

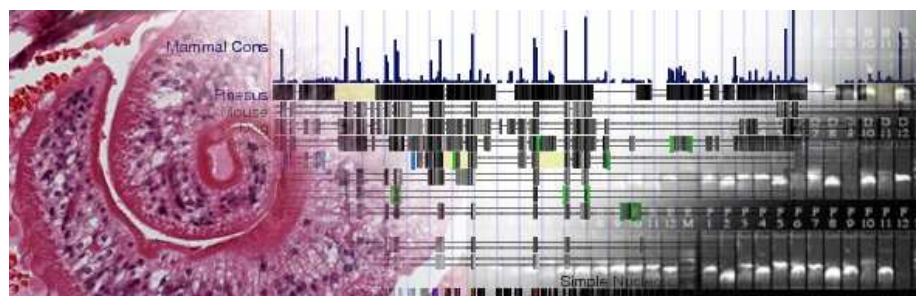
# Hľadanie génov v nepreskúmaných genómoch

Broňa Brejová

Katedra informatiky

Fakulta matematiky, fyziky a informatiky

Univerzita Komenského v Bratislave



## Hľadanie génov

**Gény:** úseky genómu kódujúce proteíny (alebo RNA molekuly)

**Hľadanie génov:** identifikácia týchto úsekov v osekvenovanom genóme

**Cieľ:** katalóg všetkých proteínov daného organizmu

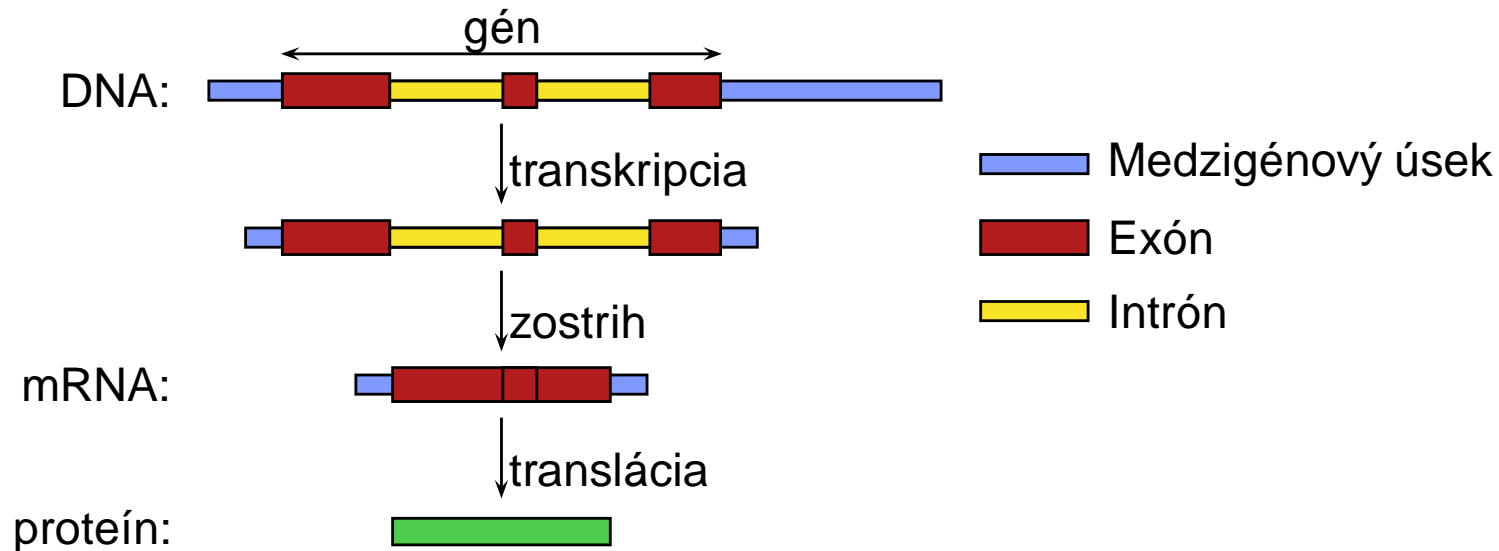
## Hľadanie génov

**Gény:** úseky genómu kódujúce proteíny (alebo RNA molekuly)

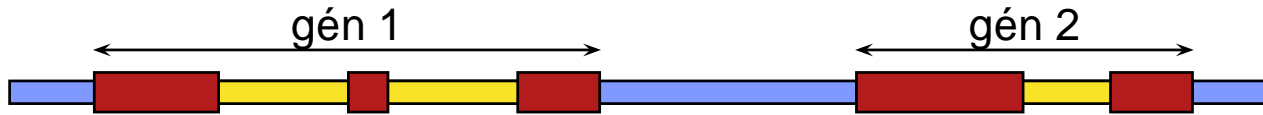
**Hľadanie génov:** identifikácia týchto úsekov v osekvenovanom genóme

**Cieľ:** katalóg všetkých proteínov daného organizmu

### Eukaryotická génová štruktúra

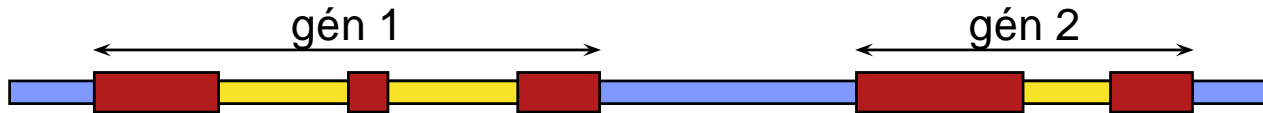


