

THE EFFECTS OF HORIZONTAL GENE TRANSFER ON THE EVOLUTION OF LAND PLANTS

A Signature Honors Project Presented to the
Honors College

East Carolina University

A Thesis Submitted
in Partial Fulfillment of the Requirements for
Graduation with Honors

Memphis C. Jandron

Greenville, NC
May 2023

Approved by:
Jinling Huang, Ph.D.
Department of Biology

THE EFFECTS OF HORIZONTAL GENE TRANSFER ON THE EVOLUTION OF LAND PLANTS

Memphis C. Jandron

A thesis submitted to the Department of Biology, East Carolina University, in partial
fulfillment of the requirements for Biology Honors Thesis

Advisor: Jinling Huang, Ph.D.

East Carolina University

Department of Biology

May 2023

I hereby declare I am the sole author of this thesis. It is the result of my own work and is not the outcome of work done in collaboration, nor has any of it been submitted elsewhere for another degree.

Signed: Memphis C. Jandron Date: 4/25/2023

Memphis C. Jandron

THE EFFECTS OF HORIZONTAL GENE TRANSFER ON THE EVOLUTION OF LAND PLANTS

Memphis C. Jandron

Department of Biology, East Carolina University, Greenville N.C. 27858-4353, U.S.A

ABSTRACT

This study sought to understand how horizontal gene transfer (HGT) has affected the evolution of land plants. Land plants evolved from charophyte green algae. A list of candidate genes derived from HGT was generated from the genome of the charophyte green alga *Chlorokybus atmophyticus* and were subjected to detailed phylogenetic analyses to investigate their evolutionary origin. The flowering plant *Arabidopsis thaliana* was used as a reference organism to understand the functions of the investigated genes, which in turn were used to assess the contribution of acquired genes to the evolution and adaptation of plants. I performed a detailed investigation on three acquired genes, including *Chrsp45S09062* in the cupredoxin superfamily that encodes a multicopper oxidase protein, *Chrsp52S06587* that encodes a calcium-dependent phosphotriesterase superfamily protein, and *Chrsp22S08802* that encodes a protein with histidine kinase 1 activity. BLASTP searches were performed against the NCBI non-redundant protein sequence database using the HGT candidate genes from *C. atmophyticus*. Identified homologs were sampled for each candidate gene from representative groups of both prokaryotes and eukaryotes and then input into the computer program MEGA to perform a multiple sequence alignment (MSA). A phylogenetic tree was generated for each gene using MEGA and the resulting MSA. The phylogenetic trees generated were analyzed to identify true cases of HGT, and biochemical functions and related biological processes were further researched on each acquired gene to determine the effects of HGT on plant evolution. In the

evolution of plants, the expected impact of HGT is a mechanism that aids in the exchange of genetic information to benefit the adaptation of land plants to their environment.

ACKNOWLEDGEMENT

I want to thank Dr. Jinling Huang for his assistance and guidance in the completion of my Signature Honors Project.

TABLE OF CONTENTS

ABSTRACT	3
BACKGROUND	9
RESEARCH GOAL	11
RESEARCH DESIGN	12
RESULTS AND DISCUSSION	19
LITERATURE CITED	25

LIST OF TABLES

Table 1. Results of <i>C. atmophyticus</i> . (<i>A. thaliana</i> information via TAIR)	24
---	----

LIST OF FIGURES

Figure 1. Flowchart of research design	15
Figure 2. Phylogenetic tree of <i>Chrsp22S08802</i>	16
Figure 3. Phylogenetic tree of <i>Chrsp52S06587</i>	17
Figure 4. Phylogenetic tree of <i>Chrsp45S09062</i>	18

BACKGROUND

Horizontal gene transfer (HGT) is the process of genetic exchange between organisms without sexual reproduction (Keeling, P. J. 2009). Although the idea of HGT is not new, the advent of genomic sequencing and the availability of many genomic sequence databases in the 21st century have revolutionized research in this field (Wickell D., Li FW. 2019). HGT is known to play a significant role in the evolution of prokaryotes, but recent studies have indicated a more extensive role of HGT in eukaryotes than previously thought. HGT has been observed between bacteria and animals, and there are limited examples of HGT from animals to bacteria, facilitated through endosymbiosis and free-living bacteria (Hotopp 2011). HGT can occur via direct contact through epiphyte, parasitism, and symbiosis, as well as through indirect contact using vectors such as pollen, fungi, viruses, and others (Gao C. 2014). *Agrobacterium* is one of the most extensively studied organisms for HGT, having transferred a tumor gene to the host-plant genome (Quispe-Huamanquispe D. et al. 2017). HGT can spread new and innovative genes essential to the fitness of organisms (Gao C. 2014), and its significance is heightened because transferred genes can have an immediate impact on the recipient taxon (Wickell D., Li FW. 2019).

