**Construct species phylogenic tree for *Saccharomyces sensu stricto* using BUSCO genes**

**Jing Li**

1. **Background**

Phylogenic relationship is very important when we research on gene origin or some other evolutionary questions. Currently, there are seven Saccharomyces species, S. cerevisiae, S. paradoxus, S. mikatae, S. kudriavzeii, S. arboricolus, S. uravum, and S. eubayanus, with good assembly genomes. However, their phylogenic relationship is not clear. Different papers present different phylogenic tree for *Saccharomyces sensu stricto* species.

BUSCO, also known as assessing genome assembly and annotation completeness with Benchmarking Universal Single-Copy Orthologs. It provides all the ancient or common gene orthologs in fungi.

1. **Research objective**

My purpose of this project is that construct a species phylogenic tree from all the BUSCO gene trees. Then compare to the current paper, to verify how the species phylogenic tree from BUSCO genes close to other research.

1. **Method**
2. **Data collection**

Collect all the genome sequence of these seven *Saccharomyces sensu stricto* species from NCBI database. Collect all the BUSCO orthologs dataset for fungi from BUSCO web.

1. **Predict conserved genes**

Predict fungi conserved genes in seven *Saccharomyces* species through BUSCO software. BUSCO map the fungi conserved gene dataset to the genome and then predict the conserved gene and protein sequence for each species. We can also check whether the genome assembly quality though the number of conserved genes in each genome.

1. **Gene alignment**

Extract all the conserved genes for each species, and then do protein sequence alignment for each gene in the seven species.

1. **Gene tree construction**

Construct gene trees for each gene with maximum likelihood approach using RAxML program. Using bootstrap analysis and search for the best-scoring ML tree.

1. **Species tree construction**

Combine all the gene trees and then generate a species tree using the ASTRAL coalescent-based species tree estimation program.

Github link:

https://github.com/lijing28101/EEOB563\_final