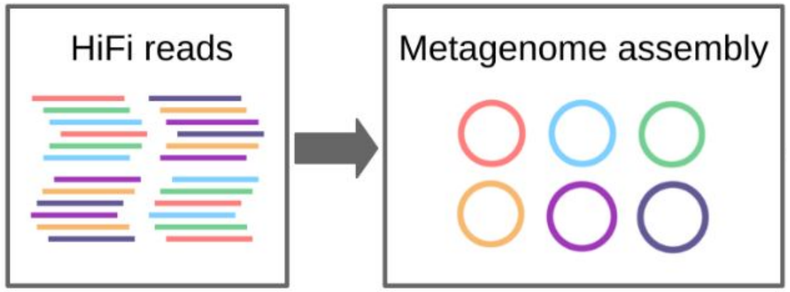
Analyzing rare mutations in metagenomes assembled using long and accurate reads

Marcus W. Fedarko, Mikhail Kolmogorov, Pavel A. Pevzner

Metagenome assembly

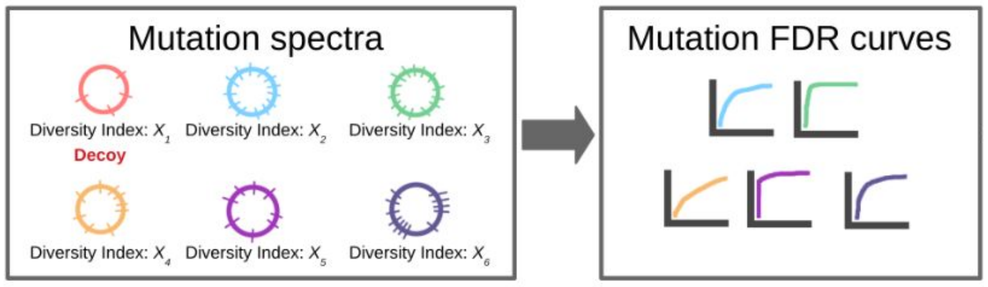
A microbiome is a community of microbes. A metagenome is the collection of these microbes' genomes.

Researchers often study which microbes are present in a microbiome using DNA sequencing.



DNA sequencers produce reads of DNA that are much shorter than most microbes' genomes. Even with long and accurate reads, we need to assemble these reads to construct our best guess of the genomes of the microbes.

Metagenome-assembled genomes (MAGs) are usually chimeras of multiple closely related genomes.

