

Pattern

Search for significant patterns in the set of sequences.

Example of output:

Total sequences: 20

Found 10 pattern(s)

Pattern 1, Length: 9, Power: 20(100%), Q:70.699721, Inf:11.5212 (2.3555)

Q2:70.699721, F0: 2.24981

Consensus: CGCABHBGG

Initial: GCTATCGG

Frequencies:

A	C	G	T	
0	950	50	0	1.7136
0	100	850	50	1.2524
0	950	50	0	1.7136
850	0	50	100	1.2524
200	0	0	800	1.2781
50	0	200	750	1.0082
200	700	50	50	0.7432
150	50	750	50	0.8460
0	50	950	0	1.7136

Sequences:

1:	126	134	+ CGCATTCGG *	6636
2:	186	194	+ CGCTATAGG *	4047
3:	239	247	+ CGCATTCGC *	5341
4:	212	220	+ CGCATGCAG *	5029
5:	251	259	+ CGCATGCGG *	5888
6:	456	464	+ CGCATGGGG *	4804
7:	183	191	+ CGGATTCTG *	4203
8:	103	111	+ CCCGTTTCGG *	4342
9:	492	500	+ CTCATTCCG *	4302
10:	468	476	+ CGCATTCGG *	6636
11:	509	517	+ CGCAATTCGG *	5845
12:	495	503	+ CGCAATTCGG *	5845
13:	219	227	+ GCCATTCGG *	4254
14:	434	442	+ CGCATTTGG *	5551
15:	280	288	+ CGCATGCGG *	5888
16:	430	438	+ CGCTATTCGG *	4759
17:	337	345	+ CGCATTAGG *	5924
18:	99	107	+ CGCATAAGG *	4810
19:	133	141	+ CGCATTCAG *	5777
20:	521	529	+ CGCATTAAG *	5065

Pattern 2, Length: 9, Power: 19(95%), Q:66.807998, Inf:11.7074 (2.3381)

Q2:66.807998, F0: 2.16649

Consensus: CGCATTCGG

Initial: GCATTCAG

Frequencies:

A	C	G	T	
0	947	53	0	1.7025
0	105	842	53	1.2258
0	947	53	0	1.7025
895	0	53	53	1.4093
158	0	0	842	1.3708
53	0	211	737	0.9785
158	737	53	53	0.8077
158	53	737	53	0.8077
0	53	947	0	1.7025

Sequences:

1:	126	134 + CGCATTCGG *	6642
3:	239	247 + CGCATTCGC *	5374
4:	212	220 + CGCATGCAG *	5117
5:	251	259 + CGCATGCGG *	5935
6:	456	464 + CGCATGGGG *	4838
7:	183	191 + CGGATTCTG *	4271
8:	103	111 + CCGTTCGG *	4367
9:	492	500 + CTCATTCCG *	4375
10:	468	476 + CGCATTCGG *	6642
11:	509	517 + CGCAATCGG *	5732
12:	495	503 + CGCAATCGG *	5732
13:	219	227 + GCCATTCGG *	4320
14:	434	442 + CGCATTTGG *	5544
15:	280	288 + CGCATGCGG *	5935
16:	430	438 + CGCTATCGG *	4494
17:	337	345 + CGCATTAGG *	5813
18:	99	107 + CGCATAAGG *	4734
19:	133	141 + CGCATTCAG *	5824
20:	521	529 + CGCATTAAG *	4995

...

Where

Total sequences: 20	- number of sequences that formed a pattern.
Found 10 pattern(s)	- number of patterns.
Pattern 1	- pattern's number.
Length: 9	- length of pattern's sequences.
Power: 20(100%)	- number and percentage of sequences that were included into pattern.
Q:70.699721	- quality of a pattern that reflects both its homogeneity and its power.
Inf:11.5212 (2.3555)	- informational content of a pattern.
Q2:70.699721	- quality of a pattern in the context of its presentation's skew in target and control sets.
F0: 2.24981	- indicates the frequency of occurrence in a target set.
Consensus: CGCABHBGG	- consensus of a pattern for 15-letter alphabet.
Initial: GCTATCGG	- initial consensus, from which the pattern was created.
Frequencies:	- pattern's matrix of frequencies. The right column represents an informational content of each pattern's position:
Sequences:	- weight of all sequences that formed a pattern.
1: 126 - 134	- start and end of sequences that formed a pattern.
+	- strand direction.
CGCATTCGG *	- sequence of a pattern. * means that this sequence was used in pattern formation.
6636	- weight of a pattern in matrix of frequencies.

Parameters:

Input	
Sequence	Input file - nucleotide sequences in FASTA-format
Output	
Result	Name of the output file
Print N best patterns pairs	Print N best patterns pairs
Options	
Search in both chain	Search for pattern in both chain

Threshold for include fragment	Threshold for include fragment to pattern.
Minimal distance for patterns in pair	Minimal distance for patterns in pair
Maximal distance for patterns in pair	Maximal distance for patterns in pair
Number of stored best patterns	Number of stored best patterns
Initial length	Initial length. Minimal value is 3, maximal value is 12.
Try to expand	Try to expand to xx position left and right. If this option is switched off, the pattern will not extend in the parties. Default value is 2, minimal value is 1, maximal value is 10.
Pair selection methods	Pair selection methods: Both pattern must present One of pattern must present