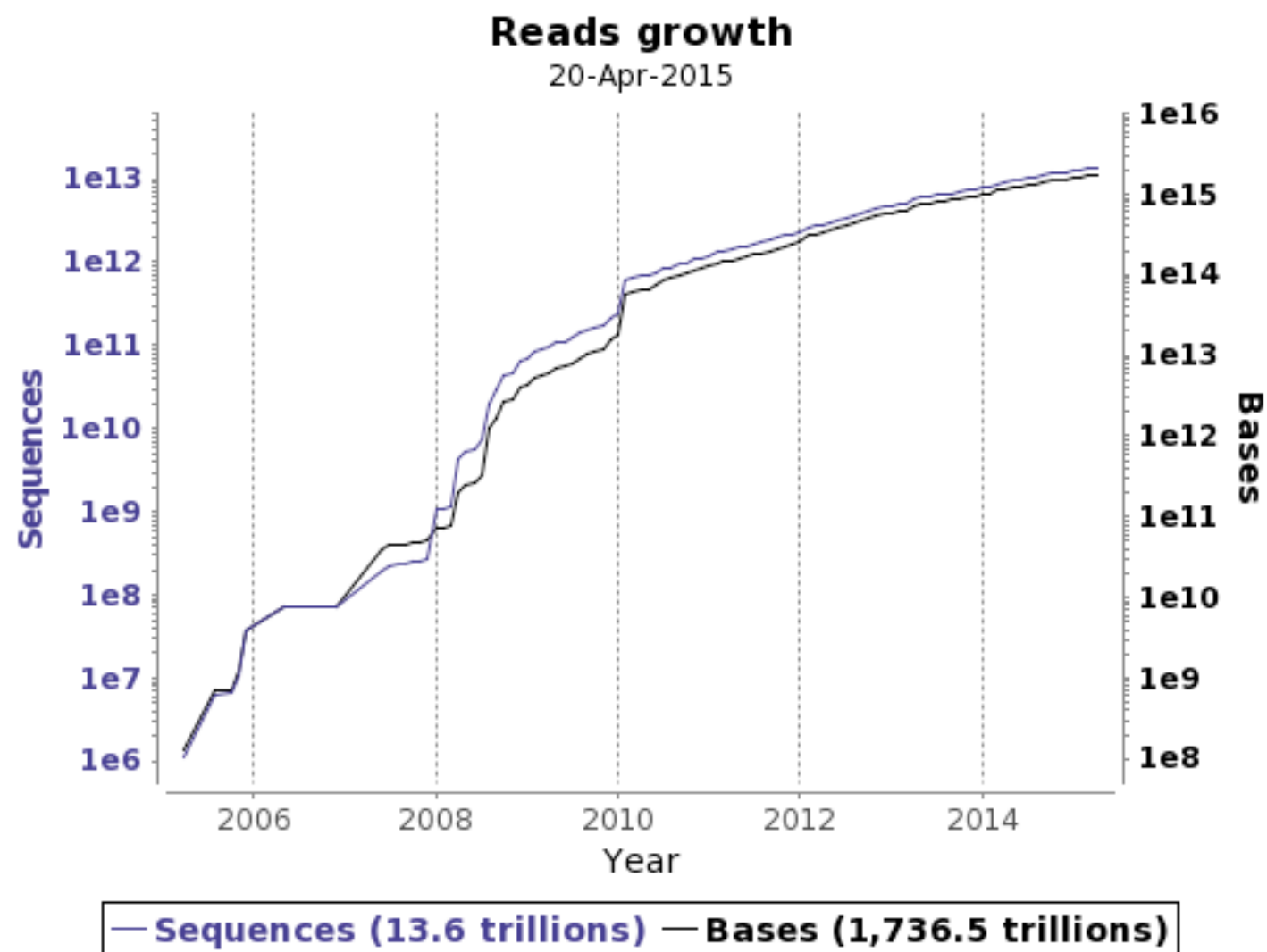




- 1 PB hudby v mp3 formátu by hrál 2000 let
- velikost paměti lidského mozku je asi 2.5 PB dat
- Google zpracoval v roce 2009 okolo 24 PB dat denně
- EBI má nyní kapacitu 60 PB dat
- CERN do roku 2012 nashromáždil 200 PB dat

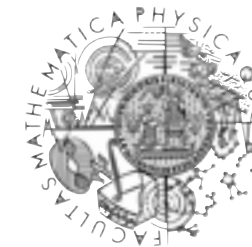


DNA databáze ENA





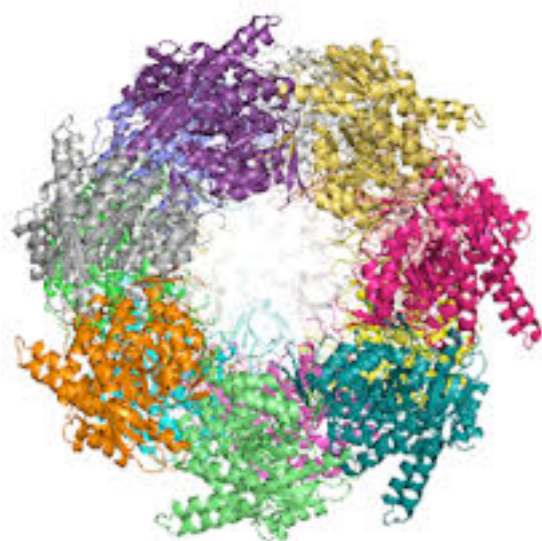
Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE



