Transcriptomic Network Analysis of Cyanobacterial-Methylotroph Interactions

Jack Orebaugh[1], Pavlo Bohutskyi, PhD[1]

1. Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA

Abstract:

A previous study demonstrated the potential for Cyanobacterial-Methylotroph cocultures to facilitate biogas processing as well as to be used in other biotechnological applications. To advance this technology, we investigated potential interactions between *Cyanobacterium stanieri* HL-69 (HL69) and *Methylotuvimicrobium alkaliphilum* 20Z (20Z) by inferring and analyzing gene co-expression networks under co-culture and axenic conditions. Five different co-expression networks were examined. These networks were inferred using gene expression profiles for 20Z axenic condition, HL-69 axenic, HL-69, 20Z coculture, HL-69 coculture, and cross-species HL-69-20Z coculture. Through the analysis of node (gene) betweenness and node normalized degree values in all five network cases, we compared adjustments in gene expression between growth conditions (axenic vs co-culture) as well as identify biological functions relevant to interspecies interactions. This analysis was done to distinguish between gene interactions within an organism and gene interactions between two organisms. Moreover, for all five cases we investigated two different network cutoff levels of 3,000 and 10,000. By shedding light on inter- and intra- species interactions, we hope to gain a better understanding of how these two organisms interact. This research will allow the investigation of further biotechnological applications of coculture systems and optimization of such applications for biotechnological purposes.

Introduction:

Within the field of biotechnology, the use of bioreactors have increased over time due to their ability to produce a variety of different bioproducts with a variety of functions from biogas to wastewater treatment[1]. Bioreactors are susceptible to contamination from a variety of different factors, and therefore are typically kept under extreme conditions (high temperature, high pH, etc.) to protect against contamination. Selecting organisms that can be utilized within bioreactors requires the organism to be able to exist and thrive in such conditions and takes careful consideration. Due to the diversity of use cases for bioreactors selecting the organism or organisms that can survive these harsh conditions for cultivation becomes critical, especially when designing a bioreactor for a specific bioproduct. With the extreme conditions typically kept within bioreactors, algae have become a popular organism of study because of their hardy nature and ability to survive in extreme conditions[2].

To create bioproducts, a precise understanding of the function and interactions that the organism of choice has with their environment is necessary to optimize the system for production. With global warming on the rise and increasing greenhouse gases, solutions need to be made for how to deal with the excess CO_2 in the atmosphere, as well as how to clean up waste products from industry. Numerous strategies have been developed to fix atmospheric CO_2, with

algae being a method of fixing atmospheric CO_2 out of different kinds of wastewater to filter it for contaminants[3]. Additionally, with the multiple use cases for algae in biotechnology, research into exploiting the properties of algae continue to be investigated[4-6].

In addition to fixing atmospheric CO_2, another objective is to transform other harmful greenhouse gases, like CH_4 and other off gasses from microbes into useful bioproducts for industry. In microbial communities, different symbiotic relationships exist where systems maintain equilibrium, and these kinds of communities have been attempted to be recreated in laboratory settings for further study[7]. Through understanding how these different synthetic communities work together, we can understand how the organism(s) functions and how best to use them to create bioproducts as well as fix common greenhouse gases.

Specifically, utilizing methylotrophic bacteria within bioreactors to facilitate the synthesis of these bioproducts. Methylotrophic bacteria, or methylotrophs, can oxidize CH_4 into different metabolites, which can be used to remove unwanted methane from a variety of different substrates[8]. Methylotrophs are also very sturdy and can grow in harsh conditions, like the ones kept within bioreactors. Due to the ability for methylotrophs and algae alike to survive in both extreme conditions and preform important bioremediation functions, investigation into coculture interactions of these two organisms were carried out. The goal of this study is to determine how these two organisms interact on their own and with each other to better understand how to optimize their growth for producing bioproducts.

