**Bioinformatical assessment of loss of heterozygosity events in *C. albicans* filamentous growth mutants**

**Emily Hovan, Class of 2021**

 *Candida albicans* is an opportunistic pathogenic yeast residing throughout the human body, often localizing in the gastrointestinal tract, vagina, and oral cavities. Generally, *C. albicans* is a harmless commensal fungus. However, environmental factors such as temperatures of 37ºC, the amino sugar N-acetylglucosamine, neutral pH, 5% CO2, and microaerophilic conditions can prompt *C. albicans* to transition from its innocuous yeast form into a virulent hyphal form (1). Hyphal formation enables locomotion, thus propelling the fungus throughout its host (2). This filamentous growth permits the fungus to invade epithelial and endothelial cells’ walls and cause illnesses such as oral thrush, vaginal infections, and deep-tissue infections in immunocompromised individuals. One factor believed to influence *C. albicans’* pathogenesis is its genetic plasticity. Having a diploid genome with 14.4 megabases organized into 8 chromosomes, *C. albicans* has notable phenotypic and genetic diversity(3). Chromosomal rearrangements often arise in response to stresses such as heat shock, host-pathogen interactions, and presence of antifungal drugs (3). Genetic modifications in *C. albicans* can manifest as loss of heterozygosity, polymorphisms, recombination, chromosomal inversions, and chromosomal aneuploidies (3). Thus, phenotypic variations and genetic diversity may play a key role in influencing *C. albicans’* commensalism and pathogenicity. This summer, I worked alongside Professor Forche to analyze loss of heterozygosity (LOH) chromosomal events across two mutant strains of *C. albicans*. We applied a bioinformatical approach to verify loss of heterozygosity events while analyzing whole genome sequence data.

To study LOH events, the Broad Institute Integrative Genomics Viewer (IGV) and the Candida Genome Database were utilized (4,5). We employed a 2-step process to determine LOH events: a comparative analysis between two *C. albicans* strains in Excel and visual confirmation of data in IGV. Previously, genome sequence data was processed and returned all loci differing from their parental strain, AF7. I studied this processed genome sequence data for all 8 of *C. albicans’* chromosomes in Excel and selected loci that indicated homozygosity across both strains. If both strains at a locus appeared homozygous for the same allele, I deemed them a LOH event. However, the whole genome sequence data needed to be verified in IGV, since the Excel data was not always reliable. Inconsistencies appeared between the Excel data and the data in IGV because the IGV data represent the raw data and the excel data represent the data after the application of normalization and further bioinformatic analysis specific for LOH. For example, there were instances where a locus appeared homozygous in the Excel spreadsheet, but was actually heterozygous upon confirmation in IGV. Additionally, sometimes the Excel data suggested that loci were homozygous, but upon analysis in IGV this assessment proved unreliable because of a low total-count of bases. Therefore, it proved essential to confirm LOH events in IGV manually. To eliminate data due to low sequence coverage, we filtered out all loci with a total count (read depth, allele A+B) of 10. Once these events were confirmed, I screened each locus in the Candida Genome Database to determine if LOH had occurred in an open reading frame (ORF) and collected information on gene function. Finally, I organized the confirmed LOH events and ORF descriptions in a new spreadsheet that may be referenced for future experiments.

 Studying genetic instability of *C. albicans* may help us comprehend the mechanisms behind its pathogenesis. Discovering the genotypic and phenotypic modifications that arise in *C. albicans* in response to environmental stresses can provide important insight into how infections develop, and in turn, how we can prevent them. Our research highlights the importance of verifying true LOH events. The processed whole genome sequence data did not always accurately indicate LOH, because it lacked total base count information and was revealed as heterozygous in IGV. Therefore, it was important to employ a secondary analysis method, such as IGV. Furthermore, our research can be applied to future experiments. For example, to confirm the mutations identified, we will utilize Sanger sequencing. We then will introduce candidate mutations in the parental strain AF7 to see if the filamentous growth phenotype of the mutant strain can be recapitulated. Ultimately, we hope our bioinformatical assessment of loss of heterozygosity events in *C. albicans* mutants can provide important framework for future experiments.

**Faculty Mentor: Anja Forche**

**Funded by the Life Sciences Summer Fellowship**

References

1. 1. Sharkey L.L., McNemar M.D., Saporito-Irwin S.M., Sypherd P.S., Fonzi W.A. 1999. *HWP1* functions in the morphological development of *Candida albicans* downstream of *EFG1*, *TUP1*, and *RBF1*. J Bacteriol. **181**(17):5273-5279
2. Koehler J.R., Hube B., Puccia R., Casadevall A., Perfect J.R. 2017. Fungi that infect humans. Microbiol Spectrum **5**(3):1-29.
3. Dadar M., Tiwari R., Karthik K., Chakraborty S., Shahali Y., Dhama K. 2018. *Candida albicans*—Biology, molecular characterization, pathogenicity, and advances in diagnosis and control—An update. Microb. Pathog. **117**(2018):128-138.
4. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov.[**Integrative Genomics Viewer**. Nature Biotechnology 29, 24–26 (2011)](http://www.nature.com/nbt/journal/v29/n1/abs/nbt.1754.html). (Free PMC article [here](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3346182/)).
5. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G (2017). The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res* **45** (D1); D592-D596.