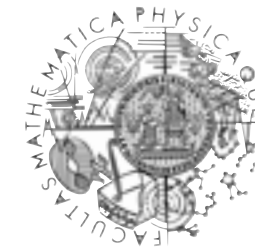
Databáze 3D struktur

108 124 struktur

- 1973-1999
- 2000-2007
- 2008-2011
- 2012-2014
- 2015

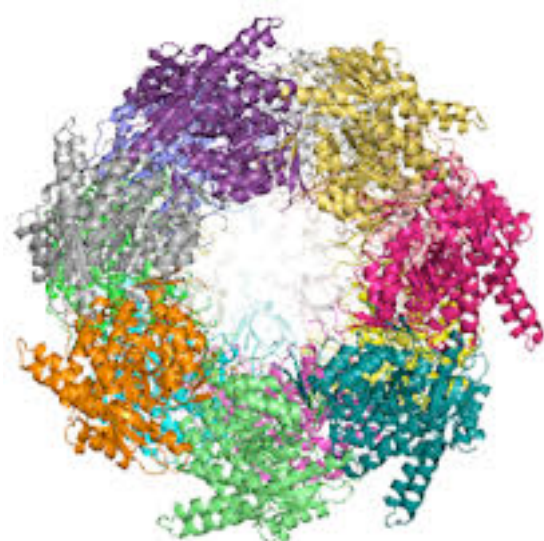
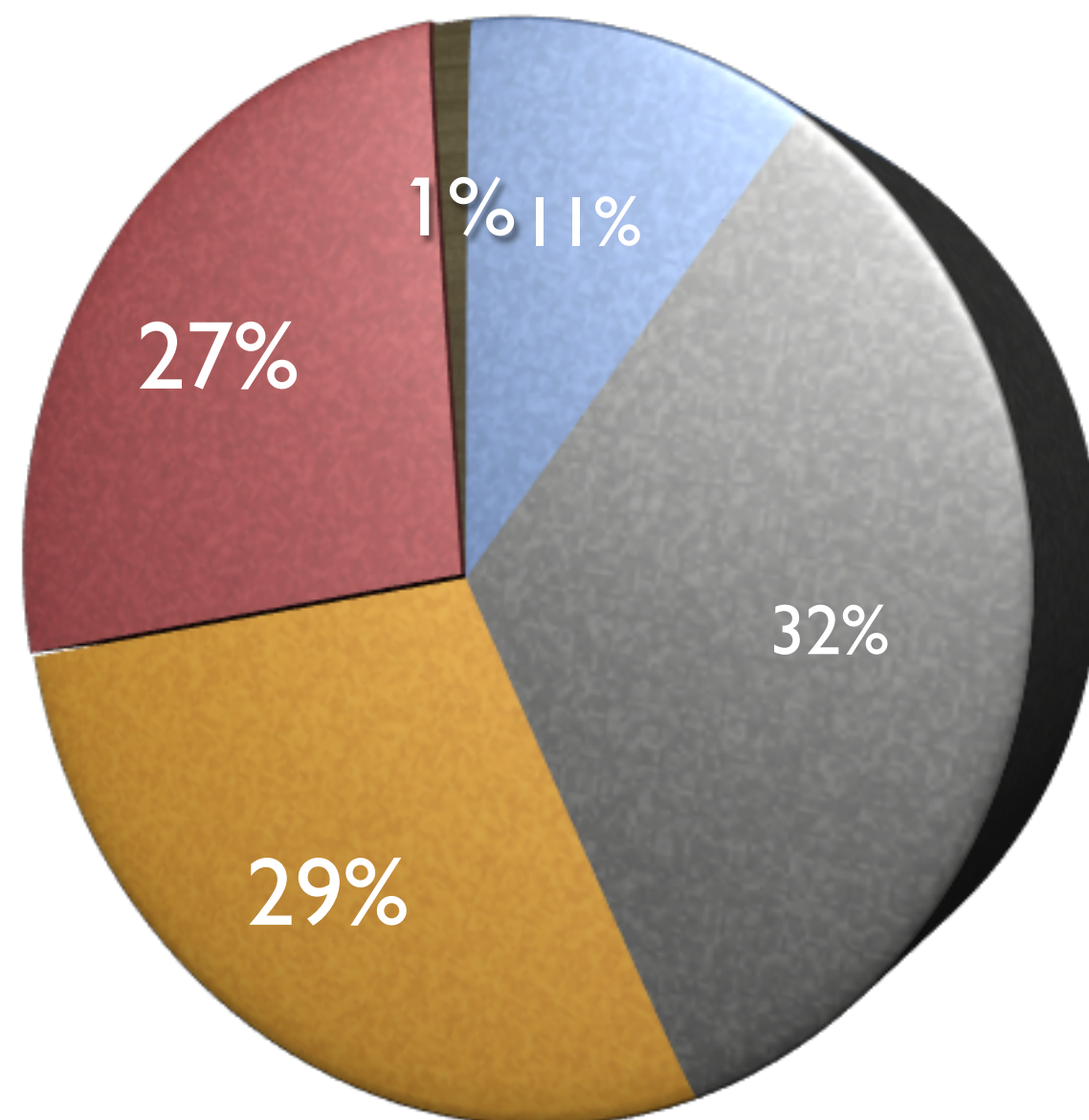
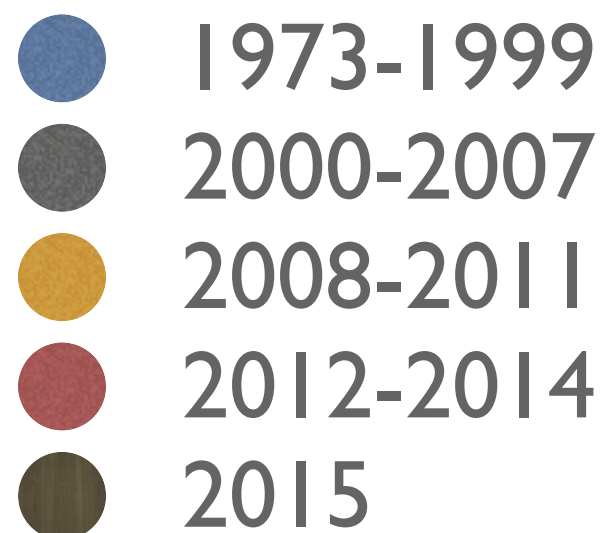