HGT has been found to have played a significant role in the early evolution of land plants, with two major episodes of transfer events identified (Ma J. et al. 2022). The genes acquired via HGT during the early evolution of land plants have been retained by their descendent lineages to a considerable extent (Ma J. et al. 2022). Although HGT events often begin as neutral or slightly detrimental in all groups that can undergo HGT, the acquired genes can evolve to become beneficial for the organism receiving the transferred gene (Soucy, S. et al.

2015). Such events are more likely to occur between closely related species with similar genomes. HGT can introduce new genes or replace existing homologs (Keeling, P., Palmer, J. 2008). While the transferred genes often give the recipient organism a clear function (Keeling, P., Palmer, J. 2008), they are rarely truly neutral.

I conducted initial research on the identification of acquired genes in the charophyte alga *Chlorokybus atmophyticus*, which were investigated further in the context of HGT and the evolution of land plants. Charophytes are the closest relatives of land plants. My investigation focused on acquired genes that are conserved in the flowering plant *Arabidopsis thaliana* due to the wealth of previous research on its genome, gene functions, and expression. My initial findings reveal that *Chrsp45S09062* encodes a protein from the cupredoxin superfamily. Homologs of this acquired cupredoxin gene are predominantly expressed in the petioles and seed capsules of *A. thaliana*. *Chrsp52S06587* encodes a calcium-dependent phosphotriesterase superfamily, its homolog in *A. thaliana* mainly being expressed in different regions of the flower, and *Chrsp22S08802* encodes a histidine kinase family gene, its homolog in *A. thaliana* being mainly expressed in various parts of the leaf. I obtained this information from The Arabidopsis Information Resource (TAIR, www.arabidopsis.org), a comprehensive database containing genetic and molecular biology data for *A. thaliana*. By investigating the effects of HGT on the evolution of land plants using these genes as a model, I aim to gain a deeper understanding of the role of HGT in plant evolution.

RESEARCH GOAL

The goal of this study is to further understand the impact of HGT on land plant evolution by identifying genes of foreign origin in the charophyte alga *C. atmophyticus*. Functions and the biological processes in which these genes participate will be assessed.

RESEARCH DESIGN

In this investigation, the first step was to identify gene candidates from the genome of the charophyte alga *C. atmophyticus*. These candidates were selected based on a variety of parameters, including BLAST results in the NCBI non-redundant protein sequence database (nr), taxonomic distribution in the nr database and the OneKP database, and homologs in related taxa. From this list of candidates, their origins were determined and true cases of HGT were identified based on phylogenetic analyses. Three genes were chosen to further understand the importance of HGT in land plants: *Chrsp45S09062*, *Chrsp52S06587*, and *Chrsp22S08802*. The functions of these genes were investigated via previously published wet-lab investigations to gain a complete picture of their impact on the evolution from charophytes to land plants (Figure 1).

To perform a detailed investigation on the candidate HGT genes, the sequences of these genes were individually run through the BLASTP searches against NCBI nr and RefSeq databases. From the resulting genes that are significantly similar ($<.01$ e-value), 30-40 different homologs from representative groups of prokaryotes (bacteria and archaea) and eukaryotes were selected for a multiple sequence alignment (MSA) using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) alignment method. MEGA software was used for the MSA and phylogenetic tree construction. The MSA was completed by finding and adjusting misalignments, trimming the ends of the sequences, and cutting out large gaps shared by multiple sequences to ensure the quality of alignment. Maximum likelihood was used to build the phylogenetic tree, which finds the topology and branch lengths of the tree with the best probability of observing sequences.