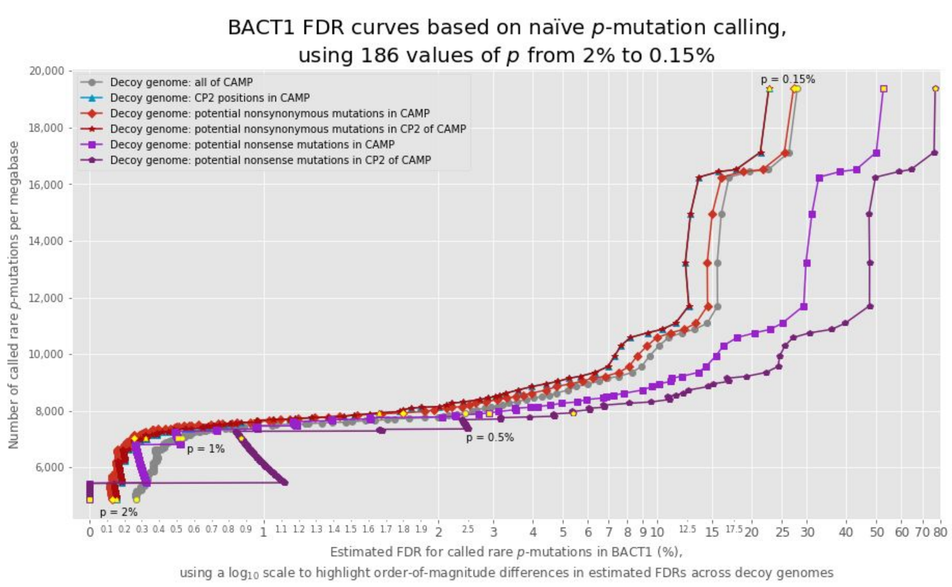


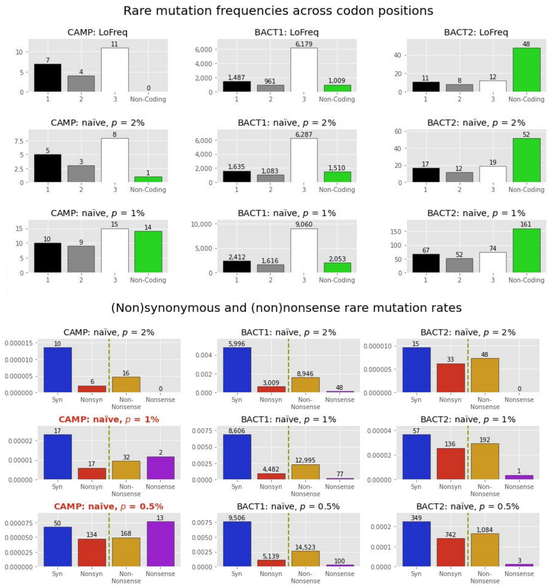
We want to figure out the exact strain-level genomes present in a MAG. To do this we need to be able to confidently identify mutations, especially rare mutations, in a MAG.

False Discovery Rate (FDR) estimation of mutation identification

Given a set of identified mutations in a MAG, what fraction of them are incorrect? To estimate this FDR for an arbitrary "target" MAG, we borrow the target-decoy approach from proteomics: select a decoy MAG, compute its mutation rate, and then estimate the FDR as the ratio of the decoy mutation rate to the target mutation rate.

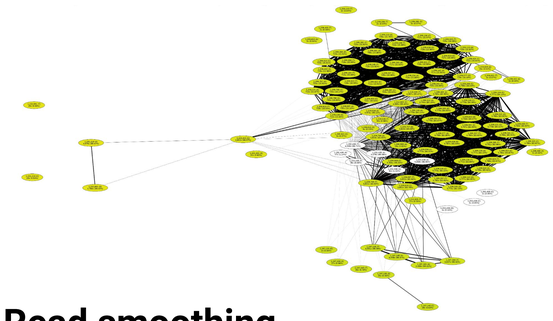
We also introduce the context-dependent target-decoy approach, exploiting the fact that we know certain types of mutations are less likely to occur than others.



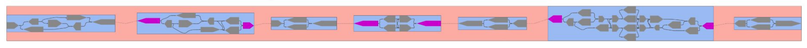


Phasing analyses

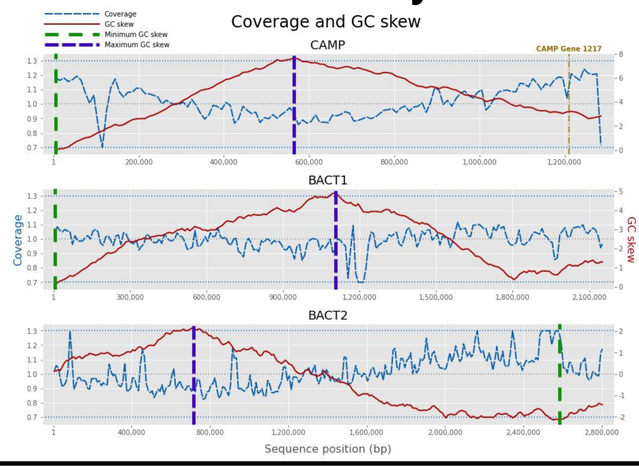
Link graphs



Read smoothing



Growth dynamics



Acknowledgements

UCSDCSE Computer Science and Engineering
UC San Diego

Nicholas Bokulich, Lee Katz, Kai Blin, Heng Li, Anton Bankevich, Alex Richter

github.com/fedarko/sheepgut
github.com/fedarko/strainFlye