Nhunt: solving the low complexity problem for nucleic acid homology search

Yury Pekov[1], Kirill Romanenkov[2], Alexey Salnikov[2], Sergei Spirin[3]

[1]*Faculty of Bioengineering and Bioinformatics,* [2]*Faculty of Computational Mathematics and Cybernetics, and*
[3]*Belozersky Institute of Physico-Chemical Biology, Moscow State University. E-mail:* `sas@belozersky.msu.ru`

We introduce Nhunt, a program for sensitive nucleic acid homology search. It is designed for searching for distant homologues of non-coding sequences, such as various non-coding RNA, genomic repeats, introns etc.

The well-known "low complexity problem" appears when the frequencies of nucleotides in two sequences under comparison significantly differ from one-fourth. If two sequences with equally shifted composition are aligned using the usual scoring scheme, then the resulting alignment would be significantly overscored. For such regions it is reasonable to use an adjusted substitution scoring matrix. Any substitution matrix can be valid in at most one background frequency context. Our goal is to transform the initial substitution matrix s into a target matrix S that is valid in a given (nonstandard) context.

The same problem exists for protein alignment and protein homology search. In program BLASTP, the matrix adjustment suggested by Yu, Wooton and Altschul [1] is used, which helps avoid the problem in protein homology search. For nucleic acid homology search, no tools of this sort are available at the moment. We have developed a number of algorithms for nucleic acid matrix adjustment and demonstrated some results of their testing on several biological examples. In most cases usage of the adjustment significantly improves quality of search.

The algorithms are implemented in the program Nhunt, which is realized on C. Executable files for Linux and the source code are available in Internet: http://mouse.belozersky.msu.ru/~bennigsen/nhunt.html . A parallel version of Nhunt and a web interface to it are included into the system «Processing biosequences on multiprocessors» (http://angel.cs.msu.su/temp-aligner/). That allows to run Nhunt on supercomputers of Moscow State University.

1. Yu, Wooton, Altschul. The compositional adjustment of amino acid susbstitution matrices. PNAS 2003; 100; 15688–15693.