

Fakulta matematiky, fyziky a informatiky, Univerzita Komenského v Bratislave

Algoritmy pre izometrickú rekonciliáciu génových stromov

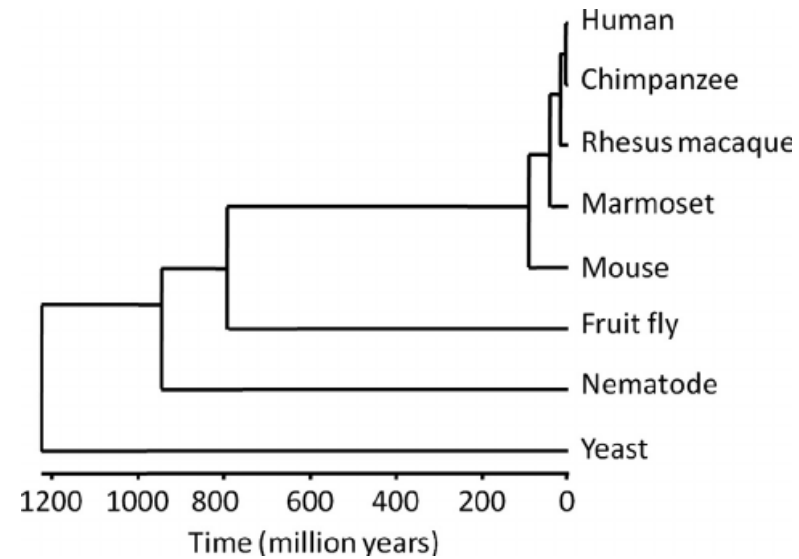
Autor: Radoslav Chládek

Školiteľka: Bronislava Brejová

Evolučné stromy

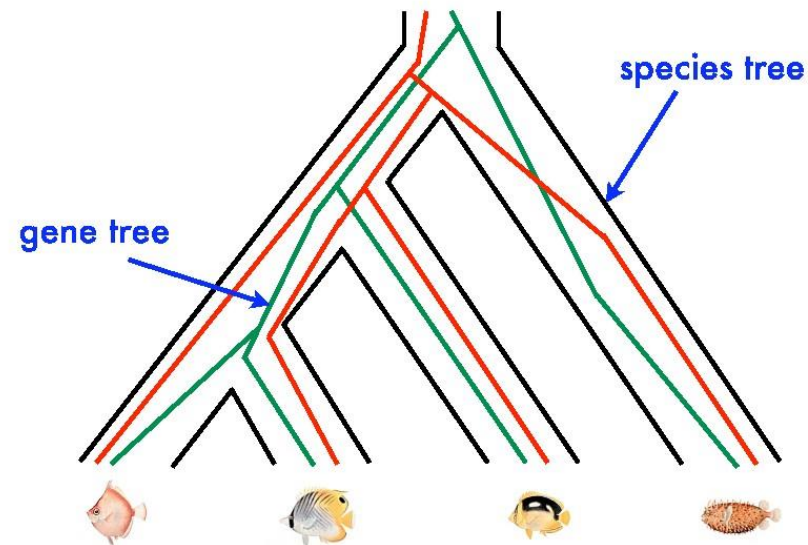
► druhový strom

- zobrazuje evolučné vzťahy medzi biologickými druhmi
- vnútorné vrcholy predstavujú speciácie

