# **Metabolic analysis of endosymbiont communities in membracid hosts using Pathway Tools**

Anson L. Clark Scholars Program 2023

Aaron Li

McKinlee Salazar

Dr. Amanda Brown

**Abstract**

Bacterial endosymbionts of the order Hemiptera play a vital role in supplementing their hosts' diets by synthesizing essential amino acids. While most hemipterans contain two or three endosymbionts that synthesize amino acids that the host is unable to obtain otherwise, previous research by Salazar et. al (2021) has demonstrated the existence of a large number of endosymbionts, including *Sulcia, Nasuia, Rickettsia, Sodalis,* *Wolbachia*, *Ophiocordyceps*, *Pectobacterium*, *Bombella* and *Asaia*, as well as a bacterium related to *Brenneria*, in species of the family Membracidae. In this study, we build upon the work of Salazar et. al by analyzing genomic data from symbionts within Brazilian membracids in combination with other related bacteria and insects from NCBI, using the Pathway Tools software suite to construct metabolic models for various endosymbionts. Comparative analysis revealed that endosymbiotic strains of bacteria showed significant signs of genome reduction. Furthermore, analysis of amino acid synthesis pathways revealed evidence of complementary network of pathways for amino acid production. The study also demonstrates the potential for interspecies pathway analysis using a multi-organism database to account for metabolite and enzyme transfer between the host insect and its endosymbionts. Further research is necessary to explore the metabolic interactions between membracids and their endosymbionts, providing insights into the broader significance of endosymbiosis in insect metabolism.

**Introduction**

Endosymbiosis plays a key role in the metabolism of many insects, especially among the order Hemiptera, commonly known as true bugs. Bacterial endosymbionts supplement the diets of hemipteran insects, as they synthesize amino acids that their hosts are normally unable to obtain through their plant sap-based diet. However, the physiological processes performed by endosymbionts and the evolution of metabolic complementarity and many of their interactions with their hosts are not well known, due to the complexity and diversity among different endosymbionts across insect species.

While most hemipteran insects, especially ones in the *Auchenorrhyncha* suborder, have been found to host two to three endosymbionts that synthesize missing amino acids,1 the family *Membracidae* is unique in that many membracid species host a wide array of endosymbionts, including ones not found in other hemipterans. Salazar et. al (2021) used genomic methods to identify a number of endosymbionts, including *Sulcia, Nasuia, Rickettsia, Sodalis,* and *Wolbachia*, as well as the yeast-like *Ophiocordyceps*,in Brazilian membracids. In addition, they identified a number of new endosymbionts that have not been found in other hemipterans, including endosymbionts that were found to be related to plant pathogens such as *Brenneria*, *Pectobacterium*, *Bombella* and *Asaia* based on BLAST hits for the16S rRNA gene.2

This study aims to expand on the work done by Salazar et. al by modelling the metabolic pathways present in membracid species and their endosymbionts. Understanding the role that these endosymbionts play in assisting their host in the production of key nutrients will allow for better understanding of the interactions between different endosymbionts and their host in both *Membracidae* and other insect species.

**Materials and Methods**

**Data Collection and Processing**

Genomic data from endosymbionts of membracid hosts was metagenomically assembled using the process outlined in Salazar et. al (2021), and the data for this study is obtained directly from their lab.2 To obtain genomic data for endosymbionts, their team collected Brazilian membracid samples and froze them at -80° C. The bacteriomes were then dissected, and the DNA was isolated and sequenced using Illumina HiSeq or NovaSeq. The resulting reads were then processed and *de novo* assembled using metaSPAdes.3 A BLAST pipeline was then used to sort and identify the assemblies.

The quality of the assembled genomes produced by this method was then improved using an iterative polishing process using Pilon.4 Annotation was performed on resulting prokaryotic genomes using Prokka.5

**NCBI Data Supplementation and Pathway Tools Analysis**

To construct metabolic models of the various endosymbionts living within the membracid samples, we used the PathoLogic module of the Pathway Tools software. PathoLogic constructs a metabolic model using an annotated genome in Genbank (.gb, .gbff, .gbk) or General Flat Format (.gff, gff3) formats as an input, using the MetaCyc database for information about enzymes, reactions and pathways,6 then stores them as a PGDB (Pathway/Genome Database). Each PGDB contains information including genes, enzymatic reactions, and their associated information (e.g., reactants and products, reaction direction, associated enzymes, reaction locations, Gibbs free energy, and more), computed pathways, and more.7

Among the various membracid samples, *Nasuia, Sulcia* and *Wolbachia* genomes from *Harmonides dispar* (Hemiptera: Membracidae), one of the membracids collected in the study, were used as inputs for PathoLogic. However, genomes for other secondary endosymbionts were highly fragmented and incomplete. As a result, for these endosymbionts, annotated genomes from NCBI were used instead.

