

# Pasta Modifications



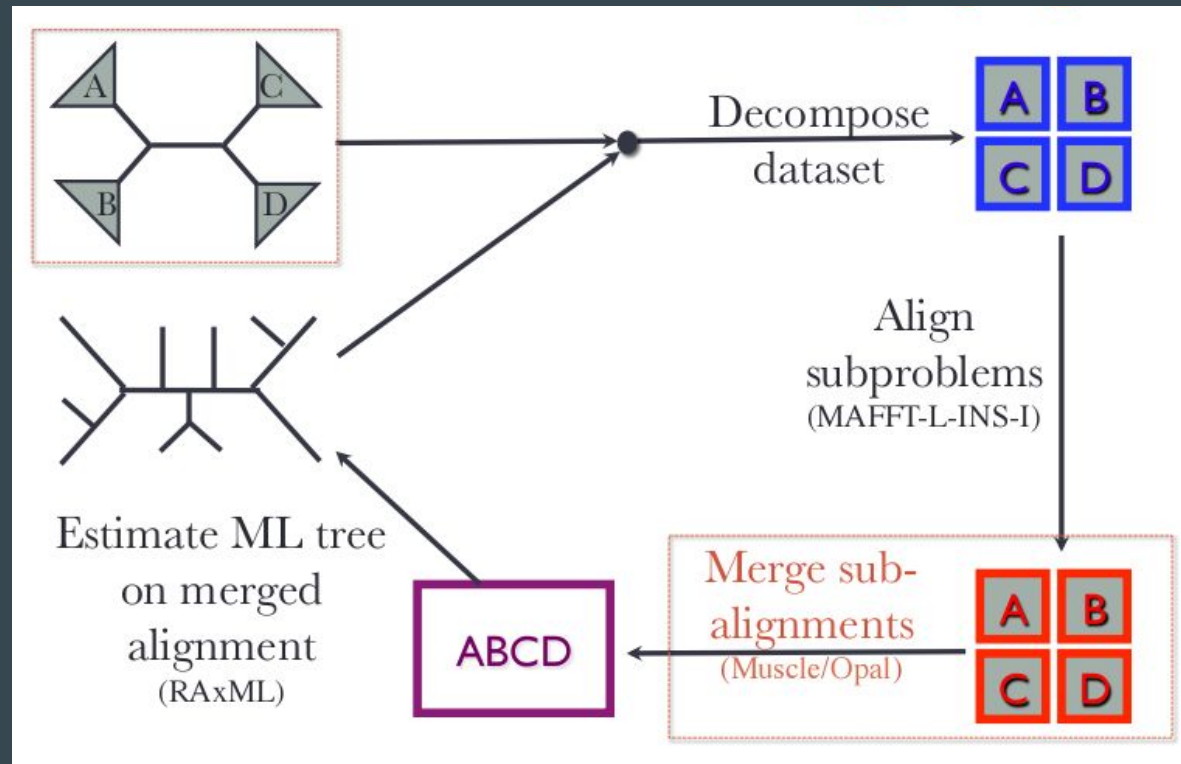
Kodi Collins  
CS 466

# Motivation: Multiple Sequence Alignment

- Evolution
  - Alleles in populations
- 3 categories of mutations
  - Advantageous
  - Deleterious
  - Neutral
- Types of Selection
  - Positive
  - Negative
  - Balancing
  - Diversifying
  - Stabilizing
- Detection of Selection
  - MSA on Coding Regions
  - Proportion synonymous and non-synonymous substitutions
  - Synonymous = Neutral
  - Differing rates means some selection
- Over-Alignment
  - Substitution favored over indels
  - Substitutions not neutral
  - False Positive Detection of Selection
- Under-Alignment
  - Assumed opposite effect but we don't know