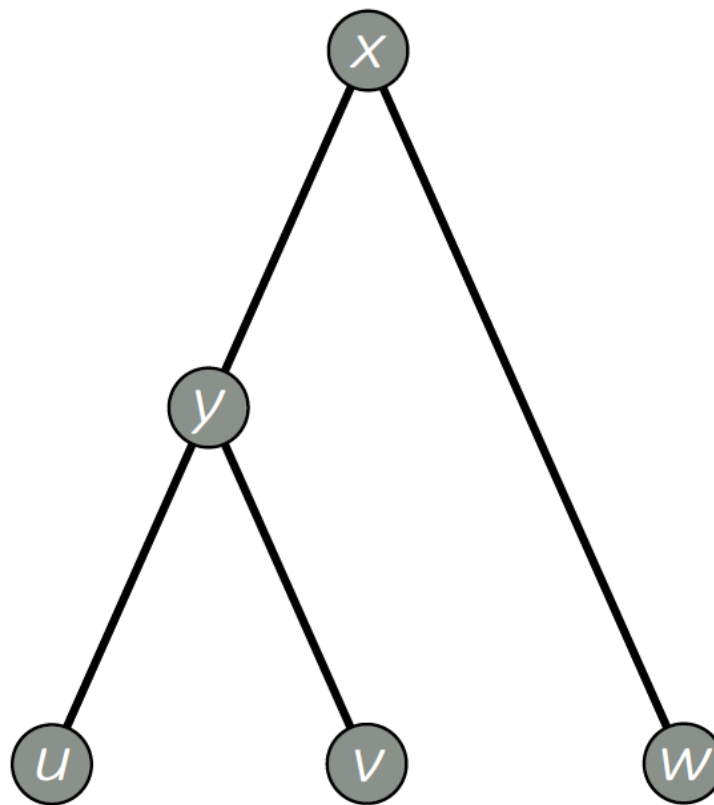
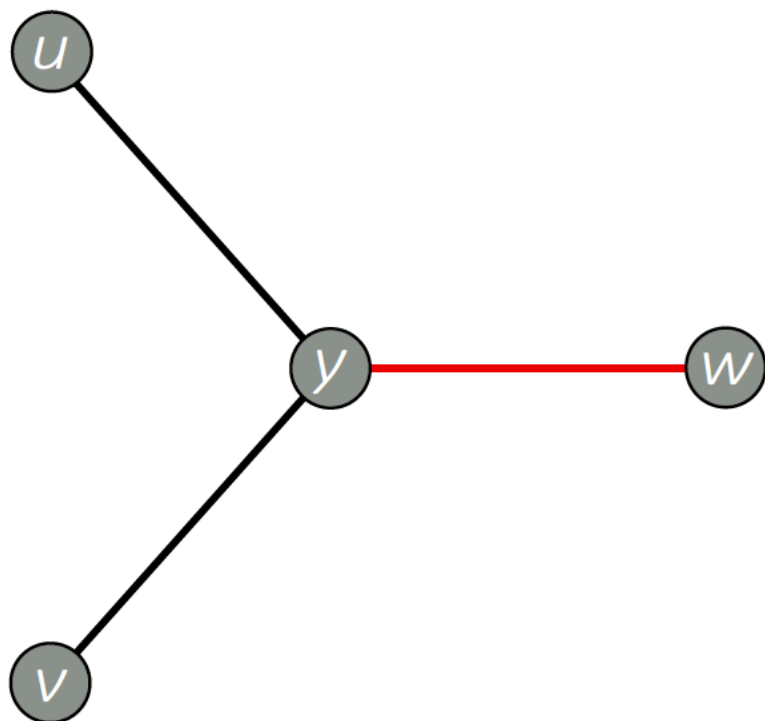


► génový strom

- zobrazuje vývoj jedného génu
- listy predstavujú rodinu génov v dnešných druhoch
- vnútorné vrcholy sú duplikácie alebo speciácie

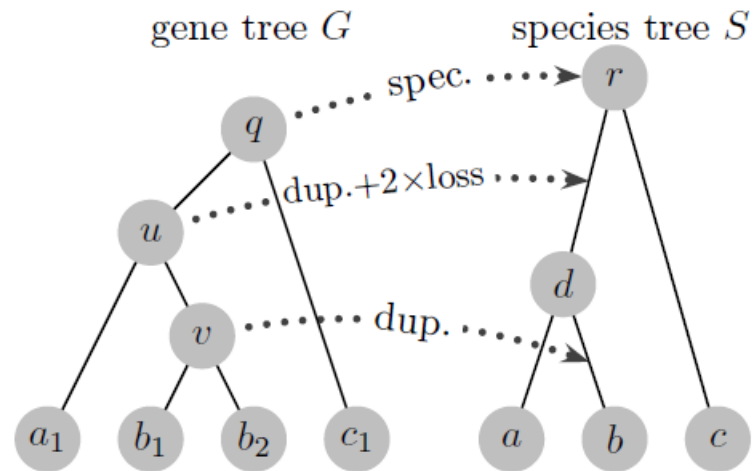
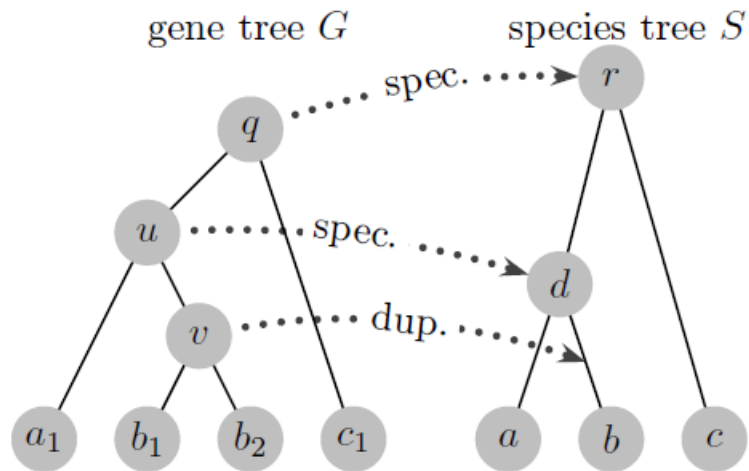