## Informácia použiteľná na hľadanie génov



- **Genomická sekvencia:** rozdiely v štatistických vlastnostiach exónov, intrónov a medzigénových úsekov  
Reprezentujeme štatistickým modelom (skryté Markovove modely)  
[Krogh a kol. 1994, Burge a Karlin 1997]

## Informácia použiteľná na hľadanie génov



- **Genomická sekvencia:** rozdiely v štatistických vlastnostiach exónov, intrónov a medzigénových úsekov  
Reprezentujeme štatistickým modelom (skryté Markovove modely)  
[Krogh a kol. 1994, Burge a Karlin 1997]
- **Evolučné stopy:** exóny sa menia pomalšie, prevaha synonymných mutácií
- **Výsledky experimentov:** overenie transkripcie, podobnosť so známymi proteínmi

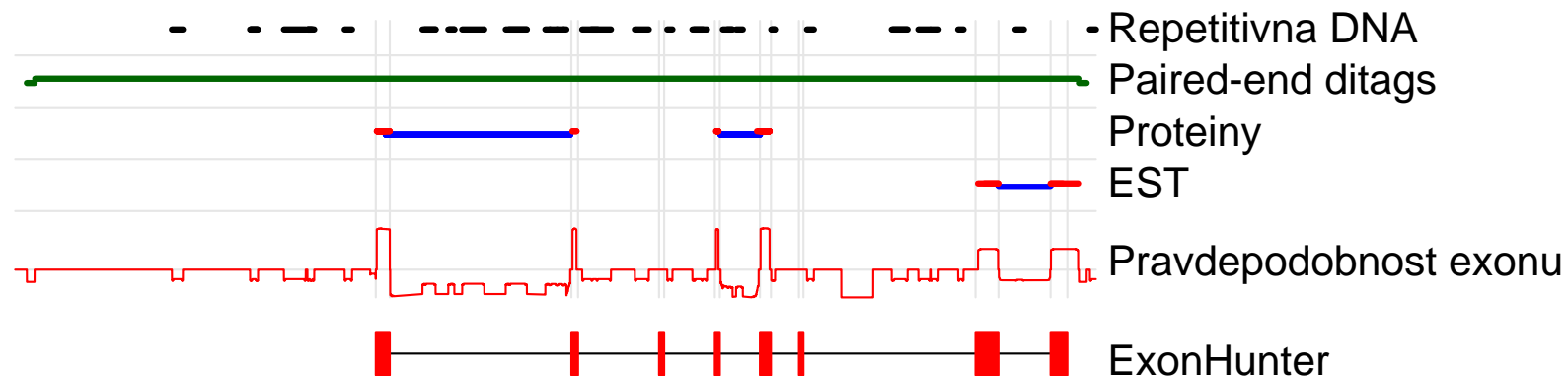
## ExonHunter [Brejová a kol. 2005 ISMB]

