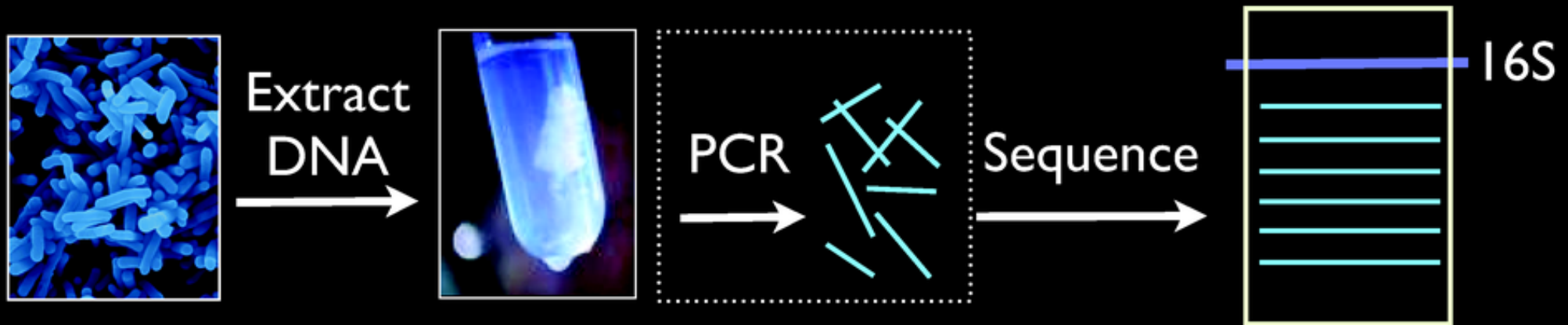


PhylOTU: Quantifying Microbial Diversity and Identifying Novel Taxa from Metagenomic Data

Thomas J. Sharpton

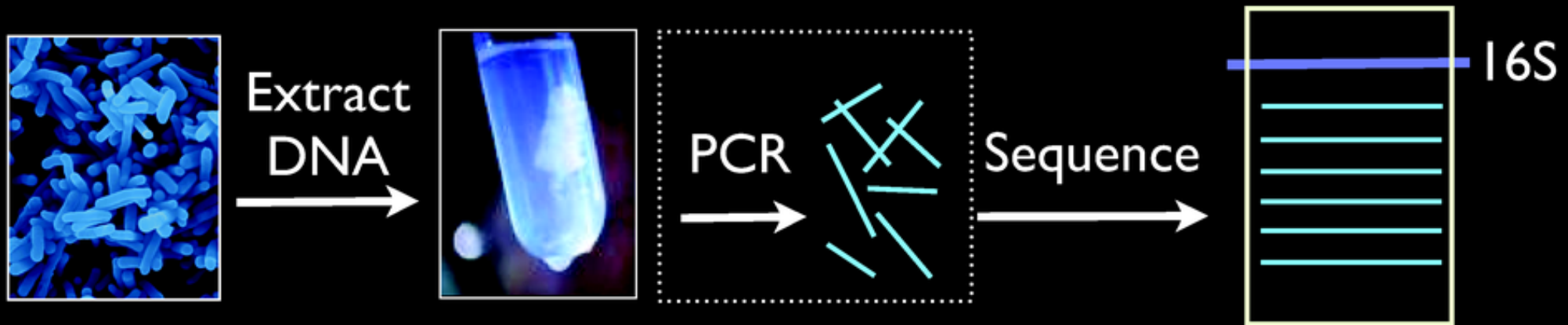
Katherine Pollard Laboratory
J. David Gladstone Institutes
University of California, San Francisco

Studying Microbes *In Situ*



PCR-Based 16S rRNA Sequencing

Studying Microbes *In Situ*



PCR-Based 16S rRNA Sequencing

The ISME Journal (2009) 3, 1365–1373
© 2009 International Society for Microbial Ecology All rights reserved 1751-7362/09 \$32.00
www.nature.com/ismej

ORIGINAL ARTICLE

Polymerase chain reaction primers miss half of rRNA microbial diversity

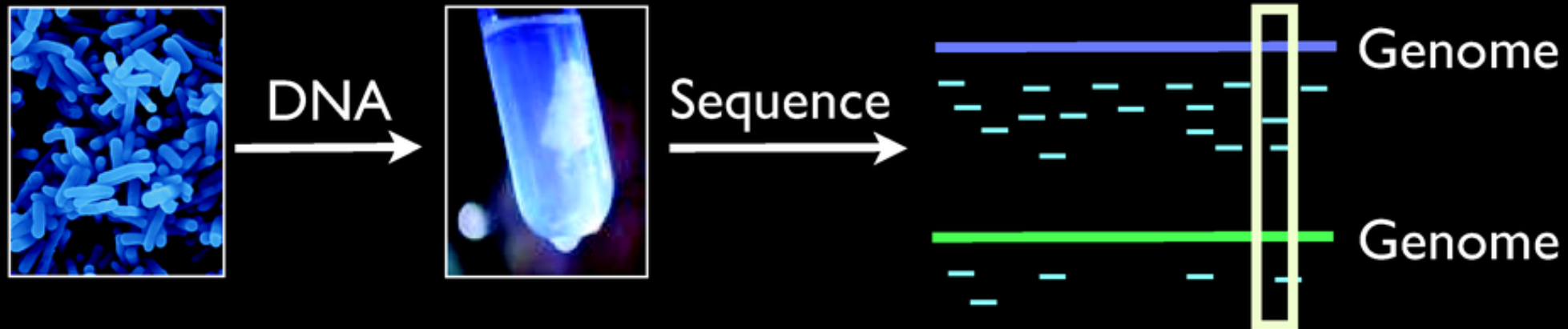
SunHee Hong^{1,4}, John Bunge², Chesley Leslin¹, Sunok Jeon³ and Slava S Epstein^{1,4}

¹Department of Biology, Northeastern University, Boston, MA, USA; ²Department of Statistical Science, Cornell University, Ithaca, NY, USA; ³Department of Environmental Science, Kangwon National University, Kangwon-Do, Korea and ⁴Marine Science Center, Northeastern University, Nahant, MA, USA

