Comprehensive Investigation of the Gut Microbial Community of *Antheraea assamensis* Helfer through Meta-omic Approach

A Thesis

Submitted in the Partial Fulfilment of the Requirement for the degree of

Doctor of Philosophy

By

Dharitri Saikia
Registration Number: 176152001

Centre for the Environment

Indian Institute of Technology Guwahati

Kamrup 781039, Assam, India 2023

Comprehensive Investigation of the Gut Microbial Community of *Antheraea assamensis* Helfer Through Meta-omic Approach

A thesis submitted

in the partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

by

Dharitri Saikia
Roll Number: 176152001

Under the supervision of

Prof. Utpal Bora
Department of Biosciences & Bioengineering,
IIT Guwahati
&
Prof. Mihir Kumar Purkait
Department of Chemical Engineering,
IIT Guwahati

March 2023

Centre for the Environment

Indian Institute of Technology Guwahati

Kamrup 781039, Assam, India

Centre for the Environment
Indian Institute of Technology Guwahati
Kamrup 781039, Assam, India

DECLARATION

This is to declare that the content embodied in this thesis entitled "**Comprehensive Investigation of the Gut Microbial Community of *Antheraea assamensis* Helfer Through Meta-omic Approach"** is the result of an investigation carried out by me under the supervision of Prof. Utpal Bora and Prof. Mihir Kumar Purkait and is submitted to the Centre for the Environment, Indian Institute of Technology Guwahati, Kamrup 781039, Assam, India for the award of the degree of **Doctor of Philosophy**. This work has not been submitted elsewhere for any degree or diploma of any institute /university to the best of my knowledge and belief.

In keeping with the general practice of scientific investigation due acknowledgment has been made wherever the work of other investigators is referred.

Dharitri Saikia

Guwahati
March 2023

Dharitri Saikia
Registration Number -176152001
Centre for the Environment
Indian Institute of Technology Guwahati,
Kamrup 781039, Assam, India

Centre for the Environment

Indian Institute of Technology Guwahati

Kamrup 781039, Assam, India

CERTIFICATE

This is to certify that the content embodied in this thesis entitled "**Comprehensive Investigation of the Gut Microbial Community of *Antheraea assamensis* Helfer Through Meta-omic Approach"** is the result of an investigation carried out by **Dharitri Saikia (Registration Number – 176152001)** under the supervision of **Prof. Utpal Bora** and **Prof. Mihir Kumar Purkait** at Centre for the Environment, Indian Institute of Technology Guwahati, Kamrup 781039, Assam, India for the award of the degree of **Doctor of Philosophy**. This work has not been submitted elsewhere for any degree or diploma of any institute/university.

6/3/2023

Prof. Utpal Bora
Department of Biosciences & Bioengineering
Centre for the Environment
Indian Institute of Technology Guwahati,
Kamrup 781039, Assam, India

6/3/2023

Prof. Mihir Kumar Purkait
Department of Chemical Engineering
Centre for the Environment
Indian Institute of Technology Guwahati,
Kamrup 781039, Assam, India

This Thesis is Dedicated to

Maa-Dita

ACKNOWLEDGEMENTS

As I am about to embark on another adventure of my life, I would like to express my deep gratitude to everyone who stood by me and helped me in the journey till now and make it into a beautiful one. First of all, I take this opportunity to express my profound gratitude and deep regards to my supervisor, Prof. Utpal Bora for his constant support, guidance and encouragement, throughout my doctoral research. His supervision, advice and constructive criticism are the key factors in shaping this work into its present form. I am thankful to him for the freedom and trust he endowed me with, his benevolent care and for providing me with a comfortable work environment. I would like to extend my sincere thanks to my co-supervisor Prof. Mihir Kumar Purkait, who continuously encouraged me to explore multiple approaches to the field and provided me with invaluable advice. I also would like to thank both Prof. Bora and Prof. Purkait for allowing me to conduct my research independently and for teaching me scientific communication. I am extremely grateful and indebted to my Doctoral Committee members, Prof. Chandan Mahanta, Prof. Ranjan Tamuli, Prof. Karuna Kalita, and Prof. Subhendu Sekhar Bag for their valuable suggestions, motivation and scientific guidance throughout my PhD which always helped me to make my work better. I would like to convey my gratitude to the Centre for the Environment and the Institutional Biotech Hub at the Centre for the Environment, Department of Biosciences and Bioengineering, Central Instrumentation Facility and Param-Ishan, IIT Guwahati for providing me with all the necessary facilities to pursue my research. I sincerely acknowledge the financial support from the Ministry of Human Resource Development (MHRD), Government of India for providing me fellowship as well as the Department of Biotechnology (DBT), Government of India for funding our laboratory. I render my deepest thanks to my seniors, Hasna ba, Deepika di, Kabiraj Da, Vimal Da, Sunita di and Papori ba for their love and guidance. I express my heartfelt gratitude to the present lab members of BERL for making this journey enjoyable and easier. Jon da, Adhiraj and Juna ba, thank you for all your suggestions, help and guidance. Pulak, thank you for standing by my side when times get hard. I can't thank you enough for your constant support and mental boosts. I also

extend my gratitude to Udangshree, Pinky, Raunak, Biju, Asma for their help and moral support. My special gratitude to Dr. Pragya Sharma Ma'am, Gulgul and Champ for their love and affection, the laughter and food we shared. The constructive advice given by Prgaya ma'am has helped me a lot in my hard times. I extend my gratitude to Dr. Kartik Neog and Dr. Jintu Dutta for their help in my research. I express my heartfelt gratitude to my loving family, Deuta, Jadav Saikia, Maa, Smt. Champa Saikia, Dollie ba and Munu for their constant encouragement and support in every possible way. My everything starts and ends with you, Dita. Thank you for the freedom and faith you empowered me with. Much love to Zoyee, Chiku and Allie for all the joy and smile they bring into my life. I am thankful to my dear friends from the Centre and other departments, Abhishek, Aquib, Shilpa, Swagata, Himadree, Prabhat, Prangan, Sudeshna, Anweshan, Karan, Rahul, Langtuk, Gariyashi, JK bhaiya and everyone else from the Centre. Much love to you all. I am also thankful to Naresh and Ashim for your constant support. Ashis, you were one of the biggest support systems. Thank you for your help and for introducing the exciting area of scientific research to me. I also take this opportunity to acknowledge the Central Muga and Eri Research and Training Institute, Niron Rabha Da and Bou, and Ramkrishna Bharali for helping me in collecting the experimental samples. I would like to express my sincere gratitude to the staff members of the Centre for the Environment, namely, Partha da, Deepmoni ma'am, Rajiv Da, Kaustav Da, Mridul Da and Supriyo Da; security guards, especially, Bipul Da, Sanjib da and Dipen da for their sustained help and support throughout my PhD work. I also thank all my previous teachers and people who have contributed immensely to my life in diverse ways. PhD wouldn't have been possible without the contribution of everyone who touched my life. Last but not least, I am always thankful to the Almighty for everything in my life.

March 2023 - Dharitri

Table of Contents

SYNOPSIS viii
Abbreviations xiv
List of Tables ix
List of figures xi
Introduction and Review of Literature 1
1.1 Lepidoptera 1
1.1.1 Feeding habits of Lepidoptera 3
1.1.2 The Saturniids 4
1.2 *Antheraea assamensis* Helfer: The Muga Silkworm 4
1.2.1 Distribution of *Antheraea assamensis* 6
1.2.2 The Host plants of *A. assamensis* 6
1.2.3 Rearing of *A. assamensis* 7
1.3 Gut microbiota, host organism and their interaction 7
1.3.1 The gut microbiota of Lepidoptera 8
1.3.2 Phyllosphere microbiota of host plants of phytophagous insects 11
1.3.3 Implication of gut microbiota for conservation 11
1.4 The Developing Concept of Metagenomics 13
1.4.1 The 16S rRNA gene as a phylogenetic marker 15
1.4.2 Shotgun Metagenomics to study the genomic potential of gut bacterial community 16
1.5 Metatranscriptomics to find out active bacterial populations under different physiological condition 18
1.6 Basic tools and techniques for metagenomic and metatranscriptomic analyses 19
1.7 Metabolomics: A powerful emerging tool to study the metabolites 25
1.8 Research gap 28
1.9 Objectives of the work 29
1.10 Organization of the Thesis 29
References 30
Structural analysis of the gut metagenome of *Antheraea assamensis* Helfer and its comparison with host leaf-associated microbiota 52
2.1 Introduction 53
2.2 Materials and methods 55
2.2.1 Sample collection 55
2.2.2 Sample preparation 56
2.2.3 Metagenomic DNA isolation 56
2.2.4 High-throughput sequencing 57
2.2.5 Amplicon sequence processing 58
2.2.5.1 Quality control 58

2.2.5.2 Alpha and Beta diversity analysis 58
2.2.5.3 Phylogenetic diversity analysis 60
2.2.5.4 Taxonomic analysis 60
2.2.5.5 Predictive functional profiling with PICRUSt2 60
2.3 Results 61
2.3.1 Amplicon Sequence Analysis 61
2.3.2 Structure of the gut microbial community of *A. assamensis* 68
2.3.3 Structure of the leaf microbial community of three different host plants of *A. assamensis* 72
2.3.4 Identification of the core and unique gut microbial taxa of *A. assamensis* reared on three different host plants 75
2.3.5 Identification of the core microbial taxa in the leaves of the host plants 76
2.3.6 Comparison between the gut and host leaf microbial composition 76
2.3.7 Alpha and beta diversity analysis 77
2.3.8 Predictive functional profiling with PICRUSt2: 80
2.4 Discussion 87
2.4.1 The unique and diverse gut microbial community of *A. assamensis* 87
2.4.2 The leaf microbial community of the host plants of *A. assamensis* 89
2.4.3 Impact of host leaf microbiome on the structure of gut microbiome of *A assamensis* 90
2.4.4 Role of larval gut microbiota in leaf digestion 90
2.4.5 Chitinase production by gut microbiome 91
2.4.6 Microbial detoxification of plant secondary metabolites in the larval gut 91
2.4.7 Methane-producing microbe in the gut of *A. assamensis* 92
2.4.8 Fat body metabolism in *A. assamensis* and role of gut microbiota 94
2.5 Conclusion 95
References 97
Functional analysis of *Antheraea assamensis* Helfer gut metagenome 108
3.1 Introduction 109
3.2 Materials and methods 111
3.2.1 Sample collection 111
3.2.2 Sample preparation 111
3.2.3 Metagenomic DNA isolation 111
3.2.4 Shotgun Sequencing 112
3.2.4.1 Shotgun sequence processing 112
3.3 Results 115
3.3.1 Quality control of the raw data 115
3.3.2 Metagenome assembly 116
3.3.3 Functional Annotation 116
3.3.4 Taxonomic analysis 130
3.4 Discussions 134

3.4.1 Functional annotation of the gut metagenome 134
3.4.2 Taxonomic annotation of the gut bacterial community of *A. assamensis*. 136
3.4.3 The gut archaeal community of *A. assamensis* 137
3.4.4 The viral population inside the gut of *A. assamensis* 138
3.5 Conclusion 139
References 140
Gene expression profiling of *Antheraea assamensis* Helfer gut metagenome 153
4.1 Introduction 154
4.2 Materials and methods 155
4.2.1 Sample collection 155
4.2.2 Gut dissection and total RNA isolation 156
4.2.3 RNA isolation 156
4.2.4 RNA sequencing 157
4.2.5 Quality control 157
4.2.6 Metatranscriptome assembly and annotation 158
4.3 Results 160
4.3.1 Quality control of the raw data and metatranscriptome assembly 160
4.3.2 Metatranscriptome annotation 160
4.3.3 Gene expression analysis of the metabolically active gut microbial community of *Antheraea assamensis* larvae 162
4.3.4 Expression analysis of transcripts for carbohydrate metabolism in the gut bacterial community of *A. assamensis* 163
4.3.5 Identification of transcripts involved in biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments 173
4.3.6 Identification of transcripts involved in stress response 178
4.3.7 Identification of transcripts involved in the metabolism of aromatic compounds 181
4.3.8 Identification of the active gut bacterial community in the gut of *A. assamensis* larvae and their function 188
4.3.9 The active gut archaeal taxa in the gut of *A. assamensis* larvae and their function 191
4.4 Discussion 195
4.4.1 Carbohydrate metabolism and the larval gut microbiota 195
4.4.2 Identification of transcripts involved in the biosynthesis of vitamins and pigments 197
4.4.3 Stress response mechanism of the larval gut microbiota 198
4.4.4 Metabolism of aromatic compounds by the gut microbial community 199
4.4.5 The active gut microbial taxa inside the gut of A. assamensis 200
4.5 Conclusion 201
Reference 203
Untargeted metabolomic analysis of *Antheraea assamensis* Helfer 217
5.1 Introduction 218
5.2 Materials and methods 222
5.2.1 Materials 222

5.2.2 Sample preparation and extraction 222
5.2.3 Metabolites extraction 223
5.2.4 Liquid Chromatography- High Resolution Mass Spectrometry (LC-HRMS) analysis 224
5.2.5 HRMS data analysis 224
5.3 Results 230
5.3.1 Raw data import and preprocessing 230
5.3.2 Feature detection 230
5.3.3 Annotation 236
5.3.3.1 Identification of Amino acids 237
5.3.3.2 Identification of carbohydrates 237
5.3.3.3 Identification of pigment molecules 241
5.3.3.4 Identification of other miscellaneous metabolites 241
5.4 Discussion 247
5.4.1 The gut metabolome of *Antheraea assamensis* larvae 247
5.4.2 The leaf metabolome of two host plants of *A. assamensis* 249
5.4.3 The silk gland metabolome of *Antheraea assamensis* 250
5.5 Conclusion 251
References 253
Summary and Prospects 264
6.1 Structural analysis of the gut metagenome of *A. assamensis* and its comparison with host leaf-associated microbiota 264
6.2 Functional analysis of *Antheraea assamensis* gut metagenome. 265
6.3 Chapter 4: Gene expression profiling of *A. assamensis* gut metagenome 266
6.4 Untargeted metabolomic analysis of *Antheraea assamensis* 266
Prospects 267
List of publication 269
List of conference/Workshop/Symposia 270
Curriculum vitae 271

SYNOPSIS

Antheraea assamensis Helfer, commonly known as the Muga silkworm, is an endemic, economically important Lepidopteran species explicitly found in the north-eastern part of India and its neighbouring areas. This semi-domesticated, multivoltine, polyphagous, edible insect produces an economically important, lustrous golden coloured silk with high tensile strength and durability. The cultivation of Muga silkworm for silk production has been in practice for ages in Assam, a state of India. Due to its extreme durability, magnificent colour and luster, Muga silk is considered one of the most expensive natural fibers in the world. The fibroin part of the Muga silk is also a promising material for tissue engineering. Unlike mulberry-feeding silkworms, *A. assamensis* is semi-domesticated and can be grown only in outdoor conditions. Lack of genetic variation and low adaptability to changing environmental conditions and limited knowledge of its host plants are the key bottlenecks towards its domestication.

The role of gut microbiota in host metabolism, growth, development, nutrition and immune regulation is becoming increasingly evident in humans and other organisms. In insects, the intricate relationship between beneficial microbes and host organisms has also played a pivotal role in their evolutionary success and diversification. Microbial colonization inside the gut of Lepidoptera is a matter of dispute likely because of their holometabolous nature, highly basic gut environment and simple structure of the gut facilitating faster food transit. However, previous studies revealed that Lepidoptera allows the inhabitation of gut microbiota which influences their vital physiological functions like nutrient acquisition, detoxification of plant secondary metabolites, and defense against pathogens.

To date, the gut microbiota of *A. assamensis* has remained largely unexplored. Culture-dependent studies followed by taxonomic identification through 16S rRNA gene sequencing have found that the most dominant bacteria in the gut of *A. assamensis* include *Bacillus* (54%), *Serratia* (24%), *Pseudomonas* (10%), and *Alcaligenes* (6%). Pathogenicity assessment of *A. assamensis* gut microbiota revealed the presence of *Pseudomonas aeruginosa*, *Ornithinibacillus bavariensis*,

Achromobacter xylosoxidans and *Staphylococcus aureus* in both healthy and diseased larvae. However, no in-depth investigation at the meta-omic level had been carried out to study the composition and functional role of the gut microbial community of *A. assamensis*. Based on the literature review, the following objectives were formulated to obtain detailed knowledge of the gut microbial community of *A. assamensis*, its functional role and its interaction with the host organism. Additionally, untargeted metabolomic analysis was carried out for identifying the global metabolome of *A. assamensis*.

1. **Objectives of the work**

Based on the identified gaps in the literature, the broad objectives of the Ph.D. thesis have been outlined as follows:

1. Structural analysis of the gut microbial community of *A. assamensis* and its comparison with the host leaf-associated microbiota.
2. Functional analysis of the gut microbial community of *A. assamensis*.
3. Gene expression profiling of the gut microbial community of *A. assamensis* and to identify the active gut microbiota.
4. Untargeted metabolomics analysis of the larval gut, host leaf and silk gland of *A. assamensis* to identify the global metabolites and their variation with change in the host plant.

Organization of the Thesis

The thesis is organized into the following six chapters:

Chapter 1: Introduction and Literature Review.

Chapter 2: Structural analysis of *A. assamensis* gut microbiome.

Chapter 3: Functional analysis of *A. assamensis* gut microbiome.

Chapter 4: Gene expression profiling of *A. assamensis* gut microbiome.

Chapter 5: Untargeted metabolomic analysis of *A. assamensis.*

Chapter 6: Summary and prospects

Chapter 1: Introduction and Literature Review

The initial part of this chapter presents a summary of the studied organism (*Antheraea assamensis* Helfer) and its ecological, scientific, socioeconomic and global importance followed by the key bottlenecks towards its domestication. The next part focuses on available knowledge on the role of gut microbiota in host metabolism, growth, development, nutrition and immune regulation in humans and other organisms and the research gap in the studied organism. Then the chapter explains the meta-omic techniques, viz., metagenomics, metatranscriptomics and metabolomics and how they could be utilized to unravel the interaction between the gut microbial community and the host organism at the genomic, transcriptomic and metabolic levels. The last part of this chapter describes the novelty of the work and how it would be helpful to the scientific community.

Chapter 2: Structural analysis of *A. assamensis* gut metagenome

This chapter reports the gut microbial community of *A. assamensis* reared separately on three different host plants and their host leaf-associated microbiota identified and characterized through amplicon sequencing of the 16S rRNA gene. Further, the impact of host leaf-associated microbiota on larval gut microbial composition, their possible functional roles and the effect of host plant leaf-associated microbiota on the gut microbial composition of *A. assamensis* was investigated. The results have indicated that the 5th instar larva of *A. assamensis* contains a highly diverse gut microbial community in terms of bacteria and archaea. Bacterial phyla like Proteobacteria, Chloroflexi, Planctomycetes, Patescibacteria, Firmicutes, Acidobacteria, Bacteroidota and Actinobacteria were predominant in the larval guts. The archaeal community of the gut primarily consisted of species from the phyla Thermoplasmata, Crenarchaeota, Nanoarchaeota, Halobacteria, and Euryarchaeota. Our study has identified 95 core operational taxonomic units (OTUs) present in the gut of *A. assamensis* larvae irrespective of their host plants. Comparative analyses of the gut and host leaf-associated microbiota showed that diet significantly affected the structure of the gut microbial community of *A. assamensis.* Finally, a total of ≃1910 enzymes were

predicted to be present through PICRUSt2 analysis in both the gut and leaf samples from which 420 MetaCyc pathways were inferred by mapping the gene families to pathways.

Chapter 3: Functional analysis of *Antheraea assamensis* gut metagenome

This chapter presents a detailed functional investigation of the gut metagenome of *A. assamensis* Helfer through shotgun metagenomic sequencing. *De novo* Metagenome assembly of 3.38×10^8 base pairs (bp) utilizing megahit v1.2.9 with *k-mer* length between 21-141 resulted in 4.2×10^5 contigs with minimum and a maximum length of 200 bp and 2.4×10^4 bp, respectively. Functional analysis of the assembled sequences with Prokka identified 72063 CDSs, 277 tRNAs and 1185 rRNAs. From the CDSs, 270 enzymes were identified with their EC numbers. The identified enzymes were used for metabolic pathway identification against MetaCyc and KEGG databases using MinPath which resulted in ~405 and 100 pathways, respectively. The coding regions were translated into protein sequences by Prokka which were then annotated against the Clusters of Orthologous Genes (COG) database using The WebMGA web server. This resulted in 1003 orthologous genes belonging to 24 COG categories. Taxonomic analysis of the shotgun data was performed using Kraken2 which revealed the presence of ~1000 microbial taxa, many of which were not detected or identified by the amplicon sequencing approach. 97% of all identified taxa consisted of the kingdom Bacteria, 2.4% Archaea and 0.06 % Viruses. The findings of this investigation will enable us to define the gut microbial population with more precision and comprehend its possible functions and interactions with the host organism.

Chapter 4: Gene expression profiling of *Antheraea assamensis* gut metagenome

Metatranscriptomics is a rapidly growing field in molecular biology that involves the study of the RNA transcripts of all the microbial organisms in a particular environment. Unlike metagenomics, which focuses on the DNA sequences of an entire microbial community, metatranscriptomics provides a snapshot of the gene expression patterns of these microorganisms. The key objective of this chapter is to identify the metabolically active gut microbial community of *A. assamensis* and their gene expression profiling through mRNA sequencing. For the metatranscriptome

analysis, total RNA from the gut of the 5th instar larvae of *A. assamensis* was isolated. polyA-mRNA enriched libraries were prepared with the help of TruSeq RNA Library Preparation Kit and sequenced. Metatranscriptome annotation of the assembled transcripts resulted in the identification of ~7000 bacterial and 2708 archaeal active protein-coding genes. Taxonomic annotation of the assembled transcripts identified 1091 bacterial and 490 archaeal taxa at the genus level. In this analysis, several key metabolically active functions have been annotated to the gut microbiota which would expand our knowledge towards the understanding of the host-microbiome interaction of *A. assamensis* larvae.

Chapter 5: Untargeted metabolomic analysis of *Antheraea assamensis*

Untargeted metabolomics of insect gut refers to the comprehensive analysis of the metabolites present in the gut of insects without prior knowledge of their identity through techniques like NMR, LCMS/MS and HRMS. This chapter describes the global metabolome of the host plant leaf, larval gut and silk gland of 5th instar larvae of *A. assamensis* and how the change in the host plant affects the gut and silk gland metabolite composition. For this, larvae reared on Som and Mejankari plants were considered. The results of this experiment identified hundreds of metabolites from the larval gut, silk gland and host plant leaf of *A. assamensis* larvae reared on the two host plants. The identified metabolites can be clustered into different categories like amino acids, sugars, fatty acids, alkaloids, secondary metabolites, pigments vitamins and others. Change in the host plant was found to change the metabolite composition of the larval gut and the silk gland. The finding of this study will help to extend our knowledge of the global metabolomes of the studied samples.

Chapter 6: Summary and prospects

Amplicon analysis of the gut microbial community of *A. assamensis* larvae reared separately on the host plants *M. bombycina, L. polyantha* and *L. citrata* showed a diverse gut microflora dominated by phylum Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria, Firmicutes, Planctomycetes, Patescibacteria, Fusobacteriota, Bacteroidota and Verrucomicrobiota. Archaeal species belonging to five different phyla viz., Thermoplasmatota, Nanoarchaeota, Halobacterota,

Euryarchaeota and Crenarchaeota occupy a significant proportion (~10%) of the total gut microbial community identified in the larval gut of *A. assamensis*. Shotgun metagenomic sequencing of the gut of *A. assamensis* larvae detected the presence of additional bacterial phyla like Armatimonadetes, Deinococcus-Thermus, Tenericutes, Spirochaetes, Nitrospirae, Thermotogae, Candidatus Saccharibacteria, Aquificae, Elusimicrobia, Thermodesulfobacteria and Chlorobi. In our investigation, it was found that a considerable fraction of the gut microbial community of the larvae contains the host leaf-associated microbiota. However, the gut microbiota of *A. assamensis* was found to be richer in terms of diversity in comparison to the host leaf-associated microbes. Functional analysis through predictive and shotgun metagenomics showed the presence of genes responsible for several key physiological functions of the larva such as host leaf digestion, chitinase production, detoxification of plant secondary metabolites and fat body metabolism. Metatranscriptomic analysis of larval gut microbiota reported several metabolically active functions with potential benefits towards the host organism such as food digestion, detoxification of plant secondary metabolite, resistance to the oxidative stress inside the gut, and vitamin biosynthesis. This study also revealed that *Acinetobacter*, *Streptomyces*, *Ruminococcus*, *Pseudomonas*, *Paenibacillus*, *Pedobacter*, *Rhizobium* and *Sphingomonas* are the most active microbes inhabiting the gut of 5th instar larvae of *A. assamensis*.

Elaborate multi-omic research considering spatiotemporal, sex, instar, and disease-related variation, as well as vertical and horizontal transmission of the gut microbial community of *A. assamensis* would help in a better understanding of the holobiont. Also, the identification and functional characterization of the fungal and single-celled eukaryotes would add a new dimension to the existing knowledge about host-microbial interaction in the future. Comparison of the gut microbiota of this endemic silkworm species with other allied species from different geographical locations would also help in understanding the potential role of gut microbiota in endemism and species conservation.

Publications:

Research Paper

1. Metagenomic and metatranscriptomic insights into the structure and function of the gut microbial community of *Antheraea assamensis* Helfer (Under review).

Book Chapter:

1. Ojha, Sunita, **Dharitri Saikia,** and Utpal Bora. "Nanopharmaceuticals: Synthesis, Characterization, and Challenges." *Nanopharmaceuticals: Principles and Applications Vol. 3*. Springer, Cham, 2020. 81-138.
2. Chetia, Hasnahana, Debajyoti Kabiraj, Biju Bharali, Sunita Ojha, Manash Pratim Barkataki, **Dharitri Saikia**, Tinka Singh, Ponnala Vimal Mosahari, Pragya Sharma, and Utpal Bora. "Exploring the benefits of endophytic fungi via omics." In *Advances in Endophytic Fungal Research*, pp. 51-81. Springer, Cham, 2019.

Conferences:

1. Kalita, J. J., Kabiraj, D., Bharali, B., Saikia, D., Bora, U. Applications of Aptamer conjugated nanomaterials in biosensing and therapeutics. International Conference on Advances in Nanotechnology (ICAN-2017), January, 2017.
2. Submitted abstract and poster at conference on “Evolutionary Emergence of Life Cycles”, Max Planck Intstitute for Evolutionary Biology, Germany, October 2018.
3. Submitted abstract on “Traditional Practices of Assam: A conservation Perspective” in Biodiverse-2018, IIT Guwahati.
4. Poster presented at SeriTech 2017 at IIT Guwahati.
5. Best oral presentation award (technical session 2, venue 4) on the topic “Understanding the Gut microbial Community of Lepidoptera through Meta-omics” in “International Conference on Challenges and Prospects of Bioresource Conservation in Eastern Himalaya with Special Reference to UN-Sustainable Development Goal” organized by Department of Zoology, Gauhati University. 3-4 May 2023.

Abbreviations

bp	:	base pair
CDS	:	Coding Sequences
EC	:	Enzyme Commission
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
COG	:	Clusters of Orthologous Genes
DNA	:	Deoxyribonucleic acid
RNA	:	Ribonucleic acid
mRNA	:	Messenger RNA
NMR	:	Nuclear Magnetic Resonance
LCMS	:	Liquid Chromatography and Mass Spectrometry
HRMS	:	High-Resolution Mass Spectrometry
NCBI	:	National Center for Biotechnology Information
LC-HRMS	:	Liquid Chromatography-High Resolution Mass Spectrometry
rRNA	:	Ribosomal RNA
OTU	:	Operational Taxonomic Unit
PCoA	:	Principal Coordinate Analysis
Min	:	Minute
in	:	Inch
cm	:	Centimeter
mm	:	Millimeter
ml	:	milliliter
µm	:	Micrometer
µL	:	Microliter
mg	:	Milligram
L	:	Liter
GoI	:	Government of India
Vern. Name	:	vernacular name
Tm	:	Melting temperature
˚C	:	Degree Celcius
ASV	:	Amplicon Sequence Variants
rDNA	:	Ribosomal DNA
D	:	Dextro
L	:	Levo
GC-MS	:	Gas Chromatography-Mass Spectrometry

NGS	:	Next Generation Sequencing
M	:	Molar
PBS	:	Phosphate Buffer Saline
PCI	:	Phenol Chloroform Isoamyl
NaCl	:	Sodium Chloride
EDTA	:	Etheylenediaminetetraacetic Acid
SDS	:	Sodium Dodecyl Sulphate
RPM	:	Revolutions per Minute
ng	:	nanogram
F	:	Forward
R	:	Reverse
PICRUSt2	:	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
QIIME2	:	Quantitative Insights Into Microbial Ecology 2
EPA-ng	:	Evolutionary Placement Algorithm-next generation
NSTI	:	Nearest Sequenced Taxon Index
k	:	Kingdom
p	:	Phylum
c	:	Class
f	:	Family
g	:	Genus
Hg	:	Mercury
ROS	:	Reactive Oxygen Species
CMERTI	:	Central Muga Eri Research and Training Institute
PCGs	:	Protein Coding Genes
Ppm	:	Parts per million
TOF	:	Time of Flight
QTOF	:	Quadrupole Time of Flight
m/z	:	Mass/Charge
RT	:	Retention Time

List of Tables

Table 1. 1 Statistics of Silk production in India (Source: Central Silk Board, GoI) 6
Table 1. 2 The rearing of Muga silkworm 7

Table 2. 1 Details of the sequencing technique of the studied samples. 57
Table 2. 2 Quality control parameters for DADA2 58
Table 2. 3 Description of the alpha diversity indices used in this study 59
Table 2. 4 Details of the samples used in the experiment and sequence statistics. 63
Table 2. 5 Identified taxa in the larval gut and host plant leaf of *A. assamensis* against the NCBI database with 97% similarity. 64
Table 2. 6 Bacterial taxa identified in the larval gut and host plant leaf of *A. assamensis* against the Greengenes database with 97% similarity. 66
Table 2. 7 Alpha diversity indices for the larval gut and host plant leaf samples 79
Table 2. 8(a) Pairwise Kruskal-Wallis test for Shannon's index between the larval gut and host plant leaf samples. 80

Table 3. 1 Basic statistics of the quality-controlled shotgun metagenomic sequences 115
Table 3. 2 Statistics of the shotgun sequence data. 117
Table 3. 3 Classification of orthologous genes identified through WebMGA webserver into different COG categories. 120
Table 3. 4 MetacCyc pathways and the number of gene families identified in the gut metagenome of *A. assamensis* 122
Table 3. 5 List of KEGG pathways identified in the functional gut metagenome of Antheraea assamensis 125

Table 4. 1 Details of the metatranscriptome assembly with Megahit. 160
Table 4. 2 Functional categories and their abundance of the gut metatranscripts identified against SEED Subsystem Database. 162
Table 4. 3 Highly expressed microbial genes identified in the gut metatranscriptome of *A. assamensis.* 165
Table 4. 4 Key enzymes identified in the gut metatranscriptome of *A. assamensis* 168
Table 4. 5 SEED Subsystem annotation of the transcripts involved in the biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments. 174
Table 4. 6 SEED Subsystem annotation of the larval gut metatranscripts involved in oxidative

stress response 179
Table 4. 7 List of expressed genes and the pathways involved in the metabolism of aromatic compounds inside the larval gut 182
Table 4. 8 Key enzymes involved in methanogenic pathways produced by the gut methanogens 193
Table 4. 9 List of major genes expressed by the gut archaeal community of *A. assamensis*..... 193

Table 5. 1 Details of the sample used for LC-HRMS analysis and the experimental parameters. 230
Table 5. 2 Details of the unique features identified in each sample..... 231
Table 5. 3 The major amino acid compounds identified in different samples 238
Table 5. 4 The major sugar compounds identified in different samples 240
Table 5. 5 Major pigment molecules identified in the metabolomes of the studied samples 242
Table 5. 6 Some of the important identified metabolites belonging to different categories. 244

List of Figures

Figure 1. 1 Lifecycle of moths and butterflies. 2
Figure 1. 2 The Milestones in the development of Metagenomics 14
Figure 1. 3 The variable and conserved regions of 16S rRNA gene 15

Figure 2. 1 Overview of the experimental design 55
Figure 2. 2 The workflow for amplicon sequence analysis 61
Figure 2. 3 Number of identified phyla against three different databases at the Phylum level. ... 62
Figure 2. 4(a) Relative abundance of the gut-associated bacterial and archaeal community of *A. assamensis* larvae reared on three different host plants at the phylum level identified against the Silva database (*p: phylum*). 70
Figure 2. 5(a) Relative abundance of the leaf-associated bacterial and archaeal communities of three different host plants of *A. assamensis* at the phylum level identified against the Silva database (*p: phylum*) 73
Figure 2. 6 Venn diagram depicting the shared and unique OTUs (a) among the larval gut samples MB-G, LC-G and LP-G; (b) among the leaf samples of three different host plants, LC-L, MB-L and LP-L; (c) between the larval gut and its host plant leaf. 76
Figure 2. 7 Alpha and beta diversity analysis of the larval gut and host plant leaf samples; (a) Rarefaction plot of the observed features, (b) alpha-diversity analysis measured by Faith's PD matrix, (c) Principal Coordinate Analysis on the unweighted UniFrac distance matrix. 78
Figure 2. 8 A rooted phylogenetic tree of the gut microbial community of *A. assamensis* with phylum level annotation as identified through 16s rRNA metagenomic sequencing. . 79
Figure 2. 9 Heat map showing the relative abundance of selected microbial enzymes in the predicted functional metagenome of larval gut and host plant leaf samples as identified through PICRUSt2 analysis. 81
Figure 2. 10 Relative abundances of selected MetaCyc pathways identified in the predicted functional metagenome of the larval guts and host plant leaves through PICRUSt2 analysis. 83
Figure 2. 11(a) Differential abundance of MetaCyc pathways between MB-G and MB-L. The first column represents the metabolic pathways, the second column represents the proportions of the pathways in each MB-G and MB-L, the third column represents the difference between proportions in percentage at 95% CI and the fourth column

represents the corrected *p*-value. 84

Figure 3. 1 Per base sequence quality of the quality-controlled shotgun metagenomic sequences 116

Figure 3. 3(a) Major gene encoding enzymes identified in the gut metagenome of *A. assamensis* through shotgun metagenomic sequencing. 118

Figure 3. 4 COG classes of the pathways identified. The coding regions were translated into protein sequences by Prokka which were then annotated against the Clusters of Orthologs (COG) database using WebMGA web server. This resulted in 1003 orthologous genes belonging to 24 COG categories as represented here. 121

Figure 3. 5 Visualization of the metabolic interaction network map based on KEGG pathways using Interactive Pathways Explorer 3 (iPATH3). (a) A total of 102 KEGG pathways were identified which were mapped against the metabolic interaction network. The nodes represent chemical compounds and the edges represent enzymatic reactions on the map. (b) Few key pathways identified in the gut metagenome of *A. assamensis* have been highlighted. 129

Figure 3. 6 The major bacterial taxa identified in the gut metagenome of *A. assamensis* larvae through shotgun sequencing. 131

Figure 3. 7 The major archaeal taxa identified in the gut metagenome of *A. assamensis* larvae 132

Figure 3. 8 Methanogens identified in the gut metagenome of *A. assamensis* larvae and their abundance 133

Figure 3. 9 The viral taxa identified in the gut metagenome of *A. assamensis* with their abundance. 134

Figure 4. 1 The workflow for metatranscriptome data analysis 159

Figure 4. 2 SEED Subsystem hierarchy of the annotated transcripts. The inner circle (or level 3) represents the expressed genes in the gut metatranscriptome belonging to higher SEED hierarchy levels such as level 2 (the middle circle) and level 1 (the outer circle). 161

Figure 4. 3 The enzyme classes responsible for carbohydrate metabolism and their abundance. The class hydrolases were found to be the highest expressed followed by oxidoreductases, transferases, isomerases and ligases. 164

Figure 4. 4 SEED Subsystem annotation of active genes involved in the biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments at hierarchy level 1. 173

Figure 4. 5 The major functional categories of the expressed genes involved in stress response.

..... 178
Figure 4. 6 The major active pathways involved in the metabolism of aromatic compounds inside the gut of *A. assamensis* larvae. 181
Figure 4. 7 The active gut microbial community of *A. assamensis and* their relative abundances 189
Figure 4. 8 The major genes expressed by *Acinetobacter baumannii* inside the gut of *A. assamensis* larvae..... 190
Figure 4. 9 The major genes expressed by *Lactobacillus* sp. inside the gut of *A. assamensis* larvae. 190
Figure 4. 10 The major genes expressed by *Serratia* sp. inside the gut of *A. assamensis* larvae. 191
Figure 4. 11 The active gut archaeal community of *A. assamensis* and their relative abundance 192

Figure 5. 1 Untargeted LC-MS Workflow 226
Figure 5. 2 Two-dimensional plot visualizing the base intensity peak/retention time of the raw data. All the chromatograms for the six samples are displayed in the same plot. The x-axis corresponds to retention time and the y-axis is the intensity level of the signal for each sample. 231
Figure 5. 3 Principal component analysis of the studied samples based on the identified unique features. The closer the samples placed to each other, the higher is the similarity between them. The PCA analysis showed that M-L shared the highest similarity with S-L, M-G shared with S-G and S-SG shared with M-SG. 232
Figure 5. 4 Histogram showing the frequencies of features in different m/z ranges. This data is generated from raw data with a range from 50 to 2000 m/z..... 233
Figure 5. 5 Scatter plot displaying each unique feature as a dot on a two-dimensional graph, with one sample represented on the X-axis and another represented on the Y-axis on a log scale. The position of each blue dot on the plot corresponds to the values of the two samples for that particular feature. The straight lines (fold lines) indicate a 2-fold difference, which is a measure of the difference in the abundance of a feature between two samples, expressed as a ratio. The fold lines represent the threshold value above which the features are considered significantly different between the two samples. The scatter plots show the relationship among the unique features detected in (a) the gut of *A. assamensis* larvae reared on Mejankari plant (M-G) and host plant (Mejanakri) leaf (M-L), (b) gut of larvae reared on Som plant and leaf of Som plant, (c) Gut of larvae

reared on Som plant and gut of larvae reared on Mejankari plant, (d) leaf of Som plant and leaf of Mejankari plant, (e) silk gland of larvae reared on Som plant and silk gland of larvae reared on Mejankari plant, (f) silk gland of larvae reared on Mejankari plant and leaf of Mejankari plant. The two silk gland samples, i.e., silk gland of larvae reared on Som plant and Mejankari plant showed the highest similarities among their unique features..... 235

Figure 5. 6 The major classes of the identified metabolites across the samples. The highest abundant metabolites were amino acids followed by sugar molecules and fatty acids. Pigment molecules, neurotransmitters, steroids, and vitamins were also annotated from some of the samples..... 236

Figure 5. 7 Number of shared identified metabolites across the samples. 237

Chapter 1

Introduction and Review of Literature

CHAPTER 1

Introduction and Review of Literature

1.1 Lepidoptera

The order Lepidoptera is one of the largest insect groups that diverged from the order Trichoptera during the Permian or Late Triassic period (van Eldijk et al., 2018a). The etymology of the term Lepidoptera roots from the ancient Greek word *Lepido* meaning scale and *petron* meaning wing (Perveen & Perveen, 2017) which refers to insects that bear wings covered with tiny scales. Based on recent research employing cladistics, morphological traits and other evolutionary data, this order has been separated into four suborders: Zeugloptera, Aglossata, Heterobathmiina, and Glossata. Only one nominotypical family of primitive moths is present in each of the first three suborders whereas the rest of the orders are classified under Glossata (Heppner, 2002). Lepidoptera contains around 157,424 existing species belonging to 43 superfamilies and 133 families (Mitter et al., 2017a). Based on their morphology, anatomy, behaviour and ecology, Carl Linnaeus divided this group into three broad categories namely, 1. micro and macro moths, 2. butterflies and 3. skippers (Perveen & Perveen, 2017). The key characteristics of all the Lepidopteran species include scaly wings, and a three-segmented body, each segment containing a pair of well-developed legs with joints and a pair of antennae. Lepidopteran species are incredibly diverse in terms of their appearance, biology, behaviour, and dietary preferences. They may be found in practically every known environment, from the arctic to the densest forests. However, no species of Lepidoptera have been reported from Antarctica to date, while a few tiny brachypterous species could exist on the northernmost point of the continent closest to Tierra del Fuego (Convey, 2005; Heppner, 2002). Like other insects, Lepidopteran species are holometabolous in nature which means they undergo complete metamorphosis. Their life cycle includes four stages namely, egg, caterpillar or larval stage, pupa or chrysalis and adult or imago stage (**Figure 1. 1**).

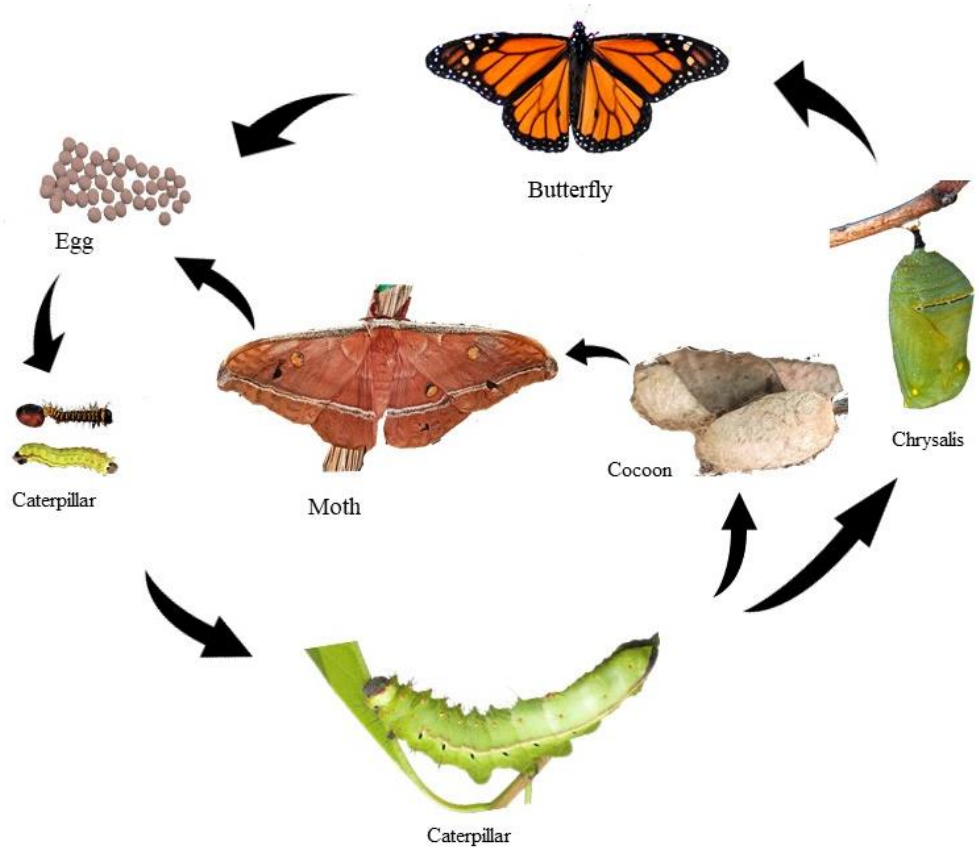


Figure 1. 1 Lifecycle of moths and butterflies.

The butterflies and moths are of great economic importance. Some species of butterflies are important pollinators for crops such as tomatoes, blueberries, and almonds (Alison et al., 2022; Singh et al., 2022; Walton et al., 2020). Their role in pollination is estimated to be worth billions of dollars annually. Certain lepidopteran species are considered pests that can damage crops (Klem & Zaspel, 2019; Suckling et al., 2017). Other moth species are used as biological control agents to manage crop pests. For example, the diamondback moth is a significant pest of cabbage and other cruciferous crops, while the parasitic wasp *Cotesia plutellae* is used to control the diamondback moth population (Sarfraz et al., 2007). Several moth species, including the domesticated silkmoth, are used to produce silk, which is an important textile material (Zurovec et al., 2002). The silk industry has a long history of contributing to the economy, providing income and employment opportunities to millions of people (Bharathi, 2016). The beauty and diversity of butterflies attract tourists to butterfly gardens, butterfly houses, and butterfly sanctuaries, which generate income for local economies (GENÇ et al., 2021; Sih Kurnianto et al., 2016). Lepidoptera is studied extensively by scientists to better understand their ecology, behavior, and physiology. This research can inform

conservation efforts, and contribute to the development of new agricultural technologies, silk production and medical applications of silk in tissue engineering (Hill et al., 2021; Pierce et al., 2002; Vepari & Kaplan, 2007; Zalucki et al., 2003).

1.1.1 Feeding habits of Lepidoptera

Lepidopteran insects have a wide range of feeding behaviour, which vary according to how well-suited a species or group is to a certain climate, location, types of available food plants, feeding strategy, gustatory, olfactory and other aspects (Song et al., 2021). Conifers and flowering plants make up the vast bulk of food sources, although some groups also consume more rudimentary plants including mosses, liverworts, ferns, and some lichens (Mejia et al., 2020). Between the adults and larvae of Lepidoptera, there are also striking differences in lifestyle. For instance, whereas the majority of adult butterflies seek nectar and the majority of adult moths at least look for moisture (although most of the moth species do not eat at the adult stage), the larvae are mostly phytophagous. Numerous caterpillars with specialised adaptations consume almost the entire plant. (Krenn, 2010) Many larvae, notably belonging to the plume moth's family (Pterophoridae), consume flowers (Frank et al., 2008), while many adults drink the nectar. Others, like yucca moths (family Incurvariidae), leaf roller moths (family Tortricidae), and several owlet moths, consume cones, fruits, and seeds. Most larvae use their biting and chewing mouthparts to consume plant matter. The majority of adults are anthophilous; they have an elongated appendage called proboscis that they utilise to drink liquids, including flower nectar (Krenn, 2010).

Caterpillars are well-known for their voracious eating habit. Most often, they consume the leaves of various plants, while some species like apefly larvae consume bugs and other tiny animals (Saikia et al., 2021). Fruit trees, vegetables, decorative plants, hardwood trees, and shrubs are just a few of the plants that leaf-eating organisms can seriously harm. For instance, the cabbage looper moth (*Trichoplusia ni*) caterpillars may eat up to three times their body weight in leaf material each day. The faecal substance they create, known as frass, can discolour leaves and make plants unsellable in addition to the harm these caterpillars do by consuming the leaves of cabbages and

related crops (Dangles et al., 2009).

1.1.2 The Saturniids

The family Saturniidae or the saturniid contains more than 2000 identified species (Nieukerken et al., 2011). This group is well known for having some of the largest species of moths found around the globe including *Citheronia regalis*, *Hyalophora Cecropia*, *Antheraea polyphemus*, *Antheraea assamensis* and *Actius luna* (Heppner, 2002). Adults have large bodies covered with hairlike scales, huge, lobed wings, and shortened mouthparts. Although they lack a frenulum, the appearance of an uninterrupted wing surface is created by the way the hindwings overlap the forewings. The eyespots or "windows" on the wings of saturniids are frequently transparent and can occasionally be vividly coloured. The degree of sexual dimorphism varies from species to species, but in general, males may be distinguished by their bigger, wider antennae. The majority of mature insects have wingspans of 1-6 inches (2.5-15 cm), however certain tropical species, such as the Atlas moth (*Attacus atlas*), can reach wingspans of 12 inches (30 cm).

1.2 *Antheraea assamensis* Helfer: The Muga Silkworm

Antheraea assamensis Helfer, commonly known as the Muga silkworm, is an endemic, economically important, Lepidopteran species explicitly found in the northeastern part of India and its neighbouring areas (Tikader et al., 2013). This semi-domesticated, multivoltine, polyphagous insect produces an economically important, lustrous golden coloured silk with high tensile strength and durability (Tikader et al., 2013) which is used to produce high-quality silk fabric.

The Muga silk industry is one of the most important cottage industries of Assam that has been in existence since time immemorial with no precise reference to its origin (Phukan, 2012). The Ahom dynasty (1228-1826) is considered to be the era of prosperity and proliferation for this industry of Assam. Currently, about 90% of the total Muga silk produced in India comes solely from this state (Borah & Borgohain, 2017). Muga silk has significant cultural importance in Assamese society, where it is considered a symbol of social status, wealth, and cultural identity. The fabric is often

used in traditional Assamese dresses, such as Mekhela Chador and Gamosa, and is an integral part of the Assamese cultural heritage (Mahan, 2012). As per the reports of the Central Silk Board, although India has seen a significant rise in raw Muga silk production in recent years, this contributes only a small fraction of the total silk production of the country (**Table 1.1**). The Muga silk industry is facing challenges due to climatic changes, pathogenic attacks, scarcity of land for growing the host plants, availability of disease-free seeds etc. The application of pesticides and insecticides in the tea gardens near Muga farms also causes immense crop loss. According to reports, Muga silk is superior in quality to cultivated mulberry silk due to its more alanine content than glycine, increased resistance towards oxidizing agents and greater content of diamino acids and dicarboxylic acid (Mohanty et al., 2006). Due to its extreme durability, magnificent colour and luster, Muga silk is one of the most expensive natural fibers in the world (Kalita & Dutta, 2020; Singh et al., 2012). The fibroin part of the Muga silk is also a promising material for tissue engineering (Das et al., 2017; Kar et al., 2013). Unlike mulberry-feeding silkworms, *A. assamensis* is semi-domesticated and can be grown only in outdoor conditions. Lack of genetic variation and low adaptability to changing environmental conditions and limited knowledge of its host plants are the key bottlenecks towards its domestication (Tikader et al., 2013).

The production of Muga silk is dependent on the Muga silkworm which feeds on the leaves of the Som and Soalu trees. However, climate change has had a significant impact on the Muga culture in recent years. One of the primary impacts of climate change on Muga culture has been the changing weather patterns. Erratic rainfall and prolonged dry spells have led to a decline in the availability of the Som and Sualu trees, which are the primary food source for the Muga silkworms. This has resulted in a decrease in the production of Muga silk. Moreover, rising temperatures have also led to the proliferation of pests and diseases, which have affected the health of the Muga silkworms. This has further impacted the production of Muga silk. Additionally, the increase in the frequency and intensity of natural disasters such as floods, landslides, and droughts have also affected the Muga culture. These disasters have caused damage to the Muga plantations and the

silkworm-rearing houses, resulting in a loss of livelihood for the farmers and weavers involved in the industry (Das, 2021; Ram et al., 2016).

Table 1. 1 Statistics of Silk production in India (Source: Central Silk Board, GoI)

Particulars	**2016-17**	**2017-18**	**2018-19**	**2019-20**	**2020-21**	**2021-22**
Mulberry Silk Production (MT)	21273	22066	25344	25239	23896	25818
Tasar Silk Yarn (MT)	3268	2988	2981	3136	2689	1466
Eri Spun Yarn (MT)	5637	6661	6910	7204	6946	7364
Muga Raw Silk (MT)	170	192	233	241	239	255

1.2.1 Distribution of *Antheraea assamensis*

A. assamensis is found in the Brahmaputra valley in Assam; Garo hills in Meghalaya; Kohima, Tuensung, Mokokchung, and Wokha districts of Nagaland; Dibang and Lohit valleys, Pampumpare and Changlang districts of Arunachal Pradesh; Tamenglong district of Manipur and the Coochbehar district of West Bengal (Tikader et al., 2013). It is also known to occur in Myanmar, Kangra and Kumaon Valleys of Himalaya, Sundaland and Sri Lanka (Tikader, 2013).

1.2.2 The Host plants of *A. assamensis*

A. assamensis is a polyphagous insect feeding primarily on *Machilus bombycina* King ex Hook. f. (synonym: *Persea bombycina* (King ex Hook. f.) Kost; vern. name: Som plant) and *Litsea polyantha* Juss. (Synonym: *Litsea monopetala* (Roxb.) Pers.; vern. name: Soalu plant). On the other hand, *Litsea citrata* Blume (Synonym: *L. cubeba* (Lour.) Pers; vern. name: Mejankari) is a secondary host plant of *A. assamensis*. Interestingly, larvae of *A. assamensis* reared on *L. citrata* produces a shining white coloured silk of superior quality instead of the typical golden colour. This species can also thrive on other plants like *Litsea salicifolia* Hook (Dighloti), *Ziziphus jujuba* Mill (Bogori), *Symplocos grandifolia* Wall. *(Bhomloti), Cyclicodaphne nitida* Roxb. (Katholua) and *Michelia champaca* Linn. (Champa) (Devi et al., 2021).

1.2.3 Rearing of *A. assamensis*

The multivoltine Muga silkworm produces 5–6 crops per year, including two commercial harvests namely, Jethua (May–June) and Kotia (October–November); two pre-seed crops, Jarua (December–January) and Aherua (June–July), and two seed crops, Chotua (February–March) and Bhodia (July–August) as described in **Table 1.2**. Pre-seed and seed crops frequently experience unfavourable climatic conditions that cause significant loss during the early stages as a result of environmental challenges, disease incidence, and insect and predator infestation.

Table 1. 2 The rearing of Muga silkworm

Name of the Crop	Month of Rearing	Crop Type
Jarua	December-February	Pre-seed
Chotua	March-April	Seed crop
Jethua	May-June	Commercial crop
Aherua	June-July	Pre-seed crop
Bhodia	August-September	Seed crop
Kotia	October-November	Main Commercial crop

1.3 Gut microbiota, host organism and their interaction

The collection of the microbial community residing inside the gastrointestinal tract of an organism is termed the gut microbiota (Thursby & Juge, 2017). The gut microbiota plays an important role in host biology, impacting many aspects of the host's health and physiology (Jandhyala et al., 2015). For example, the gut microbiota is involved in the breakdown and absorption of nutrients, particularly complex carbohydrates, that are difficult for the host organism to digest on its own (Morowitz et al., 2011). It also produces vitamins, such as vitamin K and vitamin B, that are important for the host's health and can influence the metabolism and energy balance (Morowitz et al., 2011). There is growing evidence that gut microbiota can impact mental health and brain function. Some studies have shown that the gut microbiota may play a role in conditions such as

depression, anxiety, and autism (Carabotti et al., 2015; Morais et al., 2020). The microbial system living within and on a host is crucial for the development and functionality of the host immune system (Nicholson et al., 2012; Rooks & Garrett, 2016). The development of sophisticated technologies for profiling host microbiota has given a new dimension to the understanding of host-microbiota interactions. The microbial communities and their metabolic products are crucial for developing signals for immune functionality and homeostasis. Sometimes, they may also influence the susceptibility of the host to pathogens or immune-mediated diseases. Understanding host immunity and microbial mutualism can give important knowledge about the pathogenicity of diseases (Kim et al., 2017; Shibata et al., 2017). Most of the works on gut microbiota and host immunity have been done on mammalian systems. A healthy microbiota can help in the maturation of the host immune system through microbial surface antigens and metabolites. Mammals allow a diverse and metabolically active group of microbial communities to reside inside their gut and most of the time establish an enduring mutualistic relationship (Desselberger, 2018). Whenever this dynamic beneficial relationship goes wrong due to perturbation to the microbiota or the host organism, dysbiosis occurs (Carding et al., 2015). The altered microbial community can invite a successful attack of the invasive pathogen. Like other metazoans, insects also have a similar pattern of host-gut microbiome interaction (Engel & Moran, 2013). Recent studies have been done to understand the interaction between the insect immune system and gut microbiota taking *Drosophila melanogaster* as a model organism (Broderick et al., 2014). Microbes that enter the gut lumen constantly with ingested diet include both beneficial and harmful microbes as well as some opportunistic pathogens. Like other organisms, insects are also able to discriminate between self and non-self with the help of the immune system. Insects lack adaptive immunity and have only an innate immune system. Genetic analysis of *Drosophila* has revealed the existence of two conserved innate immune systems: The Toll and IMD pathways (Tanji & Ip, 2005).

1.3.1 The gut microbiota of Lepidoptera

The order Lepidoptera under the class Insecta contains two broad groups, moths and butterflies. It

is a huge order containing 126 families, 46 superfamilies and approximately 180 000 species. Chen and his co-workers investigated the gut microbiota of *Bombyx mori* (Chen et al., 2018) and compared its community composition with two natural feeders of the mulberry plant, *Acronicta major* and *Diaphania pyloalis*. They have found that the gut bacterial community of all three samples is dominated by the active population of species belonging to phyla Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes. Species from phyla like Acidobacteria, Chloroflexi, Planctomycetes and Fusobacteria were also prevalent. *Pseudomonas, Pantoea* and *Acinetobacter* were found in all samples. A higher bacterial diversity in the foregut was observed as compared to the midgut and hindgut with 107 unique OTUs in the foregut and 80 and 74 unique OTUs in the midgut and hindgut, respectively. Variation in microbial composition and diversity was also observed throughout different developmental stages with 5th instar larvae having the maximum number of taxa. Even under the domesticated condition, the gut microbiota of *B. mori* was found to be highly diverse than its two wild relatives *A. major* and *D. pyloalis*. Another study on *Brithys crini*, a monophagous moth, revealed a developmental stage-specific variation in gut microbial composition with 29 conserved genera across the stages (Chen et al., 2018).

In some species like *Cnaphalocrocis medinalis* and *Grapholita molesta*, significantly high species richness is seen in the larval gut during the early instars (1st and 2nd) as compared to the late instars. The diversity and abundance in male pupa and adults have also been found to be higher than third to 5th instar larvae (Wang et al., 2020; Yang et al., 2020).

The role of gut microbiota in host metabolism, growth, development, nutrition and immune regulation is becoming increasingly evident in humans and other organisms (Flint et al., 2012). In insects, the intricate relationship between beneficial microbes and host organisms has also played a pivotal role in their evolutionary success and diversification (Janson et al., 2008). The favourable body parts for microbial colonization in insects are the gut, cuticle, intracellular space and hemocoel (Douglas, 2015). Microbes inhabiting the insect's gut are tolerant to both the host defense mechanism and the harsh gut environment which is often characterized by high alkalinity,

low oxygen and the presence of plant secondary metabolites (Engel & Moran, 2013b). Some microbes are selectively favoured by the host itself to reside inside their body and a biological system of commensalism between the host and residential microbes is established (Simhadri et al., 2017). The microbial community inside an insect's gut is mainly determined by diet, environmental factors, habitat, developmental stage and host phylogeny (Yun et al., 2014). These residential microorganisms can promote host fitness by providing essential amino acids, vitamins, sterols and other nutrients to the host organism. They can also protect their host from invading organisms such as pathogens and parasitoids by synthesizing toxins or modifying the internal environment of the host against them (Douglas, 2015).

As mentioned above, the order Lepidoptera started radiating from the order Trichoptera during the Permian or Late Triassic period (van Eldijk et al., 2018b) resulting in 157,424 existing species from 43 superfamilies and 133 families (Mitter et al., 2017b). Out of the total identified species of Lepidoptera, only a few (<0.1%) have been investigated for their gut microbial association (Voirol et al., 2018). Microbial colonization inside the gut of Lepidoptera is a matter of dispute (Hammer et al., 2017) likely because of their holometabolous nature, highly basic gut environment and simple structure of the gut facilitating faster food transit (Voirol et al., 2018). However, many studies have revealed that Lepidoptera allows the inhabitation of gut microbiota which influences their vital physiological functions like nutrient acquisition, detoxification of plant secondary metabolites (Jing et al., 2020; Xia et al., 2017), and defense against pathogens (Jakubowska et al., 2013; Shao et al., 2017). Studies have also established that a single bacterial inoculum is fairly adequate for establishing microbial colonization and multiplication inside the gut (Mason et al., 2020). The gut microbiome of Lepidoptera is composed of species mostly belonging to the bacterial families Enterobacteriaceae, Pseudomonadaceae and Bacillaceae with *Pseudomonas*, *Enterobacter*, *Bacillus, Staphylococcus* etc. being the most abundant genera (Voirol et al., 2018). Archaea are widespread microorganisms expanding from ordinary to extreme environments (Chaban et al., 2006). They are an essential gut inhabitant in many organisms including insects

(Engel & Moran, 2013a; Nkamga et al., 2017). However, only a few lepidopteran species have been explored for the gut archaeome (Ugwu et al., 2020; Ziganshina et al., 2018). Several analyses of the lepidopteran gut microbiome could not detect the presence of archaea (Chen et al., 2016; Tang et al., 2012). Major factors influencing the gut microbial composition of Lepidoptera include the age of the host organism, diet, and environment. However, the contribution of these factors toward the structure of the gut microbial community may not be equal and variation in one factor may change the effect of the other.

1.3.2 Phyllosphere microbiota of host plants of phytophagous insects

Likewise, a myriad of microorganisms like bacteria, fungi, viruses and protists are also associated with plants in their both above and below-ground parts (Hassani et al., 2018). Leaves are the most preferred phyllosphere region for microbial colonization. Their density on a leaf can range from 10^6 to 10^7 cells/cm^2 (Lindow & Brandl, 2003). Both epiphytic and endophytic microbial populations among and within plant species differ considerably based on nutrient composition, physical condition, morphology and anatomy of the plant parts and their carrying capacity (Lindow & Brandl, 2003). Microorganisms work as mediators in the tripartite trophic interaction among the host plant, their associated microbes and insect herbivores. Microbes associated with host plants can affect the host plant-herbivore interaction by increasing or decreasing the plant's fitness for the insect herbivore by altering their physiological, morphological and biochemical conditions. They often trigger defense activity of the plant against the insect herbivore by enhancing the emission of volatile compounds through jasmonic acid signaling, thereby inhibiting the insects from ingesting the host plant. On the other hand, the microbial communities associated with the insect herbivore help them in host plant selection and food ingestion by suppressing the defense mechanism of the host plant, neutralizing the effect of secondary metabolites, or increasing their ability to exploit host nutrients (Biere & Bennett, 2013).

1.3.3 Implication of gut microbiota for conservation

Functional analysis of the gut microbiome is also important for conservation biology. Although

changes in the microbial community affect the health and fitness of different species, studies on the implication of gut microbiota for conservation are of limited number. Species grown in laboratory condition or individual kept in captivity always carry fewer microbes than the wild and most of the time their growth and performance is reduced (Cheng et al., 2015; Kong et. al, 2014). Habitat fragmentation with limited resource availability affects the microbial diversity of the host (Amato et al., 2013; Barelli, 2015). Another factor that affects species' health, performance, disease resistance and reproduction is inbreeding (Keller & Waller, 2002). Expression of recessive deleterious mutations leads to inbreeding depression. In a population, the fitness of a species is affected by its genotype and its interaction with the microbiome. In inbred populations, individuals generally suffer from reduced microbial diversity, smaller gene pools and compromised immune systems (Ørsted et al., 2022).

The application of microbiome on tolerance to environmental perturbations is also getting importance (Engel & Moran, 2013b). Sometimes, the presence of single bacterial species, mainly endosymbionts has a large impact on the host in environmental stress tolerance (Feldhaar, 2011). Arthropods are the most species-rich (80% of all described living animal species) and successful phylum of the kingdom Animalia. Their incredible diversity and cosmopolitan existence are largely attributed to their short life cycle, genetic variability, generalization in food habits and ability to adapt and colonize in every possible habitat on Earth. Surprisingly this phylum has some species which are strictly confined to only some patches of the world and are highly vulnerable to changes in climatic and biotic factors. Our species of interest, *A. assamensis* belongs to those few species of Arthropoda which have a very narrow range of geographical distribution.

Recent research suggests that the gut microbiota may also contribute to endemism by playing a role in the coevolution of hosts and their microbial communities. For example, studies have shown that the gut microbiota of different animal species varies depending on their geographic location, suggesting that local environmental factors, such as diet and habitat, may shape the gut microbiota (Baldo et al., 2018; Härer et al., 2020; Rodpai et al., 2022; X. Wang et al., 2021). Moreover, studies

have also shown that certain gut microbiota species are specific to certain regions and may play a role in shaping the unique ecology of those regions. For instance, gut bacteria such as *Prevotella copri* have been associated with a high-fiber diet commonly found in rural areas and are more prevalent in certain indigenous populations, suggesting that the gut microbiota may contribute to endemism in some cases (Kovatcheva-Datchary, 2022).

1.4 The Developing Concept of Metagenomics

The invention of the microscope by Zacharias in 1590 and its improvement to the compound microscope by Galileo Galilei in 1609 is considered a milestone in microbial study. For almost 300 years after the discovery of the microscope, microbial studies and classification were mainly based on morphological features, growth behaviour and biochemical properties which had, though useful but only limited applications. The first microbial observations were conducted in the 1670s by Antony van Leeuwenhoek, who discovered microbes (which he named "animalcules") in saliva. Robert Koch and Julius Petri accomplished the first successful and isolation of bacteria for scientific purposes in the 1870s, even though humans have been selectively breeding bacteria and fungi for food fermentation for many centuries. Direct observation and culturing of microbes are both effective techniques although they both have drawbacks (Hugerth & Andersson, 2017). The culture-dependent method has been considered to be the gold standard technique for microbial characterization as it provides useful insights into bacterial morphology, biochemistry, physiology, and genetics. However, even Robert Koch could see that different bacteria have different growth requirements, and by the early 1900s, it was well acknowledged that the vast majority of bacteria could not be grown by conventional methods, a phenomenon subsequently became evident as the "great plate count anomaly" (Staley & Konopka, 1985). Winogradsky, a Russian microbiologist who is considered as the father of modern geomicrobiology, illuminated the concept of microbial ecology. He tried to culture Beggiatoa (in 1887), a colourless, filamentous proteobacterium that lives in a sulfur-rich environment, under conditions as close as its natural environment (Dworkin, 2012). Revolution came to microbial classification when Carl Woese proposed the use of 16S

rRNA as a tool to understand microbial classification and phylogeny (Woese & Fox, 1977). The universal presence of 16S rRNA in bacteria and archaea, its conserved function over a longer period in evolutionary time scale and its size (1500bp) which can carry sufficient information in it have made it the "gold standard" molecular marker for microbial classification. Before 16S rRNA, DNA-DNA reassociation kinetics was used for species identification which required 70% DNA-DNA similarity and $T_m \leq 5$ ˚C for the DNA-DNA heteroduplex molecule (Janda & Abbott, 2007). The idea of using the 16S rRNA gene as a marker for species identification together with the development of Sanger automated sequencing (Sanger et al., 1977) illuminated a whole new world of studying and characterizing microorganisms. The advancement of high-throughput sequencing technology with reduced cost and time has revolutionized the area of genomics.

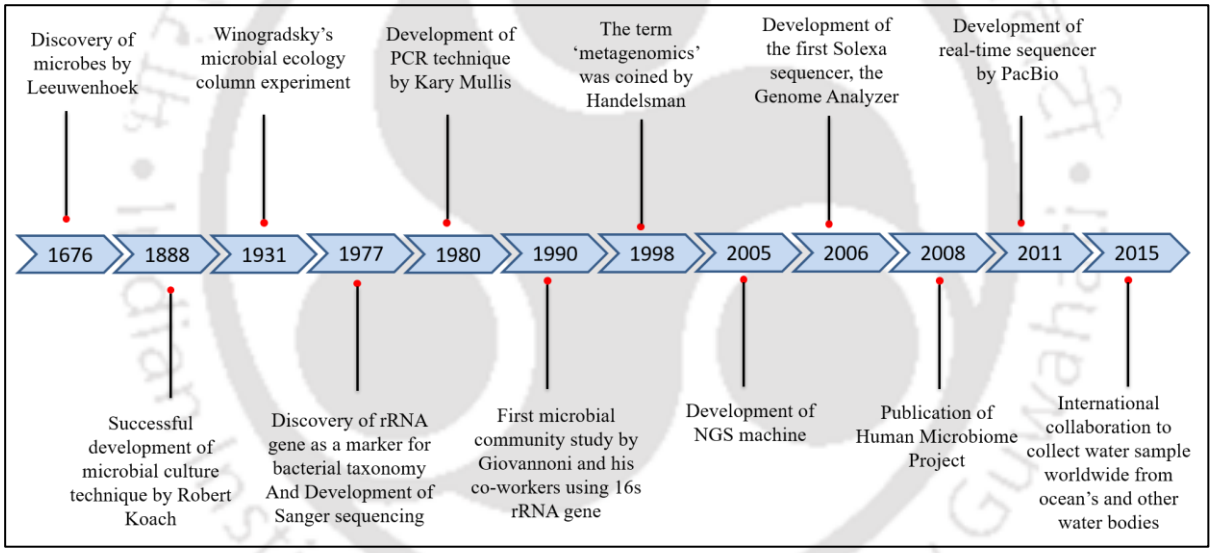


Figure 1. 2 The Milestones in the development of Metagenomics

The etymology of the term "metagenomics" can be traced back to the publication of Jo Handelseman et al. in 1998, who gave the first concept of cloning the whole metagenome to access the collective genome of the microbial community present in an environmental sample, their biosynthesis machinery and chemistry (Handelsman et al., 1998) eliminating the need of culturing them. Presently, metagenomics studies can be divided into two broad categories based on sequencing strategies viz. targeted marker gene sequencing and shotgun metagenomics (Rausch et al., 2019). Targeted marker gene sequencing or amplicon analysis involves the isolation of total DNA from the environmental sample (marker gene, e.g. 16S rRNA/ITS) using a universal primer

so that we can identify all the microbes present in the sample without the need for culturing (Yilmaz et al., 2011). This technique involves high throughput DNA sequencing technologies and computational skills to give robust information within a short period. Previously, Roche 454 pyrosequencing platform was used for amplicon sequencing which has recently been replaced by the Illumina MiSeq platform that can yield up to 25 million 2×300 paired-end reads per run enabling parallel sequencing of 400 samples with ~ 50000 paired ends reads per run per sample (Dong et al., 2017). The milestones in the development of metagenomics are depicted in **Figure 1.2**.