To understand the interactions of these organisms, gene network analysis was utilized to gain detailed knowledge about their function and cross-species interactions. Over the last several years, the field of –omics has exploded in popularity within research areas spanning numerous disciplines[9-11]. Within the arena of network biology, –omics has allowed for advances in understanding of numerous different systems with applications in industry, from drug discovery to metabolic pathways for bioproduction[5, 12]. To produce biogas and other useful bioproducts, the utilization of more and more sensitive approaches is critical to understand the complexity of the networks that we are looking to harness. In nature, organisms operate in such balance that minute changes and other nuances are incredibly difficult to reproduce in laboratory settings. However, recent advances in transcriptomics technology have allowed researchers to begin to understand these complex networks, and how they interact with the world around them[4].

For this study, we utilized two organisms, an algae, *Cyanobacterium stanieri* HL-69 (HL69) and a methylotrophic bacterium, *Methylotuvimicrobium alkaliphilum* 20Z (20Z). These organisms were chosen for their ability to survive and extreme conditions (high pH, extreme temperatures, etc.) as well as for individual characteristics, cyanobacteria for its photosynthetic properties of fixing CO_2, and methylotroph for its ability to oxidize CH_4[8]. To better understand these organisms, a gene network analysis approach was taken to get a clear understanding of the genetic interactions of these organisms.

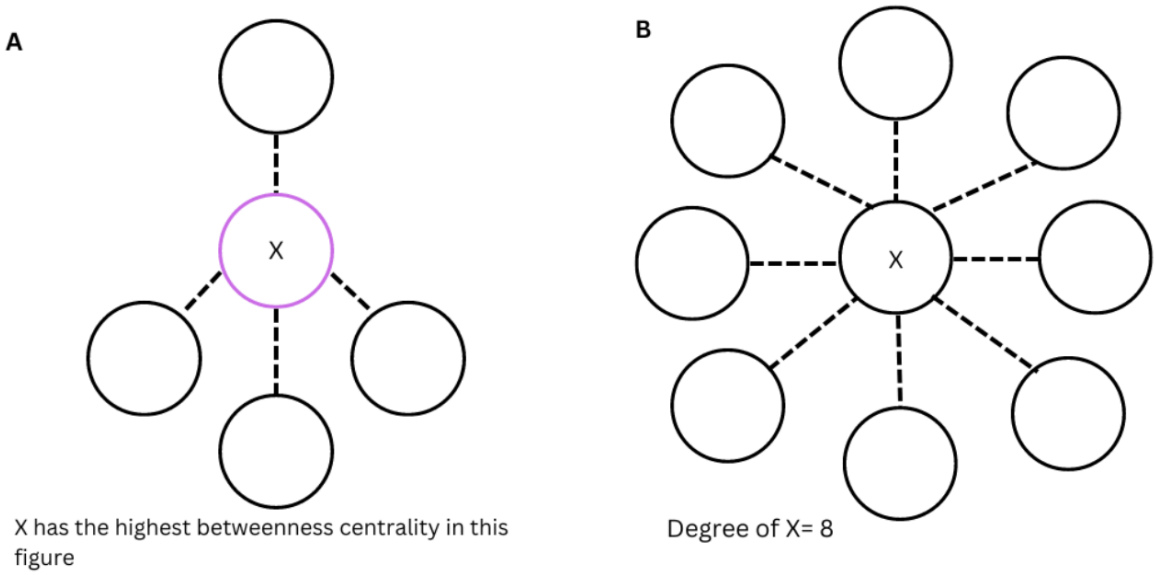
Materials and Methods:

Five different gene network cases were studied, axenic condition 20Z, axenic HL69, coculture condition 20Z, coculture HL69, and cross-species interaction HL69-20Z coculture.

These cases were studied because they allowed for baseline data to be gathered about 20Z and HL69 gene networks in singular growth conditions that could be compared to the gene expression data of the coculture growth conditions. Of note, we separated by species the gene interactions from the cross-species coculture interactions, to see what genes were differentially expressed between intra- and inter- species interactions. We analyzed the networks in Cytoscape, a network analysis visualization software, combined with a modified FastGreedy algorithm to trim gene network data to desired cutoffs[13, 14]. The modified FastGreedy script allowed for raw GENIE3 output files to be trimmed of satellite networks and allowed for focus on the main gene network responsible for the majority of interactions[15]. Additionally, this allowed us to reduce the noise within our raw data, so that greater attention could be paid to the main network interactions.

