# 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 5<sup>th</sup> July 2021

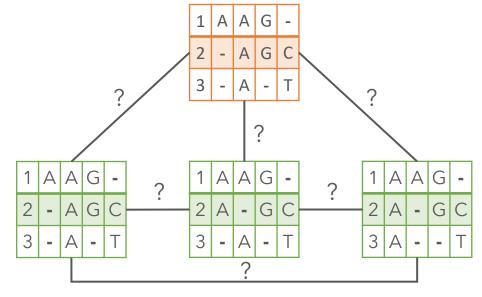


#### Motivation



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!**

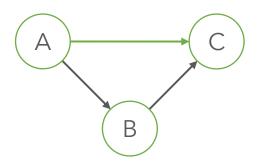
Identity of indiscernibles  $d(A,B) = 0 \leftrightarrow A=B$ 

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

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



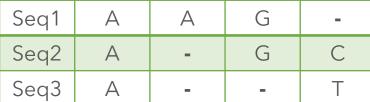




#### Why does the score need to be a metric?

### All-against-all comparison!

#### Metrics



-7	

S<sup>1</sup><sub>2</sub>

 $G^2_1$ 

S<sup>1</sup><sub>3</sub>

 $S_2^2$ 

+ Sequence Index

Seq3 S<sup>3</sup><sub>1</sub>

+ Character Index

**SSP** 

Seq1	S <sup>1</sup> <sub>1</sub>	S <sup>1</sup> <sub>2</sub>	S <sup>1</sup> <sub>3</sub>	G <sup>1</sup>
Seq2	S <sup>2</sup> <sub>1</sub>	G <sup>2</sup>	S <sup>2</sup> <sub>2</sub>	S <sup>2</sup> <sub>3</sub>
Seq3	S <sup>3</sup> <sub>1</sub>	G <sup>3</sup>	G <sup>3</sup>	S <sup>3</sup> <sub>2</sub>

Seq1

Seq2

Seq3

S<sup>1</sup><sub>1</sub>

 $S_1^3$ 

Pos

+ Tree Branch Index			1	
		Evo		



Jaccard/Hamming distance on homology sets of the MSAs

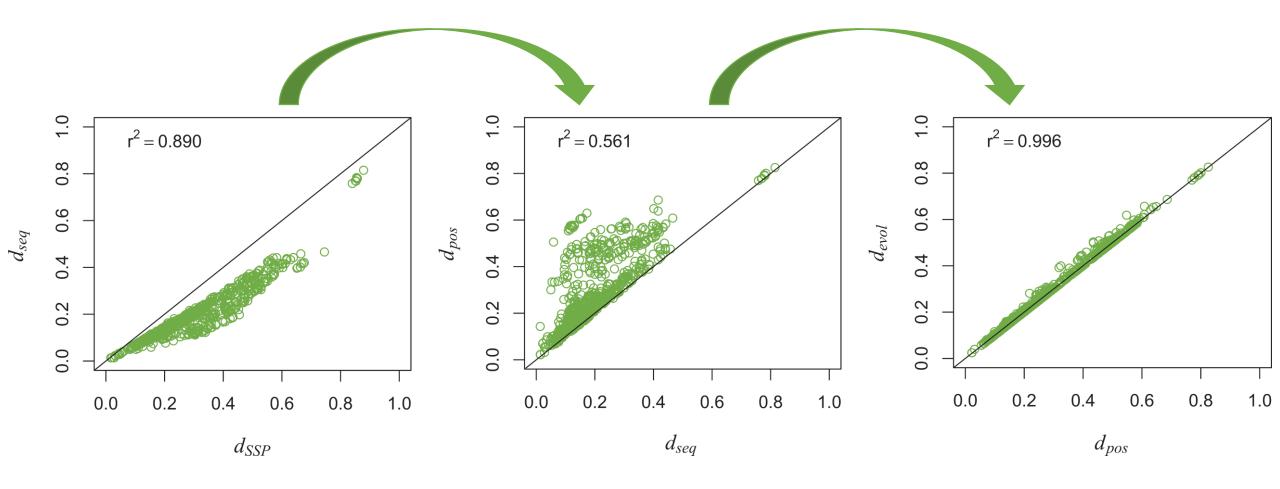
 $G_1^3$ **G**<sup>3</sup><sub>1</sub>  $S_2^3$ Seq1  $S_1$  $S_2^1$ S<sup>1</sup><sub>3</sub>  $G_{3}^{1}(4)$  $S_1^2$  $S_3^2$  $G_{1}^{2}(4)$ Seq2  $S_2^2$  $S_1^3$ S<sup>3</sup><sub>2</sub>  $G_{1}^{3}(4)$  $G_{1}^{3}(2)$ Seq3

 $G_3^1$ 

 $S^2_3$ 

## Comparison of metrics

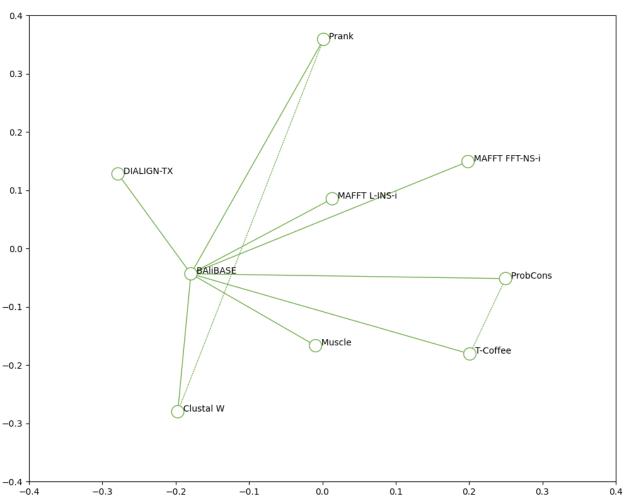




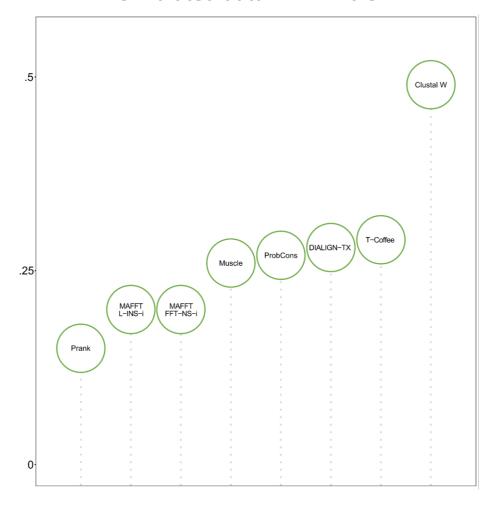
# Benchmarking of aligners $(d_{evol})$







#### Simulated data - INDELible



 $5^{th}$  July 2021

## Thank you for your attention!



Haskell implementation *MetAl* available on github!

Presentation, poster, sources and much more!

## Questions?

Join Melanie at her poster talk and our following discussion!