The chaperone GroEL



Databáze 3D struktur

108 124 struktur



The chaperone GroEL

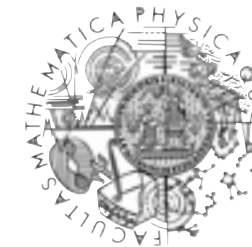


Data nedávají odpovědi

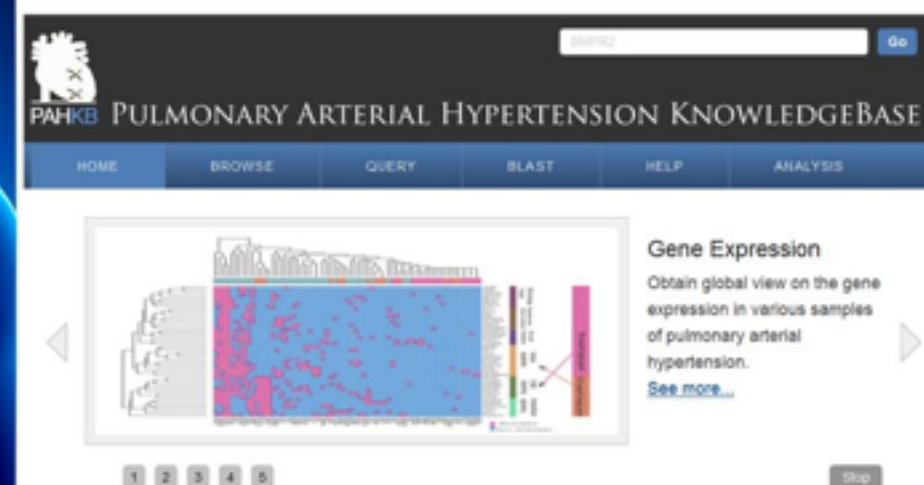
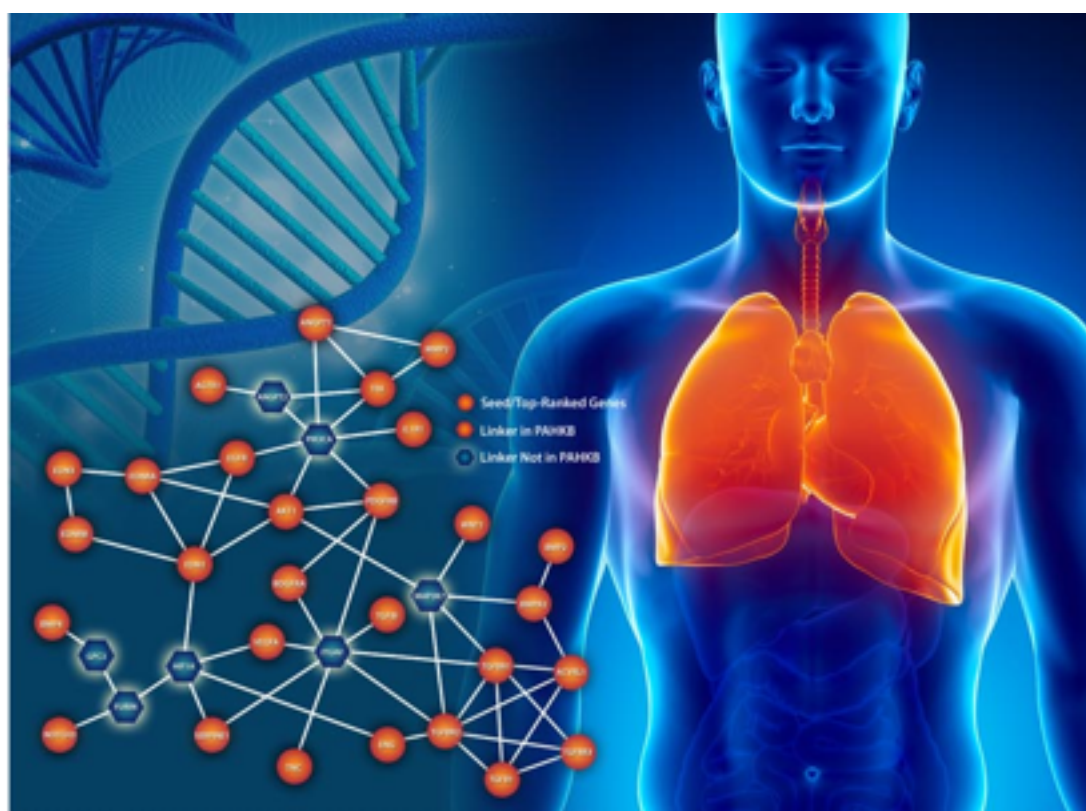
Jak využít dostupná data k našemu prospěchu?



- bioinformatika v medicíně
- bioinformatika v zemědělství
- bioinformatika v průmyslu



Bioinformatika v medicíně





OMIM[®]

Online Mendelian Inheritance in Man[®]

An Online Catalog of Human Genes and Genetic Disorders

Updated 20 April 2015

Advanced Search : [OMIM](#), [Clinical Synopses](#), [Gene Map](#)

Need help? : [Example Searches](#), [OMIM Search Help](#), [OMIM Tutorial](#)

Mirror sites : [us-east.omim.org](#), [europe.omim.org](#)

[Advanced Search ▾](#)

ALLELIC VARIANTS (23 Selected Examples):

[Table View](#)

.0001 LUNG CANCER, SOMATIC

KRAS, GLY12CYS [dbSNP:rs121913530] [ClinVar]

In a cell line of human lung cancer (211980), [Nakano et al. \(1984\)](#) identified a 34G-T transversion in exon 1 of the KRAS2 gene, resulting in a gly12-to-cys (G12C) substitution. Studies of the mutant protein showed that it had transforming abilities consistent with activation of the gene. 

