

Measuring the distance between multiple sequence alignments

Benjamin P. Blackburne
Simon Whelan

Bioinformatics, 15 February 2012
Volume 28, Issue 4, Pages 495-502
doi.org/10.1093/bioinformatics/btr701

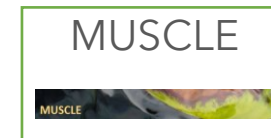
Melanie Mößer and Dominik Stroh
Aktuelle Themen der Sequenzanalyse
5th July 2021



Motivation

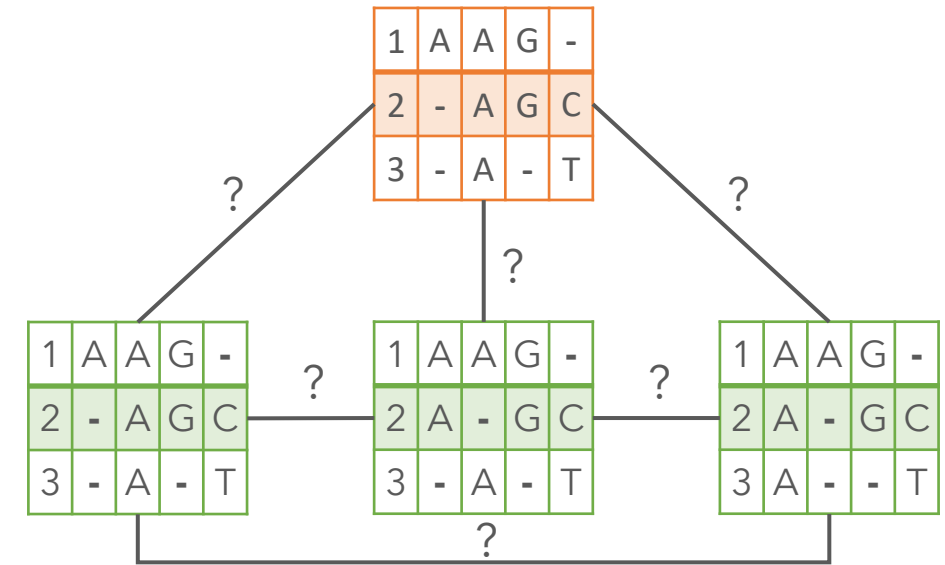
How similar are MSAs produced by different aligners to each other and to a reference?

Assessment of seven aligners on two databases:



BALiBASE
Real-world data

INDELible
Synthetic data



Introduction



Existing scores are not true metrics!

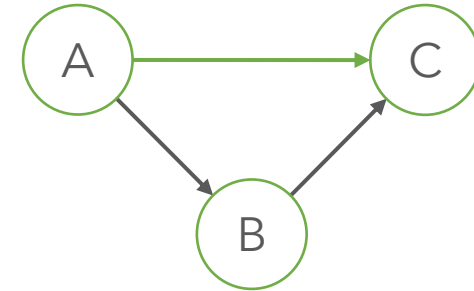
Identity of indiscernibles
 $d(A,B) = 0 \leftrightarrow A=B$



