12.01.2001

Projekt "Genefinding"

=====

Zu Bearbeiten ist eine Virus-DNA von ca. 180 Kilobasen Länge.

Die DNA liegt im FASTA-Format vor. Dieses Dateiformat sieht ungefähr folgendermaßen aus:

===

> identifizier

AATT..... (ca. 80 Zeichen/Zeile o.ä.)

:

:

:

===

Es sind folgende Aufgaben zu erledigen:

1. Datei einlesen

2. Reverse-Complement (in der Datei steht nur ungefähr die Hälfte der DNA, d.h. die andere Seite muß ergänzt werden)

3. Gene suchen

3.1. Längenrestriktion: mindestens 51 Basen

4. Regelmäßigkeiten in den 50 Basen vor dem Gen finden

Allgemeiner Aufbau eines Gens:ATG__| Vielfaches von 3 |____Stop

Gene können grundsätzlich verschachtelt sein, sind es hier aber nicht.

Die Ausgabedatei sollte pro Gen folgendes enthalten:

- Absolute Start- und Endposition

- das Gen selbst ;-)

- die 50 Basen vor dem Gen