Genomes for *Arsenophonus, Sodalis, Symbiopectobacterium* and *Rickettsia* were able to be obtained from endosymbiotic organisms inhabiting other hemipterans, however, for *Asaia, Bombella* and *Brenneria,* from free-living organisms were used due to a lack of availability of endosymbiotic variants.

Furthermore, due to the challenges presented by the assembly and annotation of eukaryotic species, a genome from a free-living parasitic strain of *Ophiocordyceps* was used instead of the metagenomically derived ones from membracids.

In addition, complete annotated genomes for membracid species were not available. As a result, an annotated *Homalodisca* *vitripennis* (Hemiptera: Cicadellidae)genome from the NCBI RefSeq database was used as a stand-in for a membracid genome, as *H. vitripennis* is currently the most closely related organism to membracid species that has a complete annotated genome. *H. vitripennis* chromosomes were downloaded separately and inputted into PathoLogic using the “Specify Replicons” tool to avoid the analysis of unmapped contigs. The sources for each of the genomes used in Pathway Tools analysis are detailed in **Table 1**.

|  |  |  |  |
| --- | --- | --- | --- |
| **Species** | **Source** | **NCBI RefSeq** | **NCBI Genbank** |
| *H. vitripennis* | N/A | GCF\_021130785.1 | GCF\_021130785.1 |
| *Arsenophonus* | *Aphis craccivora* endosymbiont | GCF\_013460135.1 | GCA\_013460135.1 |
| *Asaia bogorensis* | Orchid tree and Plumbago | GCF\_001547995.1 | GCA\_001547995.1 |
| *Bombella apis* | Bee gut bacteria | GCF\_025289935.1 | GCA\_025289935.1 |
| *Brenneria goodwinii* | Unknown | GCF\_002291445.1 | GCA\_002291445.1 |
| *Candidatus Karelsulcia muelleri* | *Harmonides dispar* endosymbiont | (unpublished, from this study) | (unpublished, from this study) |
| *Candidatus Nasuia deltocephalinicola* | *Harmonides dispar* endosymbiont | (unpublished, from this study) | (unpublished, from this study) |
| *Rickettsia* | *Cimex lectularius* endosymbiont | N/A | GCA\_020410805.1 |
| *Sodalis* | *Henestaris halophilus* endosymbiont | GCF\_900161835.1 | GCA\_900161835.1 |
| *Candidatus* *Symbiopectobacterium* | *Pseudococcus longispinus* endosymbiont | GCF\_018777385.1 | GCA\_018777385.1 |
| *Wolbachia* | *Harmonides dispar* endosymbiont | (unpublished, from this study) | (unpublished, from this study) |
| *Ophiocordyceps* (parasitic) | Extracted from deceased *Camponotus floridanus* | N/A | GCA\_012980515.1 |

**Table 1. Sources of species used in Pathway Tools Analysis. Genomes that were derived from metagenomic data from our membracid samples are highlighted in blue. Genomes that were sourced from endosymbionts on NCBI are highlighted in green, and genomes that were sourced from free-living bacteria are highlighted in red.**

**Evaluating Metabolic Data**

To evaluate metabolic complementarity in amino acid synthesis, the Pathway Tools Comparative Genome Browser was used to compare the presence of proteogenic amino acid synthesis pathways in *H. vitripennis* and membracid endosymbionts.

The approach of using a multi-organism PGDB was used to further account for the possibility of metabolite and enzyme transfer between bacteriocytes and endosymbionts. In this approach, the host genome and those of its endosymbionts are merged, which treats the system as effectively one organism. This allows PathoLogic to form pathways based on reactions in multiple organisms, and as a result, can simulate the effect of being able to freely transport metabolites and enzymes between the host insect and its endosymbionts. To test the validity of this approach, a merged PGDB consisting of the genomes of *H. vitripennis*, *Nasuia* and *Sulcia*, was constructed. A combined genome was input into PathoLogic by specifying *Nasuia* and *Sulcia* genomes as replicons along with *H. vitripennis* chromosomes, allowing PathoLogic to use reactions and enzymes from multiple different organisms in pathway inference.

