

Bioinformatické algoritmy

Preusporiadanie genómu

- Andrej Bosík
- Miroslav Jakubík

Obsah prezentácie

- Analýza a porovnanie algoritmov, ktoré sa používajú pri preusporiadavaní genómov

Trocha teórie

- Budeme skúmať číselné reťazce, a zisťovať, koľkými obráteniami podreťazcov vstupného reťazca, sa dostaneme k nejakému výstupnému reťazcu, tento počet voláme reversal distance
- Budeme hlavne hľadať minimálny reversal distance

Trocha teórie

- Skúmané reťazce(permutácie) môžu byť so znamienkom, alebo bez neho
- Pre permutácie bez znamienka je nájsť minimálny reversal distance NP-t'ážké, ale máme aproximačné algoritmy
- Pre permutácie so znamienkom máme polynomiálne optimálne algoritmy

Simple reversal sort

- Hladný, najjednoduchší, nájde riešenie pre neznamienkovú permutáciu, ale nie optimálne
- Funguje tak, že si v každom kroku zväčšuje dobrý prefix, tj. taký, ktorý už netreba meniť

Improved breakpoint reversal sort

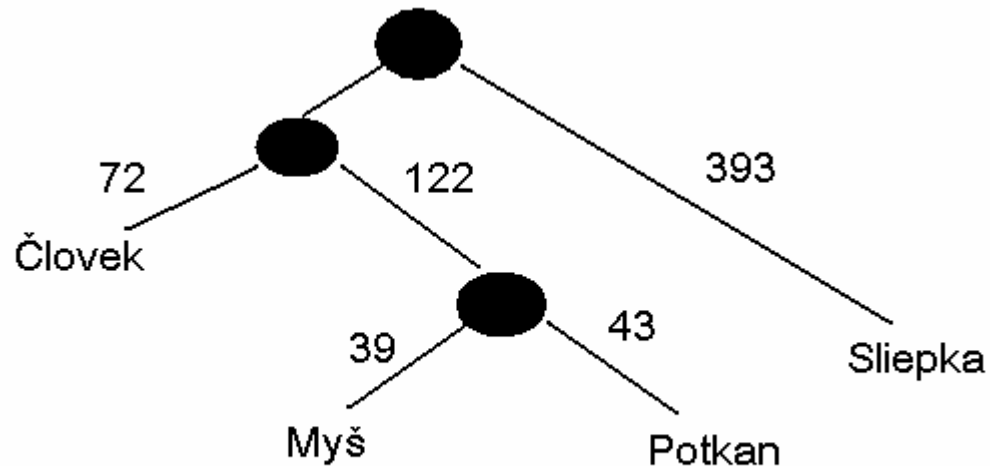
- Tento algoritmus je aproximačný, funguje pre neznamienkové permutácie
- Ďalej ibrs

HP - algoritmus pre znamienkové permutácie

- Hannenhalli – Pevzner
- Je to optimálny algoritmus

Čo sa dá zistiť pomocou týchto algoritmov

- Napr. Fylogenetické stromy



Pokus

- Mali sme data 13 rastlín, jedna rastlina mala genóm dĺžky 105
- Zisťovali sme, akú majú rastliny reversal distance od tabáku (tabák bol identita 1,2,3...)
- Pustili sme na tieto data ibrs a aj HP

Výsledky pokusu

	HP	ibrs
Tobacco		
Platycodon	13	13
Cyananthus	10	10
Codonopsis	10	12
Merciera	16	16
Wahlenbergia	14	14
Triodanus	12	13
Asyneuma	12	12
Legousia	14	15
Symphyandra	11	13
Adenophora	13	13
Campanula	11	12
Trachelium	10	10

Výsledky pokusu

- Vidíme, že ibrs má skoro rovnaké výsledky ako HP
- Skúsime väčšie vstupné data

Pokus

	dížka 160	ibrs	HP
1		67	65
2		36	35
3		72	65
4		145	145
	dížka 320		
1		51	50
2		27	26
3		68	66
4		189	175

Rýchlosť algoritmov

- Čo sa týka ibrs, tak rýchlosť nezávisí od veľkosti vstupnej permutácie, ale rýchlosť závisí hlavne od počtu breakpointov v vstupnej permutácii.
- Už pri počte cca 50 breakpointov sme čakali niekoľko minút

Rýchlosť algoritmov

- Čo sa týka HP-algoritmu, tak ten počíta veľmi rýchlo. Dávali sme mu genómy dĺžky cca 600, s počtom otočení cca 250 a vypočítal to takmer okamžite
- Našli sme, že tento algoritmus má zložitosť $O(n^3)$, ale na reálnych datach počíta oveľa rýchlejšie

Čo by sem patrilo, ale nestihli sme

- Použitie evolučných algoritmov pri výbere obrátenia
- Nájdenie nejakých významných heuristík, nám sa to nepodarilo
- Spojiť alignment s touto témou