Praktický program na hľadanie eukaryotických génov

HMM pre genómovú sekvenciu kombinovaný s ďalšími informáciami

### System "radcov" na kombinovanie externej informácie

- Pracuje s heterogénnymi a nespoľahlivými dátami
- Pre každú pozíciu určí pravdepodobnosť exónu, intrónu, medzigénového úseku
- Ak nie je externá informácia, použije iba HMM

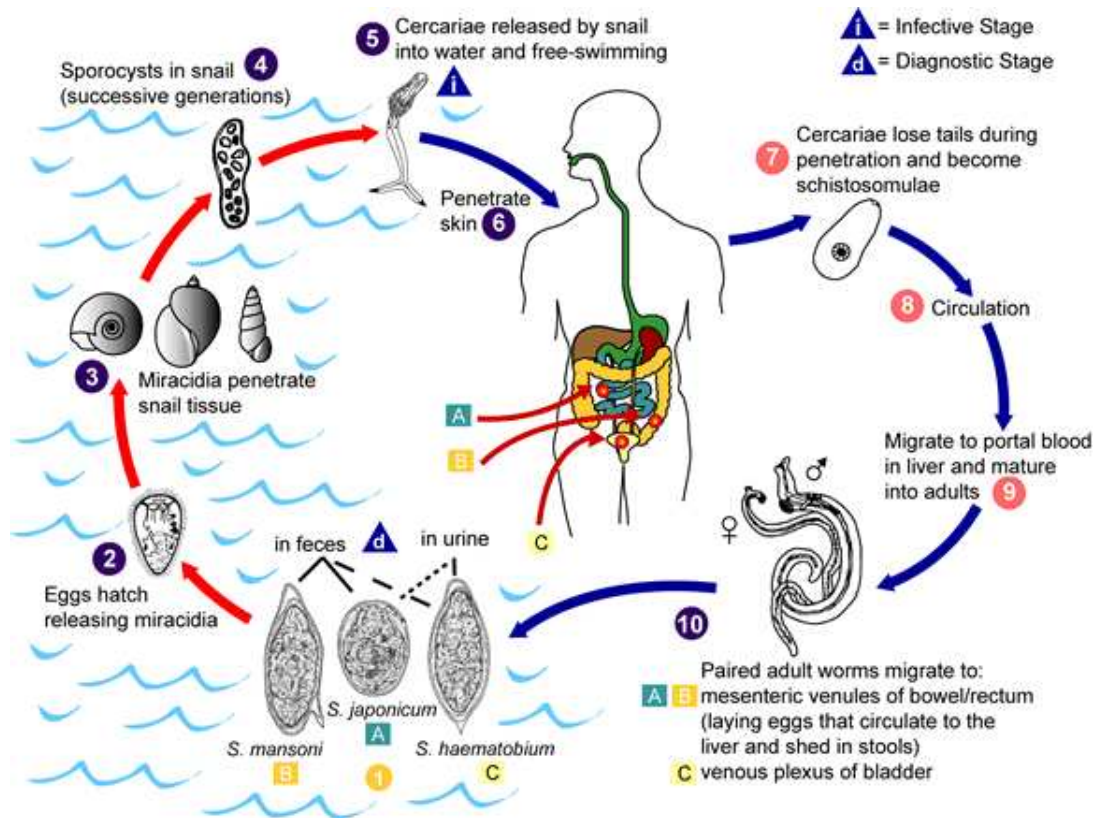


# Hľadanie génov v genóme *Schistosoma japonicum*

Parazitický červ (ploskavec) z východnej Ázie

Schistosomiáza ročne spôsobuje 14 000 úmrtí, ohrozuje 66 miliónov ľudí