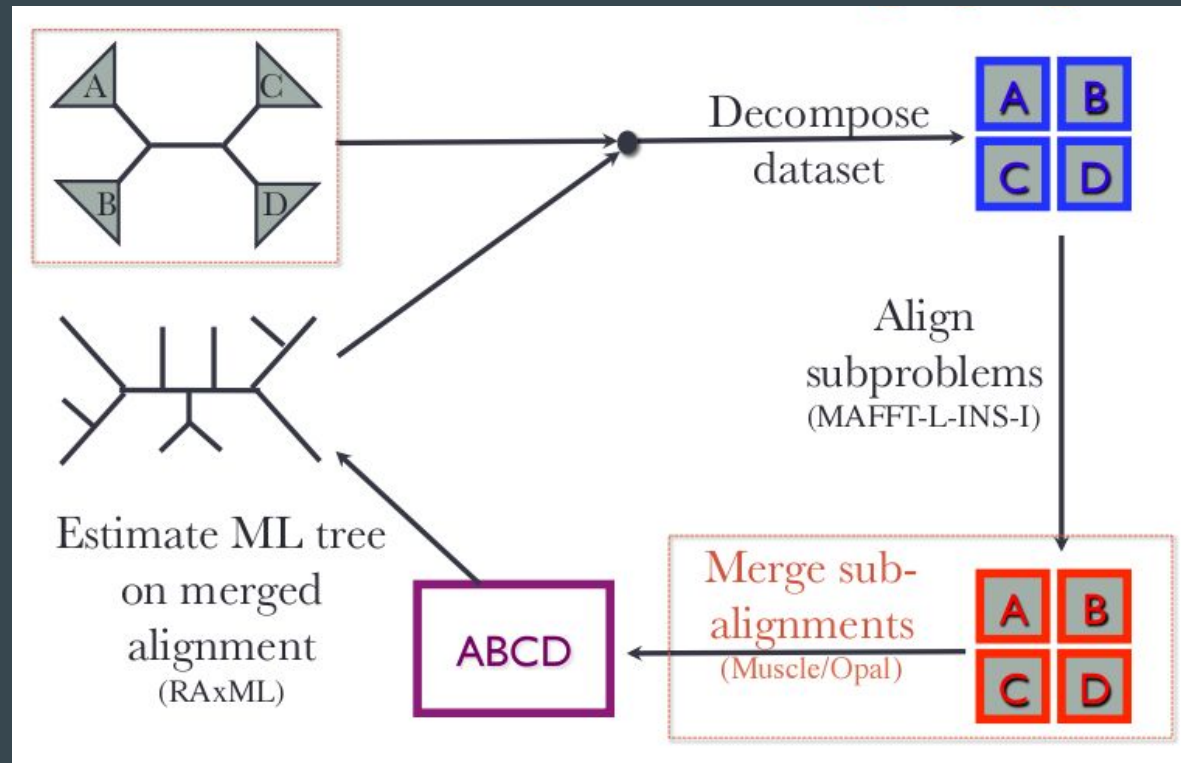
# How Pasta Works

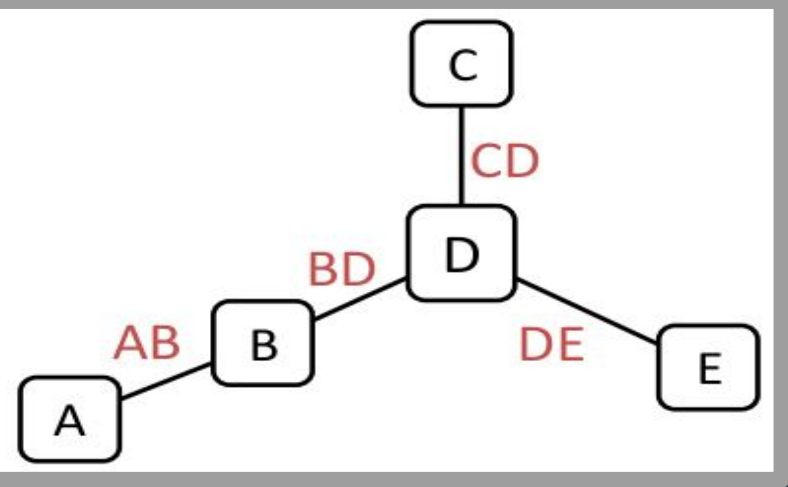
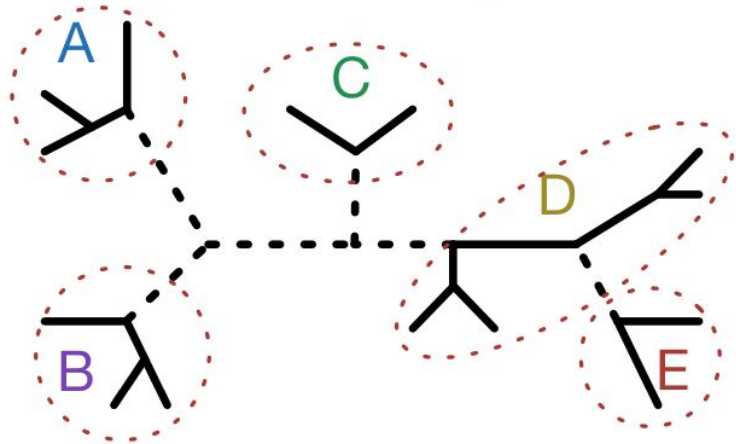
1. Build Guide Tree
2. Decompose
3. Align
4. Merge
5. Transitivity
6. Repeat 1-5



