MUSCLE Algorithm for Multiple Sequence Alignment
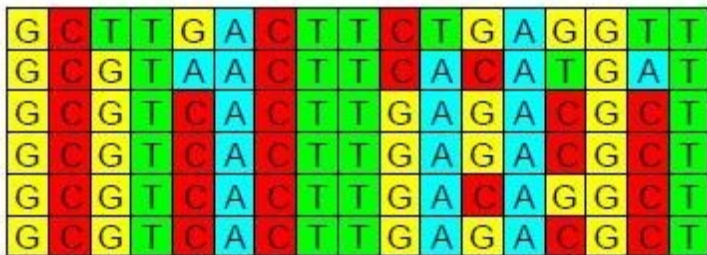
Nahian Salsabil - 1705091
Muhtasim Noor - 1705108

CSE, BUET

July 6, 2021

Multiple Sequence Alignment

Figure: Multiple Sequence Alignment

Primary Idea

- All possible solutions

Primary Idea

- All possible solutions
- Computationally Expensive

Improved Algorithm

- CLUSTAL
- MAFFT
- MUSCLE

MUSCLE Algorithm

Multiple Sequence Comparison by Log-Expectation

Source code and executable files are freely available at
http://www.drive5.com/muscle

MUSCLE Algorithm

Multiple Sequence Comparison by Log-Expectation

- Better average accuracy

Source code and executable files are freely available at
http://www.drive5.com/muscle

MUSCLE Algorithm

Multiple Sequence Comparison by Log-Expectation

- Better average accuracy

- Better speed

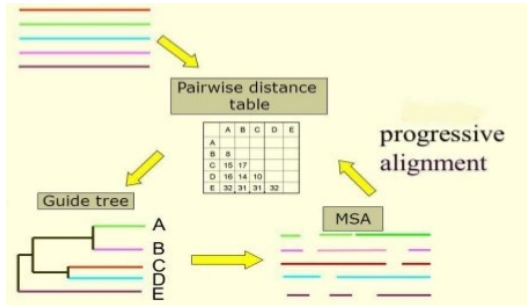Source code and executable files are freely available at
http://www.drive5.com/muscle

Progressive Alignment

Figure: Progressive Alignment

Progressive Alignment

Similarity Measure

Unaligned sequences

Progressive Alignment

Distance Matrix

Seq	1	2	3	4
1	–	–	–	–
2	1	–	–	–
3	7	8	–	–
4	11	5	2	–

Unaligned sequences

Building distance matrix

Progressive Alignment

Tree Construction

Seq	1	2	3	4
1	–	–	–	–
2	1	–	–	–
3	7	8	–	–
4	11	5	2	–

Unaligned sequences   Building distance matrix   Guide tree construction

Progressive Alignment

Aligning Sequences

Unaligned sequences

Building distance matrix

Guide tree construction

Pairwise alignment

Progressive Alignment

Aligning Sequences

Unaligned sequences

Building distance matrix
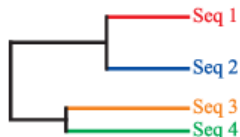
Guide tree construction

