

BioAlign requires 3 sources of information in the first stage.

- Global Sequence Similarity File (Computed by BLASTp)
- 3D Structure Similarity File (Computed by TMAAlign)
- Local Sequence Similarity File (Computed by SWAlign)

BioAlign can use topology and/or more biological information in the second stage.

- Topology: Neighborhood Expansion
- Biology: Remote Homology and Secondary Structure Motifs

NOTE: The similarity file must follow the following format

Network-1_ProteinName Network-2_ProteinName Similarity-Score

Network 1 will be the smaller network in terms of number of nodes.

Default: BioAlign uses Biology and Topology in its default settings.

How to Run:

go to the code folder in the BioAlign archive and run the following command.

`python BioAlign.py speice1 specie2 option`

Option can be 't', 'b', 'bt' ('t' for Topology, 'b' for Biology, 'bt' for Biology + Topology)

Example of mouse yeast alignment

1. `python BioAlign.py mouse yeast t` [This will use Topology after stage-1]
2. `python BioAlign.py mouse yeast b` [This will use Biology after stage-1]
3. `python BioAlign.py mouse yeast bt` [This will use Biology+Topology after stage-1]

Alignments are stored in the alignments folder with the format "specie1-specie2-option.alignment". For example, alignment of mouse and yeast will be stored with the filename "mouse-yeast-t.alignment" if 't' was given as option.

Evaluation

The "evaluation" folder contains the evaluation data and code. This folder contains the go-terms for all species that are used by "evaluation.r" code file.

How to Evaluate the Alignments:

1. `Rscript evaluation/evaluation.r specie1 specie2 option`
 2. `python evaluation/average.py specie1 specie2 option`
- option can be 't', 'b', or 'bt'

Example:

1. `Rscript evaluation/evaluation.r mouse yeast t`
2. `python evaluation/average.py mouse yeast t`

The first command is used to get AFS of the alignment file, while the last command is used to get average results of alignment.

Network Pairs

1. mouse human
2. mouse yeast
3. mouse worm
4. mouse fly
5. yeast human

All the scoring files and biological information of these pairs are available in the corresponding folders.

1. The “3D-structure-similarity” folder contains the 3D structure similarity of all pairs.
2. The “global-sequence-similarity” folder contains the global sequence similarity of all the pairs.
3. The “local-sequence-similarity” folder contains the local sequence similarity of all the pairs.
4. The “sec” folder contains the secondary-structures of all the proteins
5. The “homologs” folder contains the homologs of all the proteins

The “networks” folder contains all the PPI networks collected from the HINT database.

If you want to align these pairs, simply run the above commands.

To Align new network pairs

First compute the global and local sequence similarity, and 3D structure similarity and stored the files in corresponding folders.

Then, get the remote homologs and secondary-structure of all the proteins of networks. Save the results in corresponding folders.

Run the above commands and get the alignment file. To evaluate the alignment file, get the Go-Annotations of both network’s proteins and run the evaluation commands.

Contact

hammad.naveed@nu.edu.pk

umair.ayub@nu.edu.pk

Feel free to contact us in case of any confusions.