In a study of 106 prospectively enrolled patients with primary adenocarcinoma of the lung, [Ahrendt et al. \(2001\)](#) found that 92 (87%) were smokers. KRAS2 mutations were detected in 40 of 106 tumors (38%) and were significantly more common in smokers compared with nonsmokers (43% vs 0%; $P = 0.001$). Thirty-nine of the 40 tumors with KRAS2 mutations had 1 of 4 changes in codon 12, the most common being G12C, which was present in 25 tumors. 

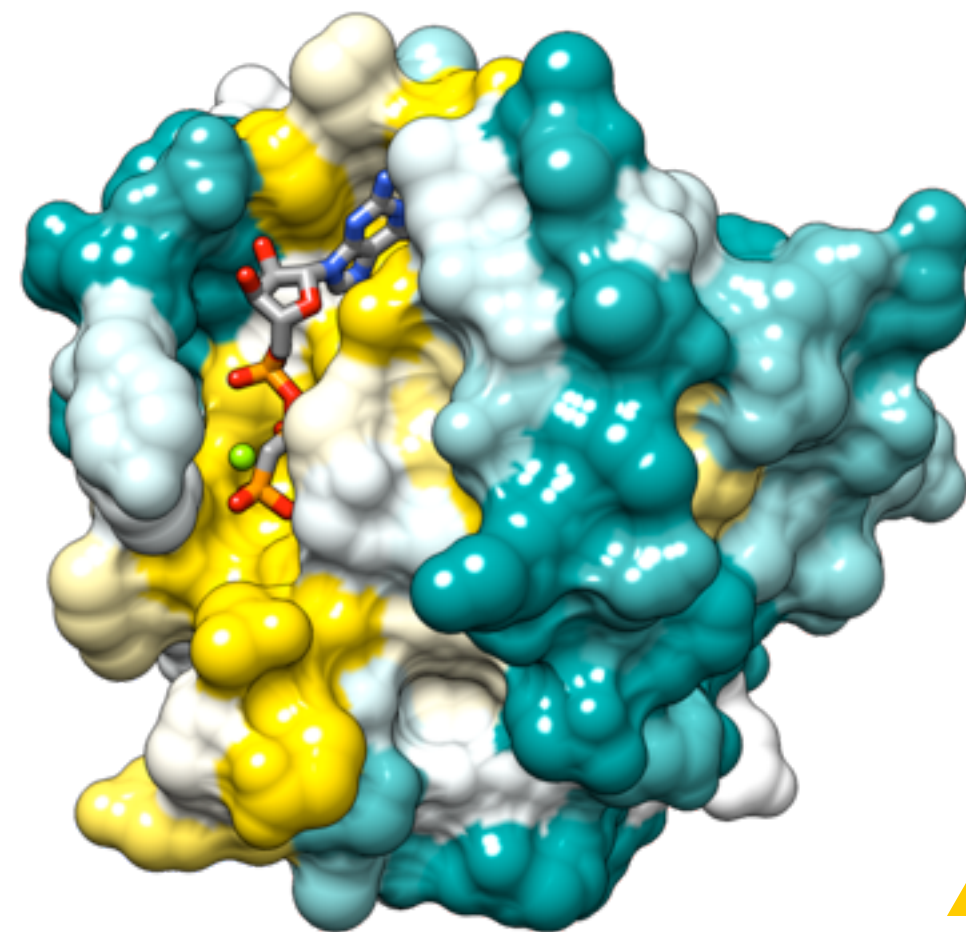
▾ Table of Contents for

- [Title](#)
- [Gene-Phenotype Relationship](#)
- [Text](#)
- [Description](#)
- [Cloning and Expression](#)
- [Gene Structure](#)
- [Mapping](#)
- [Gene Function](#)
- [Molecular Genetics](#)
- [Genotype/Phenotype Correlation](#)
- [Population Genetics](#)
- [Animal Model](#)
- [Allelic Variants](#)
- [Table View](#)
- [See Also](#)



