

SCC 172
ICC Para Biomoleculares
Alinhamento de sequências

Rosane Minghim

2012

Sequências de DNA, RNA e Proteínas

- Nucleotídeos:
 - ‘A’ , ‘C’ , ‘T’ , ‘G’
- Outros códigos (ex.)
 - Ambiguidades: ‘R’ (G A), ‘Y’ (T C), ‘K’ (G T),...
 - Uracil para RNA (correspondente a T do DNA) : ‘U’
 - Gap: ‘-’
- Ex.
 - AAATCGGCCTACGGGTTCCC
 - AACT-CG-TCTAC-GG-TCAA

Vários algoritmos existem

- Focaremos em dois:
 - Um baseado em score:

```
AAATCGGCCTACGGGT TCCC
|| | | | |
AACT-CG-TCTAC-GG-TCA
```

+1 match

-1 mismatch

-2 gap

Tarefas:

1. $\text{Score}(\text{seq1}, \text{seq2})$

Se $\text{seq1}[i] == '-'$ ou $\text{seq2}[i] == '-'$ $\text{sco} = \text{sco} - 2$

senão Se $\text{seq1}[i] == \text{seq2}[i]$ $\text{sco} = \text{sco} + 1$

senão $\text{sco} = \text{sco} - 1$

2. Localizar alinhamentos

vetor de alinhamento $[1, 1, -1, -1, -2, \dots]$ fornece alinhamentos internos parciais (sub-cadeias)

Vários algoritmos existem

- Focaremos em dois:
 - Outro baseado em matriz de pontos

	-	G	A	A	T	T	C	A	G	T	C	A
-	0	0	0	0	0	0	0	0	0	0	0	0
G	0	1							1			
G	0								1			
A	0		1	1				1				1
T	0				1	1				1		
C	0						1				1	
G	0	1							1			
A	0		1	1				1				1

Global

	A	C	C	T	G	A	G	C	T	C	A	C	C	T	G	A	G	T	T	A
A	█					█				█					█					█
C		█	█					█		█		█								
C																				
T				█					█						█			█	█	
G					█		█								█		█			
A	█					█				█						█				█
G					█		█								█		█			
C		█						█		█		█								
T				█					█						█			█	█	
C		█	█					█		█		█								
A	█					█				█						█				█
C		█	█					█		█		█								
C																				
T				█					█						█			█	█	
G					█		█								█		█			
A	█					█				█					█		█			█
G					█		█								█		█			
T				█					█						█			█	█	
T				█					█						█			█	█	
A	█					█				█					█					█

Várias tarefas

1. Preencher a matriz

seq_1 string[0..n-1]

seq_2 cadeia [0..n-1]

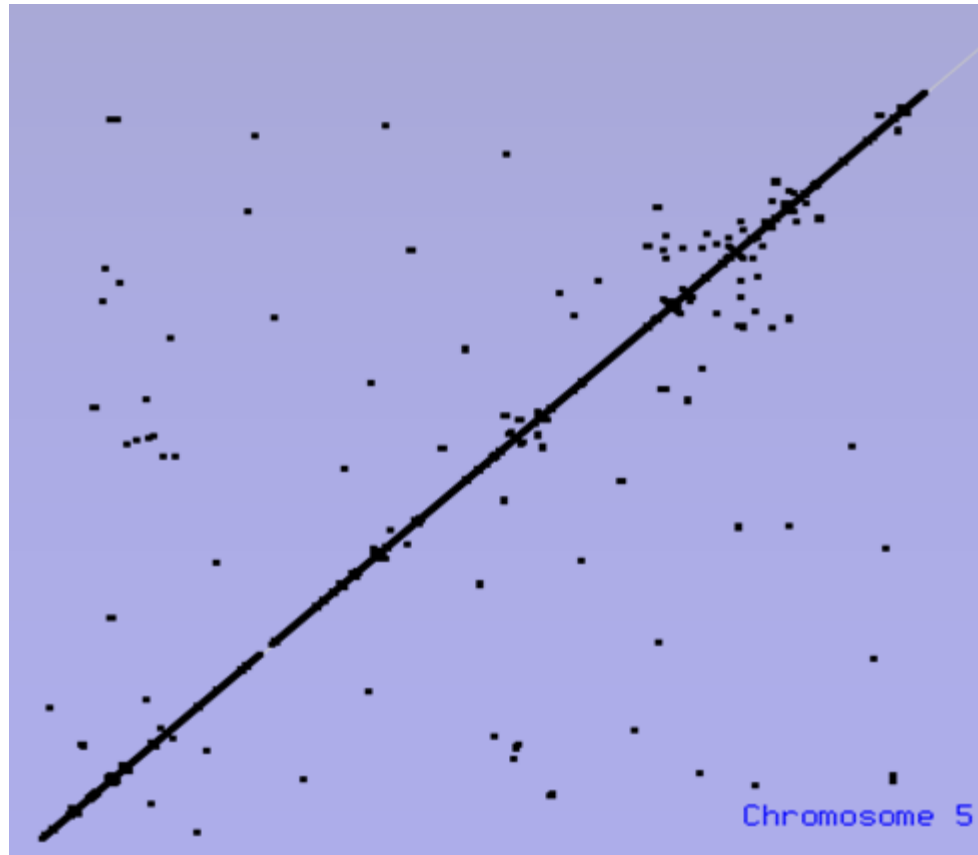
Se $seq1[i] == seq2[j]$ então $mat[i][j] = 1$

2. Encontrar diagonais para trechos de sequências

Se $mat[i][j] == 1$ olhar $mat[i-1][j-1]$ e $mat[i+1][j+1]$ até encontrar 0.

3. Diagonal principal dá o score da sequência toda.

Global dot plot



Fontes:

- www.ime.usp.br/~alexsc/align_semin.ppt
- [*bsb2007.inf.puc-rio.br/course_slides/03.pdf*](http://bsb2007.inf.puc-rio.br/course_slides/03.pdf)