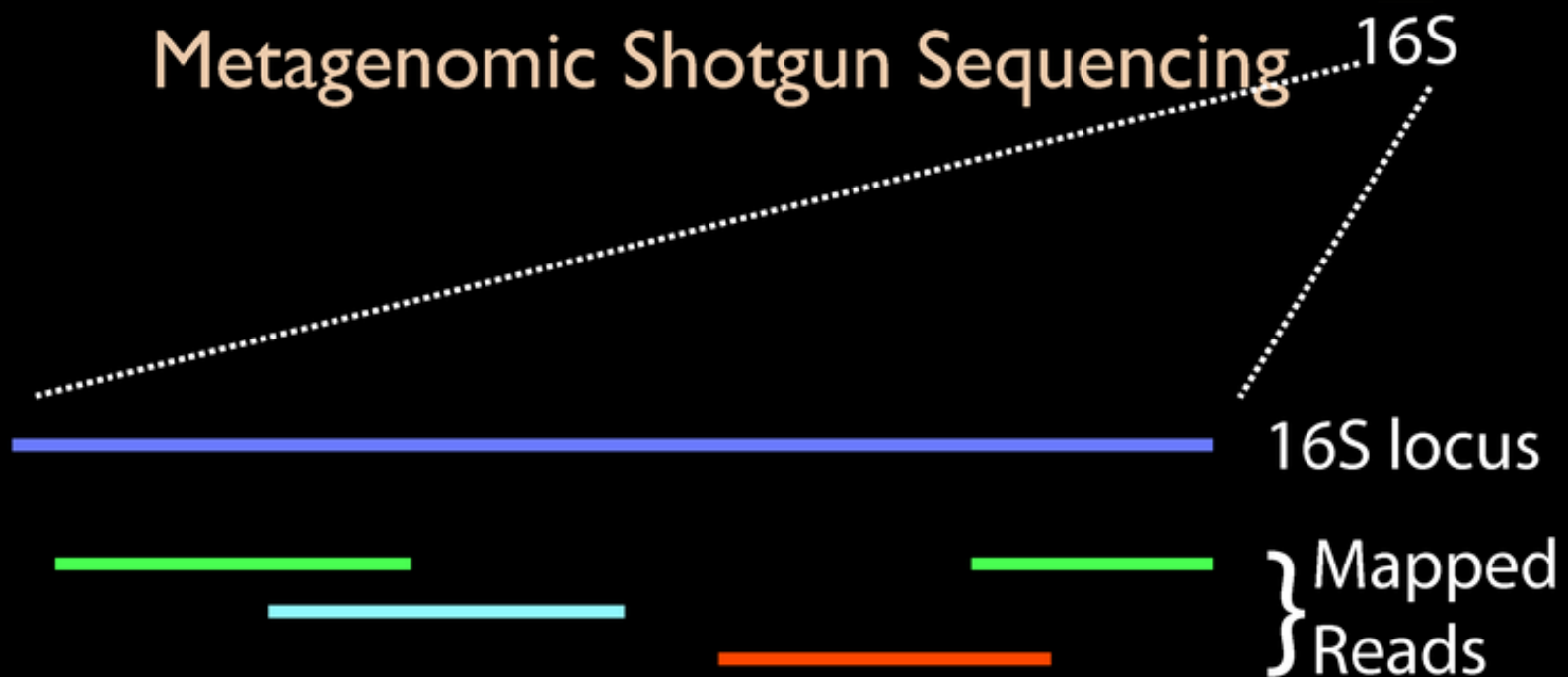
Studying Microbes *In Situ*



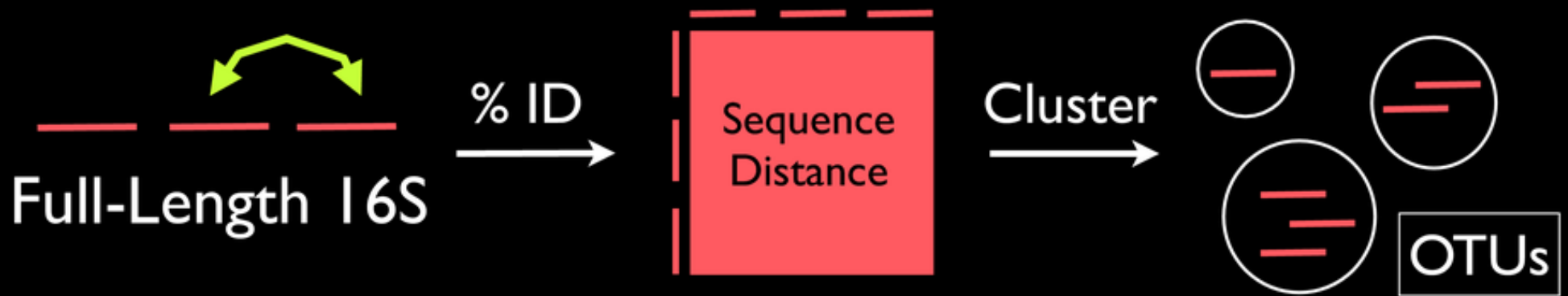
Studying Microbes *In Situ*



Metagenomic Shotgun Sequencing 16S




Operational Taxonomic Units



Operational Taxonomic Units

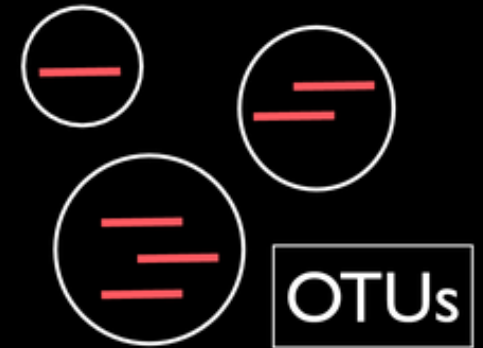
