

Übungsblatt 2

Sequenz Alignment

Deadline: 11.5.2022 um 20:00 MEZ

Bioinformatik für Biochemiestudierende

Dr. Florian Klimm

Sommersemester 2022

Aufgabe 2.1 Bewerten von Alignments

A A K M W V

A S K M V V

$$4/6 = 2/3 = 66.6\%$$

Teilaufgabe 2.1.1 Prozentuale Übereinstimmung

$$4/6 = 2/3 = 66.6\%$$

$$I = M/L * 100 = 2/3 * 100 = 66.6\%$$

wenn, $I > 30\%$ wahrscheinlich zwei homologe Proteine

$I < 20\%$ unwahrscheinlich, dass zwei homologe Proteine

$20\% < I < 30\%$ Graubereich

Teilaufgabe 2.1.2 Bewertung mit Substitutionsmatrix

	A	A	K	M	W	V
A	4	4	-1	-1	-3	0
S	1	1	0	-1	-3	-2
K	-1	-1	5	-1	-3	-2
M	-1	-1	-1	5	-1	1
V	0	0	-2	1	-3	4
V	0	0	-2	1	-3	4

$$4+1+5+5-3+4= 16$$

Aufgabe 2.4 Needleman-Wunsch Algorithmus (30%)

Finden Sie das optimale Alignment der folgenden beiden DNA Sequenzen

G C A T G -

G - A T T A

$$3/5 = 60 \%$$

Wenden Sie (manuell, nicht am Computer) den Needleman-Wunsch Algorithmus an. Als Substitutionsmatrix verwenden Sie.. wobei σ die gap penalty ist.

Vergleichen Sie die Alignments mit $\sigma = -1$ und $\sigma = 0$.

Links $\sigma = -1$

Rechts $\sigma = 0$

	-	G	C	A	T	G	
-	-	0	-1	-2	-3	-4	-5
G	-1	1	0	-1	-2	-3	
A	-2	0	0	1	0	-1	
T	-3	-1	-1	0	2	+1	
T	-4	-2	-2	-1	1	+1	
A	-5	-3	-3	-1	0	0	

	-	G	C	A	T	G
-	-	0	0	0	0	0
G	0	1	1	1	1	1
A	0	1	1	2	2	2
T	0	1	1	2	3	3
T	0	1	1	2	3	3
A	0	1	1	2	3	3

Vergleich $\sigma = -1$ und $\sigma = 0$

$\sigma = -1$:

- überwiegend negative Werte

Zahlenintervall von (-5 bis 2)

$\sigma = 0$

- überwiegend positive Werte

Aufgabe 2.5 Analyse einer unbekannten Sequenz (20%)

Sie haben die folgende Proteinsequenz sequenziert:

MVHLGPKKPQARKGSMADVPKELMDEIHQLEDMFTVDSETLRKVVKHFID
ELNKGLTKKGGNIPMIPGWVMEFPTGKESGNYLAIDLGGTNLRVVLVKLS
GNHTFDTTQSKYKLPHDMRTTKHQEELWSFIADSLKDFMVEQELLNTKDT
LPLGFTFSYPASQNKINEGILQRWTKGFDIPNVEGHDVVPLLQNEISKRE
LPIEIVALINDTVGTLIASYYTDPETKMGVIFGTGVNGAFYDVVSDIEKL
EGKLADDIPSNSPMAINCEYGSFDNEHLVLPRTKYDVAVDEQSPRPGQQA
FEKMTSGYYLGELLRLVLLELNEKGLMLKDQDLSKLKQPYIMDTSYPARI
EDDPFENLEDTDIFQKDFGVKTTLPERKLIRRLCELIGTRAARLAVCGI
DAICQKRGYKTGHIAADGSVYNKYPGFKEAAAKGLRDIYGWTGDASKDPI
TIVPAEDGSGAGAAVIAALSEKRIAEGKSLGIIGA

Nun wollen Sie diese bioinformatisch untersuchen

Genutzt wurde folgende Seite: <https://www.ebi.ac.uk/Tools/sss/ncbiblast/>

Teilaufgabe 2.5.1 BLAST

- Welche BLAST Variante nutzen Sie? Warum?

blastp Variante

- Zu welchem Organismus gehört diese Sequenz wahrscheinlich?

Saccharomyces cerevisiae (Backhefe)

- Zu welchem Protein gehört diese Sequenz wahrscheinlich?

Hexokinase-1

- Wie groß ist die Percentage Identity mit diesem Protein?

99,8%

- Wie lautet der E-Score und wie interpretieren Sie ihn?

Wir haben einen Score von 0.0 da, E-Value < 0.01 wahrscheinlich homologe Sequenzen haben

NCBI BLAST+

Protein

Nucleotide

Vectors

Web services

Help & Documentation

Bioinformatics Tools FAQ

Feedback

Tools > Sequence Similarity Searching > NCBI BLAST

Results for job ncbiblast-I20220510-204050-0080-35149919-p1m

Summary Table

Tool Output

Visual Output

Functional Predictions

Result Summary

Submission Details

Selection:

Select All Invert Clear

Apply to selection:

Annotations:

Show Hide

Alignments:

Show Hide

Entries:

Download in

fasta

format

Tools:

Align.	DB:ID	Source	Length	Score (Bits)	Identities %	Positives %	E()
<input checked="" type="checkbox"/> 1	SP:HXKA_YEAST	Hexokinase-1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HXK1 PE=1 SV=2 <i>Cross-references and related information in:</i> Gene expression Reactions & pathways Bioactive molecules Nucleotide sequences Genomes & metagenomes Enzymes Literature Samples & ontologies Molecular interactions Protein families Macromolecular structures Protein expression data Protein sequences	485	973.8	99.8	99.8	0.0
<input checked="" type="checkbox"/> 2	SP:HXKB_YEAST	Hexokinase-2 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HXK2 PE=1 SV=4 <i>Cross-references and related information in:</i> Gene expression Reactions & pathways Bioactive molecules Nucleotide sequences Genomes & metagenomes Enzymes Literature Samples & ontologies Molecular interactions	486	781.9	77.2	88.9	0.0

Teilaufgabe 2.5.2 Mutation identifizieren

Hexokinase-1

```

390          400          410
..IGTRAARLAVCGIDAICQKRGYKTGH
.: : : : : : : : : : : : : : : :
..IGTRAARLAVCGIAAICQKRGYKTGH
390          400          410
  
```

Aminosäure an der Position 401 wurde D mit A ausgetauscht (Hier findet die Punktmutation statt)

Mit der Code-Sonne kommt folgendes raus:

G **A** -> D

G **C** -> A

Eine Punktmutation C zu A an der Stelle 1202 (401 * 3)