Each network was trimmed to two different cutoffs, 3,000 and 10,000 edges. The edges, double-sided connections that come from each node (gene), showing how other genes are related to one another. These edges that connect nodes can be used to calculate how central or "important" a gene is to the overall network, which can be calculated through a variety of different topological statistics (degree, betweenness centrality, stress, etc.)[16]. These topological values were calculated via Cytoscape's "Analyze Network" function and were exported for each network case for statistical analysis and comparison between networks. This topological data was then merged with Clusters of Orthologous Genes (COG) functional group information, that was assigned to each gene based on the specific gene's name[17]. This COG annotation allowed for greater depth of analysis to be performed by analyzing not only the gross topological data but also being able to analyze the gene networks by functional group to see what functions were differentially expressed.

Figure 1.



A. Provides a visual representation of betweenness centrality. Circle X having the highest betweenness centrality due to its position within the network, therefore being the central node that information must pass

through to get to other nodes. **B.** Provides a visual representation of degree centrality. With circle X having the highest degree because it has the most connections coming from it to other circle.

For specific comparison, degree and betweenness centrality were chosen as comparative values for each gene network based on the Giovanni and Carlo 2012 paper describing the different centrality analysis statistics suitable for complex biological networks[16]. Degree, defined as the number of connections or edges, that come from one node or gene to another[16]. The higher the degree, the greater the number of connections that gene has. Betweenness centrality can be defined as how central a node is to a specific genetic pathway, and therefore, how important a gene is to the overall gene network's function[16]. The higher the betweenness centrality the more "central" a node is to the network or specific pathway that is being analyzed. These two topological analysis factors were chosen due to their ease of interpretation as well as their diagnostic significance to the overall network function and relationships. From these values, inferences could be made about the centrality and importance of a gene or functional group to the specific network or coculture interactions.

For specific comparative analysis of the two network cutoffs, a normalization was applied to the degree and betweenness values. For degree, there were two separate calculations made. The first was for axenic and coculture interactions: where each degree value was divided by (2*(total # edges/total # species-specific nodes). For the interspecies interactions, the calculation was by doing each degree value divided by (total # edges/total # of species-specific nodes). The reasoning for this type of normalization was so that when averaged the normalized degree values across all cases would equal 1 and be comparable, and interspecies data sets contain both species, 20Z and HL69, within them, whereas the axenic and coculture do not. This means that while most of the datasets have the total number of nodes of the data set being used in the calculation, the interspecies cases only have around half the number of nodes compared to the total number of edges. Therefore, alterations were made to the calculation to keep all the numbers normalized to one and comparable. Additionally, the calculation for betweenness was mu0ch more straight forward. With all cases having each betweenness value dived by the total sum betweenness values per case, equaling 1 when averaged. With this data cleaning process, it allowed for the analysis of the different growth conditions, organisms, and network cutoffs to be uniform within scaling, allowing for easier comparison downstream.

Equation 1.

$$Normalized\ Degree^{a} = \frac{Degree\ Value}{(2 * (Total\ \#\ of\ Edges)/(Total\ \#\ of\ Species - Specific\ Nodes)}$$

$$Normalized\ Betweenness\ Centrality = \frac{Betweenness\ Value}{(Sum\ of\ Betweenness\ Values)}$$

Eq. 1. These two equations show how the normalization functions were calculated during the analysis. **A.** The normalized degree function is utilized only for the axenic and coculture data, the cross-species data does not include the coefficient multiplied to the total number of edges and total number of species-specific nodes. This is because the when the species-specific nodes are calculated for the cross-species calculations, the number of edges stays the same, while effectively halving the number of nodes going into the equation. Therefore, to normalize all values to 1, the coefficient is removed, to allow for the node difference to normalize for comparison.