By default, PathoLogic assigns each genome a realistic distribution of pathways based on the taxonomy of the organism, and prunes pathways that fall outside of that range. For the construction of the multi-organism database, taxonomic pruning was disabled to account for the different species.

**Results**

Pathway Tools analysis was performed for each of the organisms detailed in **Table 1.** Before running Pathway Tools on annotated genomes, a custom python script was used to concatenate contigs for each genome. For each organism, the number of total genes, number of total pathways, number of Enzymatic reactions, number of Transport reactions, and Genome size were recorded. Genomes that had been isolated from endosymbiotic bacteria, *Sulcia, Nasuia, Wolbachia, Rickettsia* and *Arsenophonus*, tended to have lower amounts of total pathways and coding genes when compared to genomes extracted from plant pathogens such as *Brenneria*. General statistics determined by Pathway Tools analysis are shown in **Table 2.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Organism** | **Total genes​** | **Total pathways​** | **Enzymatic  Reactions​** | **Transport  Reactions​** | **Genome size (kbp)​** |
| *H. vitripennis* | 17,474 | 194 | 1,496 | 92 | 1,706,914 |
| *Arsenophonus*​ | 2,467​ | 200​ | 1,277​ | 79​ | 2,403​ |
| *Asaia*​ | 2,825​ | 250​ | 1,489​ | 57​ | 3,198​ |
| *Bombella*​ | 1,768​ | 214​ | 1,438​ | 59​ | 1,918​ |
| *Brenneria*​ | 4,690​ | 299​ | 1,624​ | 161​ | 5,360​ |
| *Sulcia*​ | 235​ | 35​ | 300​ | 3​ | 227​ |
| *Nasuia*​ | 217​ | 12​ | 118​ | 2​ | 116​ |
| *Rickettsia*​ | 1,588​ | 91​ | 685​ | 27​ | 1,612​ |
| *Sodalis*​ | 948 | 128 | 899 | 35 | 1,622 |
| *Symbiopectobacterium*​ | 4,491 | 220 | 1,389 | 99 | 4,492 |
| *Wolbachia*​ | 1,428​ | 123​ | 940​ | 36​ | 1,495​ |
| *Ophiocordyceps*​ | 7,455​ | 251​ | 1,719​ | 78​ | 30,466​ |

**Table 2. Pathway Tools Statistics for *H. vitripennis* and endosymbionts. Bacteria with genomes derived from metagenomically assembled data from membracids are highlighted in blue. Bacteria with genomes that were sourced from endosymbionts on NCBI are highlighted in green, and those with genomes that were sourced from free-living sources on NCBI are highlighted in red.**

To evaluate metabolic complementarity, Pathway Tools’ Comparative Genome browser was used to compare the presence of pathways.

Pathways involved in biosynthesis of proteinogenic amino acids were examined to determine the role of endosymbionts in supplementing the host with essential amino acids. Overall, bacterial genomes that had been obtained from endosymbiotic organisms (*Sulcia, Nasuia, Wolbachia*, *Rickettsia, Sodalis*, *Arsenophonus,* and *Symbiopectobacterium*) displayed a much lower degree of redundancy when compared to their non-endosymbiotic counterparts, with phenylalanine, glutamate, and methionine each being uniquely synthesized by *Sulcia*, *Wolbachia* and *Symbiopectobacterium* respectively. Meanwhile, *Asaia,* *Bombella*, *Brenneria,* and *Ophiocordyceps*, whose genomes had been taken from free-living variants, all had comparatively complete sets of amino acid synthesis pathways. A diagram that details the presence of proteogenic amino acid synthesizing pathways is detailed in **Figure 1**.

A colorful squares with black text

Description automatically generated

**Figure 1. Presence of Amino Acid pathways among membracid Endosymbionts**

In addition to the single-organism PGDBs, a multi-organism PGDB was constructed, consisting of the genomes of *H. vitripennis*, *Sulcia* and *Nasuia.* The combined PGDB contains significantly more pathways and enzymatic reactions than the sum of pathways and enzymatic reactions among *H. vitripennis, Sulcia* and *Nasuia* PGDBs. The data from the merged PGDB is displayed alongside that of the individual PGDBs for *H. vitripennis, Sulcia* and *Nasuia* in **Table 3.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Organism** | **Total genes​** | **Total pathways​** | **Enzymatic Reactions​** | **Transport**  **Reactions​** |
| *Combined PGDB* | 17,926 | 323 | 2,163 | 99 |
| *H. vitripennis + Sulcia + Nasuia* | 17,926 | 241 | 1,914 | 97 |
| *H. vitripennis* | 17,474 | 194 | 1,496 | 92 |
| *Sulcia*​ | 235​ | 35​ | 300​ | 3​ |
| *Nasuia*​ | 217​ | 12​ | 118​ | 2​ |