Symmetry
 $d(A,B) = d(B,A)$



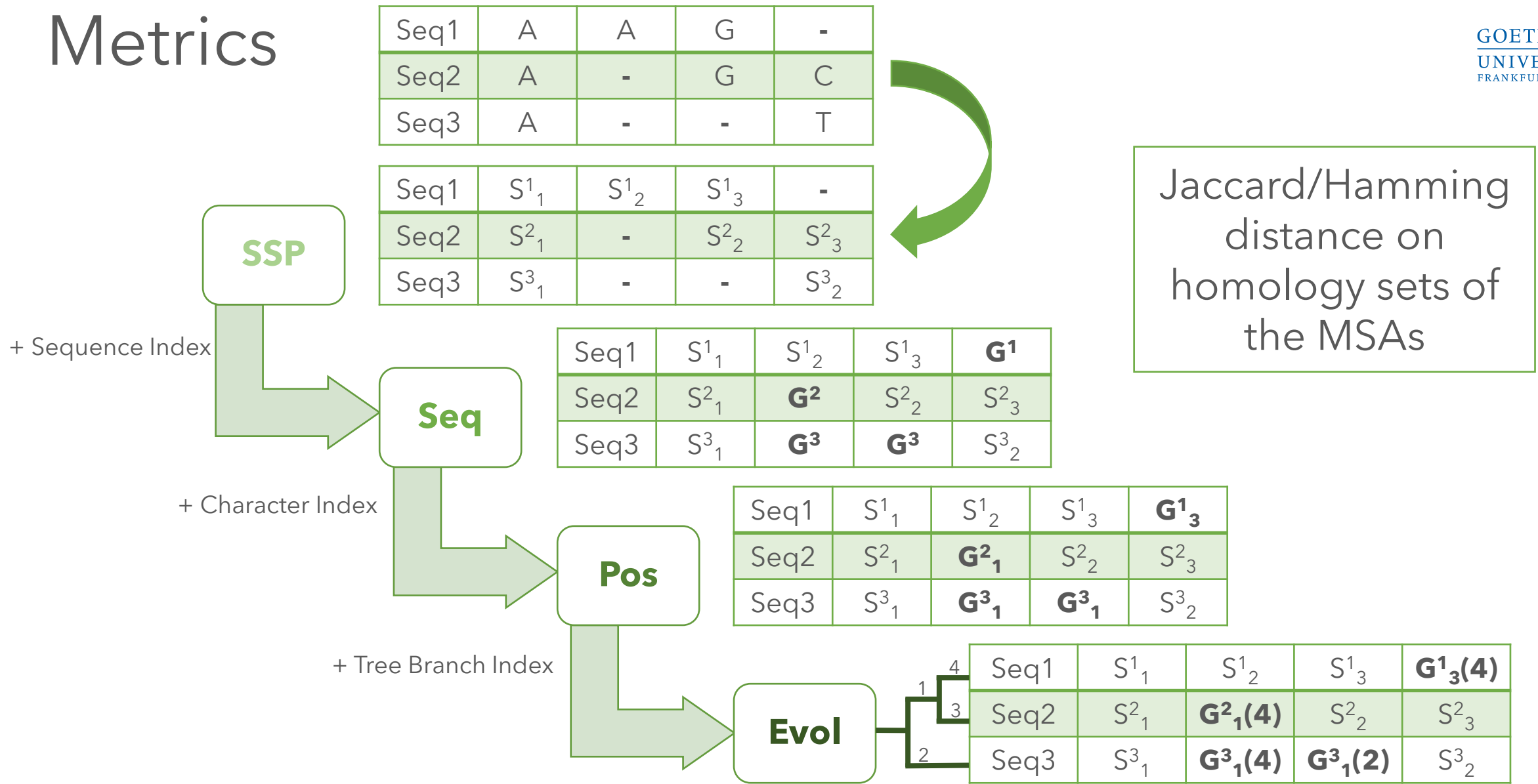
Triangle inequality
 $d(A,C) \leq d(A,B) + d(B,C)$



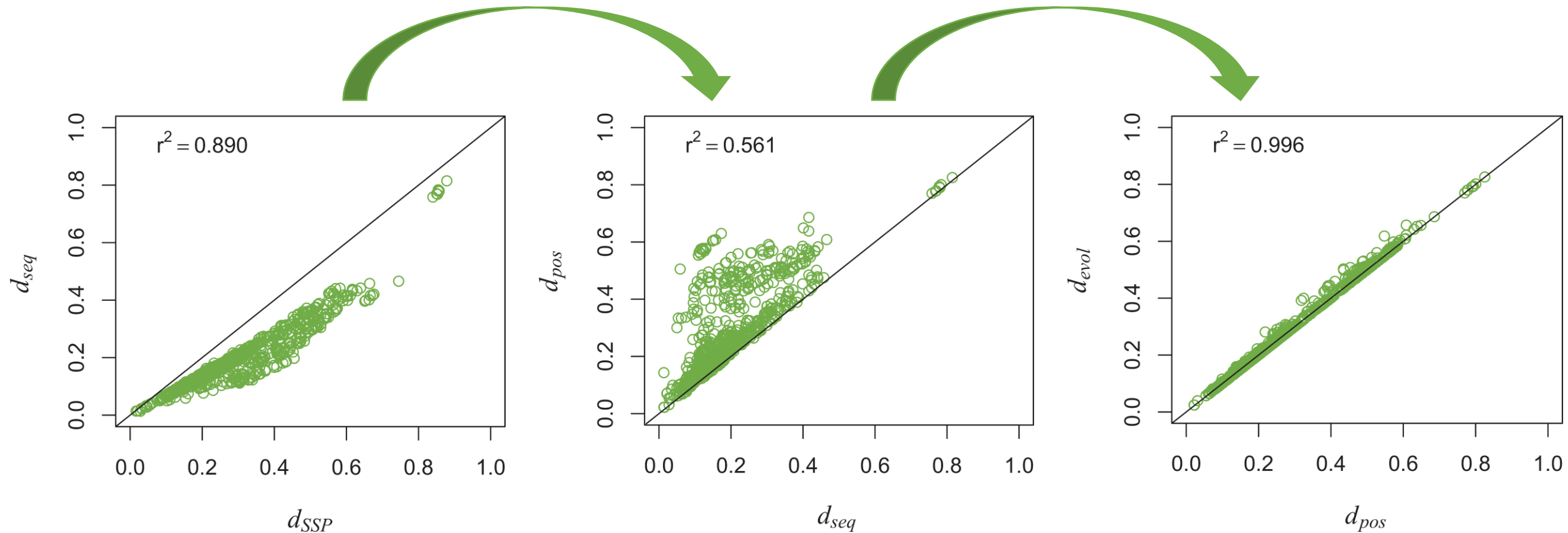
Why does the score need to be a metric?