Preliminary Results:

Analysis of the complete dataset is still on-going, and more detailed analysis of more topological analysis factors are in progress. The current data analysis shows that overall, the degree centrality values are larger in the 10,000-edge case when compared with the 3,000 edge cases. This is likely because of the greater number of edges within the main network, which allowed for a greater number of overall connections between nodes when compared to the more constrained network. Generally, the same trends are carried throughout the two different networks cutoffs regarding what functional categories have greater connections in one cutoff vs another. However, in some cases, when comparing from axenic growth to coculture and cross-species (interspecies) interactions, there is a large difference. For example, in Cell Motility, the axenic case of HL69 at 3,000 edges has .5 average degree, whereas the coculture growth case has over 1.75 average degrees. This differential expression of this motility function shows that in axenic conditions the cells are not moving as much as when in coculture growth conditions. Whereas looking at the interspecies expression of motility, the 3,000-cutoff value for HL69 was just over .75 average degrees, a drop of about a degree on the average. These kinds of fluctuations within the data as growth conditions were compared and are still be investigated.

Figure 2.

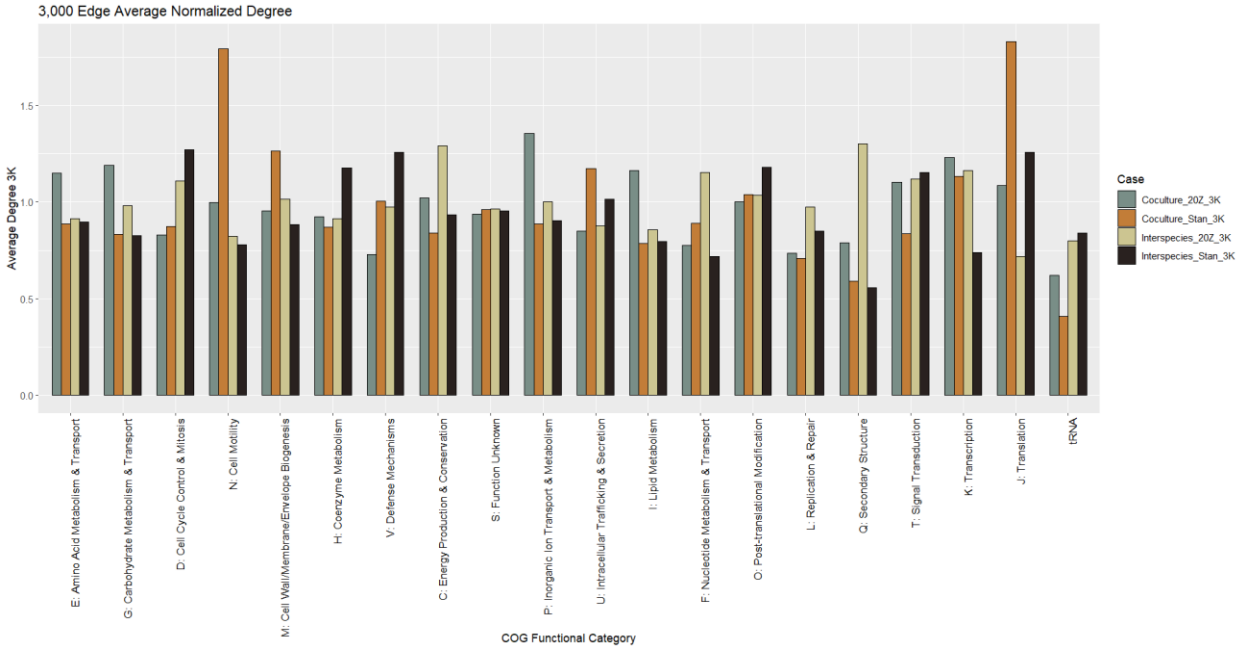


Fig. 2. This graph shows the 3,000-edge cutoff for the coculture compared to the interspecies gene networks. With the average normalized degree on the y-axis and the x-axis showing the COG functional categories.

Additionally, when comparing the translation category between the two different edge cutoffs, it is maintained that the highest degree value within this category across all gene networks is the coculture HL69, averaging >1.75 average degrees. Showing that perhaps in coculture growth conditions the HL69 organism has enriched translational response.

Within the betweenness centrality, averages hovered low, with no clear functional category hovering above the others. The main reason for the lack of clear-cut data points within

the betweenness values is speculated to be two-fold, the first being that betweenness is calculated by how used a path is to get to other nodes, and with larger and larger network size, there is no "one way" to move throughout the gene network.

