

A Simple Protocol for the Inference of RNA Global Pairwise Alignments

Felix Karg

15. Januar 2019

University of Freiburg



Content

Recap

Tree-Based

Sankoff

LocARNA

Content

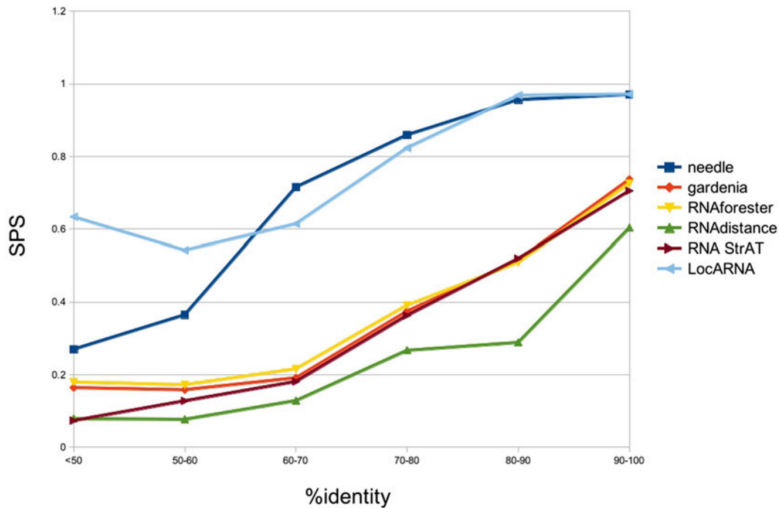
Recap

Tree-Based

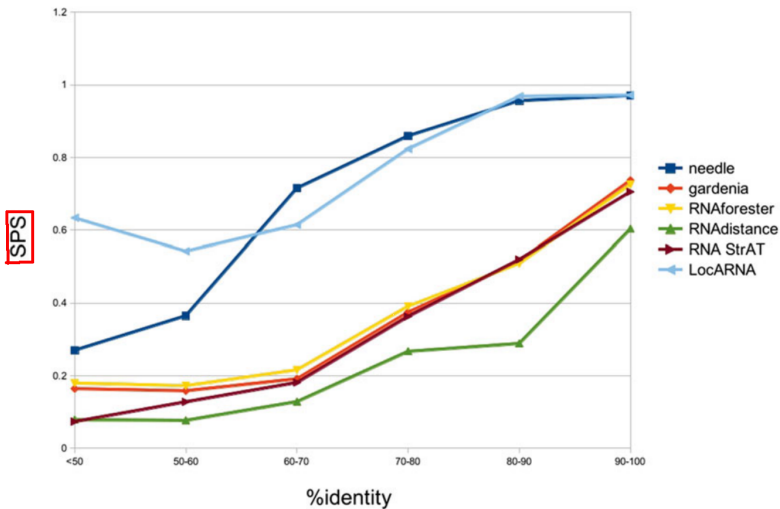
Sankoff

LocARNA

Predicted Secondary Structures



Predicted Secondary Structures



Sum of Pairs Score

Sum of Pairs Score

Used to measure the alignment of two RNA sequences

Sum of Pairs Score

Used to measure the similarity of two RNA sequences

Sequence Similarity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Similarity:

1 - (edit distance / unaligned length of shorter sequence)

Sequence Similarity - Example

A: AAGGC~~TT~~

B: AAGGC

C: AAGGCAT

Similarity:

1 - (edit distance / unaligned length of shorter sequence)

Sequence Similarity - Example

A: AAGGC~~TT~~

B: AAGGC

C: AAGGCAT

Similarity: $60\% = 1 - (2 / 5)$

$1 - (\text{edit distance} / \text{unaligned length of shorter sequence})$

Sequence Similarity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Similarity:

1 - (edit distance / unaligned length of shorter sequence)

Sequence Similarity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Similarity: $60\% = 1 - (2 / 5)$

$1 - (\text{edit distance} / \text{unaligned length of shorter sequence})$

Sequence Similarity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Similarity:

1 - (edit distance / unaligned length of shorter sequence)

Sequence Similarity - Example

A: AAGGCTT

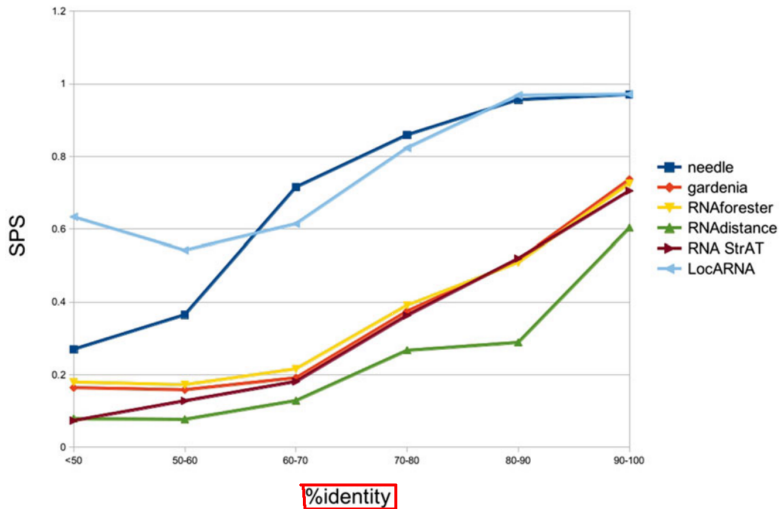
B: AAGGC

C: AAGGCAT

Similarity: $86\% = 1 - (1 / 7)$

$1 - (\text{edit distance} / \text{unaligned length of shorter sequence})$

Predicted Secondary Structures



Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity:

