Gene-based pan-genome analysis of cucumber

Rocky Shao1, Xuebo Zhao2, Zhangjun Fei2,3

1Ithaca High School, 1401 N Cayuga St, Ithaca, NY 14850, USA

2Boyce Thompson Institute, Ithaca, NY 14853, USA

3USDA-ARS Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

# **Introduction**

Global challenges such as climate change and population growth have intensified the need for more robust crop varieties. Advances in DNA sequencing technologies have made it easier to generate pangenomes from representative individuals of target species. Pangenomic studies have also helped identifying beneficial traits and provided insight into breeding efforts to develop more robust crop varieties (Schreiber et al., 2024). These studies have demonstrated success in recent plant research, uncovering evolution and diversity of different plant species (Tang et al., 2022) and identifying deleterious mutations to facilitate breeding (Wu et al., 2023). Cucumber (Cucumis sativus L.) is not only one of the most important vegetable crops worldwide, but also serves as a model plant for sex determination and vascular biology studies (Li et al., 2019). Previous studies have shown patterns in cucumber domestication, such as reduction in genetic diversity (Qi et al., 2013). These studies also highlighted key genes associated with important agronomic traits and the domestication process, providing valuable information for future breeding programs (Li et al., 2022). To capture the entire genomic diversity for future biological research, it is needed to construct a complete gene-based pan-genome from diverse cucumber accessions.

**Research Question**

This project aims to construct a gene-based pan-genome and identify core and pan genes of 39 different cucumber species. Additionally, we seek to identify candidate genes associated with traits such as stress resistance, nutritional enhancement, and yield increment through PAV (presence/absence variation) analysis. These findings will deepen our understanding of cucumber genetics, provide valuable resources for future cucumber breeding.

**Plans to Solve the Research Question**

In this project, we conduct pangenome study among 39 distinct cucumber species to identify genes linked to key agronomic traits by systematically analyzing their full proteome sequences. The pan and core genes of the cucumber species dataset will be identified using OrthoFinder (Emms et al., 2019). Then the PAV patterns of candidate genes across the genomes will be derived using a gene-based approach (He et al., 2023). Putative functions of these genes will then be annotated using bioinformatic tools including Blast2Go (Conesa et al., 2008). The final step is to conduct enrichment analysis of the identified candidate genes involved in important biological processes or pathways (Li et al., 2023).

**Anticipated Results**

Pan and core genes of cucumber orthogroups are expected to be classified. Additionally, this project aims to identify enrichment patterns of candidate genes involved in important agronomic traits, particularly stress tolerance, nutritional enhancement and yield improvement.

# **References:**

Conesa, A., et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.

Tang, D., et al. (2022). Genome evolution and diversity of wild and cultivated potatoes. *Nature*. 606, 535-541

Li, H., et al. (2022). Graph-based pan-genome reveals structural and sequence variations related to agronomic traits and domestication in cucumber. *Nature Communications*. 13, 682

Qi, J., et al. (2013). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nature Genetics*. 45, 1510-1515

Emms, D. M., Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*. 20, 238

Schreiber, M., et al. (2024). Plant pangenomes for crop improvement, biodiversity and evolution. *Nature Reviews Genetics*.

Li, N., et al. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nature Genetics*. 55, 852-860

He, Q., et al. (2023). A graph-based genome and pan-genome variation of the model plant Setaria. *Nature Genetics*. 55, 1232-1242

Li, Q., et al. (2019). A chromosome-scale genome assembly of cucumber. *GigaScience*. 8, giz072

Wu, Y., et al. (2023). Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell*. 186, 2313-2328.e15