The other potential reason for this is that the data is so varied, with the quartiles not encompassing all data points, leaving outliers high above even the maximum values within the boxplots.

Figure 3.

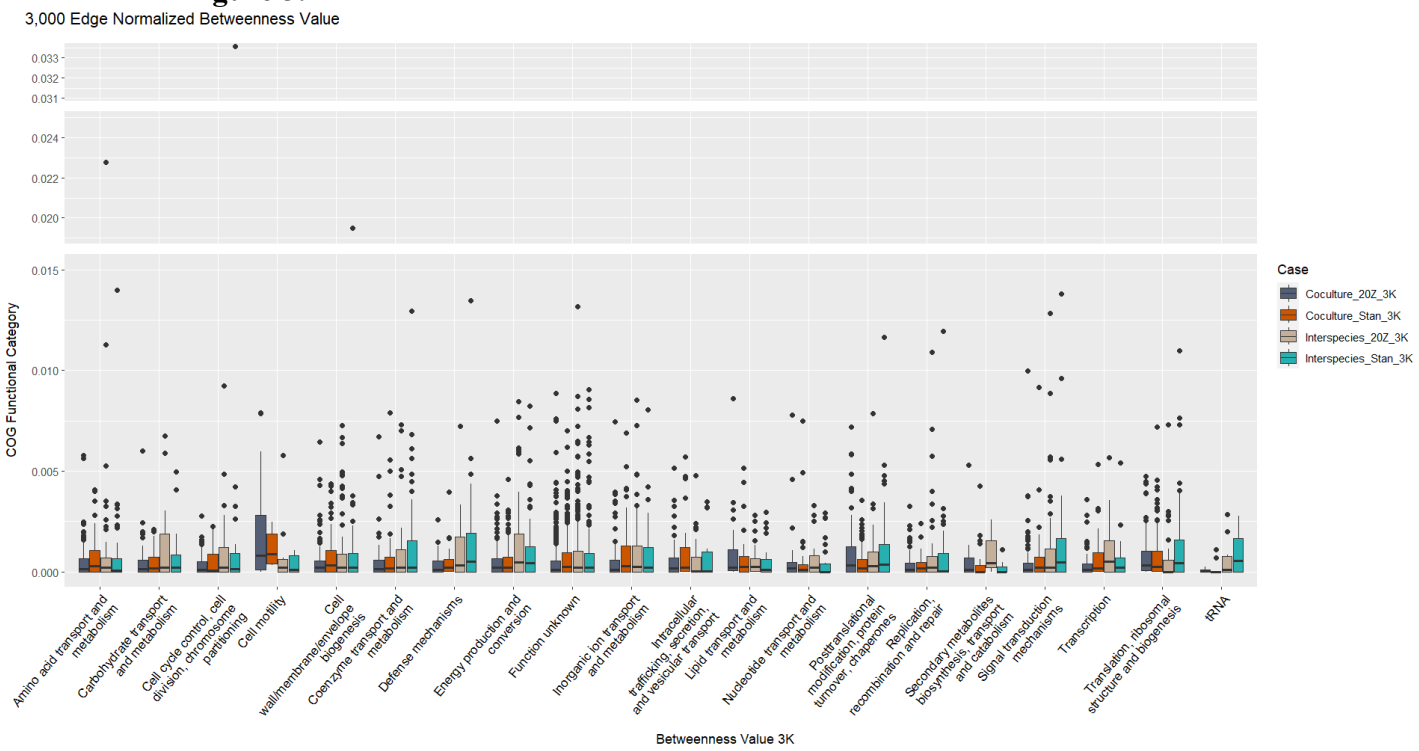


Fig. 3. This figure shows the betweenness value plotted out in a boxplot, in a 3,000-edge cutoff comparing the coculture expression to the interspecies expression data.

This makes for less straightforward interpretations of the data gathered so far and why further analysis is required to get a handle on this data. There is more analysis underway at the publishing of this report, as well as more research on network analysis of 20Z and HL69 which should further shed light on the interactions between these two organisms and how they can best be optimized for biotechnological applications.