To identify horizontally transferred genes, I analyzed the phylogenetic trees constructed (Figure 2) from each MSA based on the taxonomic distribution of the gene, the bootstrap values, and the relationship of the branches. The phylogenetic tree provides information on the evolutionary history of the relationships of the sampled sequences and the topology of each tree can help identify potential donors and recipients of the horizontally acquired gene. Bootstrap values can be influenced by a variety of factors such as sample size, evolutionary distance, and number of bootstrap replicates. Low bootstrap values were observed in the phylogenetic trees constructed for *Chrsp45S09062*, *Chrsp52S06587*, and *Chrsp22S08802* (Figures 2, 3, 4). The original gene candidate from *C. atmophyticus* and its homologs from major groups of land plants (e.g., bryophytes, lycophytes, ferns, and seed plants) were included in the tree to determine its relationships with homologs from other organisms. Genes that are likely transferred will be most closely related to homologs from their donors such as bacteria and specifically proteobacteria, instead of being to other primary photosynthetic eukaryotes (e.g., green algae and red algae). Because that chloroplasts are derived from cyanobacteria, if land plant and cyanobacterial sequences are closely related in the gene tree, the subject gene most likely evolved via gene transfer from the chloroplasts to the nucleus, a process that is commonly called intracellular gene transfer or endosymbiotic gene transfer. It is also possible, yet far less likely, that the gene may have transferred from bacteria to an ancestor of land plants, and the gene was then secondarily lost in the descendent land plant groups. Analyzing the MSA of a horizontally acquired gene across different species can provide supplemental evidence into the evolutionary history and origin of the gene. In the case of a horizontally acquired gene, the gene sequence is likely to show significant variation across different species,

as it has been acquired from a different organism rather than inherited from a common ancestor. In this investigation, the MSA did not provide strong enough evidence to make a confident conclusion; therefore, phylogenetic analysis will be the primary approach to for detecting HGT.

The analysis of gene distribution across different plant species and their monophyly can provide insight into the timing of important events in plant evolution, such as gene acquisition. If a particular gene is found in a monophyletic group of closely related species, it suggests that the gene was present in their common ancestor and has been inherited by all descendants of that ancestor. This indicates that the gene acquisition occurred before the divergence of the plant species in that group. Alternatively, if a gene is found in multiple species that are not closely related, it suggests that the gene was acquired by those species through horizontal gene transfer rather than through inheritance from a common ancestor. This indicates that the gene acquisition occurred after the divergence of those species from their common ancestor.

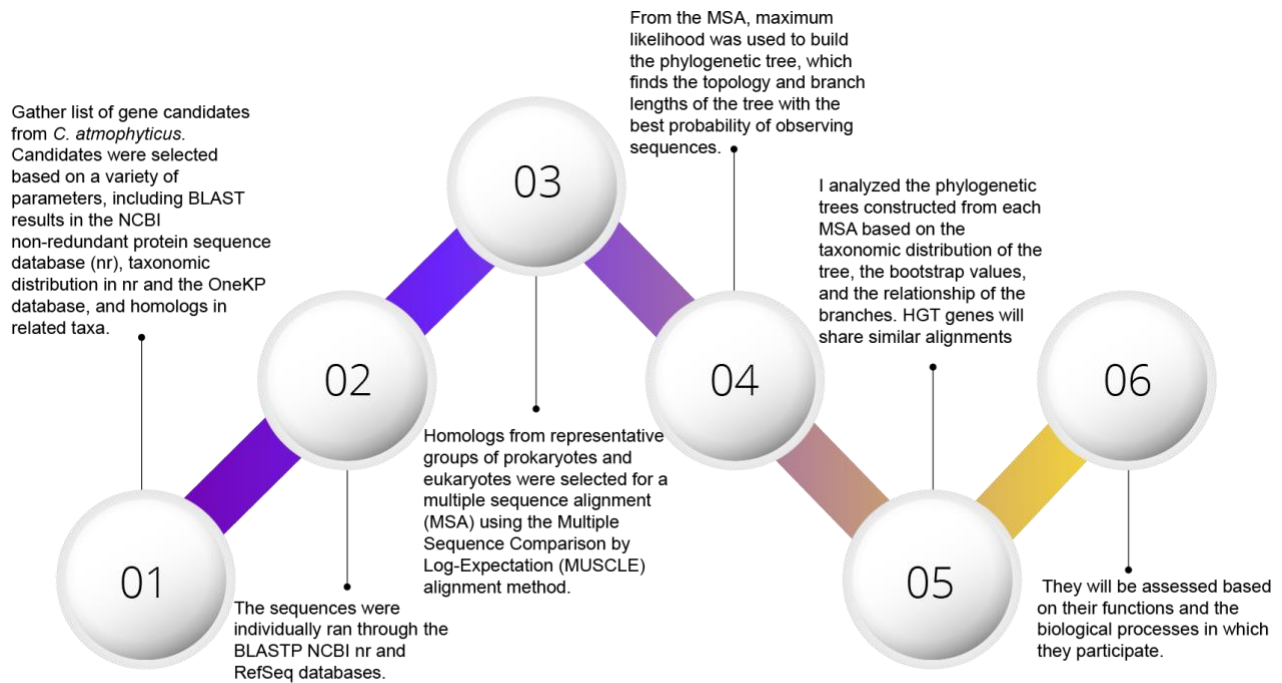


Figure 1. Flowchart of research design.

