Leveraging comparative population genomics to dissect the mechanisms of multi-stress tolerance of *I. orientalis*

Yusuke Sasaki[1,2](yusukesasaki@lbl.gov), Ping-Hung Hsieh[1,2], Jing Ke[1,2], Zhying Zhao[1,2], Zong-Yen Wu[1,2], Andrei Stecca[1,2], Sajeet Haridas[1,2], Zia Fatma[3], Huimin Zhao[3], and **Yasuo Yoshikuni**[1,2]

[1]Department of Energy Joint Genome Institute, [2]Lawrence Berkeley National Laboratory, [3]University of Illinois at Urbana-Champaign

Project Goals: Dissect phenotypic variations of 170 *I. orientalis* strains under major stress conditions through statistical analysis

Abstract:

Issatchenkia orientalis has a distinguished capability of tolerance to multiple environmental factors (e.g., low pH, heat, and chemicals consisting of lignocellulosic components) and antifungal drugs (e.g., azoles, echinocandins, and polyenes). These unique features have attracted attention to *I. orientalis* to be utilized as a next generation industrial chassis capable of producing biofuels and bioproducts from crude lignocellulosic hydrolysates. Large-scale population genomic studies will provide comprehensive views of *I. orientalis* genetic diversity, population structure, and genotype-phenotype associations that open further opportunities such as, rational strain engineering and evolutionary lineage analysis of the yeast. In contrast to the previous population genomics study of *I. orientalis* that sequenced 32 isolates (Douglass et al., 2018), we performed deep coverage genome sequencing of 170 isolates and assessed their phenotypic traits under 58 conditions. Genome sequencing of 170 *I. orientalis* isolates collected from various habitats identified 305,405 single nucleotide polymorphism (SNPs), 16,177 insertions and deletions (INDELs), and other genetic variations (e.g., ploidy, gene copy number, and pan-genome). Phenotypes under major stress conditions (9,338 variations) were quantitatively assessed by measuring cell growth changes with an automation-based high-throughput assay. We have completed the phenotypic assays under 58 conditions and collected 9,338 traits. Correlation analysis of the data shows that significant correlations with several types of azole drugs. Clustering and community detections allows to categorize each condition and strain based on the phenotypic traits, which provides a new insight of strain classification. As our future study, genome-wide association study (GWAS) allows to capture genetic variants shaping the architecture of the phenotypic traits and genetic engineering of *I. orientalis* for validation of the variants identified by GWAS proves the concept of our strategy.

References:

A.P. Douglass, B. Offei, S. Braun-Galleani, A.Y. Coughlan, A.A.R. Martos, R.A. Ortiz Merino, K.P. Byrne, K.H. Wolfe, "Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names" PLoS Pathog., 14 (2018), p. e1007138

Funding Statement:

This work was funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-FOA-0001650. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the U.S. Department of Energy.