Full-Length 16S



% ID

Sequence
Distance

Cluster



Branch
Length

Phylogenetic
Distance

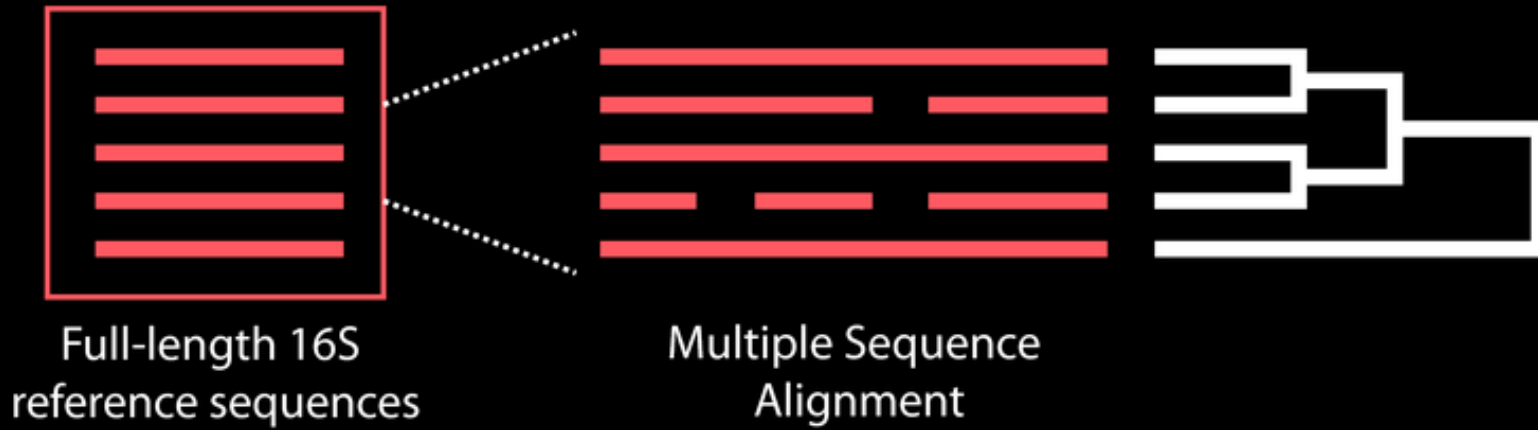
Cluster



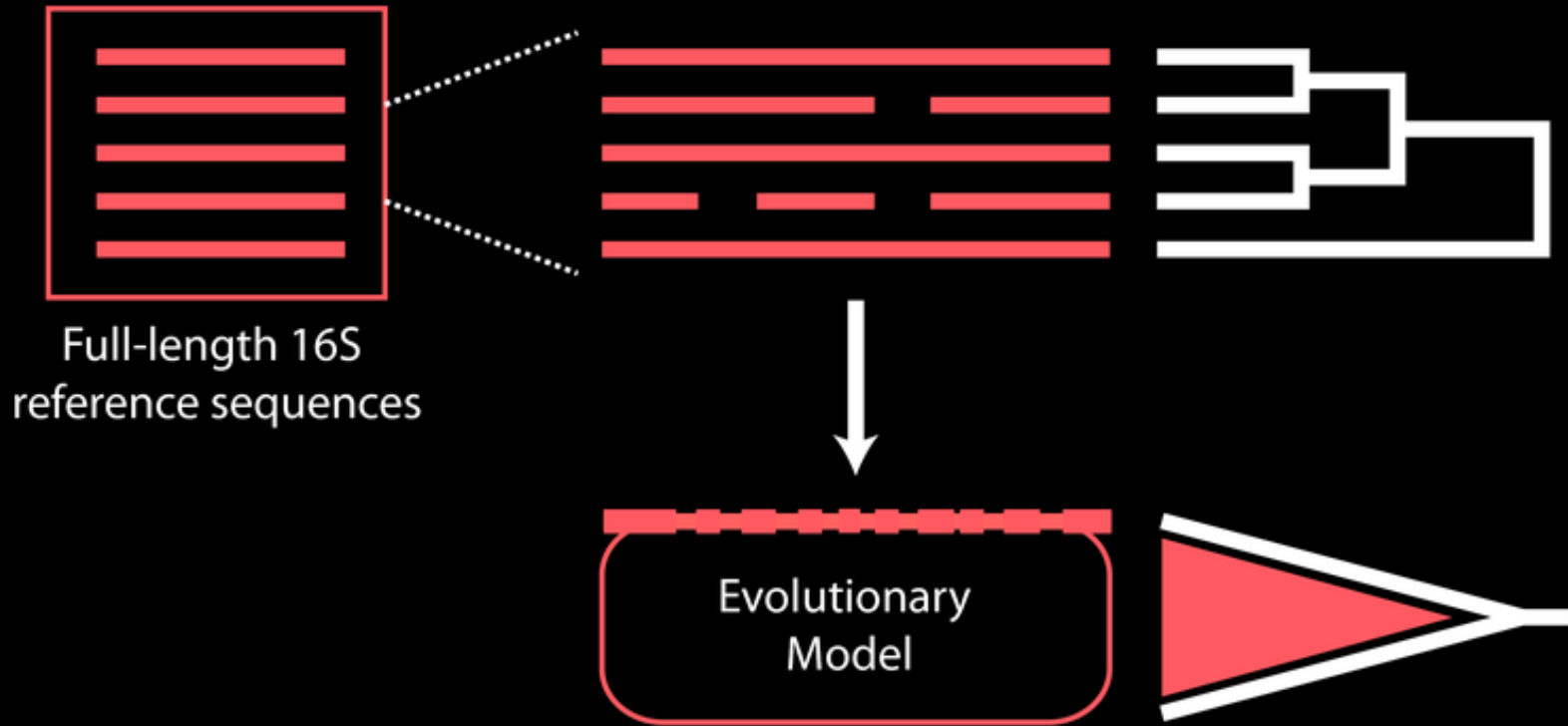
16S Read Tree



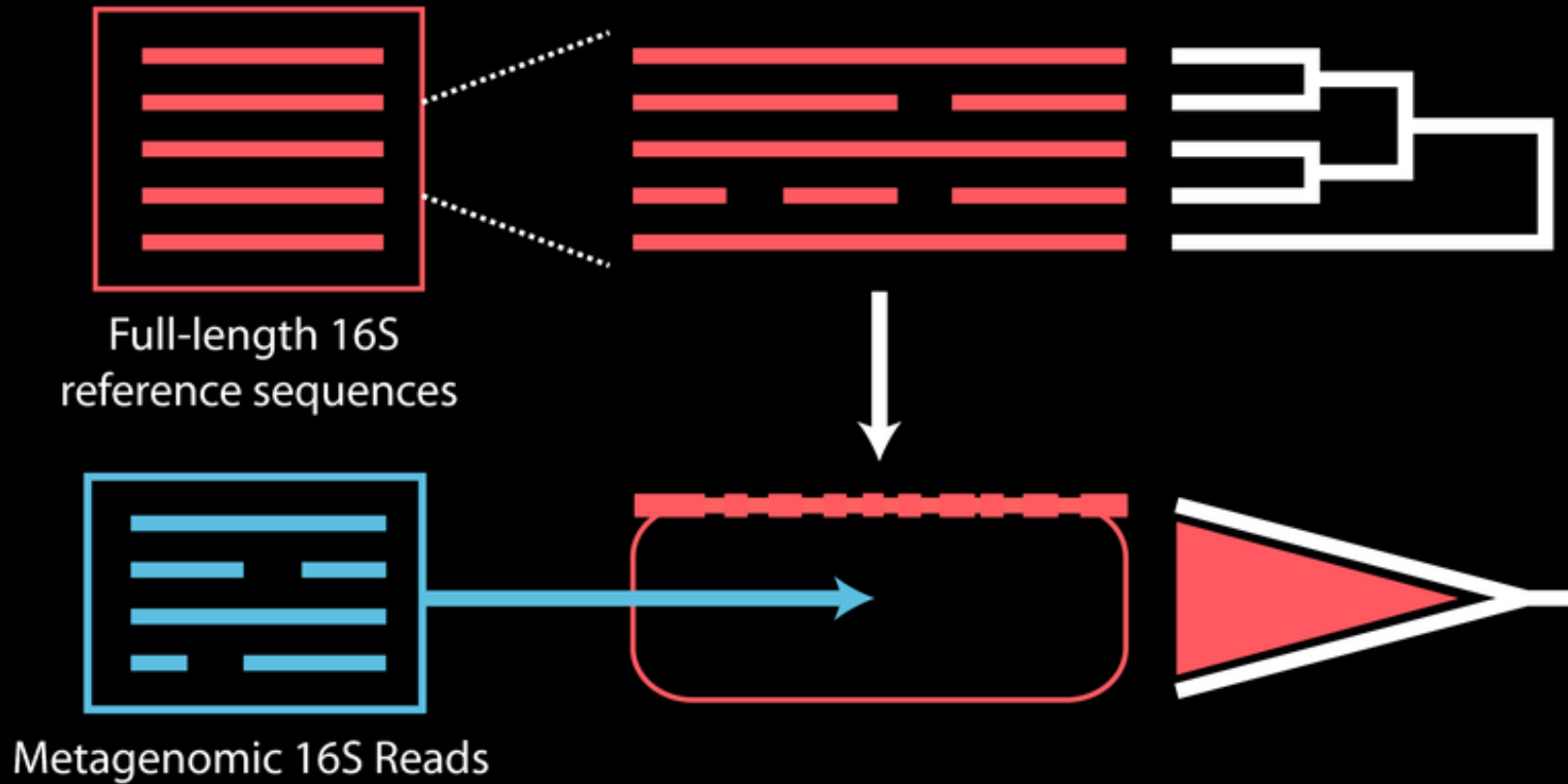
