Eukaryonten (mit Zellkern)

	Ribosom	grosse Untereinheit			kleine Untereinheit
Sedimentations-koeffizient	80S	60S			40S
Molmasse	4.200.000	2.820.000			1.400.000
# Proteine		49			33
# RNA		3			1
# Nucleotide		28S 4.718	5,8S 160	5S 120	18S 1.874

Prokaryonten (ohne Zellkern)

	Ribosom	grosse Untereinheit		kleine Untereinheit
Sedimentations-koeffizient	70S	50S		30S
Molmasse	2.500.000	1.590.000		930.000
# Proteine		34		21
# RNA		2		1
# Nucleotide		23S 2.904	5S 120	16S 1.541

rRNA

- Sie kommen in der Zelle besonders häufig vor.
- Die Teile (z.B. 16s) lassen sich leicht isolieren.
- Die Unterschiede zwischen verschiedenen Arten sind gering aber vorhanden

⇒ An ihnen lässt sich der Verwandschaftsgrad gut nachweisen.

FASTA-Format:

>gi|9931493|gb|AF282306.1|AF282306 Yersinia pestis 16S ribosomal RNA gene, partial sequence

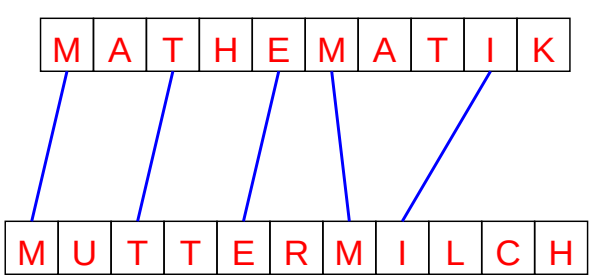
GCCTAACACATGCAAGTCGAGCGGCAGCGGGAAGTAGTTTACTACTTTGCCGGCGAGCGGCGGACGGGTG
AGTAATGTCTGGGGATCTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATGACCT
CGCAAGAGCAAAGTGGGGGACCTTAGGGCCTCACGCCATCGGATGAACCCAGATGGGATTAGCTAGTAGG
TGGGGTAATGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAACTGA
GACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGC
CATGCCGCGTGTGTGAAGAAGGCCTTCGGGTTGTAAAGCACTTTCAGCGAGGAGGAAGGGGTTGAGTTTA
ATACGCTCAATCATTGACGTTACTCGCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATA
CGGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGT
GAAATCCCCGCGCTTAACGTGGGAACTGCATTTGAAACTGGCAAGCTAGAGTCTTGTAGAGGGGGGTAGA
ATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACA
AAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCTGTAA
ACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTG
GGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGT
TTAATTCGATGCAACGCGAAGAACCTTACCTACTCTTGACATCCACAGAATTTGGCAGAGATGCTAAAGT
GCCTTCGGGAACTGTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTTGTGAAATGTTGGGTTAAGT
CCCGCAACGAGCGCAACCCTTATCCTTTGTTGCCAGCACGTAATGGTGGGAACTCAAGGGAGACTGCCGG
TGACAAACCGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGC
TACAATGGCAGATACAAAGTGAAGCGAACTCGCGAGAGCCAGCGGACCACATAAAGTCTGTCGTAGTCCG
GATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTAGATCAGAATGCTACGGTGA
ATACGTTCCCGGGCCTTGTACACA

Sequenz-Vergleich

Die Ähnlichkeit zweier Sequenzen definieren wir als die

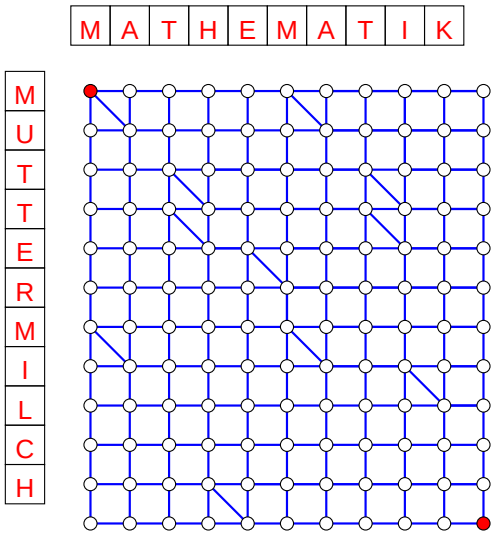
Länge einer maximalen gemeinsamen Teilsequenz.

Beispiel:



Länge 5

Modellierung als Kürzeste-Wege-Problem



Ein kürzester Weg entspricht einer maximalen Teilsequenz.

Konkrete Aufgaben

- Vergleich zwischen zwei rRNA-Codes realisieren

- Einlesen der Daten im FASTA-Format
  (im Verzeichnis ~willhalm/lehre/inf4_W00/rRNA)

Konkrete Aufgaben (Fortsetzung)

- ein Programm schreiben, das folgendes leistet:

  Gegeben:

  - eine Datei von Gen-Daten im FASTA-Format und
  - einen einzelnen Datensatz

  Ausgabe:
  Gen-Daten (Kopfzeilen) sortiert nach Ähnlichkeit

- Zwei Versionen:

  - interaktiv
  - automatisiert (zum Testen)