Sequence Searches and Alignments

Andrea Schafferhans

Sequence searches

- Usage of heuristics -> acceleration
  - FASTA
  - BLAST

- Usage of profiles -> sensitivity
  - PSI-Blast
  - HH-Search (PSI-Blast profile -> HMMs)

Search tasks

Searches of the non-redundant sequence database (NR):

- Fasta

- Blast

- PSI-Blast using standard parameters with all combinations of

  - 3 iterations

  - 5 iterations

  - default E-value cutoff (0.005)

  - E-value cutoff 10E-6;

- HHsearch

Evaluate search results

- Compare result lists
  - Overlap?
  - distribution of %identity and scores (E-values)
- Standard of truth ?
  - Structures -> HSSP

Multiple alignments

- Building up alignments iteratively
  - ClustalW (traditional)
  - MUSCLE (performance!)
  - T-Coffee

- Incorporating additional information
  - T-Coffee (many types of alignment, 3D structures)
  - Cobalt (conserved domains, user constraints)

Multiple alignment tasks

- Take 20 sequences from the database search (with pdb!)
  - 99 - 90% sequence identity
  - 89 - 60% sequence identity
  - 59 - 40% sequence identity
  - 39 - 20% sequence identity
- Use different methods (possibly use 'time'):
  - Cobalt
  - ClustalW
  - Muscle
  - T-Coffee with
    - default parameters ("t_coffee your_sequences.fasta)
    - use of 3D-Coffee

Evaluate alignments

- Document the runtimes

- View e.g. in JalView -> export graphics

    – How many conserved columns?

    – Are functionally important residues conserved?

    – How many gaps?

    – Are there gaps in secondary structure elements?