Using BLAST to Detect Horizontal Gene Transfer in Pathogenic Fungi

Christopher Good
Faculty Advisor: Theresa Grana, Ph.D.

Introduction and Objectives

Similarly to antibiotic resistance, antifungal resistance is a growing challenge for clinicians. Mechanistically, antibiotic resistance is acquired through lateral or horizontal gene transfer (HGT), as well as through genetic mutations (Huddleston, 2014). Although associated more with prokaryotes, past studies show limited evidence of HGT in eukaryotes: between a Candida and another yeast species and between Candida yeast and bacteria (Fitzpatrick, 2012), warranting additional comparative, genomic and proteomic research on the evolutionary forces behind fungal virulence.

This honors capstone project aimed to use the NCBI's Basic Local Alignment Search Tool (BLAST) to quickly and statistically compare existing biological sequence data (Altschul, *et al*. 1990; NCBI) in conjunction with EMBL Multiple Sequence Comparison by Log-Expectation (MUSCLE) alignment (Madeira, *et al*. 2019) and Molecular Evolutionary Genetics Analysis (MEGA) to visualize evolutionary relationships with incorporated bootstrapping (Kumar, *et al*. 2018). Thus, the objective of this project was to use bioinformatics tools to identify potential instances of Horizontal Gene Transfer (HGT) between pathogenic yeast and viruses, specifically HGT as an evolutionary mechanism for antifungal resistance gene (ARG) acquisition.

Methods

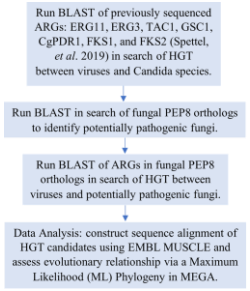
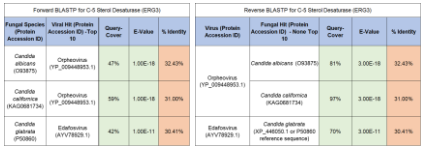


Figure 1. **Methodology flowchart used in the study:** BLAST cutoffs were ≥ 40% query coverage, E-value ≤ 10-5, and ≥80% identity based on previous studies (Lekunberri, *et al*. 2017; Rolland, *et al*. 2009). Tables 1 and 2 show respective forward (fungal → viral) and reverse (viral → fungal) BLAST results for ERG3P. Tables 3 and 4 also show respective forward (algal → viral) and reverse (viral → algal) BLAST results for ERG3P. ML Phylogeny constructed in MEGA and MUSCLE sequence alignment of the ERG3P active site are shown in Figure 2.

Results

Tables 1 and 2. **BLASTP Evidence for Fungal HGT:** Forward (fungal → viral) and reverse (viral → fungal) ERG3P results.



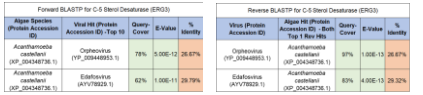
Forward BLASTP for C-5 Sterol Desaturase (ERG3)

Fungal Species (Protein Accession ID)	Viral Hit (Protein Accession ID) -Top 10	Query-Cover	E-Value	% Identity
Candida albicans (O93875)	Orpheovirus (YP_009448853.1)	47%	1.00E-18	32.43%
Candida californica (KAG0881734)	Orpheovirus (YP_009448853.1)	59%	1.00E-18	31.00%
Candida glabrata (P50860)	Edafosvirus (AYV78929.1)	42%	1.00E-11	30.41%

Reverse BLASTP for C-5 Sterol Desaturase (ERG3)

Virus (Protein Accession ID)	Fungal Hit (Protein Accession ID) - None Top 10	Query-Cover	E-Value	% Identity
Orpheovirus (YP_009448853.1)	Candida albicans (O93875)	81%	3.00E-18	32.43%
	Candida californica (KAG0881734)	97%	3.00E-18	31.00%
Edafosvirus (AYV78929.1)	Candida glabrata (XP_449350.1 or P50860 reference sequence)	70%	3.00E-11	30.41%

Tables 3 and 4. **BLASTP Evidence for Algal HGT:** Forward (algal → viral) and reverse (viral → algal) ERG3P results.



Forward BLASTP for C-5 Sterol Desaturase (ERG3)

Algae Species (Protein Accession ID)	Viral Hit (Protein Accession ID) -Top 10	Query-Cover	E-Value	% Identity
Acanthamoeba castellanii (XP_004348736.1)	Orpheovirus (YP_009448853.1)	78%	5.00E-12	26.67%
Acanthamoeba castellanii (XP_004348736.1)	Edafosvirus (AYV78929.1)	62%	1.00E-11	29.79%

Reverse BLASTP for C-5 Sterol Desaturase (ERG3)

Virus (Protein Accession ID)	Algae Hit (Protein Accession ID) - Both Top 1 Rev Hits	Query-Cover	E-Value	% Identity
Orpheovirus (YP_009448853.1)	Acanthamoeba castellanii (XP_004348736.1)	97%	1.00E-13	26.67%
Edafosvirus (AYV78929.1)	Acanthamoeba castellanii (XP_004348736.1)	83%	4.00E-13	29.32%

