

Datenbanken / Alignments

ASE Übung

26.04.2019

Datenbanken

National Center for Biotechnology Information (NCBI)

- `http://www.ncbi.nlm.nih.gov/`
- Reihe öffentlich zugänglicher Sequenz- und Meta-Datenbanken
- *NCBI Bookshelf*
- *Basic Local Alignment Search Tool (BLAST)*

Protein Data Bank (PDB)

- <https://www.rcsb.org/>,
<https://www.ebi.ac.uk/pdbe/>, ...
- Öffentlich zugängliche Datenbank von Makromolekülstrukturen (Protein / RNA)

Sequenzalignments

Arten

- Global / lokal
- Nukleotid / Protein
- Paarweise / multipel

Alignment

```
AGCAGCAGTCGTAGGAC-GCTG
  ||||| | |||  | | ||
CATAGCAGGC-TAGTGCTGATG
```

Alignment Scores

ACGAT

|||

GCG-T

match_score = 1

mismatch_penalty = -1

gap_penalty = -2

$\Rightarrow -1 + 1 + 1 + -2 + 1 = 0$

Needleman-Wunsch

Finde globales, hinsichtlich des Scores optimales, Alignment zweier Sequenzen.

Fragen

Frage 1

Sequenzdatenbanken speichern neben den reinen Sequenzen i.d.R. weitere Daten.

Geben Sie Beispiele, um welche Information es sich dabei handeln kann und wofür diese benötigt wird.

GenBank

LOCUS NC_000082 5583 bp DNA linear ...
 DEFINITION Mus musculus strain C57BL/6J chromosome 16, GRCm38.p4 ...
 ACCESSION NC_000082 REGION: 90220742..90226324

...

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; ...
 REFERENCE 1 (bases 1 to 5583)
 AUTHORS Church,D.M., Schneider,V.A., Graves,T., Auger,K., ...

...

ORIGIN
 1 gccagggcct cgtttttttg cgcggtcctt tcctgcggcg ctttccgtcc ...

...

Frage 2

Welche Parameter verwendet der /Needleman-Wunsch/ Algorithmus zum Vergleich von Nukleotidsequenzen?

Könnten die Parameter für den Vergleich von Proteinsequenzen sinnvoll erweitert werden?

Frage 3

Es besteht die Möglichkeit die Nukleotid- bzw. Aminosäuresequenzen der proteincodierenden Gene bzw. Proteine mittels Sequenzalignments zu vergleichen.

Wann ist welcher Vergleich sinnvoll?