

Jadrá zarovnaní

Broňa Brejová

20.10.2022

Opakovanie: Heuristické lokálne zarovnávanie, BLAST

Príklad: $w = 2$ (začínáme z jadier dĺžky 2).

(V praxi sa používa $w = 10$ a viac.)

		C	A	G	T	C	C	T	A	G	A
C	0	0	0	0	0	0	0	0	0	0	0
A	0	1	0	0	0	1	1	0	0	0	0
T	0	0	2	1	0	0	0	0	1	0	0
G	0	0	0	1	2	1	0	1	0	0	0
T	0	0	0	0	2	2	1	1	0	0	0
C	0	1	0	0	0	4	3	0	0	0	0
A	0	0	2	1	0	3	3	2	1	0	1
T	0	0	1	1	2	2	2	4	3	2	1
A	0	0	1	0	1	1	1	3	5	4	3

1. nájdi zhodné úseky
2. rozšír bez medzier
3. spoj medzerami

Senzitivita heuristického algoritmu

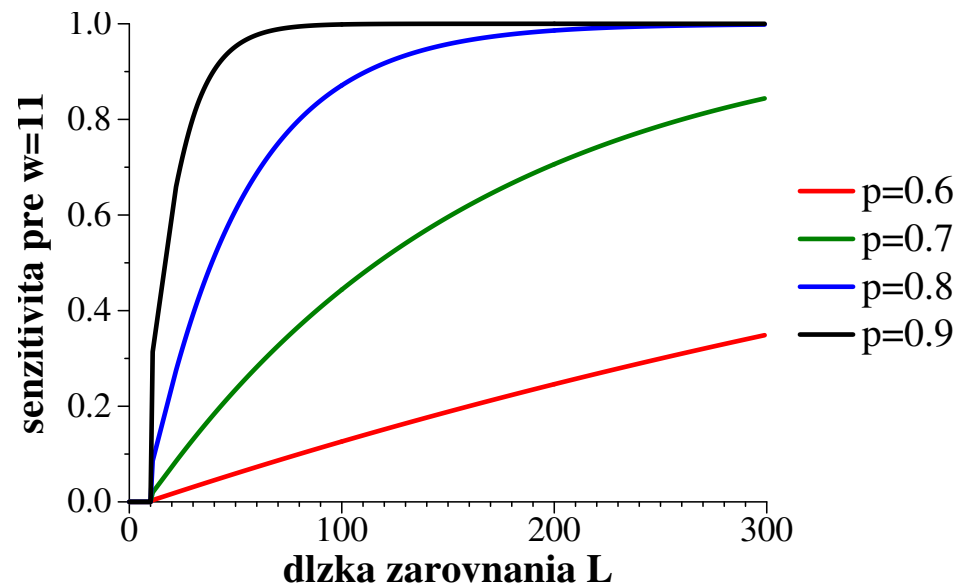
Odhad senzitivity:

Predpokladáme zarovnanie bez medzier, dĺžky L

Každá pozícia je zhoda s pravdepodobnosťou p

Senzitivita:

$$f(L, p) = \Pr(\text{zarovnanie obsahuje } w \text{ zhôd za sebou})$$



Senzitivita heuristického algoritmu

Predpokladáme zarovnanie bez medzier, dĺžky L

Každá pozícia je zhoda s pravdepodobnosťou p

Senzitivita $f(L, p) = \Pr(\text{zarovnanie obsahuje } w \text{ zhôd za sebou})$

Budeme počítať

$A[n] = 1 - f(n, p) = \Pr(\text{zarovnanie neobsahuje } w \text{ zhôd za sebou})$

Opakovanie: ako funguje hľadanie jadier

DB: ulož k -mery do hašovacej tabuľky Query: hľadaj v tabuľke

AGTGGCTGCCAGGCTGG	cGaGGCTGCCTGGtTGG
AGTGG, 0	CGAGG
GTGGC, 1	GAGGC
TGGCT, 2	AGGCT ←
GGCTG, 3	GGCTG ←
GCTGC, 4	GCTGC ←
CTGCC, 5	CTGCC ←
TGCCA, 6	TGCCT
GCCAG, 7	GCCTG
CCAGG, 8	CCTGG
CAGGC, 9	CTGGT
AGGCT, 10	TGGTT
GGCTG, 11	GGTTG
GCTGG, 12	GTTGG

Šetrenie pamät'ou: BLAT

$$w = 5, s = 3$$

AGTGGCTGCCAGGCTGG

AGTGG, 0

GGCTG, 3

TGCCA, 6

CAGGC, 9

GCTGG, 12

cGaGGCTGCctGGtTGG

CGAGG

GAGGC

AGGCT

GGCTG <-

GCTGC

CTGCC

TGCCT

GCCTG

CCTGG

CTGGT

TGGTT

GGTTG

GTTGG

Šetrenie pamäťou: minimizery

$$w = 5, s = 3$$

AGTGGCTGCCAGGCTGG

AGTGG, 0

GTGGC

TGGCT

GGCTG, 3

GCTGC, 4

CTGCC, 5

TGCCA

GCCAG

CCAGG, 8

CAGGC, 9

AGGCT, 10

GGCTG

GCTGG

cGaGGCTGCctGGtTGG

CGAGG

GAGGC

AGGCT*

GGCTG

GCTGC

CTGCC* <--

TGCCT

GCCTG

CCTGG*

CTGGT*

TGGTT

GGTTG*

GTTGG