1.4.1 The 16S rRNA gene as a phylogenetic marker

The 16S rRNA gene has been a cornerstone of sequence-based bacterial study for the past several years which has proven to be an efficient and economical method for studying

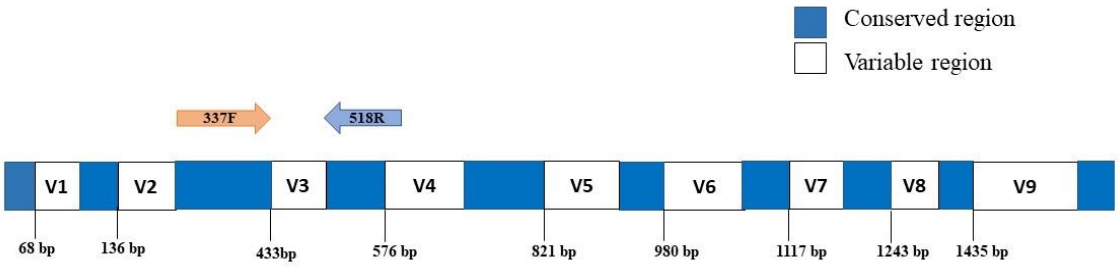


Figure 1. 3 The variable and conserved regions of the 16S rRNA gene

environmental microbial community (Johnson et al., 2019). Ribosomal RNA is universally accepted as the mainstay for studying phylogenetic relatedness among prokaryotes because of its universal occurrence and the presence of highly variable and conserved domains (Woese & Fox, 1977). The 16S rRNA gene of bacteria contains nine "hypervariable regions" (V1–V9) with signature oligonucleotide sequences characteristic of a specific phylogenetic group (**Figure 1.3**). The total length of the 16S rRNA gene is approximately 1500 base pairs which shows significant sequence variability across different bacteria (Chakravorty et al., 2007). Furthermore, rRNA genes are crucial for the phylogenetic study of distantly related species since they evolve more slowly than protein-encoding genes (Chakravorty et al., 2007).

It is possible to amplify the variable regions by using primers from the conserved areas which are

sufficient to distinguish between species using statistically reliable measures. Since every bacterial species possesses the 16S gene, connections between all bacterial species may be determined. Unknown bacteria can be categorised by comparing their 16S sequences to those that have already been deposited in the reference databases (Clarridge, 2004). The foundation of bacterial taxonomy is the analysis of 16S and 23S rRNA, particularly for the identification of nonculturable bacteria. The identification of variation in microbial composition among samples is highly dependent on the choice of the hypervariable region to be sequenced (Rausch et al., 2019).

1.4.2 Shotgun Metagenomics to study the genomic potential of gut bacterial community

Amplicon sequencing or 16S rRNA gene sequencing is an important technique to study the environmental microbial community. However, it has several limitations and is confined only to the taxonomic identification of the microbiota present in the studied environment. Often, taxonomic identification suffers from low resolution (i.e. species identification only up to the family level or above) due to insufficient sequencing depth or inadequate representation of the microbial species in the isolated environmental DNA (Rausch et al., 2019). Functional profiling through this targeted approach is difficult since it provides no direct indication of the functional capabilities of the microbial community. However, the tool Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) shows how to infer a functional profile of a microbial community directly from taxonomic profiling through marker genes and a database of reference genomes. For the detailed investigation of the functional role played by the microbial community coexisting in a particular niche, we need to perform shotgun metagenomics (Quince et al., 2017). Unlike marker-gene analysis, shotgun metagenomics aims to comprehensively analyze all the microorganisms and their genomic content present in a complex environmental sample to get detailed knowledge about their functional potential. With the declining cost of DNA sequencing, and the development of robust sequence analysis tools and computational methods, it is becoming an increasingly important part of the microbial study. Illumina HiSeq 2500, HiSeq 300 and HiSeq 4000 are some of the most preferred shotgun

sequencing platforms which can give an output of 1.5 Tb per run followed by Ion Torrent S5, Oxford Nanopore minion and pacific bioscience.

A typical shotgun metagenomic experiment starts with the isolation of a considerable amount of high-quality genomic DNA from a biological or environmental sample (Quince et al., 2017). This may be obtained by following either kit-dependent or manual isolation of metagenomic DNA. Protocols should be optimized beforehand to maintain the integrity, quality and quantity of the isolated DNA to ensure reliable, informative and less-erroneous downstream analysis. Isolated DNA is sequenced after library construction through any of the different next-generation sequencing platforms. Library construction follows the same basic steps for different sequencing platforms. These include DNA fragmentation, end-repair and manipulation, adapter ligation, size fractionation and PCR amplification of adapter-ligated DNA fragments. The process of library preparation for NGS technologies has been discussed elaborately by Head and his co-author (Head et al., 2014).

Next-generation sequencing technologies encompass massively parallel, high throughput sequencing technologies that allow an analysis of millions of reads per run with higher sequence coverage. Metagenomic technology currently utilizes short-read sequencing technologies like 454 pyrosequencing, Illumina Hiseq and Miseq, Ion torrent as well as long-read sequencing technologies like oxford nanopore and SMRT sequencing (Watts & Hurwitz, 2020). However, due to cost efficiency and higher accuracy, short-read sequencing technologies are preferred more often over long-read technologies. A detailed review of the history and development of NGS technologies and their application in omic technologies have been published by Kulski in 2016 (Kulski, 2016).

Shotgun metagenomics sequencing produces millions of reads per sample creating several Gigabytes of data depending on the sequencing depth and efficiency of the instrument (Ladoukakis et al., 2014). During the sequencing process, the sequencing signals or fluorescent peaks are assigned to nucleobases in a process called base call. The base call data are stored in individual

base call files which are converted to sequence data in FASTQ format after the completion of the sequencing process. FASTQ is a text file format that contains both the sequence bases and their respective quality scores assigned by PHRED software during the base call. It is the most common format for all sequencing platforms and is widely used for the storage and sharing of sequencing data (Cock et al., 2010). NGS technologies most often deliver their data in a paired-end format containing forward and reverse reads in two fastQ files which can be merged subsequently to obtain a longer read. Every FASTQ file contains four lines for each sequence. The first line starts with '@' and is generally the sequence identifier where description may be added optionally, the second line consists of the raw sequence letters, and the third line starts with '+' which signifies the end of the sequence letter and the start of the quality string. The Sanger FASTQ format and its variation with other sequencing platforms have been discussed by Cock and his co-authors (Cock et al., 2010). The quality score q assigned to each base is defined as

$$q = -10\, log10(P)$$

where P is the probability of estimated error during a base call. Therefore, a Phred quality score of 10 will signify a probability of 1 incorrect base calling in every 10 bases or 90% accuracy. Similarly, Phred quality scores of 20 and 30 will signify the probability of 1 incorrect base call in every 100 and 1000 bases respectively.

1.5 Metatranscriptomics to find out active bacterial populations under different physiological conditions

Technological and computational advancements in RNA sequencing and data analysis have given us the ability to investigate the genes actively expressed by the microbial community in a particular niche. This helps us a better understanding of the host-microbiome interaction and functional alteration driven by a disease condition or vice versa. After the successful completion of the human genome sequencing project, The National Institute of Health funded a major initiative towards the understanding of the enormous prokaryotic gene content of the microbial community within the human body to understand its impact on human physiology and diseases. The preliminary study of the Human Microbiome Project (HMP) focussed on the microbial community inhabiting different

parts of the healthy human body viz. nasal, oral, urogenital and gastrointestinal regions (Huttenhower et al., 2012; Manos, 2022). Human microbiome studies have found a strong correlation between gut microbial composition and pathological conditions (Huttenhower et al., 2012; Manos, 2022). Metatranscriptomics data complement metagenomics data by accurate investigation of those genes that are transcribed among all those annotated in metagenomics analysis and give an idea about their transcription level. This enables us to understand the microbial species that are functioning in that particular environment. In comparison to metagenomics, metatranscriptomics provides a more comprehensive perspective by giving information about the transcriptionally active microbial population (Benítez-Páez et al., 2014; Franzosa et al., 2014). This also gives information about which genes are the highest expressed and which are the lowest and the vital pathways required/expressed in a particular condition. To identify those genes with low expression, very high sequencing depth is required. Metatranscriptomic analysis can also be done for biomarker detection.

In general, Metatranscriptomics reads are first aligned using alignment tools such as BOWTIE, BWA and BLAST and mapped to databases like NCBI. Results are then annotated with annotation databases like GO, KEGG, COG, and Swiss-Prot (Bairoch & Apweiler, 2000; Galperin et al., 2021; Kanehisa & Goto, 2000). The shorter transcript reads are then assembled into longer fragments called contigs. Trinity, MetaVelvet, SOAPdenovo etc. (Haas et al., 2013; Luo et al., 2012; Namiki et al., 2012), are some of the popular sequence assembly tools.

1.6 Basic tools and techniques for metagenomic and metatranscriptomic analyses

There are two categories of metagenomics studies, one is marker gene analysis and the other is shotgun metagenomic analysis. Over the past two decades, several tools and techniques have been developed to identify and characterize microbial communities from environmental samples more easily and efficiently (Escobar-Zepeda et al., 2015). On the other hand, metatranscriptomics is comparatively a newer area than metagenomics and it is still developing. The basic steps of metatranscriptomic analysis are almost similar to shotgun metagenomic sequence analysis except

for the basic experimental procedure of RNA isolation from the sample (Shakya et al., 2019).

The initial step of metagenomic and metatranscriptomic sequence processing is to check the quality of the raw reads produced by the sequencing platforms. The raw reads are typically produced in 2×250 or 2×150 paired-end mode in FASTQ format (Liu et al., 2021). The FastQC tools provide an easy way to perform preliminary quality control checks of the raw sequence data. This must be done before downstream analysis of any sequence reads (Andrews, 2010). The raw paired-end sequences are categorized based on their barcode sequences. This step is known as demultiplexing. The paired-end sequence reads can subsequently be merged based on their overlapping regions to generate a sequence with a longer read length (Zhang et al., 2014). This step is followed by the removal of the barcode and primers. For amplicon sequence data, the quality control steps and downstream analyses can be performed utilizing different amplicon sequence analysis tools like QIIME2 (Bolyen et al., 2019), Mothur (Schloss et al., 2009) and Usearch (Edgar & Bateman, 2010). The quality control step is followed by the conversion of raw reads into a feature table. A crucial step in amplicon analysis is selecting the representative sequences to serve as approximations of a species. The two primary strategies for selecting representative sequences are denoising the sequences to Amplicon Sequence Variants (ASVs) and clustering to Operational Taxonomic Units (OTUs). The UPARSE algorithm groups sequences into OTUs with 97% sequence similarity (Edgar, 2013). This approach, though, could miss minute variations across species or strains. A recently created denoising algorithm, DADA2 produces ASVs as more precise representative sequences (Callahan et al., 2016). The frequency of the feature sequences in each sample may then be quantified to create a feature table (OTU/ASV table). Subsequently, the taxonomic assignment can be done to the feature sequences at different taxonomic levels such as Kingdom, Phylum, Order, Family and Genus. (Liu et al., 2021).

16S rDNA amplicon sequencing can often only be used to learn about the taxonomic composition of a sample. However, several programs are now available that can anticipate probable functional data. This prediction is based on the idea of connecting 16S rDNA sequences or taxonomy data to

functional descriptions available in the reference genome database (Liu et al., 2021). The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) software predicts that could infer functional metagenome from OTUs in the form of counts of gene content as KO assignments (Kyoto Encyclopedia of Genes and Genomes Orthology (KEGG Orthology)) (Langille et al., 2013). One of the key limitations of PICRUSt1 was that it needed closed-reference OUT-picking against Greengenes bacterial database which has not been updated since 2013. PICRUSt2, the recent version of PICRUSt uses Amplicon Sequence Variants or ASVs instead of OTUs to infer function. Another major advantage of PICRUSt2 is it utilizes Integrated Microbial Genome Database containing approximately 42000 bacterial and archaeal genomes (Douglas et al., 2020). Another open-source R package, Tax4Fun allows the users to predict the functional potential of the microbial community from amplicon data based on the SILVA database (Quast et al., 2013).

Shotgun metagenomic data analysis is a tedious process owing to its complexity and large size which necessitates high computational power and expertise to dig valuable information out of it. An in-depth review has been published by Quince and his co-workers about the process of shotgun metagenomic data analysis (Quince et al., 2017). In brief, the analysis procedure consists of the following steps:

1. Quality control and pre-processing of the raw data
2. Sequence Assembly,
3. Binning (optional)
4. Gene detection and annotation,
5. Pathway reconstruction,
6. Taxonomic and phylogenetic analysis,
7. Data comparison.

Quality control of the raw data is the foremost step for any NGS data analysis. It entails the examination of raw sequences based on the Phred quality score and the elimination and trimming

of low-quality reads and segments to avoid misleading experimental conclusions. Although the NGS technologies have brought a revolution to the area of omics studies, they possess several inherent limitations which increase the chance of sequencing artifacts including low-quality reads, dropping down the quality towards the ends, chimera formation and the presence of contamination due to incorrect amplification of the target region (Fox et al., 2014; Zhou et al., 2013). These inessential reads or fragments must be identified and filtered out before analysis. Low-quality sequence data may also be obtained due to low quality and insufficient quantity of the isolated metagenomic DNA, inappropriate library preparation and functionality and handling of the instrument. Generally, bases with a Phred score ≥ 20 or 30 are accepted for downstream analysis. Bases with Phred scores ≤ 20 or 30 are filtered out during the quality control step (Liao et al., 2017). Adapter sequences are ligated to each DNA fragment during library preparation which helps in indexing and attaching the DNA fragments to the flow cells. These sequences are considered non-biological and trimming them from sequence reads is another crucial pre-processing step of NGS data analysis (Trivedi et al., 2014).

FasQC is a widely used quality assessment tool for high-throughput sequence data. It was built by Andrews in 2010 using the Java language (Andrews, 2010). It imparts a quick assessment of the quality of the raw data in the form of graphs and tables. FastQC accepts FastQ, gunzip FastQ, SAM, or BAM formats and the results can be exported in a permanent HTML-based QC report. It can be operated in two different modes, interactive standalone mode and non-interactive mode. The standalone interactive mode is suitable for quick analysis of a small set of fastQ files whereas the non-interactive mode can be integrated into larger pipelines which facilitates analysis of a multitude of fastQ files. The quality checking results are interpreted in 11 different modules that include per base sequence quality, per sequence quality score, per base GC content, per sequence GC content, per base N content, sequence duplication level, adapter content etc. The quality of each module is coded in three different colours viz., green, yellow and red indicating very good,

reasonable and poor quality respectively. It can be run on Windows, Linux and Mac operating systems (Andrews, 2010).

Another widely-used, efficient and flexible tool developed for quality filtering and removal of adapter and PCR primers of NGS data is Trimmomatic. It can accurately handle both single-end and paired-end data. Trimmomatic follows two different approaches to spot the presence of adapter and primer sequences within the reads, viz. simple mode and palindrome mode. The simple mode works by finding approximate matches between a user-supplied primer or adapter sequence and the sequence read. It can detect these non-biological sequences in any orientation and location of the read depending on a minimum overlap. On the other hand, the palindrome mode is specially designed to detect adapters in paired-end data. Palindrome mode poses various advantages in reliable adapter detection even if the number of adapters present is very low in the read files (Bolger et al., 2014). Trimmomatic finds the adapters by following the 'seed and extend' algorithm where the small fragments of the adapter known as 'seed' (maximum 16bp long) are searched for an exact or adequate match in every possible position within the sequence read. There is a parameter called *seedMismatch* that defines the maximum number of allowed mismatches during the alignment between the adapter and the read. Based on this, an alignment score is given. The aligned regions that surpass a user-specified threshold value of the alignment score are removed. Trimmomatic also removes the low-quality 3′ end of a read following a sliding window approach where the read is scanned from the 5′ end and clips the 3′ end when the average quality of the bases of that region drops below a user-defined threshold value. The sliding window approach is particularly helpful for quality filtering of Illumina data where the quality of the 3′ end often starts falling. Maximum information quality filtering is another approach that tries to retain longer reads instead of retaining wrong bases which are often more informative than short reads. It implements three factors to determine the minimum length of read to be retained: a) length threshold, b) coverage and c) error rate (Bolger et al., 2014).

Cutadapt is a user-friendly, Python-based command-line tool that finds and trims technical sequences including adapters and PCR primers from both single-end and paired-end NGS data. It supports the most common formats FASTA and FASTQ as well as the CSFASTA and QUAL files generated by SOLiD sequencing technology. Adapters are found in an error-tolerant way allowing insertions, deletions and mismatches in the adapter sequence up to a maximum predefined error rate (0.1 by default). It also allows any IUPAC wild-card letters within the adapter sequence. Cutadapt can detect both 3′ and 5′ adapters, either full or partial in length anywhere in the sequence read. This tool can also be utilized to trim a user-specified number of bases from the beginning or end of the reads (Wang et al., 2011).

Most meta-omic sequencing platforms produce reads which are much shorter in length than the average size of the genes (Ghurye et al., 2016). Metagenomic and metatranscriptomic sequence assembly is the process of aligning and merging sequence reads in the correct order to reconstruct individual genes and genomes from the mixed sequence reads of a microbial community. Megahit is a freely available Next Generation Sequence assembly tool developed by Li and his coworkers that utilizes the succinct *de Bruijn* graph approach to assemble large and complex datasets quickly and cost-effectively (Li et al., 2015). Tools like MetaSPAdes (Nurk et al., 2017) MetaVelvet (Namiki et al., 2012) IDBA-UD (Peng et al., 2012) SOAPdenovo (Luo et al., 2012) are also widely used for assembly.

Metagenome and metatranscriptome annotation refer to the process of identifying and characterizing the genes, proteins, and other functional elements present in a metagenomic dataset (Bengtsson-Palme, 2018). This process is often facilitated by the use of bioinformatics tools and databases, such as BLAST (Altschul et al., 1990), HMMER (Finn et al., 2011), and Pfam (Finn et al., 2014), that can identify sequence similarities and assign putative functions to the predicted proteins. In addition to protein-coding genes, metagenome annotation can also reveal other functional elements in the metagenomic dataset, such as non-coding RNA genes, regulatory elements, and mobile genetic elements (Liu et al., 2021; Tobar-Tosse et al., 2013). The results of

metagenome annotation can provide insights into the functional potential of the microbial community in the sampled environment, as well as help to identify potential biotechnological applications, such as the discovery of novel enzymes or biosynthetic pathways.

1.7 Metabolomics: A powerful emerging tool to study the metabolites

Metabolomics is the comprehensive study of metabolites in biofluids, cells or tissue that can provide a deep insight to understand the biology of an organism under various physiological conditions. These metabolites can regulate vital cellular functions, immunity against the pathogen, and nutrition homeostasis etc. (Mamas et.al. 2011). It is an emerging and powerful approach that measures the concentration of small molecules necessary for metabolisms and reflects the underlying biochemical processes and state of the cells (Clish, 2015). Unlike other omics approaches, metabolomics faces analytical challenges due to variations in the physical properties of the molecules (e.g. polar, non-polar, volatile and non-volatile). Therefore, for effective analysis, metabolites are divided into subsets based on polarity, the presence of common functional groups and structural similarity (C. H. Johnson & Gonzalez, 2012). For standardization and guideline, data reporting systems such as Metabolomic Standard Initiative and open access repositories like MetaboLights and Metabolomics Workbench have been established for molecule identification (Haug et al., 2020; Sud et al., 2016). Technologies liquid chromatography-mass spectrometry, NMR, gas chromatography-mass spectrometry and X-ray crystallography can routinely measure tens to hundreds of metabolites with excellent precision (Johnson & Gonzalez, 2012).

The two main branches of metabolomics include targeted and untargeted metabolomics (Lelli et al., 2021a). Targeted metabolomics involves the analysis of a specific set of metabolites that are pre-selected based on prior knowledge of the metabolic pathways or biological processes under investigation. Targeted metabolomics is a subfield of metabolomics that focuses on the targeted quantification and analysis of a specific set of metabolites, usually within a particular biological pathway or metabolic network (Roberts et al., 2012a). The goal of targeted metabolomics is to identify and measure the levels of specific metabolites in a biological sample, such as blood, urine,

or tissue, to gain insight into the underlying biological processes or disease states. This approach uses a variety of analytical techniques, such as liquid chromatography-mass spectrometry (LC-MS), gas chromatography-mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR) spectroscopy, to measure the concentration of targeted metabolites in biological samples (Ai et al., 2021; Hosseinkhani et al., 2022). Targeted metabolomics can help in scientific research by providing a detailed understanding of the metabolic pathways and networks involved in various physiological processes. By identifying and quantifying specific metabolites, researchers can gain insight into the underlying mechanisms of diseases and disorders, and identify potential biomarkers for diagnosis and treatment (Roberts et al., 2012b). For example, targeted metabolomics can be used to identify metabolic signatures associated with cancer, diabetes, or cardiovascular disease. By measuring the levels of specific metabolites in blood or tissue samples, researchers can identify changes in metabolic pathways that may be linked to these diseases, and potentially identify new targets for drug development (Gowda et al., 2008; Tsoukalas et al., 2019). Targeted metabolomics can also be used to assess the effects of drugs or other interventions on metabolism (Saigusa et al., 2021). By measuring changes in the levels of specific metabolites in response to treatment, researchers can gain insights into the mechanisms of action of the drug and assess its effectiveness (Alarcon-Barrera et al., 2022a; Rabinowitz et al., 2011).

On the other hand, untargeted metabolomics, also known as discovery metabolomics, involves the comprehensive analysis of all metabolites present in a biological sample without prior knowledge of their identity or function. Untargeted metabolomics is useful for identifying novel metabolites, biomarkers, and metabolic pathways associated with specific biological processes or diseases (Chen et al., 2021).

Untargeted metabolomics has emerged as a powerful tool for the comprehensive analysis of the metabolome, which comprises the complete set of small-molecule metabolites in a biological sample (Chen et al., 2021). Unlike targeted metabolomics, which focuses on a predefined set of metabolites, untargeted metabolomics allows for the identification and quantification of all

detectable metabolites, without prior knowledge of their identities (Lelli et al., 2021b). This approach has a wide range of applications in various fields, including biomedical research, clinical practice, and drug development (Chen et al., 2021).

Untargeted metabolomics has numerous applications in biomedical research, including the study of disease pathogenesis, biomarker discovery, and drug development (Yang et al., 2019). By comparing the metabolite profiles of healthy and diseased samples, researchers can identify metabolic pathways that are altered in disease states, and potentially discover new biomarkers for early diagnosis and monitoring of disease progression. For example, untargeted metabolomics has been used to identify metabolic signatures associated with cancer, diabetes, and neurodegenerative diseases (Chalikiopoulou et al., 2023; Ruan et al., 2022; Thistlethwaite et al., 2022). Furthermore, untargeted metabolomics has been applied to drug discovery and development, to identify drug targets and evaluate the efficacy and safety of drugs (Alarcon-Barrera et al., 2022b).

Untargeted metabolomics has also shown great potential in clinical practice, particularly in the diagnosis and monitoring of metabolic disorders, such as inborn errors of metabolism (Ismail et al., 2019). By analyzing the metabolite profiles of patient samples, clinicians can identify metabolic abnormalities that may be indicative of a particular disorder, and potentially guide treatment decisions. For example, untargeted metabolomics has been used to identify biomarkers for inherited metabolic disorders such as phenylketonuria, maple syrup urine disease, and urea cycle disorders (Miller et al., 2015). Additionally, untargeted metabolomics has been applied to the diagnosis and monitoring of other conditions, such as cardiovascular disease, obesity, and liver disease.

Despite its potential, untargeted metabolomics still faces several challenges and limitations. One of the major challenges is the large volume and complexity of data generated by the analysis, which requires sophisticated data processing and analysis methods (Schrimpe-Rutledge et al., 2016). Variability and heterogeneity of biological samples can also lead to false positives and false negatives in metabolite detection. The lack of standardized protocols for sample collection,

storage, and preparation can introduce variability and affect the reproducibility of results (Minno et al., 2022).

Emerging trends in untargeted metabolomics research include the integration of multi-omics data, such as genomics, transcriptomics, and proteomics, to provide a more comprehensive understanding of biological systems. The development of new data processing and analysis methods, such as machine learning and artificial intelligence, can also improve the accuracy and reliability of metabolite identification and quantification. Additionally, the development of non-invasive or minimally invasive sampling methods, such as breath or saliva analysis, is gaining interest as a more convenient and cost-effective way of diagnosing and monitoring metabolic disorders. In due course, there is a growing interest in the development of non-invasive or minimally invasive untargeted metabolomics approaches, such as breath or urine analysis, which could enable the diagnosis and monitoring of metabolic disorders and other conditions more conveniently and cost-effectively.

Both targeted and untargeted metabolomics have their advantages and limitations. Targeted metabolomics is best suited for studying specific biological processes or diseases, where the metabolites of interest are already known. In contrast, untargeted metabolomics is best suited for discovering novel metabolites, metabolic pathways, and biomarkers associated with various biological processes or diseases. However, untargeted metabolomics has higher analytical complexity and data processing requirements compared to targeted metabolomics.

1.8 Research gap

To date, the gut microbiota of *A. assamensis* has remained largely unexplored. Culture-dependent studies followed by taxonomic identification through 16S rRNA gene sequencing have found that the most dominant bacteria in the gut of *A. assamensis* include *Bacillus* (54%), *Serratia* (24%), *Pseudomonas* (10%), and *Alcaligenes* (6%). Pathogenicity assessment of *A. assamensis* gut microbiota revealed the presence of *Pseudomonas aeruginosa*, *Ornithinibacillus bavariensis*, *Achromobacter xylosoxidans* and *Staphylococcus aureus* in both healthy and diseased larvae.

However, no in-depth investigation at the meta-omic level had been carried out to study the composition and functional role of the gut microbial community of *A. assamensis*. Based on the literature review, the following objectives were formulated to obtain detailed knowledge of the gut microbial community of *A. assamensis*, its functional role and its interaction with the host organism. Additionally, untargeted metabolomic analysis was carried out for identifying the global metabolome of *A. assamensis*.

1.9 Objectives of the work

The broad objectives of this research work have been outlined as follows:

1. Structural analysis of the gut microbial community of *Antheraea assamensis* and its comparison with the host leaf-associated microbiota.
2. Functional analysis of the gut microbial community of *Antheraea assamensis*.
3. Gene expression profiling of the gut microbial community of *Antheraea assamensis* and identification of the active gut microbiota.
4. Untargeted metabolomics analysis of the larval gut, host leaf and silk gland of *Antheraea assamensis* to identify the global metabolites and to study their variation with change in the host plant.

1.10 Organization of the Thesis

The thesis is organized into the following six chapters:

Chapter 1: Introduction and Literature Review.

Chapter 2: Structural analysis of the gut microbiota of *Antheraea assamensis* Helfer and its comparison with host leaf-associated microbiota.

Chapter 3: Functional analysis of *A. assamensis* gut microbiome.

Chapter 4: Gene expression profiling of *A. assamensis* gut microbiome.

Chapter 5: Untargeted metabolomic analysis of *A. assamensis.*

Chapter 6: Summary and prospects

References

Ai, Z., Zhang, Y., Li, X., Sun, W., & Liu, Y. (2021). Widely Targeted Metabolomics Analysis to Reveal Transformation Mechanism of Cistanche Deserticola Active Compounds During Steaming and Drying Processes. *Frontiers in Nutrition*, *8*, 743. https://doi.org/10.3389/FNUT.2021.742511/BIBTEX

Alarcon-Barrera, J. C., Kostidis, S., Ondo-Mendez, A., & Giera, M. (2022a). Recent advances in metabolomics analysis for early drug development. *Drug Discovery Today*, *27*(6), 1763–1773. https://doi.org/10.1016/J.DRUDIS.2022.02.018

Alarcon-Barrera, J. C., Kostidis, S., Ondo-Mendez, A., & Giera, M. (2022b). Recent advances in metabolomics analysis for early drug development. *Drug Discovery Today*, *27*(6), 1763–1773. https://doi.org/10.1016/J.DRUDIS.2022.02.018

Alison, J., Alexander, J. M., Diaz Zeugin, N., Dupont, Y. L., Iseli, E., Mann, H. M. R., & Høye, T. T. (2022). Moths complement bumblebee pollination of red clover: A case for day-And-night insect surveillance. *Biology Letters*, *18*(7). https://doi.org/10.1098/rsbl.2022.0187

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amato, K. R., Yeoman, C. J., Kent, A., Righini, N., Carbonero, F., Estrada, A., ... & Gillis, M. (2013). Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes. The ISME journal, 7(7), 1344.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Bairoch, A., & Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*, *28*(1), 45. https://doi.org/10.1093/NAR/28.1.45

Baldo, L., Riera, J. L., Mitsi, K., & Pretus, J. L. (2018). Processes shaping gut microbiota diversity in allopatric populations of the endemic lizard Podarcis lilfordi from Menorcan islets (Balearic Islands). *FEMS Microbiology Ecology*, *94*(2), 186. https://doi.org/10.1093/FEMSEC/FIX186

Barelli, C., Albanese, D., Donati, C., Pindo, M., Dallago, C., Rovero, F., ... & De Filippo, C. (2015). Habitat fragmentation is associated to gut microbiota diversity of an endangered primate: implications for conservation. Scientific reports, 5, 14862.

Bengtsson-Palme, J. (2018). Strategies for Taxonomic and Functional Annotation of Metagenomes. *Metagenomics: Perspectives, Methods, and Applications*, 55–79. https://doi.org/10.1016/B978-0-08-102268-9.00003-3

Benítez-Páez, A., Belda-Ferre, P., Simón-Soro, A., & Mira, A. (2014). Microbiota diversity and gene expression dynamics in human oral biofilms. *BMC Genomics*, *15*(1), 1–13. https://doi.org/10.1186/1471-2164-15-311/FIGURES/5

Bharathi, Prof. D. (2016). Sericulture Industry in India - A Source of Employment Generation. *International Journal of Advanced Engineering Research and Science*, *3*(10), 144–147. https://doi.org/10.22161/ijaers/310.23

Biere, A., & Bennett, A. E. (2013). Three-way interactions between plants, microbes and insects. In *Functional Ecology* (Vol. 27, Issue 3, pp. 567–573). https://doi.org/10.1111/1365-2435.12100

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. In *Nature Biotechnology* (Vol. 37, Issue 8, pp. 852–857). Nature Publishing Group. https://doi.org/10.1038/s41587-019-0209-9

Borah, M. B., & Borgohain, A. (2017). State and Muga Silk Industry in Pre-Independent Assam. In *International Journal of Humanities and Social Sciences* (Vol. 7, Issue 1). http://www.ripublication.com

Broderick, N. A., Buchon, N., & Lemaitre, B. (2014). Microbiota-induced changes in Drosophila

melanogaster host gene expression and gut morphology. MBio, 5(3), e01117-14.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods 2016 13:7*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology : Quarterly Publication of the Hellenic Society of Gastroenterology*, *28*(2), 203. /pmc/articles/PMC4367209/

Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., & Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. Microbial ecology in health and disease, 26(1), 26191.

Chaban, B., Ng, S. Y. M., & Jarrell, K. F. (2006). Archaeal habitats--from the extreme to the ordinary. *Canadian Journal of Microbiology*, *52*(2), 73–116. https://doi.org/10.1139/w05-147

Chakravorty, S., Helb, D., Burday, M., Connell, N., & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, *69*(2), 330. https://doi.org/10.1016/J.MIMET.2007.02.005

Chalikiopoulou, C., Gómez-Tamayo, J. C., & Katsila, T. (2023). Untargeted Metabolomics for Disease-Specific Signatures. *Methods in Molecular Biology (Clifton, N.J.)*, *2571*, 71–81. https://doi.org/10.1007/978-1-0716-2699-3_7

Chen, B., Du, K., Sun, C., Vimalanathan, A., Liang, X., Li, Y., Wang, B., Lu, X., Li, L., & Shao, Y. (2018). Gut bacterial and fungal communities of the domesticated silkworm (Bombyx mori) and wild mulberry-feeding relatives. *The ISME Journal 2018 12:9*, *12*(9), 2252–2262. https://doi.org/10.1038/s41396-018-0174-1

Chen, B., Teh, B.-S., Sun, C., Hu, S., Lu, X., Boland, W., & Shao, Y. (2016). Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore Spodoptera littoralis. *Scientific Reports*, *6*(1), 29505. https://doi.org/10.1038/srep29505

Chen, L., Lu, W., Wang, L., Xing, X., Chen, Z., Teng, X., Zeng, X., Muscarella, A. D., Shen, Y.,

Cowan, A., McReynolds, M. R., Kennedy, B. J., Lato, A. M., Campagna, S. R., Singh, M., & Rabinowitz, J. D. (2021). Metabolite discovery through global annotation of untargeted metabolomics data. *Nature Methods 2021 18:11*, *18*(11), 1377–1385. https://doi.org/10.1038/s41592-021-01303-3

Cheng, Y., Fox, S., Pemberton, D., Hogg, C., Papenfuss, A. T., & Belov, K. (2015). The Tasmanian devil microbiome—implications for conservation and management. Microbiome, 3(1), 76.

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, *17*(4), 840–862. https://doi.org/10.1128/CMR.17.4.840-862.2004

Clish, C. B. (2015). *Metabolomics: an emerging but powerful tool for precision medicine*. https://doi.org/10.1101/mcs.a000588

Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic acids research, 38(6), 1767-1771.

Convey, P. (2005). Recent lepidopteran records from sub-Antarctic South Georgia. *Polar Biology*, *28*(2), 108–110. https://doi.org/10.1007/S00300-004-0681-6

Dangles, O., Mesías, V., Crespo-Perez, V., & Silvain, J. F. (2009). Crop damage increases with pest species diversity: Evidence from potato tuber moths in the tropical Andes. *Journal of Applied Ecology*, *46*(5), 1115–1121. https://doi.org/10.1111/j.1365-2664.2009.01703.x

Das, N. (2021). Impact of Muga Silk (<i>Antheraea assamensis</i>) on Community Livelihood in the Brahmaputra Valley of Assam-India. *American Journal of Environmental Protection*, *10*(3), 59. https://doi.org/10.11648/j.ajep.20211003.11

Das, S., Sharma, M., Saharia, D., Sarma, K. K., Muir, E. M., & Bora, U. (2017). Electrospun silk-polyaniline conduits for functional nerve regeneration in rat sciatic nerve injury model. *Biomedical Materials (Bristol)*, *12*(4). https://doi.org/10.1088/1748-605X/aa7802

Devi, B., Chutia, M., & Bhattacharyya, N. (2021). Food plant diversity, distribution, and nutritional

aspects of the endemic golden silk producing silkworm, Antheraea assamensis – a review. *Entomologia Experimentalis et Applicata*, *169*(3), 237–248. https://doi.org/10.1111/EEA.13021

Desselberger, U. (2018). The mammalian intestinal microbiome: composition, interaction with the immune system, significance for vaccine efficacy, and potential for disease therapy. Pathogens, 7(3), 57.

Dong, X., Kleiner, M., Sharp, C. E., Thorson, E., Li, C., Liu, D., & Strous, M. (2017). Fast and simple analysis of MiSeq amplicon sequencing data with MetaAmp. *Frontiers in Microbiology*, *8*(AUG). https://doi.org/10.3389/fmicb.2017.01461

Douglas, A. E. (2015). Multiorganismal insects: Diversity and function of resident microorganisms. In *Annual Review of Entomology* (Vol. 60, pp. 17–34). Annual Reviews Inc. https://doi.org/10.1146/annurev-ento-010814-020822

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. In *Nature Biotechnology* (Vol. 38, Issue 6, pp. 685–688). Nature Research. https://doi.org/10.1038/s41587-020-0548-6

Dworkin, M. (2012). Sergei Winogradsky: a founder of modern microbiology and the first microbial ecologist. *FEMS Microbiology Reviews*, *36*(2), 364–379. https://doi.org/10.1111/J.1574-6976.2011.00299.X

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods 2013 10:10*, *10*(10), 996–998. https://doi.org/10.1038/nmeth.2604

Edgar, R. C., & Bateman, A. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/BIOINFORMATICS/BTQ461

Engel, P., & Moran, N. A. (2013a). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, *37*(5), 699–735. https://doi.org/10.1111/1574-6976.12025

Engel, P., & Moran, N. A. (2013b). The gut microbiota of insects - diversity in structure and function. In *FEMS Microbiology Reviews* (Vol. 37, Issue 5, pp. 699–735). https://doi.org/10.1111/1574-

6976.12025

Escobar-Zepeda, A., de León, A. V. P., & Sanchez-Flores, A. (2015). The road to metagenomics: From microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, *6*(DEC), 348. https://doi.org/10.3389/FGENE.2015.00348/BIBTEX

Feldhaar, H. (2011). Bacterial symbionts as mediators of ecologically important traits of insect hosts. Ecological Entomology, 36(5), 533-543.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., & Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, *42*(Database issue), D222. https://doi.org/10.1093/NAR/GKT1223

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(Web Server issue), W29. https://doi.org/10.1093/NAR/GKR367

Flint, H. J., Scott, K. P., Louis, P., & Duncan, S. H. (2012). The role of the gut microbiota in nutrition and health. *Nature Reviews Gastroenterology & Hepatology 2012 9:10*, *9*(10), 577–589. https://doi.org/10.1038/nrgastro.2012.156

Fox, E. J., Reid-Bayliss, K. S., Emond, M. J., & Loeb, L. A. (2014). Accuracy of Next Generation Sequencing Platforms. *Next Generation, Sequencing & Applications*, *1*(01). https://doi.org/10.4172/JNGSA.1000106

Frank, J. H., Frank, J. H., Thomas, M. C., Yousten, A. A., Howard, F. W., Giblin-davis, R. M., Heppner, J. B., Zuparko, R. L., Sánchez, N. E., Luna, M. G., Goula, M., Heppner, J. B., Alekseev, A. N., James, R. R., Capinera, J. L., Weintraub, P. G., Horton, D., Frank, J. H., Mccravy, K. W., … Sourakov, A. (2008). Plume Moths (Lepidoptera: Pterophoridae). *Encyclopedia of Entomology*, 2953–2959. https://doi.org/10.1007/978-1-4020-6359-6_3018

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., Izard, J., Garrett, W. S., Chan, A. T., & Huttenhower, C. (2014).

Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), E2329–E2338. https://doi.org/10.1073/PNAS.1319284111/SUPPL_FILE/PNAS.1319284111.SD06.XLS

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Alvarez, R. V., Landsman, D., & Koonin, E. v. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, *49*(D1), D274–D281. https://doi.org/10.1093/NAR/GKAA1018

GENÇ, V., SEVEN, E., & KAYMAZ, N. (2021). Determination of Butterflies' Potential in Tourism Diversification Based on a Route-Planning Case Study in Botan Valley National Park, Turkey. *Journal of Hospitality and Tourism Issues*. https://doi.org/10.51525/JOHTI.997125

Ghurye, J. S., Cepeda-Espinoza, V., & Pop, M. (2016). *Metagenomic Assembly: Overview, Challenges and Applications*.

Gowda, G. A. N., Zhang, S., Gu, H., Asiago, V., Shanaiah, N., & Raftery, D. (2008). Metabolomics-Based Methods for Early Disease Diagnostics: A Review. *Expert Review of Molecular Diagnostics*, *8*(5), 617. https://doi.org/10.1586/14737159.8.5.617

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P., & Fierer, N. (2017). Caterpillars lack a resident gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9641–9646. https://doi.org/10.1073/pnas.1707186114

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, *5*(10). https://doi.org/10.1016/S1074-5521(98)90108-9

Hara, K., Shinzato, N., Seo, M., Oshima, T., & Yamagishi, A. (2002). *Phylogenetic Analysis of*

Symbiotic Archaea Living in the Gut of Xylophagous Cockroaches (Vol. 17, Issue 4). http://wwwsoc.nii.ac.jp/jsme2/

Härer, A., Torres-Dowdall, J., Rometsch, S. J., Yohannes, E., Machado-Schiaffino, G., & Meyer, A. (2020). Parallel and non-parallel changes of the gut microbiota during trophic diversification in repeated young adaptive radiations of sympatric cichlid fish. *Microbiome*, *8*(1), 1–14. https://doi.org/10.1186/S40168-020-00897-8/FIGURES/5

Hassani, M. A., Durán, P., & Hacquard, S. (2018). Microbial interactions within the plant holobiont. In *Microbiome* (Vol. 6, Issue 1, p. 58). NLM (Medline). https://doi.org/10.1186/s40168-018-0445-0

Haug, K., Cochrane, K., Nainala, V. C., Williams, M., Chang, J., Jayaseelan, K. V., & O'Donovan, C. (2020). MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Research*, *48*(D1), D440–D444. https://doi.org/10.1093/NAR/GKZ1019

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. Biotechniques, 56(2), 61-77.

Heppner, J. B. (2002). Mexican Lepidoptera biodiversity. *Insecta Mundi*. https://digitalcommons.unl.edu/insectamundi/550

Hill, G. M., Kawahara, A. Y., Daniels, J. C., Bateman, C. C., & Scheffers, B. R. (2021). Climate change effects on animal ecology: butterflies and moths as a case study. *Biological Reviews*, *96*(5), 2113–2126. https://doi.org/10.1111/BRV.12746

Hosseinkhani, S., Arjmand, B., Dilmaghani-Marand, A., Mohammadi Fateh, S., Dehghanbanadaki, H., Najjar, N., Alavi-Moghadam, S., Ghodssi-Ghassemabadi, R., Nasli-Esfahani, E., Farzadfar, F., Larijani, B., & Razi, F. (2022). Targeted metabolomics analysis of amino acids and acylcarnitines as risk markers for diabetes by LC–MS/MS technique. *Scientific Reports 2022 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-11970-7

Hugerth, L. W., & Andersson, A. F. (2017). Analysing microbial community composition through amplicon sequencing: From sampling to hypothesis testing. *Frontiers in Microbiology*, *8*(SEP),

1561. https://doi.org/10.3389/FMICB.2017.01561/BIBTEX

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature 2012 486:7402*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Ismail, I. T., Showalter, M. R., & Fiehn, O. (2019). Inborn Errors of Metabolism in the Era of Untargeted Metabolomics and Lipidomics. *Metabolites*, *9*(10). https://doi.org/10.3390/METABO9100242

Jakubowska, A. K., Vogel, H., & Herrero, S. (2013). Increase in Gut Microbiota after Immune Suppression in Baculovirus-infected Larvae. *PLoS Pathogens*, *9*(5). https://doi.org/10.1371/journal.ppat.1003379

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, *45*(9), 2761–2764. https://doi.org/10.1128/JCM.01228-07

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Reddy, D. N. (2015). Role of the normal gut microbiota. *World Journal of Gastroenterology : WJG*, *21*(29), 8787. https://doi.org/10.3748/WJG.V21.I29.8787

Janson, E. M., Stireman, J. O., Singer, M. S., & Abbot, P. (2008). Phytophagous insect-microbe mutualisms and adaptive evolutionary diversification. In *Evolution* (Vol. 62, Issue 5, pp. 997–1012). https://doi.org/10.1111/j.1558-5646.2008.00348.x

Jing, T. Z., Qi, F. H., & Wang, Z. Y. (2020). Most dominant roles of insect gut bacteria: Digestion, detoxification, or essential nutrient provision? *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00823-y

Johnson, C. H., & Gonzalez, F. J. (2012). Challenges and Opportunities of Metabolomics. *Journal of Cellular Physiology*, *227*(8), 2975. https://doi.org/10.1002/JCP.24002

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., & Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications 2019 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-13036-1

Kalita, T., & Dutta, K. (2020). Characterisation of cocoon of different population of Antheraea assamensis (Lepidoptera: Saturniidae). *Oriental Insects*, *54*(4), 574–590. https://doi.org/10.1080/00305316.2020.1727376

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27. https://doi.org/10.1093/NAR/28.1.27

Kar, S., Talukdar, S., Pal, S., Nayak, S., Paranjape, P., & Kundu, S. C. (2013). Silk Gland Fibroin from Indian Muga Silkworm Antheraea assama as Potential Biomaterial. *Tissue Engineering and Regenerative Medicine*, *10*(4), 200–210. https://doi.org/10.1007/s13770-012-0008-6

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. Trends in ecology & evolution, 17(5), 230-241.

Kim, D., Zeng, M. Y., & Núñez, G. (2017). *The interplay between host immune cells and gut microbiota in chronic inflammatory diseases*. *49*. https://doi.org/10.1038/emm.2017.24

Klem, C. C., & Zaspel, J. (2019). Pest Injury Guilds, Lepidoptera, and Placing Fruit-Piercing Moths in Context: A Review. *Annals of the Entomological Society of America*, *112*(5), 421–432. https://doi.org/10.1093/AESA/SAZ031

Kovatcheva-Datchary, P. (2022). The Enigma of Prevotella copri. *Good Microbes in Medicine, Food Production, Biotechnology, Bioremediation, and Agriculture*, 64–68. https://doi.org/10.1002/9781119762621.CH7

Kong, F., Zhao, J., Han, S., Zeng, B., Yang, J., Si, X., & Li, Y. (2014). Characterization of the gut microbiota in the red panda (Ailurus fulgens). PloS one, 9(2), e87885.

Krenn, H. W. (2010). Feeding mechanisms of adult Lepidoptera: Structure, function, and evolution of the mouthparts. In *Annual Review of Entomology* (Vol. 55, pp. 307–327).

https://doi.org/10.1146/annurev-ento-112408-085338

Kulski, J. K. (2016). Next-generation sequencing—an overview of the history, tools, and "Omic" applications. Next generation sequencing-advances, applications and challenges, 10, 61964.

Ladoukakis, E., Kolisis, F. N., & Chatziioannou, A. A. (2014). Integrative workflows for metagenomic analysis. Frontiers in cell and developmental biology, 2, 70.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurber, R. L., Knight, R., Beiko, R. G., & Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. https://doi.org/10.1038/NBT.2676

Lelli, V., Belardo, A., Timperio, A. M., Lelli, V., Belardo, A., & Timperio, A. M. (2021a). From Targeted Quantification to Untargeted Metabolomics. *Metabolomics - Methodology and Applications in Medical Sciences and Life Sciences*. https://doi.org/10.5772/INTECHOPEN.96852

Lelli, V., Belardo, A., Timperio, A. M., Lelli, V., Belardo, A., & Timperio, A. M. (2021b). From Targeted Quantification to Untargeted Metabolomics. *Metabolomics - Methodology and Applications in Medical Sciences and Life Sciences*. https://doi.org/10.5772/INTECHOPEN.96852

Liao, P., Satten, G. A., & Hu, Y. J. (2017). PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies. *Genetic Epidemiology*, *41*(5), 375–387. https://doi.org/10.1002/GEPI.22048

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Lindow, S. E., & Brandl, M. T. (2003). Microbiology of the phyllosphere. In *Applied and Environmental Microbiology* (Vol. 69, Issue 4, pp. 1875–1883). https://doi.org/10.1128/AEM.69.4.1875-1883.2003

Liu, B., Thippabhotla, S., Zhang, J., & Zhong, C. (2021). DRAGoM: Classification and Quantification of Noncoding RNA in Metagenomic Data. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/FGENE.2021.669495/FULL

Liu, Y. X., Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. In *Protein and Cell* (Vol. 12, Issue 5, pp. 315–330). Higher Education Press Limited Company. https://doi.org/10.1007/s13238-020-00724-8

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., … Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, *1*(1), 18. https://doi.org/10.1186/2047-217X-1-18

Mahan, B. (2012). Silk Industry among the Tai-Ahom of Assam, India as an Attraction of Tourist. *International Journal of Scientific and Research Publications*, *2*(12). www.ijsrp.org

Mamas, M., Dunn, W. B., Neyses, L., & Goodacre, R. (2011). The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology, 85, 5-17.

Manos, J. (2022). The human microbiome in disease and pathology. *APMIS : Acta Pathologica, Microbiologica, et Immunologica Scandinavica*, *130*(12), 690–705. https://doi.org/10.1111/APM.13225

Mason, C. J., St Clair, A., Peiffer, M., Gomez, E., Jones, A. G., Felton, G. W., & Hoover, K. (2020). Diet influences proliferation and stability of gut bacterial populations in herbivorous lepidopteran larvae. *PloS One*, *15*(3), e0229848. https://doi.org/10.1371/journal.pone.0229848

Mejia, L. B., Arnal, P., Hallwachs, W., Haxaire, J., Janzen, D., Kitching, I. J., & Rougerie, R. (2020). A global food plant dataset for wild silkmoths and hawkmoths and its use in documenting polyphagy of their caterpillars (Lepidoptera: Bombycoidea: Saturniidae, Sphingidae). *Biodiversity Data Journal 8: E60027*, *8*, e60027-. https://doi.org/10.3897/BDJ.8.E60027

Miller, M. J., Kennedy, A. D., Eckhart, A. D., Burrage, L. C., Wulff, J. E., Miller, L. A. D., Milburn, M. v., Ryals, J. A., Beaudet, A. L., Sun, Q., Sutton, V. R., & Elsea, S. H. (2015). Untargeted

metabolomic analysis for the clinical screening of inborn errors of metabolism. *Journal of Inherited Metabolic Disease*, *38*(6), 1029–1039. https://doi.org/10.1007/S10545-015-9843-7

Minno, A. di, Gelzo, M., Caterino, M., Costanzo, M., Ruoppolo, M., & Castaldo, G. (2022). Promise of the Application of Metabolomics in Precision Medicine. *Int. J. Mol. Sci*, *2022*, 5213. https://doi.org/10.3390/ijms23095213

Mitter, C., Davis, D. R., & Cummings, M. P. (2017a). Phylogeny and Evolution of Lepidoptera. In *Annual Review of Entomology* (Vol. 62, pp. 265–283). Annual Reviews Inc. https://doi.org/10.1146/annurev-ento-031616-035125

Mitter, C., Davis, D. R., & Cummings, M. P. (2017b). Phylogeny and Evolution of Lepidoptera. In *Annual Review of Entomology* (Vol. 62, pp. 265–283). Annual Reviews Inc. https://doi.org/10.1146/annurev-ento-031616-035125

Mohanty, N., Das, H. K., Mohanty, P., & Mohanty, E. (2006). Modification of Muga Silk by Methyl-Methacrylate. II. *Http://Dx.Doi.Org/10.1080/10601329508019151*, *32*(sup7), 1103–1111. https://doi.org/10.1080/10601329508019151

Morais, L. H., Schreiber, H. L., & Mazmanian, S. K. (2020). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology 2020 19:4*, *19*(4), 241–255. https://doi.org/10.1038/s41579-020-00460-0

Morowitz, M. J., Carlisle, E. M., & Alverdy, J. C. (2011). Contributions of Intestinal Bacteria to Nutrition and Metabolism in the Critically Ill. *The Surgical Clinics of North America*, *91*(4), 771. https://doi.org/10.1016/J.SUC.2011.05.001

Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research*, *40*(20). https://doi.org/10.1093/NAR/GKS678

Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., & Pettersson, S. (2012). Host-gut microbiota metabolic interactions. *Science*, *336*(6086), 1262–1267. https://doi.org/10.1126/SCIENCE.1223813/ASSET/6CB4C145-E6AE-4557-86F1-

2BDD196B2064/ASSETS/GRAPHIC/336_1262_F1.JPEG

Nieukerken, E. J. Van, Kaila, L., Kitching, I. J., Kristensen, N. P., Lees, D. C., Minet, J., Mitter, C., Mutanen, M., Regier, J. C., Simonsen, T. J., Wahlberg, N., Yen, S.-H., Zahiri, R., Adamski, D., Baixeras, J., Bartsch, D., Bengtsson, B. Å., Brown, J. W., Bucheli, S. R., … Zwick, A. (2011). Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*, *3148*(1), 212–221–212–221. https://doi.org/10.11646/ZOOTAXA.3148.1.41

Nkamga, V. D., Henrissat, B., & Drancourt, M. (2017). Archaea: Essential inhabitants of the human digestive microbiota. *Human Microbiome Journal*, *3*, 1–8. https://doi.org/10.1016/j.humic.2016.11.005

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/GR.213959.116/-/DC1

Ørsted, M., Yashiro, E., Hoffmann, A. A., & Kristensen, T. N. (2022). Population bottlenecks constrain host microbiome diversity and genetic variation impeding fitness. PLoS Genetics, 18(5), e1010206.

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), 1420–1428. https://doi.org/10.1093/BIOINFORMATICS/BTS174

Perveen, F. K., & Perveen, F. K. (2017). Lepidoptera. *Lepidoptera*. https://doi.org/10.5772/INTECHOPEN.68337

Phukan, R. (2012). Muga silk industry of Assam in historical perspectives. Global Journal of Human-Social Science Research, 12(9-D).

Pierce, N. E., Braby, M. F., Heath, A., Lohman, D. J., Mathew, J., Rand, D. B., & Travassos, M. A. (2002). The ecology and evolution of ant association in the Lycaenidae (Lepidoptera). *Annual Review of Entomology*, *47*, 733–771. https://doi.org/10.1146/ANNUREV.ENTO.47.091201.145257

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590. https://doi.org/10.1093/NAR/GKS1219

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. In *Nature Biotechnology* (Vol. 35, Issue 9, pp. 833–844). Nature Publishing Group. https://doi.org/10.1038/nbt.3935

Rabinowitz, J. D., Purdy, J. G., Vastag, L., Shenk, T., & Koyuncu, E. (2011). Metabolomics in Drug Target Discovery. *Cold Spring Harbor Symposia on Quantitative Biology*, *76*, 235. https://doi.org/10.1101/SQB.2011.76.010694

Ram, R. L., Maji, C., & Bindroo, B. B. (2016). Impact of Climate Change on Sustainable Sericultural Development in India. In *International Journal of Agriculture Innovations and Research* (Vol. 4, Issue 6).

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., von Frieling, J., Hentschel, U., Heinsen, F. A., Höppner, M., Jahn, M. T., Jaspers, C., Kissoyan, K. A. B., Langfeldt, D., Rehman, A., Reusch, T. B. H., Roeder, T., … Baines, J. F. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, *7*(1), 1–19. https://doi.org/10.1186/S40168-019-0743-1/FIGURES/6

Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012a). Targeted Metabolomics. *Current Protocols in Molecular Biology*, *CHAPTER*(SUPPL.98), Unit30.2. https://doi.org/10.1002/0471142727.MB3002S98

Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012b). Targeted Metabolomics. *Current Protocols in Molecular Biology*, *CHAPTER*(SUPPL.98), Unit30.2. https://doi.org/10.1002/0471142727.MB3002S98

Rodpai, R., Sanpool, O., Janwan, P., Boonroumkaew, P., Sadaow, L., Thanchomnang, T., Intapan, P.

M., & Maleewong, W. (2022). Gut microbiota diversity in human strongyloidiasis differs little in two different regions in endemic areas of Thailand. *PLOS ONE*, *17*(12), e0279766. https://doi.org/10.1371/JOURNAL.PONE.0279766

Rooks, M. G., & Garrett, W. S. (2016). Gut microbiota, metabolites and host immunity. *Nature Reviews Immunology 2016 16:6*, *16*(6), 341–352. https://doi.org/10.1038/nri.2016.42

Ruan, X., Wang, Y., Zhou, L., Zheng, Q., Hao, H., & He, D. (2022). Evaluation of Untargeted Metabolomic Strategy for the Discovery of Biomarker of Breast Cancer. *Frontiers in Pharmacology*, *13*, 1722. https://doi.org/10.3389/FPHAR.2022.894099/BIBTEX

Saigusa, D., Matsukawa, N., Hishinuma, E., & Koshiba, S. (2021). Identification of biomarkers to diagnose diseases and find adverse drug reactions by metabolomics. *Drug Metabolism and Pharmacokinetics*, *37*, 100373. https://doi.org/10.1016/J.DMPK.2020.11.008

Saikia, N., Hatibaruah, D., & Dutta, R. B. (2021). Bionomics of predatory butterfly, apefly (Spalgis epius) (Lepidoptera: Lycaenidae) on mealybug, Paracoccus marginatus (Hemiptera: Pseudococcidae) in Mulberry. *~ 430 ~ Journal of Entomology and Zoology Studies*, *9*(5), 430–432. https://doi.org/10.22271/j.ento.2021.v9.i5f.8861

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463. https://doi.org/10.1073/PNAS.74.12.5463

Sarfraz, M., Keddie, A. B., & Dosdall, L. M. (2007). Biological control of the diamondback moth, Plutella xylostella: A review. *Http://Dx.Doi.Org/10.1080/09583150500136956*, *15*(8), 763–789. https://doi.org/10.1080/09583150500136956

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-

09/ASSET/91BD47E1-E1DA-4980-B8B3-5DFA9C4F1FE7/ASSETS/GRAPHIC/ZAM0230904840002.JPEG

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897–1905. https://doi.org/10.1007/S13361-016-1469-Y

Shakya, M., Lo, C. C., & Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics*, *10*(SEP), 904. https://doi.org/10.3389/FGENE.2019.00904/BIBTEX

Shao, Y., Chen, B., Sun, C., Ishida, K., Hertweck, C., & Boland, W. (2017). Symbiont-Derived Antimicrobials Contribute to the Control of the Lepidopteran Gut Microbiota. *Cell Chemical Biology*, *24*(1), 66–75. https://doi.org/10.1016/j.chembiol.2016.11.015

Shibata, N., Kunisawa, J., & Kiyono, H. (2017). Dietary and microbial metabolites in the regulation of host immunity. In *Frontiers in Microbiology* (Vol. 8, Issue NOV). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2017.02171

Sih Kurnianto, A., Yazil Wafa, I., Alifianto, F., & Kurniawan, N. (2016). Journal of Indonesian Tourism and Development Studies The Potential of Butterflies in Tourism Diversification Product: Case Study at Coban Rais Waterfall, Batu, East Java. *J.Ind. Tour. Dev. Std*, *4*(3). https://doi.org/10.21776/ub.jitode.2016.004.03.04

Simhadri, R. K., Fast, E. M., Guo, R., Schultz, M. J., Vaisman, N., Ortiz, L., Bybee, J., Slatko, B. E., & Frydman, H. M. (2017). The Gut Commensal Microbiome of Drosophila melanogaster Is Modified by the Endosymbiont Wolbachia . *MSphere*, *2*(5). https://doi.org/10.1128/MSPHERE.00287-17/FORMAT/EPUB

Singh, N., Lenka, R., Chatterjee, P., & Mitra, D. (2022). Settling moths are the vital component of pollination in Himalayan ecosystem of North-East India, pollen transfer network approach revealed. *Scientific Reports 2022 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41598-022-06635-4

Singh, Y. T., Mazumdar-Leighton, S., Saikia, M., Pant, P., Kashung, S., Neog, K., Chakravorty, R.,

Nair, S., Nagaraju, J., & Babu, C. R. (2012). Genetic Variation within Native Populations of Endemic Silkmoth Antheraea assamensis (Helfer) from Northeast India Indicates Need for In Situ Conservation. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0049972

Song, W. T., Zhu, F. F., & Chen, K. P. (2021). The molecular mechanisms and factors affecting the feeding habits of silkworm (Lepidoptera: Bombyxidae). *Journal of Asia-Pacific Entomology*, *24*(4), 955–962. https://doi.org/10.1016/J.ASPEN.2021.08.010

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*, *39*, 321–346. https://doi.org/10.1146/ANNUREV.MI.39.100185.001541

Suckling, D. M., Conlong, D. E., Carpenter, J. E., Bloem, K. A., Rendon, P., & Vreysen, M. J. B. (2017). Global range expansion of pest Lepidoptera requires socially acceptable solutions. *Biological Invasions*, *19*(4), 1107–1119. https://doi.org/10.1007/S10530-016-1325-9/FIGURES/4

Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., Edison, A., Fiehn, O., Higashi, R., Nair, K. S., Sumner, S., & Subramaniam, S. (2016). Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Research*, *44*(Database issue), D463. https://doi.org/10.1093/NAR/GKV1042

Tang, X., Freitak, D., Vogel, H., Ping, L., Shao, Y., Cordero, E. A., Andersen, G., Westermann, M., Heckel, D. G., & Boland, W. (2012). Complexity and Variability of Gut Commensal Microbiota in Polyphagous Lepidopteran Larvae. *PLOS ONE*, *7*(7), e36978. https://doi.org/10.1371/journal.pone.0036978

Tanji, T., & Ip, Y. T. (2005). Regulators of the Toll and Imd pathways in the Drosophila innate immune response. Trends in immunology, 26(4), 193-198.

Thistlethwaite, L. R., Li, X., Burrage, L. C., Riehle, K., Hacia, J. G., Braverman, N., Wangler, M. F., Miller, M. J., Elsea, S. H., & Milosavljevic, A. (2022). Clinical diagnosis of metabolic disorders using untargeted metabolomic profiling and disease-specific networks learned from profiling data.

Scientific Reports 2022 12:1, *12*(1), 1–17. https://doi.org/10.1038/s41598-022-10415-5

Thursby, E., & Juge, N. (2017). Introduction to the human gut microbiota. *Biochemical Journal*, *474*(11), 1823. https://doi.org/10.1042/BCJ20160510

Tikader, A., Vijayan, K., & Saratchandra, B. (2013). Muga silkworm, antheraea assamensis (Lepidoptera: Saturniidae) - An overview of distribution, biology and breeding. *European Journal of Entomology*, *110*(2), 293–300. https://doi.org/10.14411/eje.2013.096

Tobar-Tosse, F., Rodríguez, A. C., Vélez, P. E., Zambrano, M. M., & Moreno, P. A. (2013). Exploration of Noncoding Sequences in Metagenomes. *PLoS ONE*, *8*(3). https://doi.org/10.1371/JOURNAL.PONE.0059488

Trivedi, U. H., Cézard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., & Gharbi, K. (2014). Quality control of next-generation sequencing data without a reference. *Frontiers in Genetics*, *5*(MAY), 111. https://doi.org/10.3389/FGENE.2014.00111/BIBTEX

Tsoukalas, D., Fragoulakis, V., Sarandi, E., Docea, A. O., Papakonstaninou, E., Tsilimidos, G., Anamaterou, C., Fragkiadaki, P., Aschner, M., Tsatsakis, A., Drakoulis, N., & Calina, D. (2019). Targeted Metabolomic Analysis of Serum Fatty Acids for the Prediction of Autoimmune Diseases. *Frontiers in Molecular Biosciences*, *6*, 120. https://doi.org/10.3389/FMOLB.2019.00120/BIBTEX

Ugwu, J. A., Liu, M., Sun, H., & Asiegbu, F. O. (2020). Microbiome of the larvae of Spodoptera frugiperda (J.E. Smith) (Lepidoptera: Noctuidae) from maize plants. *Journal of Applied Entomology*, *144*(9), 764–776. https://doi.org/10.1111/jen.12821

van Eldijk, T. J. B., Wappler, T., Strother, P. K., van der Weijst, C. M. H., Rajaei, H., Visscher, H., & van de Schootbrugge, B. (2018a). *E V O L U T I O N A R Y B I O L O G Y A Triassic-Jurassic window into the evolution of Lepidoptera*.

van Eldijk, T. J. B., Wappler, T., Strother, P. K., van der Weijst, C. M. H., Rajaei, H., Visscher, H., & van de Schootbrugge, B. (2018b). *E V O L U T I O N A R Y B I O L O G Y A Triassic-Jurassic window into the evolution of Lepidoptera*. https://www.science.org

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., ... & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. science, 304(5667), 66-74.

Vepari, C., & Kaplan, D. L. (2007). Silk as a Biomaterial. *Progress in Polymer Science*, *32*(8–9), 991. https://doi.org/10.1016/J.PROGPOLYMSCI.2007.05.013

Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M., & Fatouros, N. E. (2018). Bacterial symbionts in lepidoptera: Their diversity, transmission, and impact on the host. In *Frontiers in Microbiology* (Vol. 9, Issue MAR). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2018.00556

Walton, R. E., Sayer, C. D., Bennion, H., & Axmacher, J. C. (2020). Nocturnal pollinators strongly contribute to pollen transport of wild flowers in an agricultural landscape. *Biology Letters*, *16*(5). https://doi.org/10.1098/rsbl.2019.0877

Wang, X., Ning, Y., Li, C., Gong, Y., Huang, R., Hu, M., Poulet, B., Xu, K., Zhao, G., Zhou, R., Lammi, M. J., & Guo, X. (2021). Alterations in the gut microbiota and metabolite profiles of patients with Kashin-Beck disease, an endemic osteoarthritis in China. *Cell Death & Disease 2021 12:11*, *12*(11), 1–12. https://doi.org/10.1038/s41419-021-04322-2

Wang, X., Sun, S., Yang, X., Cheng, J., Wei, H., Li, Z., Michaud, J. P., & Liu, X. (2020). Variability of Gut Microbiota Across the Life Cycle of Grapholita molesta (Lepidoptera: Tortricidae). *Frontiers in Microbiology*, *11*, 1366. https://doi.org/10.3389/FMICB.2020.01366/BIBTEX

Wang, Y., Choi, J. Y., Roh, J. Y., Liu, Q., Tao, X. Y., Park, J. bin, Kim, J. S., & Je, Y. H. (2011). Genomic sequence analysis of granulovirus isolated from the Tobacco cutworm, spodoptera litura. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0028163

Watts, G. S., & Hurwitz, B. L. (2020). Metagenomic next-generation sequencing in clinical microbiology. Clinical Microbiology Newsletter, 42(7), 53-59.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(11), 5088–5090.

https://doi.org/10.1073/PNAS.74.11.5088/SUPPL_FILE/SCIENCEOFMICROBESPODCAST.MP3

Xia, X., Gurr, G. M., Vasseur, L., Zheng, D., Zhong, H., Qin, B., Lin, J., Wang, Y., Song, F., Li, Y., Lin, H., & You, M. (2017). Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00663

Yang, Q., Zhang, A. H., Miao, J. H., Sun, H., Han, Y., Yan, G. L., Wu, F. F., & Wang, X. J. (2019). Metabolomics biotechnology, applications, and future trends: a systematic review. *RSC Advances*, *9*(64), 37245–37257. https://doi.org/10.1039/C9RA06697G

Yang, Y., Liu, X., Xu, H., Liu, Y., Ali, P., Adnan Bodlah, M., & Lu, Z. (2020). The abundance and diversity of gut bacteria of rice leaffolder Cnaphalocrocis medinalis (Guenée) across life stages. *Journal of Asia-Pacific Entomology*, *23*(2), 430–438. https://doi.org/10.1016/J.ASPEN.2020.03.006

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., … Glöckner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, *29*(5), 415. https://doi.org/10.1038/NBT.1823

Yun, J. H., Roh, S. W., Whon, T. W., Jung, M. J., Kim, M. S., Park, D. S., Yoon, C., Nam, Y. do, Kim, Y. J., Choi, J. H., Kim, J. Y., Shin, N. R., Kim, S. H., Lee, W. J., & Bae, J. W. (2014). Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. *Applied and Environmental Microbiology*, *80*(17), 5254–5264. https://doi.org/10.1128/AEM.01226-14

Zalucki, M. P., Clarke, A. R., & Malcolm, S. B. (2003). Ecology and Behavior of First Instar Larval Lepidoptera. *Https://Doi.Org/10.1146/Annurev.Ento.47.091201.145220*, *47*, 361–393.

https://doi.org/10.1146/ANNUREV.ENTO.47.091201.145220

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614. https://doi.org/10.1093/BIOINFORMATICS/BTT593

Zhou, Q., Su, X., Wang, A., Xu, J., & Ning, K. (2013). QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data. *PLoS ONE*, *8*(4), 60234. https://doi.org/10.1371/journal.pone.0060234

Ziganshina, E. E., Mohammed, W. S., Shagimardanova, E. I., Vankov, P. Y., Gogoleva, N. E., & Ziganshin, A. M. (2018). Fungal, Bacterial, and Archaeal Diversity in the Digestive Tract of Several Beetle Larvae (Coleoptera). *BioMed Research International*, *2018*, e6765438. https://doi.org/10.1155/2018/6765438

Zurovec, M., Fedic, R., & Sehnal, F. (2002). The silk of Lepidoptera European Network of Bioadhesion Expertise (ENBA) View project Clonal biotechnological sericulture View project The Silk of Lepidoptera. In *Article in Journal of Insect Biotechnology and Sericology* (Vol. 71). https://www.researchgate.net/publication/280885330

Chapter 2

Structural analysis of the gut microbiota of *Antheraea assamensis* Helfer and its comparison with host leaf-associated microbiota

CHAPTER 2

Structural analysis of the gut microbiota of *Antheraea assamensis* Helfer and its comparison with host leaf-associated microbiota

Executive summery

Antheraea assamensis Helfer is an economically important, endemic, lepidopteran insect native to Northeast India that produces a lustrous golden-coloured silk of distinct quality and durability. To date, the gut microbiota of *A. assamensis* has remained largely unexplored. The present work aims to comprehensively identify and characterize the gut microbial community of *A. assamensis* through metagenomic approach. It also aims to investigate the influence of the host leaf–associated microbiota on larval gut microbial composition and its variation to changes in host plants. The results have indicated that the 5th instar larva of *A. assamensis* contains a highly diverse gut microbial community in terms of bacteria and archaea. The host leaf-associated microbiota was found to occupy a major portion of the total larval gut microbiota. However, the diversity of the larval gut microbiota was greater than the host leaf-associated microbiota. Besides a core set of microbiota, intraspecific variation of the gut microbial community was also observed among larvae reared separately on three different host plants. Predictive functional analysis of *A. assamensis* gut microbiome revealed a significant interaction between the host and its gut microbial community including host leaf digestion, metabolite detoxification, chitinase production and fat body metabolism. The findings of this study will illustrate the structure of the gut microbial community of *A. assamensis*, their key interactions with the host organism and the role of host leaf-associated microbiota on the holobiont.

