



Align Two Sequences Using NCBI BLAST

Directly comparing two sets of custom sequences using NCBI BLAST search pages

<https://blast.ncbi.nlm.nih.gov/>

National Center for Biotechnology Information • National Institutes of Health • Department of Health and Human Services

Align Two (or more) Sequences Using BLAST

NCBI BLAST service does allow direct comparison of two custom set of sequences directly. This function can be activated by checking the “Align two or more sequences” checkbox, which brings out a new input box for the second set of sequences. This function is fully integrated with the splitd system [1] and assigns unique request ids (RIDs) to individual searches. With the assigned RID, you will be able to access the result for up to 36 hours. With an NCBI login [2], you can save a search strategy for future reference, subject to a limit on input size. With an assigned RID, you can display or download a search result in various formats to highlight different features or for local archive. The dot-matrix graph presentation is also available if the query and the subject each contains only a single input sequence. The new report format [3] provides additional functions, such as tabbed display, and link to Sequence Viewer [4] for interactive examination in the context of sequence annotation. NOTE this function is still based local alignment function provided by BLAST. This is different from what Global Alignment provides, which is based in global alignment algorithm for only a pair of input sequences.



Search Settings

In a BLAST search form, the “Align two or more sequences” checkbox (A) activates this function to display the subject sequence input box (B) while removing the elements related to database selection. The “Align Two Sequences” also adds a new set of parameters for fine tuning searches:

- blastn, megablast or discontinuous megablast algorithms are available for nucleotide searches (C);
- the “Automatically adjust parameter for short input sequences” (D) under the “Algorithm parameters” section is on by default to automatically optimize settings for this type of searches;
- organism-specific repeat filters are available for masking repeat regions in nucleotide searches (E);
- different composition-based statistics can be selected (F) to adjust the significance of the protein alignment.

The screenshot shows the NCBI BLAST search interface. Annotations A-F highlight specific features:

- A:** The checkbox labeled "Align two or more sequences" in the "Program Selection" section.
- B:** The input box for the "Subject Sequence" (accession number, gi, or FASTA sequence).
- C:** The "Optimize for" section, specifically the radio button for "Highly similar sequences (megablast)".
- D:** The checkbox "Automatically adjust parameters for short input sequences" in the "General Parameters" section.
- E:** The dropdown menu for "Species-specific repeats for:" in the "Filters and Masking" section, currently set to "Homo sapiens (Human)".
- F:** The dropdown menu for "Compositional adjustments" in the "Scoring Parameters" section, currently set to "Conditional compositional score matrix adjustment".

References

1. Bealer K, Coulouris G, Dondoshansky I, Madden T, Merezuk Y, Raytselis Y. A Fault-Tolerant Parallel Scheduler for BLAST. <ftp.ncbi.nlm.nih.gov/blast/documents/blast-sc2004.pdf>
2. My NCBI help manual. www.ncbi.nlm.nih.gov/books/NBK3843/
3. The New BLAST Result Page. ftp.ncbi.nlm.nih.gov/pub/factsheets/Howto_NewBLAST.pdf
4. The Graphical Sequence Viewer. ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_Graphical_SV.pdf