**Table 3. Pathway Tools statistics for the merged PGDB**

The Pathway Tools Comparative Genome Browser was used to determine the presence of proteogenic amino acid biosynthesis pathways that were not individually contained within any of *Nasuia*, *Sulcia* or *H. vitripennis*. The combined PGDB contained pathways for Cysteine, Glutamate and Proline that were not present in any of the species individually. However, neither the merged PGDB nor any of *H. vitripennis, Nasuia* or *Sulcia* contained pathways synthesizing arginine, cysteine, or methionine. The results for this analysis are revealed in **Figure 2.**

A chart of different colored squares

Description automatically generated

**Figure 2. Comparative analysis of the merged PGDB consisting of *H. vitripennis, Sulcia* and *Nasuia* with each organism individually.**

**Discussion**

In this study, the Pathway Tools software suite was used to model metabolic pathways in membracid endosymbionts. Genomes for endosymbionts discovered by Salazar et. al were obtained through metagenomic assembly and annotated using prokka, with genomes that were overly fragmented being substituted for annotated genomes from NCBI. These genomes were used as data for the PathoLogic, a component of Pathway Tools with metabolic modelling capabilities, and metabolic models (known as Pathway/Genome Databases or PGDBS) were constructed for each organism. Pathway Tools Comparative Genome Browser was used to evaluate the complementarity of amino acid biosynthesis pathways between each of these organisms. In addition, to account for the potential transportation of metabolites and enzymes between the endosymbionts and their insect hosts, a multi-organism database was constructed as proof of concept by merging *H. vitripennis*, *Sulcia* and *Nasuia* genomes.

Initial results from Pathway Tools point to the significance of genome reduction in pathway analysis. Bacteria whose genomes came from endosymbiotic strains displayed highly different characteristics when compared to free-living variants, with lower counts of genes and pathways across the board.

Comparative analysis further revealed a moderately complementary set of amino acid pathways among endosymbionts. The amino acids phenylalanine, glutamate, and methionine are each uniquely synthesized by *Sulcia*, *Wolbachia* and *Symbiopectobacterium* respectively. In addition, across the 20 main proteogenic amino acids, only arginine and proline are not represented among pathways in any of *H. vitripennis* and its endosymbionts. Free-living variants of *Ophiocordyceps, Asaia, Brenneria* and *Bombella* each contain biosynthesis pathways for both arginine and proline, and it is possible that the endosymbiotic strains retain these pathways.

Another possibility is that the pathways for the synthesis of these amino acids are broken up across multiple species. Indeed, the construction of a multi-organism PGDB revealed the presence of pathways split across *H. vitripennis, Sulcia* and *Nasuia*. The merged PGDB contained a significantly higher number of pathways and enzymatic reactions than the sum total of pathways and enzymatic reactions of each of the constituent members individually. Furthermore, the merged PGDB contains amino acid biosynthesis pathways for cysteine, glutamate, and proline, meaning that only arginine biosynthesis pathways are not represented among *H. vitripennis* or any of the endosymbionts. It is very possible that the addition of new organisms to the merged PGDB would reveal an interspecies arginine biosynthesis pathway. However, further testing is required to confirm if these results are a result of the removal of taxonomy-based pruning (as detailed in the Methods section) rather than the inclusion of inter-species pathways.

An interesting idiosyncrasy determined by Pathway Tools analysis is that the eukaryotic organisms (*H. vitripennis* and *Ophiocordyceps*) displayed significantly lower pathway counts than expected. Despite having significantly higher genome sizes and gene counts, *H. vitripennis* and *Ophiocordyceps* had similar numbers of enzymatic reactions as many of the free-living prokaryotic bacterium. It is possible that *MetaCyc*, the database used by Pathway Tools to infer the presence of reactions, is not fully complete with respect to eukaryotic reactions and pathways, however, further investigation is required to determine the reason behind this behavior. Further experimentation with the MetaCyc database, as well as the usage of the BioCyc database, which contains highly curated metabolic data for a multitude of organisms, could potentially resolve this problem.