Sekvenovanie genómu Chinese Human Genome Center v Šanghaji

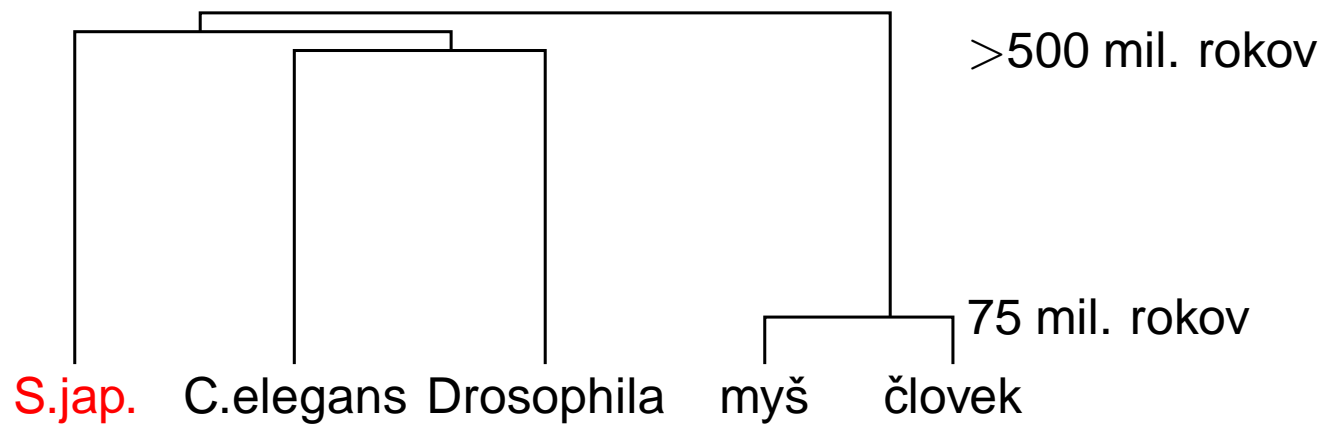


## Hľadanie génov v novo-sekvenovaných genómoch

Programy na hľadanie génov treba natrénovať na každý nový genóm

### Obvyklé postupy:

- Použijeme trénovaciu vzorku ručne anotovaných génov (neexistovala pre *S. japonicum*)
- Trénujeme na príbuznom dobre anotovanom genóme (tiež neexistuje pre *S. japonicum*)

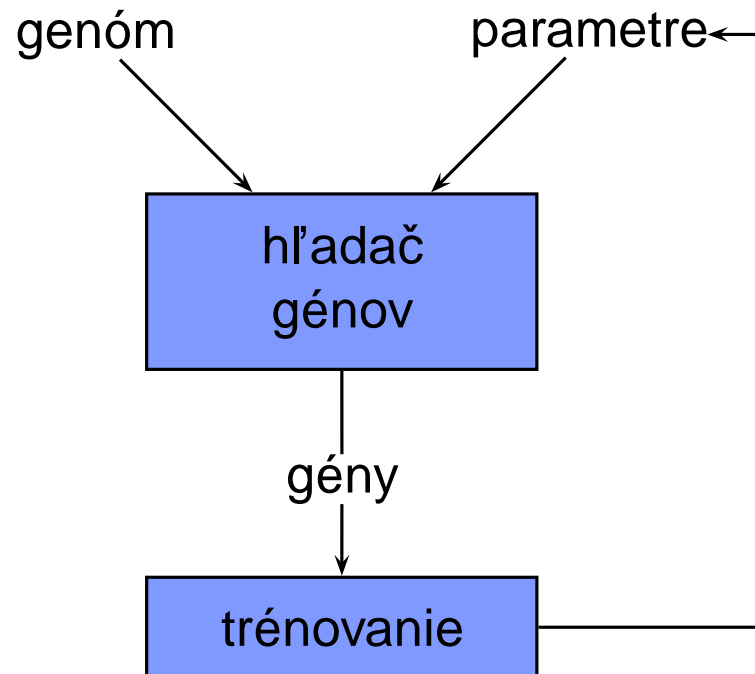




# Hľadanie génov v nových genómoch

## Predchádzajúce práce:

Iteratívne trénovanie na nájdených génoch [Korf 2004, Lomsadze 2005]