Pairwise alignment

Aligned sequences

Stages of MUSCLE

Stages of MUSCLE

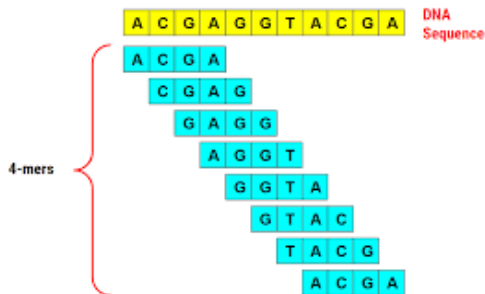- Draft Progressive

Stages of MUSCLE

- Draft Progressive
- Improved Progressive

Draft Progressive

Stage 1: Draft Progressive

Similarity Measure

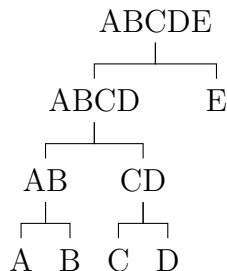## κ-mer Counting

k-mer	Count
ACGA	2
CGAG	1
GAGG	1
AGGT	1
GGTA	1
GTAC	1
TAGC	1

Stage 1: Draft Progressive

Tree Construction

UPGMA

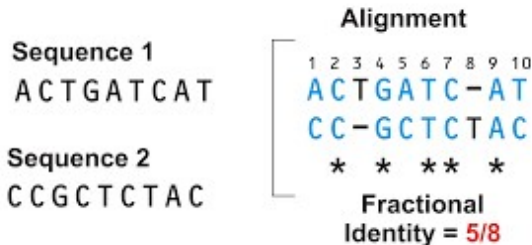|	A	B	C	D	E
A	-				
B	11	-			
C	20	30	-		
D	27	36	9	-	
E	30	33	20	27	-

Stage 1: Draft Progressive

Alignment
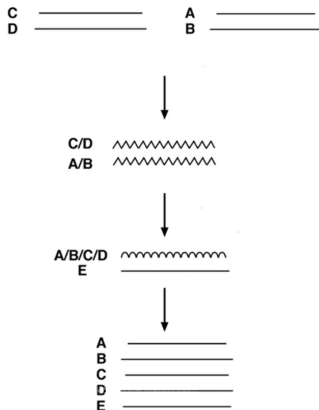
Improved Progressive

Stage 2: Improved Progressive

Similarity Measure

Fractional Identity **(D)**

Stage 2: Improved Progressive

Tree Construction

Kimura Distance Matrix

	Seq1	Seq2	Seq3	Seq4
Seq1	0	0.1	0.3	0.8
Seq2		0	0.2	0.7
Seq3			0	0.7
Seq4				0

Kimura →

	Seq1	Seq2	Seq3	Seq4
Seq1	0	0.1	0.38	2.6
Seq2		0	0.23	1.6
Seq3			0	1.6
Seq4				0

$$d_{Kimura} = -log_e(1 - D - D^2/5)$$
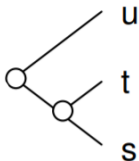
Stage 2: Improved Progressive
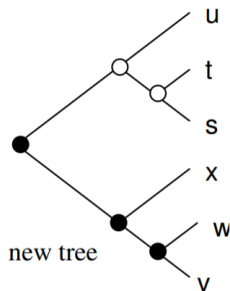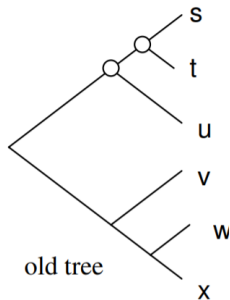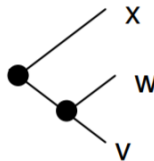
Tree Comparison

Stage 2: Improved Progressive

Tree Comparison

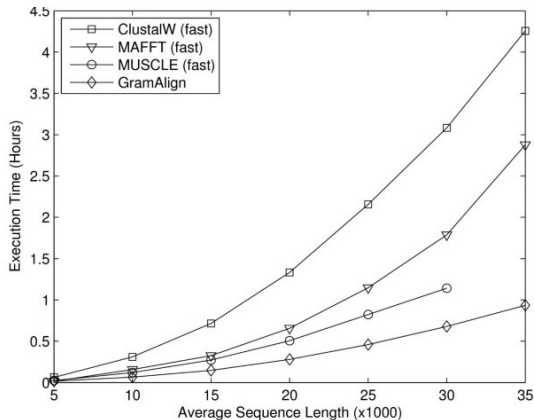
Unchanged Branch   Changed Branch

MUSCLE Algorithm

Terminates

MUSCLE Algorithm

Terminates

- When root is unchanged

MUSCLE Performance

Thank You!