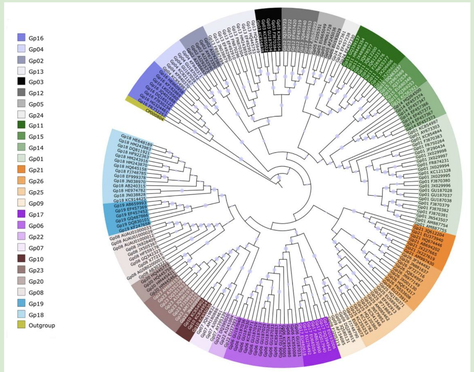
Metabolic modeling of soil *Acidobacteriota* improves with custom resource inputs

Ashley Y. Eng[1,2], Zoey R. Werbin[3], Jennifer M. Bhatnagar[3]
[1]Boston University Bioinformatics BRITE REU Program, Summer 2022,
[2]Department of Microbiology, University of Massachusetts Amherst, [3]Department of Biology, Boston University

Introduction



Acidobacteriota
- large part of soil microbiome[1]
- drivers of biogeochemical cycles[1]
- difficult to culture and poorly understood[1]

Genome scale metabolic models (GEMs)
- model metabolic activity *in silico*[2]
- model microbial processes and interactions[3]

Knowledge gap: do model-creation methods accurately reflect *Acidobacteriota* physiology
Overarching goal: create and validate metabolic models for cultured strains of *Acidobacteriota*

Fig 1. Phylogenetic tree of *Acidobacteriota*, many of which have not been cultured. Gp = group (Kielak *et al.*, 2016)

Hypotheses

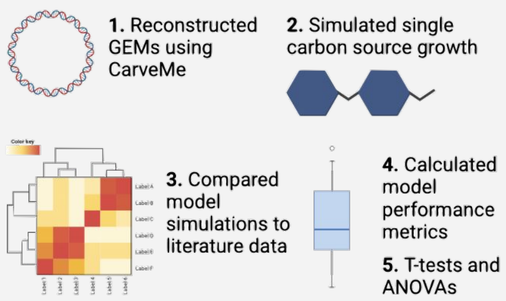
H1. Model performance (F1, accuracy, specificity, sensitivity, and precision) will be worse for *Acidobacteriota* compared to the model organisms (*B. subtilis, E. coli, P. aeruginosa, R. solanacearum, S. oneidensis*) used to benchmark GEM algorithms

H2. Across *Acidobacteriota*, larger genome sizes will require fewer gap-filled reactions (i.e., biochemical reactions added to force growth simulations that match observed growth on a specific media)

H3. The number of gap-filled reactions will be lower when complete media (i.e., generic media including all metabolites usable by the organisms) is used in comparison to custom media (i.e., media closely matching experimental conditions)

H4. Monosaccharide metabolism will be more accurate in comparison to polysaccharide metabolism

Methods



Model performance metrics: F1, accuracy, sensitivity, specificity, precision
Analyses of associations with: Genome size, media type, sugar classification

Discussion

- CarveMe might be appropriate for *Acidobacteriota*
- Model performance may be affected by other factors
- Complete media had more gap-filled reactions likely because it lacks metabolites that are unique to the custom media used for culturing *Acidobacteriota*

Results

No significant difference in performance metrics between CarveMe model organisms and *Acidobacteriota*

F1: $t(10.59) = 1.97$, $p = 0.08$
Precision: $t(5.75) = 2.29$, $p = 0.06$
Sensitivity: $t(6.86) = -0.67$, $p = 0.53$
Specificity: $t(10.1) = 1.23$, $p = 0.25$

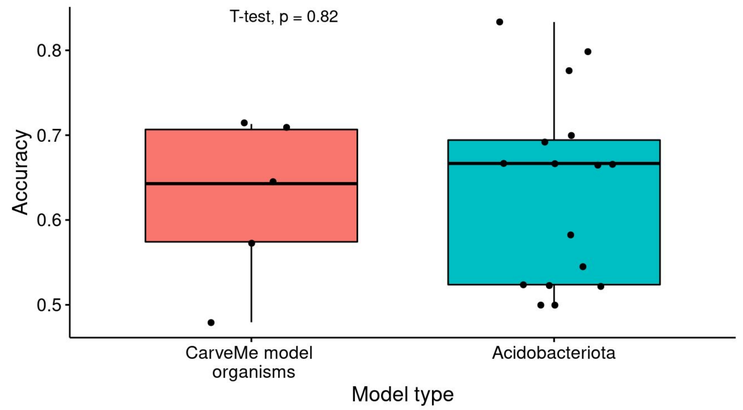


Fig 2. Accuracy was calculated for single carbon source growth simulations and experimental data (n = 5 for model organisms, n = 16 for *Acidobacteriota*, n = 21 for carbon sources)