Ras onkogen

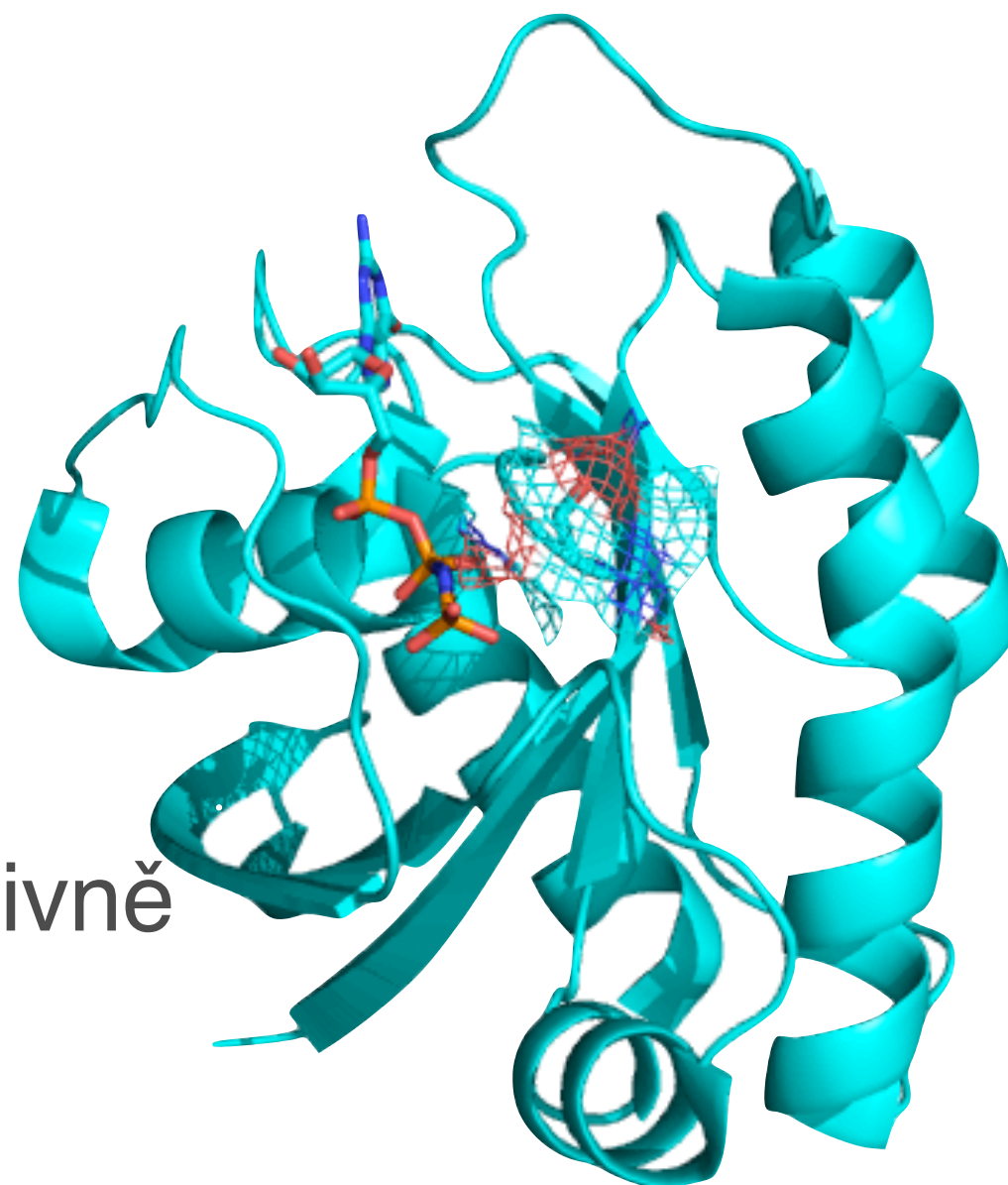
- jeden z ortologů malé GTPázy Ras
- molekulární přepínač
- mutován v 20 % lidských nádorů
- mutován až v 90 % případů u specifických typů nádorů