Our Solution: PhyloOTU



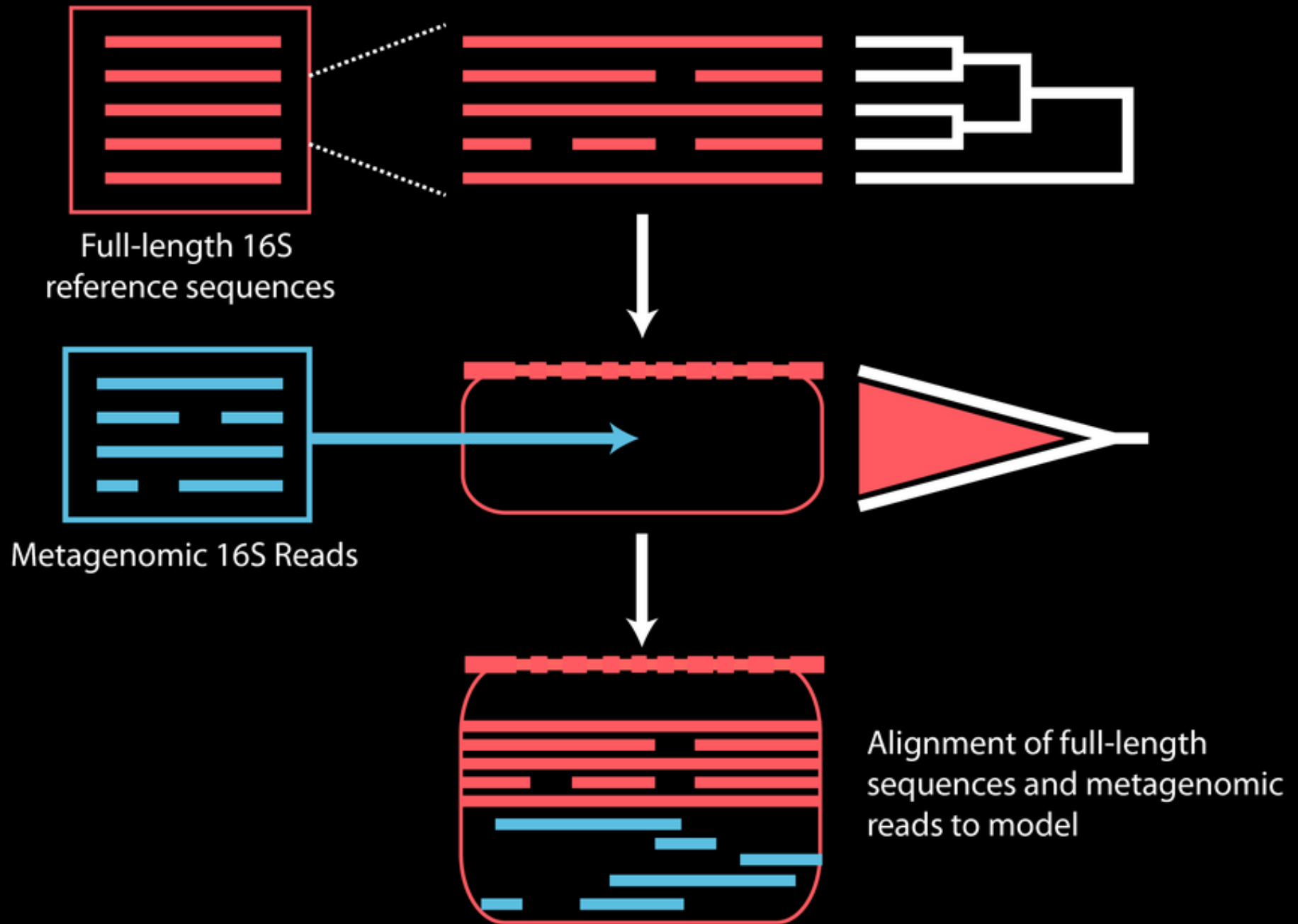
Our Solution: PhyloOTU



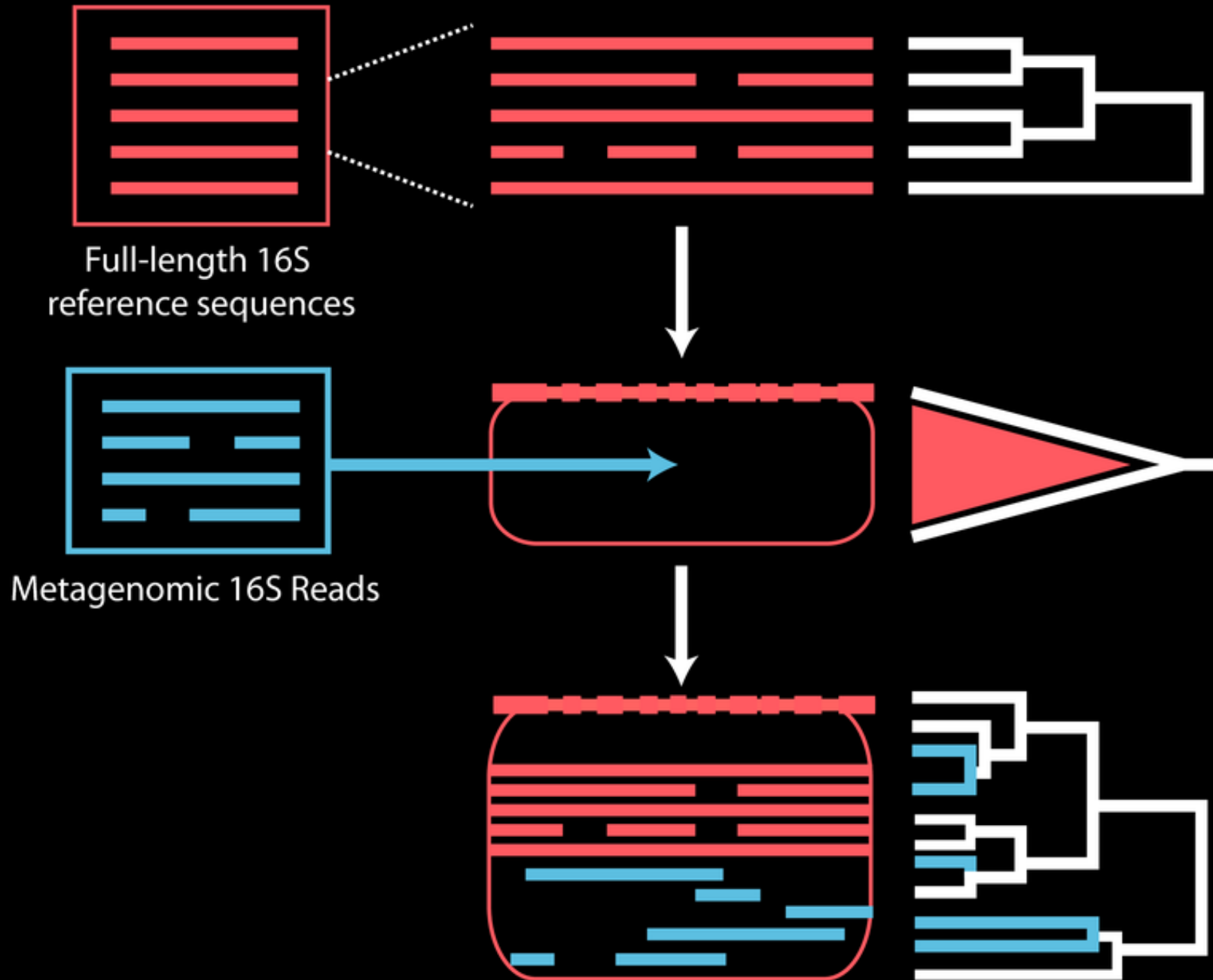
Our Solution: PhyIOTU



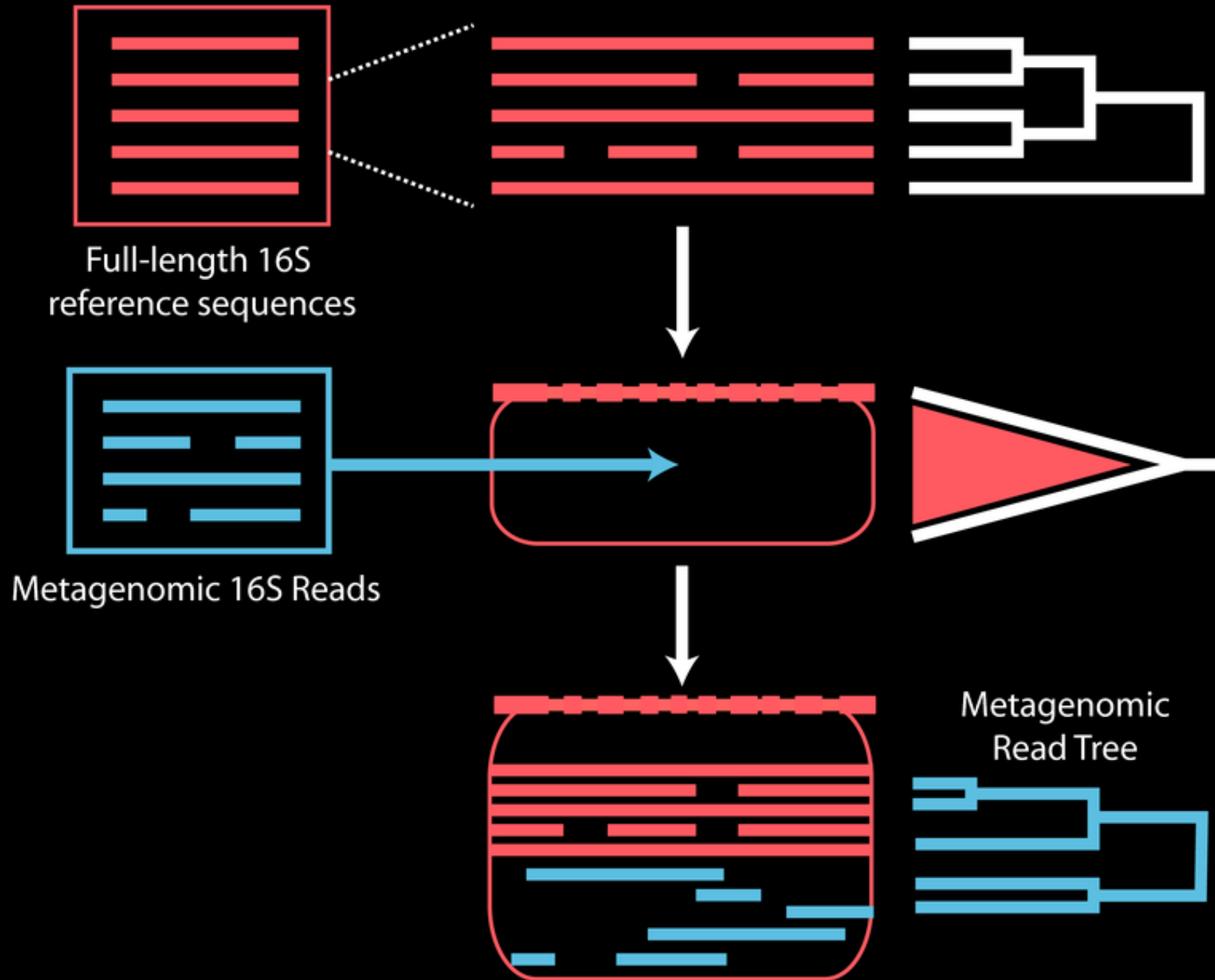
Our Solution: PhyLOTU



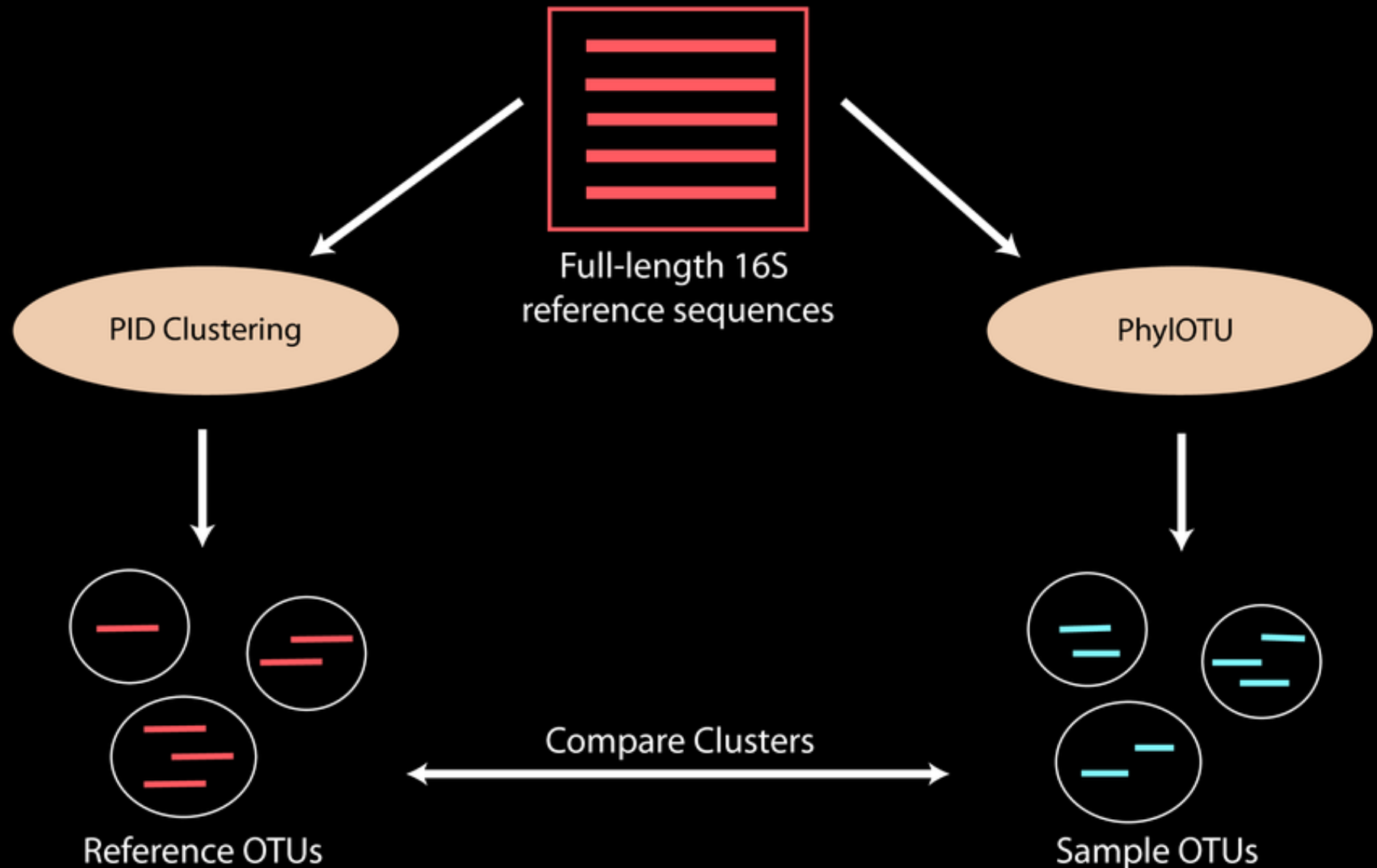