2.1 Introduction

Amplicon sequencing of the 16S rRNA gene (16S rRNA analysis) is a widely utilized culture-independent technique to investigate microbial communities with the help of high-throughput NGS technologies. Numerous universal primers targeting the partial sequences in hypervariable regions (such as V1-V2, V1-V3, V3-V4, etc.) have been designed for microbiome study owing to the limitations of short-read NGS technology (Kameoka et al., 2021). It is now well-recognised that healthy gut flora plays a significant role in the overall health of the host organism. The normal gut microbiota plays key roles in the host biology such as the metabolism of dietary protein, carbohydrates, lipids, plant polyphenols and vitamins, defense against pathogens and essential amino acid provisioning (Rowland et al., 2018). Due to the advancements in genome sequencing technologies and bioinformatics, scientists are now able to investigate gut microbes, their function, and interactions with the host (Jandhyala et al., 2015). Studies on both humans and animals have shown that the diet of the host organism is an important determinant of the structure and function of the gut microbiota (Wen & Duffy, 2017). The composition and function of the gut microbiome are also influenced by a number of other variables such as host genetics, mode of reproduction and delivery (e.g. oviparity or viviparity), feeding behaviour, biotic and abiotic environmental factors, and geography (Gacesa et al., 2022; Hasan & Yang, 2019; Wen & Duffy, 2017). The gut microbiome also plays a major role in the host's development, adaptation toward its habitat and feeding preferences (Huang et al., 2021). This is particularly true for herbivorous insects with no endogenous cellulases. These insects generally rely on gut bacteria to mediate or facilitate cellulose breakdown (Martin, 1991). Comparative metagenomic analysis of 59 mammalian species revealed that their fecal microbiota could be grouped more closely according to their food habits rather than phylogeny (Sandberg et al., 2008). It has also been established that animal and their gut microbe evolve together profoundly affecting the host's radiation towards a wide range of habitats (Muegge et al., 2011). There is always a crosstalk between the gut microbiota and the host organism

influencing each other's key biological processes which eventually leads to the establishment of an intimate mutualistic interrelationship in the long evolutionary run (Huang et al., 2021).

To date, the gut microbiota of *Antheraea assamensis* has remained largely unexplored. Culture-dependent studies followed by taxonomic identification through 16S rRNA gene sequencing have found that the most dominant bacteria in the gut of *A. assamensis* include *Bacillus* (54%), *Serratia* (24%), *Pseudomonas* (10%), and *Alcaligenes* (6%) (Gandotra et al., 2018; Haloi et al., 2016). Pathogenicity assessment of *A. assamensis* gut microbiota revealed the presence of *Pseudomonas aeruginosa*, *Ornithinibacillus bavariensis*, *Achromobacter xylosoxidans* and *Staphylococcus aureus* in both healthy and diseased larvae. However, *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Bacillus thuringiensis* reduced the survivability of the larvae beyond a certain concentration (Haloi et al., 2016).

A. assamensis is a polyphagous herbivorous insect feeding primarily on *Machilus bombycina* (synonym: *Persea bombycine*, vern. name: Som plant) and *Litsea polyantha* Juss. (Synonym: *Litsea monopetala*, vern. name: Soalu plant). On the other hand, *L. citrata* Blume (Synonym: *L. cubeba*, vern. name: Mejankari) is a secondary host plant of *A. assamensis*. None of these host plants have been studied for their leaf-associated microbes through a culture-independent approach to date.

The present work aims to identify and characterize the structure of the gut microbial community of *A. assamensis* through culture-independent methods. Here, the structure of a microbial community in an environment implies the microorganisms present in that environment and their respective abundances. We also investigated the impact of host leaf-associated microbiota on larval gut microbial composition and their possible functional roles in the host (*A. assamensis*) biology. To find out the effect of host leaf microbiota on the gut microbial composition of *A. assamensis*, we performed metagenomic sequencing of three different host plant leaves as well as the gut of larvae reared on those host plants. Subsequently, a comparative analysis of the structure and function of the gut and leaf-associated microbiota was carried out.

The overview of the experiment is represented in **Figure 2. 1**.

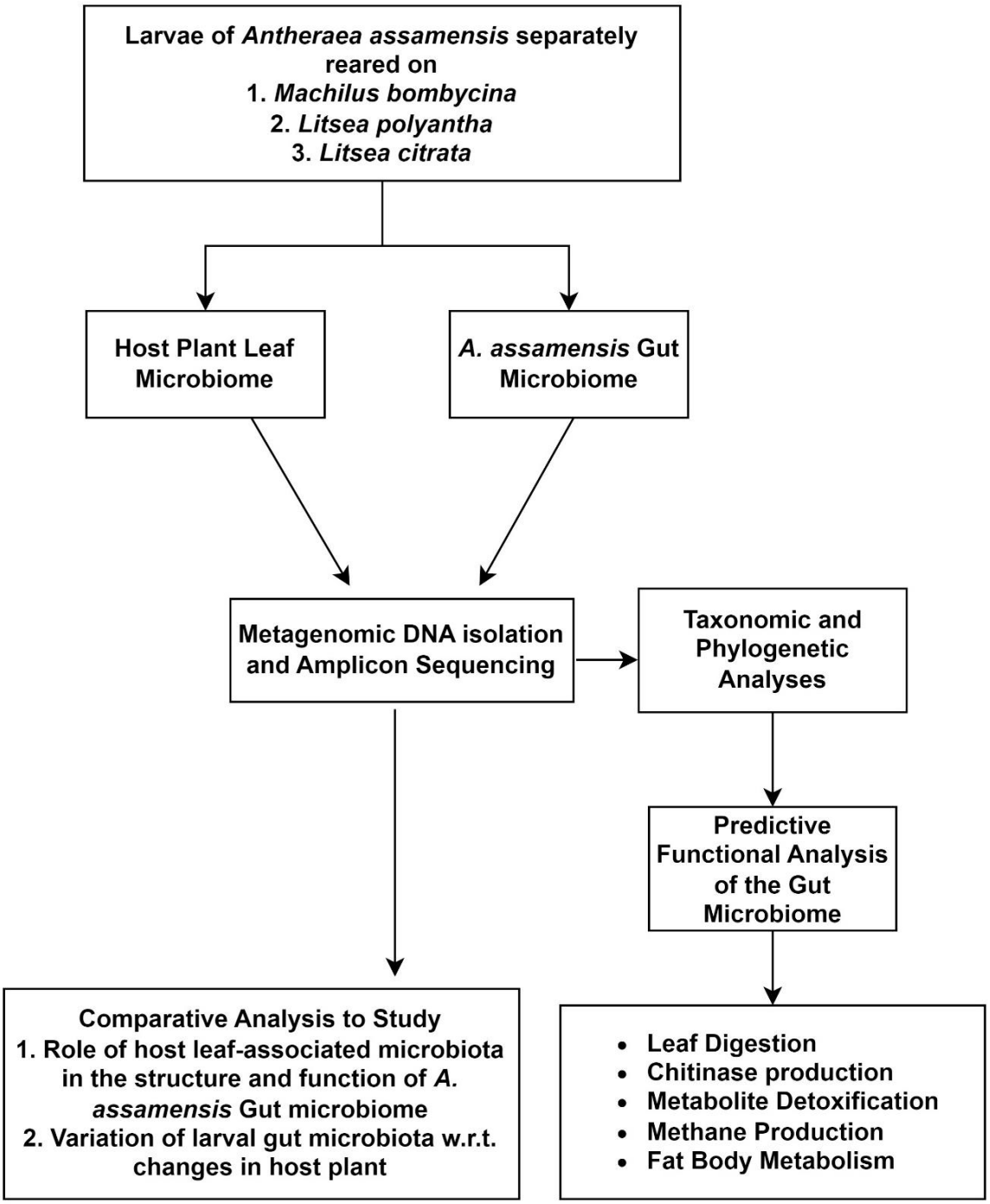


Figure 2. 1 Overview of the experimental design

2.2 Materials and methods

2.2.1 Sample collection

Healthy, fifth instar larvae of *A. assamensis* reared in natural conditions on two primary host plants, viz. *Machilus bombycina* (Som plant) and *Litsea polyantha* (Soalu plant) were collected

irrespective of sex from Central Muga and Eri Research and Training Institute (CMERTI), Jorhat, Assam (26°43'19.5"N 94°13'49.5"E). Mature leaves from the host plants *M. bombycina* and *L. polyantha* were also collected from the same location. Larvae of *A. assamensis* were reared on a secondary host plant, *Litsea citrata* (Mejankari plant) inside the Indian Institute of Technology Guwahati (IITG) campus (26°11'19.6"N 91°41'39.5" E). Healthy fifth instar larvae reared on *L. citrata* were collected irrespective of sex together with mature leaves of the host plant.

2.2.2 Sample preparation

Insect dissection was performed as described previously with some modifications (Shao et al., 2013). In brief, the insect was surface sterilized with 75% ethanol and mounted dorsoventrally on a dissection tray. Dissection was performed on the dorsal side from behind the head to the anus and the intact gut was safely removed. For each sample, guts from 5-6 individual larvae were pooled together. They were then rinsed with water and 1X PBS buffer and stored at -80°C in cryovials till metagenomic DNA was isolated. This procedure was repeated for all the gut samples.

Fresh leaves from 5-6 individual host plants were pooled together and finely ground with liquid nitrogen immediately after collection. This process was performed separately for each of the three host plants, *M. bombycina*, *L. polyantha* and *L. citrata*.

2.2.3 Metagenomic DNA isolation

Metagenomic DNA was isolated following the PCI method with some modifications (Heikrujam et al., 2020; Yuan et al., 2012). Briefly, the finely ground tissue of each sample was taken separately in 2 ml centrifuge tubes, added 500 µl of lysis buffer (1M Tris pH 8.0, 5M NaCl, 0.5M EDTA pH 8.0, 10% SDS, dH_2O) and Proteinase K (20 µL of 20 mg/ml stock solution per 1000 µl of the lysis buffer), incubated in a hot water bath at 65°C for 1 hour (mixed well by inversion at intervals). Each tube was then vortexed vigorously and centrifuged at 13000 rpm at 22°C for 15 min. The supernatants were transferred to a new 1.5 ml centrifuge

tube and added an equal volume of phenol: chloroform: isoamyl alcohol and mixed well by vortexing. The mixture was again centrifuged at the same speed and temperature for 15 mins. This step was repeated twice. Then 1/4th volume of 3M sodium acetate (pH 5.2) and double volume of 95-100% ice-cold ethanol were added to the sample and incubated overnight at 4°C. The next day, the mixture was centrifuged at 12000 rpm for 15 min at 4°C. The DNA pellet was repeatedly washed with 70% ethanol twice and finally with 90% ethanol, air-dried until the odour of ethanol disappeared, dissolved in the TE buffer and stored at 4°C. The quality and the quantity of the isolated DNA were checked using NanoDrop™ UV visible spectrophotometer.

2.2.4 High-throughput sequencing

DNA with concentration greater than 50 ng/µl and 260/280 absorbance ratio of ~1.8 was sent for sequencing at Eurofins Genomics India Pvt. Ltd. for amplicon sequencing. The V3-V4 region of the 16S rRNA gene was amplified with the primer pair 341 F: 5' CCTACGGGAGGCAGCAG 3' (V3) and 806 R: 5' GGACTACHVGGGTWTCTAAT 3' (V4). The amplified regions were sequenced using Illumina MiSeq 250×2 paired-end sequencing platform. Agarose gel electrophoresis of PCR amplified DNA was performed as a quality control measure to verify the amplified DNA. Details of the sequencing technique of the studied samples are given in **Table 2.1**.

Table 2. 1 Details of the sequencing technique of the studied samples.

Experiment type	16S rRNA amplicon sequencing
Region of amplification	V3-V4 region of 16S rRNA gene
Primer pair	341F: 5′ CCTACGGGAGGCAGCAG 3′ (V3)
	806 R: 5′ GGACTACHVGGGTWTCTAAT 3′ (V4)
Sequencing platform	Illumina MiSeq
Read length	2×250 (paire-end)

2.2.5 Amplicon sequence processing

2.2.5.1 Quality control

The forward and reverse raw read files in fastq format were checked for quality using the fastQC tool (Andrew, 2010). The primers and adapters were removed using Cutadapt (V3.4) (Martin, 2011). Quality control of the amplicon sequence data was performed utilizing DADA2, a quality control plugin implemented in QIIME2. Additionally, this quality control procedure removes the phiX reads and chimeric sequences that are often present in marker gene Illumina sequence data (Bolyen et al., 2019; Callahan et al., 2016). The paired-end sequences were denoised, dereplicated, and filtered for chimeras with *qiime dada2 denoise-paired* command. Different quality control parameters used in DADA2 are shown in **Table 2. 2**.

Table 2. 2 Quality control parameters for DADA2

Parameter	Description
--p-trunc-len-f	This determines the pisoition at which the forward reads should be truncated due to low quality. It also discards the reads shorter than the provided value. If the given value is 0, no truncation or length trimming will be performed.
--p-trunc-len-r	This determines the pisoition at which the reverse reads should be truncated due to low quality.
--p-trim-left-f	Indicates the postion at which forward read should be trimmed due to low base quality.
--p-trim-left-r	Indicates the postion at which reverse read should be trimmed due to low base quality.
--p-min-overlap	The minimum length of the overlapping region necessary for merging the forward and the reverse sequences
--p-chimera-method	Removes the chimera sequences

2.2.5.2 Alpha and Beta diversity analysis

Each sample was rarefied to a sampling depth of 100000 to eliminate the effect of heterogeneous sequence numbers among the samples. Alpha diversity is the measure of

microbial diversity and abundance in a single sample. Whereas, beta diversity measures the similarities and dissimilarities in terms of microbial diversity among different samples (Galloway-Pena et al., 2018). Both alpha and beta diversity analyses were done using the q2-diversity plugin of QIIME2. Alpha diversity was measured through the Shannon index, observed features, Simpson's index, Pielou's evenness and Good's coverage (Willis, 2019). A description of the different alpha diversity indices used in the study is given in **Table 2. 3**.

Table 2. 3 Description of the alpha diversity indices used in this study

Alpha diversity index	Description
Shannon index	It calculates the diversity and richness of microbial community in a sample considering the evenness and abundance of the present taxa (Shannon, 1949).
Chao1	This is one of the most important alpha diversity indices that measure the species richness of a sample. It is an extension of observed richness and is used for better estimation of species richness by detecting species with lower read counts (Chao, 1984).
Observed OTUs	It calculates the number of distinct operational taxonomic units (OTU) observed in a sample.
Simpson index	It measures community diversity by considering the relative abundance of the taxa (Simpson, 1949).
Pielou's index	It calculates the relative evenness of species richness in a sample. Here, the term 'species evenness' describes how each species is present in an environment close to each other in terms of number or abundance (Pielou, 1966).
Good's coverage	Calculates the percent of a species' total population that is represented in a sample (Good, 1953).

For beta diversity analysis, the values of the Bray-Curtis dissimilarity matrix were measured between 0 and 1, denoting exact and no similarity respectively, between the microbial composition of the samples (Bray & Curtis, 1957). Additionally, the UniFrac distance was used to compare microbial communities present in different samples by measuring the relative phylogenetic relatedness (or phylogenetic distance) of the community members (Lozupone &

Knight, 2005). Visualization of beta diversity was done using the Principal Coordinate Analysis (PCoA) or multidimensional scaling. Both alpha and beta diversity analyses were performed utilizing the *q2-diversity* plugin of QIIME2.

2.2.5.3 Phylogenetic diversity analysis

Phylogenetic diversity metrics including weighted and unweighted UniFrac distance, Faith's phylogenetic diversity were calculated by multiple sequence alignment using MAFFT implemented in QIIME2 (Katoh et al., 2002). These metrics require a rooted tree in addition to the number of features in each sample. The rooted phylogenetic tree was generated using the *align-to-tree-mafft-fasttree* pipeline of *q2-phylogeny* plugin. The pipeline first creates a QIIME 2 artifact of the aligned sequences by performing a multiple sequence alignment using the *mafft* programme on the sequences of each sample. The pipeline then masks (or filters) the alignment to eliminate positions with a high degree of variability. After that, the FastTree programme is used by the pipeline to create a phylogenetic tree from the masked alignment (Price et al., 2009). As FastTree generates an unrooted tree, the root of the tree is placed in the middle of the unrooted tree's longest tip-to-tip distance in the final step of this section.

2.2.5.4 Taxonomic analysis

Taxonomic classification was assigned against the filtered feature table through the q2-feature-classifier plugin implemented in QIIME2 with full-length Silva (release 138) (Quast et al., 2012), Greengenes (v13.8) (DeSantis et al., 2006) and NCBI GenBank databases trained with Naive-Bayes classifier using QIIME2 with 97% sequence similarity. The taxonomic composition was visualized using interactive bar plots.

2.2.5.5 Predictive functional profiling with PICRUSt2

The functional role of the microbial communities identified in the gut and leaf samples through amplicon sequencing was predicted utilizing PICRUSt2 (Douglas et al., 2020). The key steps in PICRUSt2 analysis included aligning the amplicon sequence variants (ASVs) with reference 16S sequences using HMMER (hmmer.org) (Eddy, 2020). The aligned sequences were then

placed in a reference phylogenetic tree using EPA-ng (Barbera et al., 2019) and GAPPA (Czech & Stamatakis, 2019). PICRUSt2 predicts the gene family abundances using CASTOR, an R package that performs hidden-state prediction through ancestral state reconstruction of the aligned sequences (Louca & Doebeli, n.d.). Metabolic pathways were inferred from gene families using MinPath (Ye & Doak, 2009) against the MetaCyc database (Caspi et al., 2016). Amplicon sequence variants with a Nearest Sequenced Taxon Index (NSTI) of less than 2 were not considered suitable for the analysis. The workflow for amplicon sequence analysis is represented in **Figure 2. 2**.

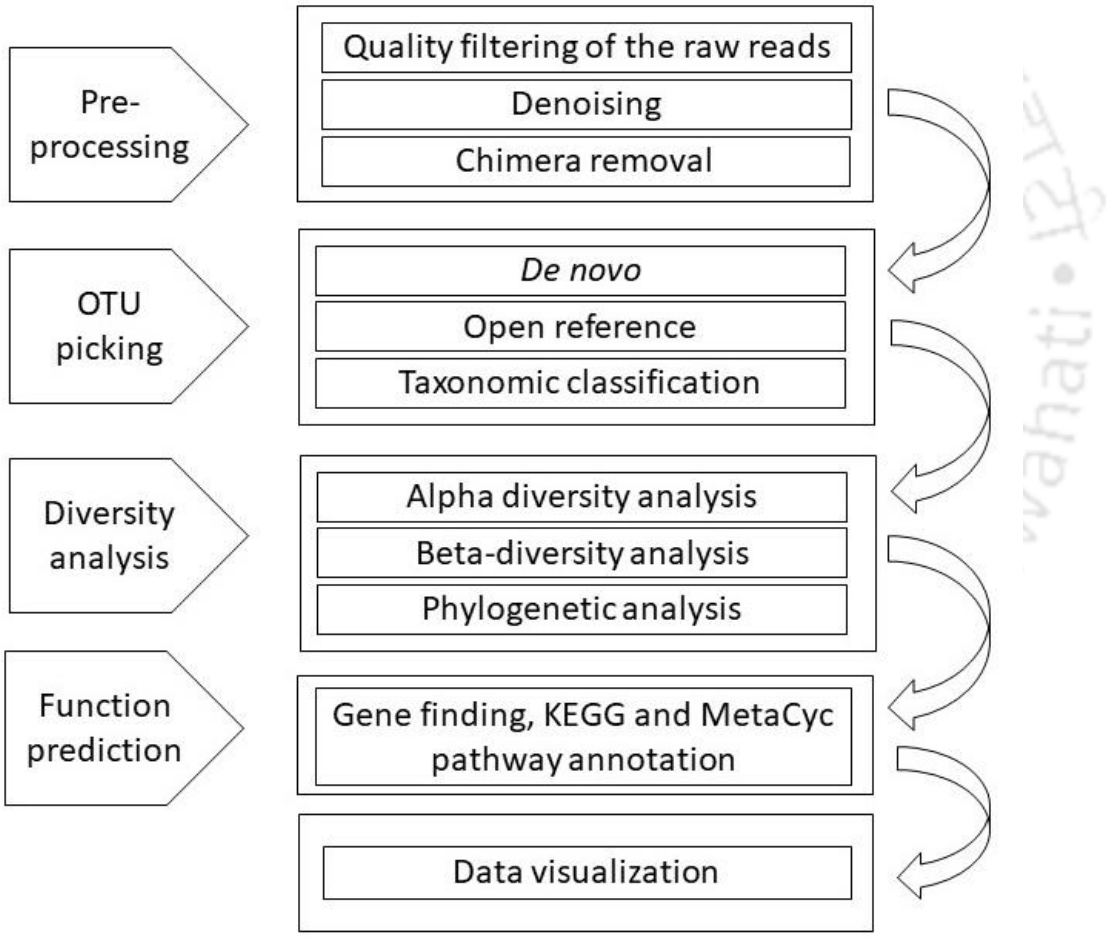


Figure 2. 2 The workflow for amplicon sequence analysis

2.3 Results

2.3.1 Amplicon Sequence Analysis

A total of 2327281 reads and an average of 387880 reads per sample were obtained through

2×250 paired-end Illumina MiSeq amplicon sequencing. The reads that passed the quality filtering step were used for downstream processing. Taxonomic assignment was done by aligning the quality-filtered sequences separately against Silva (version 138), Greengenes (13.8) and NCBI GenBank databases with ≥ 97% sequence similarity. However, for easy handling of the data and downstream processing, operational taxonomic units (OTUs) identified against only the Silva database were considered as the highest number of taxa were identified against this database in the majority of the samples (**Figure 2. 3**). Details of the studied samples and their sequence statistics are represented in **Table 2. 4**. Major bacterial taxa identified against the NCBI and Greengenes databases are represented in **Table 2. 5** and **Table 2. 6**, respectively.

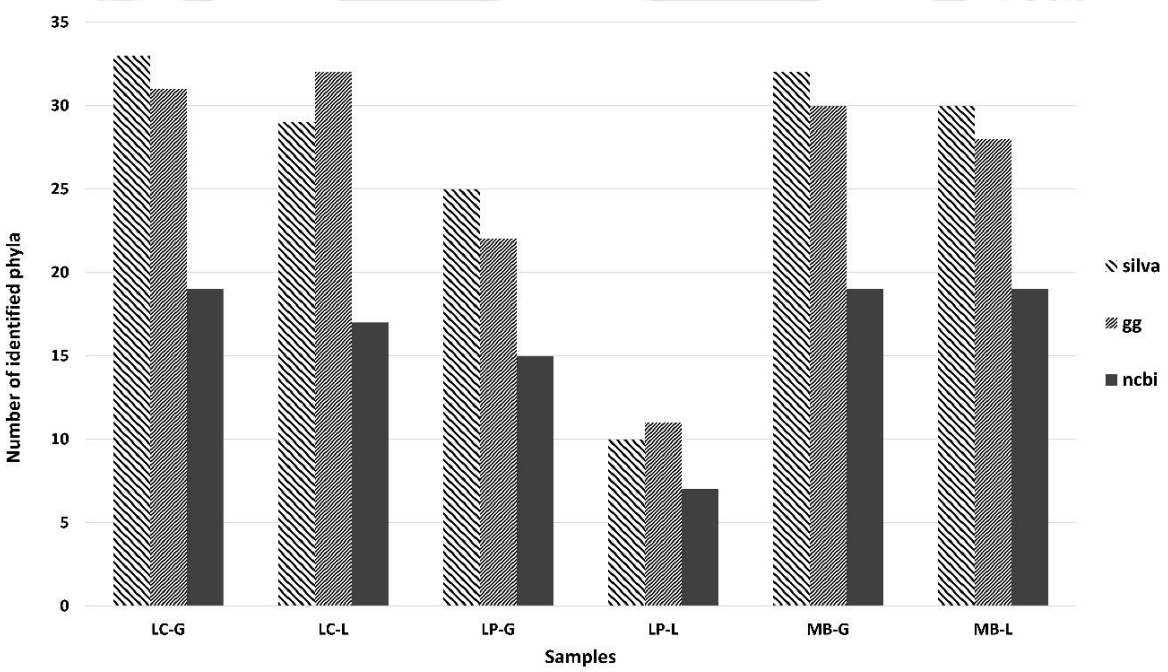


Figure 2. 3 Number of identified phyla against three different databases at the Phylum level.

Table 2. 4 Details of the samples used in the experiment and sequence statistics.

Description of the sample	Sample ID	Number of sequences received	Number of sequences retained post QC*	Percent of sequences passed QC
Gut of 5th instar larva of *A. assamensis* reared on *Litsea citrata* (Mejankari)	LC-G	277281	186641	67.31
Leaf of *Litsea citrata*	LC-L	227736	153544	67.42
Gut of 5th instar larva of *A. assamensis* reared on *Machilus bombycina* (Som plant)	MB-G	257044	173310	67.42
Leaf of *Machilus bombycina* (Som plant)	MB-L	254758	171636	67.37
Gut of 5th instar larva of *A. assamensis* reared on *Litsea polyantha* (Soalu)	LP-G	636336	486266	76.42
Leaf of *Litsea polyantha*	LP-L	674126	567912	84.24

*QC: Quality control

Table 2. 5 Identified taxa in the larval gut and host plant leaf of *A. assamensis* against the NCBI database with 97% similarity.

Identified taxa	LC-G	LC-L	LP-G	LP-L	MB-G	MB-L
p: Planctomycetes; c: Phycisphaerae; f: Tepidisphaeraceae; g: *Tepidisphaera*	23.39	13.77	0.00	0.00	24.00	21.41
p: Proteobacteria; c: Alphaproteobacteria; f: Sphingomonadaceae; g: *Sphingomonas*	20.00	25.29	1.06	18.15	19.85	19.76
p: Acidobacteria; c: Acidobacteriia; f: Bryobacteraceae; g: *Paludibaculum*	4.78	4.37	0.78	0.00	4.00	5.19
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiaceae; g: *Pedomicrobium*	4.43	5.22	0.09	0.00	3.64	4.98
p: Proteobacteria; c: Alphaproteobacteria; f: Bradyrhizobiaceae; g: *Bradyrhizobium*	4.17	5.84	1.06	0.00	4.44	4.16
p: Fusobacteria; c: Fusobacteriia; f: Fusobacteriaceae; g: *Cetobacterium*	3.45	3.72	0.07	0.00	4.12	3.38
p: Acidobacteria; c: Acidobacteriia; f: Acidobacteriaceae; g: *Acidobacterium*	2.89	3.30	0.43	0.00	2.54	2.82
p: Chloroflexi; c: Ktedonobacteria; f: Dictyobacteraceae; g: *Dictyobacter*	2.06	2.10	0.00	0.00	2.12	2.59
p: Actinobacteria; c: Acidimicrobiia; f: Ilumatobacteraceae; g: *Ilumatobacter*	1.98	1.38	0.21	0.00	1.78	1.76
p: Proteobacteria; c: Alphaproteobacteria; f: Erythrobacteraceae; g: *Tsuneonella*	1.94	2.23	0.00	0.00	1.81	2.03
p: Proteobacteria; c: Oligoflexia; f: Bdellovibrionaceae; g: *Bdellovibrio*	1.78	1.23	0.00	0.25	1.90	1.73
p: Firmicutes; c: Clostridia; f: Clostridiaceae; g: *Clostridium*	1.58	1.93	1.99	0.00	1.53	1.82
p: Actinobacteria; c: Acidimicrobiia; f: Iamiaceae; g: *Actinomarinicola*	1.36	1.53	0.59	0.00	1.34	1.82

Identified taxa	**LC-G**	**LC-L**	**LP-G**	**LP-L**	**MB-G**	**MB-L**
p: Proteobacteria; c: Alphaproteobacteria; f: Caulobacteraceae; g: *Phenylobacterium*	1.27	1.66	1.16	0.00	0.93	1.04
p: Planctomycetes; c: Planctomycetia;; f: Gemmataceae; g: *Zavarzinella*	1.16	0.54	0.00	0.00	1.55	1.24
p: Proteobacteria; c: Alphaproteobacteria; f: Devosiaceae; g: *Devosia*	0.96	1.04	1.61	0.00	1.06	0.78
p: Proteobacteria; c: Alphaproteobacteria; f: Reyranellaceae; g: *Reyranella*	0.93	0.66	32.54	2.39	0.00	0.60
p: Firmicutes; c: Clostridia; f: Peptostreptococcaceae; g: *Romboutsia*	0.92	1.50	0.74	0.00	0.89	1.23
p: Acidobacteria; c: Blastocatellia; f: Pyrinomonadaceae; g: *Brevitalea*	0.83	1.33	0.00	0.00	0.65	0.78
p: Proteobacteria; c: Alphaproteobacteria; f: Methylobacteriaceae; g: *Microvirga*	0.70	0.83	0.00	0.00	0.87	0.74
p: Proteobacteria; c: Betaproteobacteria; f: Oxalobacteraceae; g: *Massilia*	0.67	1.13	0.00	0.00	1.08	0.52
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiales; g: *Nordella*	0.65	0.80	0.00	0.00	0.55	0.82
p: Proteobacteria; c: Gammaproteobacteria; f: Xanthomonadaceae; g: *Lysobacter*	0.65	0.58	0.00	0.00	0.16	0.58
p: Acidobacteria; c: Blastocatellia; f: Blastocatellaceae; g: *Stenotrophobacter*	0.57	0.70	0.10	0.00	0.49	0.47
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiaceae; g: *Hyphomicrobium*	0.55	0.97	15.07	0.76	0.77	1.08
p: Acidobacteria; c: Acidobacteriia; f: Bryobacteraceae; g: *Bryobacter*	0.47	0.66	0.00	0.00	0.77	0.87
p: Proteobacteria; c: Alphaproteobacteria; f: Caulobacteraceae; g: *Brevundimonas*	0.46	0.46	4.66	0.00	0.40	0.46
p: Proteobacteria; c: Alphaproteobacteria; f: Methylobacteriaceae; g: *Methylobacterium*	0.00	0.00	0.98	39.38	0.00	0.57

Table 2. 6 Bacterial taxa identified in the larval gut and host plant leaf of *A. assamensis* against the Greengenes database with 97% similarity.

Identified Taxa	LC-G	LC-L	LP-G	LP-L	MB-G	MB-L
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiaceae; g: *Rhodoplanes*	21.00	22.62	1.66	0.00	20.58	20.18
p: Acidobacteria; c: Solibacteres; f: Solibacteraceae; g: Candidatus *Solibacter*	9.29	7.34	1.50	0.00	8.98	9.75
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiaceae; g: *Pedomicrobium*	6.75	7.40	0.34	0.00	5.31	7.16
p: Fusobacteria; c: Fusobacteriia; f: Fusobacteriaceae; g: *Cetobacterium*	4.89	4.54	0.09	0.00	5.73	4.71
p: Proteobacteria; c: Alphaproteobacteria; f: Sphingomonadaceae; g: *Sphingomonas*	3.85	4.39	1.11	17.34	3.34	3.25
p: Acidobacteria; c: Acidobacteriia; f: Koribacteraceae; g: Candidatus *Koribacter*	3.63	3.90	1.32	0.00	3.34	3.74
p: Proteobacteria; c: Deltaproteobacteria; f: Bdellovibrionaceae; g: Bdellovibrio	2.98	1.78	0.00	0.42	3.25	2.62
p: Firmicutes; c: Clostridia; f: Clostridiaceae; g: Clostridium	2.34	2.45	1.20	0.00	2.40	2.48
p: Proteobacteria; c: Alphaproteobacteria; f: Caulobacteraceae; g: Phenylobacterium	1.86	2.15	0.65	0.00	1.36	1.49
p: Planctomycetes; c: Planctomycetia; f: Gemmataceae; g: Gemmata	1.66	0.70	0.00	0.00	2.36	1.81
p: Proteobacteria; c: Alphaproteobacteria; f: Rhodospirillaceae; g: Reyranella	1.61	1.58	39.05	2.29	0.56	1.11
p: Proteobacteria; c: Alphaproteobacteria; f: Hyphomicrobiaceae; g: Devosia	1.37	1.34	1.96	0.00	1.51	1.14
p: Proteobacteria; c: Alphaproteobacteria; f: Bradyrhizobiaceae; g: Bradyrhizobium	1.14	1.81	0.49	0.00	1.37	0.86
p: [Thermi]; c: Deinococci; f: Trueperaceae; g: Truepera	1.07	1.04	0.00	0.00	1.34	2.13

Identified Taxa	**LC-G**	**LC-L**	**LP-G**	**LP-L**	**MB-G**	**MB-L**
p: Proteobacteria; c: Alphaproteobacteria; f: Bradyrhizobiaceae; g: Balneimonas	1.02	1.07	0.00	0.00	1.26	1.09
p: Proteobacteria; c: Alphaproteobacteria; f: Rickettsiaceae; g: Wolbachia	0.89	0.39	0.00	0.24	0.80	0.62
p: Actinobacteria; c: Actinobacteria; f: Micrococcaceae; g: Arthrobacter	0.80	0.82	0.14	0.00	0.77	0.13
p: Proteobacteria; c: Betaproteobacteria; f: Oxalobacteraceae; g: Massilia	0.74	0.92	0.00	0.00	1.23	0.30
p: Actinobacteria; c: Actinobacteria; f: Streptosporangiaceae; g: Nonomuraea	0.63	0.76	0.00	0.00	0.80	0.82
p: Armatimonadetes; c: Fimbriimonadia; f: Fimbriimonadaceae; g: Fimbriimonas	0.56	0.58	0.00	0.54	0.61	0.55
p: Proteobacteria; c: Alphaproteobacteria; f: Caulobacteraceae; g: Mycoplana	0.56	0.60	0.91	0.00	0.52	0.42
p: Proteobacteria; c: Alphaproteobacteria; f: Phyllobacteriaceae; g: Mesorhizobium	0.55	0.61	0.18	0.00	0.66	0.18
p: Proteobacteria; c: Gammaproteobacteria; f: Xanthomonadaceae; g: Lysobacter	0.45	0.25	0.00	0.00	0.05	0.34
p: Actinobacteria; c: Acidimicrobiia; f: Iamiaceae; g: Iamia	0.43	0.00	0.00	0.00	0.33	0.34
p: Proteobacteria; c: Gammaproteobacteria; f: Xanthomonadaceae; g: Thermomonas	0.41	0.26	0.00	0.00	0.08	0.47
p: Proteobacteria; c: Alphaproteobacteria; f: Beijerinckiaceae; g: Beijerinckia	0.37	0.07	0.00	33.59	0.37	0.77
p: Proteobacteria; c: Alphaproteobacteria; f: Methylobacteriaceae; g: Methylobacterium	0.00	0.00	0.66	36.91	0.00	0.84
p: Proteobacteria; c: Alphaproteobacteria; f: Rhizobiaceae; g: Agrobacterium	0.37	0.00	0.73	0.00	0.14	0.40

k: Kingdom, p: Phylum, c: Class, f: Family, g: Genus

2.3.2 Structure of the gut microbial community of *A. assamensis*

A total of 382 OTUs from 186641 reads in LC-G, 248 OTUs from 486266 reads in LP-G and 370 OTUs from 173310 reads in MB-G were assigned under the domains Bacteria and Archaea after removing the singletons, chloroplast and mitochondrial sequences. The identified OTUs of the gut samples were clustered into 34, 22 and 33 phyla in LC-G, LP-G and MB-G respectively (for sample details, kindly refer to **Table 2. 4**). Taxonomic assignment of the ASVs showed that the majority of the bacterial phyla of LC-G composed of Proteobacteria (25.84%), Chloroflexi (8.35%), Planctomycetes (12.08%), Patescibacteria (11.80%), Firmicutes (1.94%), Acidobacteria (7.62%), Bacteroidota (1.06%), Actinobacteria (4.03%), Fusobacteria (1.34%) and Verrucomicrobia (0.38%). Further characterization of the bacterial diversity at other taxonomic levels resulted in *Sphingomonas* (6.46%), WD2101 soil group (9.43%), *Candidatus kaiserbacteria* (4.42%), uncultured Xanthobacteraceae (3.85%), *Candidatus solibacter* (2.49%), LWQ8 (1.99%), *Pedomicrobium* (1.89%), *Bryobacter* (1.59%), *Bradyrhizobium* (1.36%), Actinobacterial genus IMCC26256 (1.27%), *Cetobacterium* (1.32%), uncultured Chloroflexi RBG-13-54-9 (0.39%) and KD4-96 (0.89%), uncultured Alphaproteobacteria (0.85%) and *Reyranella* (0.42%).

The archaeal diversity of LC-G consists of Thermoplasmata (6.40%), Crenarchaeota (2.33%), Nanoarchaeota (1.23%), Halobacteria (0.52%), and Euryarchaeota (0.31%) at the phylum level. At the lower taxonomic level, taxa like Marine Group II (6.21%), Nitrososphaeraceae (1.00%), *Candidatus nitrocosmicus* (0.62%), *Methanosarcina* (0.42%), uncultured Nanoarcheaon GW2011_GWC1_47_15 (0.38%) and *Methanobrevibacter* (0.23%) were present. 11.70% of the total ASVs remained unassigned in LC-G.

Similarly, the microbiota of MB-G showed the presence of the bacterial phyla Proteobacteria (24.90%), Chloroflexi (8.43%), Planctomycetes (12.24%), Patescibacteria (11.98%), Firmicutes (1.95%), Acidobacteria (7.20%), Bacteroidota (0.98%), Actinobacteria (3.76%) and Fusobacteria (1.53%). Bacterial diversity of MB-G at lower taxonomic level consists of

WD2101 soil group (9.36%), *Sphingomonas* (6.49%), *Candidatus kaiserbacteria* (4.42%), uncultured Xanthobacteraceae (3.79%), *Candidatus solibacter* (2.32%), LWQ8 (1.99%), *Pedomicrobium* (1.56%), *Bryobacter* (1.58%), *Bradyrhizobium* (1.36%), IMCC26256 (1.26%), *Cetobacterium* (1.52%), RBG-13-54-9 (0.38%), KD4-96 (1.02%), JG30-KF-CM45 (0.29%), AD3 (0.70%), *Devosia* (0.37%), SBR1031 (0.54%), *Bdellovibrio* (0.82%), TK10 (0.56%), SM1A02 (0.73%), *Clostridium* (0.35%). 12.40% of the total ASVs of MB-G remained unassigned. The archaeal community of MB-G consists of the phyla Thermoplasmata (6.85%), Crenarchaeota (2.31%), Nanoarchaeota (1.16%), Halobacteria (0.64%), and Euryarchaeota (0.33%).

Compared to the other two samples, the microbial community of LP-G was observed to be less species-rich where the majority of the reads (70.74%) remained unassigned. The remaining reads were assigned to phyla Proteobacteria (18.51%), Firmicutes (3.12%), Chloroflexi (2.31%), Acidobacteria (1.03%) and Actinobacteria (0.79%). In the genus level, the most abundant OTUs were *Reyranella* (4.50%), *Hyphomicrobium* (2.03%), *Actinobacteria* (3.59%), *Bradyrhizobium* (1.36%), *Pedomicrobium* (1.89%), *Devosia* (0.70%), and uncultured bacterium AD3 (Chloroflexi) (0.35%). Only a minor fraction of the total ASVs was categorized to the archaeal community. Taxa belonging to the archeal community consisted of phyla Thermoplasmatota (1.02%), Crenarchaeota (0.62%) and Halobacterota (0.20%). The relative abundances of the bacterial and archaeal taxa identified in the gut of *A. assamensis* at the hylum level and below were represented in **Figure 2. 4**.

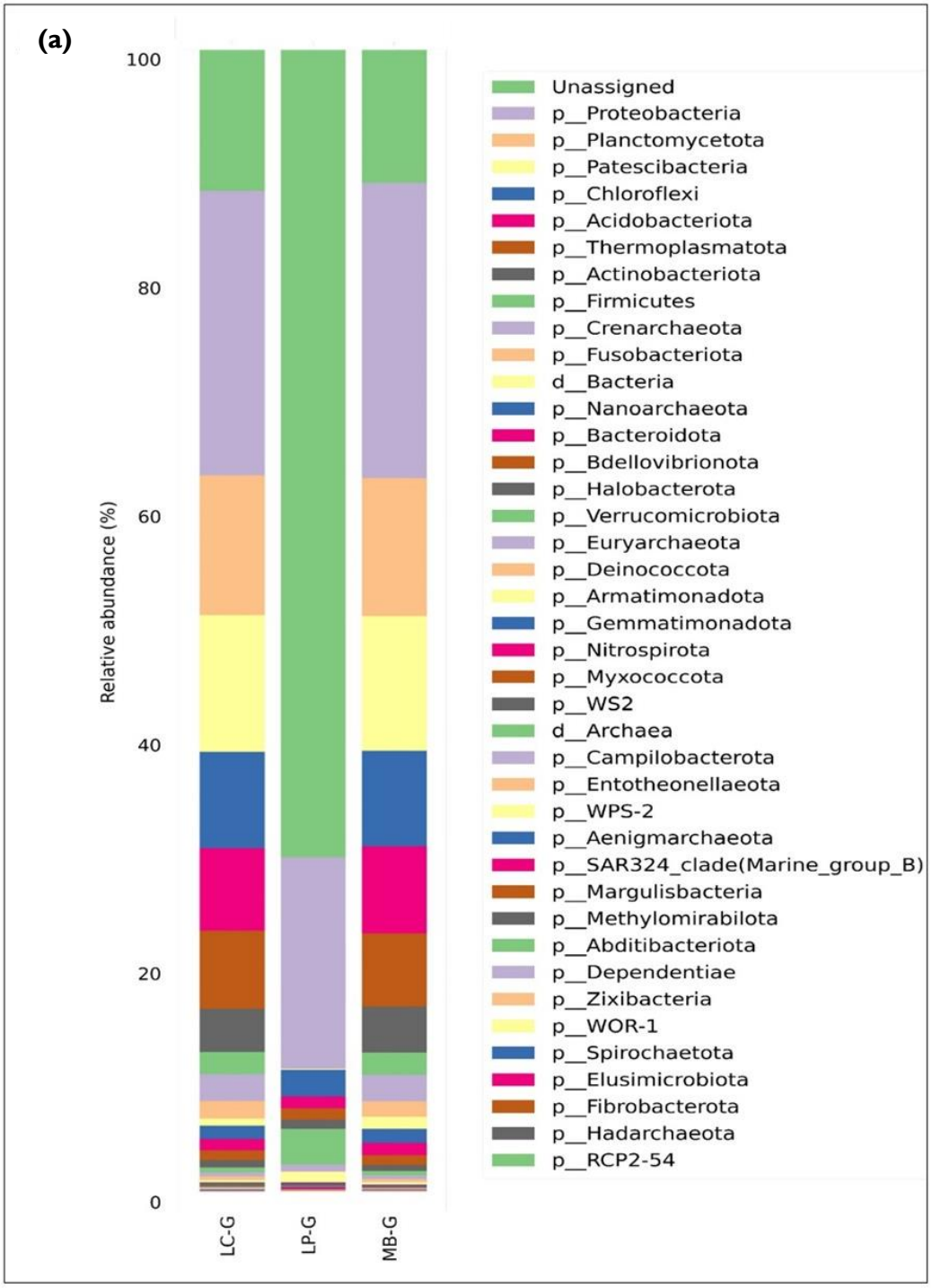


Figure 2. 4(a) Relative abundance of the gut-associated bacterial and archaeal community of *A. assamensis* larvae reared on three different host plants at the phylum level identified against the Silva database (*p: phylum*).

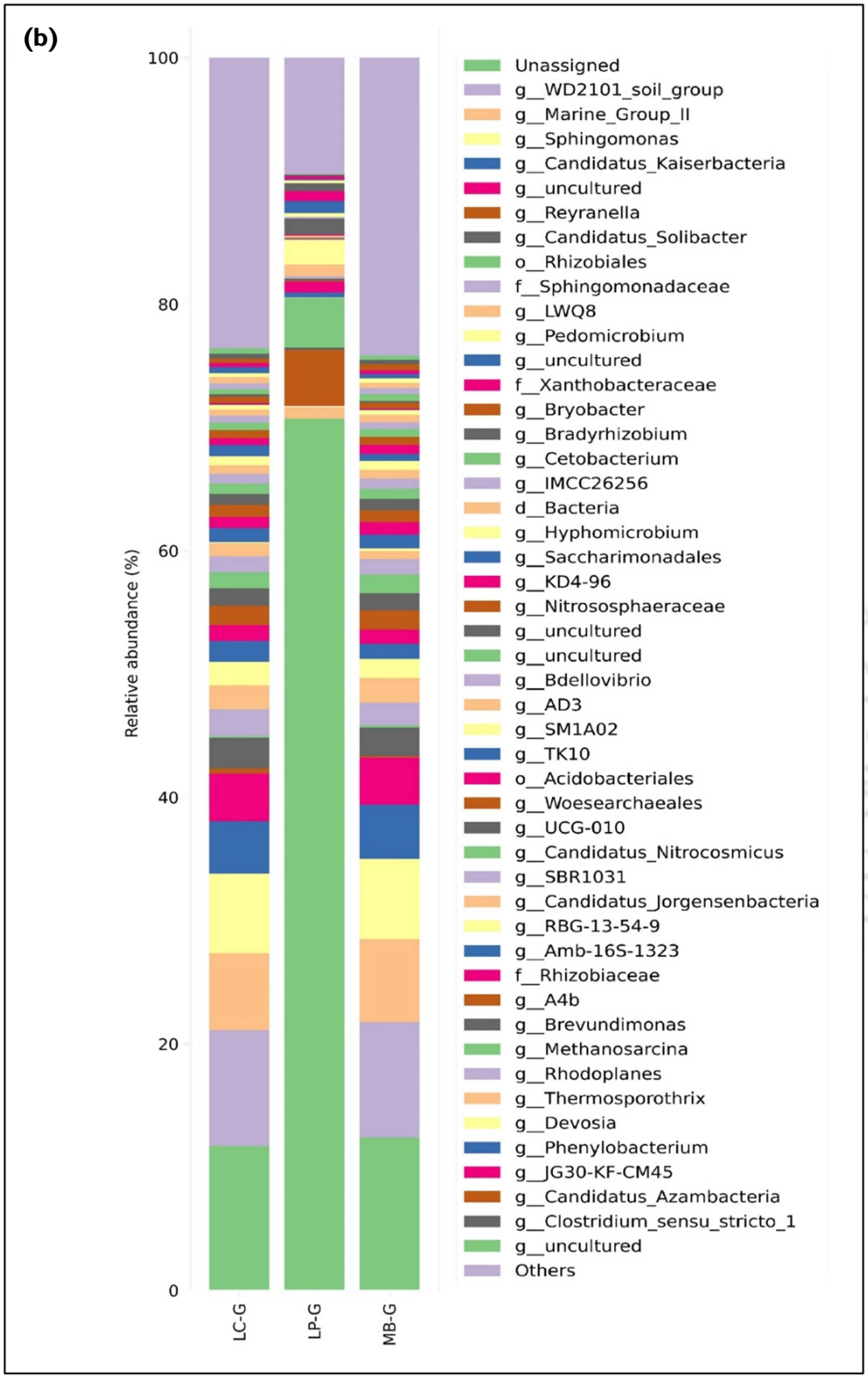


Figure 2. 4(b) Relative abundance of the gut-associated bacterial and archaeal community of *A. assamensis* larvae reared on three different host plants at the order level or below identified against the Silva database (*o: order, f: family, g: genus*)

2.3.3 Structure of the leaf microbial community of three different host plants of *A. assamensis*

At the phylum level, the bacterial community of the leaves of *L. citarata* (LC-L) contained Proteobacteria (41.75%), Chloroflexi (9.77%), Planctomycetota (9.21%), Acidobacteriota (10.43%), Patescibacteria (8.14%), Actinobacteria (5.21%), Firmicutes (3.10%), Fusobacteriota (1.82%), Bacteroidota (0.98%), Bdellovibrionota (0.71%), Armatimonadota (0.31%), Verrucomicrobiota (0.23%), Gemmatimonadota (0.17%) and Myxococcota (0.12%). Further classification of those phyla resulted into *Sphingomonas* (10.55%), Unidentified Sphingomonadaceae (3.06%), Unidentified Xanthobacteraceae (1.90%), *Bdellovibrio* (0.67%), *Reyranella* (0.59%) uncultured bacterium JG30-KF-CM45 (Chloroflexi) (0.34%) *Hyphomicrobium* (0.17%), Unidentified Rhizobiales (0.18%), Unidentified Fimbriimonadaceae (0.22%), uncultured Roseiflexaceae (0.64%), and *Wolbachia* (0.15%).

In the leaf sample of *M. bombycina* (MB-L), Proteobacteria (27.32%), Chloroflexi (8.90%), Planctomycetes (12.01%), Acidobacteria (8.51%), Patescibacteria (12.14%), Actinobacteria (4.30%), Firmicutes (2.07%), Fusobacteria (1.37%), Bacteroidota (0.91%), Bdellovibrionota (0.81%), Verrucomicrobia (0.43%), Armatimonadota (0.25%), Gemmatimonadetes (0.07%), and Myxococcus (0.05%). Other highly abundant taxa in MB-L include *Sphingomonas* (7.05%), unidentified LWQ8 (2.31%), unidentified Sphingomonadaceae (1.90%), uncultured Roseiflexaceae (0.97%), *Bdellovibrio* (0.72%), *Reyranella* (0.31%), and *Methylobacterium* (0.28%). In the leaf sample of *L. polyantha* (LP-L), 88.17% of the total ASVs remained unassigned. The assigned ASVs belonged to phyla Proteobacteria (10.69%), Acidobacteria (0.65%), Chloroflexi (0.13%), Actinobacteria (0.09%), Patescibacteria (0.06%), Firmicutes (0.03%), Bacteroidota (0.02%), Bdellovibrionota (0.03%), and Armatimonadota (0.04%). Other taxa like *Methylobacterium* (3.09%), *Methylocella* (2.28%), 1174-901-12 (Beijerinckiaceae) (1.73%), *Sphingomonas* (1.04%), unidentified Sphingomonadaceae (0.66%), Unidentified Beijarinckiaceae (0.56%), *Reyranella* (0.18%), Uncultured

Caulobacteraceae (0.14%), and *Aureimonas* (0.14%) were found to be present. The relative abundance of various taxa at the phylum level and below were represented in **Figure 2. 5**.

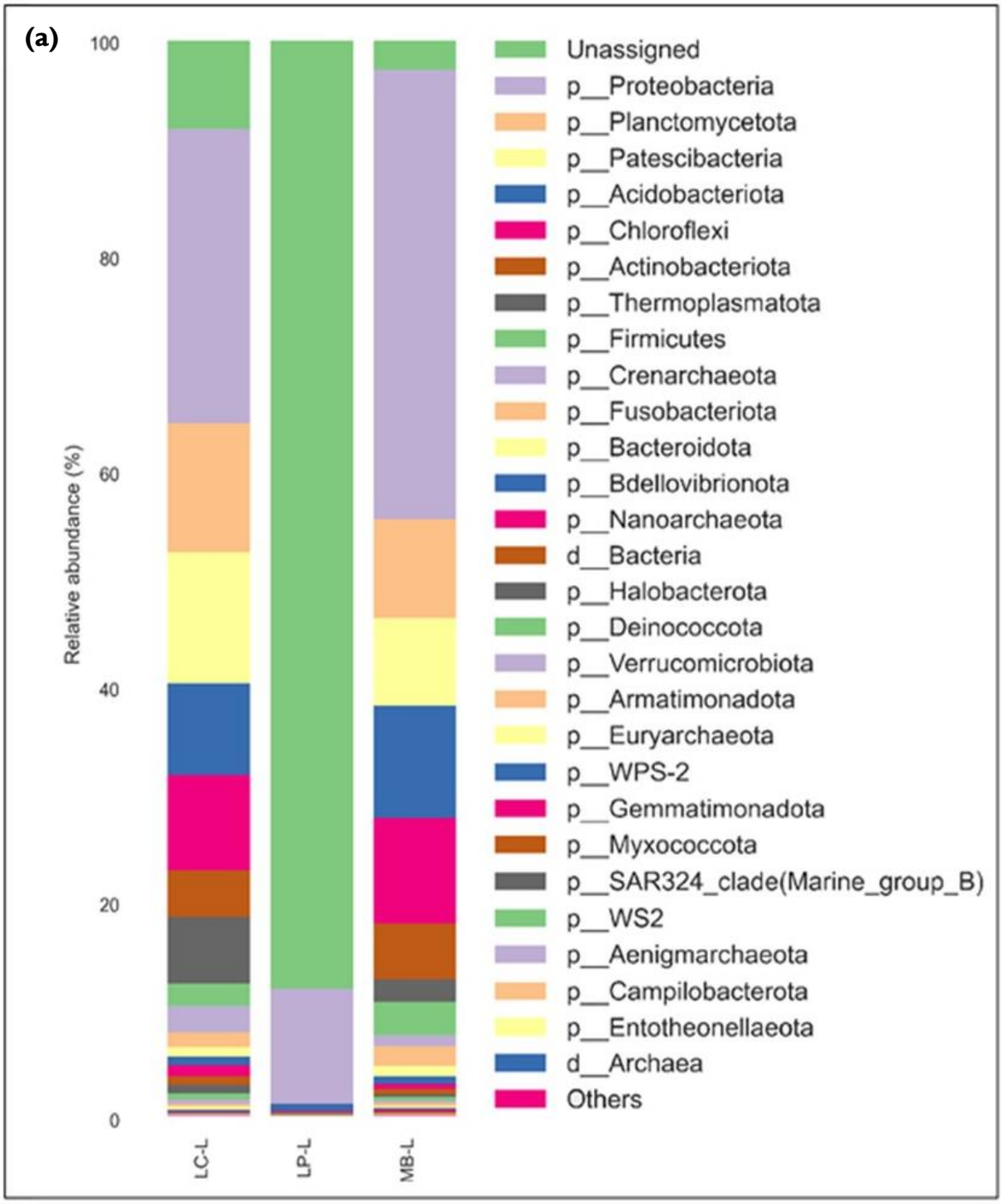


Figure 2. 5(a) Relative abundance of the leaf-associated bacterial and archaeal communities of three different host plants of *A. assamensis* at the phylum level identified against the Silva database (*p: phylum*).

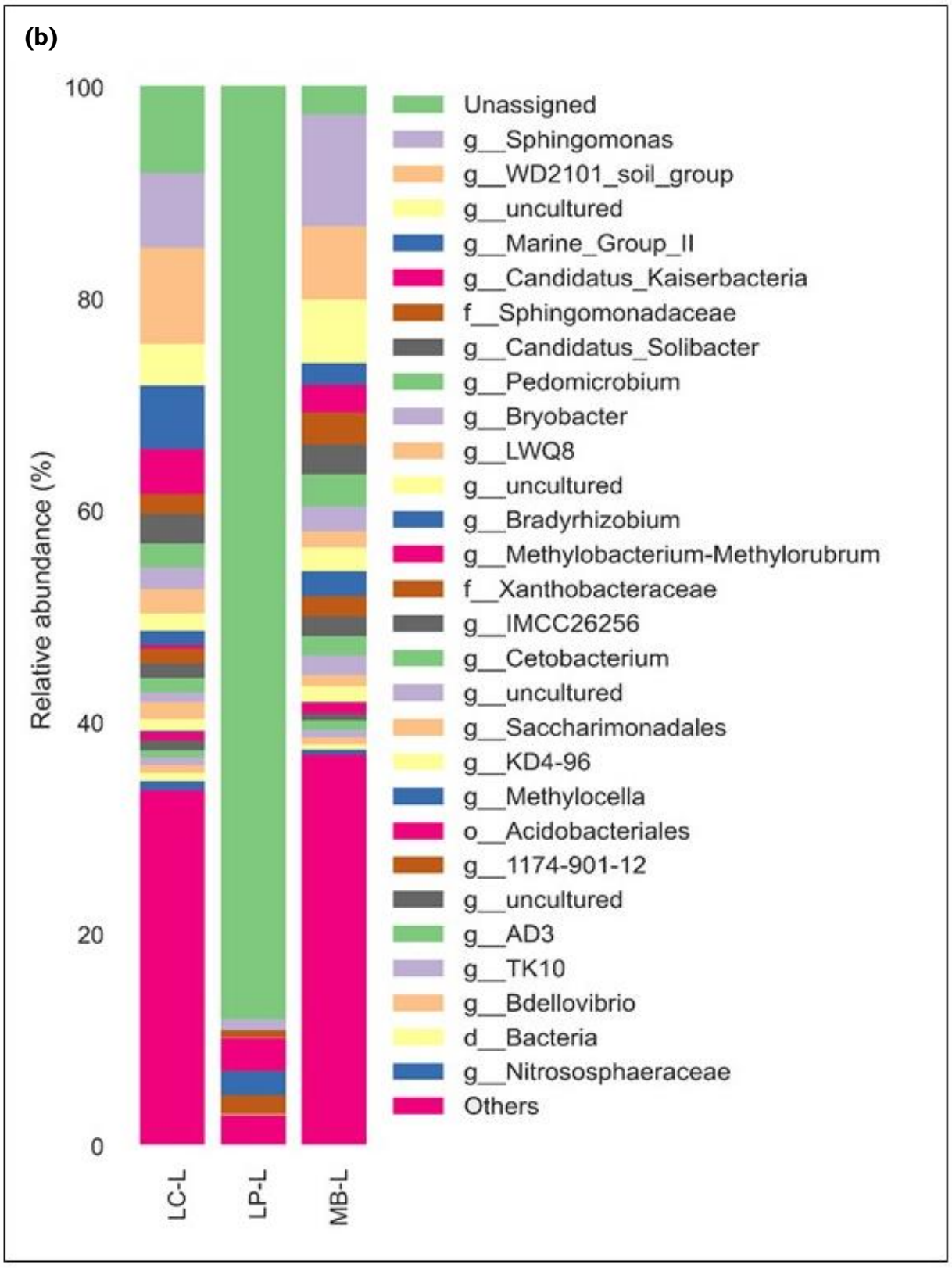


Figure 2. 5(b) Relative abundance of the leaf-associated bacterial and archaeal communities of three different host plants of *A. assamensis* at the order level or below identified against the Silva database (*o: order, f: family, g: genus*).

2.3.4 Identification of the core and unique gut microbial taxa of *A. assamensis* reared on three different host plants

The change in gut microbial composition with the change of host plant was studied in three gut samples, LC-G, MB-G and LP-G from larvae of *A. assamensis* reared on three different host plants, *L. citrata*, *L. polyantha* and *M. bombycina*, respectively. The core gut microbial community of *A. assamensis* was identified using the command *feature- table core-features* of QIIME2. We found 95 OTUs to be commonly present in MB-G, LC-G and LP-G. These OTUs mainly belong to phyla Proteobacteria, Acidobacteria, Chloroflexi, Patescibacteria, Gemmatimonadetes, Firmicutes, Chloroflexi, Actinobacteria, Acidobacteria, candidate phylum WPS-2, Thermoplasmata, Nanoarchaeota. Crenarchaeota and Euryarchaeota. At the genus level, *Sphingomonas*, *Paracoccus*, *Pseudorhodoplanes*, *Bradyrhizobium*, *Rhodomicrobium*, *Mesorhizobium*, *Pedomicrobium*, *Hyphomicrobium*, *Devosia*, *Reyranella*, *Dongia*, *Phenylobacterium*, *Brevundimonas*, *Staphylococcus*, and *Bryobacter*. were found to be shared by LC-G, LP-G and MB-G. Similarly, 280 OTUs were found to be common between LC-G and MB-G. Between LC-G and LP-G, 111 OTUs were common whereas, 114 OTUs were shared by LP-G and MB-G. the number of unique OTUs identified in MB-G, LC-G and LP-G were 71, 86 and 118, respectively.

The unique OTUs in MB-G at the genus level contained *Methanocella*, *Bifidobacterium*, *Blastococcus*, *Glutamicibacter*, *Amycolatopsis*, *Thermopolyspora*, *Rubrobacter*, *Prevotella* and *Rhodocytophaga*. Similarly, *Nitrosarchaeum*, Hadarchaeales, uncultured Methanomethylophilaceae, *Paludibaculum*, *Acidipila* and *Rothia* were present only in LC-G. In LP-G *Candidatus* Nitrosopumilus, *Halococcus*, *Candidatus* Methanoperedens, *Rhodococcus*, *Microbacterium* and *Cutibacterium* were uniquely present with a read count ≥4. The number of unique and shared OTUs among the gut samples were represented in **Figure 2. 6(a)**.

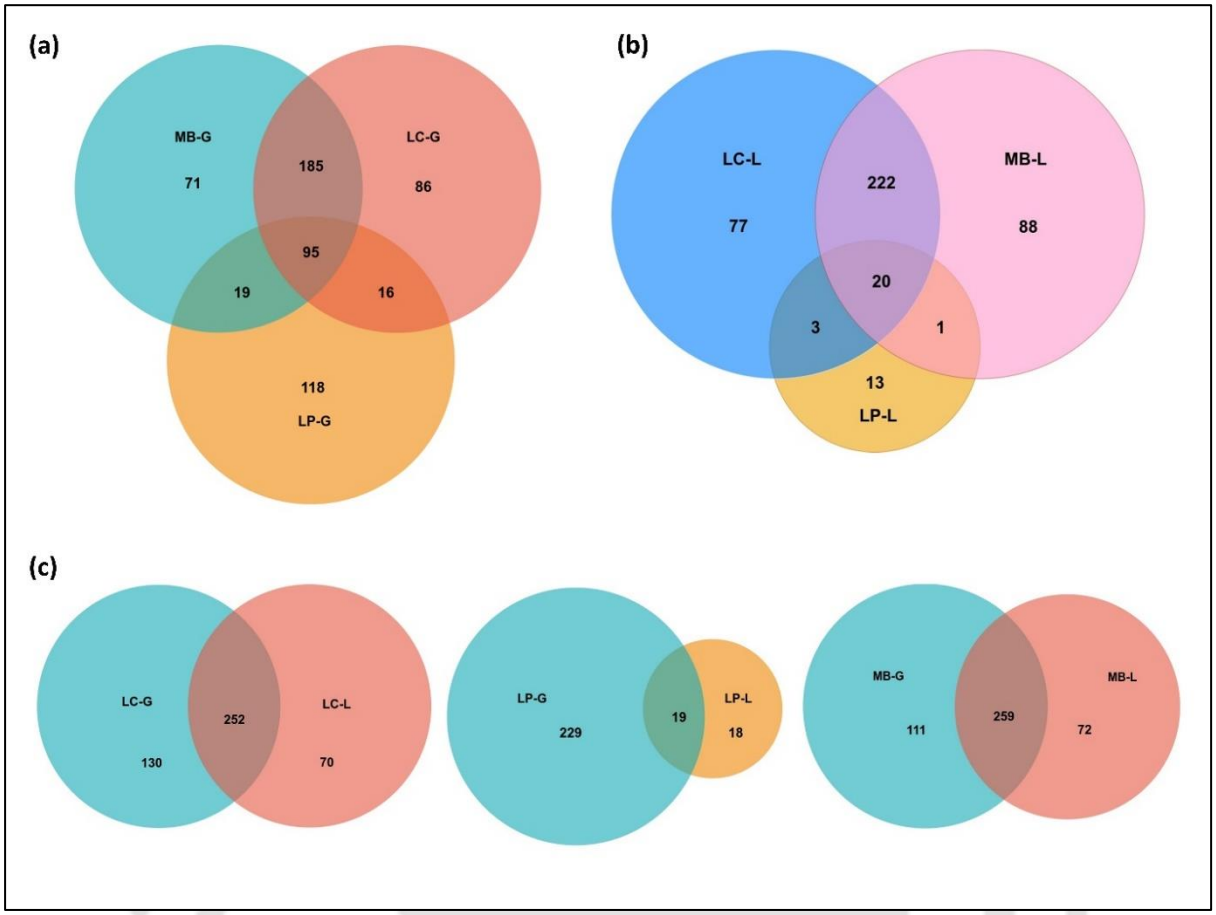


Figure 2. 6 Venn diagram depicting the shared and unique OTUs (a) among the larval gut samples MB-G, LC-G and LP-G; (b) among the leaf samples of three different host plants, LC-L, MB-L and LP-L; (c) between the larval gut and its host plant leaf.

2.3.5 Identification of the core microbial taxa in the leaves of the host plants

The core microbial community of the leaf samples of three different host plants of *A. assamensis* was done as described in the preceding section. 20 OTUs have been identified to be shared by all three leaf samples, i.e., LC-L, LP-L and MB-L (**Figure 2. 6(b)**). At the phylum level, these OTUs belong to Proteobacteria, Thermoplasmatota, Patescibacteria, Bdellovibrionota, Chloroflexi, Armatimonadota and Acidobacteriota. At the genus level, *Sphingomonas*, *Sphingomonas*, *Reyranella*, *Hyphomicrobium*, *Methylocella*, *Wolbachia* and *Blastocatella* were identified to be shared by LC-L, LP-L and MB-L.

2.3.6 Comparison between the gut and host leaf microbial composition

Comparison between the identified taxa in the larval gut and host plant leaf showed that out of

382 OTUs identified in LC-G, 252 OTUs (65.97%) were shared with its host leaf microbiota (i.e. LC-L). Similarly, 259 OTUs (70% of MB-G) were shared by MB-G and MB-L. On the other hand, we could assign only 37 OTUs from the leaf sample of *L. polyantha* (LC-L), out of which half were shared with the gut microbiota of the larvae reared on it (i.e. only 7.9% of LC-G). The number of shared OTUs between the host leaf and larval gut were represented in **Figure 2. 6(c)**. Only 11 OTUs were identified to be shared by all the gut and host leaf samples. At the phylum level, they were assigned to Proteobacteria, Thermoplasmatota, Chloroflexi, and Acidobacteriota. The core taxa (taxa present in all the gut and leaf samples) identified at the genus level consisted of *Sphingomonas*, *Reyranella*, *Hyphomicrobium* and *Blastocatella*.