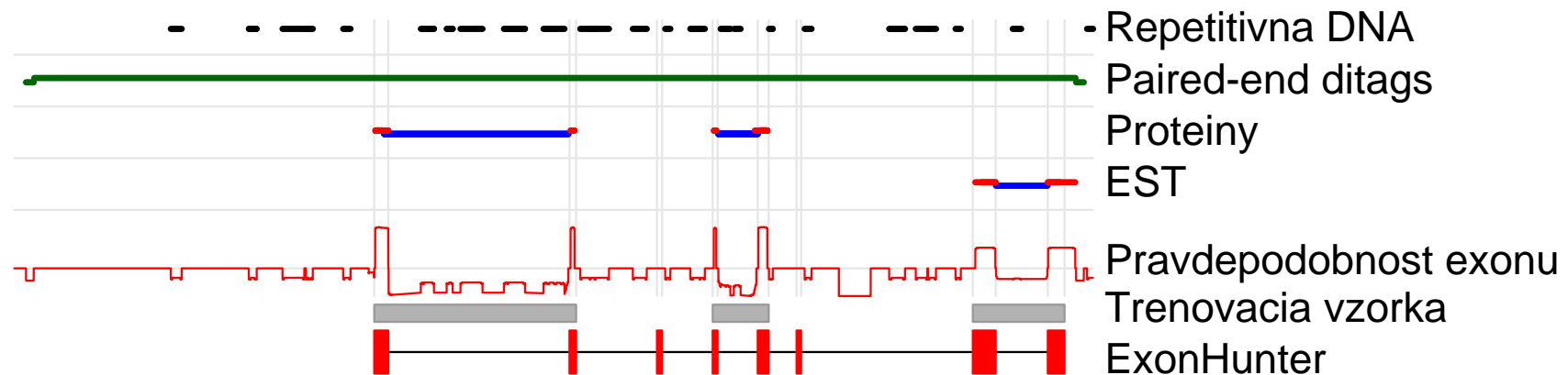
# Hľadanie génov v nových genómoch

## Predchádzajúce práce:

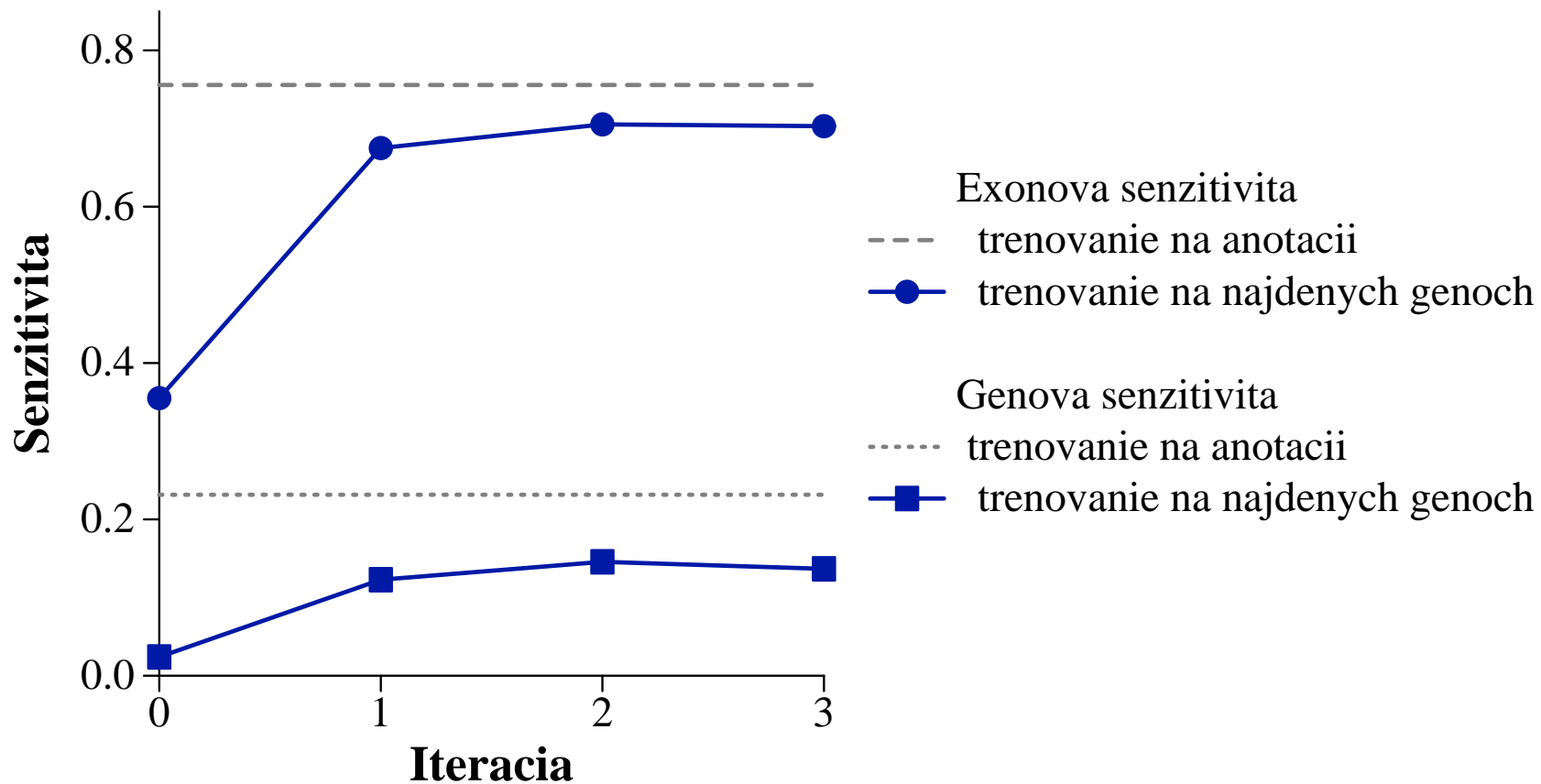
Iteratívne trénovanie na nájdených génoch [Korf 2004, Lomsadze 2005]

