

Aufgabenblatt 3

Genetischer Code; Naïves Stringmatching

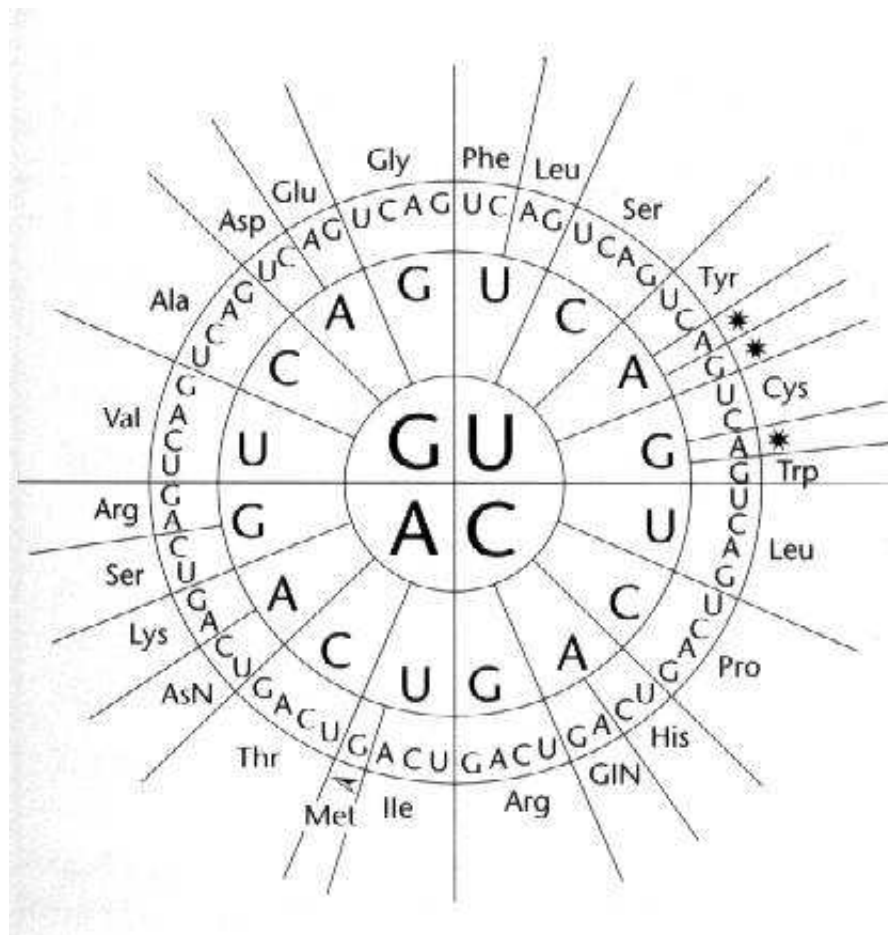
Abgabe: **16.11.2004** bis **17 Uhr** über Goya
Maximal: 20 Punkte
Namen nicht vergessen!

1. Wieviele Möglichkeiten gibt es, eine Nukleotidsequenz in korrespondierende Proteinstränge zu übersetzen? Übersetzt die folgende Nukleotidsequenz für alle möglichen Open Reading Frames (ORFs):

tctcacagcaaattaggcaggagttgctca

1+6 P = 7 P

Zur Übersetzung könnt Ihr die abgebildete Codesonne nach Bresch und Hausmann verwenden. Gelesen wird immer von innen nach außen: das Codon "GCA" entspricht also der Aminosäure *Ala(nin)*.



2. Wie würde sich die Proteinsequenz ändern, wenn die DNA von einer Hefezelle (lat. *Saccharomyces cerevisiae*) stammen würde? **3 P**

3. Implementiert den naïven Algorithmus zum Stringmatching in Java.

Der Aufruf der Java-Klasse für den naïven Algorithmus soll folgendermaßen aussehen:

```
java GruppeXAufgabe3 -p Datei1 -t Datei2
```

Zwei Beispieldateien für das Pattern bzw. Template findet Ihr auf der Webseite zur Übung, <http://www.informatik.hu-berlin.de/wbi/teaching/ws0405/> .

Erwartet wird eine Ausgabe im Format

```
Gefunden an Position 4  
Gefunden an Position 7  
Gefunden an Position 16  
Anzahl der Vergleiche: 57
```

Die Anzahl an Vergleichen ist die Anzahl der benötigten Zeichenvergleiche, um alle Vorkommen vom Pattern P in dem Template T zu finden.

Die Ausgabe bitte auch auf Eurem Lösungsblatt notieren!

6 P

4. Sucht in der Datenbank, von welchem Organismus die Sequenz von Aufgabe 1 stammt und aus welchem Gen sie kommt.

Für die Datenbanksuche kann BLAST verwendet werden, siehe

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides> .

4 P