2.3.7 Alpha and beta diversity analysis

Alpha diversity analyses of the gut and leaf samples were performed through the q2-diversity plugin of QIIME2 using various alpha-diversity indices viz., Shannon index, observed features, Simpson's index, Pielou's evenness and Good's coverage. The rarefaction plot based on observed features showed saturated sampling depth for all six samples (**Figure 2. 7(a)**). An average Good's coverage of >0.99 suggested that enough sequencing coverage has been accomplished to include the majority of the microbial species of each sample. The results of alpha diversity analysis are depicted in **Table 2. 7**. Shannon index, a commonly used alpha diversity index (Shannon, 1948) showed that MB-G was the most species-rich among the gut samples followed by LC-G and LP-G. Similarly, among the leaf samples, MB-L was found to contain the highest number of species followed by LC-L and LP-L. However, this variation in alpha diversity is not significant as indicated by the pairwise, non-parametric Kruskal-Wallis test for the Shannon index (**Table 2. 8(a)**). Faith's phylogenetic diversity metrics were calculated undertaking *de novo* multiple sequence alignment through MAFFT (Multiple

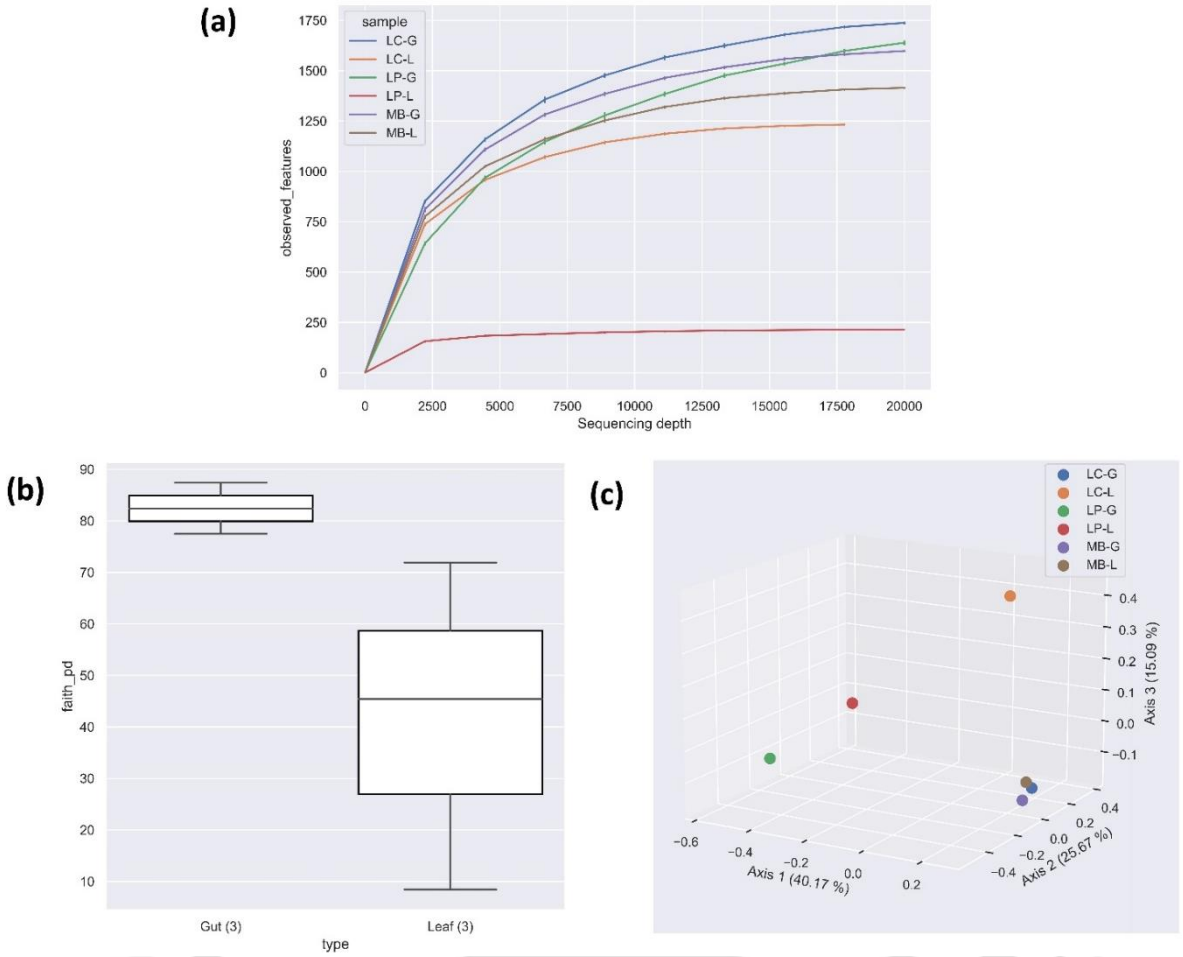


Figure 2. 7 Alpha and beta diversity analysis of the larval gut and host plant leaf samples; (a) Rarefaction plot of the observed features, (b) alpha-diversity analysis measured by Faith's PD matrix, (c) Principal Coordinate Analysis on the unweighted UniFrac distance matrix.

Alignment using Fast Fourier Transform) via the q2-alignment plugin implemented in QIIME2. Kruskal-Wallis test on Faith phylogenetic diversity vector showed a statistically significant phylogenetic difference between the microbial communities of the gut and leaf groups (**Figure 2. 7(b), Table 2. 8(b)**). A phylogenetic tree at the phylum level was constructed by importing the rooted tree and the taxonomic annotation file of QIIME2 into iTOL (Letunic & Bork, 2007) (**Figure 2. 8**).

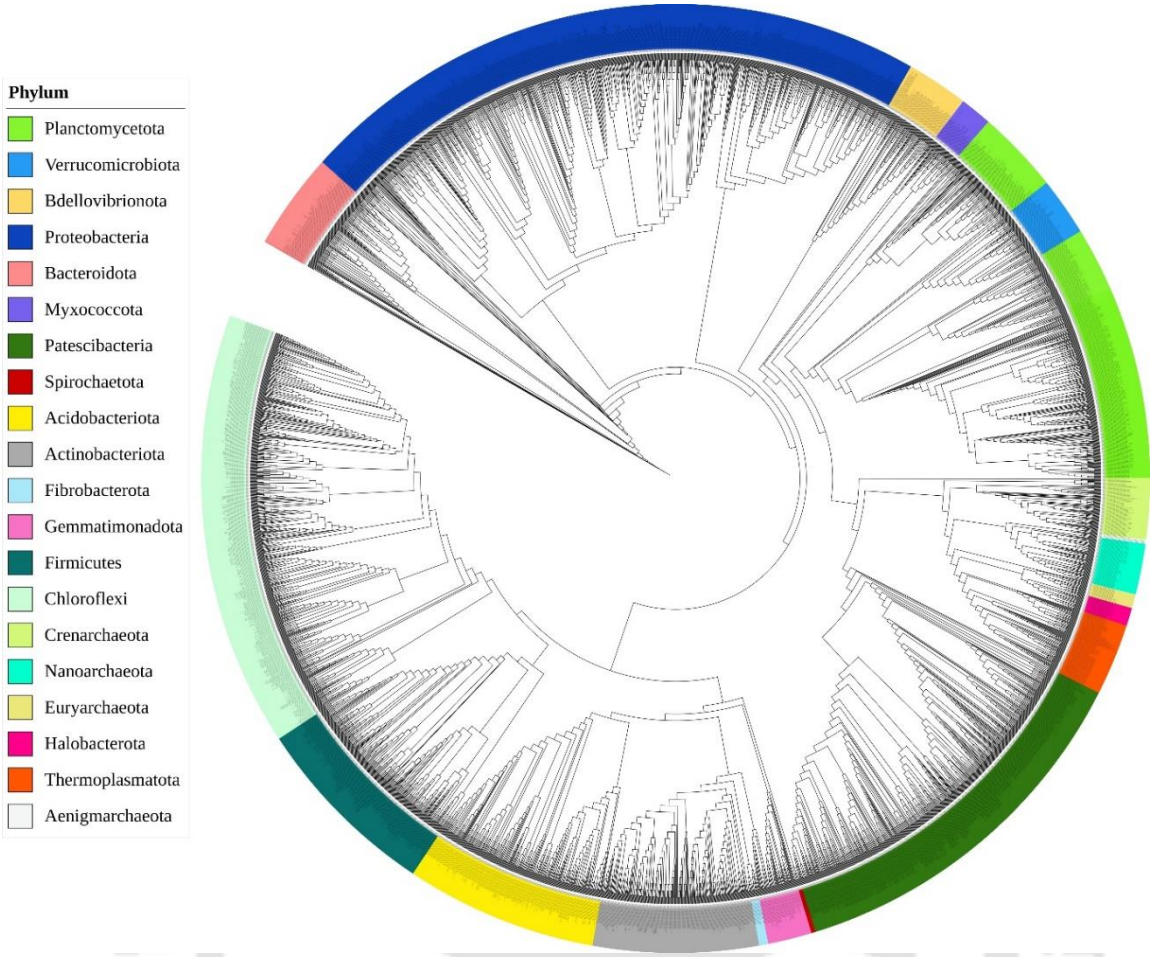


Figure 2. 8 A rooted phylogenetic tree of the gut microbial community of *A. assamensis* with phylum level annotation as identified through 16s rRNA metagenomic sequencing.

Table 2. 7 Alpha diversity indices for the larval gut and host plant leaf samples.

Sample Name	Shannon index	Simpson's index	Pielou's evenness	Good's coverage
LC-G	8.44	0.99	0.93	0.99
LC-L	8.40	0.99	0.94	0.99
LP-G	7.78	0.99	0.89	0.99
LP-L	4.95	0.93	0.72	1
MB-G	8.52	0.99	0.94	0.99
MB-L	8.42	0.99	0.93	0.99

Table 2. 8(a) Pairwise Kruskal-Wallis test for Shannon's index between the larval gut and host plant leaf samples.

Group 1	Group 2	H-value	p-value	q-value
Gut (n=3)	Leaf (n=3)	1.19	0.27	0.27

Table 2. 8(b): Pairwise Kruskal-Wallis test for Faith's phylogenetic diversity vector between the larval gut and host plant leaf samples.

Group 1	Group 2	H-Value	p-value	q-value
Gut (n=3)	Leaf (n=3)	3.85	0.04	0.04

Principal Coordinate Analysis (PCoA) was conducted among the six samples (three guts and three host plant leaves samples) based on their Euclidean distances to visualize beta diversity as shown in **Figure 2. 7(c)**. The results show that MB-G, MB-L and LC-G are clustered together with distinct positions that can be isolated from each other. This reflects that although the microbial composition of MB-G, MB-L and LC-G are distinct from each other, the dissimilarity among them is less than in the other three samples. LP-G is more similar to LP-L and LC-L occupies almost an equal distance from LP-L and MB-L. Here, axis 1, axis 2 and axis 3 account for 40.17%, 25.67% and 15.09% variation, respectively.

2.3.8 Predictive functional profiling with PICRUSt2:

The functional role of the larval gut and host plant leaf-associated microbiome was profiled by gene family prediction of the identified taxa against their nearest sequenced genome available in reference databases using the tool Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2). The results of this analysis depict the abundances of KEGG orthologs, enzymes and metabolic pathways detected in each sample. A total of ≃1910 enzymes were predicted to be present in both the gut and leaf samples from which 420 MetaCyc pathways were inferred through mapping the EC gene families to pathways. The enzymes alcohol dehydrogenase, 3-oxoacyl-[acyl-carrier-protein] reductase, L-threonine 3-dehydrogenase, D-threo-aldose 1-dehydrogenase, UDP-glucose 6-dehydrogenase, Histidinol dehydrogenase, Shikimate dehydrogenase, Glyoxylate reductase, GDP-L-fucose

synthase, Homoserine dehydrogenase, 3-hydroxybutyrate dehydrogenase, 3-hydroxyisobutyrate dehydrogenase, Isocitrate dehydrogenase (NADP(+)) and Glycerol-3-phosphate dehydrogenase (NAD(P)(+)), were among the highly expressed enzymes in all the samples (**Figure 2. 9**).

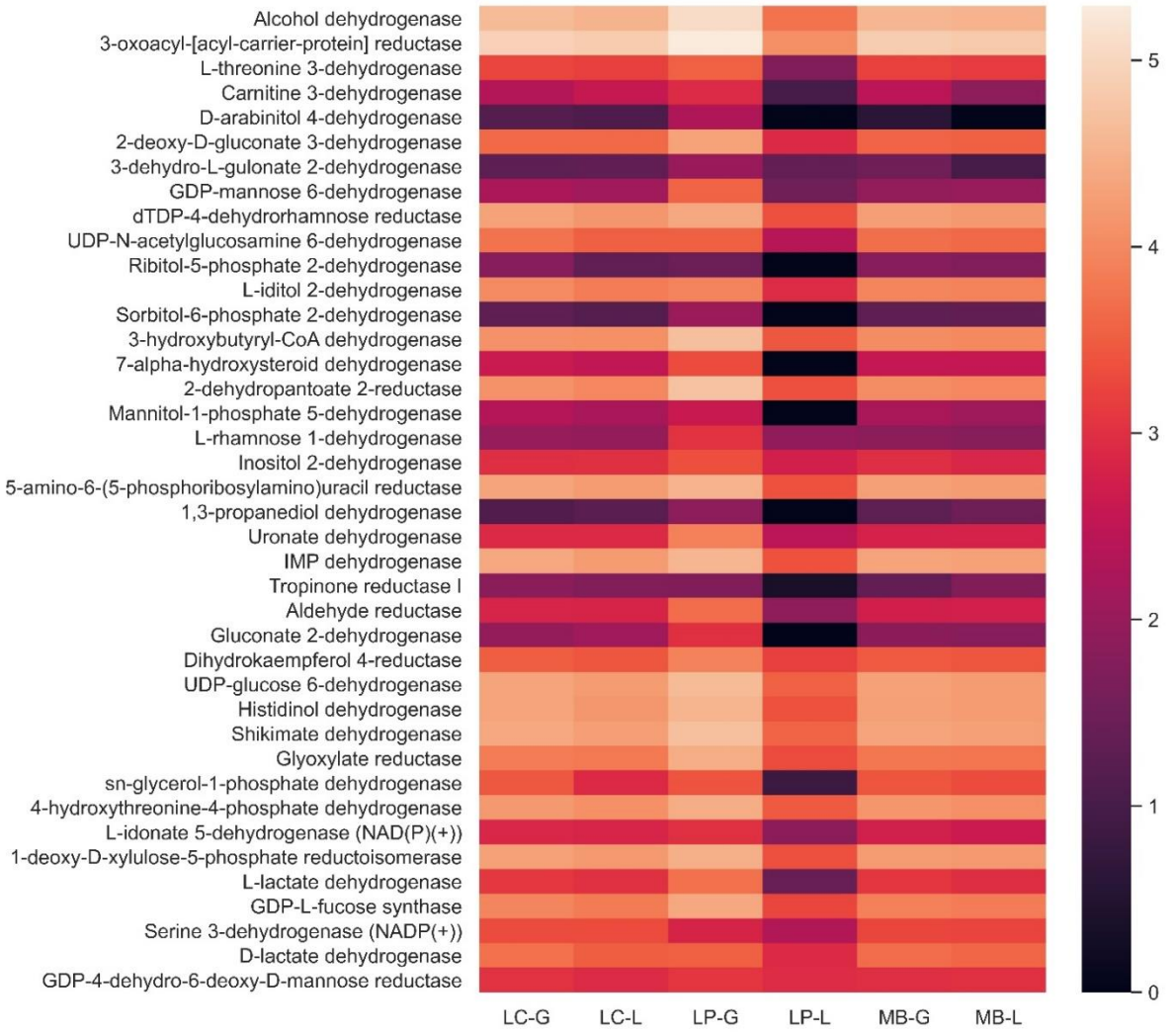


Figure 2. 9 Heat map showing the relative abundance of selected microbial enzymes in the predicted functional metagenome of larval gut and host plant leaf samples as identified through PICRUSt2 analysis.

Metabolic pathways such as glycolysis, homolactic fermentation, amino acid biosynthesis, Calvin-Benson-Bassham cycle, reductive TCA cycle, and gluconeogenesis I were abundantly present across the samples. However, pathways like starch degradation V, urea cycle, formaldehyde assimilation II (RuMP Cycle), CMP-3-deoxy-D-manno-octulosonate

biosynthesis I, Mycolate biosynthesis, and NAD salvage pathway were enriched in the gut samples than their respective host plant leaves. The most abundant MetaCyc pathways predicted to be present in the studied samples have been represented in **Figure 2. 10** as a barplot.

The differential abundance of MetaCyc pathways between the larval gut and its respective host leaf is represented in **Figure 2. 11**. Between MB-G and MB-L, superpathway for aerobic toluene degradation, L-1, 2-propanediol degradation, peptidoglycan biosynthesis V and II, Superpathway of taurine degradation, polymyxin resistance pathways were significantly more abundant in the gut (MB-G) than in the leaves (MB-L) (**Figure 2. 11(a)**). Between LP-G and LP-L, superpathway of C1 compound oxidation to CO2, mycothiol biosynthesis, superpathway of demethymenaquinol-6 biosynthesis, phylloquinol biosyntheisis, nitrate reduction I, ectoine biosynthesis, biotin biosynthesis, fucose degradation and rhamnose degradation were found to be significantly enriched in the gut (LP-G) (**Figure 2. 11(b)**). Similarly, in LC-L protocatechuate degradation I (meta-cleavage pathway), superpathway of taurine degradation, vanillin and vanillate degradation I, gallate degradation I, superpathway of ornithine degradation were found to be enriched in comparison to LC-G. On the other hand, in LC-G, meta cleavage pathway of aromatic compounds, superpathway of fucose and rhamnose degradation, aerobic toluene degradation IV, peptidoglycan biosynthesis, superpathway of methanogenesis, chitin derivatives degradation, superpathway of glycol metabolism and degradation were observed to be enrich as depicted in **Figure 2. 11(c)**.

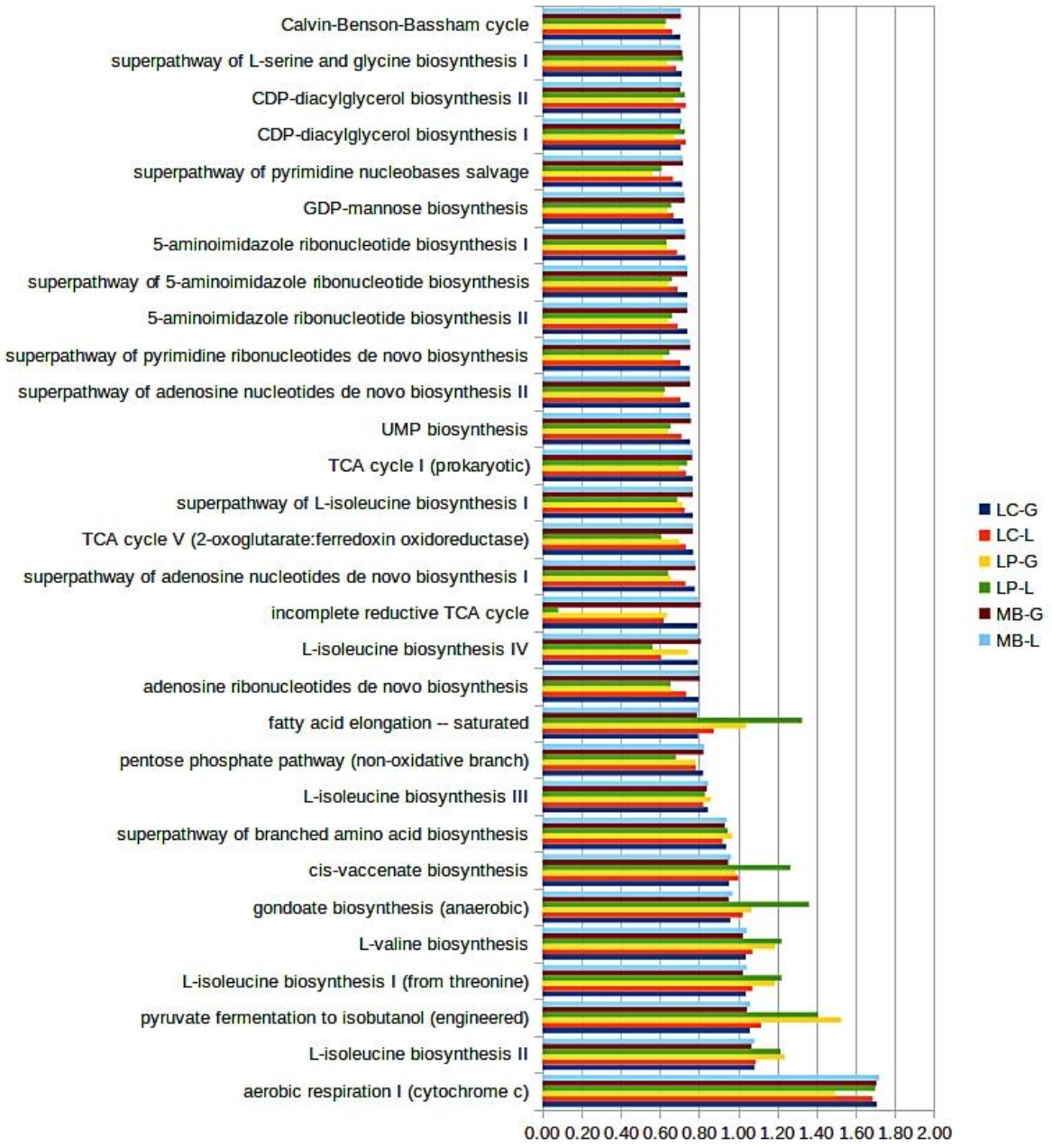


Figure 2. 10 Relative abundances of selected MetaCyc pathways identified in the predicted functional metagenome of the larval guts and host plant leaves through PICRUSt2 analysis.

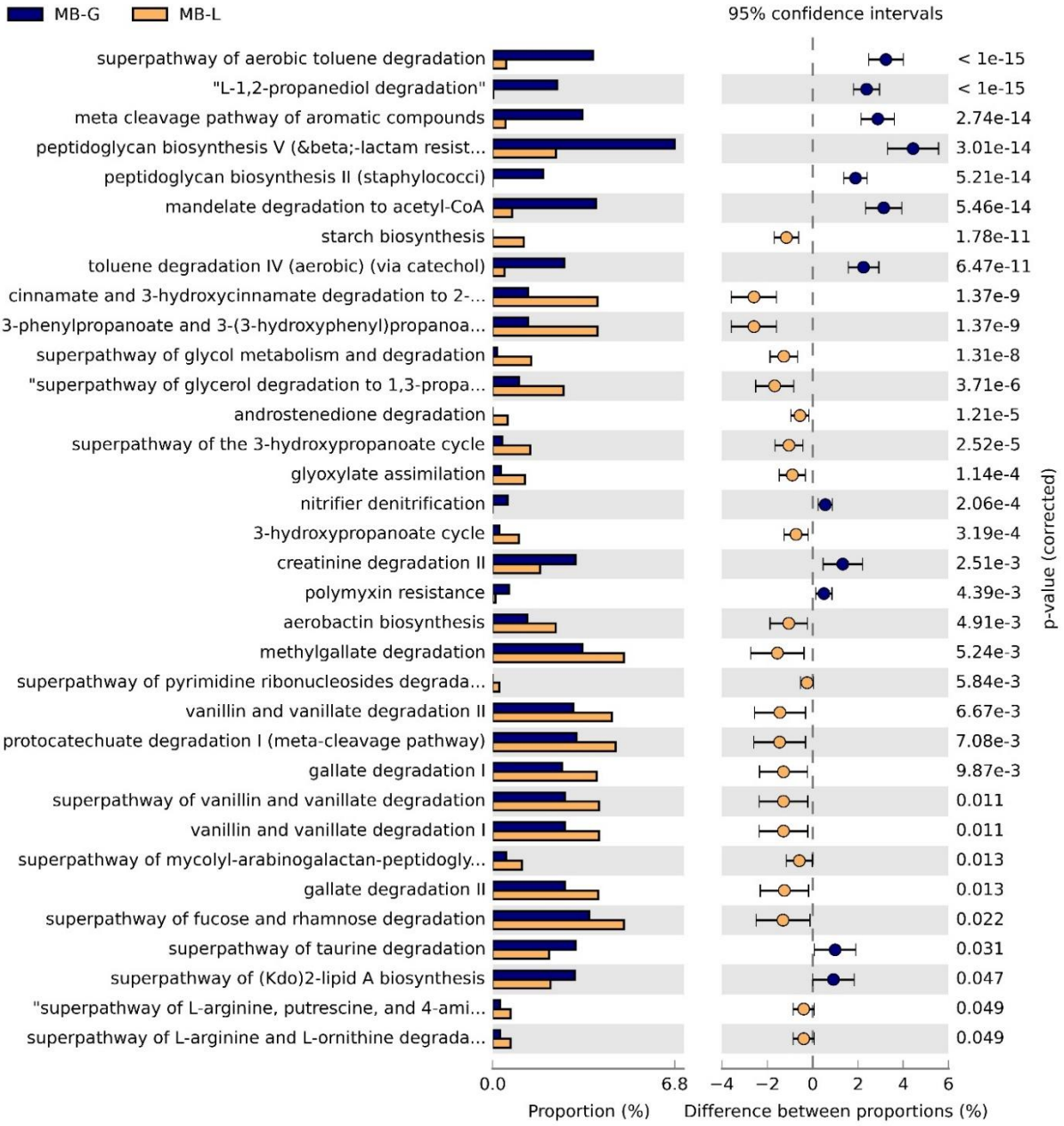


Figure 2. 11(a) Differential abundance of MetaCyc pathways between MB-G and MB-L. The first column represents the metabolic pathways, the second column represents the proportions of the pathways in each MB-G and MB-L, the third column represents the difference between proportions in percentage at 95% CI and fourth column represents the corrected *p*-value.

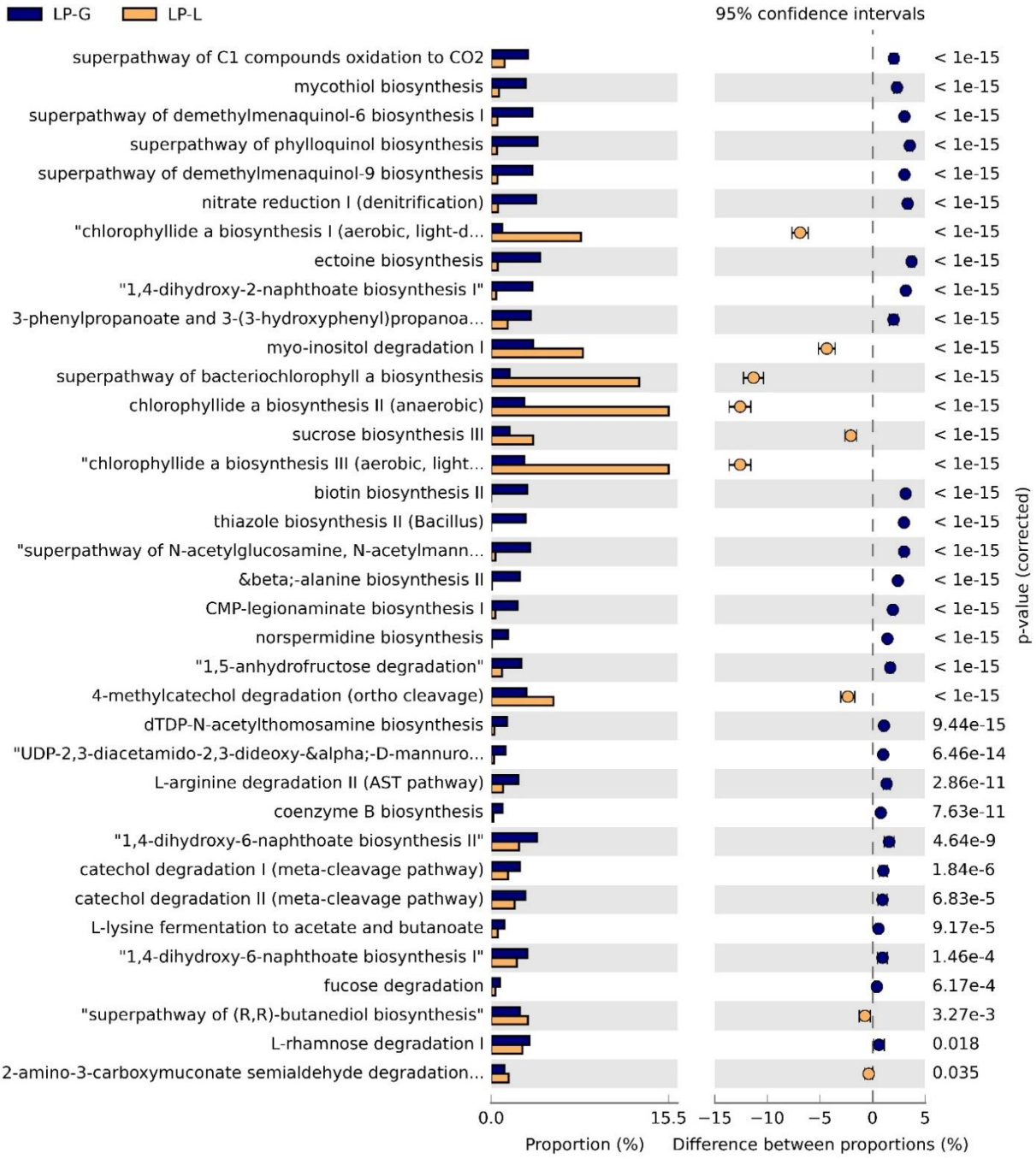


Figure 2. 11(b): Differential abundance of MetaCyc pathways between LP-G and LP-L. The first column represents the metabolic pathways, second column proportion in percentage of the pathways in each LP-G and LP-L, third column difference between proportions in percentage at 95% CI and fourth column represents the corrected *p*-value

.

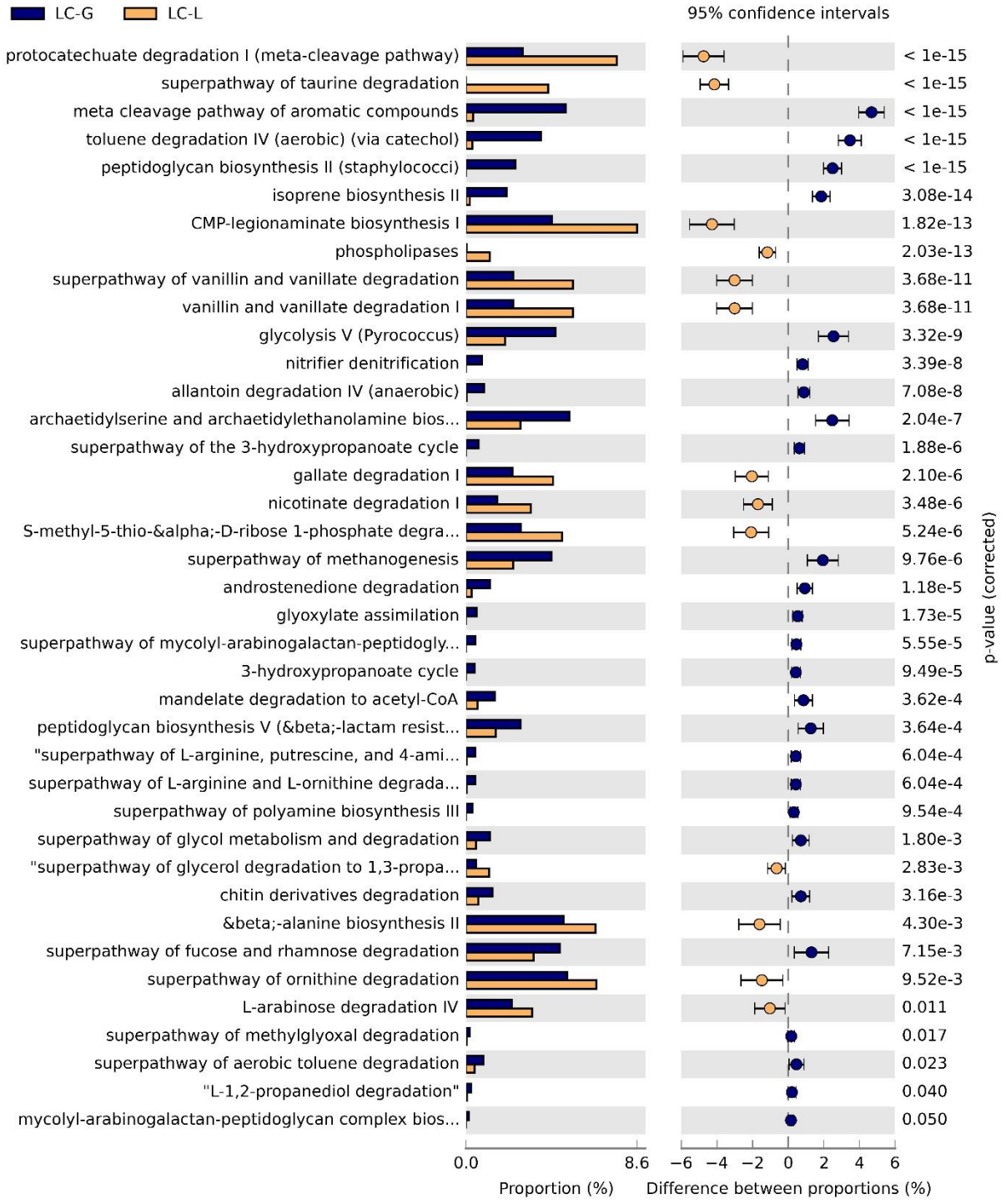


Figure 2. 11(c): Differential abundance of MetaCyc pathways between LC-G and LC-L. The first column represents the metabolic pathways, second column proportion in percentage of the pathways in each LC-G and LC-L, third column difference between proportions in percentage at 95% CI and fourth column represents the corrected *p*-value.

2.4 Discussion

2.4.1 The unique and diverse gut microbial community of *A. assamensis*

Amplicon analysis of the gut microbiome of *A. assamensis* larvae reared on *M. bombycina*, *L. polyantha* and *L. citrata* revealed a diverse microbial community dominated by Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria, Firmicutes, Planctomycetes, Patescibacteria, Fusobacteriota, Bacteroidota and Verrucomicrobiota. Bacterial phyla with no cultured representatives, commonly known as candidate phyla like WPS-2, OD1 and TM7 have also been observed amongst the larval gut samples. WPS-2 or Eremiobacterota is a poorly understood candidate phylum abundantly present in various terrestrial environments (Ji et al., 2021; Sheremet et al., 2020). The presence of this lesser-known bacterial phylum is reported in various mammals (Camanocha & Dewhirst, 2014). In insects, it is present in the gut of walking sticks (*Phasmotaenia lanyuhensis)* (Li et al., 2020). The candidate phylum OD1 or Parcubacteria has been detected through metagenomic sequencing from various environments which are exclusively anoxic. Interestingly, genome analysis of this phylum has revealed a substantial reduction in genome size containing signatures of symbiotic habit (Nelson & Stegen, 2015). OD1 is present in other insect species like xylophagous beetle larvae (Coleoptera) (Mohammed et al., 2018) and *Pectinophora gossypiella* (Lepidoptera) (Chaitra et al., 2022). In our study, the identification of anaerobic taxa like *Methanobacterium* and *Methanobrevibacter* can be correlated with earlier reports that stated the foregut and midgut environments of many herbivorous insects are nearly anoxic with oxygen concentrations varying from 0 to 7.3 mm Hg (Johnson & Barbehenn, 2000b). Low oxygen concentration in insect gut lumen mainly prevents the plant compounds such as phenols from oxidation thereby inhibiting the formation of ROS (Reactive Oxygen Species) and other oxidative compounds which may cause oxidative stress on the insect herbivore (Johnson & Barbehenn, 2000a). During the gut dissection of this study, a rapid colour change of the gut content from green to brown was observed as soon as it came in contact with air. This may indicate the presence of a

low oxygen concentration inside the gut lumen of *A. assamensis* larvae. However, oxygen concentration inside the gut lumen may vary from species to species. From our study, it is not clear at which portion of the gut the anaerobic bacteria reside as we considered the whole gut. Studies on microbial composition and measurement of oxygen concentration of the foregut, midgut and hindgut region would help in further understanding of the gut environment.

TM7 or Saccharibacteria is another poorly characterized bacterial phylum with a reduced genome (He et al., 2015). Its association with humans and other organisms is well-established (Bor et al., 2019; He et al., 2015). However, association of TM7 with insect gut is not well reported. We also observed other bacterial phyla like Myxococcota, Gemmatimonadota, Bdellovibrionota, Nitrospira and Campilobacterota in the gut of *A. assamensis* in minor abundance. From the present study, it is evident that the gut microbiota of *A. assamensis* exhibit a unique and diverse gut microbial composition with higher phylum level bacterial diversity compared to other lepidopteran species.

In our study, Archaea have been found to constitute approximately 10% of the total identified taxa in the gut of *A. assamensis* larvae reared on three different host plants. Studies on other insects like cockroaches (Hara et al., 2002), beetles (Ziganshina et al., 2018), termites (Yang et al., 2006) have revealed the association of Euryarchaeota and Crenarchaeota in the gut. These insects are also well known to harbor methanogens inside their hindgut that utilizes reduced one-carbon compound and hydrogen mainly produced during the gut fermentation of dietary fibers (Brune, 2019). In higher termites, Bathyarchaeia belonging to the phylum Crenarchaeota are involved in reductive acetogenesis from organic substrates (Loh et al., 2021). In the present study, methanogenic archaea like *Methanobacterium*, *Methanobrevibacter*, *Methanosphaera*, *Methanoculleus*, *Methanosaeta*, *Methanosarcina*, etc. are found to be present in the gut of the larvae.

Gut microbial analysis of *Spodoptera litura*, a lepidopteran pest, showed that the gut bacterial community composed mainly of Proteobacteria (~80%), Firmicutes (~4%) followed by

Bacteroidetes and Actinobacteria (Xia et. Al., 2020). Although, they mentioned about the presence archaea inside the gut (0.007% of the total gut microbe), no archaeal species was identified. Amplicon analysis of the foregut, midgut and hindgut microbiota of *Bombyx mori* identified Proteobacteria as the most dominant phyla followed by Actinobacteria, Firmicutes, Bacteroidetes, Acidobacteria, Chloroflexi, Armatimonadetes, Planctomycetes, Spirochaetae, Chlorobi (Chen et. al, 2018). However, no archaeal species were reported from *Bombyx mori* gut. Another lepidopteran species, *Brithys crini* was reported to contain Proteobacteria (52%), Bacteroidetes (36%) and Firmicutes (10%) (González-Serrano, 2020). Similar results were found from the gut microbial analysis of *Busseola fusca* (Snyman et al. 2016) and *Cnaphalocrocis medinali* (Li et al., 2022).

Comparison of the gut microbial community of *A. assamensis* with that of other Lepidopteran species based on available literature have showed that, Proteobacteria is present as the most dominant gut microbial phylum in most of the cases. Other phyla like Firmicutes, Actinobacteria, Acidobacteria, Fusobacteria and Bacteroidetes were also present in many lepidopteran species in varying abundance. In our study, many other microbial phyla from both the domain Bacteria and Archaea have been identified and reported (Figure 2.4(a) and 2.4(b)).

2.4.2 The leaf microbial community of the host plants of *A. assamensis*

The leaves of the genera *Machilus* and *Litsea* have not been explored yet for microbial diversity and their functional potential through culture-independent methods. Here, for the first time, we conducted a culture-independent investigation of the leaf-associated bacterial and archaeal community of *M. bombycina*, *L. polyantha* and *L. citrata* and their predictive function. The leaves of both *M. bombycina* and *L. citrata* harbored a rich bacterial and archaeal community. The lower diversity in *L. polyantha* leaves is possibly due to the antimicrobial activity of its phytochemical constituents (Ahmmad et al., 2012; Khanal & Bhatt, 2020) or insufficient representation of the microbial community due to degradation of the quality of the isolated DNA caused by the presence of diverse plant secondary metabolites like flavonoids, alkaloids,

phenols, terpenes and quinones (Kumar Shasany & Darokar, 1999).

2.4.3 Impact of host leaf microbiome on the structure of gut microbiome of *A assamensis*

In our investigation, comparative analyses of the gut and host leaf-associated microbiota showed that diet significantly affects the gut microbial community structure of *A. assamensis* (refer to section 2.3.6 and **Figure 2. 6(c)**). A considerable fraction of the gut microbial community of *A. assamensis* larvae contributed to the core gut microbial community (38%, 25% and 24% of the total identified OTUs in LP-G, MB-G and LC-G respectively; (**Figure 2. 6(a)**). However, many unique OTUs were present in the gut but not in the leaf or vice versa (**Figure 2. 6(c)**). Our results indicated that the host leaf-associated microbiota plays a key role in structuring the gut microbial composition of the larvae. This is also suggested by previous reports studying the effect of diet on the gut microbial composition of other lepidopteran insects (Hammer et al., 2017; Martínez-Solís et al., 2020; Mason Id et al., 2020).

2.4.4 Role of larval gut microbiota in leaf digestion

Starch and cell wall polysaccharides are the key constituents of green plants which can contribute to half of a leaf's dry weight (Kirsch et al., 2019). The cell wall polysaccharides can be digested easily by various hydrolytic enzymes present in the insect herbivore's gut which include amylases, glycosidases, proteases and lipases (Zibaee, 2012). Glycosidases or glycosyl hydrolases (EC 3.2.1) are enzymes available widely in all living organisms and catalyze the hydrolysis of glycosidic bonds present in complex sugars (Cerqueira et al., 2012). Insect α-amylase (EC 3.2.1.1) is a sub-subclass of glycosyl hydrolases that break down alpha-1,4 glycosidic bonds of a polysaccharide molecule (da Lage, 2018). On the other hand, the proteases (or peptidases, EC 3.4) break down the peptide bonds and lipases (EC 3.1.1.3) hydrolyze the ester bonds present in fatty acids (Zibaee, 2012). In our study, genes responsible for producing different types of glycosidases have been identified in the gut metagenome through predictive functional analysis. These enzymes include beta-glucosidase, beta-N-

acetylhexosaminidase, alpha-glucosidase, alpha-L-fucosidase, alpha-amylase, cellulase, alpha and beta-galactosidase, alpha-L-arabinofuranosidase, alpha-L-rhamnosidase, chitinase, endo-1,4-beta-xylanase, xylan 1,4-beta-xylosidase, beta-mannosidase, pectinase etc. These enzymes play pivotal roles in breaking down the complex biological compounds present in plant cells into simpler forms for further breakdown and assimilation by the insect herbivore.

2.4.5 Chitinase production by gut microbiome

Chitin is one of the most abundant biopolymers on earth consisting of a linear chain of β-(1,4)-N-acetyl-d-glucosamine (GlcNAc) (Khajuria et al., 2010). It is a major component of the exoskeleton of the phylum Arthropoda. In insects, it produces the cuticular lining of the foregut and hindgut lumen, the peritrophic membrane of the midgut and the trachea. For insects to grow and develop from one instar to the next, a portion of the chitin in the old cuticle must be degraded, followed by the synthesis of new chitin during molting (Merzendorfer & Zimoch, 2003). Insects regulate this process through repeated production and degradation of chitin through enzymes chitin synthase and chitinases, respectively. Microorganisms like Proteobacteria, Actinobacteria, Firmicutes are well known producers of chitinases or chitinolytic enzymes (Veliz et al., 2017). In the present study, genes producing chitinases (EC 3.2.1.14) are found to be present in the predicted gut metagenome of the larvae reared on all the three host plants in amplicon analysis. This has been further confirmed by the results of shotgun metagenome analysis in our next objective. However, no chitin synthase gene was found in either of the amplicon or shotgun analyses of our study.

2.4.6 Microbial detoxification of plant secondary metabolites in the larval gut

Secondary metabolites are allelochemicals produced by plants that serve a variety of important roles including defense against pathogenic microbes and phytophagous insects, stress response as well as modulation of organismal interaction (Pang et al., 2021). Plant secondary metabolites (PSMs) can be divided into different broad molecular categories including alkaloids, phenolics, terpenoids, flavonoids and steroids (Pang et al., 2021). Gut microbes of phytophagous

organisms are known to detoxify the plant secondary metabolites in the gut immediately after ingestion to facilitate their degradation (Kohl et al., 2018). Metagenomics analysis to study the microbial detoxification of small mammalian herbivore have revealed the presence of genes such as oxalyl-CoA decarboxylase associated with oxalate degradation and aryl-alcohol dehydrogenase that degrades aromatic hydrocarbons (Kohl et al., 2018). Both these genes have been found to be present in the predicted metagenomes of larval guts and host plant leaves. Phytochemical analysis of the leaf constituents of nine species of Lauraceae has revealed the presence of compounds like neochlorogenic acid, afzelin, laurolitsine, catechin, epicatechin, coclaurine, quercetin etc. (Oh et al., 2022) which mainly belong to alkaloids, sesquiterpenoids, norisoprenoids, phenolics, tocopherols, flavonoids and proanthocyanidins (Alejo-Armijo et al., 2017). Phenolic compounds can impair the digestive system of insect herbivore. When these chemicals are oxidised into quinones, they stimulate the generation of reactive oxygen species (ROS), especially at high pH levels like those seen in the lepidopteran gut (Voirol et al., 2018). Metagenomic sequencing of the gut microbial community of diamondback moth has revealed that ROS detoxifying enzymes such as catalases can be produced by gut bacteria, especially *Enterococcus* spp. (Xia et al., 2017). In our study, the catalase-producing genes detected in the predicted gut metagenome mainly belong to the bacteria like *C. kaiserbacteria*, *Pedomicrobium*, *Bradyrhizobium*, and *Phenylobacterium.*

2.4.7 Methane-producing microbe in the gut of *A. assamensis*

Methane is the second most important greenhouse gas and a potent cause of climate change and global warming (Mar et al., 2022). It is also a key component of natural gas which can be considered a promising replacement for fossil fuels (Enzmann et al., 2018). Methanogens are microbes that produce methane as their metabolic byproduct in an environment with low oxygen concentration by using CO_2, H_2, or small organic compounds such as formate, acetate and methylamine (Balch et al., 1979). All currently known methanogens are strictly confined to a phylogenetically diverse group of Euryarchaeota under the domain Archaea (Buan, 2018;

Thauer et al., 2008). The association of methanogens inside the digestive tract of herbivorous animals including ruminants is well established. These microbes help the host organism in the digestion of plant biomass which results in the production of methane (St-Pierre & Wright, 2013a). Among the arthropods, only a few organisms including termites, cockroaches, millipedes, and scarab beetle larvae are reported to contain methanogens (Horváthová et al., 2021). Amplicon analysis of the present study revealed the presence of the phylum Euryarchaeota in all the samples except LP-L. In the host leaf samples LC-L and MB-L, only *Methanobrevibacter* and *Methanosphaera* are detected whereas *Methanobacterium*, *Methanobrevibacter* and *Methanosphaera* are detected in the three gut samples. We observed that, *Methanobacterium* is not present in any of the host leaf-associated microbiota but detected in the gut of the larvae irrespective of their diet. Methanobacteriales are hydrogenotrophic archaea with varying lifestyles in contrasting environments abundantly present in a diverse range of anoxic habitats (Browne et al., 2017). *Methanococcus maripaludis* under the class Methanococci is a well-studied, strictly anaerobic, hydrogenotrophic archaeon mostly characterized by marine environments (Goyal et al., 2016). The association of this species with the human intestine has been reported recently (Coker et al., 2020). On the other hand, *Methanobacterium*, *Methanobrevibacter*, *Methanosphaera etc.* are well identified previously from diverse gut environments including insects (Brune, 2010; Gaci et al., 2014; Hackstein & van Alen, 1996; St-Pierre & Wright, 2013b).

We found that there exists an intricate relationship between healthy larvae of *A. assamensis* and methanogenic archaeal flora. Genes like Tetrahydromethanopterin S-methyltransferase, Nitrogenase, CoB--CoM heterodisulfide reductase, coenzyme F420 hydrogenase, and formate dehydrogenase are found to be associated with the methanogens present in the predicted metagenome of the larval gut. Among the genes predicted from the gut archaeome, Formylmethanofuran dehydrogenase and methylenetetrahydromethanopterin dehydrogenase were previously reported to be associated with hydrogenotrophic methanogenesis pathway (C.

J. Zhang et al., 2020). The identification of the methanogens and their probable functional potential in the gut of *A. assamensis* larvae should provide a key avenue towards future research in understanding the association of methanogenic archaeal flora in the gut of lepidopteran larvae and the contribution of Lepidoptera in global methane emission.

2.4.8 Fat body metabolism in *A. assamensis* and role of gut microbiota

Insect fat body is a loose mass of multifunctional tissue found beneath the integument surrounding the alimentary canal and the reproductive organ (Arrese & Soulages, 2010; Skowronek et al., 2021). One of the key functions of the fat body is to store energy reserves in the adipocytes in the form of triglycerides and glycogens and release it at the time of energy demand, especially during the non-feeding stage (Arrese & Soulages, 2010). Although insects vary in their amount of energy reserves in the fat body, lipid makes up more than 50% of the dry weight of the adipocytes (Arrese & Soulages, 2010; Gilby, 2003). On the other hand, glycogen content of the adipocytes is significantly lower than the lipids and it fluctuates depending on factors like feeding activity, life stage, environmental conditions etc. (Anand & Lorenz, 2008; Lorenz & Anand, 2004). Approximately, 90% of the total stored lipid of the adipocytes is composed of triglycerides produced from non-lipid precursors such as carbohydrates and proteins present in the diet through the process of lipogenesis analogous to that of mammalian system (Arrese & Soulages, 2010). Research on microbial association with the fat body in brown planthopper, *Nilaparvata lugens* have shown the presence of Firmicutes and Proteobacteria with 99% abundance followed by Actinobacteria, Bacteroidetes and Fusobacteria and other bacterial classes (J. H. Zhang et al., 2019).

In our study, acetyl-CoA carboxylase gene which is necessary for the process of lipogenesis has been identified from the gut microbial metagenome. This enzyme catalyzes the carboxylation of acetyl-CoA to malonyl-CoA which is considered to be the first committed step in fatty acid biosynthesis (Towle et al., 1997). Other two enzymes, glycerol-3-phosphate 1-O-acyltransferase and stearoyl-CoA 9-desaturase required for the formation and maturation

of triglycerides have also been identified in the predicted gut metagenomes of the larvae of *A. assamensis* reared on three different host plants. These enzymes have been previously reported to be induced by a high-carbohydrate diet in mammals(Towle et al., 1997). Other enzymes such as Glucose-6-phosphate dehydrogenase and phosphogluconate dehydrogenase which fulfill the electron requirement during fatty acid synthesis have also been detected in the predicted gut metagenome of *A. assamensis* larvae. Further, triglycerol lipase, the initial enzyme involved in lipid mobilization from the fat body as reported in *Manduca sexta* (Arrese & Wells, 1994) has been found to be consistently present in the predicted larval gut metagenomes. Other enzymes such as glycogen phophorylase required for glycogen mobilization (Arrese & Soulages, 2010) and alpha, alpha-trehalose-phosphate synthase (UDP-forming) and trehalose-phosphatase involved in trehalose production from glycogen (Skowronek et al., 2021), have also been detected to be abundantly contained by the larval gut microbiome of *A. assamensis*. The above findings of our result suggest that the gut microbiota of *A. assamensis* may play an intricate role in the regulation and metabolism of energy storage inside the fat body cells.

2.5 Conclusion

The structural and functional diversities of the gut microbial community of an organism are dependent on various factors such as habitat conditions, age/instar, host phylogeny, diet, physiological conditions, sex etc. In our study, we have explored the diversity, potential function, host plant-associated variation as well as the contribution of the host leaf microbiome to the composition of the gut microbial community of *A. assamensis* Helfer. This study reveals that *A. assamensis* Helfer contains a highly diverse gut microbial community in terms of bacteria, archaea and viruses. The host plant leaf-associated microbiota contributes largely to the gut microbial composition. We also observed intraspecific variation in gut microbial composition with respect to the change in the host plant. Interestingly, the larval gut microbiota was found to be richer in microbial diversity than their respective host leaf-associated microbes. This indicates that the gut microbiota of *A. assamensis* larvae is not solely diet-derived.

However, irrespective of the diet, the larvae reared on three different host plants carried a core gut microbiota consisting of taxa like *Sphingomonas*, *Paracoccus*, *Pseudorhodoplanes*, *Bradyrhizobium*, *Rhodomicrobium*, *Mesorhizobium*, *Pedomicrobium*, *Hyphomicrobium*, *Devosia*, *Reyranella*, *Dongia*, *Phenylobacterium*, *Brevundimonas*, *Staphylococcus*, *Bryobacter* etc. at the genus level. Functional investigation of the gut microbial community through predictive metagenomic analysis revealed various physiological and metabolic interactions between the host organism and the gut microbiota including leaf digestion, chitinase production, detoxification of plant secondary metabolites and fat body metabolism. Limitations associated with this study such as small sample size and low resolution of the 16S rRNA amplicon sequence data should be taken into consideration in future research. Comprehensive and elaborate multi-omic research considering spatiotemporal, sex, instar as well as disease-related variation would help in a better understanding of the holobiont. Also, the identification and functional characterization of the fungal and single-celled eukaryotes would add a new dimension to the knowledge about host-microbial interaction in the future. Comparison of the gut microbiota of this endemic silkworm species with other allied species from different geographical locations would also help in understanding the potential role of gut microbiota in endemism and species conservation.

References

Ahmmad, A., Torequl Islam, M., Sultana, I., Mahmood, A., Akther Hossain, J., Homa, Z., Ibrahim, M., & Mohi Uddin Chowdhury, M. (2012). Pharmacological and Phytochemical Screening of Ethanol Extract of Litsea monopetala (Roxb.) Pers. *IOSR Journal of Pharmacy*, *2*, 398–402.

Alejo-Armijo, A., Altarejos, J., & Salido, S. (2017). Phytochemicals and biological activities of laurel tree (Laurus nobilis). Natural product communications, 12(5), 1934578X1701200519.

Anand, A. N., & Lorenz, M. W. (2008). Age-dependent changes of fat body stores and the regulation of fat body lipid synthesis and mobilisation by adipokinetic hormone in the last larval instar of the cricket, Gryllus bimaculatus. *Journal of Insect Physiology*, *54*(10–11), 1404–1412. https://doi.org/10.1016/J.JINSPHYS.2008.08.001

Arrese, E. L., & Soulages, J. L. (2010). Insect Fat Body: Energy, Metabolism, and Regulation. *Annu. Rev. Entomol*, *55*, 207–225. https://doi.org/10.1146/annurev-ento-112408-085356

Arrese, E. L., & Wells, M. A. (1994). Purification and properties of a phosphorylatable triacylglycerol lipase from the fat body of an insect, Manduca sexta. *Journal of Lipid Research*, *35*(9), 1652–1660. https://doi.org/10.1016/S0022-2275(20)41163-0

Balch, W. E., Fox, G. E., Magrum, L. J., Woese, C. R., & Wolfe, R. S. (1979). Methanogens: reevaluation of a unique biological group. *Microbiological Reviews*, *43*(2), 260–296. https://doi.org/10.1128/MR.43.2.260-296.1979

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology*, *68*(2), 365–369. https://doi.org/10.1093/sysbio/syy054

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.

https://doi.org/10.1038/s41587-019-0209-9

Bor, B., Bedree, J. K., Shi, W., McLean, J. S., & He, X. (2019). Saccharibacteria (TM7) in the Human Oral Microbiome. *Journal of Dental Research*, *98*(5), 500–509. https://doi.org/10.1177/0022034519831671

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological monographs*, *27*(4), 326-349.

Browne, P., Tamaki, H., Kyrpides, N., Woyke, T., Goodwin, L., Imachi, H., Bräuer, S., Yavitt, J. B., Liu, W. T., Zinder, S., & Cadillo-Quiroz, H. (2017). Genomic composition and dynamics among Methanomicrobiales predict adaptation to contrasting environments. *ISME Journal*, *11*(1), 87–99. https://doi.org/10.1038/ismej.2016.104

Brune, A. (2010). Methanogenesis in the Digestive Tracts of Insects. *Handbook of Hydrocarbon and Lipid Microbiology*, 707–728. https://doi.org/10.1007/978-3-540-77587-4_56

Brune, A. (2019). Methanogenesis in the Digestive Tracts of Insects and Other Arthropods. *Biogenesis of Hydrocarbons*, 229–260. https://doi.org/10.1007/978-3-319-78108-2_13

Buan, N. R. (2018). Methanogens: pushing the boundaries of biology. *Emerging Topics in Life Sciences*, *2*(4), 629. https://doi.org/10.1042/ETLS20180031

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Camanocha, A., & Dewhirst, F. E. (2014). Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions. *Journal of Oral Microbiology*, *6*, 10.3402/jom.v6.25468. https://doi.org/10.3402/jom.v6.25468

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., & Karp, P. D. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*,

44(D1), D471–D480. https://doi.org/10.1093/nar/gkv1164

Cerqueira, N., Brás, N., Ramos, M. J., Fernandes, P. A., Cerqueira, N., Brás, N., Ramos, M. J., & Fernandes, P. A. (2012). Glycosidases – A Mechanistic Overview. *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology*. https://doi.org/10.5772/52019

Chaitra, H. S., Singh, A., Pandiyan, K., & Kalia, V. K. (2022). Sex Biased Variance in the Structural and Functional Diversity of the Midgut Bacterial Community of Last Instar Larvae of Pectinophora gossypiella (Lepidoptera: Gelechiidae). *Microbial Ecology*, *83*(4), 1112–1122. https://doi.org/10.1007/s00248-021-01829-1

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of statistics*, 265-270.

Chen, B., Du, K., Sun, C., Vimalanathan, A., Liang, X., Li, Y., ... & Shao, Y. (2018). Gut bacterial and fungal communities of the domesticated silkworm (Bombyx mori) and wild mulberry-feeding relatives. The ISME journal, 12(9), 2252-2262.

Coker, O. O., Ka, W., Wu, K., Wong, S. H., Sung, J. J. Y., & Yu, J. (2020). Altered Gut Archaea Composition and Interaction With Bacteria Are Associated With Colorectal Cancer. *Gastroenterology*, *159*, 1459-1470.e5. https://doi.org/10.1053/j.gastro.2020.06.042

Czech, L., & Stamatakis, A. (2019). Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples. *PLOS ONE*, *14*(5), e0217050. https://doi.org/10.1371/journal.pone.0217050

da Lage, J.-L. (2018). The Amylases of Insects. *International Journal of Insect Science*, *10*, 117954331880478. https://doi.org/10.1177/1179543318804783

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... & Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology, 72(7), 5069-5072.

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome

functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Eddy, S. R. (2020). *HMMER User's Guide Biological sequence analysis using profile hidden Markov models*. http://hmmer.org

Enzmann, F., Mayer, F., Rother, M., & Holtmann, D. (2018). Methanogens: biochemical background and biotechnological applications. *AMB Express*, *8*(1), 1–22. https://doi.org/10.1186/S13568-017-0531-X/FIGURES/6

Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M. A. Y., Bolte, L. A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., Dekens, J. A. M., Lenters, V. C., Björk, J. R., Swarte, J. C., Swertz, M. A., Jansen, B. H., Gelderloos-Arends, J., Jankipersadsing, S., Hofker, M., … Weersma, R. K. (2022). Environmental factors shaping the gut microbiome in a Dutch population. *Nature*, *604*(7907), 732–739. https://doi.org/10.1038/s41586-022-04567-7

Gaci, N., Tottey, W., Brugère, J.-F., Borrel, G., William, P. O., & Archaea, B. J. (2014). TOPIC HIGHLIGHT WJG 20 th Anniversary Special Issues (17): Intestinal microbiota Archaea and the human gut: New beginning of an old story. *World J Gastroenterol*, *20*(43), 16062–16078. https://doi.org/10.3748/wjg.v20.i43.16062

Galloway-Pena, J., Stevenson, B., Rodrigues, R. R., Wagner, B. D., Grunwald, G. K., Zerbe, G. O., Mikulich-Gilbertson, S. K., Robertson, C. E., Zemanick, E. T., & Harris, J. K. (2018). On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Frontiers in Microbiology | Www.Frontiersin.Org*, *1*, 1037. https://doi.org/10.3389/fmicb.2018.01037

Gandotra, S., Kumar, A., Naga, K., Bhuyan, P. M., Gogoi, D. K., Sharma, K., & Subramanian, S. (2018). Bacterial community structure and diversity in the gut of the muga silkworm, Antheraea assamensis (Lepidoptera: Saturniidae), from India. *Insect Molecular Biology*, *27*(5), 603–619. https://doi.org/10.1111/imb.12495

Gilby, A. R. (2003). Lipids and Their Metabolism in Insects. *Https://Doi.Org/10.1146/Annurev.En.10.010165.001041*, *10*(1), 141–160.

https://doi.org/10.1146/ANNUREV.EN.10.010165.001041

González-Serrano, F., Pérez-Cobas, A. E., Rosas, T., Baixeras, J., Latorre, A., & Moya, A. (2020). The gut microbiota composition of the moth Brithys crini reflects insect metamorphosis. Microbial ecology, 79, 960-970.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, *40*(3-4), 237-264.

Goyal, N., Zhou, Z., & Karimi, I. A. (2016). Metabolic processes of Methanococcus maripaludis and potential applications. *Microbial Cell Factories*, *15*(1), 1–19. https://doi.org/10.1186/S12934-016-0500-0/TABLES/3

Hackstein, J. H. P., & van Alen, T. A. (1996). Fecal methanogens and vertebrate evolution. *Evolution*, *50*(2), 559–572. https://doi.org/10.1111/j.1558-5646.1996.tb03868.x

Haloi, K., Kalita, M. K., Nath, R., & Devi, D. (2016). Characterization and pathogenicity assessment of gut-associated microbes of muga silkworm Antheraea assamensis Helfer (Lepidoptera: Saturniidae). *Journal of Invertebrate Pathology*, *138*, 73–85. https://doi.org/10.1016/j.jip.2016.06.006

Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P., & Fierer, N. (2017). Caterpillars lack a resident gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9641–9646. https://doi.org/10.1073/pnas.1707186114

Hara, K., Shinzato, N., Seo, M., Oshima, T., & Yamagishi, A. (2002). *Phylogenetic Analysis of Symbiotic Archaea Living in the Gut of Xylophagous Cockroaches* (Vol. 17, Issue 4). http://wwwsoc.nii.ac.jp/jsme2/

Hasan, N., & Yang, H. (2019). Factors affecting the composition of the gut microbiota, and its modulation. In *PeerJ* (Vol. 2019, Issue 8). PeerJ Inc. https://doi.org/10.7717/peerj.7502

Heikrujam, J., Kishor, R., & Behari Mazumder, P. (2020). The Chemistry Behind Plant DNA Isolation Protocols. In *Biochemical Analysis Tools - Methods for Bio-Molecules Studies*. IntechOpen. https://doi.org/10.5772/intechopen.92206

He, X., McLean, J. S., Edlund, A., Yooseph, S., Hall, A. P., Liu, S. Y., Dorrestein, P. C., Esquenazi, E., Hunter, R. C., Cheng, G., Nelson, K. E., Lux, R., & Shi, W. (2015). Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(1), 244–249. https://doi.org/10.1073/PNAS.1419038112/SUPPL_FILE/PNAS.1419038112.ST04.DOCX

Horváthová, T., Šustr, V., Chroňáková, A., Semanová, S., Lang, K., Dietrich, C., Hubáček, T., Ardestani, M. M., Lara, A. C., Brune, A., & Šimek, M. (2021). Methanogenesis in the Digestive Tracts of the Tropical Millipedes Archispirostreptus gigas (Diplopoda, Spirostreptidae) and Epibolus pulchripes (Diplopoda, Pachybolidae). *Applied and Environmental Microbiology*, *87*(15), 1–13. https://doi.org/10.1128/AEM.00614-21

Huang, K., Wang, J., Huang, J., Zhang, S., Vogler, A. P., Liu, Q., Li, Y., Yang, M., Li, Y., & Zhou, X. (2021). Host Phylogeny and Diet Shape Gut Microbial Communities Within Bamboo-Feeding Insects. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.633075

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M., & Reddy, N. (2015). role of the normal gut microbiota. *World J Gastroenterol*, *21*(29), 8787–8803. https://doi.org/10.3748/wjg.v21.i29.8787

Ji, M., Williams, T. J., Montgomery, K., Wong, H. L., Zaugg, J., Berengut, J. F., Bissett, A., Chuvochina, M., Hugenholtz, P., & Ferrari, B. C. (2021). Candidatus Eremiobacterota, a metabolically and phylogenetically diverse terrestrial phylum with acid-tolerant adaptations. *ISME Journal*, *15*(9), 2692–2707. https://doi.org/10.1038/s41396-021-00944-8

Jing, T. Z., Qi, F. H., & Wang, Z. Y. (2020). Most dominant roles of insect gut bacteria: Digestion, detoxification, or essential nutrient provision? *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00823-y

Johnson, K. S., & Barbehenn, R. v. (2000a). Oxygen levels in the gut lumens of herbivorous insects. In *Journal of Insect Physiology* (Vol. 46). www.elsevier.com/locate/jinsphys

Johnson, K. S., & Barbehenn, R. v. (2000b). Oxygen levels in the gut lumens of herbivorous insects. *Journal of Insect Physiology*, *46*(6), 897–903. https://doi.org/10.1016/S0022-1910(99)00196-1

Kameoka, S., Motooka, D., Watanabe, S., Kubo, R., Jung, N., Midorikawa, Y., Shinozaki, N. O., Sawai, Y., Takeda, A. K., & Nakamura, S. (2021). Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1–V2 and V3–V4 primer sets. *BMC Genomics*, *22*(1), 1–10. https://doi.org/10.1186/S12864-021-07746-4/FIGURES/3

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/NAR/GKF436

Khajuria, C., Buschman, L. L., Chen, M. S., Muthukrishnan, S., & Zhu, K. Y. (2010). A gut-specific chitinase gene essential for regulation of chitin content of peritrophic matrix and growth of Ostrinia nubilalis larvae. *Insect Biochemistry and Molecular Biology*, *40*(8), 621–629. https://doi.org/10.1016/j.ibmb.2010.06.003

Khanal, S., & Bhatt, B. D. (2020). Study on Biological Activity of Litsea monopetala From Panchthar District of Nepal. *Journal of Institute of Science and Technology*, *25*(2), 113–118. https://doi.org/10.3126/jist.v25i2.33747

Kirsch, R., Kunert, G., Vogel, H., & Pauchet, Y. (2019). Pectin digestion in herbivorous beetles: Impact of pseudoenzymes exceeds that of their active counterparts. *Frontiers in Physiology*, *10*(MAY). https://doi.org/10.3389/fphys.2019.00685

Kohl, K. D., Oakeson, K. F., Orr, T. J., Miller, A. W., Forbey, J. S., Phillips, C. D., Dale, C., Weiss, R. B., & Dearing, M. D. (2018). Metagenomic sequencing provides insights into microbial detoxification in the guts of small mammalian herbivores (Neotoma spp.). *FEMS Microbiology Ecology*, *94*(12). https://doi.org/10.1093/femsec/fiy184

Kumar Shasany, A., & Darokar, M. (1999). *Rapid Isolation of DNA from Dry and Fresh Samples of Plants Producing Large Amounts of Secondary Metabolites and Essential Oils*.

https://doi.org/10.1023/A:1007528101452

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), 127–128. https://doi.org/10.1093/bioinformatics/btl529

Li, Y.-H., Huang, Y.-F., Chen, T.-H., Wu, S.-S., Tang, H.-C., Hsiao, C.-Y., Huang, L.-C., Chang, J.-C., Chiu, K.-P., & Nai, Y.-S. (2020). Comparison of gut microbiota of healthy and diseased walking sticks, Phasmotaenia lanyuhensis. *Archives of Insect Biochemistry and Physiology*, *105*(4), e21749. https://doi.org/10.1002/arch.21749

Li, C., Han, G., Sun, J., Huang, L., Lu, Y., Xia, Y., ... & Xu, J. (2022). The Gut Microbiota Composition of Cnaphalocrocis medinalis and Their Predicted Contribution to Larval Nutrition. Frontiers in Microbiology, 13, 909863.

Loh, H. Q., Hervé, V., & Brune, A. (2021). Metabolic Potential for Reductive Acetogenesis and a Novel Energy-Converting [NiFe] Hydrogenase in Bathyarchaeia From Termite Guts – A Genome-Centric Analysis. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.635786

Lorenz, M. W., & Anand, A. N. (2004). Changes in the biochemical composition of fat body stores during adult development of female crickets, Gryllus bimaculatus. *Archives of Insect Biochemistry and Physiology*, *56*(3), 110–119. https://doi.org/10.1002/ARCH.20002

Louca, S., & Doebeli, M. (n.d.). *Efficient comparative phylogenetics on large trees*. https://doi.org/10.1093/bioinformatics/btx701

Lozupone, C., & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, *71*(12), 8228-8235.

