



Insights from 20 years of bacterial genome sequencing

Contact: Miriam Land (865241-6231, landml@ornl.gov)

Funding Source: DOE Office of Biological and Environmental Research, Genomic Science Program

Background

- In the two decades since the first bacterial genome was completely sequenced, the technical improvements and subsequent increases in biological knowledge have been dramatic. Second and third generation sequencing have greatly reduced the cost of sequencing.

Approach

- A review was done on the progress made in sequencing & analyzing microbial genomes in the last 20 years.

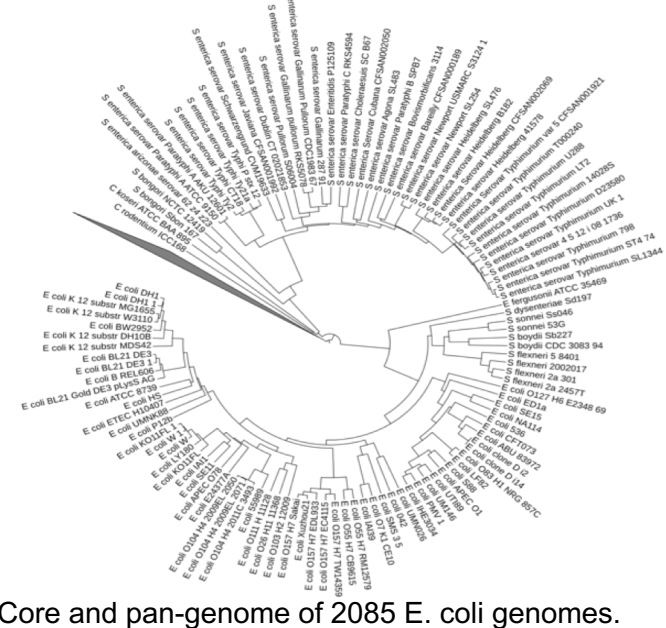
Outcomes

- The availability of large number of genomes from the same species has created opportunities for in-depth analysis of the species, including many organisms important for plant-microbe interactions. Phylogenetic trees can be done with finer detail and the concept of pan and core genomes has expanded our concept of species.

Significance

- As demonstrated by the *Escherichia coli* genome, even with thousands of sequenced genomes within a same species, the core set of genes levels off quickly, at about 3100 gene families. However, the pan-genome continues to grow. There are more than 100,000 *E. coli* gene families, which is larger than the human genome

A branching pattern of *E. coli* and *Shigella* on an alignment-free whole proteome phylogeny.



Core and pan-genome of 2085 *E. coli* genomes.