Discussion:

This research can help further inform biotechnical applications of co-culture systems, as well as help to improve current process. With the explosion of recent microbial research and the push for renewable energy, all potential avenues of biofuel and bioproducts should be investigated. With the promising results coming out of bioreactor products, such as wastewater cleanup and other products, having a greater understanding of the transcriptomic functions of these two hardy organisms, 20Z and HL69, will hopefully lead to useful products being created from their interactions[3]. Having a better understanding of the different gene interactions that

occur between and within these organisms will allow for future research endeavors to study the flow of specific desirable metabolites within the network, as well as how to engineer this coculture growth to produce whatever desired bioproducts, such as biogas that are of interest. Furthermore, with a greater understanding of these networks, sequestering of greenhouse gases, such as excess CH_4, atmospheric CO2, as well as other harmful waste products could be removed from the environment and transformed into more useful products.

Acknowledgments:

The author JO would like to author PB, for his amazing mentorship and guidance, Additionally JO would like to thank PNNL's SULI program for allowing him the opportunity to work on some amazing research as well as the greater DOE for its involvement in his education and continued professional development.

Funding:

DOE BER project "Harnessing photoautotroph-methanotroph interactions for biogas conversion to fuels and chemicals using binary consortia" Award Number DE-SC0019181.

Predictive Phenomics Science & Technology Initiative at PNNL. Project "Microbial community engineering to reduce C/N footprints in biomanufacturing".

References

1. J. Drewnowski, A. Remiszewska-Skwarek, S. Duda and G. Łagód, Processes **7** (5), 311 (2019).
2. P. Varshney, P. Mikulic, A. Vonshak, J. Beardall and P. P. Wangikar, Bioresource Technology **184**, 363-372 (2015).
3. A. Ahmad, F. Banat, H. Alsafar and S. W. Hasan, Science of The Total Environment **806**, 150585 (2022).
4. P. Bohutskyi, R. S. McClure, E. A. Hill, W. C. Nelson, W. B. Chrisler, J. R. Nuñez, R. S. Renslow, M. A. Charania, S. R. Lindemann and A. S. Beliaev, Algal Research **42**, 101580 (2019).
5. S. Akash, B. Sivaprakash, N. Rajamohan and D.-V. N. Vo, Environmental Chemistry Letters **21** (3), 1477-1497 (2023).
6. H. Chen, T. Li and Q. Wang, Planta **249** (1), 195-219 (2019).
7. R. S. McClure, C. C. Overall, J. E. McDermott, E. A. Hill, L. M. Markillie, L. A. McCue, R. C. Taylor, M. Ludwig, D. A. Bryant and A. S. Beliaev, Nucleic Acids Res **44** (18), 8810-8825 (2016).
8. Z. J. Johnson, D. D. Krutkin, P. Bohutskyi and M. G. Kalyuzhnaya, Methods Enzymol **650**, 185-213 (2021).
9. E. Mathé, J. L. Hays, D. G. Stover and J. L. Chen, Yearbook of medical informatics **27** (01), 211-222 (2018).

10. P. Tolani, S. Gupta, K. Yadav, S. Aggarwal and A. K. Yadav, in *Advances in Protein Chemistry and Structural Biology*, edited by R. Donev and T. Karabencheva-Christova (Academic Press, 2021), Vol. 127, pp. 127-160.
11. S. Wu, D. Chen and M. P. Snyder, Current Opinion in Chemical Biology **66**, 102101 (2022).
12. H. N. Kadarmideen and N. S. Watson-Haigh, Bioinformation **8** (18), 855-861 (2012).
13. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, Genome Res **13** (11), 2498-2504 (2003).
14. A. Clauset, M. E. J. Newman and C. Moore, Physical Review E **70** (6), 066111 (2004).
15. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters and S. Aerts, Nature Methods **14** (11), 1083-1086 (2017).
16. S. Giovanni and L. Carlo, in *New Frontiers in Graph Theory*, edited by Z. Yagang (IntechOpen, Rijeka, 2012), pp. Ch. 16.
17. M. Y. Galperin, K. S. Makarova, Y. I. Wolf and E. V. Koonin, Nucleic Acids Res **43** (Database issue), D261-269 (2015).