Mar, K. A., Unger, C., Walderdorff, L., & Butler, T. (2022). Beyond CO2 equivalence: The impacts of methane on climate, ecosystems, and health. *Environmental Science and Policy*, *134*, 127–136. https://doi.org/10.1016/J.ENVSCI.2022.03.027

Martínez-Solís, M., Collado, M. C., & Herrero, S. (2020). Influence of Diet, Sex, and Viral

Infections on the Gut Microbiota Composition of Spodoptera exigua Caterpillars. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00753

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://journal.embnet.org/index.php/embnetjournal/article/view/200/479

Martin, M. M. (n.d.). *The evolution of cellulose digestion in insects*. https://royalsocietypublishing.org/

Mason Id, C. J., Clair, A. S., Peiffer, M., Gomez, E., Jones, A. G., Felton, G. W., & Hoover, K. (2020). *Diet influences proliferation and stability of gut bacterial populations in herbivorous lepidopteran larvae*. https://doi.org/10.1371/journal.pone.0229848

Merzendorfer, H., & Zimoch, L. (2003). Chitin metabolism in insects: Structure, function and regulation of chitin synthases and chitinases. In *Journal of Experimental Biology* (Vol. 206, Issue 24, pp. 4393–4412). https://doi.org/10.1242/jeb.00709

Mohammed, W. S., Ziganshina, E. E., Shagimardanova, E. I., Gogoleva, N. E., & Ziganshin, A. M. (2018). Comparison of intestinal bacterial and fungal communities across various xylophagous beetle larvae (Coleoptera: Cerambycidae). *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-27342-z

Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., González, A., Fontana, L., Henrissat, B., Knight, R., & Gordon, J. I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, *332*(6032), 970–974. https://doi.org/10.1126/SCIENCE.1198719

Nelson, W., & Stegen, J. (2015). The reduced genomes of Parcubacteria (OD1) contain signatures of a symbiotic lifestyle. *Frontiers in Microbiology*, *6*. https://www.frontiersin.org/article/10.3389/fmicb.2015.00713

Oh, M., Park, H. S., Um, S., Yang, T. J., & Kim, S. H. (2022). A comparative phytochemical study of nine Lauraceae species by using chemometric data analysis. *PLoS ONE*, *17*(9 September).

https://doi.org/10.1371/journal.pone.0273616

Pang, Z., Chen, J., Wang, T., Gao, C., Li, Z., Guo, L., Xu, J., & Cheng, Y. (2021). Linking Plant Secondary Metabolites and Plant Microbiomes: A Review. In *Frontiers in Plant Science* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fpls.2021.621276

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, *13*, 131-144.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, *26*(7), 1641–1650. https://doi.org/10.1093/MOLBEV/MSP077

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research, 41(D1), D590-D596.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. In *European Journal of Nutrition* (Vol. 57, Issue 1). Dr. Dietrich Steinkopff Verlag GmbH and Co. KG. https://doi.org/10.1007/s00394-017-1445-8

Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A., & Burge, C. B. (2008). Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science*, *320*(5883), 1643–1647. https://doi.org/10.1126/science.1155390

Shannon, C. E. (n.d.). A Mathematical Theory of Communication. In *The Bell System Technical Journal* (Vol. 27).

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shao, Y., Arias-Cordero, E. M., & Boland, W. (2013). Identification of metabolically active bacteria in the gut of the generalist spodoptera littoralis via DNA stable isotope probing using 13c-glucose. *Journal of Visualized Experiments*, *81*. https://doi.org/10.3791/50734

Shelomi, M., Lo, W. S., Kimsey, L. S., & Kuo, C. H. (2013). Analysis of the gut microbiota of walking sticks (Phasmatodea). *BMC Research Notes*, *6*(1). https://doi.org/10.1186/1756-0500-6-368

Sheremet, A., Jones, G. M., Jarett, J., Bowers, R. M., Bedard, I., Culham, C., Eloe-Fadrosh, E. A., Ivanova, N., Malmstrom, R. R., Grasby, S. E., Woyke, T., & Dunfield, P. F. (2020). Ecological and genomic analyses of candidate phylum WPS-2 bacteria in an unvegetated soil. *Environmental Microbiology*, *22*(8), 3143–3157. https://doi.org/10.1111/1462-2920.15054

Simpson, E.H. (1949). "Measurement of diversity". Nature. (163): 688.

Skowronek, P., Wójcik, Ł., & Strachecka, A. (2021). Fat Body—Multifunctional Insect Tissue. *Insects 2021, Vol. 12, Page 547*, *12*(6), 547. https://doi.org/10.3390/INSECTS12060547

Snyman, M., Gupta, A. K., Bezuidenhout, C. C., Claassens, S., & Van den Berg, J. (2016). Gut microbiota of Busseola fusca (Lepidoptera: Noctuidae). World Journal of Microbiology and Biotechnology, 32, 1-9.

St-Pierre, B., & Wright, A. D. G. (2013a). Diversity of gut methanogens in herbivorous animals. *Animal*, *7*(SUPPL.1), 49–56. https://doi.org/10.1017/S1751731112000912

St-Pierre, B., & Wright, A. D. G. (2013b). Diversity of gut methanogens in herbivorous animals. *Animal*, *7*(SUPPL.1), 49–56. https://doi.org/10.1017/S1751731112000912

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., & Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews Microbiology 2008 6:8*, *6*(8), 579–591. https://doi.org/10.1038/nrmicro1931

Towle, H. C., Kaytor, E. N., & Shih, H.-M. (1997). REGULATION OF THE EXPRESSION OF LIPOGENIC ENZYME GENES BY CARBOHYDRATE. *Annu. Rev. Nutr*, *17*, 405–438. www.annualreviews.org

Veliz, E. A., Martínez-Hidalgo, P., & Hirsch, A. M. (2017). Chitinase-producing bacteria and their role in biocontrol. In *AIMS Microbiology* (Vol. 3, Issue 3, pp. 689–705). AIMS Press. https://doi.org/10.3934/microbiol.2017.3.689

Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M., & Fatouros, N. E. (2018). Bacterial symbionts in lepidoptera: Their diversity, transmission, and impact on the host. In *Frontiers in Microbiology* (Vol. 9, Issue MAR). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2018.00556

Wang, J., Chen, H., & Tang, M. (2017). Community structure of gut bacteria of Dendroctonus armandi (Coleoptera: Curculionidae: Scolytinae) larvae during overwintering stage. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-14724-y

Wen, L., & Duffy, A. (2017). Factors influencing the gut microbiota, inflammation, and type 2 diabetes. *Journal of Nutrition*, *147*(7), 1468S-1475S. https://doi.org/10.3945/jn.116.240754

Willis, A. D. (2019). Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*, *10*(OCT). https://doi.org/10.3389/fmicb.2019.02407

Xia, X., Gurr, G. M., Vasseur, L., Zheng, D., Zhong, H., Qin, B., Lin, J., Wang, Y., Song, F., Li, Y., Lin, H., & You, M. (2017). Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00663

Xia, X., Lan, B., Tao, X., Lin, J., & You, M. (2020). Characterization of Spodoptera litura gut bacteria and their role in feeding and growth of the host. Frontiers in microbiology, 11, 1492.

Yang, H., Peng, J.-X., Liu, K.-Y., & Hong, H.-Z. (2006). [Diversity and function of symbiotic microbes in the gut of lower termites]. *Wei sheng wu xue bao = Acta microbiologica Sinica*, *46*(3), 496–499.

Ye, Y., & Doak, T. G. (2009). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLOS Computational Biology*, *5*(8), e1000465. https://doi.org/10.1371/journal.pcbi.1000465

Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., & Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033865

Zhang, C. J., Zhang, C. J., Pan, J., Liu, Y., Duan, C. H., Duan, C. H., & Li, M. (2020). Genomic and transcriptomic insights into methanogenesis potential of novel methanogens from mangrove sediments. *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00876-z

Zhang, J. H., Yu, N., Xu, X. X., & Liu, Z. W. (2019). Community structure, dispersal ability and functional profiling of microbiome existing in fat body and ovary of the brown planthopper, Nilaparvata lugens. *Insect Science*, *26*(4), 683–694. https://doi.org/10.1111/1744-7917.12575

Zibaee, A. (2012). Digestive enzymes of large cabbage white butterfly, Pieris brassicae L. (Lepidoptera: Pieridae) from developmental and site of activity perspectives. *Italian Journal of Zoology*, *79*(1), 13–26. https://doi.org/10.1080/11250003.2011.607190

Ziganshina, E. E., Mohammed, W. S., Shagimardanova, E. I., Vankov, P. Y., Gogoleva, N. E., & Ziganshin, A. M. (2018). Fungal, Bacterial, and Archaeal Diversity in the Digestive Tract of Several Beetle Larvae (Coleoptera). *BioMed Research International*, *2018*, e6765438. https://doi.org/10.1155/2018/6765438

Chapter 3

Functional analysis of *Antheraea assamensis* Helfer gut metagenome

CHAPTER 3

Functional analysis of *Antheraea assamensis* Helfer gut metagenome

Executive summary

Over the past decade, it has become evident that gut microbiota is a crucial factor in the health and disease of most organisms. Amplicon analysis of the gut microbial community of *A. assamensis* in our previous objective has revealed more than 250 microbial taxa and their predictive functional role. The present work aims to identify and characterize the functional potential of the gut microbiome of *A. assamensis* through shotgun metagenomic sequencing. For this, we have considered the larvae of *A. assamensis* reared on *Machilus bombycina*, the most commonly used host plant for commercial Muga silk production. The findings of this study will illustrate the functional role played by the gut microbial community of *A. assamensis* in host biology. The assembled shotgun sequences were annotated into 72063 coding sequences (CDSs). From the CDSs, 270 enzymes were identified with their EC numbers. The coding sequences were annotated into 1003 orthologous genes belonging to 24 COG categories. Taxonomic analysis of the assembled shotgun sequences resulted in 97% Bacteria, 2.4% Archaea and 0.06 % Viruses. The findings of this investigation will enable us to define the gut microbial population with more precision and comprehend its possible functions and interactions with the host organism.

3.1 Introduction

Over the past decade, it has become evident that the gut microbiota is a crucial factor in the health and disease of most organisms. The area of metagenomics has travelled a long way with the rapid advancement of high throughput sequencing technologies resulting in the ability to identify and quantify all microbial species present inside a host species (Lee & Hase, 2014). Marker gene sequencing approaches have given incredible insights into the host-microbiome relationships. However, this approach targets only the marker genes and mainly focuses on the diversity and abundance of the microbial community in a particular environment (Galloway-Peña & Hanson, 2020). On the other hand, shotgun metagenomic approaches can uncover the functional potential of a microbiome by capturing every microbial genome present in a sample through untargeted sequencing techniques (Jovel et al., 2016). It can reveal the whole genetic repertoire of bacteria, archaea, fungi and viruses from the microbiome, their diversity and abundances in a sample (Qin et al., 2010; L. C. Xia et al., 2011). Following the initial study design, a typical shotgun metagenomics study entails five steps: (a) sample collection, processing, and sequencing; (b) pre-processing of sequence reads; (c) sequence analysis to find out functional, and genomic features of the microbiome as well as their diversity and abundance (d) post-processing statistical and biological analysis; and (e) data validation. Each stage may be carried out in a variety of experimental and computational ways, leaving researchers with a bewildering array of options. Shotgun metagenomics has limits despite appearing to be simple because of potential experimental biases and the complexity of computational analyses and their interpretations (Quince et al., 2017). Functional metagenomics has developed over the past two decades as a powerful tool to investigate the functional diversity and metabolic potential of microbial communities (Chistoserdova, 2010; Lam et al., 2015). Functional metagenomics allows researchers to uncover the metabolic capabilities of microorganisms within a community, including those that may not be easily cultured in a laboratory (Sharpton, 2014). It is particularly valuable for identifying novel enzymes, pathways and microbial species

that may have potential applications (Hirakata et al., 2021; Schuele et al., 2022; Srinivas et al., 2022; Verma et al., 2018; Vijayvargiya et al., 2019). Additionally, functional metagenomics can be used to study the impact of environmental changes on microbial communities, such as pollution, climate change, or host-microbe interactions (Boolchandani et al., 2017a, 2017b; Mendes et al., n.d.; Neelakanta & Sultana, 2013; Visconti et al., 2019).

Insect gut metagenome functional analysis is the study of the collective genetic material of microorganisms present in the digestive system of insects, with a focus on their functional capabilities. The gut microbiome of insects plays an essential role in nutrient acquisition, digestion, and immune function, making the study of insect gut microbiomes crucial for understanding insect biology and ecology (Jing et al., 2020). The microbial communities within the insect gut provide the host with essential nutrients, aid in digestion, and help protect against pathogen invasion(Jing et al., 2020). Understanding the functional capabilities of these microbial communities is therefore crucial for improving our understanding of insect biology, ecology, and evolution (Krishnan et al., 2014; Wang et al., 2021; Xia et al., 2017a).

Amplicon analysis of the gut microbial community of *A. assamensis* in our previous objective has revealed more than 250 microbial taxa and their predictive functional role. The present objective aims to identify and characterize the functional potential of the gut microbiome of *A. assamensis* through shotgun metagenomic sequencing. For this, we have considered the larvae of *A. assamensis* reared on *Machilus bombycina*, the most commonly used host plant for commercial Muga silk production. The findings of this study will illustrate the functional role played by the gut microbial community of *A. assamensis* in host biology. The assembled shotgun sequences were annotated into 72063 coding sequences (CDSs). From the CDSs, 270 enzymes were identified with their EC numbers. The coding sequences were annotated into 1003 orthologous genes belonging to 24 COG categories. Taxonomic analysis of the assembled shotgun sequences resulted in 97% Bacteria, 2.4% Archaea and 0.06 % Viruses.

3.2 Materials and methods

3.2.1 Sample collection

For shotgun sequencing, fifth instar, healthy larvae reared on the primary host plant *M. bombycina* were considered as it is the most widely used host plant for silkworm cultivation and collected from CMERTI, Jorhat, Assam, India.

3.2.2 Sample preparation

With some modifications, insect dissection was carried out as previously reported (Shao et al., 2013). In short, the insect was put dorsoventrally on a dissection tray after being surface disinfected with 75% ethanol. The complete gut was securely removed during the dorsal side dissection, which was done from behind the head to the anus. Guts from five to six different larvae were combined to prepare the sample. After being washed with water and 1X PBS buffer, they were placed in cryovials and kept at -80° C until metagenomic DNA was extracted.

3.2.3 Metagenomic DNA isolation

Metagenomic DNA was obtained using a modified version of the PCI approach (38,39). In essence, the finely ground gut tissue was placed in a 2 ml centrifuge tubes, 500 ml of lysis buffer (1M Tris pH 8.0, 5M NaCl, 0.5M EDTA pH 8.0, 10% SDS, dH2O), and Proteinase K (20 L of 20 mg/ml stock solution per 1000 l of the lysis buffer) was added. The tube was incubated in a water bath at 65 °C for an hour with constant mixing by inversion at intervals. Each tube was then vortexed well and centrifuged for 15 minutes at 13000 rpm and 22 °C. The supernatants were moved to a fresh 1.5 ml centrifuge tube, and an equal volume of phenol, chloroform, and isoamyl alcohol was added. A vortex was used to thoroughly mix the mixture. The mixture was centrifuged once again for 15 minutes at the same speed and temperature as in the above step. This process was carried out twice. The sample was then treated with a double volume of 95–100% ice-cold ethanol and a 1/4th volume of 3M sodium acetate (pH 5.2) before being incubated at 4° C for an overnight period. The mixture was

centrifuged at 12000 rpm for 15 minutes at 4 °C the next day. The DNA pellet was repeatedly cleaned with 70% ethanol followed by 90% ethanol. It was then air-dried until the ethanol evaporated, dissolved in TE buffer, and kept at 4 °C. The amount and quality of the isolated DNA were evaluated using a NanoDrop™ UV visible spectrophotometer.

3.2.4 Shotgun Sequencing

The shotgun metagenomic sequencing was performed at Eurofins Genomic India Private Limited. In brief, RNA-free NGS grade DNA isolated from the gut of *A. assamensis* was checked for quality using agarose gel electrophoresis and NanoDrop before sequencing. Quantification of the DNA was carried out using Qubit 3.0 fluorometer. The indexed metagenome library was prepared and validated using Agilent High Sensitivity D1000 ScreenTape System. The prepared library was sequenced using 2×150 base pair paired-end sequencing through the Illumina NextSeq500 platform.

3.2.4.1 Shotgun sequence processing

3.2.4.1.1 Quality control of the raw sequence data and de novo assembly of the metagenome

Raw data were first checked for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Andrews, 2001). Low-quality base pairs and sequence ends were trimmed off using Trimmomatic V0.39 (Bolger et al., 2014) in paired-end mode with the sliding windows of 4 bases to remove bases with an average Phred quality score below 25. Cutadapt V 3.4 was used to remove the adapter sequences (Martin, 2011). The presence of the rRNA sequences was removed by mapping the resulting reads to the Silva rRNA database using bowtie2 (Quast et al., 2013). Similarly, the presence of host sequences was also detected and removed using bowtie2 (Langmead & Salzberg, 2012) and Samtools V1.9 (H. Li et al., 2009). The mapped reads were discarded and unmapped reads were retained for downstream processes and re-examined using FastQC for the presence of essential quality features.

The quality-filtered paired-end reads were used for metagenome assembly. Metagenome assembly was done using megahit v1.2.9 (Li et al., 2015), a *de novo* tool for NGS sequence assembly that utilizes the succinct *de Bruijn* graphs (Bowe et al., 2012) to assemble the sequences. Multiple k-mer lengths were used to assemble the sequences with minimum and maximum *k*-mer lengths of 21 and 141, respectively.

3.2.4.1.2 Metagenome annotation

Genome annotation is the procedure of finding and labeling every relevant feature on a genome sequence to derive structural and functional information about the proteins or genes it carries (Abril & Castellano, 2019). Most of the available genome annotation pipelines use homology search methods to obtain information from the available genome of the nearest ancestor in the reference database (Richardson & Watson, 2013). Functional annotation of the assembled contig was done with Prokka v1.12 (Seemann, 2014a) and Prodigal v2.6.3 (Hyatt et al., 2010). Prokka requires assembled DNA sequences in FASTA format as input produced by *de novo* assembly tools. It depends on external tools such as Prodigal, RNAmmer, Aragorn, SignalP and Infernal for feature prediction within the contigs (Seemann, 2014b). Annotation of the protein-coding genes (PCGs) is carried out in two different steps. First, the coordinates of the candidate genes are identified using Prodigal without describing the putative gene products. In the second step, Prokka predicts the gene products by comparing it with databases of different sizes in a hierarchical manner at the level of protein sequences to find out the most significant match (Seemann, 2014b). WebMGA, a web server for metagenomic analysis (Wu et al., 2011) was utilized for COG annotation of the identified protein sequences. Based on the gene annotation results, metabolic pathways were constructed using MinPath (Ye & Doak, 2009) against MetaCyc and KEGG (Kyoto Encyclopedia of Genes and Genomes) databases. MinPath is a tool for constructing pathways that follows a parsimony approach to determine the minimal set of biological pathways that must exist in a biological system to explain the existence of input protein sequences. Visualization of the metabolic network map based on KEGG pathways was

done with Interactive Pathways Explorer 3 (iPATH3) (Darzi et al., 2018).

3.2.4.1.3 Taxonomic analysis

One of the key concerns of metagenomic analysis is to deduce the microbial composition of a community and its relative abundances (Menzel et al., 2016). BLAST is the most widely used method used for taxonomic assignment to unknown sequences which classifies the sequences by searching against single or multiple reference databases (Altschul et al., 1990). To outperform BLAST's accuracy, several approaches to sequence categorization have been put forth that make use of sequence alignment and machine learning algorithms. The MEGAN software uses BLAST to search each sequence against multiple databases and the sequences are assigned with the lowest common ancestor of the best matches (Huson et al., 2007). The well-known Naive Bayes Classifier (NBC) employs a Bayesian rule to the ranges of *k*-mers within a genome (Rosen et al., 2008). However, many of these tools are computationally expensive requiring substantial CPU time and memory. Kraken2 is the recent version of Kraken (Wood & Salzberg, 2014) and uses the *k*-mer based approach for the taxonomic assignment of the query sequences. It is faster, highly precise and requires less CPU memory and time for taxonomic classification (Wood et al., 2019b). In our analysis, taxonomic annotation of the assembled sequences was done using Kraken2 (Wood et al., 2019a) against a database built by combining the NCBI Taxonomy Database and the NCBI Reference Sequence Databases for complete genomes for the domains Bacteria, Archaea and Viruses. Kraken2 is a highly efficient software that associates *k*-mers also known as short genomic substrings with the lowest common ancestor (LCA) taxa. Each sequence (or pair of sequences, in the case of paired reads) that Kraken 2 classifies results in a single line of output with five tab-delimited entities containing information about the taxonomic classification. The results of the taxonomic analysis were visualized utilizing Krona (Ondov et al., 2011).

3.3 Results

3.3.1 Quality control of the raw data

Out of 11339193 raw reads in the sample sequenced for shotgun metagenomics (SMG), 10367458 (91.43%) fulfilled the quality trimming criteria and passed the quality control steps. The reference whole genome sequence data is not available for *Antheraea assamensis.* Therefore, to remove the host sequences, the available whole genome sequence of four related species, namely, *Antheraea pernyi*, *Antheraea mylitta*, *Bombyx mori* and *Bombyx mandarina* were combined and mapped against the metagenomic sequences using Bowtie2. This showed 6.11% overall alignment rate with 172242 sequences (0.87%) aligned exactly 1 time and 448242 sequences (2.27%) aligned more than 1 time. 19122494 sequences (96.86%) remained unmapped which were retained for the downstream analysis. The FastQC report of the quality filtered data is represented in **Table 3. 1** and **Figure 3. 1**.

Table 3. 1 Basic statistics of the quality-controlled shotgun metagenomic sequences

Measure	**Value**
File format	fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	9341333
Sequences flagged as poor quality	0
Sequence length	21-143
%GC	31

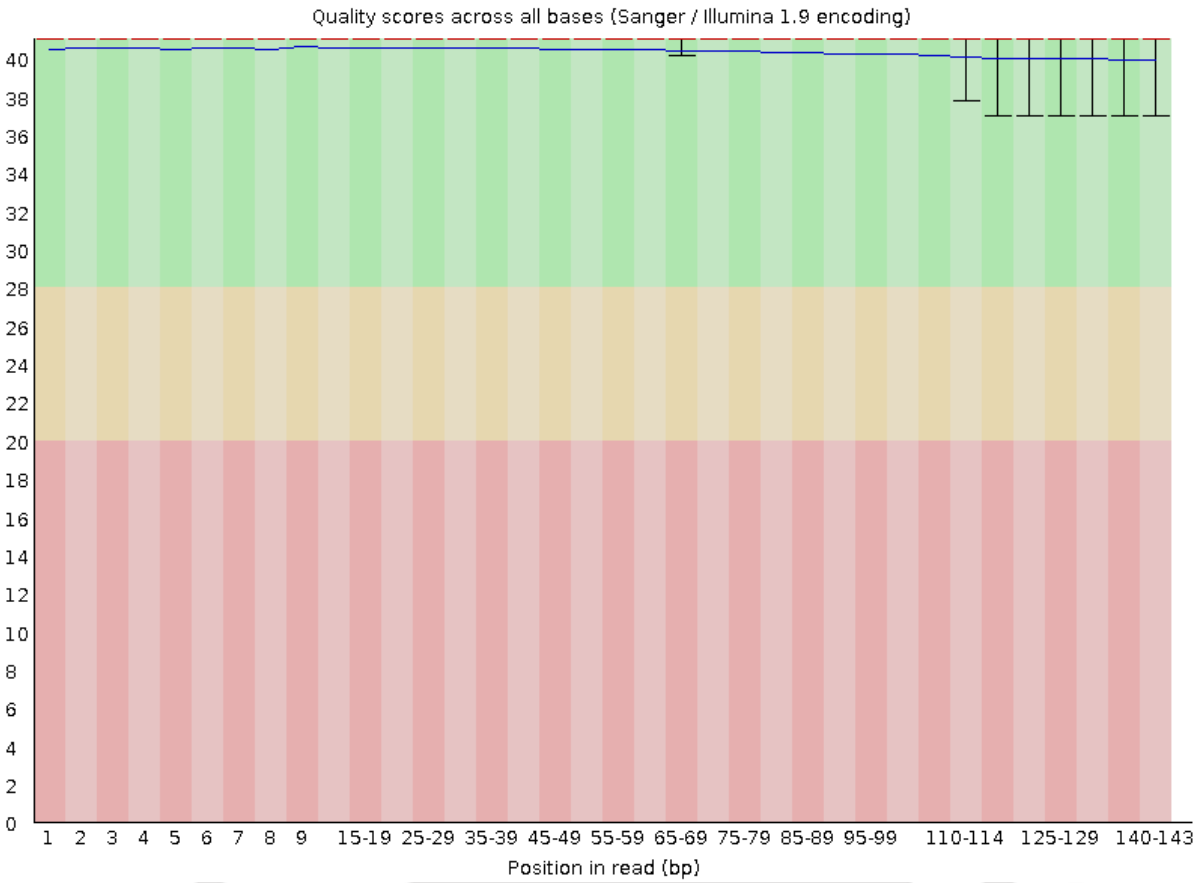


Figure 3. 1 Per base sequence quality of the quality-controlled shotgun metagenomic sequences

3.3.2 Metagenome assembly

De novo Metagenome assembly of 338694821 base pairs using megahit v1.2.9 with *k-mer* length between 21-141 base pairs resulted in 422117 contigs with minimum and a maximum length of 200 bp and 24867 bp respectively. The average contig length of the assembly was 716 base pairs with N50 of 803 base pairs. The sequence statistics of the shotgun sequencing data are represented in **Table 3. 2**.

3.3.3 Functional Annotation

Functional analysis of the assembled sequences with Prokka identified 72063 CDSs, 277 tRNAs and 1185 rRNAs. From the CDSs, 270 enzymes were identified with their EC numbers. The key enzymes identified and their abundances have been depicted in **Figure 3. 2(a).** Categorization of the identified enzymes into the six enzyme classes based on the abundance of enzymes belonging to each class has been represented in **Figure 3. 2(b)**. The identified

enzymes were used for metabolic pathway identification against MetaCyc and KEGG databases using MinPath which resulted in ~405 and 100 pathways respectively (Supplementary File 5). The most abundant metabolic pathways in the gut metagenome of *A. assamensis* identified against MetaCyc database include superpathway of histidine, purine, and pyrimidine biosynthesis, ethene biosynthesis, superpathway of purine nucleotides *de novo* biosynthesis I, colanic acid building blocks biosynthesis, mycolate biosynthesis, superpathway of N-acetylneuraminate degradation, superpathway of anaerobic sucrose degradation, superpathway of heme b biosynthesis from glycine and others.

Table 3. 2 Statistics of the shotgun sequence data.

Attributes	**Gut metagenome**
Sequence size (GB)	4.2
No. of reads	11339193
Sequence strategy	Illumina
Library insert size	2×150
Average read length (bp)	150
Number of reads removed during quality control	1997860
Number of reads retained after quality control	9341333
Number of contigs after assembly	458840
Total bases in contigs	328973024
Largest contig length (bp)	24867
Average contig size (bp)	716
N50	803
GC content	32%

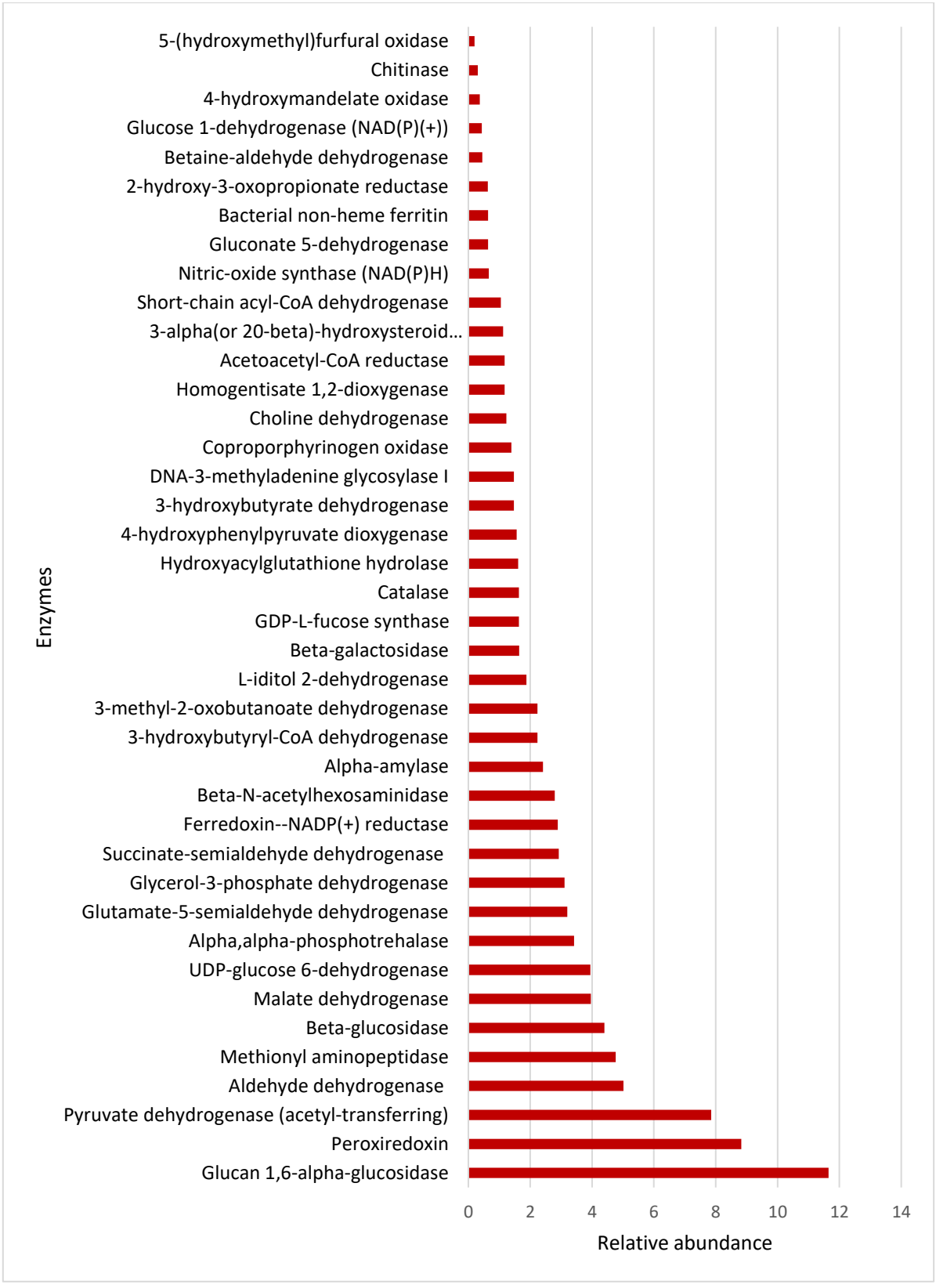


Figure 3. 2(a) Major gene encoding enzymes identified in the gut metagenome of *A. assamensis* through shotgun metagenomic sequencing.

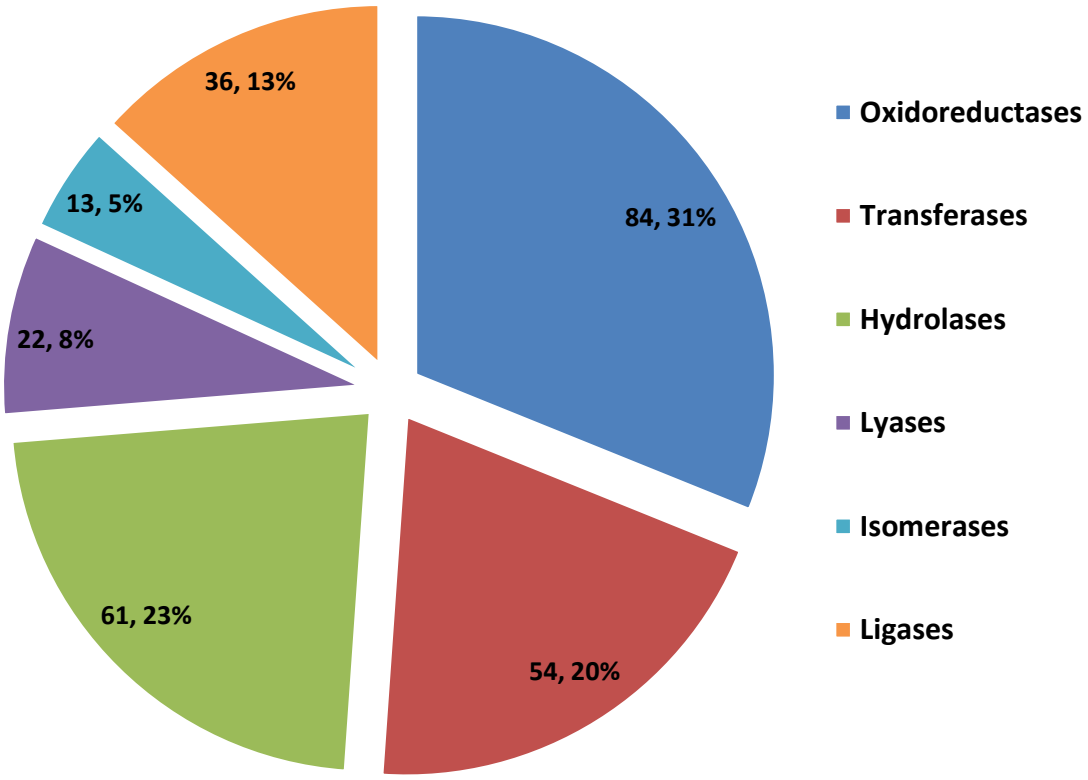


Figure 3. 2(b) Categorization of the identified enzymes in the gut metagenome of *A. assamensis* into different enzyme classes and their abundances.

The major pathways annotated against KEGG database contain pathways for biosynthesis of secondary metabolites, microbial metabolism in diverse environments, amino sugar and nucleotide sugar metabolism, purine metabolism, glycolysis, glyoxylate and dicarboxylate metabolism, pyruvate metabolism, glycine, serine and threonine metabolism, cysteine and methionine metabolism. Gene families involved in pathways like styrene degradation, vitamin B6 metabolism, terpenoid backbone biosynthesis, limonene and pinene degradation, steroid hormone biosynthesis, geraniol degradation, carbapenem biosynthesis, novobiocin biosynthesis, steroid degradation and streptomycin biosynthesis were also observed to be present with less abundance. The coding regions were translated into protein sequences by Prokka which were then annotated against the Clusters of Orthologous Genes (COG) database using WebMGA web server (Wu et al., 2011). This resulted in 1003 orthologous genes belonging to 24 COG categories as represented in **Table 3. 3** and **Figure 3. 3**.

Table 3. 3 Classification of orthologous genes identified through WebMGA webserver into different COG categories.

Class	Description	No. of families	Coverage	Abundance
J	Translation, ribosomal structure and biogenesis	245	0.477551	0.084013
A	RNA processing and modification	25	0.64	0.004786
K	Transcription	231	0.242424	0.055655
L	Replication, recombination and repair	238	0.264705	0.143972
B	Chromatin structure and dynamics	19	0.789473	0.008272
D	Cell cycle control, cell division, chromosome partitioning	72	0.388888	0.021565
Y	Nuclear structure	2	1	0.000523
V	Defense mechanisms	46	0.130434	0.008724
T	Signal transduction mechanisms	152	0.151315	0.031464
M	Cell wall/membrane/envelope biogenesis	188	0.159574	0.013056
N	Cell motility	96	0.03125	0.000705
Z	Cytoskeleton	12	1	0.016553
W	Extracellular structures	1	0	0
U	Intracellular trafficking, secretion, and vesicular transport	158	0.189873	0.078966
O	Posttranslational modification, protein turnover, chaperones	203	0.389162	0.150363
C	Energy production and conversion	258	0.368217	0.090217
G	Carbohydrate transport and metabolism	230	0.260869	0.041841
E	Amino acid transport and metabolism	270	0.292592	0.062768
F	Nucleotide transport and metabolism	95	0.452631	0.024241
H	Coenzyme transport and metabolism	179	0.251396	0.022105
I	Lipid transport and metabolism	94	0.404255	0.112667
P	Inorganic ion transport and metabolism	212	0.165094	0.023883
Q	Secondary metabolites biosynthesis, transport and catabolism	88	0.261363	0.090721
R	General function prediction only	702	0.200854	0.1877
S	Function unknown	1347	0.04974	0.032886

The majority of the COGs were assigned to the category of general function prediction. The well-characterized COGs were categorized into distinct categories such as secondary metabolites biosynthesis, transport and catabolism; lipid transport and metabolism; amino acid

transport and metabolism; energy production and conversion; inorganic ion transport and metabolism; intracellular trafficking, secretion, and vesicular transport; post translational modification, protein turnover, chaperones; replication, recombination and repair; defence mechanism etc.

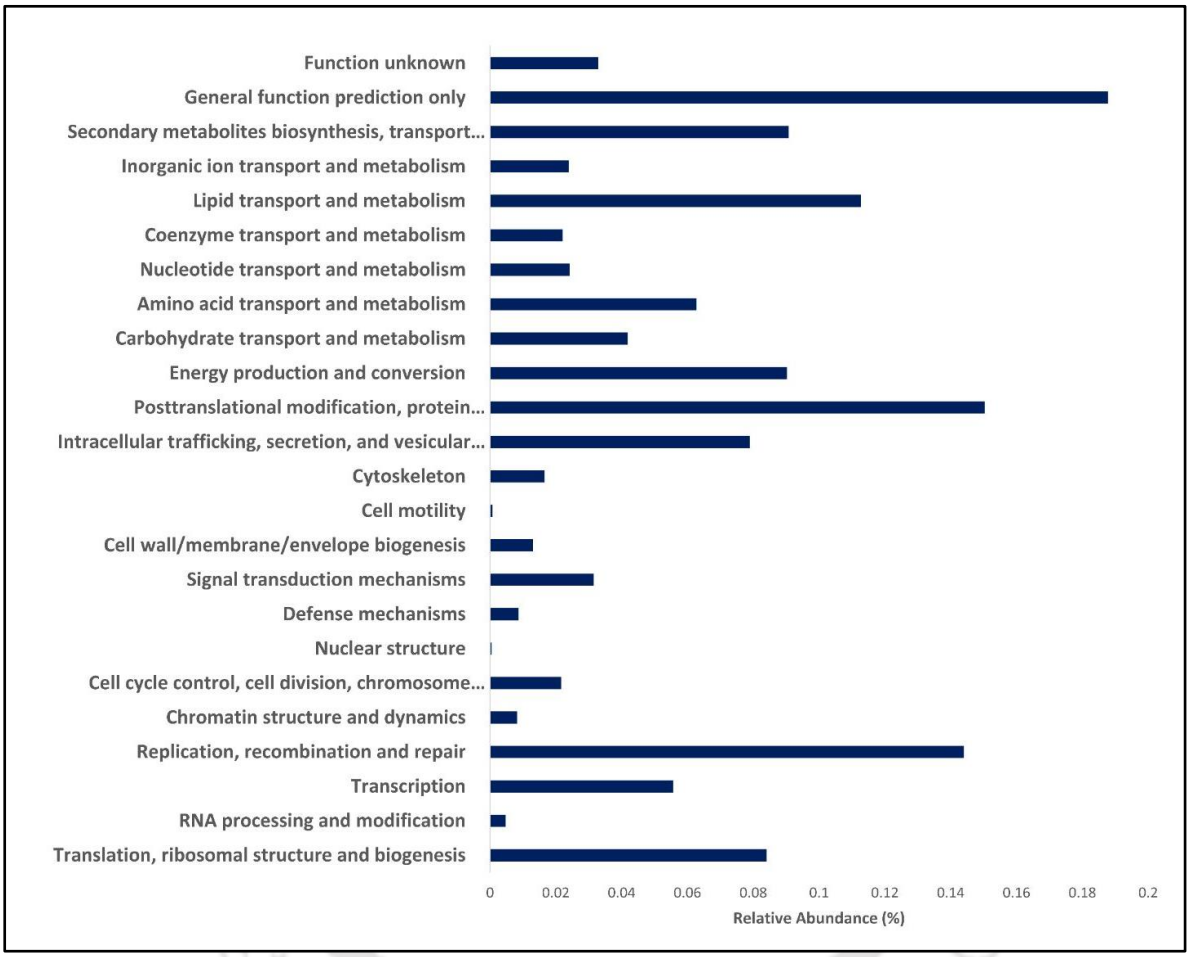


Figure 3. 3 COG classes of the pathways identified. The coding regions were translated into protein sequences by Prokka which were then annotated against the Clusters of Orthologs (COG) database using the WebMGA web server. This resulted in 1003 orthologous genes belonging to 24 COG categories as represented here.

Based on the gene annotations, parsimonious pathways were constructed using MinPath against MetaCyc and KEGG databases. Visualization of the metabolic network map based on KEGG pathways was done using Interactive Pathways Explorer version 3 (**Figure 3. 4**). The pathways identified against the MEtaCyc and KEGG databases are listed in **Table 3. 4** and **Table 3. 5**, respectively.

Table 3. 4 MetacCyc pathways and the number of gene families identified in the gut metagenome of *A. assamensis*.

MetaCyc pathway	Description	Family-total	Family-found
PRPP-PWY	Superpathway of histidine, purine, and pyrimidine biosynthesis	37	10
PWY-7124	Ethene biosynthesis	21	9
DENOVOPURINE2-PWY	DENOVOPURINE2-PWY	18	8
PWY-841	Superpathway of purine nucleotides de novo biosynthesis I	17	8
PWY-6886	1-butanol autotrophic biosynthesis	14	6
COLANSYN-PWY	Colanic acid building blocks biosynthesis	9	6
PWYG-321	Mycolate biosynthesis	23	6
P441-PWY	Superpathway of N-acetylneuraminate degradation	21	6
PWY-7345	Superpathway of anaerobic sucrose degradation	14	5
PWY-5918	Superpathway of heme b biosynthesis from glutamate	9	5
PWY-5920	Superpathway of heme b biosynthesis from glycine	8	5
ANAGLYCOLYSIS-PWY	ANAGLYCOLYSIS-PWY	10	4
PWY-7695	Aurofusarin biosynthesis	6	4
P341-PWY	Glycolysis V (Pyrococcus)	9	4
ANAEROFRUCAT-PWY	Homolactic fermentation pathway	11	4
PWY-7575	Superpathway of candicidin biosynthesis	8	4
PWY-6947	Superpathway of cholesterol degradation II (cholesterol dehydrogenase)	23	4
PWY66-409	Superpathway of purine nucleotide salvage	11	4
PWY-7659	Viridicatumtoxin biosynthesis	7	4
PWY-7692	Bikaverin biosynthesis	3	3
PWY-7673	Fusarin C biosynthesis	5	3

ARGSYN-PWY	L-arginine biosynthesis I	8	3
PWY-6397	Mycolyl-arabinogalactan-peptidoglycan complex biosynthesis	12	3
PWY-7287	Novobiocin biosynthesis	15	3
P221-PWY	Octane oxidation	4	3
SUCSYN-PWY	Sucrose biosynthesis I	8	3
PWY-6277	Superpathway of 5-aminoimidazole ribonucleotide biosynthesis	5	3
PWY-6928	Superpathway of cholesterol degradation I (cholesterol oxidase)	15	3
PWY-5464	Superpathway of cytosolic glycolysis (plants), pyruvate dehydrogenase and TCA cycle (PWY-5464)	11	3
PWY-6977	Superpathway of erythromycin biosynthesis	16	3
PWY-7323	Superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis	9	3
PWY-561	Superpathway of glyoxylate cycle and fatty acid degradation	13	3
PWY-7228	Superpathway of guanosine nucleotides de novo biosynthesis I	5	3
PWY-6125	Superpathway of guanosine nucleotides de novo biosynthesis II	5	3
PWY-5004	Superpathway of L-citrulline metabolism	11	3
PWY-5327	Superpathway of L-lysine degradation	31	3
THREOCAT-PWY	Superpathway of L-threonine metabolism	14	3
PWY-7110	Superpathway of megalomicin A biosynthesis	11	3
PWY-5840	Superpathway of menaquinol-7 biosynthesis	7	3
PWY-5838	Superpathway of menaquinol-8 biosynthesis I	9	3
PWY-7211	Superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis	14	3
PWY0-162	Superpathway of pyrimidine ribonucleotides de novo biosynthesis	9	3

SULFATE-CYS-PWY	Superpathway of sulfate assimilation and cysteine biosynthesis	9	3
PWY-7328	Superpathway of UDP-glucose-derived O-antigen building blocks biosynthesis	6	3
PWY-7250	[2Fe-2S] iron-sulfur cluster biosynthesis pathway	3	2
PWY-7565	Aspyridone A biosynthesis pathway	4	2
PWY-5903	Bacillibactin biosynthesis pathway	4	2
PWY-8047	Bryostatin biosynthesis pathway	7	2
PWY-8036	Chaxamycin biosynthesis pathway	5	2
PWY-6981	Chitin biosynthesis	9	2
PWY-7669	Equisetin biosynthesis	3	2
PWY-7094	Fatty acid salvage pathway	6	2
BETSYN-PWY	Glycine betaine biosynthesis I pathway (Gram-negative bacteria)	2	2
PWY-7653	Griseofulvin biosynthesis	4	2
PWY-8049	Pederin biosynthesis pathway	6	2
PWY-6627	Salinosporamide A biosynthesis pathway	9	2
PWY-8035	Streptovaricin biosynthesis pathway	6	2
PWY-8133	Succinoglycan biosynthesis pathway	3	2
PWY-6396	Superpathway of 2,3-butanediol biosynthesis	7	2
PWY-6285	Superpathway of fatty acids biosynthesis (E. Coli)	11	2
POLYAMINSYN3-PWY	Superpathway of polyamine biosynthesis II	7	2
PWY-6395	Superpathway of seleno-compound metabolism	5	2
PWY-821	Superpathway of sulfur amino acid biosynthesis	9	2
PWY-7820	Teichuronic acid biosynthesis (B. Subtilis 168)	6	2

Table 3. 5 List of KEGG pathways identified in the functional gut metagenome of Antheraea assamensis

Pathway ID	Description	Pathway ID	Description
map00010	Glycolysis / Gluconeogenesis	map00561	Glycerolipid metabolism
map00020	Citrate cycle (TCA cycle)	map00562	Inositol phosphate metabolism
map00030	Pentose phosphate pathway	map00564	Glycerophospholipid metabolism
map00040	Pentose and glucuronate interconversions	map00590	Arachidonic acid metabolism
map00051	Fructose and mannose metabolism	map00592	alpha-Linolenic acid metabolism
map00052	Galactose metabolism	map00600	Sphingolipid metabolism
map00053	Ascorbate and aldarate metabolism	map00603	Glycosphingolipid biosynthesis - globo and isoglobo series
map00061	Fatty acid biosynthesis	map00604	Glycosphingolipid biosynthesis - ganglio series
map00062	Fatty acid elongation	map00620	Pyruvate metabolism
map00071	Fatty acid degradation	map00625	Chloroalkane and chloroalkene degradation
map00130	Ubiquinone and other terpenoid-quinone biosynthesis	map00627	Aminobenzoate degradation
map00140	Steroid hormone biosynthesis	map00630	Glyoxylate and dicarboxylate metabolism

map00220	Arginine biosynthesis	map00640	Propanoate metabolism
map00230	Purine metabolism	map00643	Styrene degradation
map00240	Pyrimidine metabolism	map00650	Butanoate metabolism
map00250	Alanine, aspartate and glutamate metabolism	map00660	C5-Branched dibasic acid metabolism
map00254	Aflatoxin biosynthesis	map00670	One carbon pool by folate
map00260	Glycine, serine and threonine metabolism	map00680	Methane metabolism
map00261	Monobactam biosynthesis	map00710	Carbon fixation in photosynthetic organisms
map00270	Cysteine and methionine metabolism	map00720	Carbon fixation pathways in prokaryotes
map00280	Valine, leucine and isoleucine degradation	map00730	Thiamine metabolism
map00281	Geraniol degradation	map00750	Vitamin B6 metabolism
map00290	Valine, leucine and isoleucine biosynthesis	map00760	Nicotinate and nicotinamide metabolism
map00310	Lysine degradation	map00770	Pantothenate and CoA biosynthesis
map00330	Arginine and proline metabolism	map00780	Biotin metabolism
map00332	Carbapenem biosynthesis	map00785	Lipoic acid metabolism
map00340	Histidine metabolism	map00790	Folate biosynthesis
map00350	Tyrosine metabolism	map00860	Porphyrin and chlorophyll metabolism

map00360	Phenylalanine metabolism	map00900	Terpenoid backbone biosynthesis
map00362	Benzoate degradation	map00903	Limonene and pinene degradation
map00365	Furfural degradation	map00908	Zeatin biosynthesis
map00380	Tryptophan metabolism	map00910	Nitrogen metabolism
map00400	Phenylalanine, tyrosine and tryptophan biosynthesis	map00920	Sulfur metabolism
map00401	Novobiocin biosynthesis	map00930	Caprolactam degradation
map00404	Staurosporine biosynthesis	map00940	Phenylpropanoid biosynthesis
map00410	beta-Alanine metabolism	map00950	Isoquinoline alkaloid biosynthesis
map00430	Taurine and hypotaurine metabolism	map00960	Tropane, piperidine and pyridine alkaloid biosynthesis
map00450	Selenocompound metabolism	map00966	Glucosinolate biosynthesis
map00460	Cyanoamino acid metabolism	map00970	Aminoacyl-tRNA biosynthesis
map00470	D-Amino acid metabolism	map00980	Metabolism of xenobiotics by cytochrome P450
map00480	Glutathione metabolism	map00981	Insect hormone biosynthesis
map00500	Starch and sucrose metabolism	map00983	Drug metabolism - other enzymes
map00511	Other glycan degradation	map00984	Steroid degradation
map00513	Various types of N-glycan biosynthesis	map00998	Biosynthesis of various secondary metabolites - part 2

map00520	Amino sugar and nucleotide sugar metabolism	map01051	Biosynthesis of ansamycins
map00521	Streptomycin biosynthesis	map01053	Biosynthesis of siderophore group nonribosomal peptides
map00522	Biosynthesis of 12-, 14- and 16-membered macrolides	map01055	Biosynthesis of vancomycin group antibiotics
map00531	Glycosaminoglycan degradation	map01100	Metabolic pathways
map00541	O-Antigen nucleotide sugar biosynthesis	map01110	Biosynthesis of secondary metabolites
map00550	Peptidoglycan biosynthesis	map01120	Microbial metabolism in diverse environments

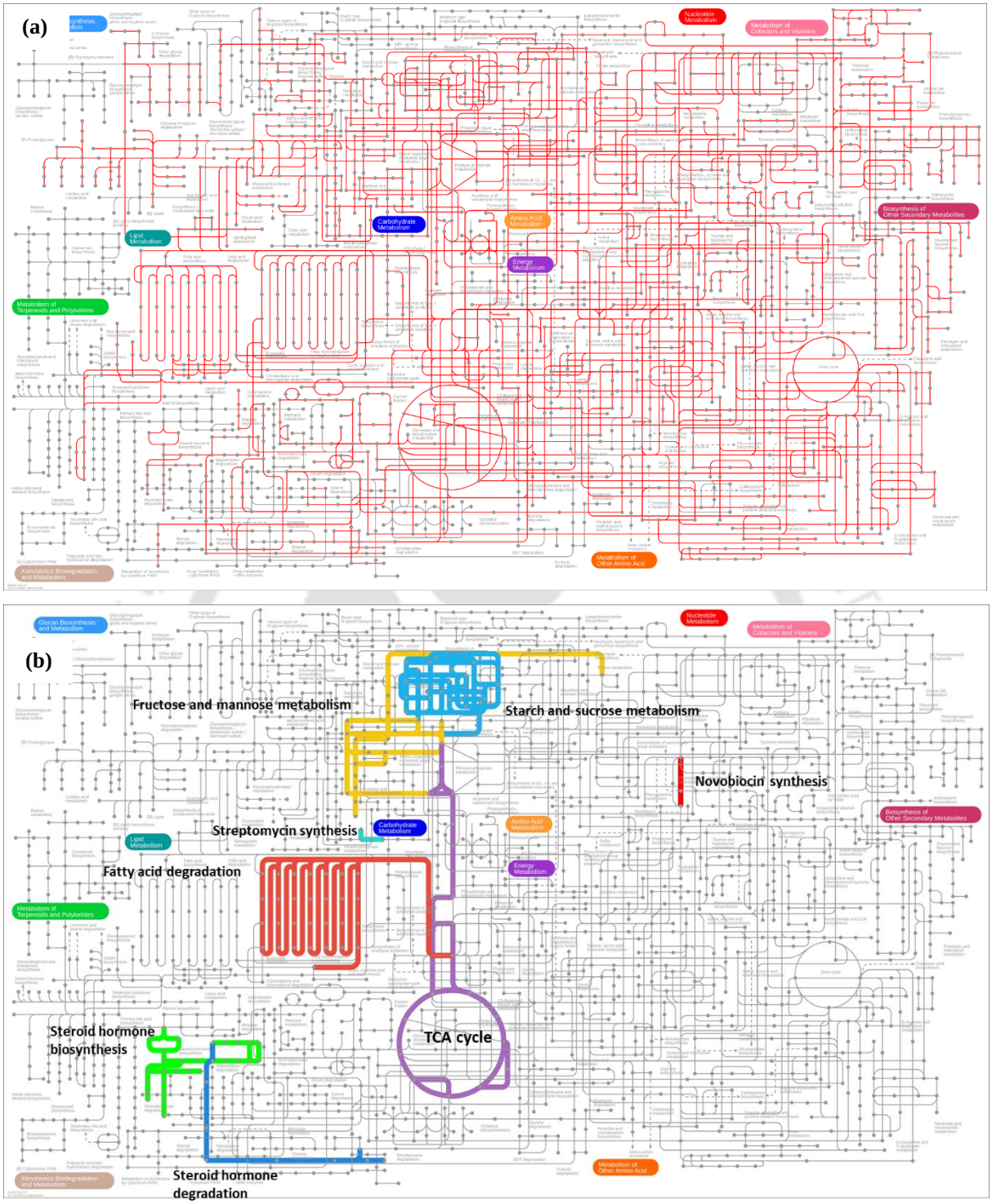


Figure 3. 4 Visualization of the metabolic interaction network map based on KEGG pathways using Interactive Pathways Explorer 3 (iPATH3). (a) A total of 102 KEGG pathways were identified which were mapped against the metabolic interaction network. The nodes represent chemical compounds and the edges represent enzymatic reactions on the map. (b) Few key pathways identified in the gut metagenome of *A. assamensis* have been highlighted.

3.3.4 Taxonomic analysis

Taxonomic analysis of the shotgun data was performed using Kraken2, a fast and feasible tool for taxonomic classification that uses exact *k*-mer matching to achieve high accuracy (Wood et al., 2019a). This revealed the presence of several taxa which were not detected or identified by the amplicon sequencing approach. 97% of all the identified taxa consisted of the kingdom Bacteria, 2.4% Archaea and 0.06 % Viruses. Among the bacterial community, phyla like Proteobacteria, Firmicutes, Actinobacteria, Chloroflexi, Armatimonadetes Acidobacteria, Verrucomicrobia, Planctomycetes, Spirochaetes, Nitrospirae, Fusobacteria, Thermotogae, Chlamydiae, Tenericutes, Deinococcus-Thermus were predominant. Among the Proteobacteria, the class Gammaproteobacteria was the highest abundance followed by Alphaproteobacteria, Betaproteobacteria and Deltaproteobacteria, respectively. The phylum Firmicutes contained taxa from classes like Bacilli, Clostridia, Negativicutes, Tissierellia, Erysipelotrichia and Limnochordia. Similarly, Actinobacteria contained Actinomycetia, Acidimicrobiia, Rubrobacteria, and Coriobacteriia. Taxa such as *Vibrio anguillarum*, *Serratia surfactantfaciens*, *S. marcescens*, *S. symbiotica*, *Pantoea* sp., *Pseudomonas* sp., *Acinetobacter baumannii*, *Moraxella osloensis*, *Psychrobacter* sp., *Stenotrophomonas maltophilia* (opportunistic pathogen), *Xanthomonas euvesicatoria*, *Rhodanobacter* sp. *Klebsiella pneumoniae*, *Salmonella enterica*, Escherichia *coli* and *Enterobacter* have been identified with significant abundance. **Figure 3. 5** represents the major bacterial taxa identified in the gut metagenome of *A. assamensis* larvae through shotgun sequencing.

Similarly, the majority of the archaeal species identified consisted of the phyla, Thaumarchaeota, Euryarchaeota, Candidatus Thermoplasmatota and Crenarchaeota in the

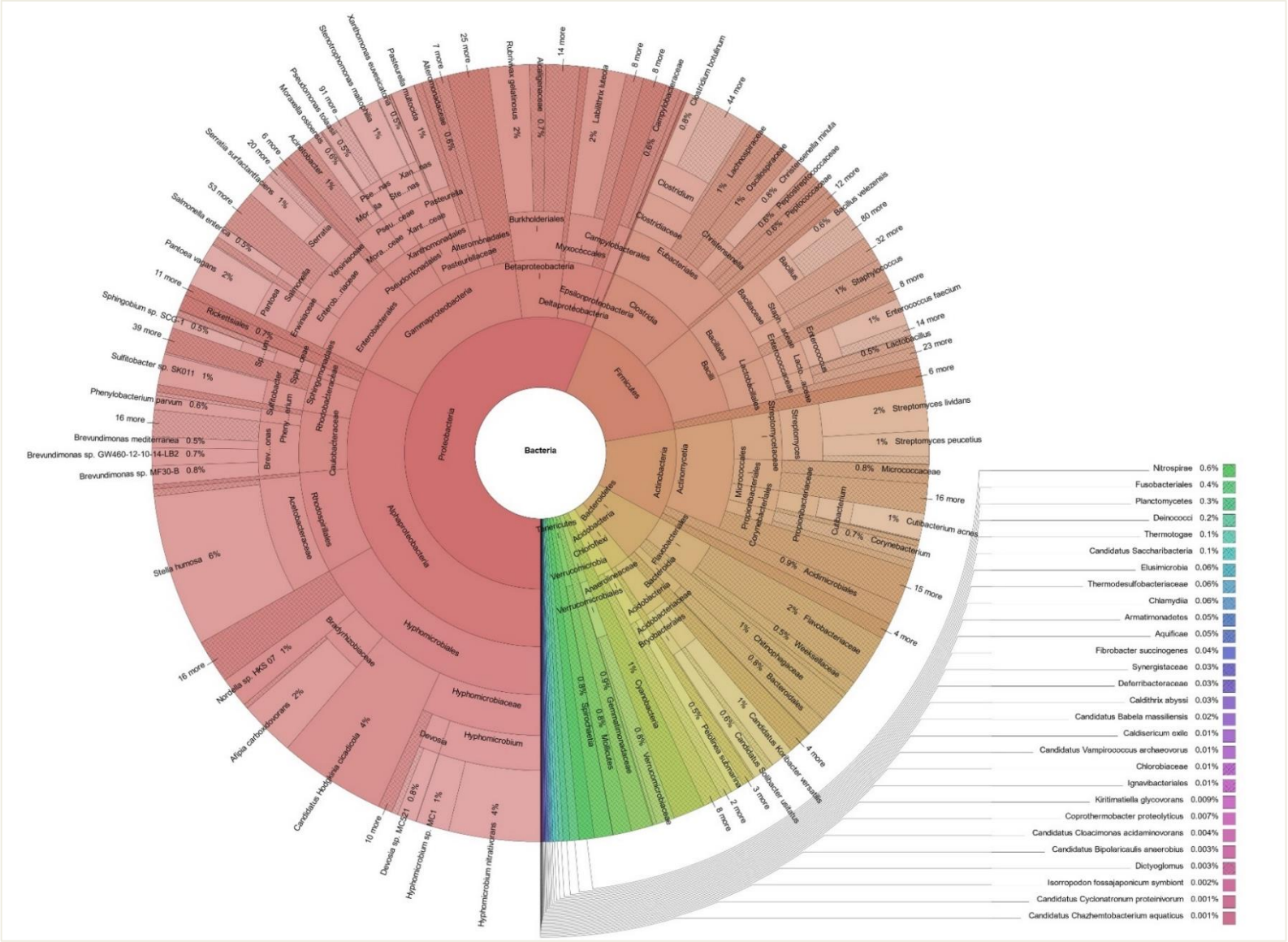


Figure 3. 5 The major bacterial taxa identified in the gut metagenome of *A. assamensis* larvae through shotgun sequencing.

descending order of their abundance (**Figure 3. 6**). Thaumarchaeota consisted of the key genera *Nitrosopumilus* and Candidatus *Nitrosopelagicus*. Similar to the amplicon sequence results, methanogens were found to be the key constituents of the gut euryarchaeal community of *A. assamensis*.

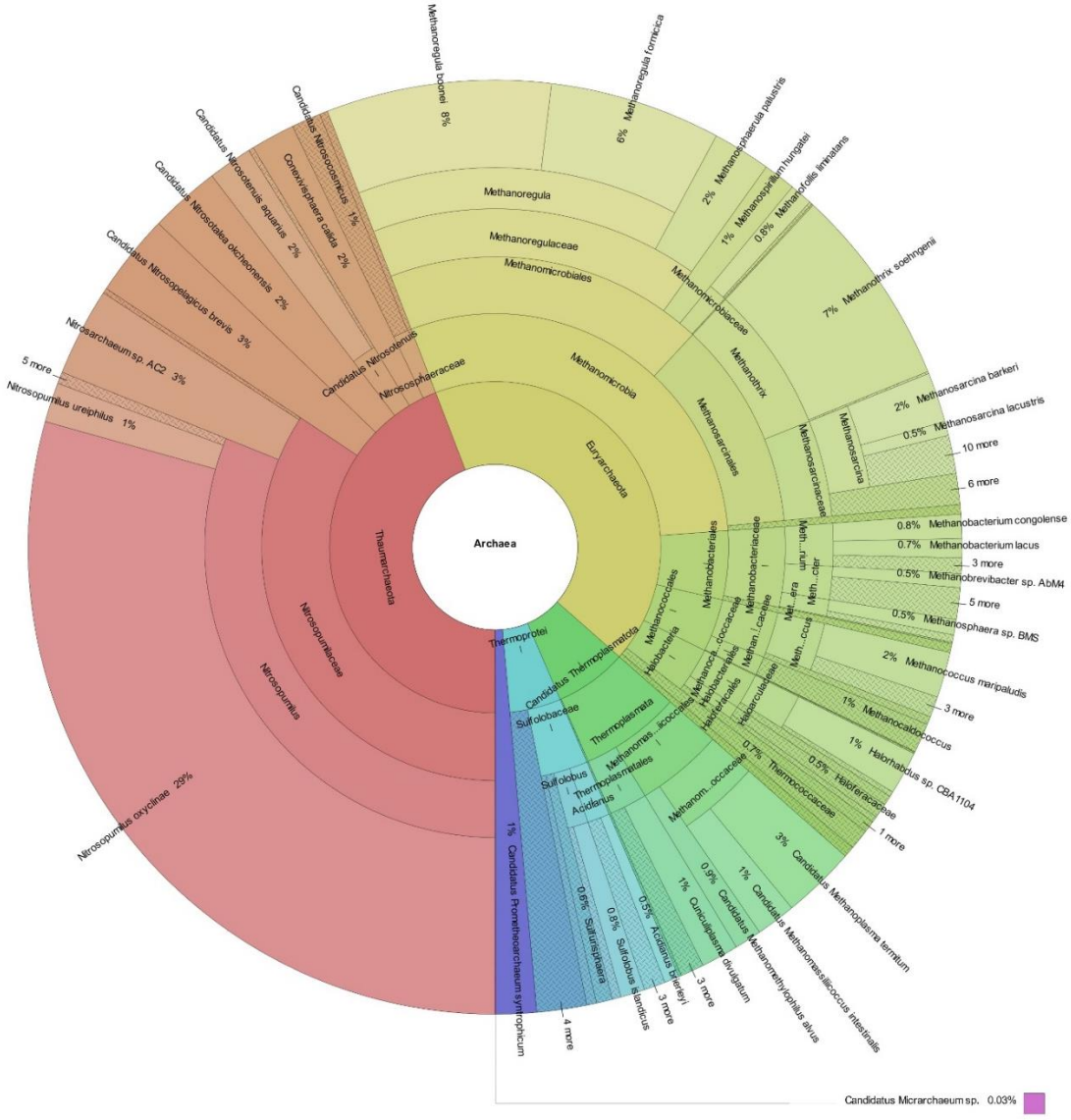


Figure 3. 6 The major archaeal taxa identified in the gut metagenome of *A. assamensis* larvae

The gut methanogens include *Methanoregula*, *Methanosphaerula*, *Methanospirillum*, *Methanofollis*, *Methanothrix*, *Methanosarcina*, *Methanohalophilus*, *Methanocella*, *Methanobacterium*, *Methanobrevibacter*, *Methanosphaera*, *Methanococcus*, and *Methanocaldococcus*. The methanogens identified in the gut metagenome of *A. assamensis*

larvae at different taxonomic levels through shotgun sequencing is depicted in **Figure 3. 7**.

Figure 3. 7 Methanogens identified in the gut metagenome of *A. assamensis* larvae and their abundance

At the kingdom level, the viral taxa were identified as Bamfordvirae, Orthornavirae, and Shotokuvirae. At the phylum level, Preplasmiviricota, Nucleocytoviricota, Uroviricota, Kitrinoviricota, and Peploviricota were abundant. At the species level, Agrotis segetum nucleopolyhedrovirus A, Perigonia lusca nucleopolyhedrovirus, Adoxophyes orana granulovirus, Choristoneura fumiferana granulovirus, Spodoptera litura granulovirus, Cotesia congregata bracovirus, Adoxophyes honmai entomopoxvirus had been detected. The identified viral taxa and their abundances in the studied gut metagenome has been depicted in **Figure 3.8.**

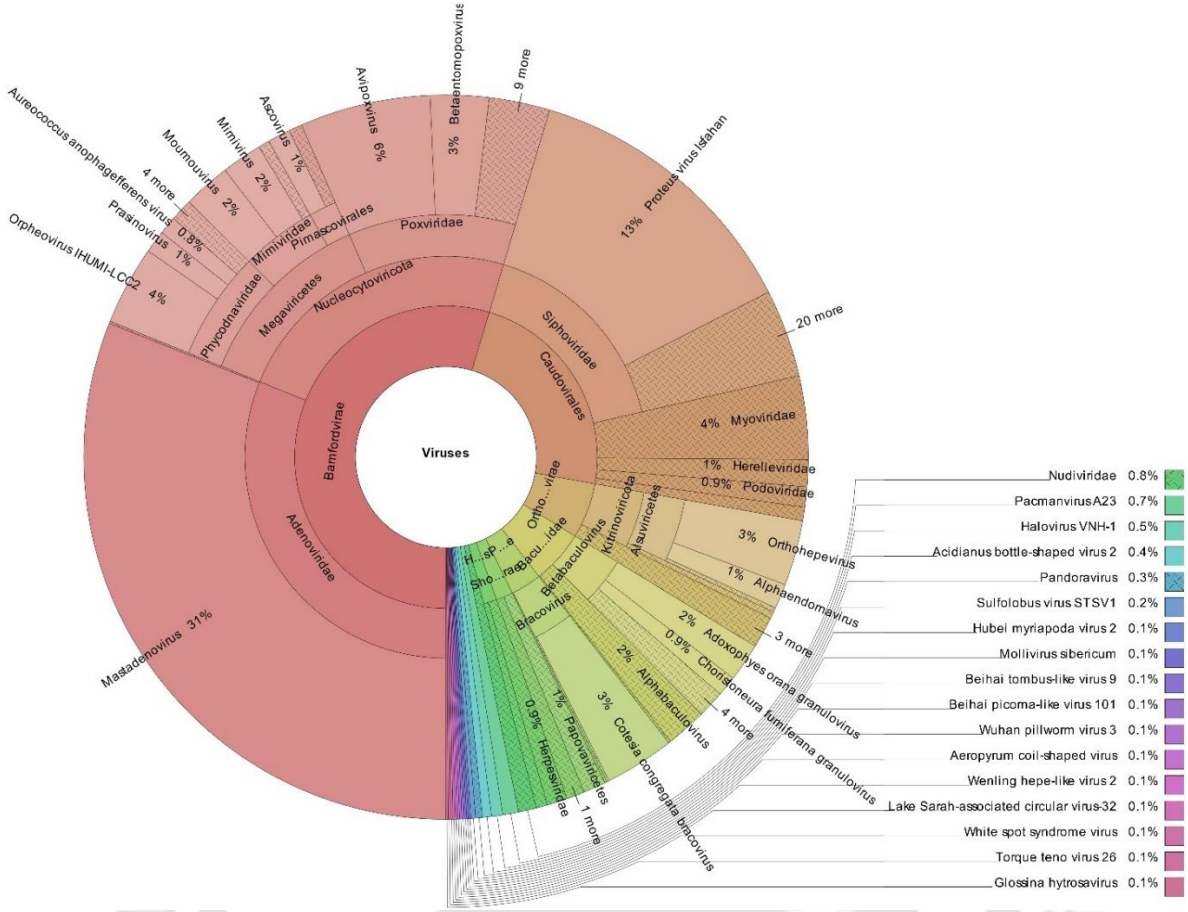


Figure 3. 8 The viral taxa identified in the gut metagenome of *A. assamensis* with their abundance.

3.4 Discussions

3.4.1 Functional annotation of the gut metagenome

Shotgun metagenomic analysis of the gut microbial community of *Antheraea assamensis* identified 270 genes encoding enzymes out of which 205 were also detected in the predicted gut metagenome through the PICRUSt2 analysis in our previous objective. Identification of genes encoding starch and cell wall polysaccharide degrading enzymes such as alpha-amylase; chitinase; beta-glucosidase; beta-galactosidase; beta-N-acetylhexosaminidase; glucan 1,6-alpha-glucosidase; glucan 1,4-beta-glucosidase; and alpha,alpha-phosphotrehalase reflects that gut metagenome may play a pivotal role in food digestion of the herbivorous larvae (kindly refer **Figure 3. 2(a)**) (Cerqueira et al., 2012; da Lage, 2018; Zibaee, 2012). The presence of

other genes producing enzymes such as chitinase, catalase and acetyl-CoA carboxylase shows the potential role of the gut microbiome in the degradation of chitin, detoxification of plant secondary metabolites and fat body metabolism, respectively (Merzendorfer & Zimoch, 2003; Towle et al., 2003; X. Xia et al., 2017b).