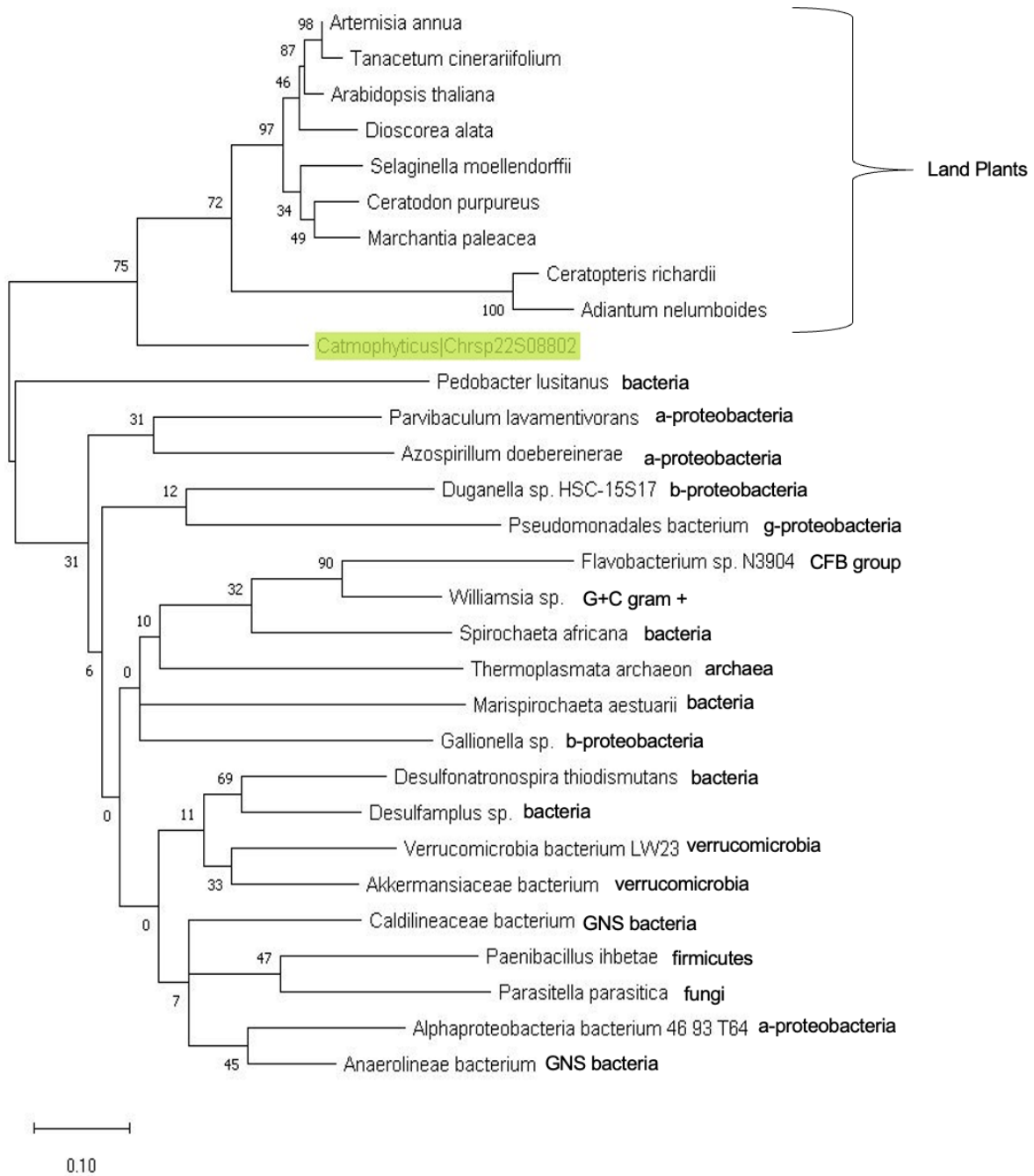


Figure 2. Phylogenetic tree *Chrsp22S08802*
 Highlighted – query sequence

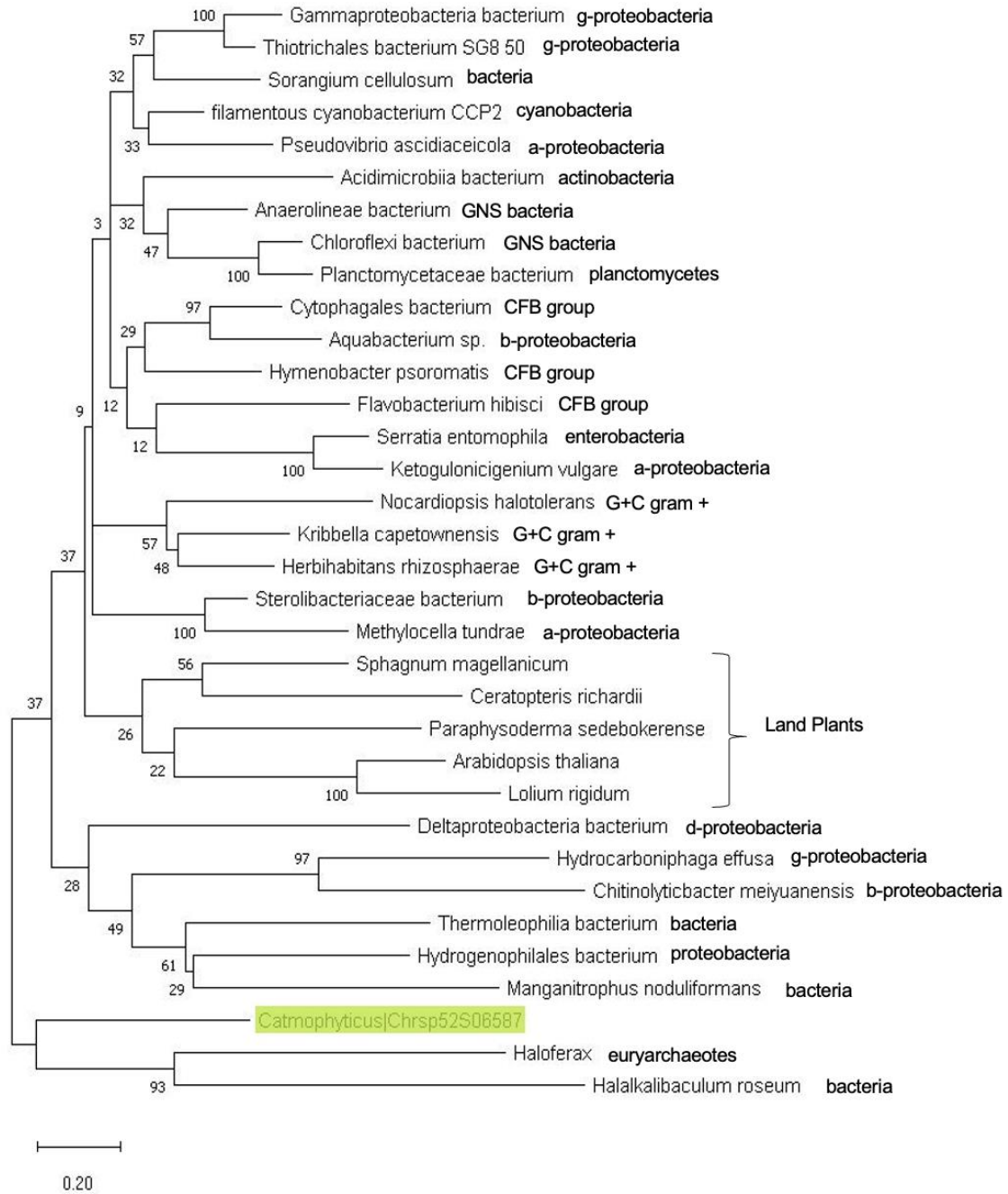


Figure 3. Phylogenetic tree of *Chrsp52S06587*

Highlighted – query sequence

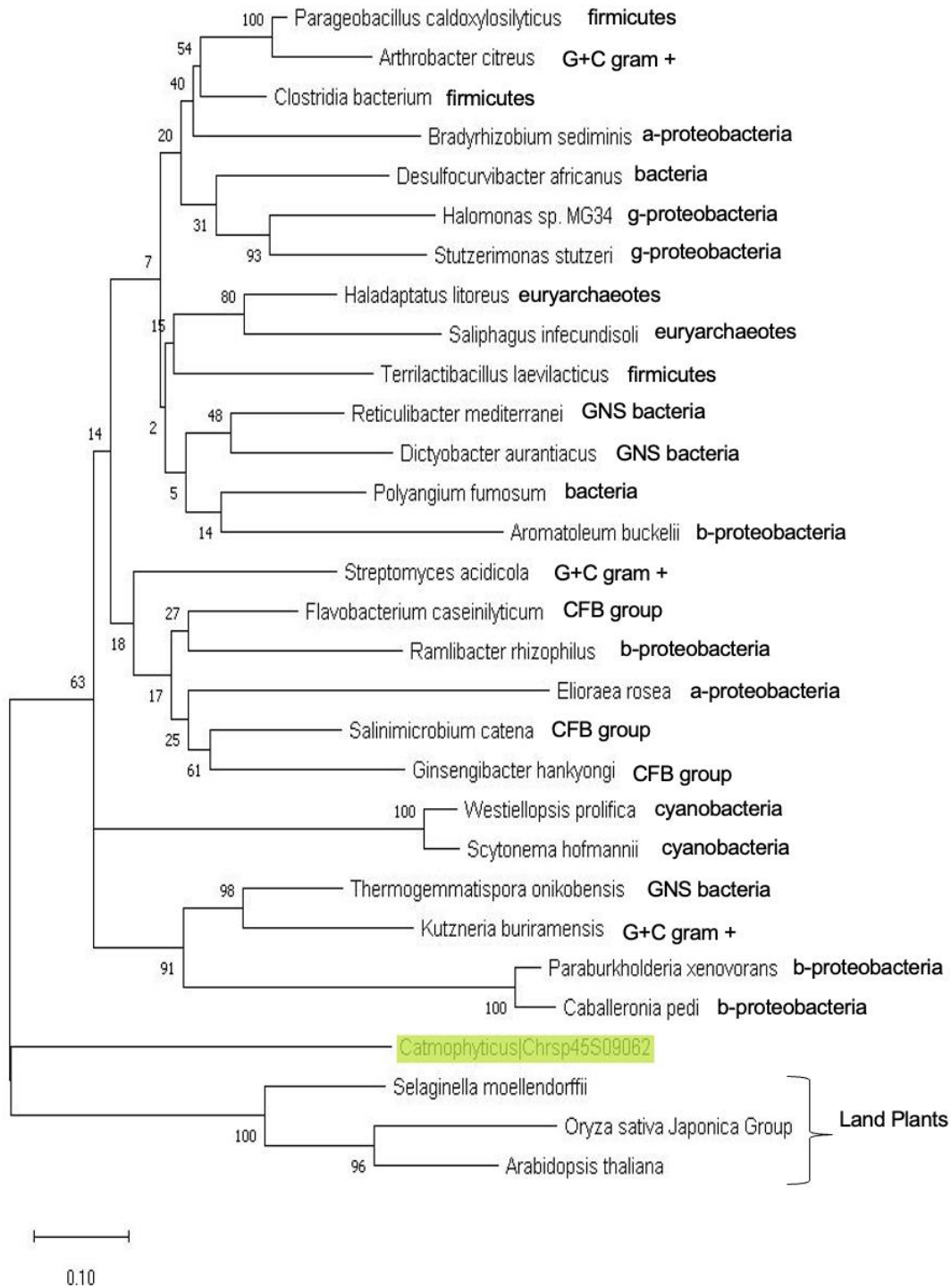


Figure 4. Phylogenetic tree *Chrsp45S09062*
 Highlighted – query sequence

RESULTS AND DISCUSSION

I employed phylogenetic analysis to identify horizontally transferred genes in *C. atmophyticus*. This involved comparing the gene sequences of interest from various organisms and constructing an evolutionary tree to determine their relationships.

There are 26 gene candidates from *C. atmophyticus* that I investigated the gene origin, transfer mechanism, and the function of the homolog in *A. thaliana* that were not included in this study. I concluded that 20 of these candidates were acquired horizontally and present in the flowering plant *A. thaliana*. In all cases, the donor of the transferred gene was either a bacterium or specifically a proteobacterium. The remaining six were most likely transferred via chloroplasts due to the large representation of cyanobacteria in the taxonomic distribution, and further investigation is required in these six candidates to confirm their method of transfer. Three transferred genes identified in my analyses, including *Chrsp22S08802*, *Chrsp45S09062*, and *Chrsp52S06587*, were subjected to further analyses of their biochemical functions and the biological processes in which they are involved (Table 1).

