

# 1 Bioinformatics

- Hat mit Biologie zu tun
- Lösen biologischer Fragestellungen
- Verwendung biologische Prinzipien
- E V O L U T I O N
- Spezielle Art von Daten (Rauschen)
- Verschiedene Methoden aus Computertheorie, Statistik

# 2 Sequenzanalyse

## 2.1 Sequenz

- Sequenz: Fixe Abfolge von Dingen
- In der Biologie/Bioinformatik synonym zu String verwendet
- dh. Subsequenz in Bioinformatik ist aufeinanderfolgend!
- Bioinformatik: Buchstaben (aus alphabet  $\alpha$ )
- $\alpha$  aus biologischen (chemischen) Bausteinen
- Verschieden Alphabete: AS, Nukleobasen
- Biologie: Sequenzen können miteinander verwandt sein
- Unterschiede in Sequenzen (z.B: SNP) Informationen zB über Probleme
- Gemeinsamkeiten helfen, Verwandtschaft zu klären

## 2.2 Sequenzunterschiede

- naiv: Hamming Distanz: Zählen der Unterschiede
- Viel besser: Levinshtein Distanz:
  - Zahl der Veränderungen, die notwendig sind um eine Sequenz in die andere zu überführen
  - In Biologie gibt es: Insertion, Deletion, Mutation
    - \* Eine Sequenz (willkürlich als Basis, dann sind:
    - \* Insertion: In der anderen Sequenz fehlen Buchstaben
    - \* Deletion: in der anderen Sequenz gibt es mehr Buchstaben
    - \* Mutation: In der anderen Sequenz steht ein anderer Buchstabe
  - Es gibt sehr viele mögliche Verteilungen der Veränderungen
  - zB. 2x 100 Buchstaben lange Sequenzen:  $\approx 10^{60}$ , wenn man Insertion und Deletion nicht gleichzeitig vorkommen lässt

### 3 Dynamic Programming

Lösung von Problemen durch Zusammensetzung von Lösungen von einfach zu lösenden Teilproblemen. Die kleinsten Teilprobleme haben triviale Lösungen. Bellman's Prinzip: Geht immer dann wenn sich ein Problem so in Teilprobleme zerlegen lässt, dass die optimale Lösung des Gesamtproblems sich durch Kombination der optimalen Lösungen der Teilprobleme ergibt.

Meistens 3 Stufen:

- Initialisierung (triviale Probleme)
- Forward recursion
- Backtrack

### 4 Alignment

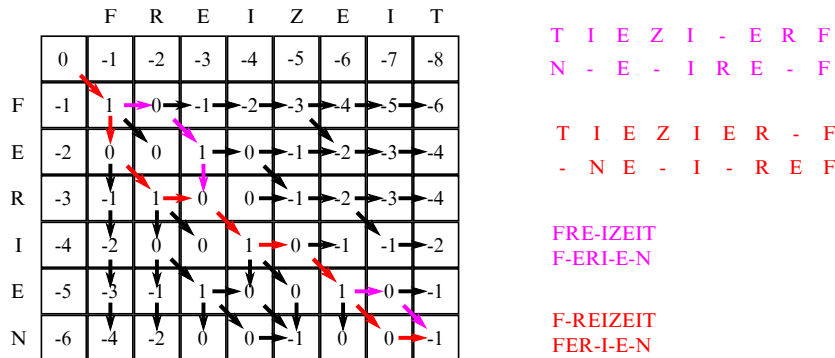
#### 4.1 Needleman Wunsch Algorithmus

Globales Alignment

Man braucht:

Gap-scores (Insertion und Deletion)

Sowie (Mis)match scores



Backtrack: Von links unten (Optimum). Dann immer dem Weg folgen, der zum aktuellen Ergebnis geführt hat. Meist sind mehrere Wege möglich

#### 4.2 Semiglobales alignment

Zur Suche einer kurzen in einer langen Sequenz.

Gap kosten in der ersten und letzten Reihe werden auf 0 gesetzt.