KEGG pathway annotation (**Table 3. 5**) elucidated the key microbial pathways present within the gut microbial metagenome such as fatty acid biosynthesis and degradation, steroid hormone biosynthesis, geraniol degradation, benzoate degradation, starch and sucrose metabolism, methane metabolism, insect hormone biosynthesis, staurosporine biosynthesis and steroid degradation. Fatty acids play a major role in energy storage and utilization, and serve as a precursor molecule for pheromone and eicosanoid biosynthesis in insects (Arrese & Soulages, 2010). Our results show that the gut microbial community play a crucial role in both the synthesis and degradation of fatty acids in *A. assamensis* larvae. Staurosporine is a microbial alkaloid exhibiting anticancer properties through the inhibition of protein kinase C (Tamaoki et al., 1986). It is synthesized by *Streptomyces* sp. from the key building blocks tryptophan, glucose and two methyl groups from methionine (Salas et al., 2005). Both the amplicon and shotgun analyses of the gut microbial community revealed the presence of *Streptomyces* species inside the gut of *A. assamensis* larvae. Identification of staurosporine biosynthesis pathway inside the gut metagenome of *A. assamensis* is interesting and can be a potential topic for in-depth future studies. Insect metamorphosis is regulated by the principal steroid hormone, ecdysteroid or ecdysone(Kirimura et al., 1962). This is synthesised by the insect inside a specialised endocrine organ named as prothoracic gland from dietary cholesterol (Niwa & Niwa, 2016). Host-associated microbiota are well reported to influence the synthesis of steroid hormones in several animals including humans(Doden & Ridlon, 2021; Qi et al., 2021; Tetel et al., 2018). The presence of steroid hormone synthesis and degradation pathways within the gut metagenome of *A. assamensis* shows the potential association of gut microbiota in larval gut

microbiota in steroid hormone metabolism. Several pathways involved in the biosynthesis of antibiotics have also been identified. For example, Novobiocin biosynthesis, Streptomycin biosynthesis, Monobactam biosynthesis and Carbapenem biosynthesis (**Table 3. 5**).

3.4.2 Taxonomic annotation of the gut bacterial community of *A. assamensis*.

Taxonomic annotation of the assembled reads from shotgun metagenomic sequences was performed in Kraken2 against NCBI taxonomy database. More than 1000 bacterial species (97% of the total identified microbes) were identified in this study with taxa identified up to the genus level. The phylum Proteobacteria was found to the highest abundant followed by Firmicutes, Actinobacteria, Chloroflexi, Tenericutes, Deinococcus-Thermus, Armatimonadetes, Bacteroidetes, Chlorobi and others (**Figure 3. 5**). Proteobacteria is a phylum of bacteria that is found in a wide variety of environments, including soil, water, and the gastrointestinal tracts of animals, including humans. Within the human gut, Proteobacteria typically account for 1-5% of the total bacterial population, although this can vary depending on factors such as diet, age, and health status (Rizzatti et al., 2017). However, in insects, Protobacteria was reported to be highly dominant occupying approximately 60% of the total gut microbial population (Yun et al., 2014). Inside the gut, they play many key functional roles like maintaining the health of the gut and host, breakdown and fermentation of dietary fibers. However, their function varies from one individual to the other based on age, diet, sex and other factors (Bradley & Pollard, 2017). One of the most interesting functions of gut proteobacteria is that they prepare the gut environment for colonization by strictly anaerobic microorganisms by lowering the redox potential of the gut and consuming the oxygen (Moon et al., 2018). The gut environment of many insects has either no or very low oxygen concentration (Bradley & Pollard, 2017). The presence of Proteobacteria inside the gut in a high abundance can be a key factor determining the high diversity of gut microbial community inside the gut of many lepidopteran species including our organism of interest. Firmicutes is a

phylum of bacteria that includes many species found in the gut of insects, as well as other animals. These bacteria play important roles in the digestion and metabolism of their host insects and are often considered mutualistic symbionts. Some examples of Firmicutes found in the guts of *A. assamensis* include *Lactobacillus*, *Enterococcus*, and *Clostridium* species. These bacteria are often acquired by insects from their environment and play a crucial role in several biological processes of the host organism (Siddiqui et al., 2022; Zheng et al., 2021).

3.4.3 The gut archaeal community of *A. assamensis*

From the shotgun metagenomic analysis of the gut microbial community of *A. assamensis*, we could identify 70 archaeal taxa up to the genus level. Archaea are a type of microorganism that is genetically distinct from bacteria and eukaryotes (Esko et al., 2009). They are present in many different environments, including the human gut (Kim et al., 2020). The gut archaeal community is less-studied than the bacterial community in the gut, but recent research has shown that archaea may play an important role in gut health (Gaci et al., 2014; Hoegenauer et al., 2022; van de Pol et al., 2017). In particular, the presence of certain archaeal species has been associated with a reduced risk of inflammatory bowel disease (IBD) (Houshyar et al., 2021a; Krawczyk et al., 2021). Methanogenic archaea are a type of archaea that are commonly found in the gut. These microorganisms produce methane gas as a by-product of their metabolism (Guindo et al., 2020; Triantafyllou et al., 2014). While methane production can cause discomfort and bloating in some individuals, recent research has suggested that methanogenic archaea may also have beneficial effects on gut health. For example, some studies have shown that methanogenic archaea can help to regulate the levels of hydrogen gas in the gut, which can reduce the risk of inflammation and damage to the intestinal lining. Methanogenic archaea have also been shown to have anti-inflammatory effects in the gut (Houshyar et al., 2021b; Poles et al., 2019). In our study, methanogens like *Methanoregula*, *Methanothrix*, *Methanosarcina*, *Methanobacterium*, *Methanococcus*, *Methanobrevibacter*,

were found to be abundantly present (**Figure 3. 7**). Association of methanogens with insect gut and methane emission is well-established (Brune, 2010, 2019; Horváthová et al., 2021a, 2021b; Oonincx et al., 2010). Future exploration estimating and quantifying the methane production by *A. assamensis* can provide insight into the contribution of this species to global methane emission. Besides the methanogens, the gut environment of *A. assamensis* was found to be inhabited by other archaeal species like *Nitrosopumilus*, *Nitrosarchaeum*, *Halomicrobium*, *Thermofilum*, *Haloplanus*, and *Halosimplex*. Some of these species were reported from few animals including humans (Abaramak et al., 2021; Klammsteiner et al., 2020; Satari et al., 2021). However, their functional role is yet to be well understood.

3.4.4 The viral population inside the gut of *A. assamensis*

Insects may acquire viruses from virus-contaminated food sources, environment and through vertical transmission (Ma et al., 2021). The association of viruses with the invertebrate intestine is less studied in comparison to other animals (Harvey & Holmes, 2022). Recently, hundreds of viruses have been identified through large-scale meta-transcriptomic analysis of diverse invertebrate taxa (Ma et al., 2021; Shi et al., 2018). Baculoviruses and entomopoxviruses are well known to infect the order Lepidoptera (Ma et al., 2021). Among the baculoviruses, genera from *Alphabaculovirus* and *Betabaculoviruses* are the most common (Ma et al., 2021). In our study, we have identified the genera *Alphabaculovirus* and *Betabaculovirus*. At the species level, *Alphabaculovirus* contained *Agrotis segetum* nucleopolyhedrovirus A and *Perigonia lusca* nucleopolyhedrovirus. Similarly, *Betabaculovirus* contained A*doxophyes orana* granulovirus, *Choristoneura fumiferana* granulovirus and *Spodoptera litura* granulovirus. These viruses are well-reported to infect other Lepidopteran species (Ardisson-Araújo et al., 2016; Escasa et al., 2006; Gueli Alletti et al., 2018; Nakai et al., 2015; Y. Wang et al., 2011). Adoxophyes honmai entomopoxvirus belonging to the genus *Betaentomopoxvirus* is also found to be present inside the gut of *A. assamensis* larvae.

3.5 Conclusion

Amplicon analysis of the gut microbial community of *A. assamensis* in our previous objective delineated approximately 250 microbial species including bacteria and archaea. However, most of the microbial taxa could be identified only up to the family or genus level. The functional metagenomic analysis of the gut microbial community of *Antheraea assamensis* through shotgun metagenomic sequencing identified ~1100 bacteria, 78 archaea and 50 viral genera. Functional annotation of the identified genes revealed several crucial functions of the larval gut microbiome which may help in the key physiological processes of the host organism. However, functional analysis of microbial community through metagenomics can only identify the genes, genomes and pathways. The expression of the genes encoded by the collective microbial community can be performed through the analysis of the RNA transcripts which will be the key focus of our next objective.

References

Abaramak, G., Porras-Domínguez, J. R., van Rensburg, H. C. J., Lescrinier, E., Öner, E. T., Kırtel, O., & van den Ende, W. (2021). Functional and molecular characterization of the halomicrobium sp. Ibsba inulosucrase. *Microorganisms*, *9*(4). https://doi.org/10.3390/MICROORGANISMS9040749/S1

Abril, J. F., & Castellano, S. (2019). Genome Annotation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*, 195–209. https://doi.org/10.1016/B978-0-12-809633-8.20226-4

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ardisson-Araújo, D. M. P., Lima, R. N., Melo, F. L., Clem, R. J., Huang, N., Báo, S. N., Sosa-Gómez, D. R., & Ribeiro, B. M. (2016). Genome sequence of Perigonia lusca single nucleopolyhedrovirus: Insights into the evolution of a nucleotide metabolism enzyme in the family Baculoviridae. *Scientific Reports*, *6*. https://doi.org/10.1038/srep24612

Arrese, E. L., & Soulages, J. L. (2010). INSECT FAT BODY: ENERGY, METABOLISM, AND REGULATION. *Annu Rev Entomol*, *55*, 207–225. https://doi.org/10.1146/annurev-ento-112408-085356

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). Retrieved October 25, 2022, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boolchandani, M., Patel, S., & Dantas, G. (2017a). Functional metagenomics to study

antibiotic resistance. *Methods in Molecular Biology*, *1520*, 307–329. https://doi.org/10.1007/978-1-4939-6634-9_19/TABLES/1

Boolchandani, M., Patel, S., & Dantas, G. (2017b). Functional metagenomics to study antibiotic resistance. *Methods in Molecular Biology*, *1520*, 307–329. https://doi.org/10.1007/978-1-4939-6634-9_19/TABLES/1

Bowe, A., Onodera, T., Sadakane, K., & Shibuya, T. (2012). *Succinct de Bruijn Graphs* (B. Raphael & J. Tang, Eds.; pp. 225–235). Springer. https://doi.org/10.1007/978-3-642-33122-0_18

Bradley, P. H., & Pollard, K. S. (2017). Proteobacteria explain significant functional variability in the human gut microbiome. *Microbiome*, *5*(1), 1–23. https://doi.org/10.1186/S40168-017-0244-Z/TABLES/2

Brune, A. (2010). Methanogenesis in the Digestive Tracts of Insects. *Handbook of Hydrocarbon and Lipid Microbiology*, 707–728. https://doi.org/10.1007/978-3-540-77587-4_56

Brune, A. (2019). Methanogenesis in the Digestive Tracts of Insects and Other Arthropods. *Biogenesis of Hydrocarbons*, 229–260. https://doi.org/10.1007/978-3-319-78108-2_13

Cerqueira, N., Brás, N., Ramos, M. J., Fernandes, P. A., Cerqueira, N., Brás, N., Ramos, M. J., & Fernandes, P. A. (2012). Glycosidases – A Mechanistic Overview. *Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology*. https://doi.org/10.5772/52019

Chistoserdova, L. (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnology Letters*, *32*(10), 1351–1359. https://doi.org/10.1007/s10529-010-0306-9

da Lage, J.-L. (2018). The Amylases of Insects. *International Journal of Insect Science*, *10*, 117954331880478.

https://doi.org/10.1177/1179543318804783/ASSET/IMAGES/LARGE/10.1177_1179543318804783-FIG1.JPEG

Darzi, Y., Letunic, I., Bork, P., & Yamada, T. (2018). iPath3.0: interactive pathways explorer v3. *Nucleic Acids Research*, *46*(W1), W510–W513. https://doi.org/10.1093/nar/gky299

Doden, H. L., & Ridlon, J. M. (2021). Microbial hydroxysteroid dehydrogenases: From alpha to omega. In *Microorganisms* (Vol. 9, Issue 3, pp. 1–24). MDPI AG. https://doi.org/10.3390/microorganisms9030469

Escasa, S. R., Lauzon, H. A. M., Mathur, A. C., Krell, P. J., & Arif, B. M. (2006). Sequence analysis of the Choristoneura occidentalis granulovirus genome. *The Journal of General Virology*, *87*(Pt 7), 1917–1933. https://doi.org/10.1099/VIR.0.81792-0

Esko, J. D., Doering, T. L., & Raetz, C. R. (2009). Eubacteria and Archaea. *Principles and Applications of Soil Microbiology, Third Edition*, 111–148. https://doi.org/10.1016/B978-0-12-820202-9.00005-8

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., & Brugère, J. F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology : WJG*, *20*(43), 16062. https://doi.org/10.3748/WJG.V20.I43.16062

Galloway-Peña, J., & Hanson, B. (2020). Tools for Analysis of the Microbiome. In *Digestive Diseases and Sciences* (Vol. 65, Issue 3, pp. 674–685). Springer. https://doi.org/10.1007/s10620-020-06091-y

Gueli Alletti, G., Carstens, E. B., Weihrauch, B., & Jehle, J. A. (2018). Agrotis segetum nucleopolyhedrovirus but not Agrotis segetum granulovirus replicate in AiE1611T cell line of Agrotis ipsilon. *Journal of Invertebrate Pathology*, *151*, 7–13. https://doi.org/10.1016/J.JIP.2017.10.005

Guindo, C. O., Drancourt, M., & Grine, G. (2020). Digestive tract methanodrome: Physiological roles of human microbiota-associated methanogens. *Microbial*

Pathogenesis, *149*, 104425. https://doi.org/10.1016/J.MICPATH.2020.104425

Harvey, E., & Holmes, E. C. (2022). Diversity and evolution of the animal virome. *Nature Reviews Microbiology 2022 20:6*, *20*(6), 321–334. https://doi.org/10.1038/s41579-021-00665-x

Hirakata, S., Sakiyama, Y., Yoshimura, A., Ikeda, M., Takahata, K., Tashiro, Y., Yoshimura, M., Arata, H., Yonezawa, H., Kirishima, M., Higashi, M., Hatanaka, M., Kanekura, T., Yagita, K., Matsuura, E., & Takashima, H. (2021). The application of shotgun metagenomics to the diagnosis of granulomatous amoebic encephalitis due to Balamuthia mandrillaris: a case report. *BMC Neurology*, *21*(1), 1–6. https://doi.org/10.1186/S12883-021-02418-Y/FIGURES/3

Hoegenauer, C., Hammer, H. F., Mahnert, A., & Moissl-Eichinger, C. (2022). Methanogenic archaea in the human gastrointestinal tract. *Nature Reviews Gastroenterology & Hepatology 2022 19:12*, *19*(12), 805–813. https://doi.org/10.1038/s41575-022-00673-z

Horváthová, T., Šustr, V., Chroňáková, A., Semanová, S., Lang, K., Dietrich, C., Hubáček, T., Ardestani, M. M., Lara, A. C., Brune, A., & Šimek, M. (2021a). Methanogenesis in the Digestive Tracts of the Tropical Millipedes Archispirostreptus gigas (Diplopoda, Spirostreptidae) and Epibolus pulchripes (Diplopoda, Pachybolidae). *Applied and Environmental Microbiology*, *87*(15), 1–13. https://doi.org/10.1128/AEM.00614-21

Horváthová, T., Šustr, V., Chroňáková, A., Semanová, S., Lang, K., Dietrich, C., Hubáček, T., Ardestani, M. M., Lara, A. C., Brune, A., & Šimek, M. (2021b). Methanogenesis in the Digestive Tracts of the Tropical Millipedes Archispirostreptus gigas (Diplopoda, Spirostreptidae) and Epibolus pulchripes (Diplopoda, Pachybolidae). *Applied and Environmental Microbiology*, *87*(15), 1–13. https://doi.org/10.1128/AEM.00614-21

Houshyar, Y., Massimino, L., Lamparelli, L. A., Danese, S., & Ungaro, F. (2021a). Going Beyond Bacteria: Uncovering the Role of Archaeome and Mycobiome in Inflammatory

Bowel Disease. *Frontiers in Physiology*, *12*, 2169. https://doi.org/10.3389/FPHYS.2021.783295/BIBTEX

Houshyar, Y., Massimino, L., Lamparelli, L. A., Danese, S., & Ungaro, F. (2021b). Going Beyond Bacteria: Uncovering the Role of Archaeome and Mycobiome in Inflammatory Bowel Disease. *Frontiers in Physiology*, *12*. https://doi.org/10.3389/FPHYS.2021.783295

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, *17*(3), 377–386. https://doi.org/10.1101/gr.5969107

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Jing, T. Z., Qi, F. H., & Wang, Z. Y. (2020). Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome 2020 8:1*, *8*(1), 1–20. https://doi.org/10.1186/S40168-020-00823-Y

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A. L., Madsen, K. L., & Wong, G. K. S. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Frontiers in Microbiology*, *7*(APR). https://doi.org/10.3389/fmicb.2016.00459

Kim, J. Y., Whon, T. W., Lim, M. Y., Kim, Y. B., Kim, N., Kwon, M. S., Kim, J., Lee, S. H., Choi, H. J., Nam, I. H., Chung, W. H., Kim, J. H., Bae, J. W., Roh, S. W., & Nam, Y. do. (2020). The human gut archaeome: Identification of diverse haloarchaea in Korean subjects. *Microbiome*, *8*(1), 1–17. https://doi.org/10.1186/S40168-020-00894-X/FIGURES/6

Kirimura, J., Saito, M., & Kobayashi, M. (1962). Steroid Hormone in an Insect, Bombyx mori. *Nature 1962 195:4842*, *195*(4842), 729–730. https://doi.org/10.1038/195729b0

Klammsteiner, T., Walter, A., Bogataj, T., Heussler, C. D., Stres, B., Steiner, F. M., Schlick-

Steiner, B. C., Arthofer, W., & Insam, H. (2020). The Core Gut Microbiome of Black Soldier Fly (Hermetia illucens) Larvae Raised on Low-Bioburden Diets. *Frontiers in Microbiology*, *11*, 993. https://doi.org/10.3389/FMICB.2020.00993/BIBTEX

Krawczyk, A., Salamon, D., Kowalska-Duplaga, K., Bogiel, T., & Gosiewski, T. (2021). Association of Fungi and Archaea of the Gut Microbiota with Crohn's Disease in Pediatric Patients—Pilot Study. *Pathogens*, *10*(9). https://doi.org/10.3390/PATHOGENS10091119

Krishnan, M., Bharathiraja, C., Pandiarajan, J., Prasanna, V. A., Rajendhran, J., & Gunasekaran, P. (2014). Insect gut microbiome - An unexploited reserve for biotechnological application. *Asian Pacific Journal of Tropical Biomedicine*, *4*(Suppl 1), S16. https://doi.org/10.12980/APJTB.4.2014C95

Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D., & Charles, T. C. (2015). Current and future resources for functional metagenomics. *Frontiers in Microbiology*, *6*(OCT), 1196. https://doi.org/10.3389/FMICB.2015.01196/BIBTEX

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, W. J., & Hase, K. (2014). Gut microbiota–generated metabolites in animal health and disease. *Nature Chemical Biology 2014 10:6*, *10*(6), 416–424. https://doi.org/10.1038/nchembio.1535

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Ma, E., Zhu, Y., Liu, Z., Wei, T., Wang, P., & Cheng, G. (2021). Interaction of Viruses with the Insect Intestine. *Https://Doi.Org/10.1146/Annurev-Virology-091919-100543*, *8*, 115–131. https://doi.org/10.1146/ANNUREV-VIROLOGY-091919-100543

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://journal.embnet.org/index.php/embnetjournal/article/view/200/479

Mendes, L. W., Palma Perez Braga, L., Navarrete, A. A., Goss De Souza, D., Silva, G. G. Z., & Tsai, S. M. (n.d.). *Using Metagenomics to Connect Microbial Community Biodiversity and Functions*. https://doi.org/10.21775/cimb.024.103

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11257

Merzendorfer, H., & Zimoch, L. (2003). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology*, *206*(24), 4393–4412. https://doi.org/10.1242/JEB.00709

Moon, C. D., Young, W., Maclean, P. H., Cookson, A. L., & Bermingham, E. N. (2018). Metagenomic insights into the roles of Proteobacteria in the gastrointestinal microbiomes of healthy dogs and cats. *MicrobiologyOpen*, *7*(5). https://doi.org/10.1002/MBO3.677

Nakai, M., Harrison, R. L., Uchida, H., Ukuda, R., Hikihara, S., Ishii, K., & Kunimi, Y. (2015). Isolation of an adoxophyes orana granulovirus (AdorGV) occlusion body morphology mutant: Biological activity, genome sequence and relationship to other isolates of AdorGV. *Journal of General Virology*, *96*(4), 904–914. https://doi.org/10.1099/JGV.0.000023/CITE/REFWORKS

Neelakanta, G., & Sultana, H. (2013). The Use of Metagenomic Approaches to Analyze Changes in Microbial Communities. *Microbiology Insights*, *6*, 37.

https://doi.org/10.4137/MBI.S10819

Niwa, Y. S., & Niwa, R. (2016). Transcriptional regulation of insect steroid hormone biosynthesis and its role in controlling timing of molting and metamorphosis. *Development, Growth & Differentiation*, *58*(1), 94–105. https://doi.org/10.1111/DGD.12248

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. https://doi.org/10.1186/1471-2105-12-385

Oonincx, D. G. A. B., van Itterbeeck, J., Heetkamp, M. J. W., van den Brand, H., van Loon, J. J. A., & van Huis, A. (2010). An Exploration on Greenhouse Gas and Ammonia Production by Insect Species Suitable for Animal or Human Consumption. *PLOS ONE*, *5*(12), e14445. https://doi.org/10.1371/JOURNAL.PONE.0014445

Poles, M. Z., Juhász, L., & Boros, M. (2019). Methane and Inflammation - A Review (Fight Fire with Fire). *Intensive Care Medicine Experimental* , *7*(1), 1–10. https://doi.org/10.1186/S40635-019-0278-6/FIGURES/1

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., … Zoetendal, E. (2010). A human gut microbial gene catalog established by metagenomic sequencing. *Nature*, *464*(7285), 59. https://doi.org/10.1038/NATURE08821

Qi, X., Yun, C., Pang, Y., & Qiao, J. (2021). *The impact of the gut microbiota on the reproductive and metabolic endocrine system*. https://doi.org/10.1080/19490976.2021.1894070

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing

and web-based tools. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1219

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. In *Nature Biotechnology* (Vol. 35, Issue 9, pp. 833–844). Nature Publishing Group. https://doi.org/10.1038/nbt.3935

Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*, *14*(1), 1–12. https://doi.org/10.1093/bib/bbs007

Rizzatti, G., Lopetuso, L. R., Gibiino, G., Binda, C., & Gasbarrini, A. (2017). Proteobacteria: A Common Factor in Human Diseases. *BioMed Research International*, *2017*. https://doi.org/10.1155/2017/9351507

Rosen, G., Garbarine, E., Caseiro, D., Polikar, R., & Sokhansanj, B. (2008). Metagenome Fragment Classification Using N-Mer Frequency Profiles. *Advances in Bioinformatics*, 12. https://doi.org/10.1155/2008/205969

Salas, A. P., Zhu, L., Sánchez, C., Braña, A. F., Rohr, J., Méndez, C., & Salas, J. A. (2005). Deciphering the late steps in the biosynthesis of the anti-tumour indolocarbazole staurosporine: Sugar donor substrate flexibility of the StaG glycosyltransferase. *Molecular Microbiology*, *58*(1), 17–27. https://doi.org/10.1111/j.1365-2958.2005.04777.x

Satari, L., Guillén, A., Latorre-Pérez, A., & Porcar, M. (2021). Beyond Archaea: The Table Salt Bacteriome. *Frontiers in Microbiology*, *12*, 3157. https://doi.org/10.3389/FMICB.2021.714110/BIBTEX

Schuele, L., Lizarazo-Forero, E., Strutzberg-Minder, K., Schütze, S., Löbert, S., Lambrecht, C., Harlizius, J., Friedrich, A. W., Peter, S., Rossen, J. W. A., & Couto, N. (2022). Application of shotgun metagenomics sequencing and targeted sequence capture to detect circulating porcine viruses in the Dutch-German border region. *Transboundary and Emerging Diseases*, *69*(4), 2306–2319. https://doi.org/10.1111/TBED.14249

Seemann, T. (2014a). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seemann, T. (2014b). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shao, Y., Arias-Cordero, E. M., & Boland, W. (2013). Identification of Metabolically Active Bacteria in the Gut of the Generalist Spodoptera littoralis via DNA Stable Isotope Probing Using 13C-Glucose. *Journal of Visualized Experiments : JoVE*, *81*, 50734. https://doi.org/10.3791/50734

Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in Plant Science*, *5*(JUN). https://doi.org/10.3389/FPLS.2014.00209

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., Wang, W., Eden, J. S., Shen, J. J., Liu, L., Holmes, E. C., & Zhang, Y. Z. (2018). The evolutionary history of vertebrate RNA viruses. *Nature 2018 556:7700*, *556*(7700), 197–202. https://doi.org/10.1038/s41586-018-0012-7

Siddiqui, J. A., Khan, M. M., Bamisile, B. S., Hafeez, M., Qasim, M., Rasheed, M. T., Rasheed, M. A., Ahmad, S., Shahid, M. I., & Xu, Y. (2022). Role of Insect Gut Microbiota in Pesticide Degradation: A Review. *Frontiers in Microbiology*, *13*, 860. https://doi.org/10.3389/FMICB.2022.870462/BIBTEX

Srinivas, M. ;, O'sullivan, O. ;, Cotter, P. D. ;, Sinderen, D. V. ;, Kenny, J. G., Srinivas, M., O'sullivan, O., Cotter, P. D., van Sinderen, D., & Kenny, J. G. (2022). The Application of Metagenomics to Study Microbial Communities and Develop Desirable Traits in Fermented Foods. *Foods 2022, Vol. 11, Page 3297*, *11*(20), 3297. https://doi.org/10.3390/FOODS11203297

Tamaoki, T., Nomoto, H., Takahashi, I., Kato, Y., Morimoto, M., & Tomita, F. (1986). Staurosporine, a potent inhibitor of phospholipidCa++dependent protein kinase.

Biochemical and Biophysical Research Communications, *135*(2), 397–402. https://doi.org/10.1016/0006-291X(86)90008-2

Tetel, M. J., de Vries, G. J., Melcangi, R. C., Panzica, G., O'mahony, S. M., di Neuroscienze, D., & Montalcini, R. L. (n.d.). *Steroids, stress and the gut microbiome-brain axis*. https://doi.org/10.1111/jne.12548

Towle, H. C., Kaytor, E. N., & Shih, H. M. (2003). REGULATION OF THE EXPRESSION OF LIPOGENIC ENZYME GENES BY CARBOHYDRATE. *Https://Doi.Org/10.1146/Annurev.Nutr.17.1.405*, *17*, 405–433. https://doi.org/10.1146/ANNUREV.NUTR.17.1.405

Triantafyllou, K., Chang, C., & Pimentel, M. (2014). Methanogens, Methane and Gastrointestinal Motility. *Journal of Neurogastroenterology and Motility*, *20*(1), 31. https://doi.org/10.5056/JNM.2014.20.1.31

van de Pol, J. A. A., Best, N. van, Mbakwa, C. A., Thijs, C., Savelkoul, P. H., Ilja, I. C., Hornef, M. W., Mommers, M., & Penders, J. (2017). Gut colonization by methanogenic archaea is associated with organic dairy consumption in children. *Frontiers in Microbiology*, *8*(MAR), 355. https://doi.org/10.3389/FMICB.2017.00355/BIBTEX

Verma, M. K., Ahmed, V., Gupta, S., Kumar, J., Pandey, R., Mandhan, V., & Chauhan, N. S. (2018). Functional metagenomics identifies novel genes ABCTPP, TMSRP1 and TLSRP1 among human gut enterotypes. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-19862-5

Vijayvargiya, P., Jeraldo, P. R., Thoendel, M. J., Greenwood-Quaintance, K. E., Garrigos, Z. E., Rizwan Sohail, M., Chia, N., Pritt, B. S., & Patel, R. (2019). Application of metagenomic shotgun sequencing to detect vector-borne pathogens in clinical blood samples. *PLOS ONE*, *14*(10), e0222915. https://doi.org/10.1371/JOURNAL.PONE.0222915

Visconti, A., le Roy, C. I., Rosa, F., Rossi, N., Martin, T. C., Mohney, R. P., Li, W., de Rinaldis, E., Bell, J. T., Venter, J. C., Nelson, K. E., Spector, T. D., & Falchi, M. (2019). Interplay between the human gut microbiome and host metabolism. *Nature Communications 2019 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-12476-z

Wang, Y., Choi, J. Y., Roh, J. Y., Liu, Q., Tao, X. Y., Park, J. bin, Kim, J. S., & Je, Y. H. (2011). Genomic sequence analysis of granulovirus isolated from the Tobacco cutworm, spodoptera litura. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0028163

Wang, Y. T., Shen, R. X., Xing, D., Zhao, C. P., Gao, H. T., Wu, J. H., Zhang, N., Zhang, H. D., Chen, Y., Zhao, T. Y., & Li, C. X. (2021). Metagenome Sequencing Reveals the Midgut Microbiota Makeup of Culex pipiens quinquefasciatus and Its Possible Relationship With Insecticide Resistance. *Frontiers in Microbiology*, *12*, 228. https://doi.org/10.3389/FMICB.2021.625539/BIBTEX

Wood, D. E., Lu, J., & Langmead, B. (2019a). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wood, D. E., Lu, J., & Langmead, B. (2019b). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1891-0

Wood, D. E., & Salzberg, S. L. (2014). *Kraken: ultrafast metagenomic sequence classification using exact alignments*. http://ccb.jhu.edu/software/kraken/.

Wu, S., Zhu, Z., Fu, L., Niu, B., & Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, *12*(1), 444. https://doi.org/10.1186/1471-2164-12-444

Xia, L. C., Cram, J. A., Chen, T., Fuhrman, J. A., & Sun, F. (2011). Accurate genome relative abundance estimation based on shotgun metagenomic reads. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0027992

Xia, X., Gurr, G. M., Vasseur, L., Zheng, D., Zhong, H., Qin, B., Lin, J., Wang, Y., Song, F.,

Li, Y., Lin, H., & You, M. (2017a). Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Frontiers in Microbiology*, *8*, 663. https://doi.org/10.3389/FMICB.2017.00663/BIBTEX

Xia, X., Gurr, G. M., Vasseur, L., Zheng, D., Zhong, H., Qin, B., Lin, J., Wang, Y., Song, F., Li, Y., Lin, H., & You, M. (2017b). Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Frontiers in Microbiology*, *8*, 663. https://doi.org/10.3389/FMICB.2017.00663/BIBTEX

Ye, Y., & Doak, T. G. (2009). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLOS Computational Biology*, *5*(8), e1000465. https://doi.org/10.1371/journal.pcbi.1000465

Yun, J. H., Roh, S. W., Whon, T. W., Jung, M. J., Kim, M. S., Park, D. S., Yoon, C., Nam, Y. do, Kim, Y. J., Choi, J. H., Kim, J. Y., Shin, N. R., Kim, S. H., Lee, W. J., & Bae, J. W. (2014). Insect Gut Bacterial Diversity Determined by Environmental Habitat, Diet, Developmental Stage, and Phylogeny of Host. *Applied and Environmental Microbiology*, *80*(17), 5254. https://doi.org/10.1128/AEM.01226-14

Zheng, X., Zhu, Q., Zhou, Z., Wu, F., Chen, L., Cao, Q., & Shi, F. (2021). Gut bacterial communities across 12 Ensifera (Orthoptera) at different feeding habits and its prediction for the insect with contrasting feeding habits. *PLoS ONE*, *16*(4 April). https://doi.org/10.1371/journal.pone.0250675

Zibaee, A. (2012). Digestive enzymes of large cabbage white butterfly, Pieris brassicae L. (Lepidoptera: Pieridae) from developmental and site of activity perspectives. *Italian Journal of Zoology*, *79*(1), 13–26. https://doi.org/10.1080/11250003.2011.607190

Chapter 4

Gene expression profiling of *Antheraea assamensis* Helfer gut metagenome

CHAPTER 4

Gene expression profiling of *Antheraea assamensis* Helfer gut metagenome

Executive summary

Gene expression analysis is the process of determining which genes are actively being transcribed and translated into functional proteins in a particular cell or tissue sample. The analysis of gene expression can provide insight into a wide range of biological processes, including cell differentiation, development, and disease condition. . Metatranscriptomics is a rapidly growing field in molecular biology that involves the study of the RNA transcripts of all the microbial organisms in a particular environment. Unlike metagenomics, which focuses on the DNA sequences of an entire microbial community, metatranscriptomics provides a snapshot of the gene expression patterns of these microorganisms. The key objective of this chapter is to identify the metabolically active gut microbial community of *A. assamensis* and their gene expression profiling through mRNA sequencing. For the metatranscriptome analysis, total RNA from the gut of 5th instar larvae of *A. assamensis* was isolated. polyA-mRNA enriched libraries were prepared with the help of TruSeq RNA Library Preparation Kit and sequenced. Metatranscriptome annotation of the assembled transcripts resulted in the identification of ~7000 bacterial and 2708 archaeal active protein-coding genes. Taxonomic annotation of the assembled transcripts identified 1091 bacterial and 490 archaeal taxa at the genus level. In this analysis, several key metabolically active functions have been annotated to the gut microbiota which would expand our knowledge towards the understanding of the host-microbiome interaction of *A. assamensis* larvae.

4.1 Introduction

Gene expression analysis is the process of determining which genes are actively being transcribed and translated into functional proteins in a particular cell or tissue sample (Segundo-Val & Sanz-Lozano, 2016; st. Laurent et al., 2013). The analysis of gene expression can provide insight into a wide range of biological processes, including cell differentiation, development, and disease condition. By comparing the expression of genes between different cell types or under different conditions, researchers can identify genes that are associated with specific functions or physiological states (Costa-Silva et al., 2017; Nichols & Davenport, 2021; Rodriguez-Esteban & Jiang, 2017; st. Laurent et al., 2013). Metatranscriptomics is a rapidly growing field in molecular biology that involves the study of the RNA transcripts of all the microbial organisms in a particular environment. Unlike metagenomics, which focuses on the DNA sequences of an entire microbial community, metatranscriptomics provides a snapshot of the gene expression patterns of these microorganisms (Aguiar-Pulido et al., 2016). This information can be used to identify which genes are being actively transcribed and to infer the functional roles of these microorganisms in their environment (Shakya et al., 2019). One of the key advantages of metatranscriptomics is that it can provide insight into the activities of microorganisms without the need for culturing or isolation(Shakya et al., 2019). This is particularly useful for studying complex microbial communities in environmental samples, such as soil, water, or the gut microbiome. In these communities, there may be many different microbial species interacting with each other, and metatranscriptomics can reveal the functional roles of each species. The process of metatranscriptomics involves extracting RNA from a microbial community, sequencing the RNA, and then analyzing the resulting data (Reck et al., 2015). One of the challenges of metatranscriptomics is the large amount of data that is generated. This data must be carefully curated and analyzed using bioinformatics tools to identify which transcripts are present and to assign them to specific microorganisms (Pascault et al., 2015; Shakya et al.,

2019). Metatranscriptomics has already been used to study a wide range of microbial communities, including those in marine environments, gut of human and other animals, and soil (Abbà et al., 2022; Abu-Ali et al., 2018; Carvalhais et al., 2012; Johansson et al., 2013; Peters et al., 2019; Peterson & Scharf, 2016; Yang et al., 2022). In these studies, researchers have identified functional roles for previously unknown microorganisms and have shed light on the complex interactions between different species and with the host organism.

Objective 1 and objective 2 of our study have identified more than 1000 microbial species and the genes associated with the gut of *Antheraea assamensis* through metagenomic analysis (kindly refer to Chapter 2 and Chapter 3). The key objective of this chapter is to identify the metabolically active gut microbial community of *A. assamensis* and their gene expression profiling through mRNA sequencing. For the metatranscriptome analysis, total RNA from the gut of 5th instar larvae of *A. assamensis* was isolated. polyA-mRNA enriched libraries were prepared with the help of TruSeq RNA Library Preparation Kit and sequenced. Metatranscriptome annotation of the assembled transcripts resulted in the identification of ~7000 bacterial and 2708 archaeal active protein-coding genes. Taxonomic annotation of the assembled transcripts identified 1091 bacterial and 490 archaeal taxa at the genus level. In this analysis, several key metabolically active functions have been annotated to the gut microbiota which would expand our knowledge towards the understanding of the host-microbiome interaction of *A. assamensis* larvae.

4.2 Materials and methods

4.2.1 Sample collection

Healthy, fifth instar larvae of *Antheraea assamensis* reared on *Machilus bombycina* in outdoor conditions (70–80% relative humidity and temperature ~ 25 °C,) were collected irrespective of sex from Central Muga Eri Training and Research Institute (CMERTI), Jorhat, Assam. The larvae were immediately transported to the laboratory and flash frozen alive with liquid

nitrogen and stored at -80 °C in RNAlater until further processing.

4.2.2 Gut dissection and total RNA isolation

The gut from individual flash-frozen larvae was dissected out and total RNA was isolated by the standard Trizol method as described by Chetia et al. (Chetia et al., 2017; Thomas et al., 2019). The Trizol method is a widely used protocol for the isolation of high-quality RNA from various biological samples, including cells, tissues, and even virus-infected material (Vennapusa et al., 2020). A brief outline of the Trizol method is discussed below.

4.2.3 RNA isolation

The required chemicals for RNA isolation include TRIzol Reagent, PBS, Chloroform: isoamyl alcohol 24:1, DNase, absolute ethanol (molecular biology grade), UltraPure phenol:chloroform: isoamyl alcohol 24:24:1, nuclease-free pure water, sodium acetate (3M) at pH 5.5 and isopropyl alcohol.

The total RNA from insect gut was isolated following the standard TRIzol method (Rio et al., 2010). In brief, the freshly dissected gut tissues from 5-6 individual larvae were initially washed with PBS solution and homogenized using liquid nitrogen and the TRIzol reagent. Trizol facilitates the lysis of cells and dissolves the RNA. The homogenized tissue mixed with TRIzol was vortexed and incubated at room temperature for 5 minutes. Then, chloroform: isoamyl alcohol solution (4:1 ratio) was added to the sample so that the final proportion becomes four parts TRIzol and one part chloroform: isoamyl alcohol. The sample should turn murky once the chloroform-containing solution has been added. The reaction mixture was then vortexed for 15s, incubated at room temperature for 5 minutes and then centrifuged at 12000*g*, at 4 °C for 15 minutes. Then the aqueous top layer containing the RNA was transferred to a fresh tube. Cautions should be taken not to contaminate the upper aqueous layer with the interface layer which contains DNA. The RNA was precipitated by adding isopropyl alcohol so that the final volume becomes a 1:1 ratio of isopropyl alcohol: aqueous layer. It was then mixed by vortexing for 15 s, incubated at room temperature for 10 min and then centrifuged at 12000 *g*, for 15 minutes at 4 °C. The supernatant

was then discarded and the pellet was washed with 75% ethanol twice. The pellet was then air-dried and dissolved in 100 microlitres of ultrapure water. The isolated RNA was quantified using NanoDrop ™ spectrophotometer.

4.2.4 RNA sequencing

RNA with a quantity of more than 1 µg and A_{260}/A_{280} ratio ≥ 1.8 and RIN number ≥ 8 was considered for the downstream process. The cDNA libraries for paired-end sequencing were prepared from the fragmented mRNA and sequenced through the Illumina HiSeq™ 2000 sequencing platform at Genotypic Technology, Bangalore.

4.2.5 Quality control

Paired-end raw data with a read length of 100 bp were first checked for quality using FastQC (Andrews, 2010). The forward and reverse reads were merged using a fast and accurate paired-end read merging tool, PEAR (J. Zhang et al., 2014). This resulted in a single sequence file with paired-end reads merged at the overlapping region. It also produces two separate files for the forward and the reverse unmerged reads, respectively. However, for our downstream analysis, only the merged read file was considered. The low-quality base pairs and sequence ends were trimmed off using Trimmomatic V0.39 (Bolger et al., 2014) with the sliding windows of 4 bases to remove bases with an average quality score below 15. Cutadapt V 3.4 was used to remove the adapter sequences (Martin, 2011). The presence of the host sequences was detected and removed using bowtie2 (Langmead & Salzberg, 2012) and Samtools V1.9 (H. Li et al., 2009). The rRNA sequences were removed by mapping the processed reads against SILVA, GreenGenes and NCBI databases utilizing the program SortMeRNA (Kopylova et al., 2012) version 2.1, a robust rRNA filtering command line tool. SortMeRNA produces two output files, one is *file.ribosome* (file containing those reads identified by SortMeRNA as ribosomal sequences) and the other one is *file.ribodepleted*. The *file.ribodepleted* file was used

for the next step. The *file.ribosome* was used for taxonomic identification.

4.2.6 Metatranscriptome assembly and annotation

The merged, cleaned, adapter removed and rRNA sequence filtered reads were assembled using the *de novo* assembly tool into putative transcripts. Most microbes can't always be fully characterized using a reference genome. *De novo* assemblers offer a reference scaffold representing longer, expressed genome segments that can provide a reference set of transcripts that enables users to find homologs more easily and efficiently, create taxonomic origin, and also work as a reference for studying the expression profile (Shakya et al., 2019). The processed reads were assembled using Megahit v1.2.9 (Li et al., 2015) as described previously (Tláskal et al., 2021). The next step of assembly is to annotate the transcript to its corresponding match in a reference database. The major computational bottleneck of metatranscriptome assembly is that it generally contains tens of millions of reads which makes the usual annotation processes, such as BLASTX, too slow to perform and less sensitive. Here we have used DIAMOND (Buchfink et al., 2014), which is an open-source tool that uses a double indexing-based algorithm for protein alignment 20000 times faster than BLASTX. Normalization of transcript expression was done using the reads per kilobase of transcript per million mapped reads (FPKM) as described by He and his coworkers (He et al., 2019). Annotation was performed against NCBI RefSeq and SEED Subsystems databases. One of the most comprehensive databases for all purposes, RefSeq is known to have high-quality annotations. With SEED Subsystems, it is possible to organize certain functions into hierarchies, allowing related functions to be categorized under terms like "cellular respiration" or "protein biosynthesis." The annotated metatranscriptome files were then condensed and simplified following the Simple Annotation of Metatranscriptomes by Sequence Analysis 2 (SAMSA2) aggregation pipeline (Westreich et al., 2018). The aggregation step of SAMSA2 converts the large files returned from DIAMOND annotation as per the best match for each read in the reference

database. This step generates two files that contain the annotations of the active microbes clustered into distinct taxa and the second file contains the annotations clustered by function. Functions of specific organisms were identified using python scripts. Taxonomic annotation was done against the NCBI taxonomy database using kraken2 (Wood et al., 2019).

The whole pipeline for metatranscriptome assembly has been depicted in **Figure 4. 1**.

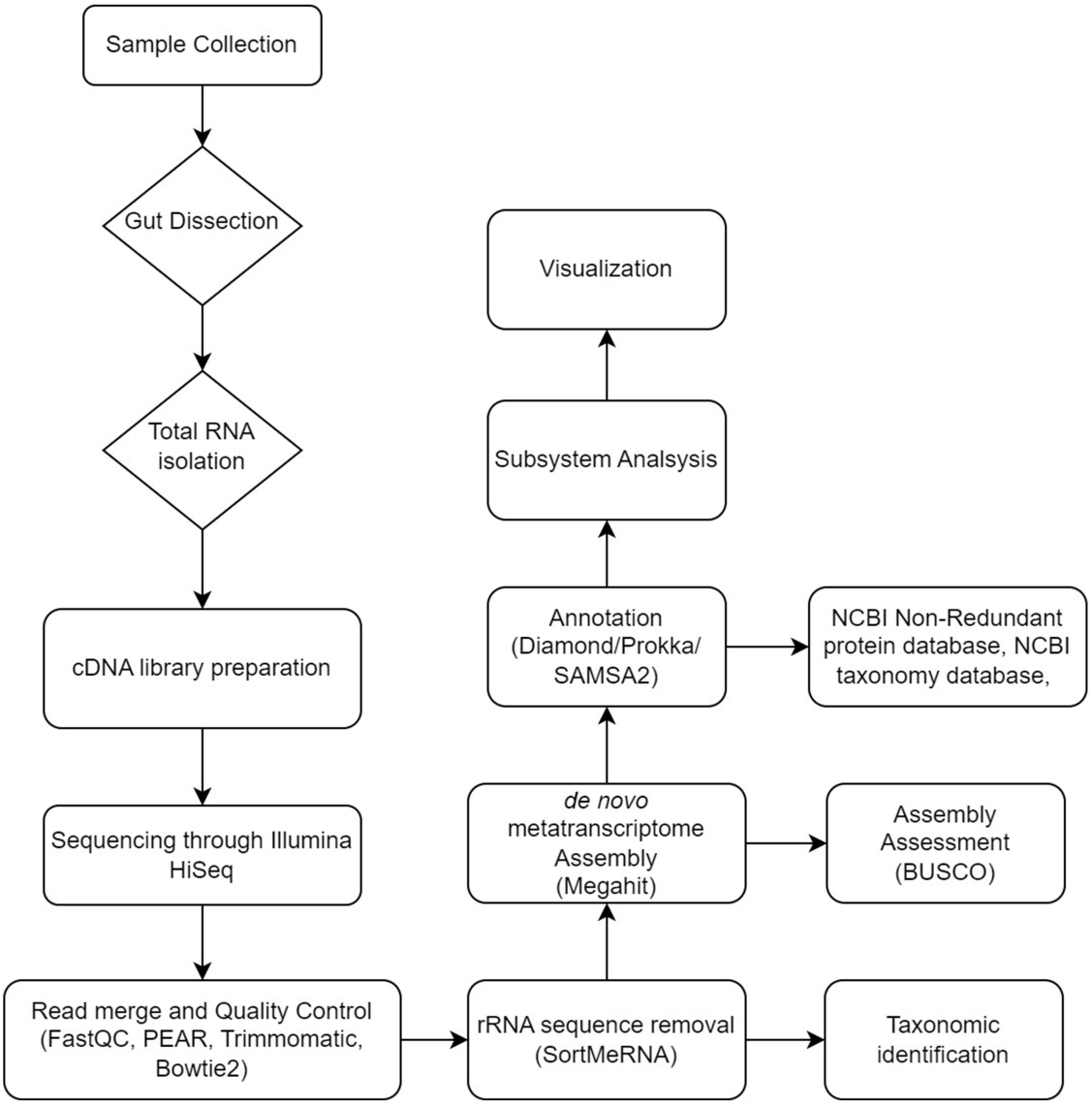


Figure 4. 1 The workflow for metatranscriptome data analysis

4.3 Results

4.3.1 Quality control of the raw data and metatranscriptome assembly

To explore the active microbial community and their gene expression inside the gut of healthy fifth instar larvae, we conducted metatranscriptome analysis through total RNA isolation of the larval gut followed by cDNA library construction. Sequencing was performed using Illumina HiSeq™ 2000 sequencing platform in paired-end mode. We received 7.7 Gb of raw transcriptome data containing 70.37 paired-end reads out of which 45.13 million reads passed the quality control step. *De novo* metatranscriptome assembly of 45137930 numbers of merged and quality filtered reads with Megahit v1.2.9 resulted in 31708 contigs. The details of the metatranscriptome assembly are represented in **Table 4.1.**

Table 4. 1 Details of the metatranscriptome assembly with Megahit.

Attribute	**Value**
Number of sequences passed the quality control step	45137930
Number of contigs generated	31708
k-mer length	21-99bp
Total number of bases	36091206 bp
Minimum contig length	200 bp
Maximum contig length	16979 bp
N50	1737 bp

4.3.2 Metatranscriptome annotation

The assembled transcripts of the gut metatranscriptome of *A. assamensis* were annotated against the NCBI non-redundant databases for bacteria and archaea using diamond v2.0.8.146 as described in SAMSA2 (Westreich et al., 2018) documentation. Annotation of the assembled transcripts against two custom databases created from NCBI RefSeq non-redundant protein sequences for bacteria and archaea. This resulted in the identification of ~7000 bacterial, and 2708 archaeal active protein-encoding genes in the metatranscriptome. Taxonomic annotation of the most active microbiota identified 1091 bacterial and 490 archaeal species with a relative

abundance of more than 0.015. The SEED Subsystem (Overbeek et al., 2014.) analysis resulted in three increasing levels of hierarchy for 2385 annotated functions which can be divided into 23 categories as described in **Table 4. 1**. A snapshot of the SEED Subsystem hierarchies of the annotated transcripts are depicted in **Figure 4. 2**.

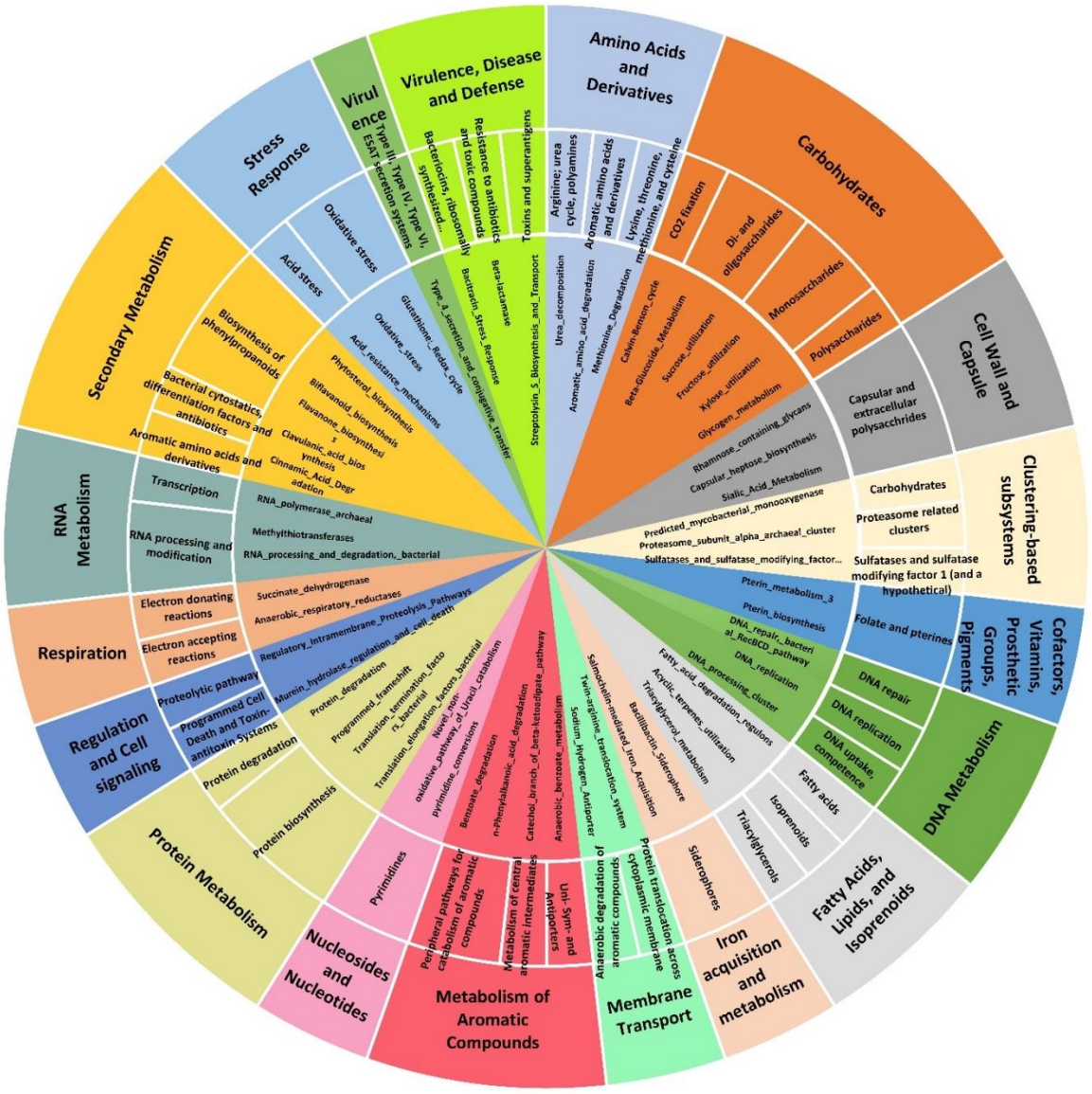


Figure 4. 2 SEED Subsystem hierarchy of the annotated transcripts. The inner circle (or level 3) represents the expressed genes in the gut metatranscriptome belonging to higher SEED hierarchy levels such as level 2 (the middle circle) and level 1 (the outer circle).

Table 4. 2 Functional categories and their abundance of the gut metatranscripts identified against SEED Subsystem Database.

Functional Category	Relative Abundance
Protein Metabolism	32.3297
Regulation and Cell signaling	11.0509
Respiration	9.6662
Carbohydrates	8.5566
Stress Response	5.1583
RNA Metabolism	4.1357
Amino Acids and Derivatives	2.9379
Cofactors, Vitamins, Prosthetic Groups, Pigments	2.0032
Fatty Acids, Lipids, and Isoprenoids	0.9766
Nucleosides and Nucleotides	0.9482
Cell Wall and Capsule	0.7868
Clustering-based subsystems	0.5878
DNA Metabolism	0.2597
Membrane Transport	0.2222
Virulence, Disease and Defense	0.1783
Metabolism of Aromatic Compounds	0.1300
Secondary Metabolism	0.0892
Miscellaneous	0.0336
Virulence	0.0300
Sulfur Metabolism	0.0170
Iron acquisition and metabolism	0.0075
Motility and Chemotaxis	0.0004
Phages, Prophages, Transposable elements, Plasmids	0.0002

4.3.3 Gene expression analysis of the metabolically active gut microbial community of *Antheraea assamensis* larvae

Gene expression analysis of the gut bacterial community resulted in the annotation of 7230 specific functions to the transcripts expressed by 1091 active bacterial taxa. In the gut metatranscriptome of *A. assamensis*, the highest expressed bacterial genes include molecular

chaperone DnaK, cytochrome c oxidase subunit I, cytochrome P450, ubiquitin, serine protease, translation elongation factor EF-1 subunit alpha, trypsin, NADP-dependent oxidoreductase, peptidase, alpha-amylase, aldehyde dehydrogenase, catalase, glutathione peroxidase and carboxylesterase. The unidentified proteins were annotated as hypothetical protein. The top 100 most abundantly expressed genes in the gut metatranscriptome of *A. assamensis* have been listed in **Table 4. 3**.

SEED Subsystem analysis of the annotated functions identified the active gut microbial genes responsible for different functions and categorized them into different SEED families such as protein metabolism; carbohydrate metabolism; cell wall and capsule protein; cofactors, vitamins, prosthetic groups, pigments metabolism, DNA metabolism; fatty acid lipid and isoprenoid metabolism; iron acquisition and metabolism; membrane transport; metabolism of aromatic compounds; nucleic acid metabolism; respiration; cell cycle regulation and cell signalling; secondary metabolism; stress response and virulence disease and defense. Some important SEED families and their transcripts identified in the gut metatranscript have been discussed below.

4.3.4 Expression analysis of transcripts for carbohydrate metabolism in the gut bacterial community of *A. assamensis*

The gut bacterial transcripts of *A. assamensis* larvae responsible for carbohydrate metabolism include carbohydrate esterase family 4, aldehyde dehydrogenase, putative esterase, alpha and beta-glucosidase, trehalose synthase, chitinase, alpha-amylase and others. The abundance of different enzyme classes responsible for carbohydrate metabolism were represented in **Figure 4. 3**. As described in section 2.4.4, enzymes responsible for host leaf digestion such as alpha and beta amylase, beta-glucosidase, chitinase and beta mannosidase were found to be highly expressed by the gut bacterial community. However, we could not detect xylanase, pectinase or rhamonosidase within the larval gut metatranscriptome. Methylglyoxal is well known as a toxic-byproduct of glycolysis and other metabolic pathways found in many organisms

including prokaryotes, invertebrates and mammals (Lee & Park, 2017; Wyllie & Fairlamb, 2011). Methylglyoxal found in honey bees was previously reported to show antibacterial properties (Mavric et al., 2008). Methylglyoxal metabolism has been found to be a highly active metabolism within the gut metatranscriptome in our study. Two enzymes involved in methylglyoxal metabolism, viz., aldehyde dehydrogenase and methylglyoxal synthase were expressed by the gut bacteria. Further details of the major active genes involved in carbohydrate metabolism have been listed in **Table 4. 4**.

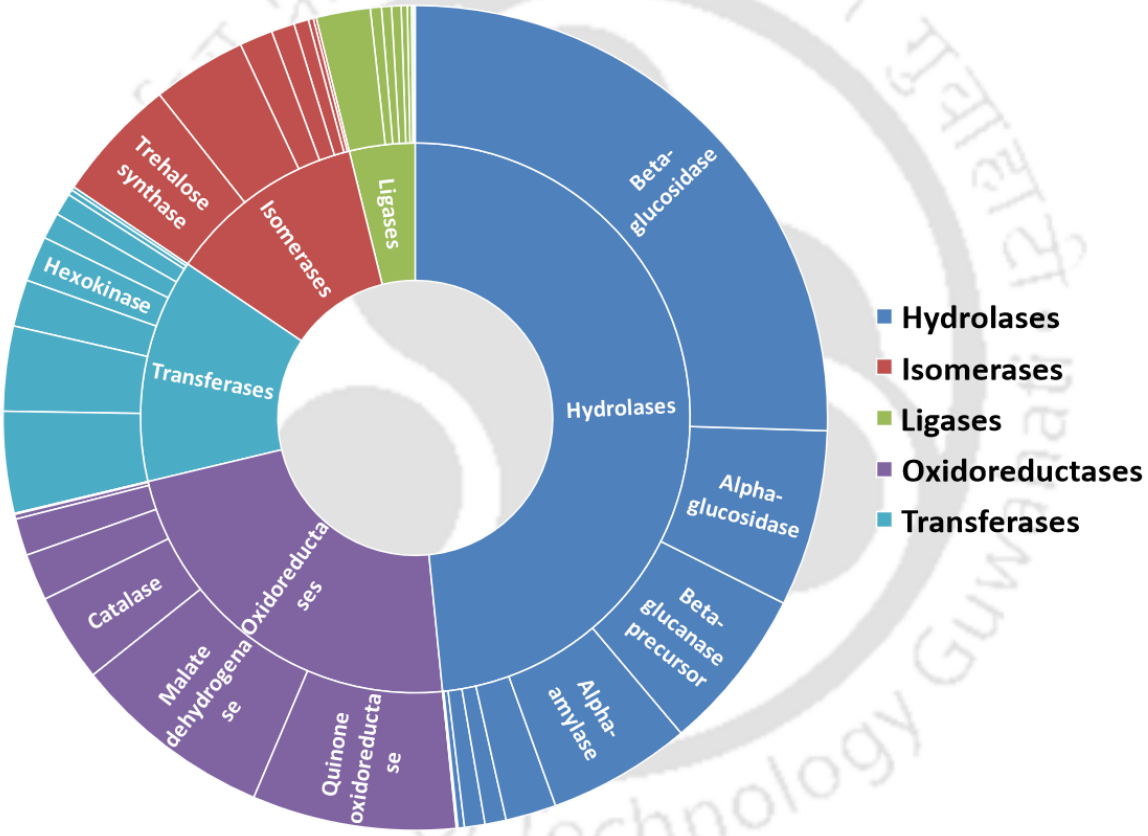


Figure 4. 3 The enzyme classes responsible for carbohydrate metabolism and their abundance. The class hydrolases were found to be the highest expressed followed by oxidoreductases, transferases, isomerases and ligases.

Table 4. 3 Highly expressed microbial genes identified in the gut metatranscriptome of *A. assamensis.*

Function	Relative Abundance	Function	Relative Abundance
Hypothetical protein	16.05	Glycosidase	0.30
Molecular chaperone DnaK	8.06	ATP synthase subunit A	0.29
Cytochrome c oxidase subunit I	3.44	Alpha-glucosidase	0.29
Cytochrome P450	3.15	ATPase AAA	0.29
Ubiquitin	2.81	RNA-binding protein	0.29
Serine protease	2.45	Phosphoglycerate dehydrogenase	0.29
Translation elongation factor EF-1 subunit alpha	2.19	MULTISPECIES: cytochrome P450	0.28
Trypsin	1.96	Aminopeptidase N	0.28
NADP-dependent oxidoreductase	1.94	Peptidase C14	0.28
Translation elongation factor EF-1 subunit alpha	1.24	Serine/threonine protein kinase	0.25
Peptidase	1.10	Phosphogluconate dehydrogenase NADP (+)-dependent, decarboxylating)	0.24
Alpha-amylase	0.88	Autotransporter domain-containing protein	0.24
Peptidase S1	0.85	Cell division protein FtsH	0.24
Molecular chaperone DnaK, partial	0.81	UV excision repair protein Rad23	0.23
Cytochrome c oxidase subunit 3	0.80	Molecular chaperone HtpG	0.23

IPTL-CTERM sorting domain-containing protein	0.77	Secreted protein containing Ubiquitin	0.23
Peptidylprolyl isomerase	0.76	Short-chain dehydrogenase	0.22
Methylmalonate-semialdehyde dehydrogenase (CoA acylating)	0.74	Aminopeptidase	0.21
Aldehyde dehydrogenase family protein	0.67	Malate dehydrogenase	0.21
Carboxylesterase	0.66	NADH-quinone oxidoreductase subunit H	0.21
Carboxylesterase/lipase family protein	0.65	Cytochrome b	0.21
Sodium-dependent transporter	0.59	ATP synthase subunit alpha	0.20
MULTISPECIES: molecular chaperone DnaK	0.57	NAD(P)-dependent oxidoreductase	0.20
MULTISPECIES: serine protease	0.54	AAA family ATPase	0.19
ATP-dependent helicase	0.52	Type I glyceraldehyde-3-phosphate dehydrogenase	0.19
Cytochrome c oxidase subunit II	0.48	ATP synthase subunit beta	0.18
Betaine-aldehyde dehydrogenase	0.48	DEAD/DEAH box helicase	0.18
Alpha-amlyase	0.48	Adenosylhomocysteinase	0.17
MULTISPECIES: hypothetical protein	0.48	F0F1 ATP synthase subunit alpha	0.17
Beta-glucosidase	0.47	Glu/Leu/Phe/Val dehydrogenase	0.16
Nucleotide sugar dehydrogenase	0.46	Sulfate adenylyltransferase	0.16
MULTISPECIES: cytochrome c oxidase subunit I	0.43	Elongation factor G	0.16
ABC transporter ATP-binding protein	0.43	Molecular chaperone DnaJ	0.16

Nucleoside-diphosphate kinase	0.42	Elongation factor 4	0.16
ATPase	0.42	Histidine triad nucleotide-binding protein	0.16
MULTISPECIES: trypsin	0.39	Cold-shock protein	0.15
Enoyl-CoA hydratase	0.38	RNA helicase	0.15
Beta-glucuronidase	0.36	Fructose-bisphosphate aldolase class I	0.15
Thioredoxin	0.34	3-phosphoserine/phosphohydroxythreonine aminotransferase	0.14
Actin, cytoplasmic 2	0.34	Elongation factor Tu	0.14
Acyl-CoA dehydrogenase	0.34	Long-chain-acyl-CoA synthetase	0.14
Glutamine--fructose-6-phosphate aminotransferase	0.33	T9SS C-terminal target domain-containing protein	0.14
Carboxylesterase family protein	0.31	Alkaline ceramidase	0.13
Adenylyl-sulfate kinase	0.31	Aconitate hydratase	0.10
Phosphoserine transaminase	0.31	Peroxiredoxin	0.09
Aldehyde dehydrogenase	0.31	Glucohydrolase	0.07
V-type ATP synthase subunit B	0.31	Carnitine dehydratase	0.06
Glycosidase	0.29	Catalase	0.051
Alpha-glucosidase	0.28	Glutathione peroxidase	0.05
Phosphoglycerate dehydrogenase	0.2	Glutamate dehydrogenase	0.04

Table 4. 4 Key enzymes identified in the gut metatranscriptome of *A. assamensis*

Expressed gene	Function	Number of Transcripts annotated	Relative abundance
Carbohydrate esterase family 4	Predicted carbohydrate hydrolases	284932	26.07
Aldehyde dehydrogenase (EC 1.2.1.3)	Methylglyoxal metabolism	221099	20.23
Putative esterase	Alpha-amylase locus in *Streptococcoccus*	120551	11.03
Beta-glucosidase (EC 3.2.1.21)	Beta-glucoside metabolism	94558	8.65
Alpha-glucosidase (EC 3.2.1.20)	D-Galacturonate and D-Glucuronate Utilization	25660	2.35
Maltodextrin glucosidase (EC 3.2.1.20)	Maltose and Maltodextrin Utilization	22671	2.07
Branched-chain alpha-keto acid dehydrogenase, E1 component, beta subunit (EC 1.2.4.4)	Dehydrogenase complexes	20544	1.88
Trehalose synthase (EC 5.4.99.16)	Trehalose biosynthesis	18102	1.66
Phosphoenolpyruvate synthase (EC 2.7.9.2)	Pyruvate metabolism I: anaplerotic reactions, PEP	16807	1.54
Glycogen phosphorylase (EC 2.4.1.1)	Glycogen metabolism	14759	1.35
Trehalose-6-phosphate hydrolase (EC 3.2.1.93)	Trehalose Uptake and Utilization	14679	1.34
Beta-hexosaminidase (EC 3.2.1.52)	Chitin and N-acetylglucosamine utilization	14208	1.30
Triosephosphate isomerase (EC 5.3.1.1)	Glycolysis and Gluconeogenesis	13674	1.25

Transaldolase (EC 2.2.1.2)	Fructose utilization	11893	1.09
Methylmalonate-semialdehyde dehydrogenase [inositol] (EC 1.2.1.27)	Inositol catabolism	9678	0.89
Glycerol kinase (EC 2.7.1.30)	Glycerol and Glycerol-3-phosphate Uptake and Utilization	9315	0.85
Pyruvate kinase (EC 2.7.1.40)	Glycerate metabolism	6803	0.62
Hexokinase (EC 2.7.1.1)	Mannose metabolism	6605	0.60
Putative beta-glucuronidase	L-rhamnose utilization	6365	0.58
Hydroxypyruvate reductase (EC 1.1.1.81)	Glycolate, glyoxylate interconversions	6096	0.56
Family 13 glycosyl hydrolase, row 724	Sucrose utilization	5692	0.52
Phosphoglycerate mutase (EC 5.4.2.1)	Entner-doudoroff pathway	4845	0.44
Chitinase (EC 3.2.1.14)	Chitin and N-acetylglucosamine utilization	3082	0.28
3-hydroxybutyryl-coa dehydratase (EC 4.2.1.55)	Acetyl-coA fermentation to Butyrate	3042	0.28
Pyruvate dehydrogenase E1 component beta subunit (EC 1.2.4.1)	Dehydrogenase complexes	2899	0.27
2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2)	Dehydrogenase complexes	2854	0.26
2-deoxy-D-gluconate 3-dehydrogenase (EC 1.1.1.125)	D-Galacturonate and D-Glucuronate Utilization	2835	0.26
Phosphoenolpyruvate carboxykinase [GTP] (EC	Pyruvate metabolism I: anaplerotic	2577	0.24

4.1.1.32)	reactions, PEP		
Cytosol aminopeptidase pepA (EC 3.4.11.1)	Dehydrogenase complexes	2543	0.23
Dihydrolipoamide succinyltransferase component (E2) of 2-oxoglutarate dehydrogenase complex (EC 2.3.1.61)	Dehydrogenase complexes	2428	0.22
6-phospho-beta-glucosidase (EC 3.2.1.86)	Beta-glucoside metabolism	2368	0.22
Phosphoglycerate kinase (EC 2.7.2.3)	Glycolysis and Gluconeogenesis	2357	0.22
Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	Dehydrogenase complexes	2353	0.22
Acetyl-coenzyme A synthetase (EC 6.2.1.1)	Pyruvate metabolism II: acetyl-coA, acetogenesis from pyruvate	2331	0.21
L-fuconate dehydratase (EC 4.2.1.68)	L-fucose utilization temp	2310	0.21
UDP-glucose 4-epimerase (EC 5.1.3.2)	Lactose and Galactose Uptake and Utilization	2230	0.20
Sucrose-6-phosphate hydrolase (EC 3.2.1.26)	Sucrose utilization	2150	0.20
Aldose 1-epimerase (EC 5.1.3.3)	Lactose and Galactose Uptake and Utilization	2130	0.19
Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94)	Glycerol and Glycerol-3-phosphate Uptake and Utilization	2069	0.19
Glycogen phosphorylase (EC 2.4.1.1)	Maltose and Maltodextrin Utilization	2051	0.19
Family 1 glycosyl hydrolase, row 954	Unknown oligosaccharide utilization	1919	0.18

	Sde 1396		
Galactose-1-phosphate uridylyltransferase (EC 2.7.7.10)	Lactose and Galactose Uptake and Utilization	1854	0.17
Methylmalonate-semialdehyde dehydrogenase (EC 1.2.1.27)	Isobutyryl-coA to Propionyl-coa Module	1718	0.16
Malate synthase (EC 2.3.3.9)	Glyoxylate bypass	1589	0.15
2,5-Diketo-D-gluconate reductase A (EC 1.1.1.274)	D-gluconate and ketogluconates metabolism	1521	0.14
2-Hydroxy-3-oxopropionate reductase (EC 1.1.1.60)	Glycerate metabolism	1520	0.14
Galactokinase (EC 2.7.1.6)	Lactose and Galactose Uptake and Utilization	1454	0.13
Sorbitol-6-phosphate 2-dehydrogenase (EC 1.1.1.140)	D-Sorbitol(D-Glucitol) and L-Sorbose Utilization	1358	0.12
Aconitate hydratase (EC 4.2.1.3)	Serine-glyoxylate cycle	1355	0.12
Alpha-amylase (EC 3.2.1.1)	Trehalose biosynthesis	1326	0.12
Phosphoglycerate mutase (EC 5.4.2.1)	Glycolysis and Gluconeogenesis	1326	0.12
Hydroxypyruvate isomerase (EC 5.3.1.22)	Glycerate metabolism	1307	0.12
3-Hydroxyacyl-coA dehydrogenase (EC 1.1.1.35)	Acetyl-coA fermentation to Butyrate	1306	0.12
Pyruvate kinase (EC 2.7.1.40)	Pyruvate metabolism I: anaplerotic reactions, PEP	1304	0.12
2,3-Butanediol dehydrogenase, S-alcohol forming,	Acetoin, butanediol metabolism	1290	0.12

(S)-acetoin-specific (EC 1.1.1.76)			
D-xylose proton-symporter xyle	Xylose utilization	1227	0.11
Endo-1,4-beta-xylanase A precursor (EC 3.2.1.8)	Xylose utilization	1170	0.11
Beta-galactosidase (EC 3.2.1.23)	Lactose and Galactose Uptake and Utilization	1012	0.09
Enolase (EC 4.2.1.11)	Glycolysis and Gluconeogenesis, including Archaeal enzymes	1167	0.11

4.3.5 Identification of transcripts involved in biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments

In the gut metatranscriptome of *Antheraea assamensis* larvae, 88 enzymes involved in the biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments have been identified to be actively expressed which can be categorized into eight broad categories based on their function. These classes include biotin biosynthesis, coenzyme A biosynthesis, folate and pterins metabolism, lipoic acid metabolism, NAD and NADP cofactor biosynthesis, pyridoxin (vitamin B6) biosynthesis, ubiquinone biosynthesis and tetrapyrroles metabolism (**Figure 4. 4**). Among these classes, the highest number of transcripts were annotated as genes responsible for pyridoxine biosynthesis followed by folate and pterins metabolism is the highest expressed followed by tetrapyrrole metabolism and NAD and NADP cofactor biosynthesis. Details of the genes involved in cofactors, vitamins, prosthetic groups and pigment synthesis and metabolism are represented with their abundance in **Table 4. 5**.