I was able to make conclusions about the type of gene transfer that the gene candidates underwent by analyzing the MSA in conjunction with the phylogenetic tree. When analyzing the MSA of each candidate, I found significant similarities among the sequences of the gene in different organisms, but the phylogenetic tree indicated that the organisms were distantly related, suggesting that the gene was acquired through horizontal inheritance rather than vertical. For example, 39 fully conserved regions were identified between the recipient and proteobacterial donor in the MSA of *Chrsp45S09062*. The conclusions made are consistent with features commonly identified in horizontally transferred genes and hold true for the three gene

candidates highlighted in this study. *Chrsp22S08802* and *Chrsp52S06587* exhibited many fully conserved regions in the MSA. The 20 candidates previously mentioned that are not included in this investigation exhibited these same traits resulting in a similar conclusion that these candidates were horizontally acquired. In the three gene candidates highlighted in this investigation, it was unclear as to the exact bacterial donor group.

In the cases of *Chrsp45S09062*, *Chrsp52S06587*, and *Chrsp22S08802*, the timing of the gene acquisition events can be inferred by examining the distribution of genes across different species and whether those genes are part of a monophyletic group. Monophyletic groups were identified among the land plants in *Chrsp22S08802* and *Chrsp45S09062* (Figures 2 & 4) indicating the gene transfer event may have occurred after the divergence of those species from their common ancestor. When analyzing the gene tree of *Chrsp52S06587* no monophyletic groups were identified (Figure 3). The lack of a monophyletic group provides less clarity as to the timing of the gene transfer event but does not rule out the gene being transferred horizontally.

As previously mentioned, this investigation aimed to further understand the impact of these horizontally acquired genes on the evolution of land plants. To fully understand the extent of the influence HGT had on the evolution of land plants, I analyzed the functions and biological processes for *A. thaliana* homologs of three identified genes using The Arabidopsis Information Resource (TAIR, www.arabidopsis.org) (Table 1). The homolog of *Chrsp22S08802* in *A. thaliana* encodes a protein in the histidine kinase family. The histidine kinase family plays a crucial role in signal transduction pathways in *A. thaliana*. These proteins are involved in the two-component system, which is a type of signal transduction pathway that allows bacteria,

fungi, and plants to sense and respond to environmental stimuli. Histidine kinases are known to be involved in a wide range of biological processes of plants, including growth and development, hormone responses, light responses, osmotic stress responses, and circadian rhythms. For example, some types of histidine kinases are involved in cytokinin signaling, which regulates plant growth and development. Other histidine kinases are involved in ethylene signaling, which regulates plant responses to environmental stresses such as drought and pathogens (Ueguchi, C., et al. 2001). The acquisition of this histidine kinase gene family has likely played a role in the development of plant defense mechanisms against pathogens and environmental stressors.

The homolog of *Chrsp45S09062* in *A. thaliana* encodes a protein in the cupredoxin superfamily. The cupredoxin superfamily is a group of proteins that contain a conserved copper-binding site and are involved in electron transfer reactions. In *A. thaliana*, the cupredoxin superfamily includes several members, such as stellacyanin and uclacyanin. Stellacyanin is a small copper-containing protein that plays a role in copper homeostasis and oxidative stress responses in plants. It is involved in transporting copper ions from the cell wall to the cytoplasm, where they can be used for various biochemical processes. Uclacyanin is a copper-containing protein that is localized in the apoplast, the extracellular space outside of the plasma membrane. It is involved in plant defense responses against pathogens and environmental stresses, likely through its ability to generate reactive oxygen species and facilitate lignin biosynthesis (Harrison, M. & Dennison, C. 2004).

The homolog of *Chrsp52S06587* in *A. thaliana* encodes for a protein in the calcium-dependent phosphotriesterase superfamily (CDP). In plants, the CDP superfamily proteins have

been found to be involved in various biological processes. One of these functions of CDPs in plants is their role in the detoxification of harmful chemicals, such as herbicides and insecticides. CDPs have been shown to be involved in the degradation of various xenobiotics that are commonly used in agriculture (Bray, E. 2004). In addition, CDPs have also been implicated in the detoxification of heavy metals, such as cadmium and copper, which can be toxic to plants. CDPs have also been found to play a role in abiotic stress response in plants. For example, a CDP gene in *A. thaliana*, *At3g11410*, has been shown to be upregulated in response to salt stress (Weeda, S., et al. 2014p). The presence of these enzymes has allowed plants to adapt to changing environmental conditions and has likely played a role in the evolution of plants' ability to thrive in diverse environments.