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity:

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity: 100%

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity:

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity: 100%

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

B: AAGGC

C: AAGGCAT

Identity:

Identical nucleotides / shorter sequence length

Sequence Identity - Example

A: AAGGCTT

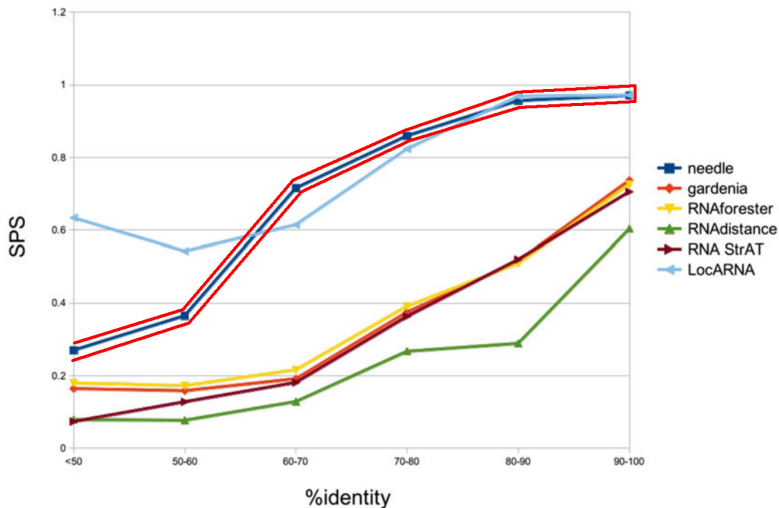
B: AAGGC

C: AAGGCAT

Identity: $85\% = 6 / 7$

Identical nucleotides / shorter sequence length

Predicted Secondary Structures



Needleman-Wunsch-Algorithm

Needleman-Wunsch

match = 1

mismatch = -1

gap = -1

		G	C	A	T	G	C	U	
		0	-1	-2	-3	-4	-5	-6	-7
G	-1	1	0	-1	-2	-3	-4	-5	
A	-2	0	0	1	0	-1	-2	-3	
T	-3	-1	-1	0	2	1	0	-1	
T	-4	-2	-2	-1	1	1	0	-1	
A	-5	-3	-3	-1	0	0	0	-1	
C	-6	-4	-2	-2	-1	-1	1	0	
A	-7	-5	-3	-1	-2	-2	0	0	

Content

Recap

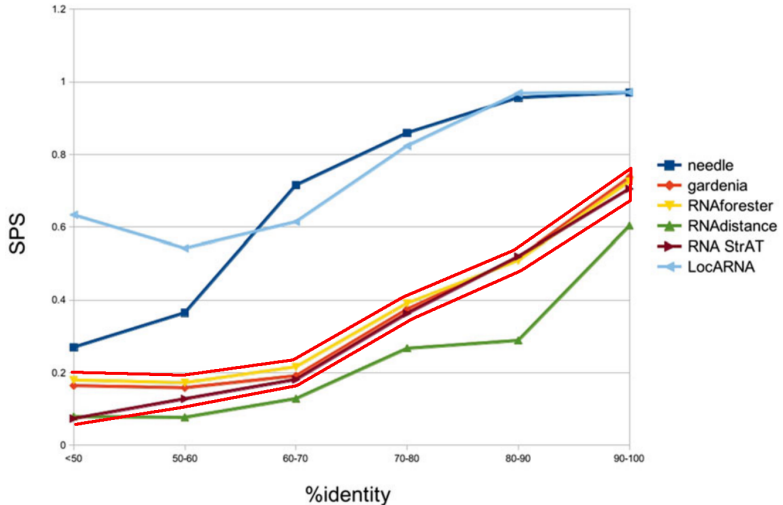
Tree-Based

Sankoff

LocARNA

Tree-based

Predicted Secondary Structures



Content

Recap

Tree-Based

Sankoff

LocARNA

Sankoff-Algorithm

- Dynamic Programming

Sankoff-Algorithm

- Dynamic Programming
- Runtime $O(n^6)$

Sankoff-Algorithm

- Dynamic Programming
- Runtime $O(n^6)$
- Space needed: $O(n^4)$

Sankoff-Algorithm

- Dynamic Programming
- Runtime $O(n^6)$
- Space needed: $O(n^4)$

What does it do ... ?

Global free energy minimisation.

Sankoff-Algorithm

Sankoff-Algorithm

- Base match

Sankoff-Algorithm

- Base match
- Base insertion

Sankoff-Algorithm

- Base match
- Base insertion
- Base deletion

Sankoff-Algorithm

- Base match
- Base insertion
- Base deletion
- Base **pair** match

Content

Recap

Tree-Based

Sankoff

LocARNA

intro