## Náš prístup:

- Iteratívne trénujeme na nájdených génoch
- **Použijeme iba tie, ktoré sú podporené externými informáciami**

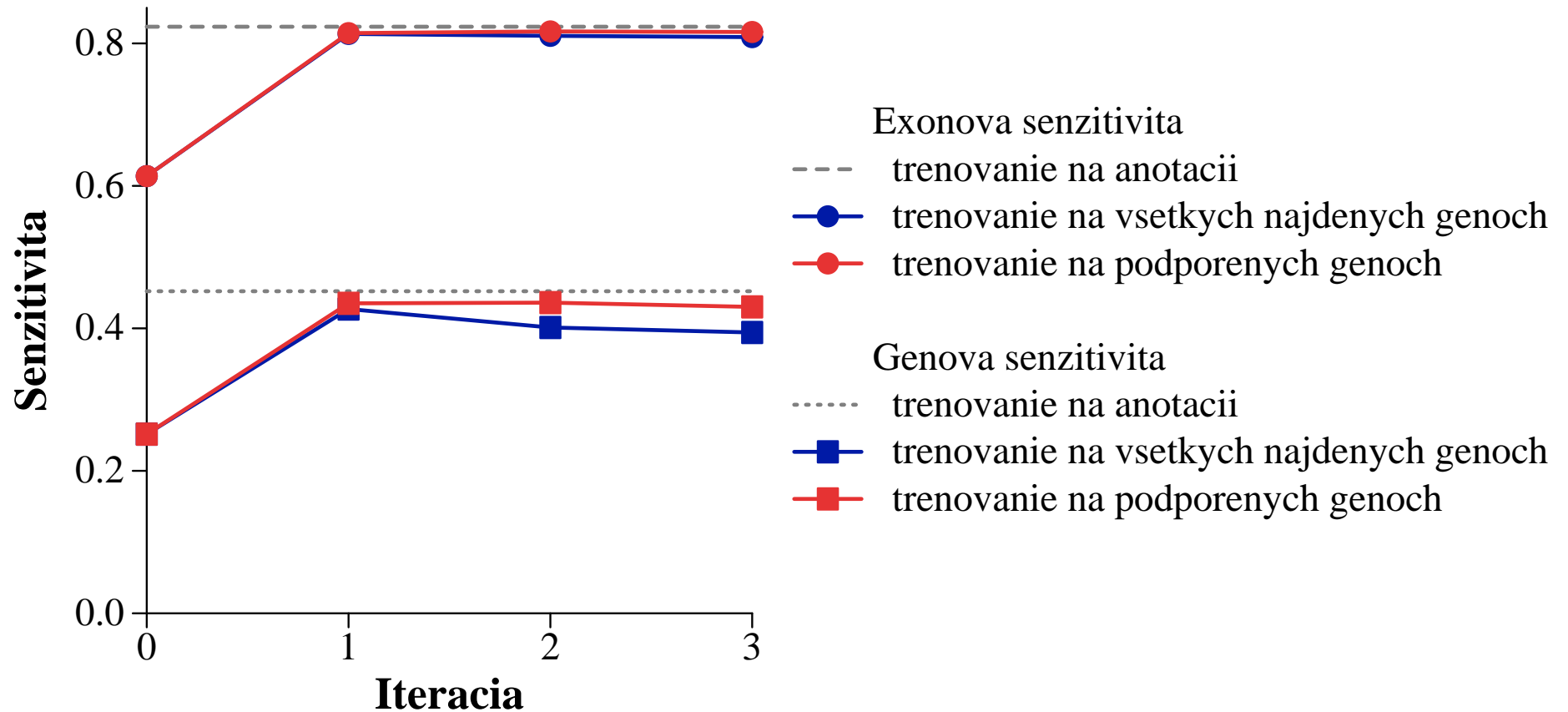