Nezakorenený/zakorenený strom



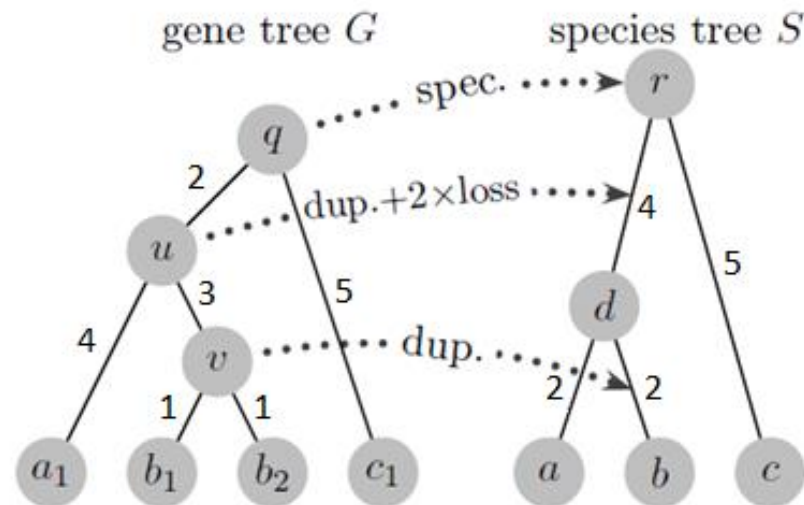
Rekonciliácia génového stromu

- ▶ vstup: druhový strom S , génový strom G , mapovanie listov G do listov S
- ▶ úloha: namapovať vnútorné vrcholy G do vrcholov alebo hrán S



Súčasný stav problematiky

- ▶ viacero algoritmov pre najúspornejšiu (parsimonious) rekonciliáciu (=len topológia)
 - ▶ problém sa študuje od r. 1979
- ▶ izometrická rekonciliácia = okrem topológie sa berú do úvahy aj dĺžky hrán
 - ▶ problém prvý krát zavedený v Ma et al. (2008)
 - ▶ Brejová et al. (2017)
 - ▶ poskytuje vylepšenie algoritmu z Ma et al., algoritmy pre ďalšie varianty problému



Ciel' práce

- ▶ navrhnuť algoritmy pre izometrickú rekonciliáciu
 - ▶ hrany vstupných stromov môžu mať svoje **dĺžky určené intervalom**
 - ▶ zakorenený/nezakorenený génový a druhový strom
 - ▶ nájsť všetky možné riešenia, najúspornejšie riešenie
- ▶ motivácia
 - ▶ v doterajších riešeniach bola dĺžka hrán fixná
 - ▶ skutočné dĺžky hrán nie je možné určiť presne, iba s určitou chybou
 - ▶ => doterajšie algoritmy na skutočných vstupoch často nenájdu riešenie

Výsledky

Vyjadrenie problému vo forme LP

$$X_v - X_u \leq w(u, v)_{\max} \quad \forall (u, v) \in E(G)$$

$$X_v - X_u \geq w(u, v)_{\min} \quad \forall (u, v) \in E(G)$$

$$Y_b - Y_a \leq w(a, b)_{\max} \quad \forall (a, b) \in E(S)$$

$$Y_b - Y_a \geq w(a, b)_{\min} \quad \forall (a, b) \in E(S)$$

$$X_u \leq Y_{\Phi_{\text{lca}}(u)} \quad \forall u \in V(G)$$

$$X_u = Y_{\mu(u)} \quad \forall u \in \text{leaves}(G)$$

$$Y_r = 0$$

Efektívnejšie algoritmy (1)

- ▶ zakorenený S s fixnými dĺžkami hrán, G nepresné dĺžky

- ▶ algoritmus pre zakorenený G s časovou zložitou $O(N)$:

- ▶ prechod zdola nahor

$$x[u]_{\min} = \max_{v \in \text{children}(u)} (x[v]_{\min} - w(u, v)_{\max})$$

$$x[u]_{\max} = \min(D(\Phi_{\text{lca}}(u)), \min_{v \in \text{children}(u)} (x[v]_{\max} - w(u, v)_{\min}))$$

- ▶ prechod zhora nadol

$$X[u]_{\min} = \max(x[u]_{\min}, X[w]_{\min} + w(w, u)_{\min})$$

$$X[u]_{\max} = \min(x[u]_{\max}, X[w]_{\max} + w(w, u)_{\max})$$

- ▶ najúspornejšie riešenie

Efektívnejšie algoritmy (2)

- ▶ nezakorenený G ($O(N^2)$)
 - ▶ pre každú hranu z G vyrieši problém pre polo-zakorenený (semi-rooted) strom
 - ▶ používa predošlý $O(N)$ algoritmus + lineárny program konštantnej veľkosti
 - ▶ nevie nájsť najúspornejšie riešenie
- ▶ počet duplikácií a strát spôsobených izometrickou rekonziliáciou - $O(N \log N)$
- ▶ najúspornejšia rekonziliácia pre nezakorenený G - $O(N^2 h^2 \log N)$
 - ▶ opäť rieši problém pre každý polo-zakorenený strom
 - ▶ rozkladá intervaly mapovania vrcholov na podintervaly s rovnakou úspornosťou
 - ▶ vyberie najlepšiu možnú kombináciu riešení

Zhrnutie

- ▶ algoritmy pre izometrickú rekonciliáciu
 - ▶ dĺžky hrán určené intervalmi
- ▶ všeobecné riešenie - lineárny program
- ▶ najúspornejšie riešenie (strom S je zakorenený s presnými dĺžkami hrán)
 - ▶ $O(N)$ algoritmus pre zakorenený strom G
 - ▶ $O(N^2h^2\log N)$ algoritmus pre nezakorenený strom G

Ďakujem za pozornosť