All-against-all comparison!

Metrics

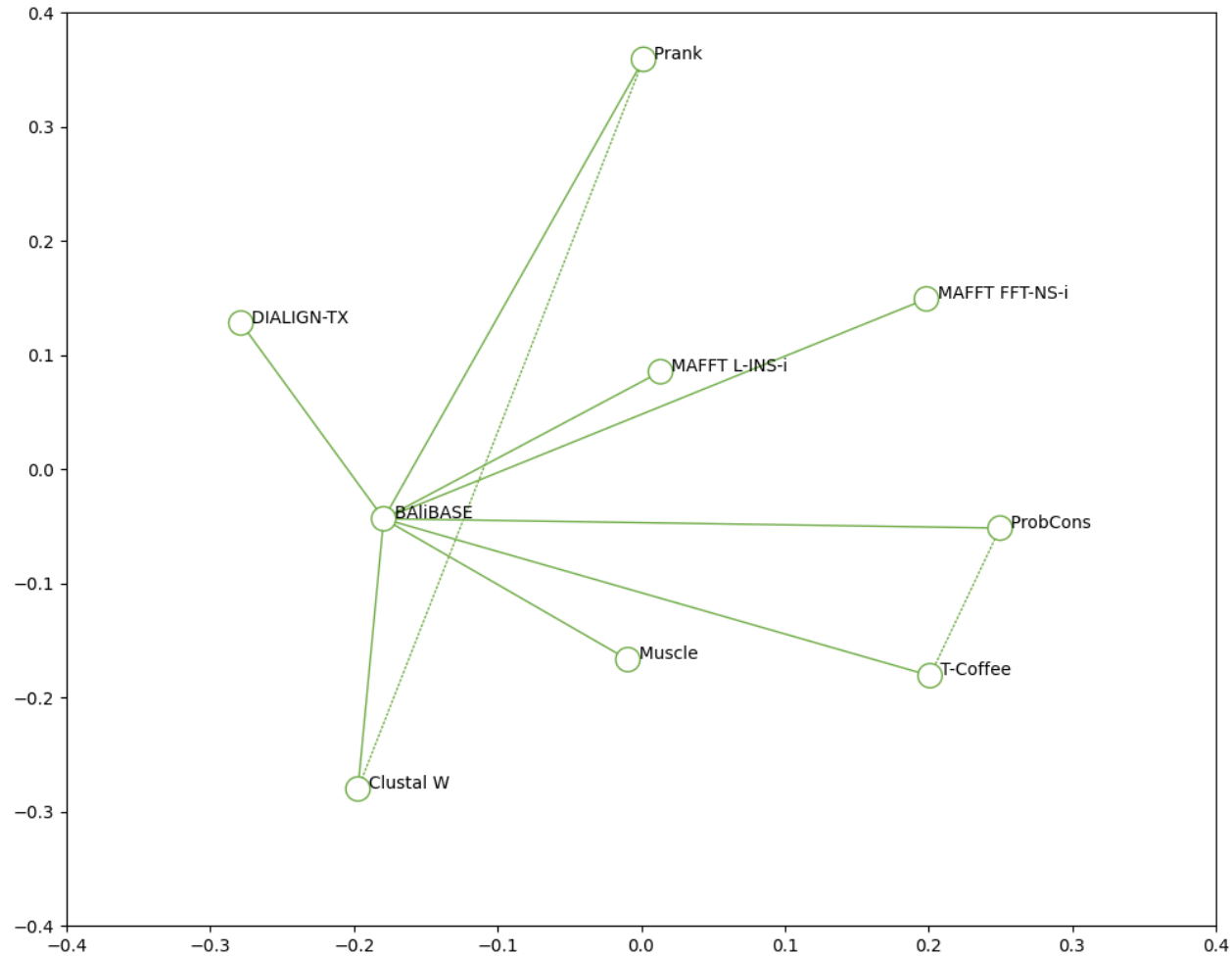


Comparison of metrics

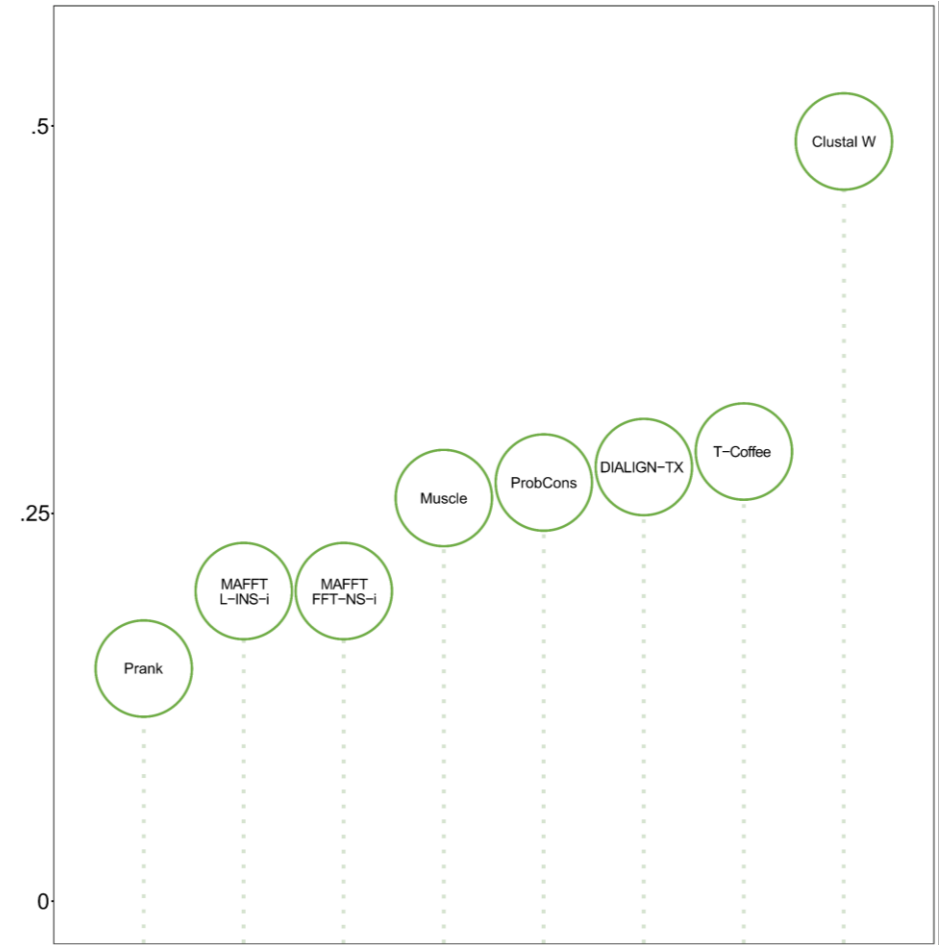


Benchmarking of aligners (d_{evol})

Real-world data - BALiBASE



Simulated data - INDELible



Thank you for your attention!

Haskell implementation *MetAI*
available on github!

*Presentation, poster,
sources and much more!*

Questions?

*Join Melanie at her poster talk and
our following discussion!*