## Výsledky na modelovom organizme *Caenorhabditis elegans*



Iteratívne tréovanie ako v [Korf 2004]

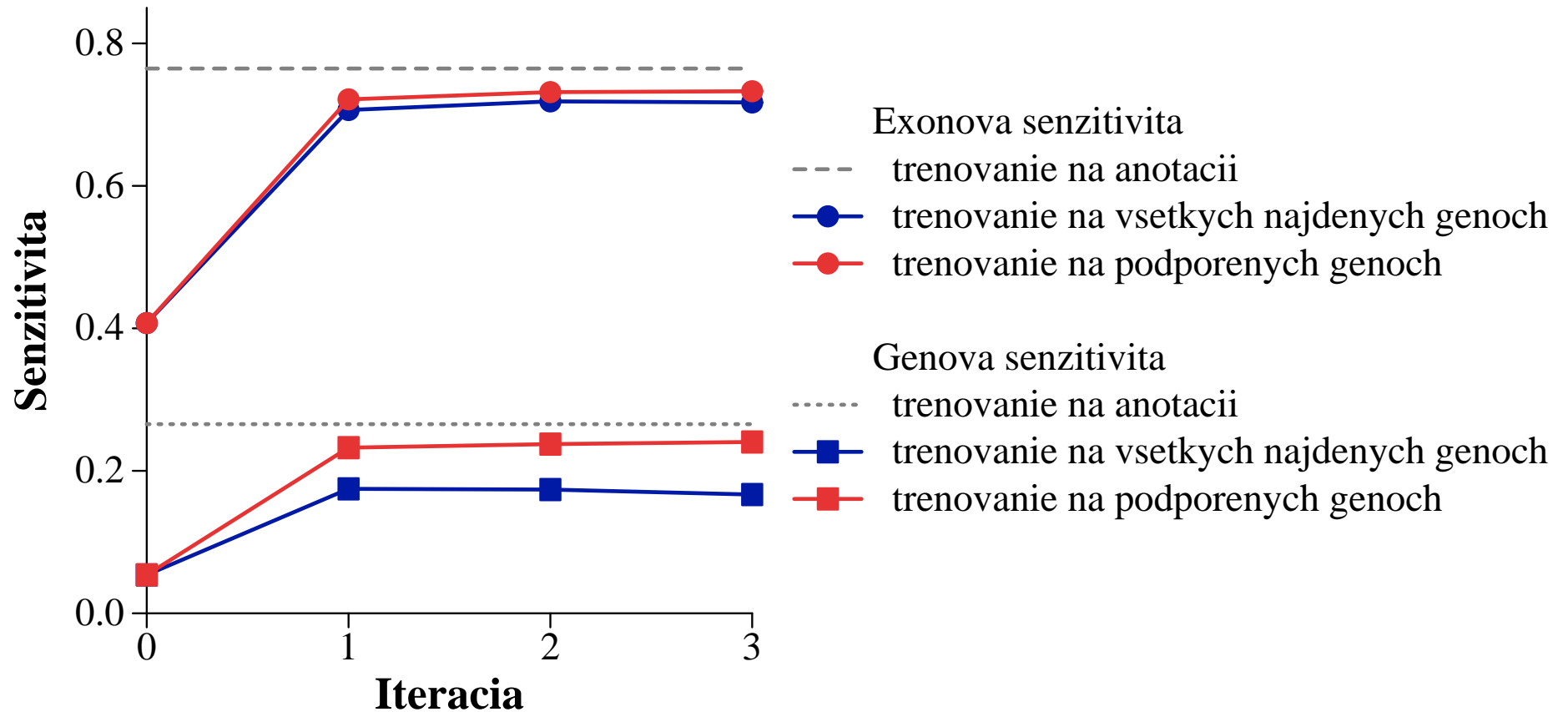
## Výsledky na modelovom organizme *Caenorhabditis elegans*