Our Solution: PhyLOTU



Our Solution: PhyloOTU



Statistical Assessment of PhylOTU Clustering



Statistical Assessment of PhylOTU Clustering

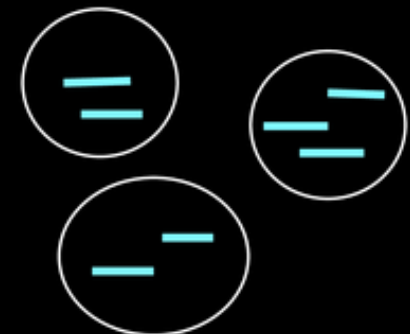
True Positive Rate: 91.83%

True Negative Rate: 99.81%



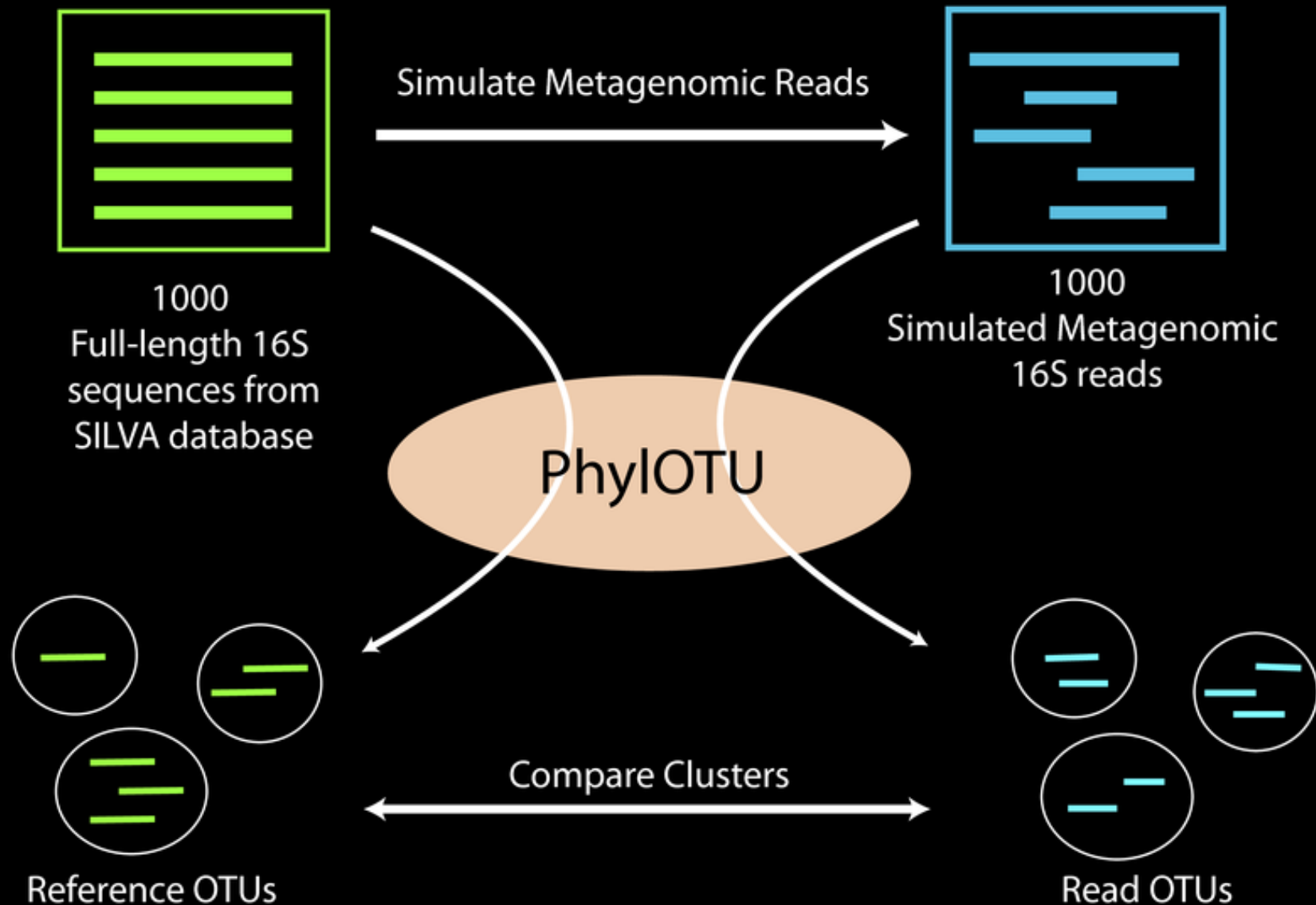