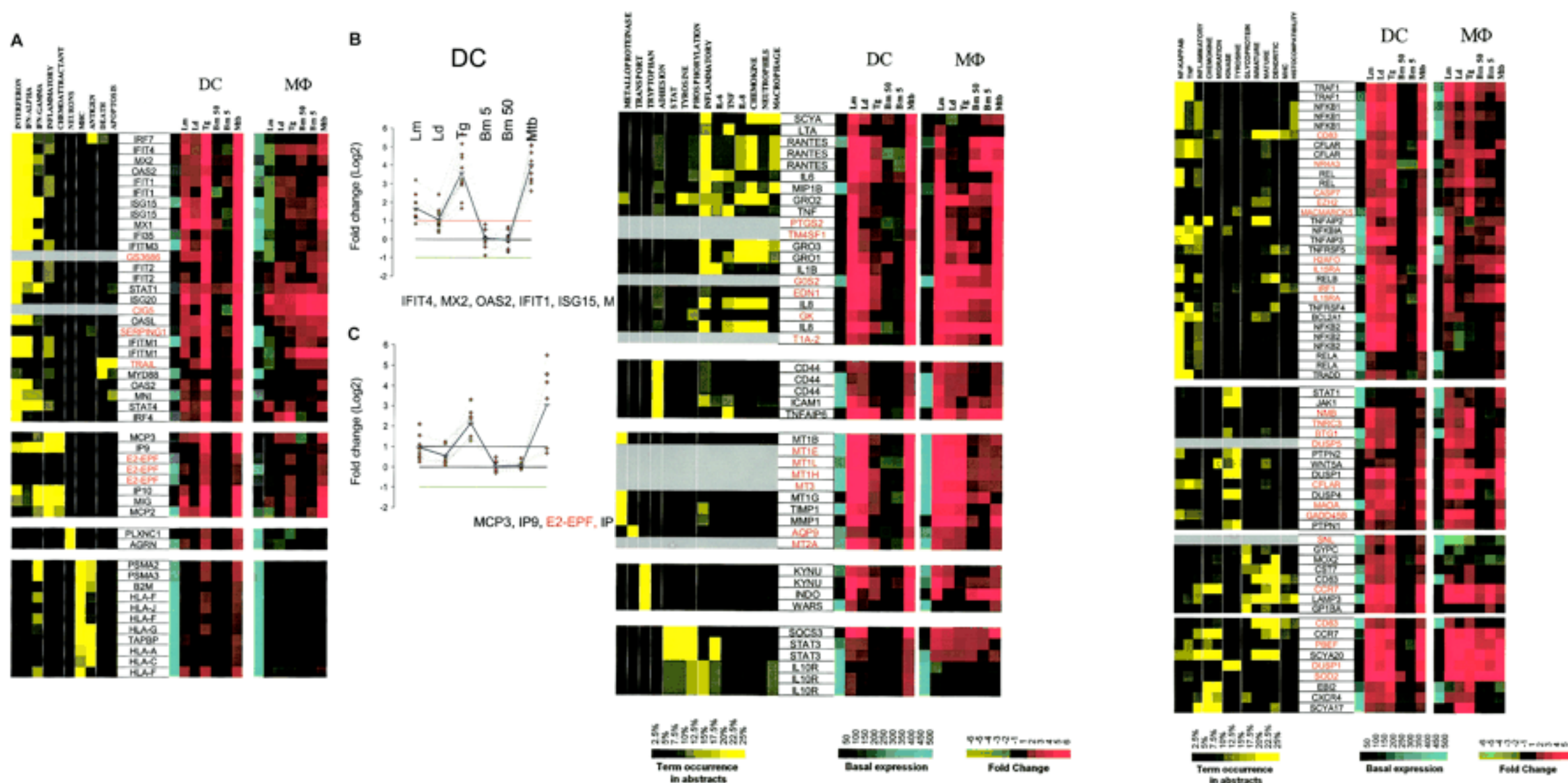


Až 25% případů nádorů plic má mutaci G12C v K-Ras

- obtížně zasažitelný léky
- hladký povrch
- zastupitelnost ortologů
- mutace glycinu činí protein konstitutivně aktivní
- vývoj inhibitoru mutace G12C -> personalisovaná medicína