S použitím všetkých informácií, vrátane proteínov a mRNA z *C. elegans*.

Toľko informácií nie je vždy k dispozícii

## Výsledky na modelovom organizme *Caenorhabditis elegans*



S použitím informácie iba z iných organizmov  
(*C. remanei*, *Drosophila*, človek)

## Hľadanie génov v *S. japonicum*

16 687 nájdených génov v 398M genomickej sekvencii

**Zlepšenie presnosti iteráciou** (951 génov s mRNA sekvenciou)

| Iterácia | Génová senzitivita | Exónová senzitivita |
|----------|--------------------|---------------------|
| 0        | 31%                | 71%                 |
| 1        | 38%                | 75%                 |
| 2        | 38%                | 75%                 |

- ExonHunterové výsledky tvorili základ anotácie
- Umožnili ďalšie štúdie *S.japonicum* proteínov
- Môžu pomôcť pri vývoji vakcín, liečiv, diagnostiky
- Straty a expazie rodín génov kvôli parazitickému životu

## Pod'akovanie

### **Finding Genes in *Schistosoma Japonicum*: Annotating Novel Genomes with Help of Extrinsic Evidence**

B Brejová\*, T Vinař\*, Y Chen,<sup>2</sup> S Wang,<sup>2</sup> G Zhao,<sup>2</sup> DG Brown,<sup>1</sup> M Li,<sup>1</sup> Y Zhou<sup>2</sup> Nucleic Acids Research, 37(7):e52. April 2009.

<sup>1</sup> U. of Waterloo, Canada

<sup>2</sup> Chinese National Human Genome Center at Shanghai

**The *Schistosoma japonicum* genome reveals features of host-parasite interplay.** Schistosoma japonicum Genome Sequencing and Functional Analysis Consortium, F. Liu, Y. Zhou, Z.Q. Wang, G. Lu, H. Zheng, P.J. Brindley, D.P. McManus, D. Blair, Q.H. Zhang, Y. Zhong, S. Wang, Z.G. Han, Z. Chen. Nature, 460(7253):345-351. 2009.

**Finančná podpora:** FP7 International Reintegration grant

**Ďalšie projekty** <http://compbio.fmph.uniba.sk/>

**Genómové projekty** (orangutan, panda, marmoset, huby)

Baylor College of Medicine, Washington U. St Luis,

Yan Zhou (Chinese National Human Genome Center at Shanghai)

**Hľadanie génov, algoritmy pre HMM**

**RNA** (štrukturálne motívy, vizualizácia)

Andrej Lupták (UC Irvine), Jozef Nosek (Prírodovedecká fakulta UK)

**Evolúcia génov a genómov** (duplikácie, génová štruktúra, mtDNA)

NHGRI, Adam Siepel (Cornell U.), Web Miller (Penn State),

Jozef Nosek, Ľubo Tomáška (Prírodovedecká fakulta UK)

**Hľadanie homológov**

Ľubo Tomáška (Prírodovedecká fakulta UK)