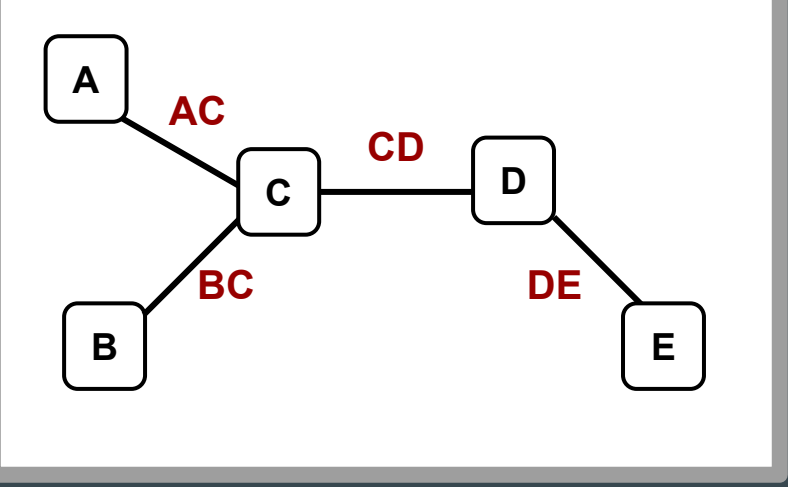
# How Pasta Works

1. Build Guide Tree
2. Decompose
3. Align
4. Merge
5. Transitivity
6. Repeat 1-5





<b>A</b>	<b>B</b>	= 3	<b>B</b>	<b>D</b>	= 4.5
<b>A</b>	<b>C</b>	= 2	<b>B</b>	<b>E</b>	= 3.5
<b>A</b>	<b>D</b>	= 5	<b>C</b>	<b>D</b>	= 2.5
<b>A</b>	<b>E</b>	= 4	<b>C</b>	<b>E</b>	= 4
<b>B</b>	<b>C</b>	= 1.5	<b>D</b>	<b>E</b>	= 1



# Ways to Score

## Percentage of gaps:

- List of number of gaps in each sequence
- Divide by length of the Opal Alignment
- Comparison by median and largest value

## Other Potential Considerations:

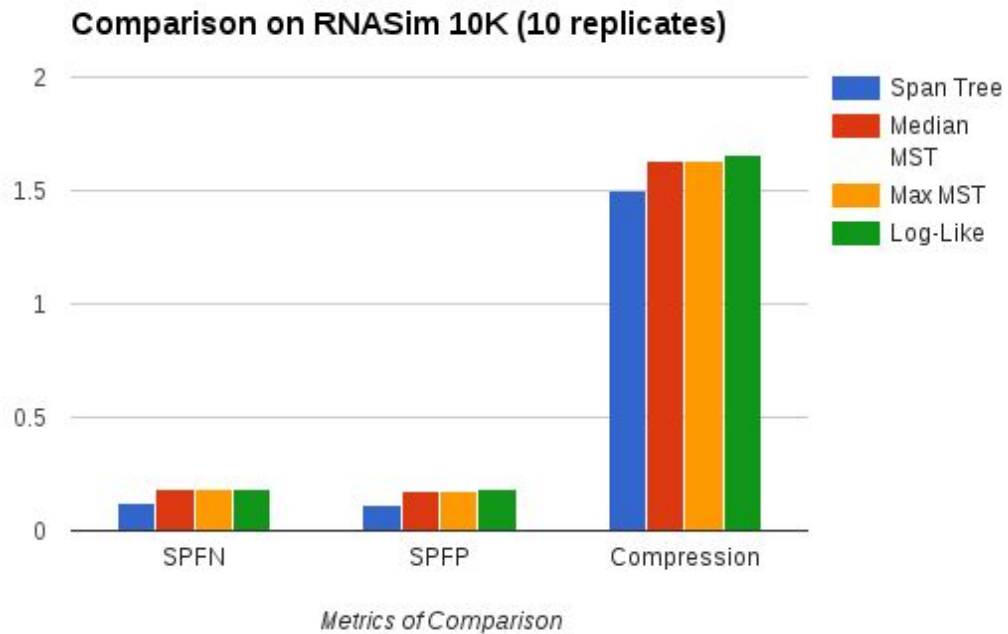
- Sum-of-Pair Score
- Distance-based: FastME
- Profile HMMs

## Maximum Likelihood:

- Build a ML tree on each Opal Alignment
- Compare Log-Likelihood Value
- Maximum Spanning Tree

# Results:

- Mixed Results
  - Default Pasta Best
  - no/little improvement
- Next Steps
  - Local alignments where transitivity 'fails'
  - Use Muscle not Opal
  - ...



# Sources:

Mirarab, S., N. Nguyen, and T. Warnow, 2014. "PASTA: ultra-large multiple sequence alignment." Proceedings RECOMB 2014. An extended version of this paper appears in the Journal of Computational Biology.

Warnow, Tandy. *Computational Phylogenetics: An Introduction to Designing Methods for Phylogeny Estimation*. N.p.: Cambridge U Press, 2017. Print.

Mirarab, S. Presentation on Pasta at RECOMB 2014: <http://www.cs.utexas.edu/~phylo/software/pasta/pasta.pdf>