Reference OTUs

Compare Clusters

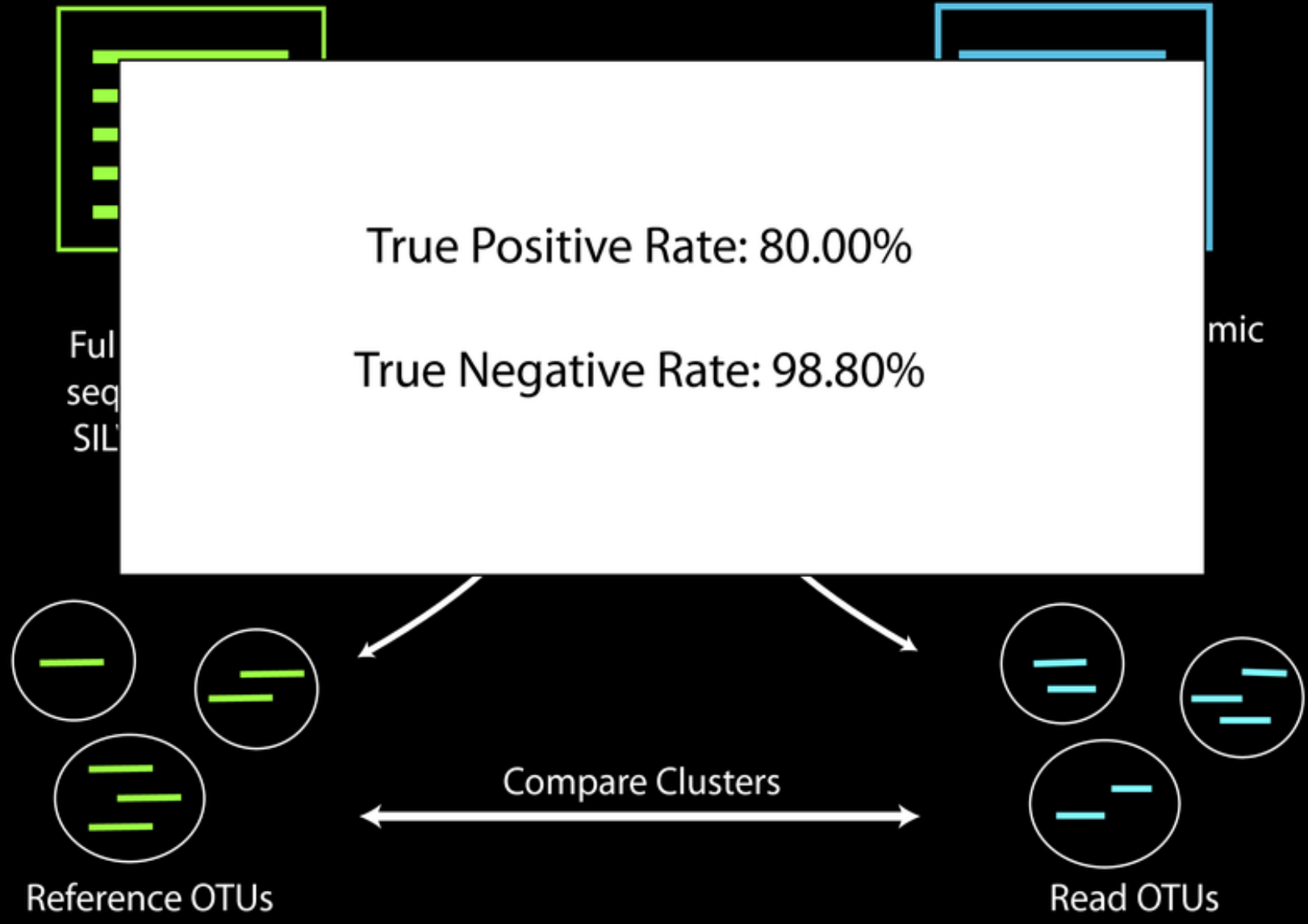


Sample OTUs

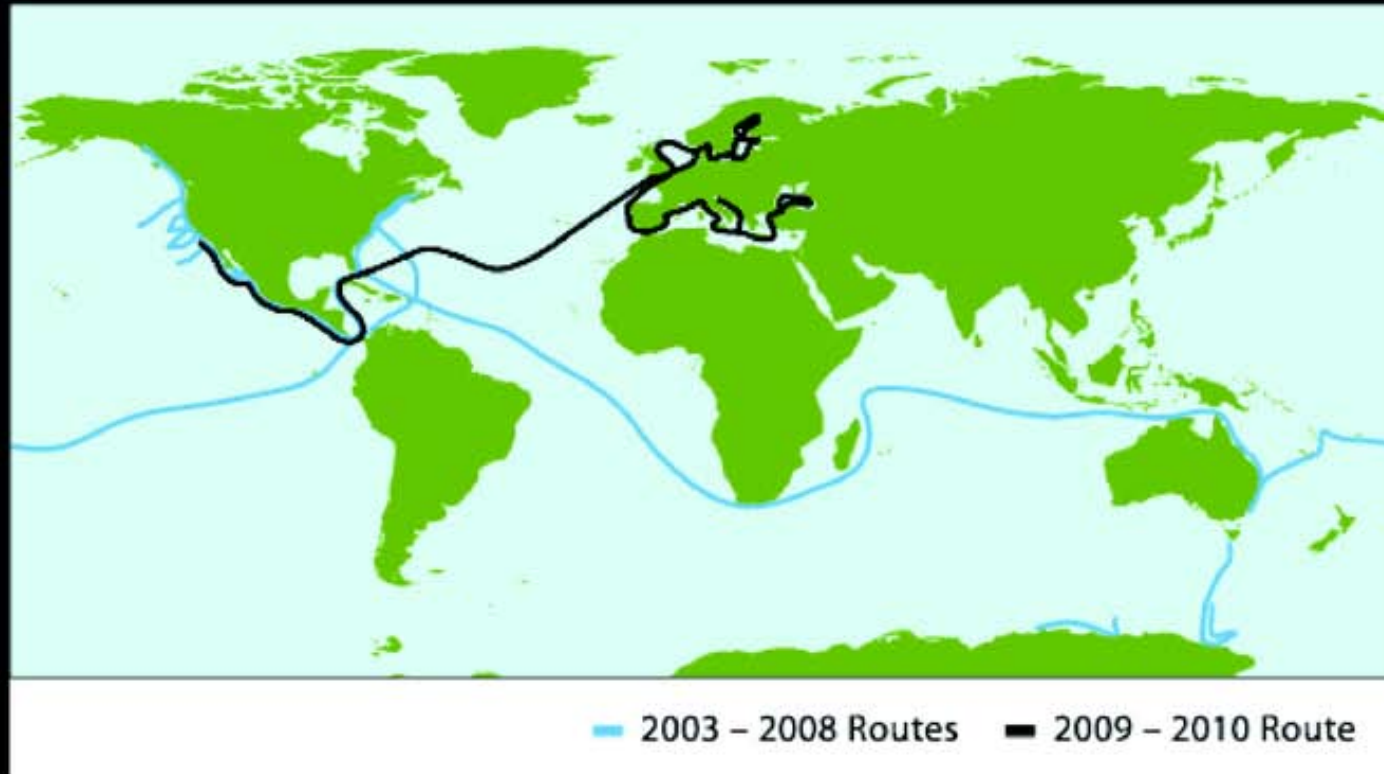
Statistical Assessment through Simulation



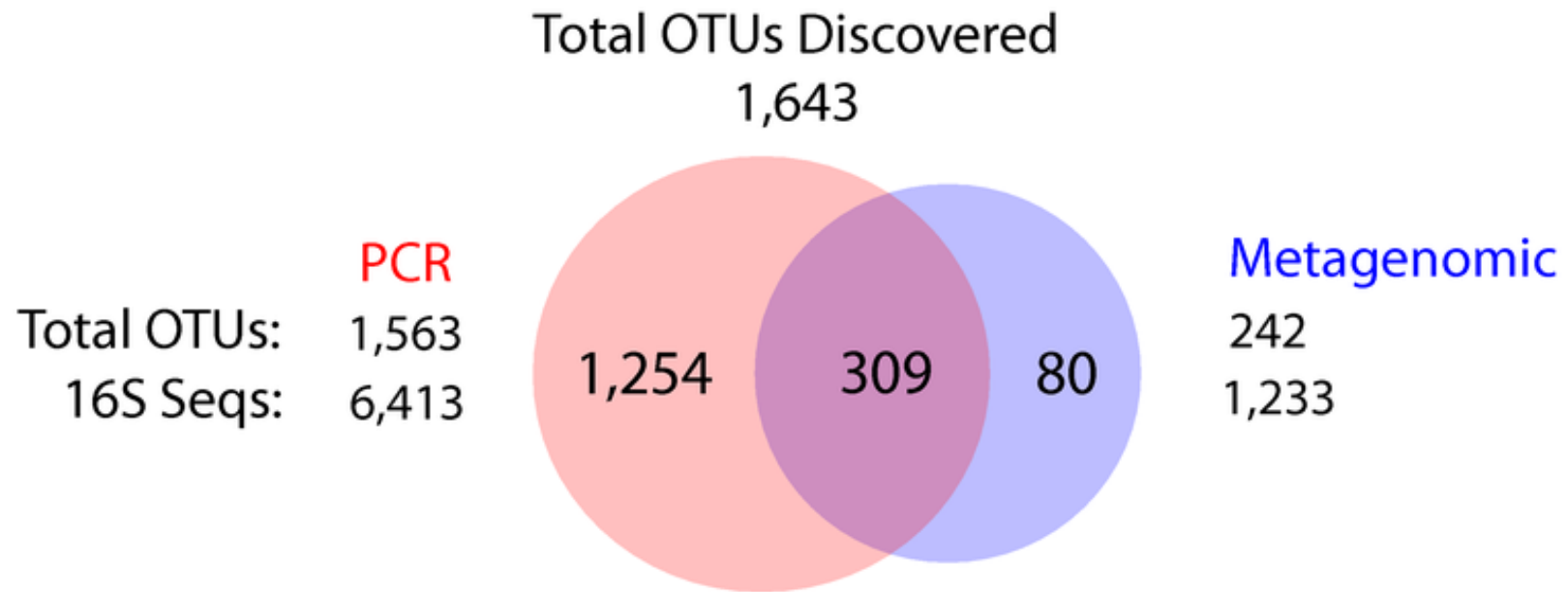
Statistical Assessment through Simulation