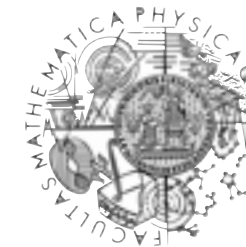


Identifikace pathogena z buněčné odpovědi

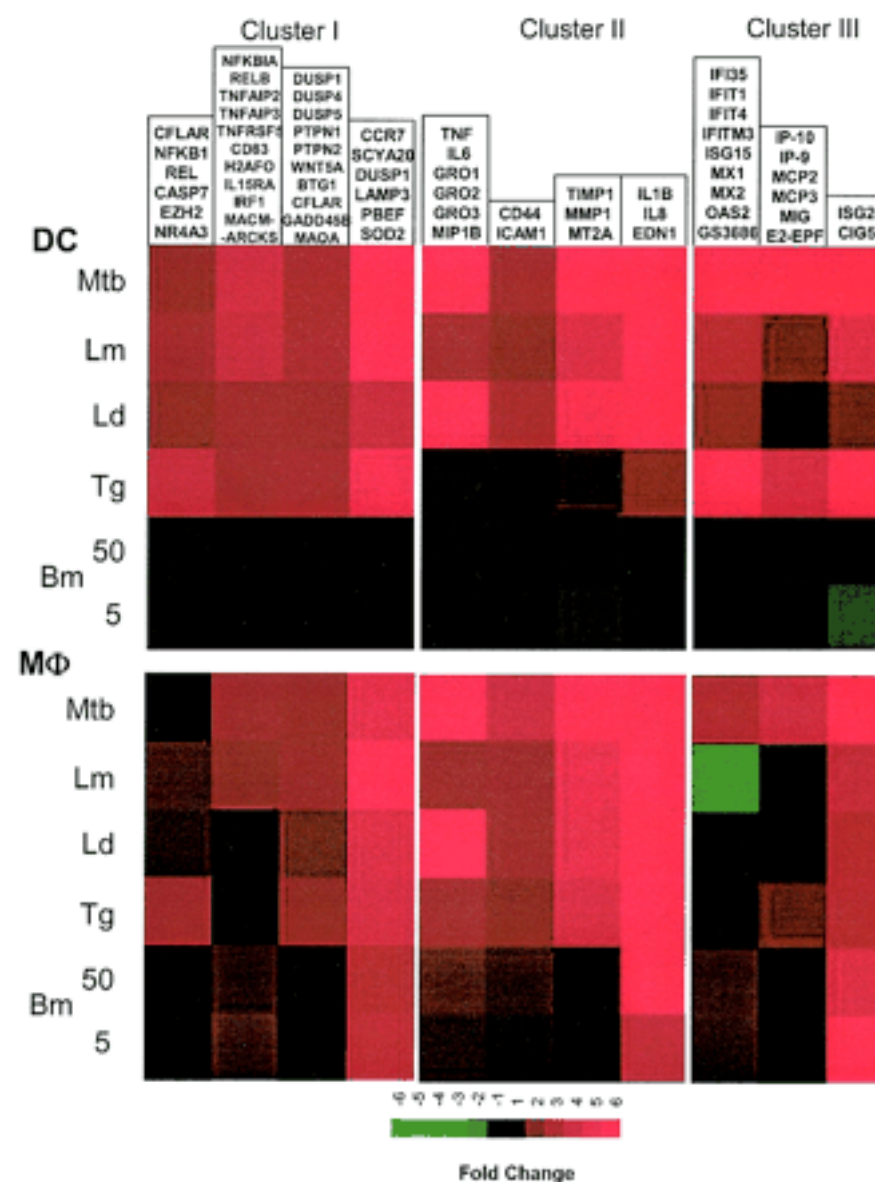




Identifikace pathogena z buněčné odpovědi



Identifikace pathogena z buněčné odpovědi





Bioinformatika a zemědělství

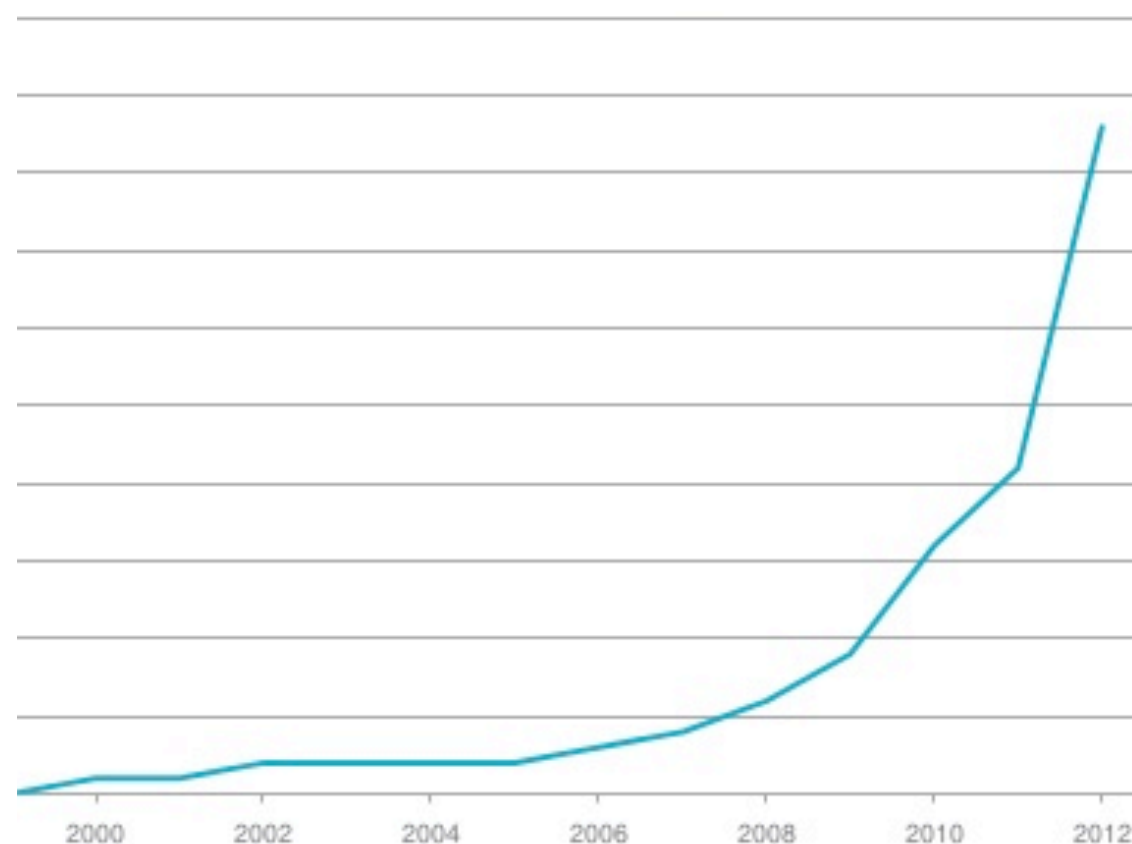




Sekvenování významných plodin a zvířat

- rýže, pšenice, kukuřice, cukrová třtina
- rajče, banán, jahoda, víno
- dodnes asi 90 druhů rostlin
- kráva, ovce, treska
- kuře, prase

Published Plant Genomes





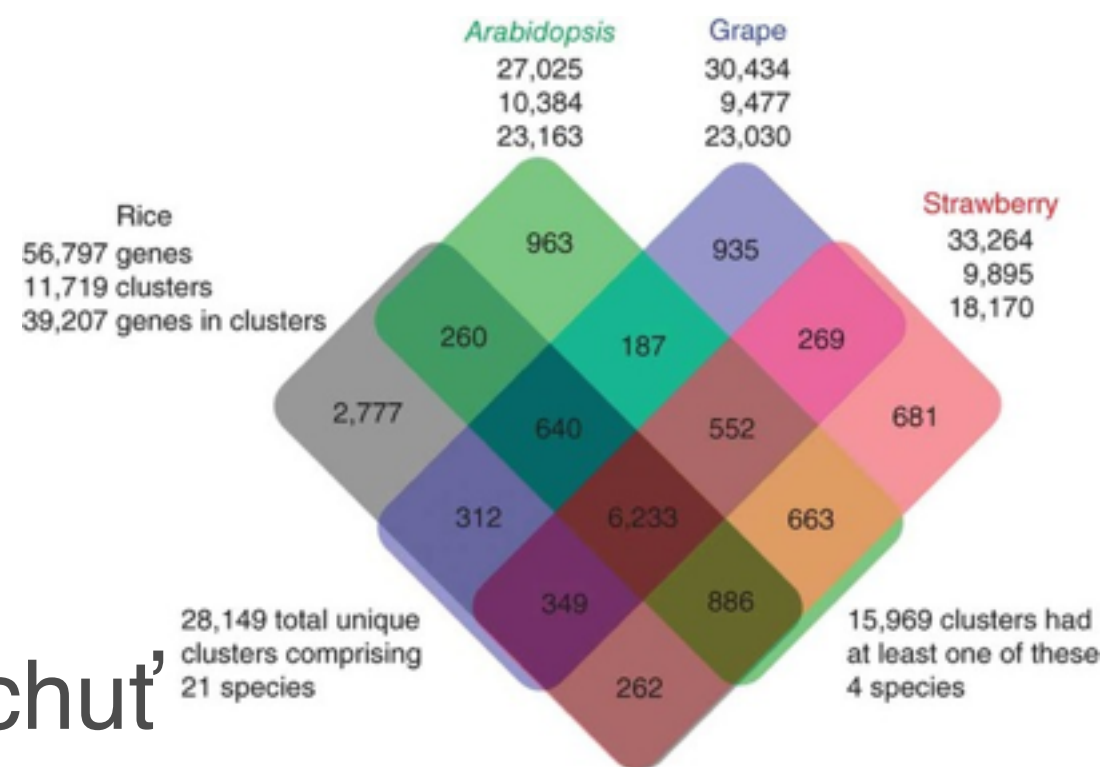
Cíle

- zvýšit odolnost plodin i živočichů vůči stresu abiotickému (sucho, salinita) i biotickému (škůdci, patogeny)
- zvýšit produkci
- snížit náklady