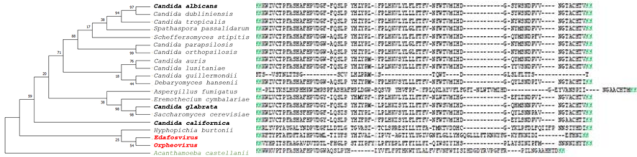


Figure 2. **ML Phylogeny showing relatedeness of ERG3P in Fungi and Viruses of Interest:** Tree with the highest log likelihood (-10678.16) inferred by MEGA using a JTT model of substitution. The numbers shown at internal nodes indicate the percentage of trees with the same taxa grouped together. Proximate to each species name is the determined sequence of the ERG3P active site motif (Kamthan, *et al*. 2017), which was collectively aligned in EMBL MUSCLE for further, visual analysis of the evolutionary history.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.
Andreani, J., Khalil, J., Baptiste, E., Hasni, I., Michelle, C., Raoult, D., Levasseur, A., and La Scola, B. 2018. Orpheovirus IHUMI-LCC: A New Virus among the Giant Viruses. *Frontiers in microbiology* 8: 2643.
Farr, D.F. and Rossman, A.Y. Fungal Databases, *U.S. National Fungus Collections*, Agricultural Research Service, United States Department of Agriculture
Fitzpatrick, D.A. 2012. Horizontal gene transfer in fungi. *FEMS Microbiology Letters* 329: 1–8.
Huddleston, J.R. 2014. Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes. *Infection and drug resistance* 7: 167–176.
Kamthan, A., Kamthan, M., and Datta, A. 2017. Expression of C-5 sterol desaturase from an edible mushroom in fission yeast enhances its ethanol and thermotolerance. *PloS one* 12
Köhler, J.R., Casadevall, A., and Perfect, J. 2014. The spectrum of fungi that infects humans. *Cold Spring Harbor perspectives in medicine* 5.
Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549
Lekunberri, I., Subirats, J., Borrego, C.M., and Balcázar, J.L. 2017. Exploring the contribution of bacteriophages to antibiotic resistance. *Environmental Pollution* 220: 981-984.
Madeira, F., Park, Y.M., Lee, J., *et al*. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47.
National Center for Biotechnology Information (NCBI). Basic Local Alignment Search Tool. *National Library of Medicine* (US), Bethesda, MD.
Rolland, T., Neuvéglise, C., Sacerdot, C., and Dujon, B. 2009. Insertion of Horizontally Transferred Genes within Conserved Syntenic Regions of Yeast Genome. *PloS one*
Schulz, F., Alteio, L., Goudeau, D., Ryan, E.M., Yu, F.B., Malmstrom, R.R., Blanchard, J., and Woyke, T. 2018. Hidden diversity of soil giant viruses. *Nature Communications* 1-9.
Thomas, G., Bain, J.M., Budge, S., Brown, A., and Ames, R.M. 2020. Identifying Candida albicans Gene Networks Involved in Pathogenicity. *Frontiers in genetics* 11: 375

Discussion

BLAST showed more support for ERG3 HGT between edafosvirus and *C. glabrata* and between orpheovirus and *C. albicans* and *C. californica*, respectively. Since both viruses are "giant viruses" known to infect algae, BLAST was used to analyze HGT in their confirmed algal host species (Andreani, *et al*. 2017; Schulz, *et al*. 2018) and to produce an outgroup in the ML phylogeny. BLAST showed support for ERG3 HGT between both viruses and *A. castellanii*. A limitation in this study was that only the top 10 forward BLAST (viral) hits were searched for reverse BLAST. However, this served as a necessary data cutoff.

However, the ML phylogeny contradicts the BLAST results and shows more support for ERG3 HGT between both viruses and *H. burtonii*, another fungal species that is not known to be pathogenic to humans (Farr and Rossman). This association could be due to the shorter sequence length of ERG3P (≈ 370 amino acids). However, BLAST did not support ERG3 HGT between this fungal species and the viruses of interest.

Conclusions and Acknowledgements

In conclusion, while BLAST showed limited evidence for ERG3 HGT between three Candida species and two viruses, the ML phylogeny fails to support this evolutionary event for ERG3P. For the purposes of HGT, BLAST might be better suited to certain organisms (i.e. prokaryotes) and its use should be reinforced as a non-definitive predictor of such evolutionary events.

Considering that < 30% of *C. albicans* genes are characterized (Thomas, *et al*. 2020) and this value is likely less for other Candida species, this suggests a greater understanding of numerous uncharacterized genes could reveal new evidence of HGT in fungi. One should also consider the hypothesis that pathogenic fungi evolved to target amoeba before more complex eukaryotes (Köhler, *et al*. 2015), which may indirectly explain the observed BLAST results for ERG3 HGT between *C. castellanii* and both viruses.

I would like to thank Vinayak Mathur, Ph.D. at Cabrini University for his expert guidance, as well as the Department of Biological Sciences and Honors Program at the University of Mary Washington.