To summarize, I performed phylogenetic analyses to identify 20 horizontally transferred genes in *C. atmophyticus*. These acquired genes have also been retained in major land plant groups such as flowering plants. Additional analyses of functions and the biological processes for *Arabidopsis* homologs of three identified genes suggest that horizontally acquired genes have likely played a crucial role in plant adaptation to diverse environmental conditions and have contributed to the evolution of various plant structures and metabolic pathways. While our understanding of HGT in eukaryotes is still evolving, my research findings indicate that the potential impact of horizontally acquired genes highlights the crucial role of this evolutionary mechanism in the adaptation of eukaryotic lineages. The next steps of this investigation could include increasing the sample size of sequences and taxa included in the MSA, as well as using more advanced bioinformatic and statistical analysis methods, such as compositional analysis

and gene expression analysis, to yield more precise results regarding HGT traits, timeframe of transfer, and node transferred from.

Genes acquired via HGT in <i>Chlorokybus atmophyticus</i>		
Gene ID	Donor	Function via <i>Arabidopsis thaliana</i>
Chrsp22S08802	Proteobacteria	Histidine kinase
Chrsp45S09062	Proteobacteria	Cupredoxin
Chrsp52S06587	Bacteria	Calcium-dependent phosphotriesterase

Table 1. Results of *C. atmophyticus* gene candidates (*A. thaliana* information via TAIR)

LITERATURE CITED

- Bray, E. A. (2004). Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. *Journal of Experimental Botany*, 55(407), 2331–2341.
<https://doi.org/10.1093/jxb/erh270>
- Cho, Anna, Constructing Phylogenetic Trees Using Maximum Likelihood (2012). Scripps Senior Theses. 46. https://scholarship.claremont.edu/scripps_theses/46
- Dunning Hotopp, J. C. (2011). Horizontal gene transfer between bacteria and animals. *Trends in Genetics*, 27(4), 157–163. <https://doi.org/10.1016/j.tig.2011.01.005>
- Gao, C., Ren, X., Mason, A. S., Liu, H., Xiao, M., Li, J., & Fu, D. (2013). Horizontal gene transfer in plants. *Functional & Integrative Genomics*, 14(1), 23–29.
<https://doi.org/10.1007/s10142-013-0345-0>
- Harrison, M., & Dennison, C. (2004). Characterization of *Arabidopsis thaliana* stellacyanin: A comparison with umecyanin. *Proteins*, 55(2), 426–435.
<https://doi.org/10.1002/prot.20017>
- Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, 9(8), 605–618. <https://doi.org/10.1038/nrg2386>
- Keeling, P. J. (2009). Functional and ecological impacts of horizontal gene transfer in eukaryotes. *Current Opinion in Genetics & Development*, 19(6), 613–619.
<https://doi.org/10.1016/j.gde.2009.10.001>
- Ma, J., Wang, S., Zhu, X., Sun, G., Chang, G., Li, L., Hu, X., Zhang, S., Zhou, Y., Song, C. P., & Huang, J. (2022). Major episodes of horizontal gene transfer drove the evolution of land plants. *Molecular Plant*, 15(5), 857–871. <https://doi.org/10.1016/j.molp.2022.02.001>

- Quispe-Huamanquispe, D. G., Gheysen, G., & Kreuze, J. F. (2017). Horizontal Gene Transfer Contributes to Plant Evolution: The Case of *Agrobacterium* T-DNAs. *Frontiers in Plant Science*, 8. <https://doi.org/10.3389/fpls.2017.02015>
- Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 16(8), 472–482. <https://doi.org/10.1038/nrg3962>
- Ueguchi, C., Koizumi, H., Suzuki, T., & Mizuno, T. (2001). Novel Family of Sensor Histidine Kinase Genes in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 42(2), 231–235. <https://doi.org/10.1093/pcp/pce015>
- Wang, H., Li, Y., Zhang, Z., & Zhong, B. (2023). Horizontal gene transfer: Driving the evolution and adaptation of plants. *Journal of Integrative Plant Biology*. <https://doi.org/10.1111/jipb.13407>
- Weeda, S., Zhang, N., Zhao, X., Ndip, G., Guo, Y., Buck, G. A., Fu, C., & Ren, S. (2014). *Arabidopsis* Transcriptome Analysis Reveals Key Roles of Melatonin in Plant Defense Systems. *PLOS ONE*, 9(3), e93462. <https://doi.org/10.1371/journal.pone.0093462>
- Wickell, D. A., & Li, F. (2019). On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist*, 225(1), 113–117. <https://doi.org/10.1111/nph.16022>
- Yue, J., Hu, X., Sun, H., Yang, Y., & Huang, J. (2012). Widespread impact of horizontal gene transfer on plant colonization of land. *Nature Communications*, 3(1). <https://doi.org/10.1038/ncomms2148>