Global Ocean Survey Data

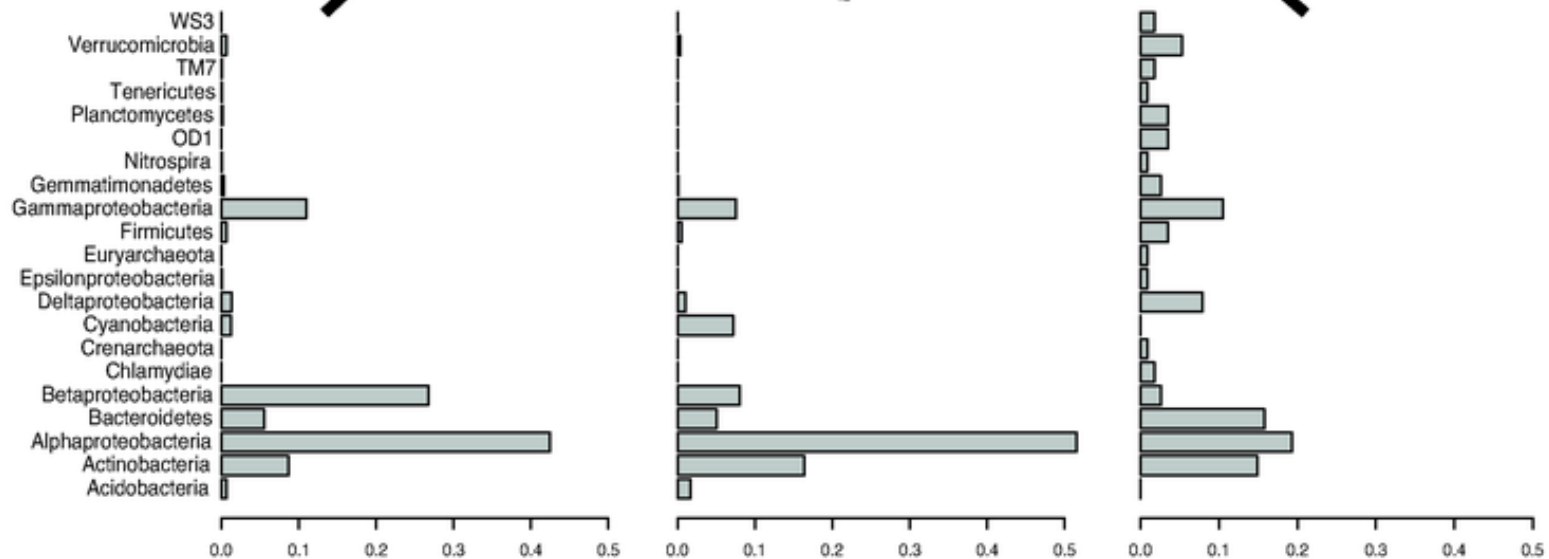


Global Ocean Survey OTUs



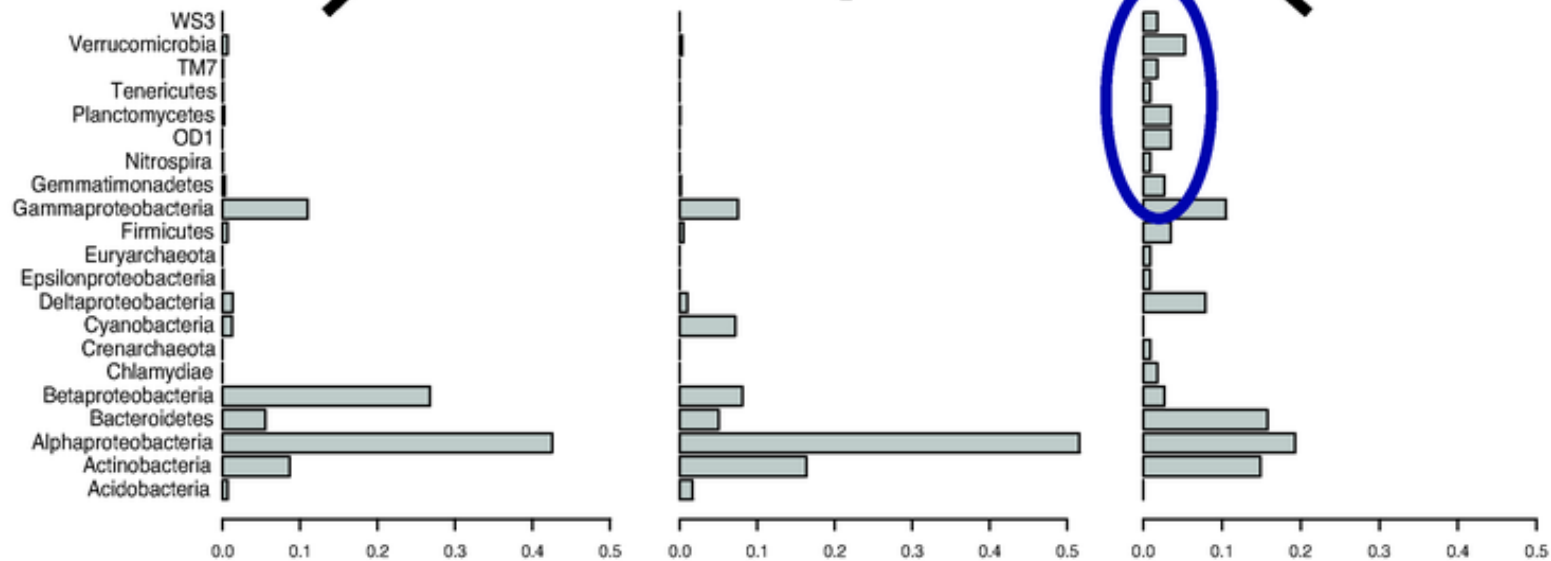
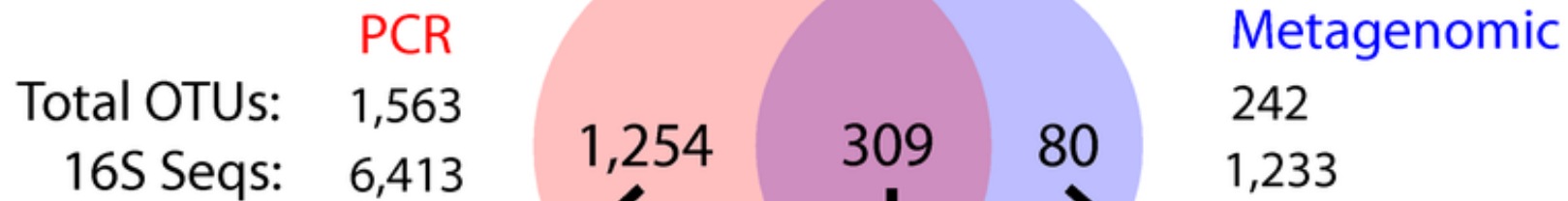
Global Ocean Survey OTUs

Total OTUs Discovered
1,643



Global Ocean Survey OTUs

Total OTUs Discovered
1,643



Summary

PhylOTU is a clusters non-overlapping 16S sequences into OTUs with reasonable accuracy

Can use metagenomic 16S sequences to identify microbial taxa that may be missed by targeted sequencing approaches

Currently works on 454-like sequence length reads or greater

Additional details can be found in our manuscript:
Sharpton et al. PLoS Computational Biology 2011 7(1)