**Future Work**

Due to the lack of complete genomes in many of the prokaryotic endosymbionts that were gathered from membracids, many bacterial genomes were replaced by reference genomes from NCBI for metabolic modelling. Genomes of endosymbiotic variants of *Arsenophonus, Sodalis, Rickettsia* and *Symbiopectobacterium* from hemipteran hosts were used as stand-ins. However, due to the rapid genetic drift that endosymbionts undergo,8 it is very possible that endosymbionts in membracid hosts have significantly different metabolic capabilities than the models used as stand-ins. Furthermore, the genomes from free-living *Asaia*, *Bombella* and *Brenneria* variants displayed very different properties from the genomes from endosymbiotic organisms. It is very likely that the pathways that are present in those species are not reflective of the actual compounds produced by their endosymbiotic variants.

Considering these potential discrepancies, an important next step is to further process and annotate genomic data from other endosymbionts, to allow for a more accurate understanding of the products that the endosymbionts provide for their host.

Similarly, while *H. vitripennis* functions as a stand-in for the membracid host, it likely does not fully represent their metabolic capabilities. hemipterans tend to have highly complementary pathway sets with their endosymbionts (Bennet and Moran), and as a result, amino acid biosynthesis pathways sets are highly varied across hemipterans as redundant pathways are lost. Especially when considering that none of *H. vitripennis*, *Sulcia* and *Nasuia* possess capabilities to synthesize necessary amino acids like arginine or methionine, but *H. vitripennis* pathways in alanine, asparagine, aspartate, glycine, and isoleucine are also pathways present in *Sulcia* or *Nasuia*, it is reasonable to assume that the amino acid biosynthesis pathway sets of membracid species are not only significantly different from those of *H. vitripennis* but are less redundant. As a result, it is crucial to successfully assemble and annotate a complete Membracidae genome to improve the accuracy of the analysis.

The overall effect of *Ophiocordyceps* as an endosymbiont is still unexplored within the realm of this study. The effects of genetic drift and genome reduction on endosymbionts are well documented for prokaryotic organisms but remains a mystery in eukaryotic endosymbionts, despite similar fungal endosymbionts being found in multiple other of hemipterans, including *Aphidoidea* and *Fulgoroidea*.8 Further study of *Ophiocordyceps* endosymbionts in membracids could elucidate the origin and function of similar eukaryotic endosymbionts.

While the method of constructing a multi-organism PGDB is time-consuming, it elucidates the presence of pathways that are completed across species. Further analysis of specific pathways revealed by this method is necessary to determine reactions that are shared across organisms. However, this method shows significant promise for the future, and it may prove valuable to utilize this approach with other organisms.

**Conclusion**

This study used the Pathway Tools software suite to model metabolic pathways in membracid endosymbionts. The genomic data from endosymbionts discovered by Salazar et. al (2021) was supplemented by reference data from NCBI, and was used to construct metabolic models for each organism. Analysis of general Pathway Tool statistics revealed that endosymbiotic strains lower counts of genes and pathways compared to free-living strains, suggesting genome reduction in the endosymbionts studied. Comparative analysis of the pathways revealed fairly complete sets of amino acid biosynthesis pathways for free-living bacteria, as well as a moderately complementary set of amino acid pathways among endosymbionts, with some amino acids uniquely synthesized by specific bacteria. The construction of a multi-organism PGDB further demonstrated the potential for interspecies pathway analysis.

Understanding the metabolic interactions between membracids and their endosymbionts is crucial for gaining insights into the broader significance of endosymbiosis in insect metabolism. This research not only contributes to our understanding of the intricate relationships between insects and their bacterial and fungal partners but also sets the stage for future investigations into the functional roles of these endosymbionts.

**References**

1. Bennett GM, Moran NA. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol*. 2013;5(9):1675-88. doi:10.1093/gbe/evt118

2. Salazar MM, Pupo MT, Brown AMV. Co-Occurrence of Viruses, Plant Pathogens, and Symbionts in an Underexplored Hemipteran Clade. *Front Cell Infect Microbiol*. 2021;11:715998. doi:10.3389/fcimb.2021.715998

3. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res*. May 2017;27(5):824-834. doi:10.1101/gr.213959.116

4. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963

5. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. Jul 15 2014;30(14):2068-9. doi:10.1093/bioinformatics/btu153

6. Caspi R, Billington R, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res*. Jan 08 2020;48(D1):D445-D453. doi:10.1093/nar/gkz862

7. Karp PD, Midford PE, Billington R, et al. Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform*. Jan 18 2021;22(1):109-126. doi:10.1093/bib/bbz104

8. Bennett GM, Moran NA. Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A*. Aug 18 2015;112(33):10169-76. doi:10.1073/pnas.1421388112