

How test to see if aligner is working?

How test to see if aligner is working?

Simulations

- Create genome
 - emulate characteristics of real genomes or use a real genome
- Create virtual reads
 - randomly generate 300 bp reads
 - with or without errors to see how those affect accuracy
- Assemble virtual reads into genome and compare with “ground truth”

What does this system miss?

Simulations are only as good as the simulated data

- Optimize for the data you create
- Organisms with unusual genomic structure might not align well
- Not everyone using the same simulated data to validate

How address these issues?

Assemblathon!

- Assemblathon 3.0 is currently ongoing.
- Data generated by the 10K genome consortium on a new genome
- Each team assembles from the data
- Assemblies are compared and contrasted

Any questions about genome
assembly?