Number of gap-filled reactions was significantly lower for custom media vs. complete media

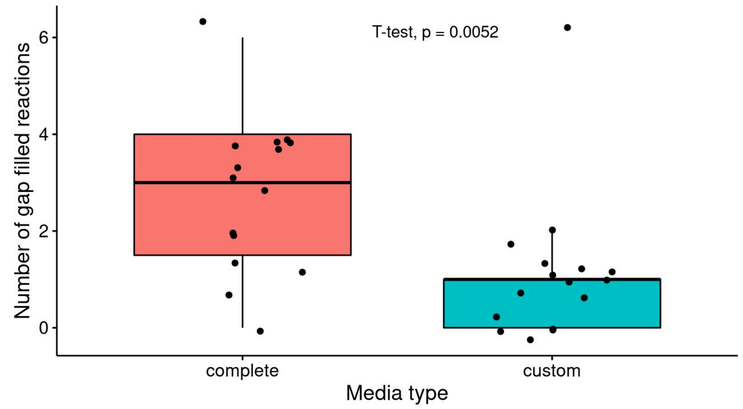


Fig 4. Number of gap-filled reactions for each GEM gap filled on custom and complete media reconstructed for *Acidobacteriota* (n = 16 for *Acidobacteriota)*.

No relationship between genome size and the number of gap-filled reactions

Variable	t-test	p-value
Genome size	$t(30) = -0.58$	0.57
Media type	$t(30) = -1.18$	0.25
genome:media	$t(30) = 0.45$	0.65

Table 1. Regression statistics for genome size on number of gap-filled reactions.

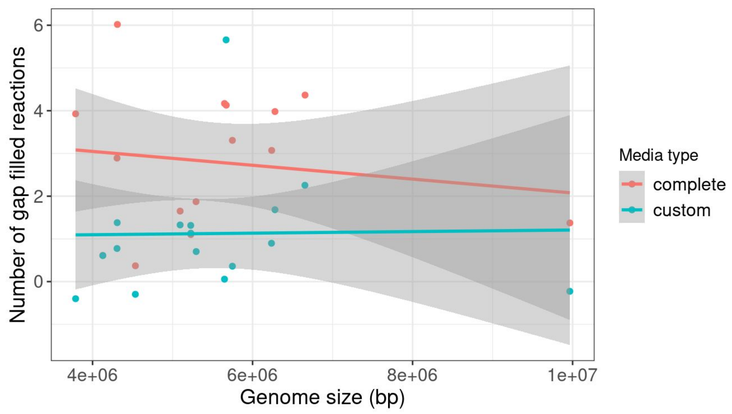


Fig 3. Trendline fitted for data of genome size vs. number of gap-filled reactions of each GEM gap filled on custom and complete media reconstructed for *Acidobacteriota* (n = 16 for *Acidobacteriota*).

Model performance did not depend on sugar classification

F1: $F(2,11) = 0.43$, $p = 0.66$
Precision: $F(2, 12) = 0.35$, $p = 0.71$
Sensitivity: $F(2, 15) = 0.50$, $p = 0.62$
Specificity: $F(2, 13) = 1.16$, $p = 0.35$

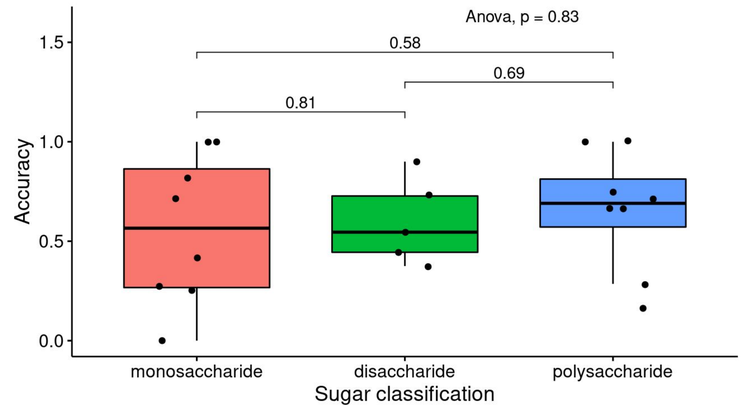


Fig 5. Accuracy was calculated for single carbon source growth simulations and experimental data. Single carbon sources were classified into sugar types (n = 16 for *Acidobacteriota*, n = 21 for carbon sources).

References

1. Kalam, S. *et al. Front Microbiol.* 11 (2020).
2. Machado D. *et al. Nucleic Acids Res.* 46(15):7542–53 (2018).
3. O'Brien EJ *et al.*. *Cell.* 161(5):971–87 (2015).

Acknowledgements

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program