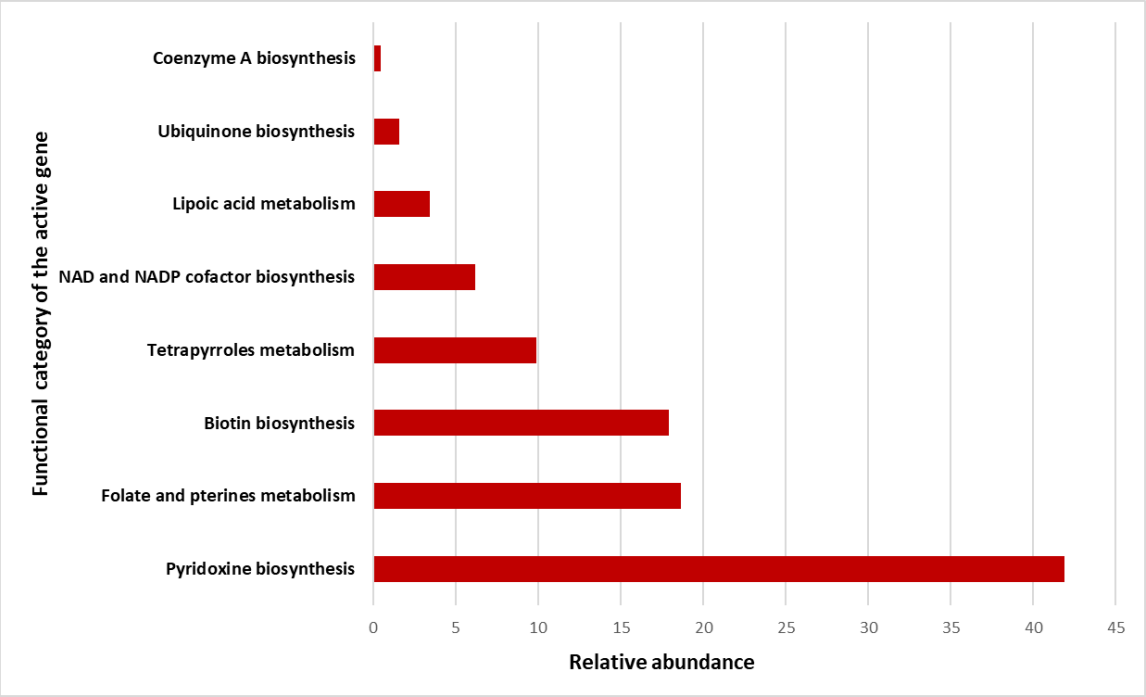


Figure 4. 4 SEED Subsystem annotation of active genes involved in the biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments at hierarchy level 1.

Table 4. 5 SEED Subsystem annotation of the transcripts involved in the biosynthesis and metabolism of cofactors, vitamins, prosthetic groups and pigments.

Pathway	Expressed Gene	Number of transcripts annotated
Biotin biosynthesis	Long-chain-fatty-acid--CoA ligase (EC 6.2.1.3)	70410
	Adenosylmethionine-8-amino-7-oxononanoate aminotransferase (EC 2.6.1.62)	1344
	8-amino-7-oxononanoate synthase (EC 2.3.1.47)	671
	3-ketoacyl-CoA thiolase (EC 2.3.1.16)	91
Coenzyme A Biosynthesis	Dephospho-CoA kinase (EC 2.7.1.24)	878
	Phosphopantothenoylcysteine decarboxylase (EC 4.1.1.36)	726
	2-dehydropantoate 2-reductase (EC 1.1.1.169)	58
	Pantothenate:Na+ symporter (TC 2.A.21.1.1)	26
Folate Biosynthesis	Transaldolase (EC 2.2.1.2)	4487
	Folylpolyglutamate synthase (EC 6.3.2.17)	1394
	Dihydrofolate reductase (EC 1.5.1.3)	73
Molybdenum cofactor biosynthesis	Molybdenum transport ATP-binding protein ModC (TC 3.A.1.8.1)	247
	GTP cyclohydrolase I (EC 3.5.4.16) type 1	19
Pterin biosynthesis	Phenylalanine-4-hydroxylase (EC 1.14.16.1)	598
	Aldo-keto reductase family 1 member B10 (EC 1.1.1.-)	332

	Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96)	265
	6-pyruvoyl tetrahydrobiopterin synthase (EC 4.2.3.12)	249
	Aldo-keto reductase family 1, member C3 (EC 1.1.1.-)	189
	Dihydropteridine reductase (EC 1.5.1.34)	98
	Sepiapterin reductase (EC 1.1.1.153)	70
Pterin metabolism	Acetoacetyl-CoA synthetase (EC 6.2.1.16)	9203
	Aromatic-amino-acid aminotransferase (EC 2.6.1.57)	5677
	Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96)	2914
	Homogentisate 1,2-dioxygenase (EC 1.13.11.5)	2201
	Fumarylacetoacetase (EC 3.7.1.2)	1532
	Maleylacetoacetate isomerase (EC 5.2.1.2)	744
	Phenylalanine-4-hydroxylase (EC 1.14.16.1)	695
	Nitric oxide synthase oxygenase (EC 1.-.-.-)	619
	Sepiapterin reductase (EC 1.1.1.153)	241
	6-pyruvoyl tetrahydrobiopterin synthase (EC 4.2.3.12)	157
	4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)	95
Lipoic acid metabolism	Lipoate-protein ligase A	13061
	Lipoate synthase	458
	Octanoate-[acyl-carrier-protein]-protein-N-octanoyltransferase	243
NAD and NADP cofactor	Kynurenine 3-monooxygenase (EC 1.14.13.9)	9743

biosynthesis global	NAD kinase (EC 2.7.1.23)	5114
	Nicotinate phosphoribosyltransferase (EC 2.4.2.11)	5054
	Niacin transporter NiaP	1696
	Tryptophan 2,3-dioxygenase (EC 1.13.11.11)	987
	Glutamine amidotransferase chain of NAD synthetase	936
	ADP-ribose pyrophosphatase (EC 3.6.1.13)	924
	N-Ribosylnicotinamide phosphorylase (EC 2.4.2.1)	403
	L-aspartate oxidase (EC 1.4.3.16)	106
	Nicotinamidase (EC 3.5.1.19)	54
Pyridoxin (Vitamin B6) Biosynthesis	Phosphoserine aminotransferase (EC 2.6.1.52)	96520
	D-3-phosphoglycerate dehydrogenase (EC 1.1.1.95)	68295
	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	2719
	Pyridoxamine 5'-phosphate oxidase (EC 1.4.3.5)	1223
	Pyridoxal kinase (EC 2.7.1.35)	621
Menaquinone and Phylloquinone Biosynthesis	O-succinylbenzoic acid--CoA ligase (EC 6.2.1.26)	743
	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase (EC 4.2.99.20)	400
Ubiquinone Biosynthesis	Ubiquinone/menaquinone biosynthesis methyltransferase UbiE (EC 2.1.1.-)	1787
	Ubiquinone biosynthesis monooxygenase UbiB	1768
	3-demethylubiquinone-9 3-methyltransferase (EC 2.1.1.64)	1468
	4-hydroxybenzoate polyprenyltransferase (EC 2.5.1.-)	281

Cobalamin synthesis	Cobalamin biosynthesis protein BluB	29
Coenzyme B12 biosynthesis	Vitamin B12 ABC transporter, ATPase component BtuD	1060
	Alpha-ribazole-5'-phosphate phosphatase (EC 3.1.3.73)	501
	Duplicated ATPase component CbrU of energizing module of predicted cobalamin ECF transporter	172
	5,6-dimethylbenzimidazole synthase, flavin destructase family	131
Heme and Siroheme Biosynthesis	Coproporphyrinogen III oxidase, aerobic (EC 1.3.3.3)	14701
	5-aminolevulinate synthase (EC 2.3.1.37)	11283
	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)	4161
	Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)	3261
	Porphobilinogen deaminase (EC 2.5.1.61)	2189
	Porphobilinogen synthase (EC 4.2.1.24)	1768
	Protoporphyrinogen IX oxidase, aerobic (EC 1.3.3.4)	53
Carotenoid Biosynthesis	Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29)	937
	Neurosporene desaturase (EC 1.-.-.-)	59
	Methoxyneurosporene dehydrogenase (EC 1.14.99.-)	39
	Phytoene synthase (EC 2.5.1.32)	28
	Beta-carotene ketolase (EC 1.14.-.-)	2

4.3.6 Identification of transcripts involved in stress response

The major active bacterial stress response genes in the gut metatranscriptome identified against the SEED Subsystem database include genes responsible for heat shock, oxidative stress, osmotic stress and detoxification in the decreasing order of their number of annotated transcripts (**Figure 4. 5**). Among the genes involved in heat shock response, chaperone protein DnaK was found to be the highest expressed. Osmotic stress response was found to be regulated by the synthesis of enzymes like Betaine aldehyde dehydrogenase (EC 1.2.1.8), Choline dehydrogenase (EC 1.1.99.1), Aquaporin Z and Choline-sulfatase (EC 3.1.6.6). Similarly, the oxidative stress response was regulated mainly by Alkyl hydroperoxide reductase subunit C-like protein, glutathione S-transferase-like protein, NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12), Hydroxyacylglutathione hydrolase (EC 3.1.2.6), Catalase (EC 1.11.1.6), Lactoylglutathione lyase (EC 4.4.1.5) and others as described in **Table 4. 6**.

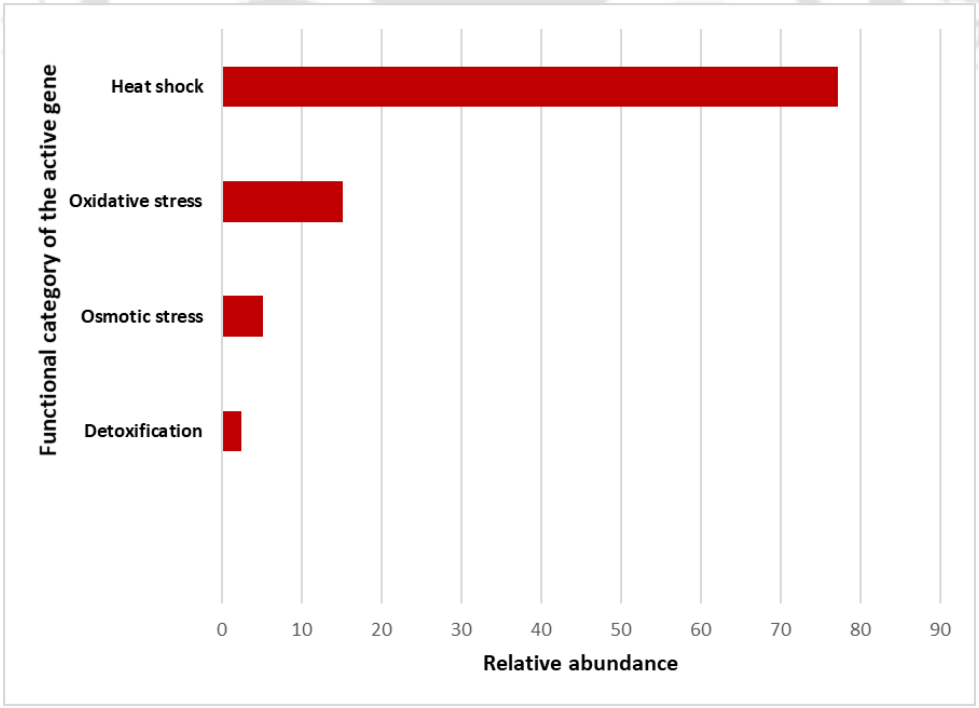


Figure 4. 5 The major functional categories of the expressed genes involved in stress response**.**

Table 4. 6 SEED Subsystem annotation of the larval gut metatranscripts involved in oxidative stress response

Gene	Relative Abundance	Function
Uncharacterized monothiol glutaredoxin ycf64-like	2.66	Glutaredoxins
Glutaredoxin-related protein	0.73	Glutaredoxins
5-oxoprolinase (EC 3.5.2.9)	4.23	Glutathione: Biosynthesis and gamma-glutamyl cycle
Gamma-glutamyltranspeptidase (EC 2.3.2.2)	3.06	Glutathione: Biosynthesis and gamma-glutamyl cycle
Uncharacterized glutathione S-transferase-like protein	10.82	Glutathione: Non-redox reactions
Hydroxyacylglutathione hydrolase (EC 3.1.2.6)	9.24	Glutathione: Non-redox reactions
Lactoylglutathione lyase (EC 4.4.1.5)	7.97	Glutathione: Non-redox reactions
Glutathione S-transferase, zeta (EC 2.5.1.18)	5.10	Glutathione: Non-redox reactions
Glutathione S-transferase (EC 2.5.1.18)	4.16	Glutathione: Non-redox reactions
Glutathione S-transferase, unnamed subgroup (EC 2.5.1.18)	1.24	Glutathione: Non-redox reactions
Glutathione S-transferase family protein	0.84	Glutathione: Non-redox reactions

Glutathione peroxidase (EC 1.11.1.9)	3.79	Glutathione: Redox cycle
Glutathione reductase (EC 1.8.1.7)	1.76	Glutathione: Redox cycle
Alkyl hydroperoxide reductase subunit C-like protein	16.70	Oxidative stress
Catalase (EC 1.11.1.6)	8.16	Oxidative stress
Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)	1.46	Oxidative stress
Superoxide dismutase [Mn] (EC 1.15.1.1)	1.04	Oxidative stress
NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)	10.31	Redox-dependent regulation of nucleus processes
NAD-dependent protein deacetylase of SIR2 family	2.43	Redox-dependent regulation of nucleus processes
Superoxide dismutase [Cu-Zn] (EC 1.15.1.1)	4.32	Regulation of Oxidative Stress Response

4.3.7 Identification of transcripts involved in the metabolism of aromatic compounds

The identified transcripts involved in the metabolism of aromatic compounds inside the gut of *A. assamensis* larvae can be categorized into three broad categories, namely, anaerobic degradation of aromatic compounds, metabolism of central aromatic intermediates and peripheral pathways for catabolism of aromatic compounds. The abundance of the major pathways is depicted in **Figure 4. 6**. The highest number of transcripts were annotated for n-Phenylalkanoic acid degradation pathway followed by salicylate and gentisate catabolism and benzoate catabolism. The details of the genes involved in each pathway and their abundances are shown in **Table 4. 7**.

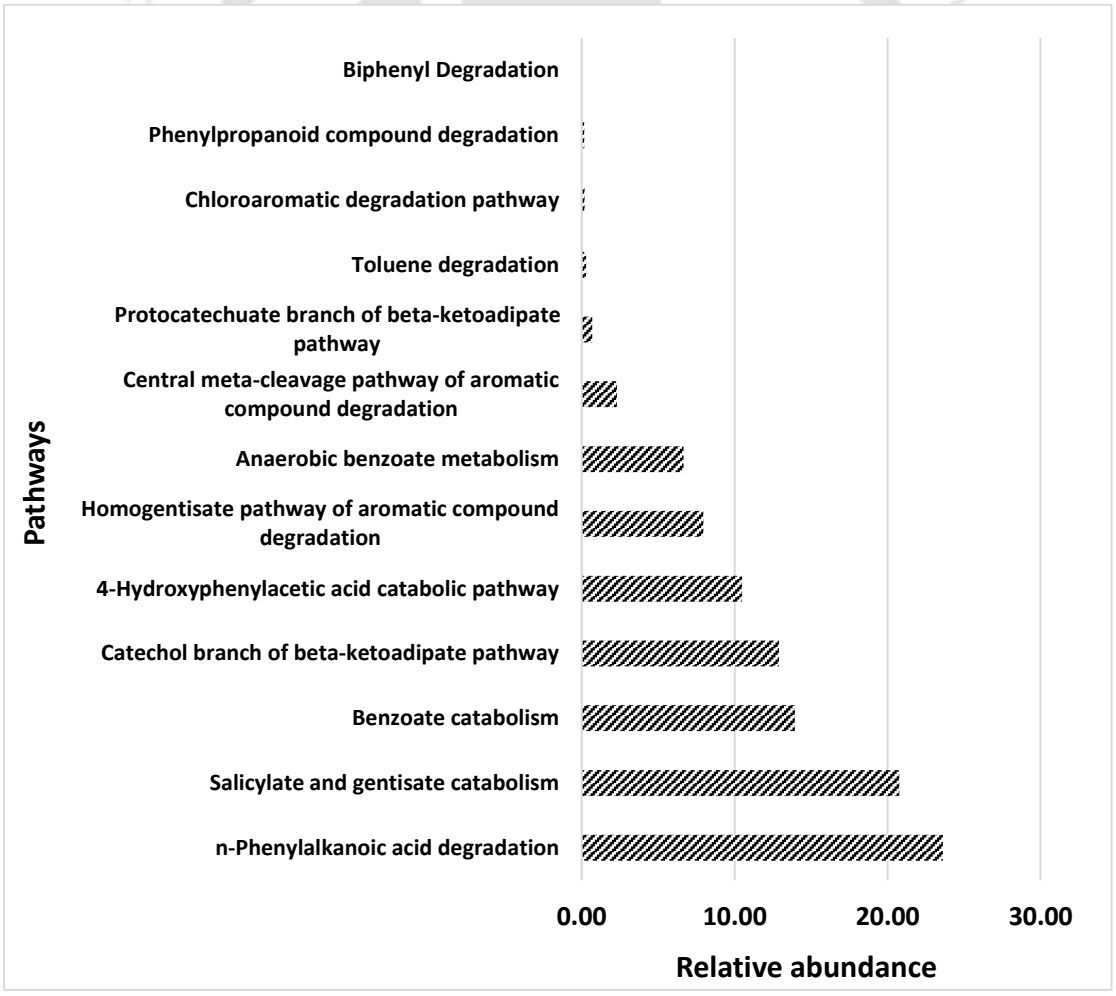


Figure 4. 6 The major active pathways involved in the metabolism of aromatic compounds inside the gut of *A. assamensis* larvae.

Table 4. 7 List of expressed genes and the pathways involved in the metabolism of aromatic compounds inside the larval gut

Category	Pathway	Expressed gene	No. Of transcripts annotated	Relative abundance
Anaerobic degradation of aromatic compounds	Anaerobic benzoate metabolism	Glutaryl-CoA dehydrogenase (EC 1.3.99.7)	1727	6.40
		Cyclohex-1-ene-1-carboxyl-CoA hydratase (EC 4.2.1.17)	59	0.22
		2-hydroxycyclohexanecarboxyl-CoA dehydrogenase (EC 1.1.1.-)	10	0.04
		3-hydroxybutyryl-CoA dehydrogenase (EC 1.1.1.157)	2	0.01
Metabolism of central aromatic intermediates	4-Hydroxyphenylacetic acid catabolic pathway	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60)	2221	8.23
		5-carboxymethyl-2-oxo-hex-3- ene-1,7-dioate decarboxylase (EC 4.1.1.68)	421	1.56
		5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)	188	0.70
	Catechol branch of beta-ketoadipate pathway	Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)	2015	7.47
		Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)	1065	3.95
		3-oxoadipate CoA-transferase subunit B (EC 2.8.3.6)	232	0.86
		Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)	80	0.30
		3-oxoadipate CoA-transferase subunit A (EC 2.8.3.6)	80	0.30
		mandelate racemase/muconate lactonizing enzyme family protein	9	0.03
	Central meta-cleavage pathway of aromatic compound degradation	2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-)	290	1.08
		5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC	183	0.68

		5.3.3.10)		
		5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60)	150	0.56
	Homogentisate pathway of aromatic compound degradation	Homogentisate 1,2-dioxygenase (EC 1.13.11.5)	1087	4.03
		Aromatic-amino-acid aminotransferase (EC 2.6.1.57)	1044	3.87
		Transcriptional regulator, IclR family	13	0.05
	Protocatechuate branch of beta-ketoadipate pathway	Beta-ketoadipyl CoA thiolase (EC 2.3.1.-)	112	0.42
		Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)	61	0.23
		Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)	12	0.04
	Salicylate and gentisate catabolism	Fumarylacetoacetate hydrolase family protein	3089	11.45
		Fumarylacetoacetase (EC 3.7.1.2)	1691	6.27
		4-hydroxybenzoate transporter	466	1.73
		Maleylacetoacetate isomerase (EC 5.2.1.2)	328	1.22
		salicylate esterase	25	0.09
		AreA (Esterase for aryl ester catabolic pathway) (EC 3.1.1.-)	2	0.01
Peripheral pathways for catabolism of aromatic compounds	Benzoate catabolism	Acetyl-CoA C-acyltransferase (EC 2.3.1.16)	16	0.06
		1,2-dihydroxycyclohexa-3,5-diene-1-carboxylate dehydrogenase (EC 1.3.1.25)	2589	9.60
		benzoate MFS transporter BenK	1141	4.23
		Benzoate 1,2-dioxygenase beta subunit (EC 1.14.12.10)	16	0.06
		2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (EC 3.7.1.-)	4	0.01

4.3.8 Identification of the active gut bacterial community in the gut of *A. assamensis* larvae and their function

Total of 1091 bacterial taxa have been identified to be metabolically active inside the gut of the 5th instar larvae. The most active gut bacterium of *A. assamensis* larvae identified through metatranscriptome analysis was *Acinetobacter baumannii* from the family Moraxellaceae with the highest number of transcripts annotated to it. The second most active taxon was *Streptomyces* followed by *Ruminococcus albus*, *Pseudomonas syringae*, *Paenibacillus*, *Pedobacter*, *Rhizobium*, *Sphingomonas* and others as depicted in **Figure 4. 7**. To find out the functional activity or the genes expressed by the active bacteria, we performed functional annotation of the selective organisms using a python script "DIAMOND_specific_organism_retreiver.py". this program requires three input files, viz., the Diamond annotation result against RefSeq, the name of the specific organism and the RefSeq database file. This analysis produces an output file containing each transcript originating from the selected organism. The key proteins expressed by *Acinetobacter baumannii* in the gut of *A. assamensis* larvae include molecular chaperone DnaK, sodium-dependent transporter, methylmalonate-semialdehyde dehydrogenase, peptidylprolyl isomerase, Zn-dependent hydrolase, sulfate adenylyltransferase, alpha/beta hydrolase, aldehyde dehydrogenase, glutathione peroxidase, hydroxyacylglutathione hydrolase, catalases and others (**Figure 4. 8**). *Lactobacillus*, a well-known probiotic species identified in most animal species expresses genes like molecular chaperone DnaK, peptidase, aminopeptidase N, beta-glucuronidase, alpha-glucosidase, methionine adenosyltransferase, sulfate adenylyltransferase, glucohydrolase, peroxiredoxin, catalase and alpha-xylosidase (**Figure 4. 9**). The genus Serratia was found to be present in the larval gut of *A. assamensis* from shotgun metagenomic sequencing. The metatranscriptome analysis has revealed that this microbe is highly active metabolically inside the gut of the 5th instar larva (**Appendix I**).

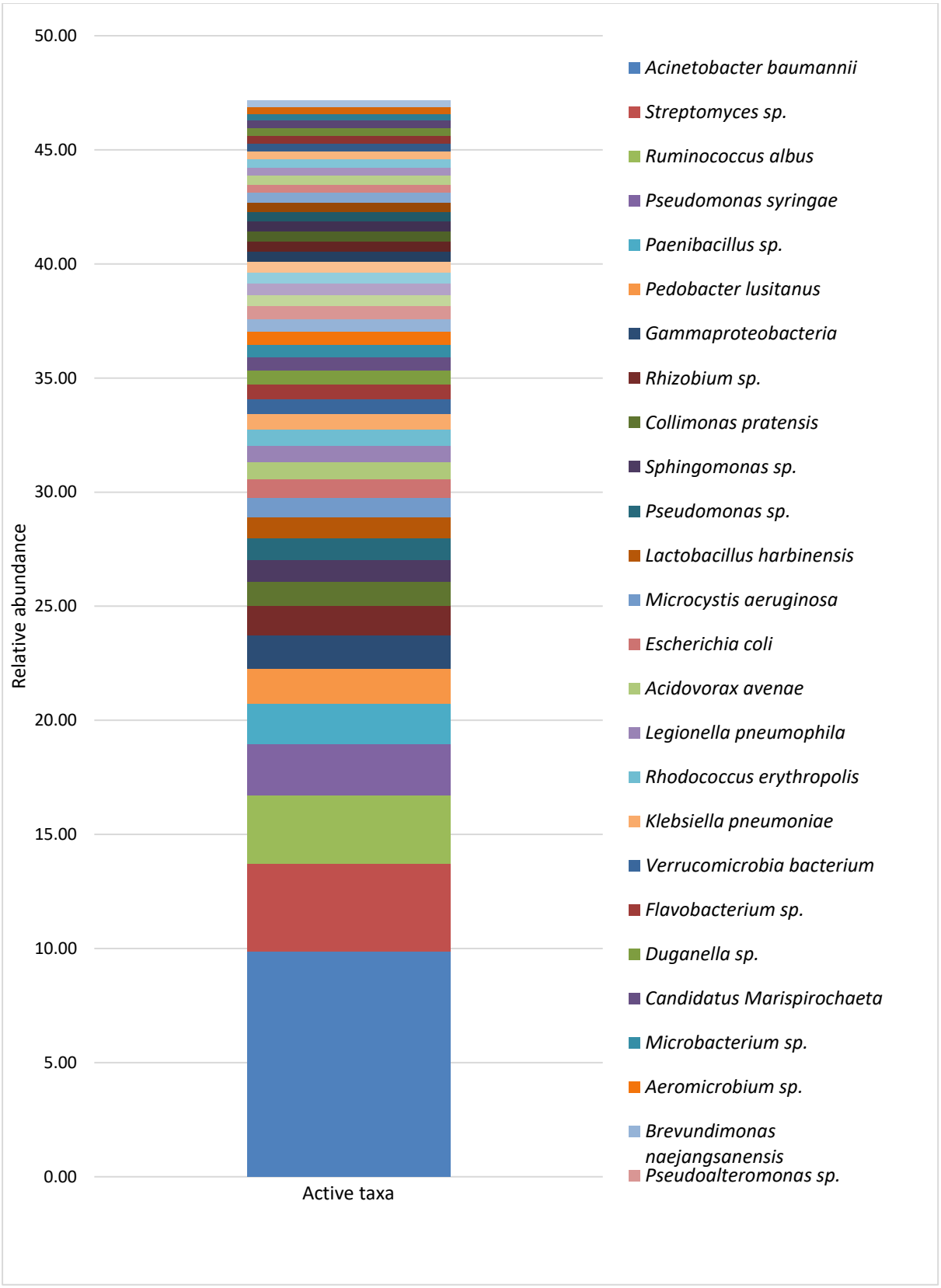


Figure 4. 7 The active gut microbial community of *A. assamensis and* their relative abundances

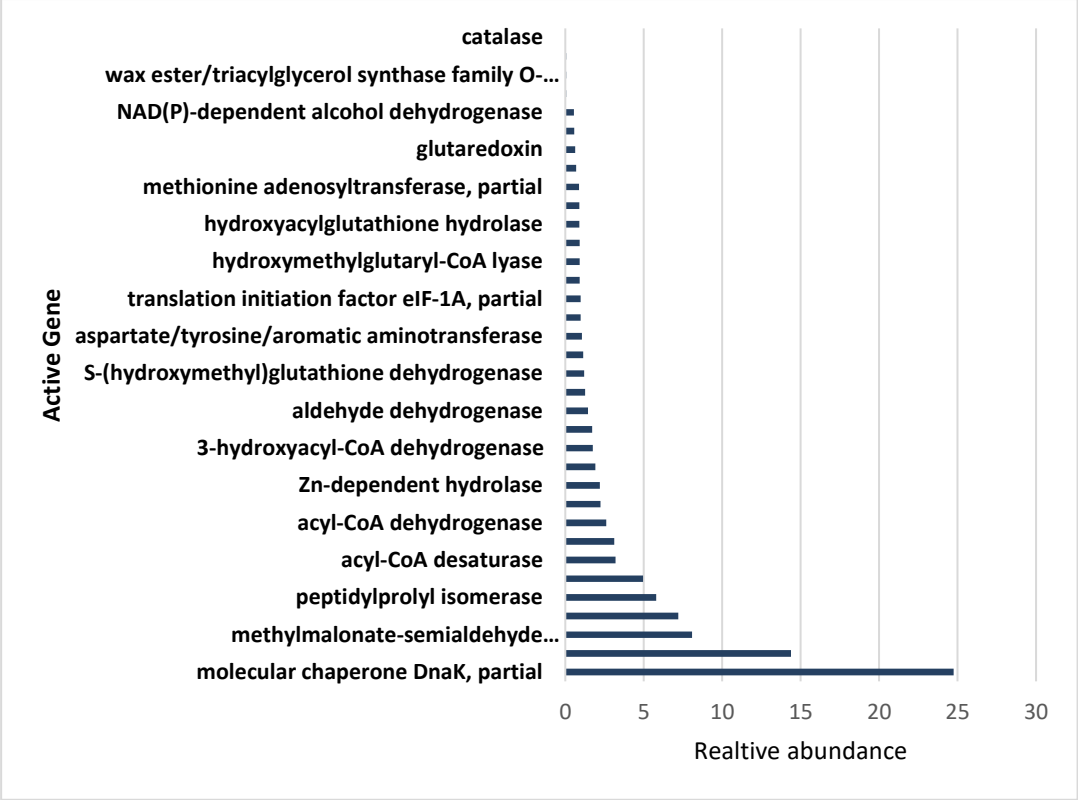


Figure 4. 8 The major genes expressed by *Acinetobacter baumannii* inside the gut of *A. assamensis* larvae.

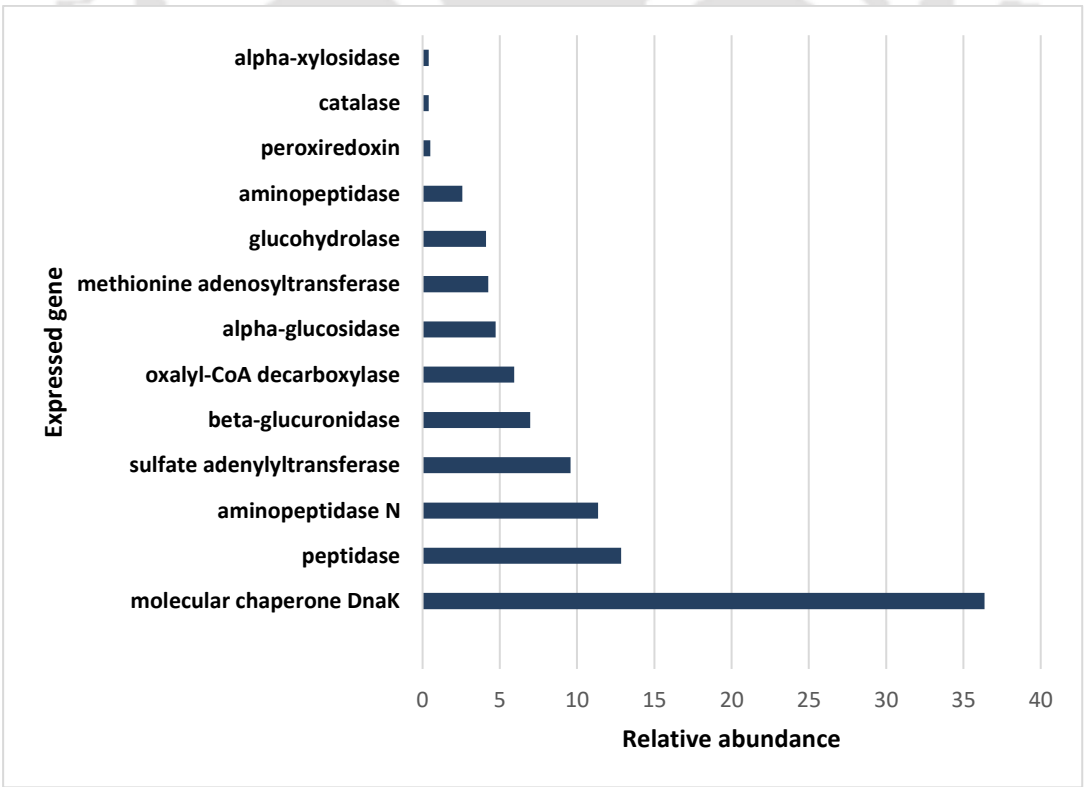


Figure 4. 9 The major genes expressed by *Lactobacillus* sp. inside the gut of *A. assamensis* larvae.

The annotated genes found to be expressed by *Serratia* include carboxylesterase family protein, chitinase, aromatic aminotransferase, N-methyltryptophan oxidase, carboxylesterase, alkaline phosphatase, aquaporin, carbonate dehydratase, methylmalonate-semialdehyde dehydrogenase, glutathione peroxidase, beta-glucosidase, and catalase (**Figure 4. 10**). The major function of other selected bacterial taxa has been given in **Appendix** I.

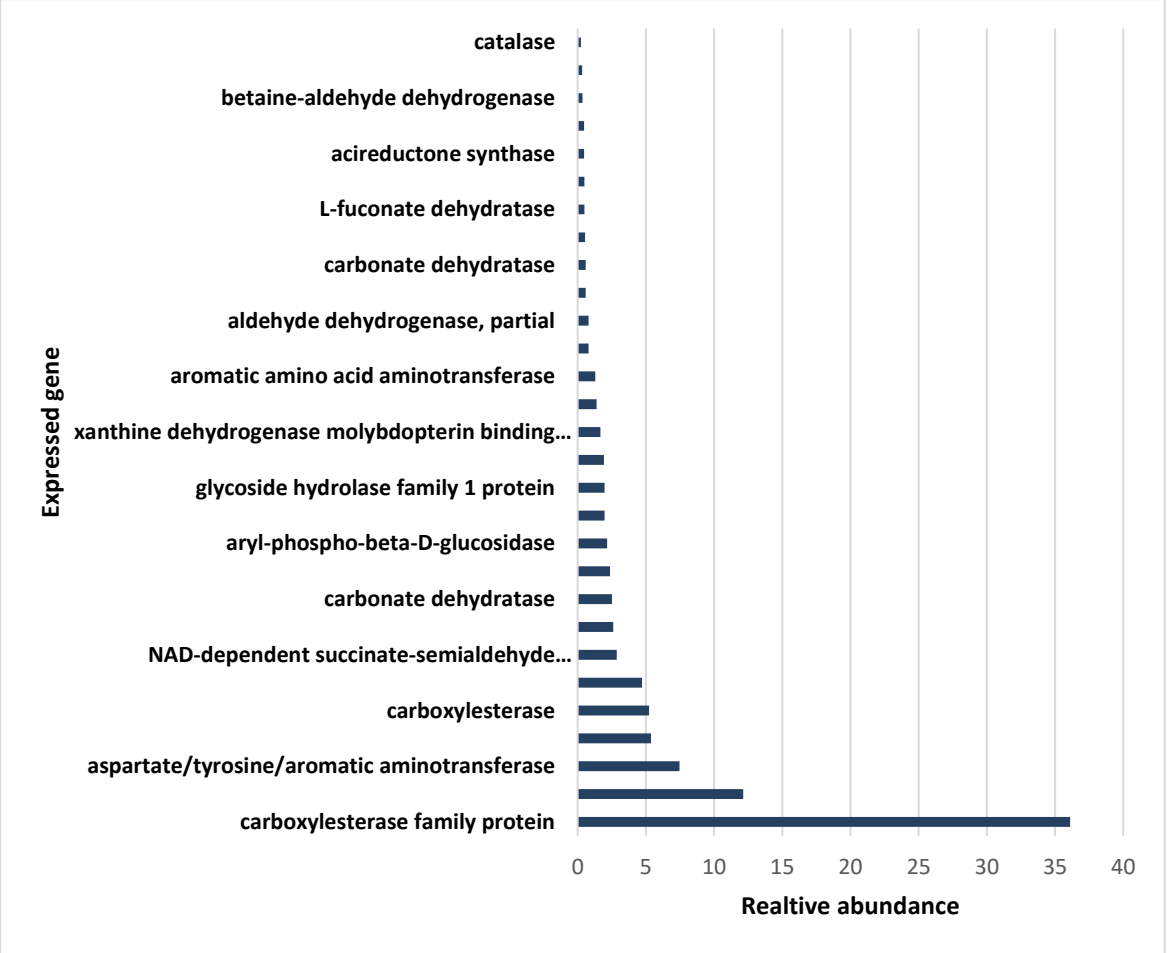


Figure 4. 10 The major genes expressed by *Serratia* sp. inside the gut of *A. assamensis* larvae.

4.3.9 The active gut archaeal taxa in the gut of *A. assamensis* larvae and their function

In the gut metatranscriptome of of *A. assamensis* larvae, 490 metabolically active archaeal taxa have been identified. Among these, the highest number of archaeal transcripts were annotated

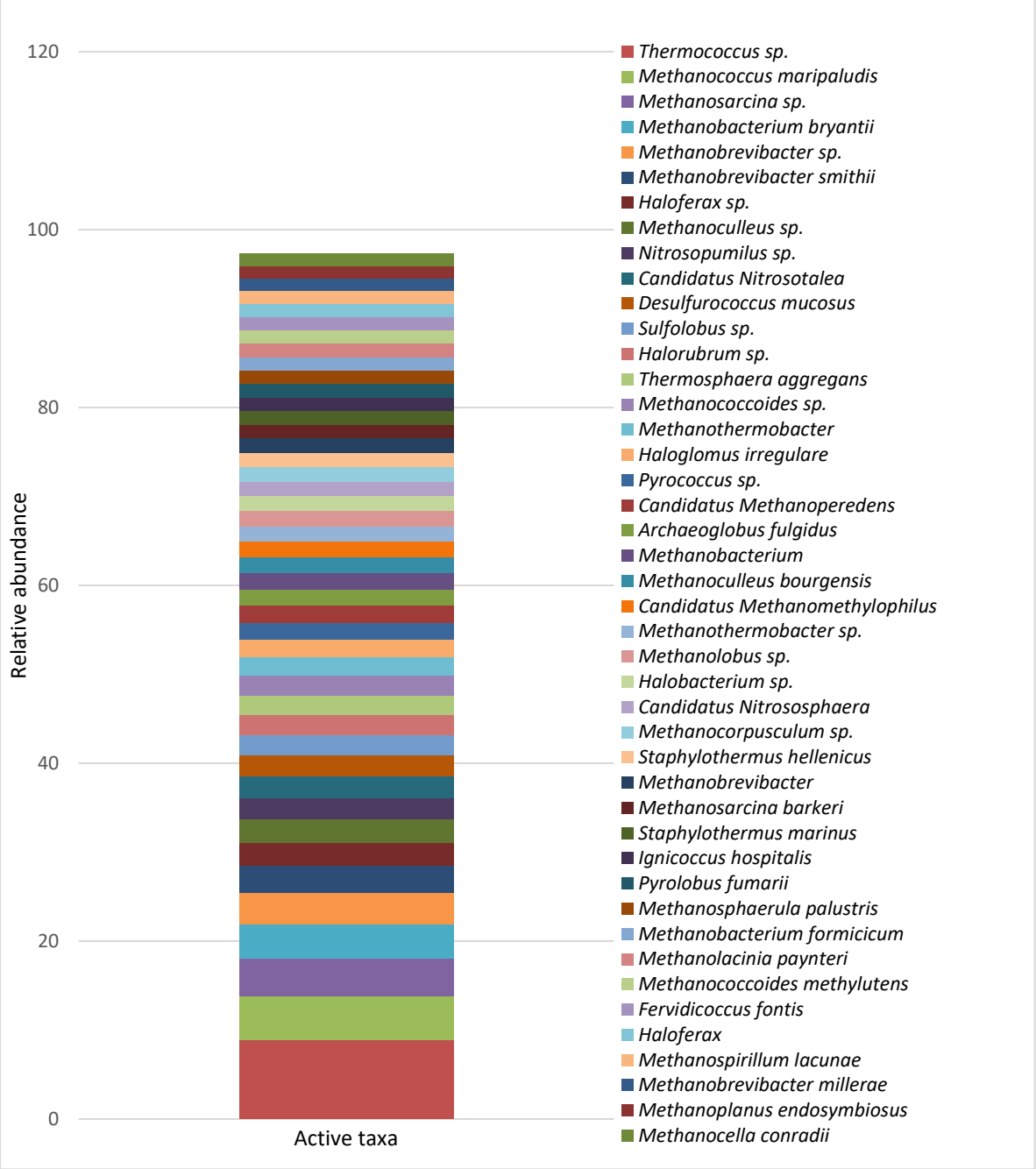


Figure 4. 11 The active gut archaeal community of *A. assamensis* and their relative abundance

to *Thermococcus* species followed by *Methanococcus maripaludis*, *Methanosarcina*, *Methanobacterium bryantii*, *Methanobrevibacter smithii*, *Haloferax sp.*, *Methanoculleus* sp., *Nitrosopumilus* sp., *Sulfolobus*, *Halorubrum* sp. and others as depicted in **Figure 4. 11**. Most of the active archaeal taxa belong to the methanogens such as *Methanococcus*, *Methanosarcina*, *Methanobacterium*, *Methanobrevibacter*, *Methanoculleus*, *Methanolobus*,

Methanothermobacter, *Methanoplanus*, *Methanospirillum*, *Methanosphaera* and *Methanocella* were found to be present and metabolically active inside the larval gut. Methyl-coenzyme M reductase glutamine C-methyltransferase, is the only enzyme required for all three major methanogenic pathways, viz., hydrogenotrophic, acetoclastic and methylotrophic pathways (Berghuis et al., 2019). Other key enzymes produced by methanogens such as formylmethanofuran dehydrogenase, bifunctional methylenetetrahydrofolate dehydrogenase, and F420H2 dehydrogenase were identified to be expressed by the gut methanogens of *A. assamensis* (**Table 4. 8).** A total of 2708 archaeal functions have been annotated to the metatranscriptome. Major functions expressed by the archaeal community have been given in **Table 4. 9**. The highest number of archaeal transcripts were annotated for translation elongation factor EF-1 alpha subunit followed by molecular chaperone DnaK, cytochrome P450, NADP-dependent oxidoreductase, aldehyde dehydrogenase family protein, alpha/beta hydrolase, carboxylesterase family protein and alpha-amylase family protein. The key enzymes involved in methanogenic pathways are also found to be expressed by the gut archaeal community.

Table 4. 8 Key enzymes involved in methanogenic pathways produced by the gut methanogens

Enzyme	**Number of transcripts annotated**	**Producing methanogen**
F420H2 dehydrogenase subunit FpoL	3283	*Methanosarcina*
Bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	273	*Methanosarcina*
Formylmethanofuran dehydrogenase subunit A	17	*Methanobacterium*, *Methanosarcina*
Methyl-coenzyme M reductase glutamine C-methyltransferase	16	*Methanothermobacter*, *Methanococcus*, *Methanobacter*

Table 4. 9 List of major genes expressed by the gut archaeal community of *A. assamensis*

Function	Relative abundance
Translation elongation factor EF-1 subunit alpha	12.78
Molecular chaperone DnaK	11.79
NADP-dependent oxidoreductase	2.57
Aldehyde dehydrogenase family protein	1.80
Proteasome-activating nucleotidase	1.19
Carboxylesterase family protein	0.76
Peptidylprolyl isomerase	0.75
Sodium-dependent transporter	0.74
Nucleoside-diphosphate kinase	0.69
Alpha-glucosidase	0.65
Glutamine--fructose-6-phosphate transaminase (isomerizing)	0.62
Archaeal proteasome endopeptidase complex subunit alpha	0.43
Phosphoglycerate dehydrogenase	0.43
Enoyl-CoA hydratase-related protein	0.34
Acyl-CoA dehydrogenase family protein	0.33
Methionine adenosyltransferase	0.20
Alpha/beta hydrolase	0.19
Thermosome subunit alpha	0.17
Glutamine-hydrolyzing GMP synthase	0.17
Adenosylhomocysteinase	0.17
Thiolase family protein	0.16
Alpha-ketoacid dehydrogenase subunit beta	0.14
Alpha-amylase family protein	0.14
UDP-glucose/GDP-mannose dehydrogenase family protein	0.14
Thermosome subunit beta	0.12
Glycoside hydrolase family 16 protein	0.11
GTP-binding protein	0.11
GDP-mannose 4,6-dehydratase	0.10
Phosphoenolpyruvate synthase	0.10
Carbamoyl-phosphate synthase (glutamine-hydrolyzing) large subunit	0.08
Catalase	0.07

Thiolase domain-containing protein	0.07
Glutamate dehydrogenase GdhB	0.07
Dihydrolipoyl dehydrogenase	0.06
Adenylosuccinate synthase	0.05
Fumarylacetoacetate hydrolase family protein	0.05
Peroxiredoxin	0.05
Thioredoxin domain-containing protein	0.05
Hydantoinase B/oxoprolinase family protein	0.05
Acetyl-CoA C-acetyltransferase	0.05
Biotin/lipoate A/B protein ligase family protein	0.05
Bifunctional N(6)-L-threonylcarbamoyladenine synthase/serine/threonine protein kinase	0.05
Archaeal proteasome endopeptidase complex subunit alpha	0.04
Cold-shock protein	0.04
L-lactate dehydrogenase	0.04
Pyruvate dehydrogenase (acetyl-transferring) E1 component subunit alpha	0.04
Pyridoxal phosphate-dependent aminotransferase	0.04
Aspartate carbamoyltransferase	0.04
Epoxide hydrolase	0.04
Hydroxymethylglutaryl-CoA reductase (NADPH)	0.02

4.4 Discussion

4.4.1 Carbohydrate metabolism and the larval gut microbiota

Annotation of the assembled transcripts identified ~7000 bacterial and 2700 archaeal active protein-encoding genes against the NCBI non-redundant protein database for bacteria and archaea. SEED Subsystem analysis enabled annotation of 2385 functions belonging to both bacteria and archaea which can be classified into 23 broad categories or subsystems (kindly refer to **Figure 4. 2** and **Table 4. 2**). Here, we have found that the majority of the transcripts were annotated to genes responsible for the basic biochemical processes of the microbes such as protein metabolism, cell signaling and regulation, respiration and nucleic acid metabolism.

However, many transcripts were annotated to genes responsible for interesting functional categories that may help us to understand the host-microbiome interaction. These categories have been discussed briefly below.

Carbohydrate metabolism is a basic biochemical process that ensures a continuous energy supply to a living cell (Dashty, 2013). Heterotrophic bacteria obtain the necessary energy through the oxidation of organic compounds such as carbohydrates, proteins and lipids (Peter Jurtshuk, 1996). As described in **section 2.4.4**, the leaf of green plants is mainly composed of cell wall polysaccharides and starch which work as the substrate for nutrition acquisition for the host as well as the host-associated microbes (Read & Holmes, 2017). It is also reported that gut microbiota enhances the capacity of the host to acquire nutrition from food material by digesting the non-digestible substrates (Valdes et al., 2018). In the metatranscriptome analysis of the gut microbial community of *A. assamensis* larvae, we found that several genes encoding carbohydrate metabolizing enzymes were being expressed by both the gut bacterial and the archaeal community (**Table 4. 4** and **Table 4. 9**). Carbohydrate esterase family 4 contains a group of enzymes that catalyze the de-acylation of polysaccharides(Armendáriz-Ruiz et al., 2018). This includes enzymes like chitin deacetylases, acetyl xylan esterases, xylanases and peptidoglycan N-acetylglucosamine deacetylases (Caufrier et al., 2003). Carbohydrate esterase family 4 proteins were found to be expressed by the gut microbial community at a high level with 284932 transcripts annotated to it. Methylglyoxal (MG) is a highly reactive compound that is produced as a by-product of various metabolic pathways, including glycolysis and lipid peroxidation. MG is known to be cytotoxic and can cause damage to cellular macromolecules, including proteins, DNA, and lipids. Therefore, cells have developed several mechanisms to detoxify MG and prevent its accumulation. (de Bari et al., 2020). The enzyme aldehyde dehydrogenase is known to metabolize methylglyoxal into pyruvate (Schumacher et al., 2018). The methylglyoxal metabolism has been found to be highly active among the larval gut

microbial community (**Table 4. 4**). Other functions like alpha and beta glucoside metabolism, maltose and maltodextrin utilization and fructose utilization have also been found to be highly expressed. These enzymes' ability to break down complex sugar molecules during food digestion was well reported from the gut of Lepidoptera (Ghadamyari et al., 2010; Huber et al., 2021; Zibaee et al., 2009). Bacterial species like *Acinetobacter*, *Lactobacillus*, *Serratia* were found to produce these enzymes (Appendix) which may assist the host in food digestion (Zhang et al., 2022). Trehalose is a non-reducing disaccharide widely present in living begins including insects. In insects, it is the major sugar component present in the hemolymph and plays a crucial role as an instant source of energy (Shukla et al., 2015). It also acts as a necessary source of energy during diapause in many insects including *Antheraea pernyi* (Y. N. Li et al., 2020). The gene encoding for trehalose synthase was found to be expressed by the gut microbial species of *Antheraea assamensis* in our study (**Table 4. 4**).

4.4.2 Identification of transcripts involved in the biosynthesis of vitamins and pigments

Biotin, also known as vitamin B_7 or vitamin H, is an essential nutrient for insects, as it plays a vital role in various physiological processes such as growth and development, reproduction, cuticle formation and detoxification (Douglas, 2017a, 2017b; Ren et al., 2021). However, insects are unable to synthesize biotin *de novo* and mostly derive it from microorganisms associated with their gut and other tissues (Douglas, 2017b). they also acquire biotin-synthesizing genes from microbes through horizontal gene transfer (Douglas, 2017b). In our metatranscriptome analysis, we have found that biotin biosynthesis is significantly active within the gut microbial community (Figure 4.3). Pyridoxine or vitamin B_6 is another essential contributing factor to the health of silkworms (Horie & Watanabe, 1983) which has also been detected to be expressed by gut microbiome (**Figure 4. 3** and **Table 4. 5**). Pterin-based pigmentation is a well-known phenomenon among animals. In Lepidoptera, pterins are often found in the scales that cover the wings of butterflies and moths (Andrade & Carneiro, 2021).

They can also impart colour to other body parts. In our analysis, folate and pterin metabolism is found to be expressed at a significant level (**Figure 4. 3**). Cocoon colouration in insect silk is mainly determined by the presence of pigment molecules like carotenoids and flavonoids (Sakudoh et al., 2007; Zhu & Zhang, 2014). Carotenoids were reported to be the key colouring component of yellow-coloured domestic silk (Ma et al., 2016). In our analysis, some key enzymes related to carotenoid biosynthesis were found to be expressed within the gut microbiome. These enzymes include phytoene synthase, geranylgeranyl pyrophosphate synthetase, neurosporene desaturase, methoxyneurosporene dehydrogenase, phytoene desaturase etc. (**Table 4. 5**). The role of these enzymes in carotenoid biosynthesis is well-reported (Ding et al., 2019; Hausmann & Sandmann, 2000; Tao et al., 2005; Zhou et al., 2022; Zhu & Zhang, 2014).

4.4.3 Stress response mechanism of the larval gut microbiota

The gut environment of Lepidopteran insects is unfavourable for microbial colonization due to its simple linear structure, faster food transit, high pH and low oxygen concentration (Hammer et al., 2017). However, microbial colonization inside the gut and their functional role in host physiology in Lepidoptera is well-established (X. Zhang et al., 2022). Microbial adaptation inside the hostile gut environment depends on the mechanism of action of stress response genes (Guan & Liu, 2020). Several genes responsible for heat shock, oxidative stress, osmotic stress and detoxification were found to be expressed by the larval gut microbiota (**Figure 4. 4** and **Table 4. 6**). When bacterial cells are subjected to elevated temperatures, a set of heat-shock proteins (hsps) are immediately and temporarily induced to reduce the protein damage (Yura et al., 1993). Glutathione is a naturally occurring molecule found in the body that acts as an antioxidant. It works by neutralizing free radicals and other harmful molecules and protects cell from damage (Yura et al., 1993). The host organism, i.e. the larva produces reactive oxygen species as a part of its defence mechanism to prevent microbial invasion (Engel & Moran, 2013).

The gut microbiota of *A. assamensis* have been found to produce several key enzymes for glutathione biosynthesis (**Table 4. 6**). Other two crucial enzymes, viz., catalase (Nandi et al., 2019) and superoxide dismutase (Sheng et al., 2014) which counter the effects of oxidative stress were observed to be significantly expressed.

4.4.4 Metabolism of aromatic compounds by the gut microbial community

Plant secondary metabolites are a diverse group of organic compounds that are not directly involved in the growth, development, or reproduction of plants, but play important roles in their survival and interactions with the environment. These compounds are often synthesized in response to biotic or abiotic stresses, such as herbivory, drought, or infection, and can have a variety of functions, such as defense against predators, the attraction of pollinators, or communication with other plants (Guerriero et al., 2018; Pang et al., 2021; Twaij & Hasan, 2022). There are four major classes of plant secondary metabolites, viz., phenolics, terpenoids, flavonoids, and sulphur-containing compounds (Guerriero et al., 2018). Plant secondary metabolites can have various effects on microorganisms, depending on the specific compound and the microbe involved. Some plant secondary metabolites have antimicrobial properties, meaning they can inhibit the growth or kill microorganisms such as bacteria, fungi, or viruses. These compounds may act by disrupting cell membranes or metabolic processes in the microorganisms, or by interfering with their DNA or RNA (Divekar et al., 2022). Herbivorous organisms employ several mechanisms to neutralize the effect of plant secondary metabolites (kindly refer section 2.4.6). In the metatranscriptome analysis of the gut microbial community of *A. assamensis* larvae, we have found that n-phenylalkanoic acid degradation pathway, salicyclate and gentisate catabolism, benzoate catabolism etc. are highly active (**Figure 4. 5**). In insects, some enzymes are reported to be only microbe derived which can degrade phenolic compounds and other xenobiotics (Jing et al., 2020). Here, we could identify more than 30 active genes encoding enzymes responsible for the catabolic pathways of various toxic plant

secondary metabolites (**Table 4. 7**).

4.4.5 The active gut microbial taxa inside the gut of A. assamensis

From our metatranscriptome analysis of the gut microbial community, we could identify more than 1000 active bacterial species. The highest active bacterium was found to be *Acinetobacter baumannii*. This gram-negative, non-motile, aerobic bacillus species was reported to be an opportunistic pathogen (Antunes et al., 2014; Howard et al., 2012; van der Kolk et al., 2019). *Acinetobacter* is known to be present abundantly among different insect groups (Vivero et al., 2016; Xue et al., 2021). In the 5th instar healthy larvae of *A. assamensis*, most of the genes expressed by this taxon is related to the basic life processes of the bacterium. Besides, it was also found to produce enzymes like catalase, glutaredoxin, methionine adenosyltransferase and aldehyde dehydrogenase which may impart benefit to the host (**Figure 4. 8**). The other four highest active taxa, namely, *Streptomyces*, *Ruminococcus albus*, *Pseudomonas syringae*, and *Paenibacillus* were previously reported from other insect groups (Brennan et al., 2004; Chevrette et al., 2019; Flury et al., 2016; Grady et al., 2016; J. Li et al., 2021; Sáez-Nieto et al., 2017; Smee et al., 2021).

Among the active archaeal species of the *A. assamensis* larvae, methanogens were found to be prevalent (**Figure 4. 11**). Methanogens are gut inhabitants of many organisms including human and herbivorous animals (Brune, 2010; Gaci et al., 2014; Horváthová et al., 2021; Kim et al., 2020; van de Pol et al., 2017). Methane production was reported in arthropods during the process of anaerobic carbohydrate fermentation (Bayon, 1980; Hackstein & Stumm, 1994). In our study, 3283 transcripts were annotated for the enzymes F420H2 dehydrogenase subunit FpoL (**Table 4. 8**). F420H2 dehydrogenase is an enzyme that catalyzes the oxidation of reduced coenzyme F420 (F420H2) to generate F420 and H+ ions. F420H2 is a key electron carrier in several metabolic pathways of anaerobic bacteria and archaea, and F420H2 dehydrogenase is essential for maintaining the balance of redox reactions in these organisms (Abken &

Deppenmeier, 1997). The function of F420H2 dehydrogenase is to transfer electrons from F420H2 to other electron acceptors, such as cytochromes or ferredoxins, to generate a proton motive force that drives ATP synthesis or other energy-dependent processes. This enzyme is also involved in the biosynthesis of complex molecules, such as methane and certain amino acids, and plays a role in the detoxification of harmful compounds in some microorganisms (Bäumer et al., 2000). Methylenetetrahydrofolate dehydrogenase (MTHFD), a bifunctional enzyme that also possesses 5,10-methenyltetrahydrofolate cyclohydrolase activity plays a critical role in the metabolism of folate, a vitamin that is essential for the synthesis of DNA, RNA, and proteins (Buchenau & Thauer, 2004; Christensen & MacKenzie, 2008). MTHFD is responsible for catalyzing the conversion of 5,10-methylenetetrahydrofolate to 5,10-methenyltetrahydrofolate, which is an important intermediate in folate-dependent metabolic pathways and was reported to be produced by *Methanosarcina* (Buchenau & Thauer, 2004). Other two essential enzymes for methanogenesis, formylmethanofuran dehydrogenase subunit A and methyl-coenzyme M reductase glutamine C-methyltransferase were also found to be expressed. However, only 17 and 16 transcripts were annotated against these two enzymes.

4.5 Conclusion

Metatranscriptome analysis of the gut microbial community of *A. assamensis* delineated the genes expressed by the microbiome inside the gut of 5th instar larvae. We could also identify the metabolically active bacterial and archaeal species inhabiting the larval gut. The active microbial enzymes may induce several beneficial effects on the host such as food digestion, detoxification of plant secondary metabolite, resistance to the oxidative stress inside the gut, and vitamin synthesis. This research would help the scientific community to identify the active gut microbial community of healthy *A. assamensis* larvae and understand their functional role in host biology. Future research focussing on pathogens, instar, host plant and location-associated variation in gut microbial gene expression would enhance our knowledge of the

mechanism of actions of the expressed microbial genes and their complex interaction with the host transcriptome. From the metagenomic and metatranscriptomic analysis of the gut microbial community of *A. assamensis* larvae, we can conclude that this lepidopteran species contain a very rich microbial community consisting of more than 1000 bacterial species and hundreds of archaeal species which are metabolically active and play a crucial role in host biology.

Reference

Abbà, S., Rossi, M., Vallino, M., Galetto, L., Marzachì, C., & Turina, M. (2022). Metatranscriptomic Assessment of the Microbial Community Associated With the Flavescence dorée Phytoplasma Insect Vector Scaphoideus titanus. *Frontiers in Microbiology*, *13*(866523). https://doi.org/10.3389/FMICB.2022.866523/FULL

Abken, H. J., & Deppenmeier, U. (1997). Purification and properties of an F420H2 dehydrogenase from Methanosarcina mazei Gö1. *FEMS Microbiology Letters*, *154*(2), 231–237. https://doi.org/10.1016/S0378-1097(97)00330-3

Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., Drew, D. A., Dulong, C., Rimm, E., Izard, J., Chan, A. T., & Huttenhower, C. (2018). Metatranscriptome of human fecal microbial communities in a cohort of adult men. *Nature Microbiology*, *3*(3), 356. https://doi.org/10.1038/S41564-017-0084-4

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., & Narasimhan, G. (2016). Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. *Evolutionary Bioinformatics Online*, *12*(Suppl 1), 5. https://doi.org/10.4137/EBO.S36436

Andrade, P., & Carneiro, M. (2021). Pterin-based pigmentation in animals. *Biology Letters*, *17*(8). https://doi.org/10.1098/RSBL.2021.0221

Antunes, L. C. S., Visca, P., & Towner, K. J. (2014). Acinetobacter baumannii: evolution of a global pathogen. *Pathogens and Disease*, *71*(3), 292–301. https://doi.org/10.1111/2049-632X.12125

Armendáriz-Ruiz, M., Rodríguez-González, J. A., Camacho-Ruíz, R. M., & Mateos-Díaz, J. C. (2018). Carbohydrate esterases: An overview. *Methods in Molecular Biology*, *1835*, 39–68. https://doi.org/10.1007/978-1-4939-8672-9_2/TABLES/5

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence

Data. (n.d.). Retrieved October 25, 2022, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bäumer, S., Ide, T., Jacobi, C., Johann, A., Gottschalk, G., & Deppenmeier, U. (2000). The F420H2 Dehydrogenase fromMethanosarcina mazei Is a Redox-driven Proton Pump Closely Related to NADH Dehydrogenases. *Journal of Biological Chemistry*, *275*(24), 17968–17973. https://doi.org/10.1074/JBC.M000650200

Bayon, C. (1980). Volatile fatty acids and methane production in relation to anaerobic carbohydrate fermentation in Oryctes nasicornis larvae (Coleoptera: Scarabaeidae). *Journal of Insect Physiology*, *26*(12), 819–828. https://doi.org/10.1016/0022-1910(80)90098-0

Berghuis, B. A., Yu, F. B., Schulz, F., Blainey, P. C., Woyke, T., & Quake, S. R. (2019). Hydrogenotrophic methanogenesis in archaeal phylum Verstraetearchaeota reveals the shared ancestry of all methanogens. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(11), 5037–5044. https://doi.org/10.1073/PNAS.1815631116/-/DCSUPPLEMENTAL

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brennan, Y. L., Callen, W. N., Christoffersen, L., Dupree, P., Goubet, F., Healey, S., Hernández, M., Keller, M., Li, K., Palackal, N., Sittenfeld, A., Tamayo, G., Wells, S., Hazlewood, G. P., Mathur, E. J., Short, J. M., Robertson, D. E., & Steer, B. A. (2004). Unusual Microbial Xylanases from Insect Guts. *Applied and Environmental Microbiology*, *70*(6), 3609. https://doi.org/10.1128/AEM.70.6.3609-3617.2004

Brune, A. (2010). Methanogenesis in the Digestive Tracts of Insects. *Handbook of Hydrocarbon and Lipid Microbiology*, 707–728. https://doi.org/10.1007/978-3-540-

77587-4_56

Buchenau, B., & Thauer, R. K. (2004). Tetrahydrofolate-specific enzymes in Methanosarcina barkeri and growth dependence of this methanogenic archaeon on folic acid or p-aminobenzoic acid. *Archives of Microbiology*, *182*(4), 313–325. https://doi.org/10.1007/S00203-004-0714-0

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature Methods 2014 12:1*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Carvalhais, L. C., Dennis, P. G., Tyson, G. W., & Schenk, P. M. (2012). Application of metatranscriptomics to soil environments. *Journal of Microbiological Methods*, *91*(2), 246–251. https://doi.org/10.1016/J.MIMET.2012.08.011

Caufrier, F., Martinou, A., Dupont, C., & Bouriotis, V. (2003). Carbohydrate esterase family 4 enzymes: Substrate specificity. *Carbohydrate Research*, *338*(7), 687–692. https://doi.org/10.1016/S0008-6215(03)00002-8

Chetia, H., Kabiraj, D., Singh, D., Mosahari, P. V., Das, S., Sharma, P., Neog, K., Sharma, S., Jayaprakash, P., & Bora, U. (2017). De novo transcriptome of the muga silkworm, Antheraea assamensis (Helfer). *Gene*, *611*, 54–65. https://doi.org/10.1016/j.gene.2017.02.021

Chevrette, M. G., Carlson, C. M., Ortega, H. E., Thomas, C., Ananiev, G. E., Barns, K. J., Book, A. J., Cagnazzo, J., Carlos, C., Flanigan, W., Grubbs, K. J., Horn, H. A., Hoffmann, F. M., Klassen, J. L., Knack, J. J., Lewin, G. R., McDonald, B. R., Muller, L., Melo, W. G. P., … Currie, C. R. (2019). The antimicrobial potential of Streptomyces from insect microbiomes. *Nature Communications*, *10*(1). https://doi.org/10.1038/S41467-019-08438-0

Christensen, K. E., & MacKenzie, R. E. (2008). Mitochondrial methylenetetrahydrofolate

dehydrogenase, methenyltetrahydrofolate cyclohydrolase, and formyltetrahydrofolate synthetases. *Vitamins and Hormones*, *79*, 393–410. https://doi.org/10.1016/S0083-6729(08)00414-7

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE*, *12*(12). https://doi.org/10.1371/JOURNAL.PONE.0190152

Dashty, M. (2013). A quick look at biochemistry: Carbohydrate metabolism. *Clinical Biochemistry*, *46*(15), 1339–1352. https://doi.org/10.1016/J.CLINBIOCHEM.2013.04.027

de Bari, L., Scirè, A., Minnelli, C., Cianfruglia, L., Kalapos, M. P., & Armeni, T. (2020). Interplay among Oxidative Stress, Methylglyoxal Pathway and S-Glutathionylation. *Antioxidants 2021, Vol. 10, Page 19*, *10*(1), 19. https://doi.org/10.3390/ANTIOX10010019

Ding, B. Y., Niu, J., Shang, F., Yang, L., Chang, T. Y., & Wang, J. J. (2019). Characterization of the Geranylgeranyl Diphosphate Synthase Gene in Acyrthosiphon pisum (Hemiptera: Aphididae) and Its Association With Carotenoid Biosynthesis. *Frontiers in Physiology*, *10*, 1398. https://doi.org/10.3389/FPHYS.2019.01398/BIBTEX

Divekar, P. A., Narayana, S., Divekar, B. A., Kumar, R., Gadratagi, B. G., Ray, A., Singh, A. K., Rani, V., Singh, V., Singh, A. K., Kumar, A., Singh, R. P., Meena, R. S., & Behera, T. K. (2022). Plant Secondary Metabolites as Defense Tools against Herbivores for Sustainable Crop Protection. *International Journal of Molecular Sciences*, *23*(5), 2690. https://doi.org/10.3390/IJMS23052690/S1

Douglas, A. E. (2017a). The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Current Opinion in Insect Science*, *23*, 65–69. https://doi.org/10.1016/J.COIS.2017.07.012

Douglas, A. E. (2017b). The B vitamin nutrition of insects: the contributions of diet, microbiome and horizontally acquired genes. *Current Opinion in Insect Science*, *23*, 65–69. https://doi.org/10.1016/J.COIS.2017.07.012

Engel, P., & Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, *37*(5), 699–735. https://doi.org/10.1111/1574-6976.12025

Flury, P., Aellen, N., Ruffner, B., Péchy-Tarr, M., Fataar, S., Metla, Z., Dominguez-Ferreras, A., Bloemberg, G., Frey, J., Goesmann, A., Raaijmakers, J. M., Duffy, B., Höfte, M., Blom, J., Smits, T. H. M., Keel, C., & Maurhofer, M. (2016). Insect pathogenicity in plant-beneficial pseudomonads: phylogenetic distribution and comparative genomics. *The ISME Journal 2016 10:10*, *10*(10), 2527–2542. https://doi.org/10.1038/ismej.2016.5

Gaci, N., Borrel, G., Tottey, W., O'Toole, P. W., & Brugère, J. F. (2014). Archaea and the human gut: New beginning of an old story. *World Journal of Gastroenterology : WJG*, *20*(43), 16062. https://doi.org/10.3748/WJG.V20.I43.16062

Ghadamyari, M., Hosseininaveh, V., & Sharifi, M. (2010). Partial biochemical characterization of α- and β-glucosidases of lesser mulberry pyralid, Glyphodes pyloalis Walker (Lep.: Pyralidae). *Comptes Rendus Biologies*, *333*(3), 197–204. https://doi.org/10.1016/J.CRVI.2009.12.011

Grady, E. N., MacDonald, J., Liu, L., Richman, A., & Yuan, Z. C. (2016). Current knowledge and perspectives of Paenibacillus: A review. *Microbial Cell Factories*, *15*(1), 1–18. https://doi.org/10.1186/S12934-016-0603-7/TABLES/1

Guan, N., & Liu, L. (2020). Microbial response to acid stress: mechanisms and applications. *Applied Microbiology and Biotechnology*, *104*(1), 51–65. https://doi.org/10.1007/S00253-019-10226-1/FIGURES/4

Guerriero, G., Berni, R., Muñoz-Sanchez, J. A., Apone, F., Abdel-Salam, E. M., Qahtan, A.

