What is MegaBLAST?

MEGABLAST Search

Mega BLAST uses a greedy algorithm [1] for the nucleotide sequence alignment search. This program is optimized for aligning sequences that differ slightly as a result of sequencing or other similar "errors". When larger word size is used (see explanation below), it is up to 10 times faster than more common sequence similarity programs. Mega BLAST is also able to efficiently handle much longer DNA sequences than the blastn program of traditional BLAST algorithm.

Default parameters

Word size.

Word size is roughly the minimal length of an identical match an alignment must contain if it is to be found by the algorithm. Mega BLAST is most efficient with word sizes 16 and larger, although word size as low as 8 can be used.
If the value W of the word size is divisible by 4, it guarantees that all perfect matches of length W + 3 will be found and extended by Mega BLAST search, however perfect matches of length as low as W might also be found, although the latter is not guaranteed. Any value of W not divisible by 4 is equivalent to the nearest value divisible by 4 (with 4i+2 equivalent to 4i).

Gapping parameters

By default, non-affine gapping parameters are assumed. This means that the gap opening penalty is 0, and gap extension penalty E can be computed from match reward r and mismatch penalty q by the formula: $E = r/2 - q$. The non-affine version of Mega BLAST requires significantly less memory and is also significantly faster, however affine gapping parameters can also be used, preferrably with larger word sizes. Non-affine gapping parameters tend to yield alignments with more gaps, but the gap lengths are shorter.

X-dropoff value

As in BLAST, this value provides a cutoff threshold for the extension algorithm tree exploration. When the score of a given branch drops below the current best score minus the X-dropoff, the exploration of this branch stops. However the actual values of the X-dropoff for Mega BLAST and for traditional nucleotide BLAST algorithms are not necessarily compatible, i.e. with the same word size, match, mismatch and gapping penalties and with the same X-dropoff, the two algorithms might produce different results, which can be remedied by changing the X-dropoff value for one of the algorithms.

[1] Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.