Acknowledgements

Collaborators

UCSF

Katherine Pollard
Samantha Riesenfeld
Joshua Ladau
Rebecca Lamb

UC Davis

Jonathan Eisen

U of Oregon

Jessica Green
Steven Kembel
James O'Dwyer

<http://iseem.org>

Source Code Freely Available at
<http://github.com/sharpton/PhylOTU>

Community Software and Database Developers

RDP

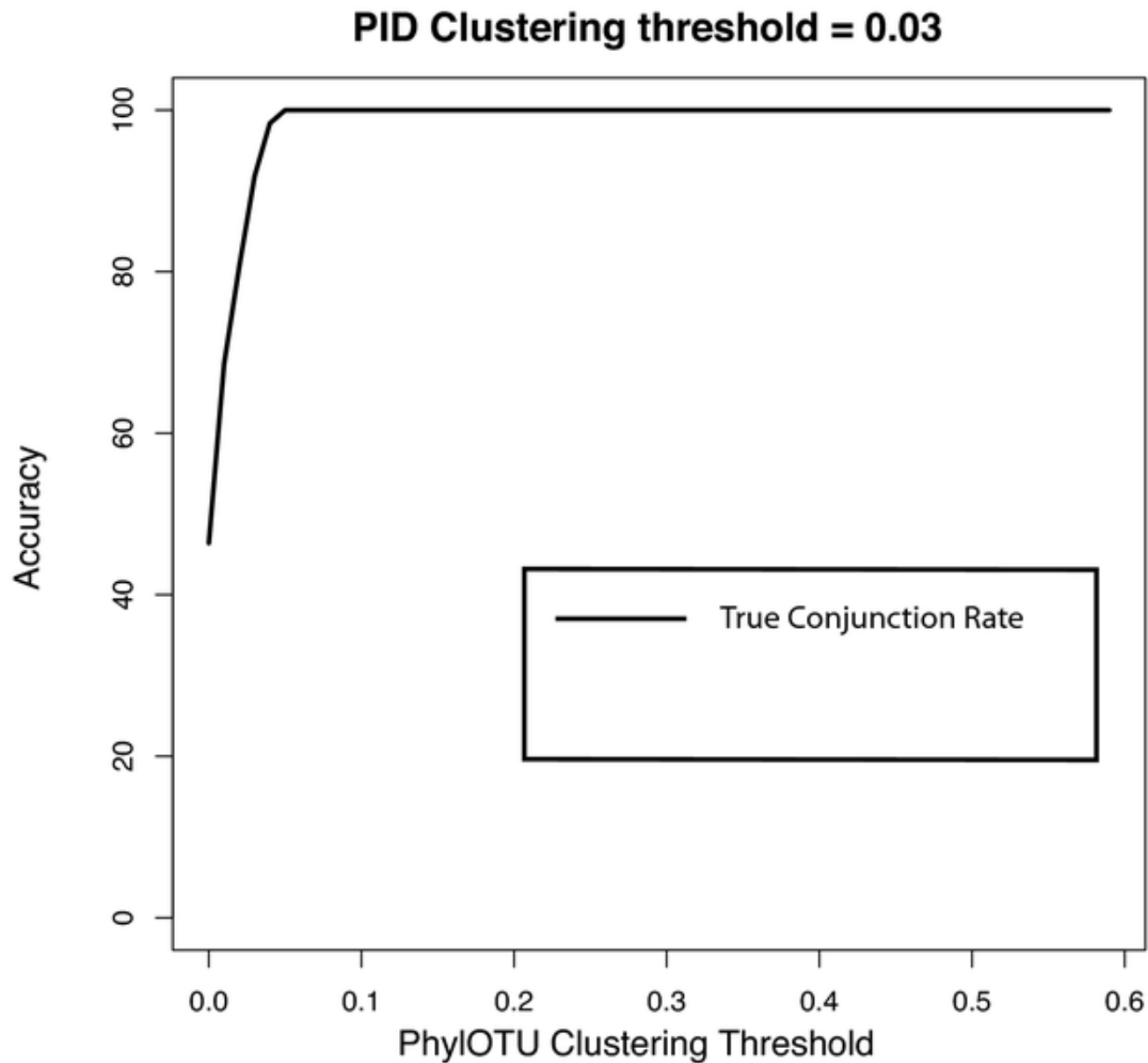
Schloss Lab

Eddy Lab

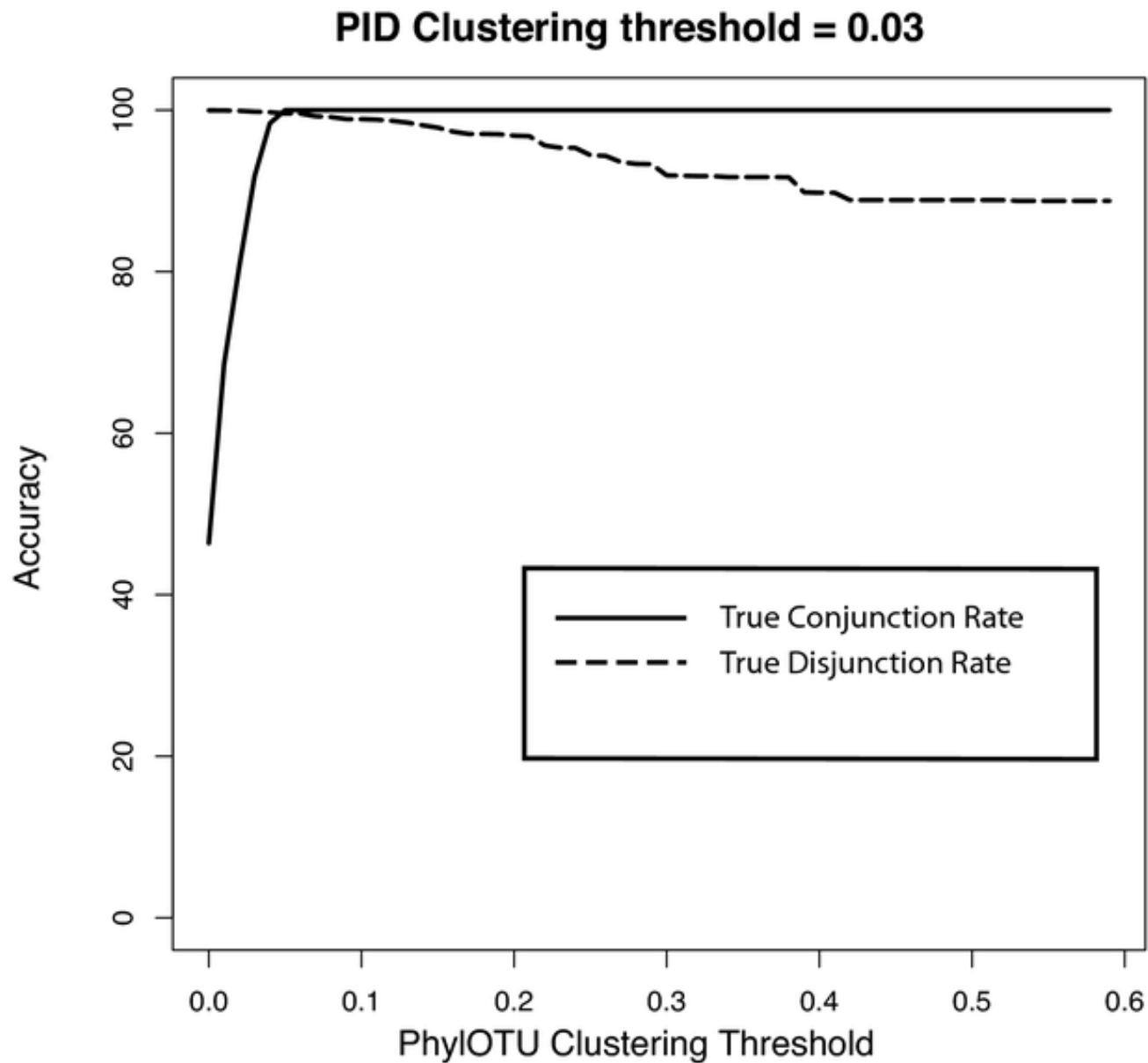
Sun Lab



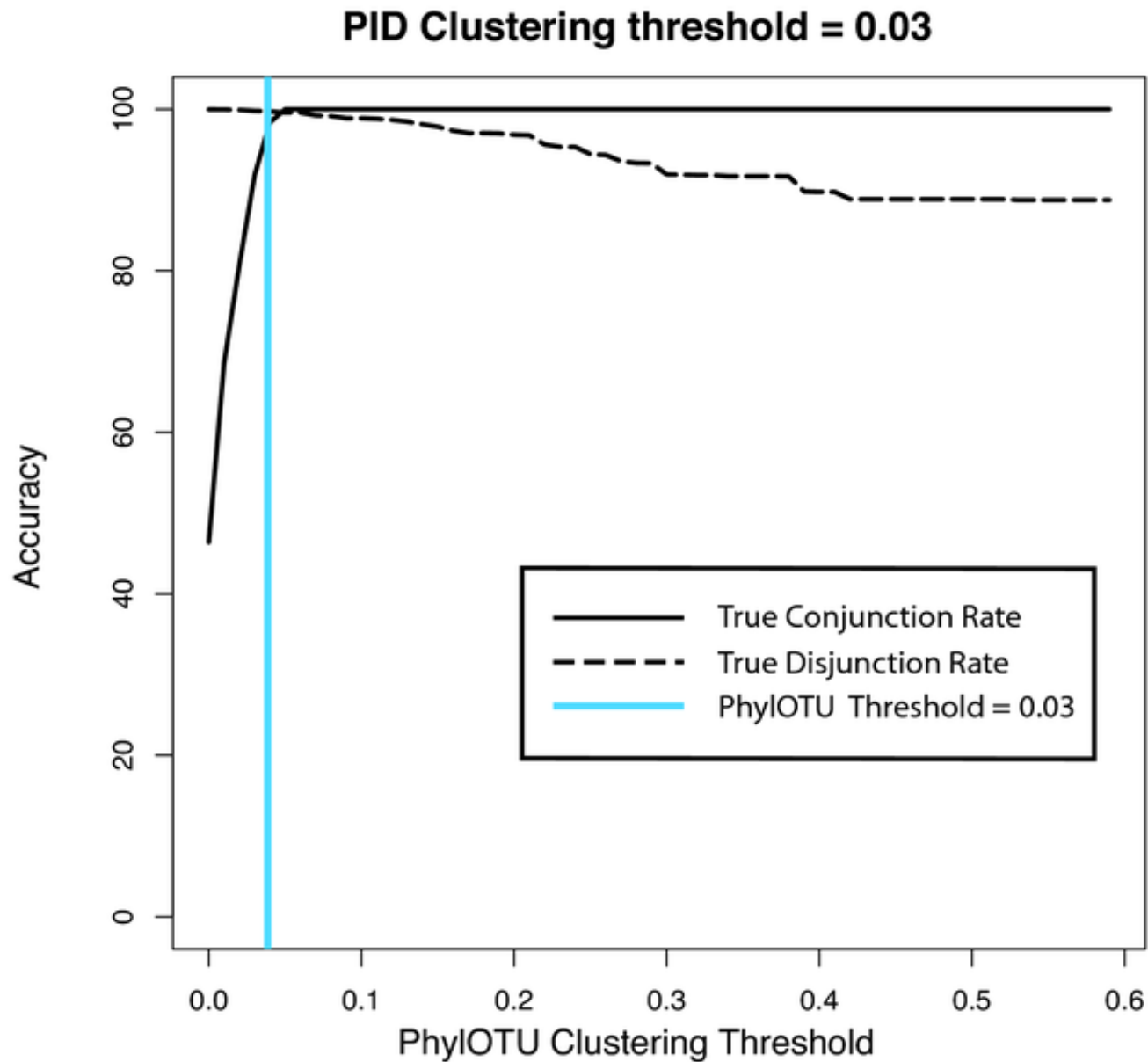
PhyIOTU clusters recapitulate PID clusters



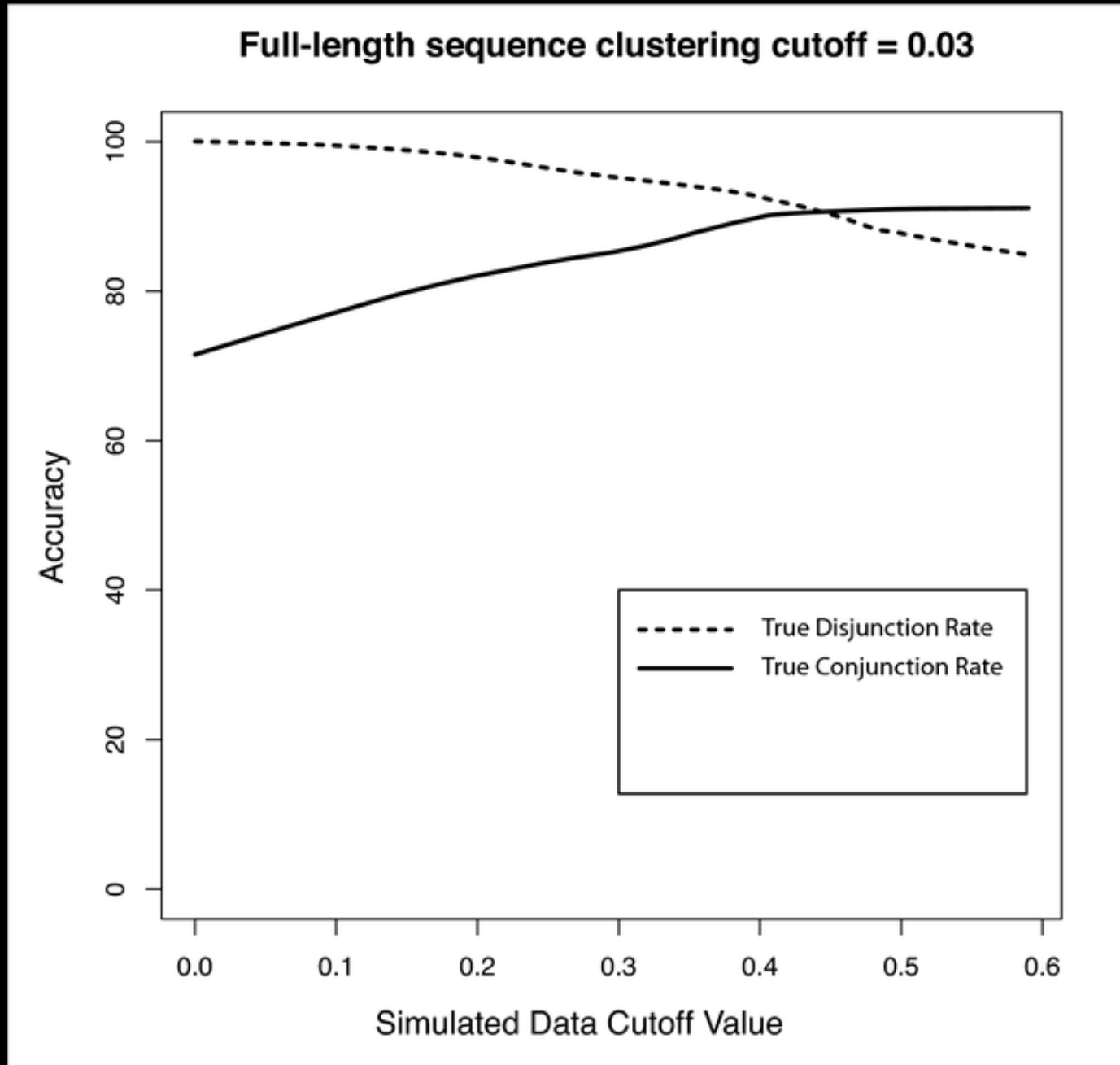
PhylOTU clusters recapitulate PID clusters



PhyIOTU clusters recapitulate PID clusters



Accuracy of Clustering Reads



Accuracy of Clustering Reads