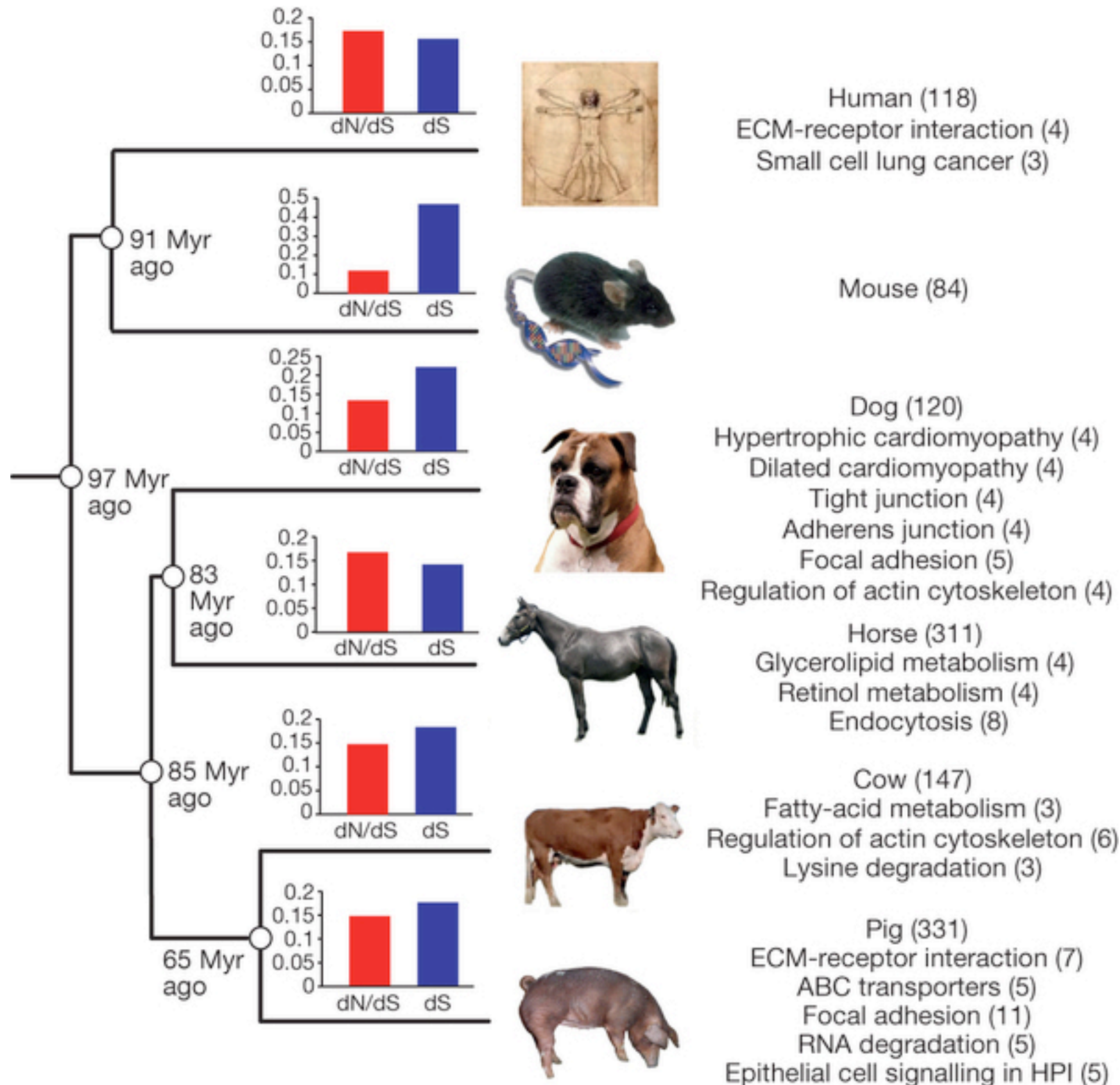
Jahoda

- jen 250 let stará plodina
- malý genom (240 MB)
- >33,000 genů
- 681 “jahodích” genů - vůně, chuť





Rychle se vyvíjející geny





Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE



Bioinformatika a průmysl



Těžba ropy

- část ropných ložisek je vždy nevytěžitelná
- vysoká viskozita neumožňuje vytažení ropy
- **Microbial Enhanced Oil Recovery (MEOR)** - technika manipulace s mikrobiálními kulturami ložisek
- in silico identifikace vhodných genů a kmenů bakterií k redukci nevytěžitelných zásob





Shrnutí



- metodické pokroky v biologii vedou k explozi dostupných dat
- uložení a zpracování těchto dat je obrovskou výzvou
- data jsou základem experimentů
- využití těchto dat je naprosto nepředstavitelné bez zapojení informatických přístupů
- bioinformatika je využívána již dnes v medicíně, zemědělství i průmyslu
- pocítujeme však nedostatek kvalifikovaných odborníků, kteří mohou dát dostupným biologickým datům smysl
- rutinní sekvenování lidí dále zvýší potávku po bioinformaticích



Biology faces a quantum leap into the incomprehensible

“In the past, we have seen biological explanations as commonsense. They have explained how genes direct the manufacture of proteins or account for the appearance of disease through the behaviour of bacteria and viruses. But such simplicity is likely to disappear in the near future.”

Paul Nurse

(<http://www.youtube.com/watch?v=Te8bXvinilA>)





Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE



děkuji za pozornost



Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE





Přírodovědecká fakulta
UNIVERZITY KARLOVY V PRAZE

