How good are our genomes? – developing quality scores for 32,000 genomes based on second and third generation sequencing and genomic analysis tools

Background:

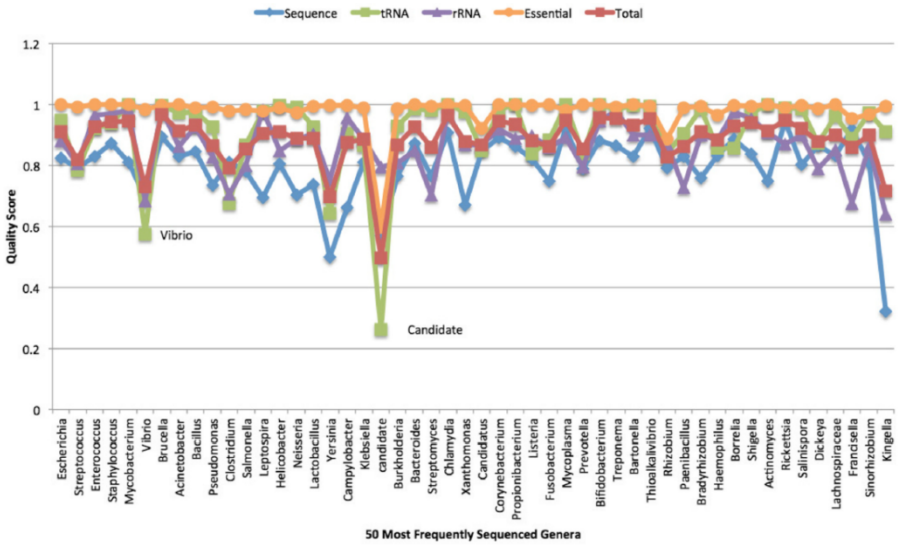
- More than 80% of the microbial genomes in GenBank and three other major public databases, are of 'draft quality'.
- In this study available microbial DNA sequences were examined for complete, draft and sequence read archive genomes.

Approach:

- Scores were assigned using four categories: the completeness of the assembly, the presence of full-length rRNA genes, tRNA composition and the presence of a set of 102 conserved genes in prokaryotes.
- Most (~88%) of the genomes had quality scores of 0.8 or better and can be safely used for standard comparative genomics analysis.

Outcomes:

- The score can be used to set thresholds for screening data when analyzing publicly available genomes and reference data is either not available or not applicable.
- The scores highlighted organisms for which commonly used tools do not perform well.
- With few exceptions, most of the 30,000 genomes have nearly all the 102 essential genes.



Quality scores for 50 most abundant genera. Average quality scores for sequence, tRNAs, rRNAs, essential genes, and total plotted for each of the 50 most represented genera. The genera are presented in order of abundance from *Escherichia* on the left to *Kingella* on the right.

Significance:

- This information can be used to improve tools and to serve computational biologists as more diverse organisms are sequenced.
- The BioEnergy Science Center (BESC) and the Plant Microbe Interface (PMI) projects are using novel "non-model" microbes. These approaches allow these projects to judge confidence in genomes and genomic tools as they study more diverse microbes. BESC is using consolidated bioprocessing microbes and PMI is investigating non-model microbes associated with plants. .

Land, M.L., Hyatt, D., Jun, S-R., Kora, G., Hauser, L.J., Lukjancenko, O., Ussery, D.W., "Quality scores for 32,000 genomes," *Standards in Genomic Sciences 9(20): 1-10*

OAK RIDGE National Laboratory
Managed by UT-Battelle for the Department of Energy