

1 Multiple Sequence Alignment

1.1 Scoring

Score einer Spalte ist Summe aller Paarweisen scores = Sum of pairs score (SPS)

- Paar enthaelt ein gap: paarweiser score = gapscore
- Paar enthaelt 2 gaps: paarweiser score=0

1.2 DP solution

$$A(i, j, \dots, n) = \begin{cases} A(i-1, j-1, \dots, n-1) + S(i, j, \dots, n) \\ A(i, j-1, \dots, n-1) + S(-, j, \dots, n) \\ A(i-1, j, \dots, n-1) + S(i, -, \dots, n) \\ \vdots \\ A(i-1, j-1, \dots, n) + S(i, j, \dots, -) \\ \vdots \\ A(i, j, \dots, (n-1), n-1) + S(-, -, \dots, -, n) \end{cases} \quad (1)$$

Das ist exponentiell in der Anzahl von sequenzen (m): $\mathcal{O}(n^m 2^m)$

1.2.1 Carillo-Lipman

Einschraenkung der Loesungsmenge durch Beschaenkung:

- Untere Schranke U aus anderer Berechnungsmethode (s.u.)
- Summe der Scores der paarweisen Alignmens P gibt obere Schranke
- Matrizen mit bestmoeglicher Loesung fuer PWA, dass entsprechendes Feld enthaelt

Dann koennen fuer ein Paar (a,b) nur Felder in der Loesung sein, die:

$$U \leq A_{a,b}(i, j) + \sum_{x,y} P(x, y) - P(a, b) \quad (2)$$

Alle anderen Kombinationen von (i,j) sind fuer Alignment a,b unmoeglich. So wird ein Grossteil der Moeglichkeiten ausgeschlossen. Bis zu 5 Sequenzen von ca 500 AS Laenge.

2 Heuristische Methoden

Heuristic findet schnell gute Loesung, garantiert aber nicht, beste Loesung zu finden

2.1 Sternforemiges Alignment

- Finde Sequenz, die durchschnittlich am naechsten zu allen anderen Sequenzen ist
- Nimm diese als Geruest
- Aligne alle anderen Sequenzen dazu (PWA)

2.2 Progressive Alignment

- Alignen entlang eines Baumes
- Distanzen zwischen Sequenzen durch PWA ermittelt
- Mit diesen Distanzen wird Guide tree erstellt
- Unterschiede in Art, wie Sequenzen oder Alignments dazukommen
- Problem: Einmal falsch, immer falsch

2.2.1 Feng-Doolittle

- Sequenzen werden normal zu Sequenzen alignt
- Sequenzen zu jeder Sequenz eines alignments alignt, bestes als guide
- bei 2 Alignments Bestes inter-Alignment PWA als guide
- Gaps werden durch neutralen character (X) ersetzt.
- NT: nur paarweise alignments.

2.2.2 Profile alignment

- Alignment zweier Alignments
- NW alignment, bei dem Sequenzen durch 2 Subalignments ersetzt werden
- Gaps werden immer in alle Spalten des Subalignments
- Scoring ist SPS.
- Nur die "Cross" terme optimieren, da sich die anderen nicht aendern.
- Anstatt gap scores auch SPS, allerdings mit nur gap Subspalten

2.3 Iterative Alignment

Versucht Problem das progressive Alignment: Einmal Fehler, immer Fehler zu loesen

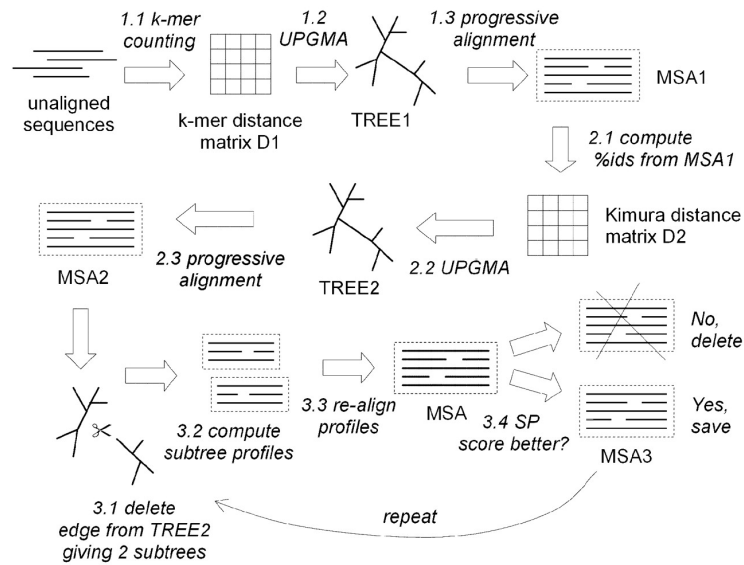
- Ausgehend von MSA
- Sequenz(en) werden aus alignment entfernt
- Entstandene nur gap Spalten werden aus Alignment entfernt
- Sequenz(en) werden wieder dazualigt
- Iteratives wiederholen, bis Abbruch (Laufplaenge oder Qualitaet)

2.4 ClustalW

Adressiert verschieden Probleme progressiver Alignments

- Gewichtet Sequenzen (Branchlaengen durch Occupancies)
- Contextabhaengiges gap scoring
 - GO niedriger, wenn schon gaps existieren
 - GO hoeher, wenn gaps nahe (+- 8AS)
 - GO niedriger bei Sequenzstuecken mit mindestens 5 hydrophilen AS (D,E,G,K,N, Q, P, R, S)
 - GO sonst AS abhaengig
 - GO hoeher fuer aehnlichere Sequenzen (niedriger fuer divergente)
 - GO steigt mit $\log(\text{Laenge})$ der kuerzeren Sequenzlaenge
 - GE hoeher bei stark unterschiedlich langen Sequenzen (weniger seehr lange gaps)
- Scoring Matrix abhaengig von Sequenzaehnlichkeit im Alignment
- Iterative Schritte bei jedem zwischenalignment moeglic

2.5 MUSCLE



2.6 T-Coffee

- Alignmentscores aus Library von Alignments
 - PWAs
 - lokale PWAs
 - Alignments ueber zwischensequenzen
- Dann wird mit $g_o=0$ und $g_e=0$ aligniert
- Teuer wegen Erstellung der Libraries