A., Alatar, A. A., Cantini, C., Cai, G., Hausman, J. F., Siddiqui, K. S., Hernández-Sotomayor, S. M. T., & Faisal, M. (2018). Production of Plant Secondary Metabolites: Examples, Tips and Suggestions for Biotechnologists. *Genes*, *9*(6), 309. https://doi.org/10.3390/GENES9060309

Hackstein, J. H. P., & Stumm, C. K. (1994). Methane production in terrestrial arthropods (methanogens/symbiouis/anaerobic protsts/evolution/atmospheric methane). *Proc. Nati. Acad. Sci. USA*, *91*, 5441–5445. https://www.pnas.org

Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P., & Fierer, N. (2017). Caterpillars lack a resident gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), 9641–9646. https://doi.org/10.1073/PNAS.1707186114/SUPPL_FILE/PNAS.1707186114.SAPP.PDF

Hausmann, A., & Sandmann, G. (2000). A single five-step desaturase is involved in the carotenoid biosynthesis pathway to beta-carotene and torulene in Neurospora crassa. *Fungal Genetics and Biology : FG & B*, *30*(2), 147–153. https://doi.org/10.1006/FGBI.2000.1212

He B, Jin S, Cao J, et al (2019) Metatranscriptomics of the Hu sheep rumen microbiome reveals novel cellulases. Biotechnol Biofuels 12:1–15. https://doi.org/10.1186/S13068-019-1498-4/FIGURES/5

Horie, Y., & Watanabe, K. (1983). Effects of dietary pyridoxine on larval growth, free amino acid pattern in haemolymph and uric acid excretion in the silkworm, Bombyx mori. *Insect Biochemistry*, *13*(2), 205–212. https://doi.org/10.1016/0020-1790(83)90085-9

Horváthová, T., Šustr, V., Chroňáková, A., Semanová, S., Lang, K., Dietrich, C., Hubáček, T., Ardestani, M. M., Lara, A. C., Brune, A., & Šimek, M. (2021). Methanogenesis in the Digestive Tracts of the Tropical Millipedes Archispirostreptus gigas (Diplopoda,

Spirostreptidae) and Epibolus pulchripes (Diplopoda, Pachybolidae). *Applied and Environmental Microbiology*, *87*(15), 1–13. https://doi.org/10.1128/AEM.00614-21

Howard, A., O'Donoghue, M., Feeney, A., & Sleator, R. D. (2012). Acinetobacter baumannii: An emerging opportunistic pathogen. *Virulence*, *3*(3), 243. https://doi.org/10.4161/VIRU.19700

Huber, M., Roder, T., Irmisch, S., Riedel, A., Gablenz, S., Fricke, J., Rahfeld, P., Reichelt, M., Paetz, C., Liechti, N., Hu, L., Bont, Z., Meng, Y., Huang, W., Robert, C. A. M., Gershenzon, J., & Erb, M. (2021). A beta-glucosidase of an insect herbivore determines both toxicity and deterrence of a dandelion defense metabolite. *ELife*, *10*. https://doi.org/10.7554/ELIFE.68642

Jing, T. Z., Qi, F. H., & Wang, Z. Y. (2020). Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome 2020 8:1*, *8*(1), 1–20. https://doi.org/10.1186/S40168-020-00823-Y

Johansson, H., Dhaygude, K., Lindström, S., Helanterä, H., Sundström, L., & Trontti, K. (2013). A Metatranscriptomic Approach to the Identification of Microbiota Associated with the Ant Formica exsecta. *PLOS ONE*, *8*(11), e79777. https://doi.org/10.1371/JOURNAL.PONE.0079777

Kim, J. Y., Whon, T. W., Lim, M. Y., Kim, Y. B., Kim, N., Kwon, M. S., Kim, J., Lee, S. H., Choi, H. J., Nam, I. H., Chung, W. H., Kim, J. H., Bae, J. W., Roh, S. W., & Nam, Y. do. (2020). The human gut archaeome: Identification of diverse haloarchaea in Korean subjects. *Microbiome*, *8*(1), 1–17. https://doi.org/10.1186/S40168-020-00894-X/FIGURES/6

Kopylova, E., Noé, L., & Lè Ne Touzet, H. (2012). *SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data*. *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, C., & Park, C. (2017). Bacterial Responses to Glyoxal and Methylglyoxal: Reactive Electrophilic Species. *International Journal of Molecular Sciences 2017, Vol. 18, Page 169*, *18*(1), 169. https://doi.org/10.3390/IJMS18010169

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, J., Sang, M., Jiang, Y., Wei, J., Shen, Y., Huang, Q., Li, Y., & Ni, J. (2021). Polyene-Producing Streptomyces spp. From the Fungus-Growing Termite Macrotermes barneyi Exhibit High Inhibitory Activity Against the Antagonistic Fungus Xylaria. *Frontiers in Microbiology*, *12*, 689. https://doi.org/10.3389/FMICB.2021.649962/BIBTEX

Li, Y. N., Liu, Y. B., Xie, X. Q., Zhang, J. N., & Li, W. L. (2020). The Modulation of Trehalose Metabolism by 20-Hydroxyecdysone in Antheraea pernyi (Lepidoptera: Saturniidae) During its Diapause Termination and Post-Termination Period. *Journal of Insect Science*, *20*(5), 1–12. https://doi.org/10.1093/JISESA/IEAA108

Ma, M., Hussain, M., Dong, S., & Zhou, W. (2016). Characterization of the pigment in naturally yellow-colored domestic silk. *Dyes and Pigments*, *124*, 6–11. https://doi.org/10.1016/J.DYEPIG.2015.08.003

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing

reads. *EMBnet.Journal*, *17*(1), 10–12. https://journal.embnet.org/index.php/embnetjournal/article/view/200/479

Mavric, E., Wittmann, S., Barth, G., & Henle, T. (2008). Identification and quantification of methylglyoxal as the dominant antibacterial constituent of Manuka (Leptospermum scoparium) honeys from New Zealand. *Molecular Nutrition & Food Research*, *52*(4), 483–489. https://doi.org/10.1002/MNFR.200700282

Nandi, A., Yan, L. J., Jana, C. K., & Das, N. (2019). Role of Catalase in Oxidative Stress- and Age-Associated Degenerative Diseases. *Oxidative Medicine and Cellular Longevity*, *2019*. https://doi.org/10.1155/2019/9613090

Nichols, R. G., & Davenport, E. R. (2021). *The relationship between the gut microbiome and host gene expression: a review*. *140*, 747–760. https://doi.org/10.1007/s00439-020-02237-0

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (n.d.). *The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)*. https://doi.org/10.1093/nar/gkt1226

Pang, Z., Chen, J., Wang, T., Gao, C., Li, Z., Guo, L., Xu, J., & Cheng, Y. (2021). Linking Plant Secondary Metabolites and Plant Microbiomes: A Review. *Frontiers in Plant Science*, *12*, 300. https://doi.org/10.3389/FPLS.2021.621276/BIBTEX

Pascault, N., Loux, V., Derozier, S., Martin, V., Debroas, D., Maloufi, S., Humbert, J. F., & Leloup, J. (2015). Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. *Genetica*, *143*(2), 157–167. https://doi.org/10.1007/S10709-014-9783-4/FIGURES/3

Peter Jurtshuk, Jr. (1996). Bacterial Metabolism. *Medical Microbiology*. https://www.ncbi.nlm.nih.gov/books/NBK7919/

Peters, B. A., Wilson, M., Moran, U., Pavlick, A., Izsak, A., Wechter, T., Weber, J. S., Osman, I., & Ahn, J. (2019). Relating the gut metagenome and metatranscriptome to immunotherapy responses in melanoma patients. *Genome Medicine*, *11*(1), 1–14. https://doi.org/10.1186/S13073-019-0672-4/FIGURES/5

Peterson, B. F., & Scharf, M. E. (2016). Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions. *BMC Genomics*, *17*(1), 1–12. https://doi.org/10.1186/S12864-016-3126-Z/FIGURES/6

Read, M. N., & Holmes, A. J. (2017). Towards an integrative understanding of diet-host-gut microbiome interactions. *Frontiers in Immunology*, *8*(MAY), 538. https://doi.org/10.3389/FIMMU.2017.00538/BIBTEX

Reck, M., Tomasch, J., Deng, Z., Jarek, M., Husemann, P., & Wagner-Döbler, I. (2015). Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics*, *16*(1), 1–18. https://doi.org/10.1186/S12864-015-1694-Y/FIGURES/7

Ren, F. R., Bai, B., Hong, J. S., Huang, Y. Z., & Luan, J. B. (2021). A microbiological assay for biotin determination in insects. *Insect Science*, *28*(2), 415–418. https://doi.org/10.1111/1744-7917.12827

Rio DC, Ares M, Hannon GJ, Nilsen TW (2010) Purification of RNA using TRIzol (TRI Reagent). Cold Spring Harb Protoc 5:. https://doi.org/10.1101/PDB.PROT5439

Rodriguez-Esteban, R., & Jiang, X. (2017). Differential gene expression in disease: A comparison between high-throughput studies and the literature. *BMC Medical Genomics*, *10*(1), 1–10. https://doi.org/10.1186/S12920-017-0293-Y/FIGURES/6

Sáez-Nieto, J. A., Medina-Pascual, M. J., Carrasco, G., Garrido, N., Fernandez-Torres, M. A., Villalón, P., & Valdezate, S. (2017). Paenibacillus spp. isolated from human and environmental samples in Spain: detection of 11 new species. *New Microbes and New Infections*, *19*, 19. https://doi.org/10.1016/J.NMNI.2017.05.006

Sakudoh, T., Sezutsu, H., Nakashima, T., Kobayashi, I., Fujimoto, H., Uchino, K., Banno, Y., Iwano, H., Maekawa, H., Tamura, T., Kataoka, H., & Tsuchida, K. (2007). From the Cover: Carotenoid silk coloration is controlled by a carotenoid-binding protein, a product of the Yellow blood gene. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 8941. https://doi.org/10.1073/PNAS.0702860104

Schumacher, D., Morgenstern, J., Oguchi, Y., Volk, N., Kopf, S., Groener, J. B., Nawroth, P. P., Fleming, T., & Freichel, M. (2018). Compensatory mechanisms for methylglyoxal detoxification in experimental & clinical diabetes. *Molecular Metabolism*, *18*, 143. https://doi.org/10.1016/J.MOLMET.2018.09.005

Segundo-Val, I. S., & Sanz-Lozano, C. S. (2016). Introduction to the Gene Expression Analysis. *Methods in Molecular Biology (Clifton, N.J.)*, *1434*, 29–43. https://doi.org/10.1007/978-1-4939-3652-6_3

Shakya, M., Lo, C. C., & Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. In *Frontiers in Genetics* (Vol. 10, Issue SEP). Frontiers Media S.A. https://doi.org/10.3389/fgene.2019.00904

Sheng, Y., Abreu, I. A., Cabelli, D. E., Maroney, M. J., Miller, A. F., Teixeira, M., & Valentine, J. S. (2014). Superoxide dismutases and superoxide reductases. *Chemical Reviews*, *114*(7), 3854–3918. https://doi.org/10.1021/CR4005296/ASSET/IMAGES/MEDIUM/CR-2013-005296_0036.GIF

Shukla, E., Thorat, L. J., Nath, B. B., & Gaikwad, S. M. (2015). Insect trehalase: physiological significance and potential applications. *Glycobiology*, *25*(4), 357–367. https://doi.org/10.1093/GLYCOB/CWU125

Smee, M. R., Real-Ramirez, I., Arias, C. Z., & Hendry, T. A. (2021). Epiphytic Strains of Pseudomonas syringae Kill Diverse Aphid Species. *Applied and Environmental*

Microbiology, *87*(11), 1–11. https://doi.org/10.1128/AEM.00017-21

st. Laurent, G., Shtokalo, D., Tackett, M. R., Yang, Z., Vyatkin, Y., Milos, P. M., Seilheimer, B., McCaffrey, T. A., & Kapranov, P. (2013). On the importance of small changes in RNA expression. *Methods*, *63*(1), 18–24. https://doi.org/10.1016/j.ymeth.2013.03.027

Tao, L., Schenzle, A., Odom, J. M., & Cheng, Q. (2005). Novel carotenoid oxidase involved in biosynthesis of 4,4'-diapolycopene dialdehyde. *Applied and Environmental Microbiology*, *71*(6), 3294–3301. https://doi.org/10.1128/AEM.71.6.3294-3301.2005

Thomas, J. M., Bryson, K. M., & Meier, J. L. (2019). Nucleotide resolution sequencing of N4-acetylcytidine in RNA. *Methods in Enzymology*, *621*, 31–51. https://doi.org/10.1016/BS.MIE.2019.02.022

Tláskal, V., Brabcová, V., Větrovský, T., López-Mondéjar, R., Monteiro, L. M. O., Saraiva, J. P., da Rocha, U. N., & Baldrian, P. (2021). Metagenomes, metatranscriptomes and microbiomes of naturally decomposing deadwood. *Scientific Data 2021 8:1*, *8*(1), 1–7. https://doi.org/10.1038/s41597-021-00987-8

Twaij, B. M., & Hasan, M. N. (2022). Bioactive Secondary Metabolites from Plant Sources: Types, Synthesis, and Their Therapeutic Uses. *International Journal of Plant Biology 2022, Vol. 13, Pages 4-14*, *13*(1), 4–14. https://doi.org/10.3390/IJPB13010003

Valdes, A. M., Walter, J., Segal, E., & Spector, T. D. (2018). Role of the gut microbiota in nutrition and health. *BMJ*, *361*, 36–44. https://doi.org/10.1136/BMJ.K2179

van de Pol, J. A. A., Best, N. van, Mbakwa, C. A., Thijs, C., Savelkoul, P. H., Ilja, I. C., Hornef, M. W., Mommers, M., & Penders, J. (2017). Gut colonization by methanogenic archaea is associated with organic dairy consumption in children. *Frontiers in Microbiology*, *8*(MAR), 355. https://doi.org/10.3389/FMICB.2017.00355/BIBTEX

van der Kolk, J. H., Endimiani, A., Graubner, C., Gerber, V., & Perreten, V. (2019). Acinetobacter in veterinary medicine, with an emphasis on Acinetobacter baumannii.

Journal of Global Antimicrobial Resistance, *16*, 59–71. https://doi.org/10.1016/J.JGAR.2018.08.011

Vennapusa, A. R., Somayanda, I. M., Doherty, C. J., & Jagadish, S. V. K. (2020). A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber. *Scientific Reports 2020 10:1*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-73958-5

Vivero, R. J., Jaramillo, N. G., Cadavid-Restrepo, G., Soto, S. I. U., & Herrera, C. X. M. (2016). Structural differences in gut bacteria communities in developmental stages of natural populations of Lutzomyia evansi from Colombia's Caribbean coast. *Parasites and Vectors*, *9*(1), 1–20. https://doi.org/10.1186/S13071-016-1766-0/FIGURES/8

Westreich, S. T., Treiber, M. L., Mills, D. A., Korf, I., & Lemay, D. G. (2018). SAMSA2: A standalone metatranscriptome analysis pipeline. *BMC Bioinformatics*, *19*(1), 1–11. https://doi.org/10.1186/S12859-018-2189-Z/TABLES/2

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1891-0

Wyllie, S., & Fairlamb, A. H. (2011). Methylglyoxal metabolism in trypanosomes and leishmania. *Seminars in Cell & Developmental Biology*, *22*(3), 271. https://doi.org/10.1016/J.SEMCDB.2011.02.001

Xue, H., Zhu, X., Wang, L., Zhang, K., Li, D., Ji, J., Niu, L., Wu, C., Gao, X., Luo, J., & Cui, J. (2021). Gut Bacterial Diversity in Different Life Cycle Stages of Adelphocoris suturalis (Hemiptera: Miridae). *Frontiers in Microbiology*, *12*, 670383. https://doi.org/10.3389/FMICB.2021.670383/FULL

Yang, Y., Sun, J., Chen, C., Zhou, Y., Lee, C., Dover, V., Wang, C., Qiu, J.-W., & Qian, P.-Y. (2022). Metagenomic and metatranscriptomic analyses reveal minor-yet-crucial roles of gut microbiome in deep-sea hydrothermal vent snail. *Animal Microbiome 2021 4:1*, *4*(1),

1–17. https://doi.org/10.1186/S42523-021-00150-Z

Yura, T., Nagai, H., & Mori, H. (1993). Regulation of the heat-shock response in bacteria. *Annual Review of Microbiology*, *47*, 321–350. https://doi.org/10.1146/ANNUREV.MI.47.100193.001541

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614. https://doi.org/10.1093/BIOINFORMATICS/BTT593

Zhang, X., Zhang, F., & Lu, X. (2022). Diversity and Functional Roles of the Gut Microbiota in Lepidopteran Insects. *Microorganisms*, *10*(6). https://doi.org/10.3390/MICROORGANISMS10061234

Zhou, X., Rao, S., Wrightstone, E., Sun, T., Lui, A. C. W., Welsch, R., & Li, L. (2022). Phytoene Synthase: The Key Rate-Limiting Enzyme of Carotenoid Biosynthesis in Plants. *Frontiers in Plant Science*, *13*, 977. https://doi.org/10.3389/FPLS.2022.884720/BIBTEX

Zhu, L., & Zhang, Y. Q. (2014). Identification and Analysis of the Pigment Composition and Sources in the Colored Cocoon of the Silkworm, Bombyx mori, by HPLC-DAD. *Https://Doi.Org/10.1673/031.014.31*, *14*(31), 1–10. https://doi.org/10.1673/031.014.31

Zibaee, A., Bandani, A. R., & Ramzi, S. (2009). Enzymatic properties of α- and β-glocusidases extracted from midgut and salivary glands of rice striped stem borer, Chilo suppressalis Walker (Lepidoptera: Pyralidae). *Comptes Rendus Biologies*, *332*(7), 633–641. https://doi.org/10.1016/J.CRVI.2009.02.009

Chapter 5

Untargeted metabolomic analysis of *Antheraea assamensis* Helfer

CHAPTER 5

Untargeted metabolomic analysis of *Antheraea assamensis* Helfer

Executive summary

Metabolomics is a field of research that focuses on the identification and quantification of the small molecules, known as metabolites, present in biological systems. Metabolites are the end products of cellular processes, and their levels can provide insight into the metabolic state of an organism. Untargeted metabolomics of insect gut refers to the comprehensive analysis of the metabolites present in the gut of insects without prior knowledge of their identity till date, no comprehensive investigation has been carried out through untargeted metabolomics to profile the global metabolome of Muga silkworm or to study its variation in different conditions. The key goal of this objective is to identify the global metabolome of the host plant leaf, gut and silk gland of 5th instar larvae of *A. assamensis* and how the change in the host plant affects the gut and silk gland metabolite composition. For this, we considered larvae reared on Som and Mejankari plants. The results of this experiment identified hundreds of metabolites from the larval gut, silk gland and host plant leaf of *A. assamensis* larvae reared on Som and Mejankari plants. The identified metabolites can be identified into different categories like amino acids, sugars, fatty acids, alkaloids, secondary metabolites, pigments vitamins and others. Change in the host plant was found to change the metabolite composition of the larval gut and the silk gland. The finding of this study will help to extend our knowledge of the metabolomes of the studied samples, which have not been reported to date.

5.1 Introduction

Metabolomics is a field of research that focuses on the identification and quantification of the small molecules, known as metabolites, present in biological systems (Clish, 2015b). Metabolites are the end products of cellular processes, and their levels can provide insight into the metabolic state of an organism (Rinschen et al., 2019). It is an interdisciplinary field that combines analytical chemistry, biochemistry, and systems biology, among others, to study the complex interactions between metabolites and their environment (Alvares, 1978; Piazza et al., 2018). Metabolomics is a relatively new field of research, and it has emerged as a result of recent advancements in analytical techniques, such as liquid chromatography-mass spectrometry (LC-MS), gas chromatography-mass spectrometry (GC-MS), and nuclear magnetic resonance (NMR) spectroscopy (Clish, 2015a). These techniques allow researchers to detect and quantify metabolites in complex biological samples, such as blood, urine, and tissues.

Metabolomics can be broadly classified into two categories: targeted and untargeted metabolomics (Lelli et al., 2021). Targeted metabolomics focuses on the identification and quantification of a specific set of metabolites, such as those involved in a particular metabolic pathway or disease (Roberts et al., 2012). Untargeted metabolomics, on the other hand, aims to identify and quantify as many metabolites as possible in a biological sample without prior knowledge of their identity or function (Schrimpe-Rutledge et al., 2016; Snart et al., 2015).

Metabolomics has a wide range of applications in various fields, including biomedical research, environmental science, agriculture, and food science (Q. Yang et al., 2019). In biomedical research, metabolomics can be used to study the metabolic changes associated with various diseases, such as cancer, diabetes, and heart disease (Kim et al., 2016). Metabolomics can also be used to identify biomarkers for disease diagnosis and prognosis (Zhang et al., 2015). In environmental science, metabolomics can be used to study the impact of environmental factors, such as pollutants and toxins, on the metabolism of organisms (Bedia, 2022; Lankadurai et al.,

2013). Metabolomics can also be used to monitor the health of ecosystems and to identify new species of microorganisms. In agriculture, metabolomics can be used to study the metabolic pathways involved in crop growth and development. This can also be used to identify the metabolites that contribute to the nutritional quality of crops and to develop new crop varieties with improved nutritional profiles (Ibarra-Estrada et al., 2016). Overall, metabolomics is a powerful tool for studying the complex interactions between metabolites and their environment. It has the potential to revolutionize our understanding of the metabolic processes that underlie various biological systems and to develop new diagnostic and therapeutic approaches for various diseases.

Untargeted metabolomics of insect gut refers to the comprehensive analysis of the metabolites present in the gut of insects without prior knowledge of their identity. The gut microbiota of insects plays an important role in nutrient acquisition and immune defense, and untargeted metabolomics has emerged as a powerful tool to investigate the metabolic interactions between insects and their gut microbiota (Muñoz-Benavent et al., 2021).

In recent years, there has been a growing interest in studying the gut metabolites of insects as a means of understanding the ecology and physiology of insects, as well as developing new insecticides and biological control agents. Untargeted metabolomics has been used to identify the metabolites produced by the gut microbiota of various insects, including fruit flies, beetles, and moths (Muñoz-Benavent et al., 2021; Roques et al., 2020; Snart et al., 2015; Späth et al., 2022; L. Yang et al., 2022). Untargeted metabolomics has also been used to investigate the impact of environmental factors, such as diet and temperature, on insect gut metabolites (Oliveira et al., 2022; Raza et al., 2020). For example, one study found that the gut metabolites of fruit flies were significantly affected by changes in temperature, with metabolites involved in energy metabolism and amino acid biosynthesis showing the most significant changes (Moghadam et al., 2018).

Silk is a natural protein fiber that has been used for centuries in a wide range of applications,

from clothing to medicine (Sutherland et al., 2009). Silk production is a complex process that involves the biosynthesis of silk proteins in specialized glands found in silk-producing insects such as silkworms, spiders, and caterpillars (Sehnal & Sutherland, 2008). In recent years, untargeted metabolomics has emerged as a powerful tool for investigating the metabolic pathways and networks involved in silk production in silk-producing insects and their silk gland (Cao et al., 2020).

The silk gland is a specialized organ found in silk-producing insects that synthesizes and secretes silk proteins. The silk gland is composed of several regions, each with a specific function in the synthesis and processing of silk proteins. The gland is capable of producing a wide range of different types of silk, each with unique physical and chemical properties (Sehnal & Akai, 1990).

Several studies have used untargeted metabolomics to investigate the metabolites involved in silk production in silk-producing insects and their silk gland. For example, in silkworms, untargeted metabolomics has identified several amino acids, lipids, and sugars that are involved in the biosynthesis of silk proteins. Similarly, in spiders, untargeted metabolomics has been used to identify the metabolites involved in the production of spider silk, including glycolipids, fatty acids, and polyamines (Q. Chen et al., 2015; Wang et al., 2019; Wu et al., 2022).

One of the main applications of untargeted metabolomics in silk-producing insects has been to investigate the impact of environmental factors on the metabolites present in the silk gland and silk produced by silk-producing insects. Studies have shown that changes in temperature, humidity, and diet can significantly affect the metabolites present in the silk gland and silk produced by silk-producing insects (Dong et al., 2017; Oliveira et al., 2022).

The Muga silk produced by *A. assamensis* is known for its unique properties, including its golden color and durability (Tikader et al., 2013). The silk gland of *A. assamensis* is responsible for the biosynthesis and secretion of silk proteins, which are used to produce Muga silk. The silk gland of *A. assamensis* is composed of the anterior, middle and posterior silk gland, each

with a specific function in the biosynthesis and processing of silk proteins (Goswami & Devi, 2021). The untargeted metabolomics approach involves the comprehensive analysis of metabolites in biological samples without prior knowledge of their identity. This approach has been used to identify the metabolites involved in the biosynthesis of silk proteins in the silk gland of *A. assamensis*. Untargeted metabolomics can provide insights into the metabolic processes that occur during the life cycle of Muga silkworm, silk synthesis and the effects of biotic and abiotic factors. However, to date, no comprehensive investigation has been carried out through untargeted metabolomics to profile the global metabolome of the Muga silkworm or to study its variation in different conditions. The key goal of this objective is to identify the global metabolome of the host plant leaf, gut and silk gland of 5^{th} instar larvae of *A. assamensis* and how the change in the host plant affects the gut and silk gland metabolite composition. For this, we considered larvae reared on Som and Mejankari plants. The results of this experiment identified hundreds of metabolites from the larval gut, silk gland and host plant leaf of *A. assamensis* larvae reared on Som and Mejankari plants. The identified metabolites can be identified into different categories like amino acids, sugars, fatty acids, alkaloids, secondary metabolites, pigments vitamins and others. Change in the host plant was found to change the metabolite composition of the larval gut and the silk gland. The finding of this study will help to extend our knowledge on the metabolomes of the studied samples, which have not been reported till date.

5.2 Materials and methods

5.2.1 Materials

Reagent required:

a. Acetonitrile (HPLC grade)

b. Absolute methanol (HPLC grade)

c. Formic acid (analytical grade, 98-100%)

d. Phosphoric acid (85% in water, w/v)

e. Ultrapure water

f. Liquid nitrogen

Sample extraction solution:

99.875% methanol solution acidified with 0.125% formic acid.

Mobile phase: Two eluents were used

a. 0.1% formic acid (v/v) in ultrapure water

b. 0.1% formic acid (v/v) in acetonitrile.

Equipment required

a. Mortar and pestle

b. Micropipettes and tips

c. Disposable sterile syringe

d. Disposable nylon membrane syringe filters (0.2 µm pore size, 10mm diameter, should be resistant to the extraction solution)

e. Crimp cap autosampler vial (1-2 ml)

f. High-resolution mass spectrometry (HRMS)

5.2.2 Sample preparation and extraction

Antherea assamensis larvae were reared in natural conditions on two different host plants, *Machilus bombycina* and *Litsea citrata* inside the Indian Institute of Technology Guwahati

campus (26.2027° N, 91.7004° E). The gut and silk glands of the 5th instar larvae were prepared for metabolome extraction as described by Vinayavekhin & Saghatelian with some modification (Vinayavekhin & Saghatelian, 2010). Briefly, the gut of 5-6 numbers of 5th instar larvae of *A. assamensis* reared on the above two host plants were collected and freeze killed immediately. Freeze killing was done by pouring adequate amount of liquid nitrogen to cover and freeze the insect (Zheng et al., 2019). The whole gut was dissected out and flash-frozen. The frozen tissue was homogenized in liquid nitrogen in a sterile pre-cooled mortar and pestle. The homogenized powder was transferred into a storage container resistant to liquid nitrogen. Tissue was kept well frozen during the homogenization process using liquid nitrogen. The intact silk glands were dissected out carefully from 5-6 larvae reared on *M. bombycina* and immediately frozen with liquid nitrogen. Similarly, the silk gland of larvae reared on *Litsea citrata* were also dissected out and flash-frozen. The silk glands of 5-6 larvae were pooled to make one sample for each of the host plants and homogenized into a fine powder using freeze-cooled mortar-pestle and liquid nitrogen. The fresh leaves of the two host plants, i.e., *M. bombycina* and *L. citrata* were also collected and flash frozen immediately after collection as described by Perez de Souza and his coauthors (Perez de Souza et al., 2019). The leaf tissue homogenization was done similarly to the gut and the silk glands. All samples were divided into aliquots and kept at -80°C until metabolite extraction.

5.2.3 Metabolites extraction

The sample extract was prepared by adding ice-cold sample extraction solution to 100 mg of the frozen powder in a 1:3 (sample: extraction solution) ratio. The mixture was then vortexed immediately for 30 seconds. The extract was stored on ice until all the samples were ready. This process was repeated for the silk gland and the leaf samples. Each sample was then sonicated for 15 minutes at maximum frequency continuously at room temperature. The sonicated mixture was centrifuged at room temperature at maximum speed. The supernatant was filtered through a 0.2 µm nylon membrane filter using a disposable syringe into a 1.8 ml

crimp auto-sampler vial. The extract was then analyzed with HRMS. The extraction solution without a sample was used as a control to run the HRMS analysis.

5.2.4 Liquid Chromatography- High-Resolution Mass Spectrometry (LC-HRMS) analysis

An Agilent 6520 LC-HRMS system equipped with a quadrupole–time of flight mass spectrometer (Q-TOF) (Agilent Technologies, Santa Clara, CA, USA) was used for the instrumental analysis (Barco et al., 2022; Lebeau-Roche et al., 2021; Peces et al., 2017; Ríos Peces et al., 2017; Rochat et al., 2018). The instrument is capable of detecting a mass range of ~ 20-20000 Daltons (TOF) with up to 40,000 mass resolution, sub-1-ppm MS and 2-ppm MS/MS mass accuracy. Ionization (±eV) technology used in the instrument was Agilent Jet Stream Technology, Orthogonal Spray Ionization Source and Electrospray & Atmospheric Pressure Chemical Ionization. The installed detector was MS/MS using a quadrupole, a hexapole – collision cell – and a time-of-flight (TOF) which can detect up to 20 MS or 10 MS/MS spectra per second. A liquid chromatography (LC) inlet system of Agilent Technologies Series 1200 binary solvent delivery system with thermostated autosampler compartment (vials & 96 well plates) was used. Samples were run in both positive and negative mode tuning masses with the parameter of draw speed 200.0 µL/min, eject speed 400.0 µL/min, wait time after draw 1.2 seconds and sample flush-out factor of 5.0 times injection volume. After setting the parameters and the protocol, the samples were allowed to run for the analysis which took around an hour. All the detected peaks were exported for further analysis.

5.2.5 HRMS data analysis

Mzmine 3.3.0 software was used for the processing of HRMS raw data. It is an open-source software tool for mass spectrometry-based metabolomics data analysis (Pluskal et al., 2010; Schmid et al., 2023). It is a powerful and user-friendly platform that enables researchers to preprocess, visualize, and analyze large metabolomics datasets with ease. The typical MS data processing workflow comprises raw data file import, mass detection, filtering/smoothing

(optional), peak picking, peak list deisotoping, alignment, gap filling, and annotation. The MZmine 3.3.0 modules cover all these workflow stages and also include additional functionality for the visualization and interpretation of the results.

The workflow proposed herein is intended as a general pipeline for untargeted LC-MS (or LC−MS/MS) data preprocessing. The main goal is essentially to turn the highly complex LC-MS raw data into a list of features, and corresponding signal intensity detected across the analysed samples. Such feature lists can then be exported for further downstream analysis (e.g., identification, search against spectral libraries, statistical analysis, etc.). A schematic representation of the workflow is shown in **Figure 5. 1**.

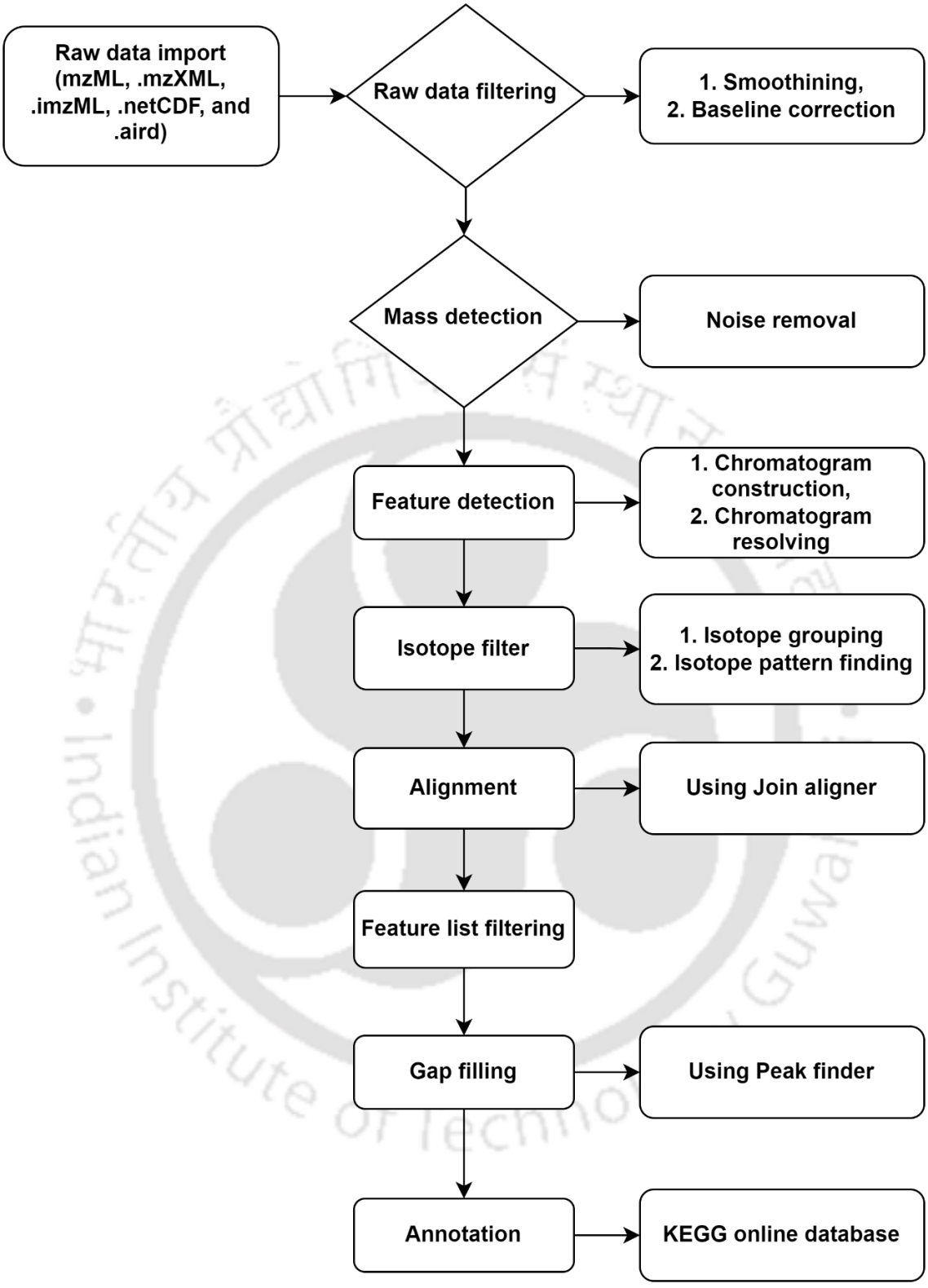


Figure 5. 1 Untargeted LC-MS Workflow

Raw data were generated from HRMS (Q-TOF/MS) analysis Agilent 6520 system instrument in the form of ".d" extended format. This ".d" extended format was further converted to .mzML / .imzML in a converter software called MSConvert (ProteoWizard) as MZmine 3.3.0 support only .mzML, .mzXML, .imzML, .netCDF, and .aird data format type. In the conversion, the peak picking module was adapted. The MsConvert uses peak picking to improve the quality of the mass spectrometry data by extracting the most significant peaks and reducing the amount of noise and irrelevant information, which can help downstream analysis tools to identify and quantify the compounds of interest.

The converted .mzXML files were imported to the MZmine 3.3.0 software by clicking the "Raw data import" tool under the "Raw Data methods" tab and selecting all files from the computer. After selecting all of the raw data files by clicking on the "Mass Detection" tool using the following path: "Raw data methods" → "Feature Detection" → "Mass Detection". In the default parameters, the following adjustments were made, mass detector- Centroid (noise level 1.0E3).

In MZmine 3.3.0, the centroid mass detector is a data processing tool that identifies and extracts individual peaks from raw mass spectrometry data. Centroiding is the process of converting the raw mass spectra into a set of discrete, individual peaks, each representing a single compound or ion in the sample. For feature detection and chromatogram building, the "ADAP chromatogram builder" module was adapted which firstly identifies the m/z values of interest based on a user-defined list of target compounds or ions. It then extracts the raw intensity values of those m/z values from the raw mass spectrometry data and constructs a chromatogram for each target compound or ion. In the process following adjustments were made, min group size in # scan: 5; group intensity threshold: 2.5e3; min highest intensity: 2.5e3; m/z tolerance: 0.020 or 10 ppm and kept other parameters in default.

To resolve the chromatogram, "Local minimum resolver" was adapted by using the following path: "Feature detection" → "Chromatogram resolving" → "Local minimum resolver". Here,

the feature-resolving step enables the separation of co-eluting and overlapping chromatography peaks and as such is one of the pivotal steps in data preprocessing. In the process following adjustments were made, Chromatographic threshold: 70%; Minimum absolute height: 2.0; Min ratio of peak top/edge: 1.2; Peak duration range (min/mobility): 0.0- 10.0; Min # of data points: 5. To remove the redundant features, such as the ones generated due to the presence of isotopologues, 13C isotope filter (isotope grouper) was applied where this removes 13C isotope features from the feature list using the isotope finder (which is sensitive in the detection of possible isotope signals). To perform the process the following path was used: "Feature list methods" → "Isotopes" → "13C isotope filter" and the parameter were changed as follows, m/z tolerance: 0.02 or 10 ppm; RT tolerance: 0.20; max charge: 2.

In mass spectrometry, when a compound is ionized, it may produce a series of peaks corresponding to the different isotopes of the ion. For example, the ion of a compound with one carbon atom and two nitrogen atoms (C_1N_2) may produce a series of peaks at m/z 42, 43, and 44, corresponding to the isotopes of carbon (C_{12} and C_{13}) and nitrogen (N_{14} and N_{15}). The isotope pattern finder tool is used to detect these isotopic peaks and help confirm the identity of the compound. Here, the path for the isotope pattern finder is as follows, "Feature list methods" → "Isotopes" → "Isotope pattern finder". In the parameter, the adjusted values included m/z tolerance: 0.02 or 10 ppm; maximum charge of isotope m/z: 2.

Further, corresponding features across all samples are aligned using the join aligner module. This module aligns the detected peaks across different samples through a match score, which is calculated based on the mass and retention time of each peak and ranges of tolerance stipulated in the parameter setup dialog. Path involved in the process is "feature list methods" → "Alignment" → "Join aligner". In the module, various parameters were adjusted for the best result in the following values: m/z tolerance: 0.02 or 10 ppm; weight for m/z: 3 (default); Retention Time tolerance: 0.2; weight for RT: 1 (default).

Quality control of the features was done by using the "feature list filtering" module which

removes low-quality or noisy features from the feature list based on specific criteria or thresholds. The path followed was: "feature list methods" → "feature list filtering" → "feature list rows filter", and the adjusted parameter was as follows: min peak in a row: 0.100; m/z: 100.00- 2500.00; Retention time: .05- 1.00 min. To improve data completeness, better statistical analysis, facilitate downstream analyses and better identification of unknown metabolites the "gap filling" module was used. This module involves imputing missing values in the data matrix. Here, the missing values could occur in metabolomics data due to various reasons, such as instrument errors, sample handling issues, or analytical limitations. To perform the process the following path was used: "Feature list methods" → "Gap filling" → "Peak finder". In the module, the following values were set in the parameters as follows: Intensity tolerance: 20%; m/z tolerance: 0.02 or 10 ppm; RT tolerance: 0.20. The final feature list was carried forward for the annotation against the KEGG database. The final annotated features are exported for further analysis and visualization.

5.3 Results

5.3.1 Raw data import and preprocessing

The raw data files generated through LC-HRMS analysis (.d format) were converted into mzXML format and imported to MZmine3 software. The details of the samples used in the experiment, injection volume for LC-HRMS analysis, m/z range, the total number of scans and total ion chromatogram (TIC) detected are given in **Table 5. 1**. The total ion chromatogram of the raw data for each sample has been represented as a base peak intensity/retention time plot in **Figure 5. 2**.

Table 5. 1 Details of the sample used for LC-HRMS analysis and the experimental parameters.

Sample Name	Description	Injection Volume (µL)	m/z range	Retention time range (Minute)	Total No. of scan	Total Ion Chromatogram (TIC)
M-L	Mejankari leaf	0.1	100-2500	0.00-1.00	57	1.1e6
S-L	Som leaf					1.1e7
M-G	Gut of larvae reared on Mejankari plant					2.1e7
S-G	Gut of larvae reared on Som plant					1.6e7
M-SG	Silk gland of larvae reared on Mejankari plant					2.5e7
S-SG	Silk gland of larvae reared on Som plant					2.0e7

5.3.2 Feature detection

After chromatogram building from the raw data, the number of the total unique features identified in the leaf samples, M-L and S-L were 109 and 669, respectively. Similarly, in M-G, 647 and in S-G, 654 unique features were identified. The silk gland samples, M-SG and S-

SG were found to contain 737 and 674 unique features, respectively. The number of unique features retained after chromatogram deconvolution and deisotoping is represented in **Table 5. 2**.

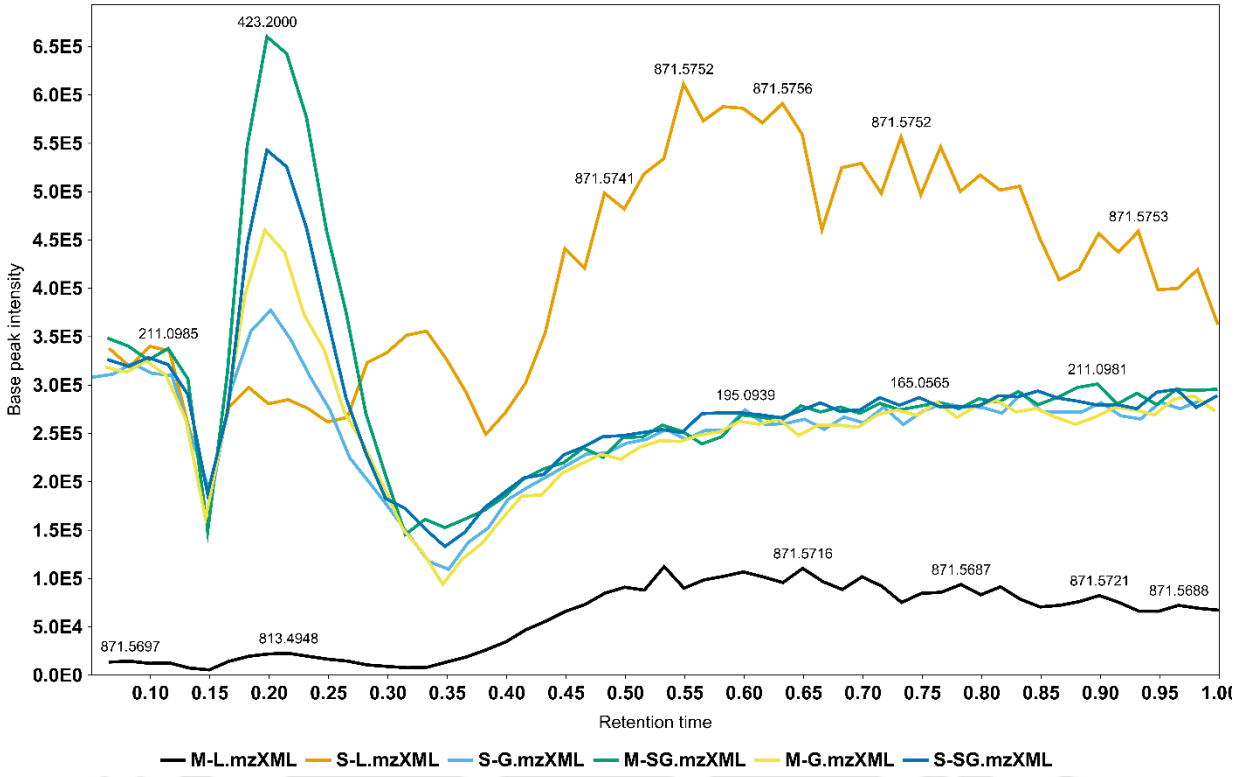


Figure 5. 2 Two-dimensional plot visualizing the base intensity peak/retention time of the raw data. All the chromatograms for the six samples are displayed in the same plot. The x-axis corresponds to retention time and the y-axis is the intensity level of the signal for each sample.

Table 5. 2 Details of the unique features identified in each sample

Sample	Number of chromatograms built	Feature retained after deconvolution	Feature retained after deisotoping	Total number of identified metabolites
M-L	652	324	109	96
S-L	1024	826	667	365
M-G	1204	749	647	363
S-G	1123	923	654	359
M-SG	1286	876	737	327
S-SG	1185	912	674	348

Principal component analysis (PCA) based on the identified unique features showed that

sample S-L showed the highest similarity with M-L. Similarly, S-G and S-SG shared the highest similarity with M-G and M-SG, respectively (**Figure 5. 3**).

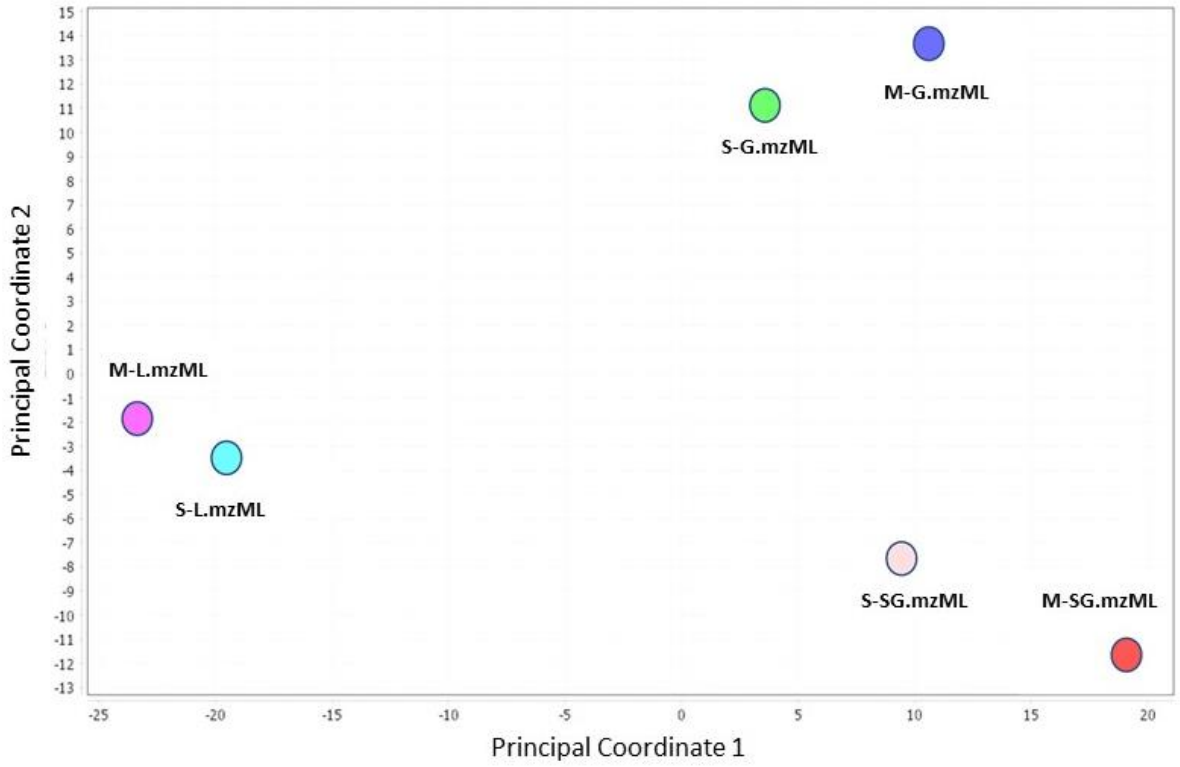


Figure 5. 3 Principal component analysis of the studied samples based on the identified unique features. The closer the samples placed to each other, the higher is the similarity between them. The PCA analysis showed that M-L shared the highest similarity with S-L, M-G shared with S-G and S-SG shared with M-SG.

The frequency distribution of unique features in different mass range is shown in **Figure 5. 4**. This shows that the frequency distribution is the highest in the m/z range of 635.00-830.00 in most of the samples (except S-L). the majority of the features were detected within the m/z range between 50.00-1000.00 for all the six samples.

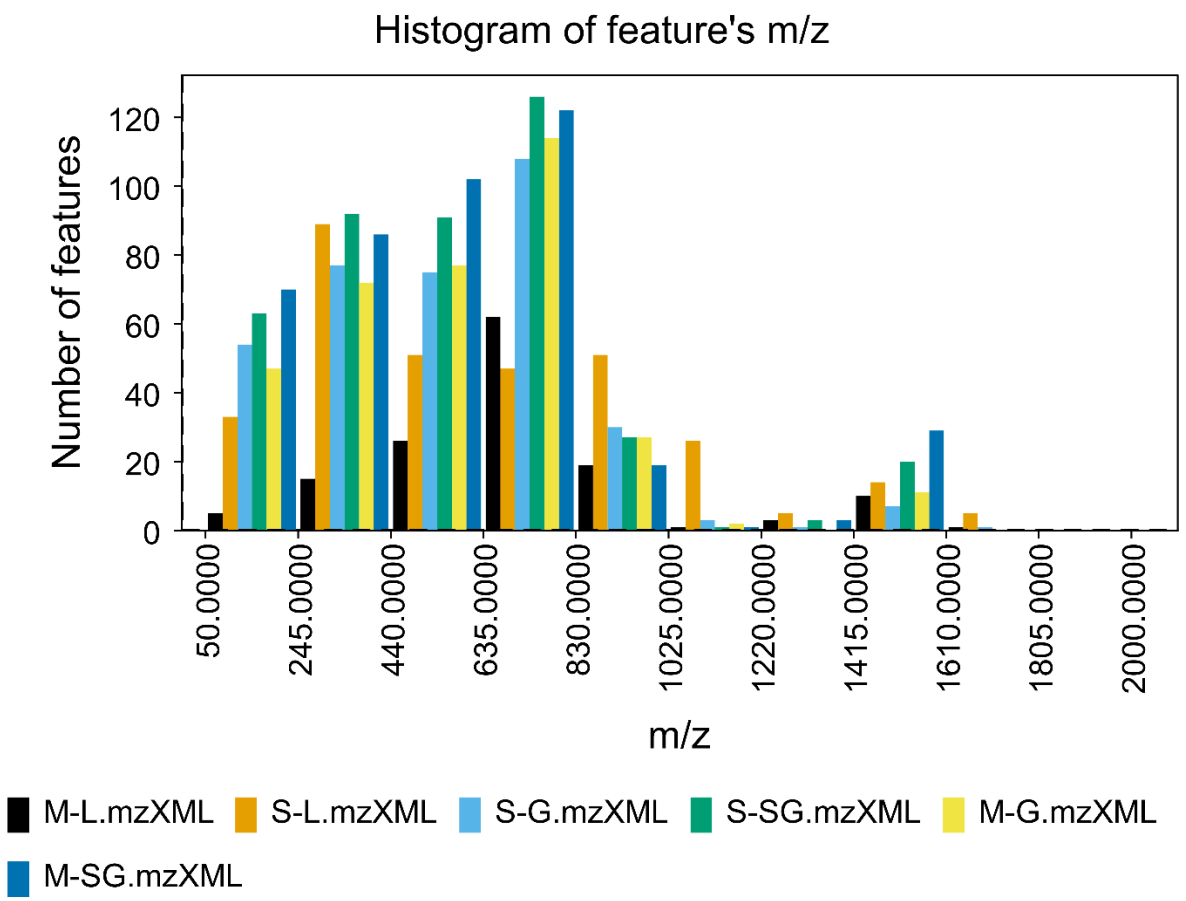
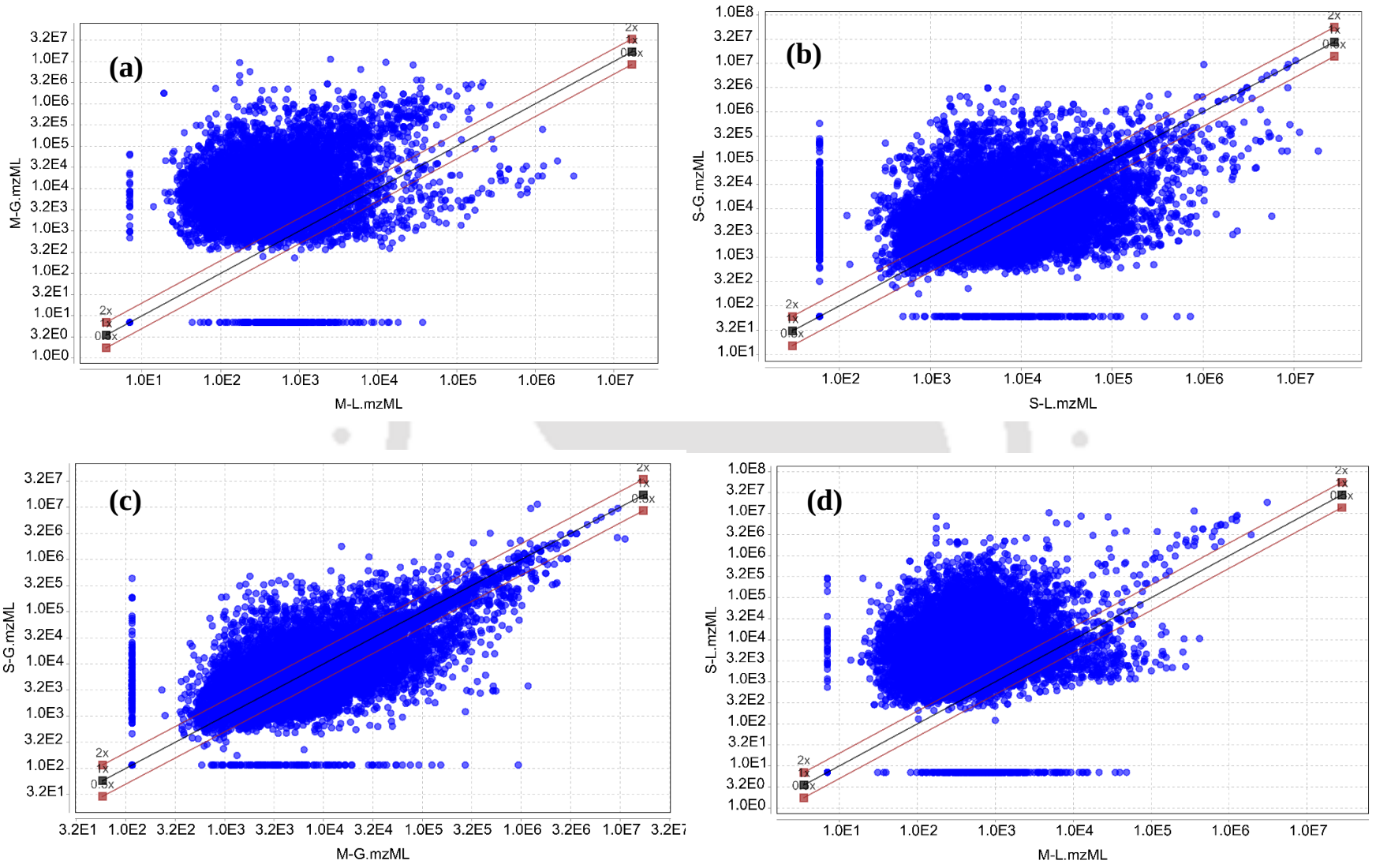


Figure 5. 4 Histogram showing the frequencies of features in different m/z ranges. This data is generated from raw data with a range from 50 to 2000 m/z.

The relationship among the unique features identified in between different sample sets is shown in **Figure 5. 5**. through scatter plots. A scatter plot shows each unique feature as a dot on a two-dimensional graph, with one sample represented on the X-axis and another represented on the Y-axis. The position of each dot on the plot corresponds to the values of the two samples for that particular feature. By examining the pattern of dots on the plot, it is possible to infer the relationship between the variables. For example, if the dots are clustered in a straight line, this suggests a strong linear relationship between the samples, while if they are scattered randomly, this suggests little or no relationship.



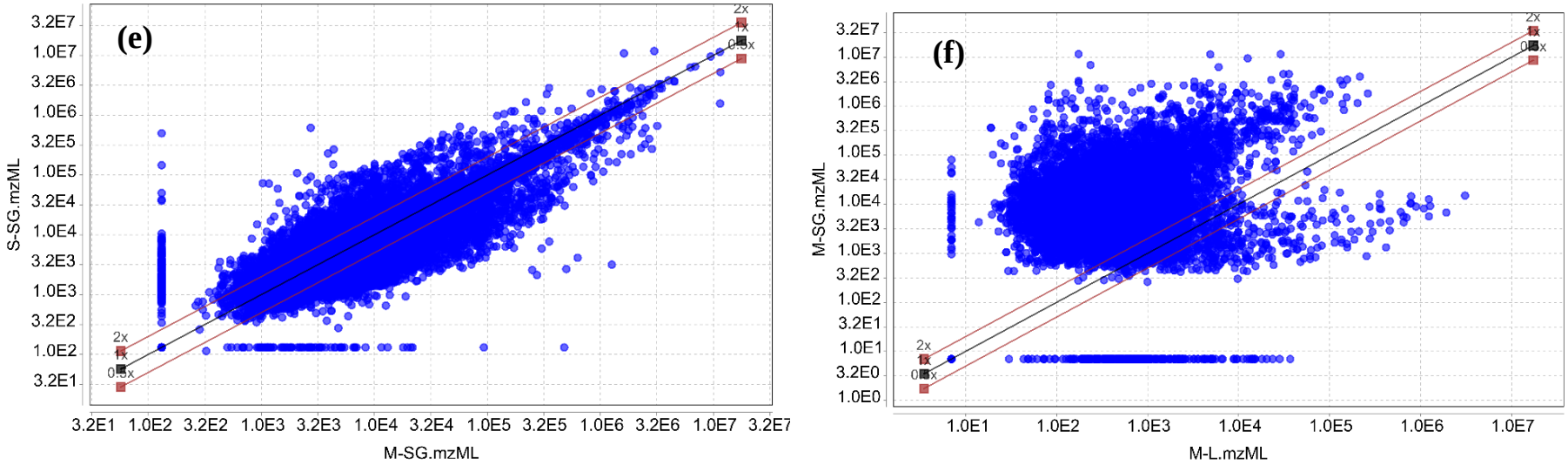


Figure 5. 5 Scatter plot displaying each unique feature as a dot on a two-dimensional graph, with one sample represented on the X-axis and another represented on the Y-axis on a log scale. The position of each blue dot on the plot corresponds to the values of the two samples for that particular feature. The straight lines (fold lines) indicate a 2-fold difference, which is a measure of the difference in the abundance of a feature between two samples, expressed as a ratio. The fold lines represent the threshold value above which the features are considered significantly different between the two samples. The scatter plots show the relationship among the unique features detected in (a) the gut of *A. assamensis* larvae reared on Mejankari plant (M-G) and host plant (Mejanakri) leaf (M-L), (b) the gut of larvae reared on Som plant and leaf of Som plant, (c) Gut of larvae reared on Som plant and gut of larvae reared on Mejankari plant, (d) leaf of Som plant and leaf of Mejankari plant, (e) silk gland of larvae reared on Som plant and silk gland of larvae reared on Mejankari plant, (f) silk gland of larvae reared on Mejankari plant and leaf of Mejankari plant. The two silk gland samples, i.e., the silk gland of larvae reared on the Som plant and Mejankari plant showed the highest similarities among their unique features.

5.3.3 Annotation

A total number of 6493 features were detected across the samples out of which 1028 metabolites were annotated against the KEGG online database. The identified metabolites were composed of 24% amino acids, 21% sugars, 13% organic acids, 9% fatty acid, 7% polyamines, 6% phosphoric acids, 5% polyol, 3% vitamins, 2% pigment, 1% steroids, 1% neurotransmitters and 8% others as represented in **Figure 5. 6**. The major categories identified in each sample have been discussed below.

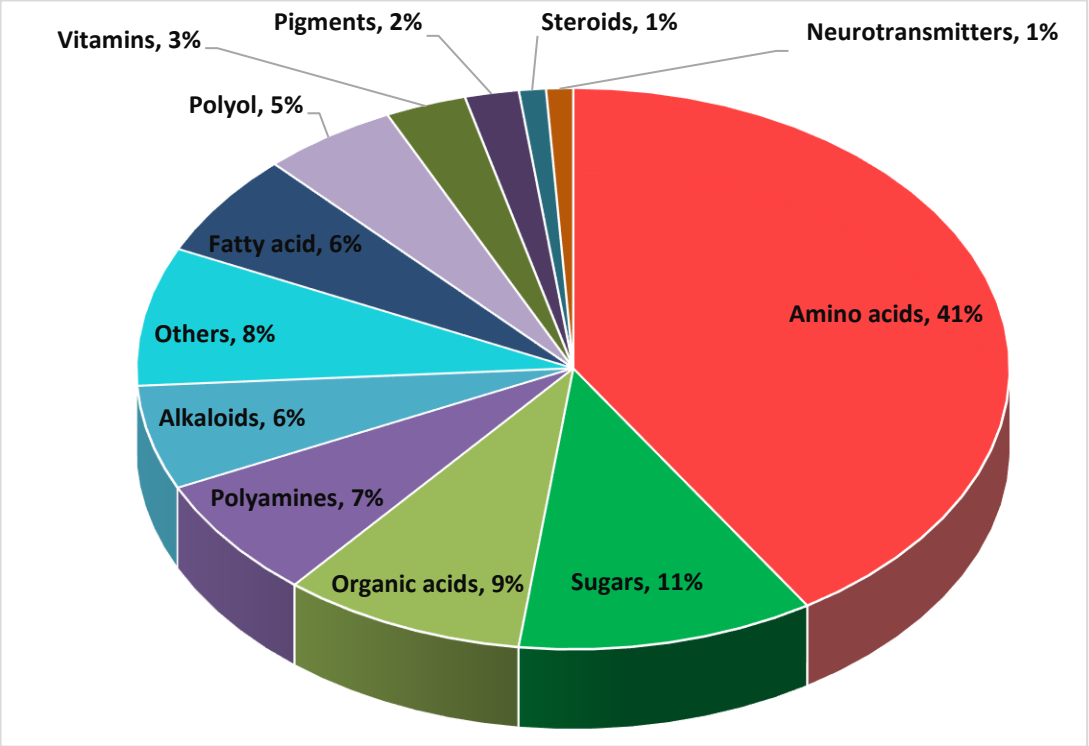


Figure 5. 6 The major classes of the identified metabolites across the samples. The highest abundant metabolites were amino acids followed by sugar molecules and fatty acids. Pigment molecules, neurotransmitters, steroids, vitamins were also annotated from some of the samples.

The number shared identified metabolites across the studied samples is represented in **Figure 5. 7**. The highest number of shared metabolites were observed between M-SG and S-SG (167 metabolites) followed by S-G and M-G (165 metabolites).

Sample	M-L	S-L	M-G	S-G	M-SG	S-SG
M-L	96					
S-L	32	365				
M-G	24	102	363			
S-G	29	115	165	359		
M-SG	31	95	132	123	327	
S-SG	38	99	128	117	167	348

Figure 5. 7 Number of shared identified metabolites across the samples.

5.3.3.1 Identification of Amino acids

The amino acids molecules were found to be largely present in all the samples. Amino acids likeL-Tyrosine, L-Tryptophanamide; 4-Hydroxy-L-threonine; 3,4-Dihydroxy-L-phenylalanine; N4-(Acetyl-beta-D-glucosaminyl) asparagine; histidylleucine and N(alpha)-Benzyloxycarbonyl-L-leucine were found to be present in most of the samples. The highest number of amino acids were detected from S-SG (total 23 amino acids) followed by S-G (18), M-G (18), M-SG (15) and S-L (11). Only 5 amino acids were detected in M-L. Further details of the amino acids identified in the studied samples are given in **Table 5. 3**.

5.3.3.2 Identification of carbohydrates

The major sugar molecules identified in the metabolomes of different samples were Menthyl O-beta-D-glucoside, Pelargonidin 3-O-glucoside, D-Glucosamine, Nuatigenin 3beta-D-glucopyranoside, Sterol 3-beta-D-glucoside, 1D-1-Guanidino-3-amino-1,3-dideoxy-scyllo-inositol, N-Acetyl-beta-D-glucosaminylamine, D-Glucosamine, Mannitol and Raffinose. The distribution across the samples is shown in **Table 5. 4**.

Table 5. 3 The major amino acid compounds identified in different samples

m/z	Retention time	Identified metabolite	M-L	S-L	M-G	S-G	M-SG	S-SG	Molecular Formula
273.175	0.868	(2R,3R)-3-Methylglutamyl-5-semialdehyde-N6-lysine	-	+	-	-	+	+	C12H23N3O4
187.137	0.685	(E)-2-Butenyl-4-methyl-threonine	-	-	+	-	-	-	C9H17NO3
761.533	0.654	1-Hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine	-	-	-	+	+	+	C40H76NO10P
803.563	0.543	2,3-Bis-(O-geranylgeranyl)-sn-glycero-1-phospho-L-serine	-	+	-	+	-	-	C46H78NO8P
199.082	0.679	3-(3,4-Dihydroxypyridin-1-yl)-L-alanine	-	-	+	-	+	+	C8H11N2O4
197.062	0.113	3,4-Dihydroxy-L-phenylalanine	-	+	+	+	-	+	C9H11NO4
195.091	0.904	3-Methyl-L-tyrosine	-	-	+		-	-	C7H9N5O2
135.047	0.113	4-Hydroxy-L-threonine	-	+	+	+	+	+	$C_4H_9NO_4$
245.137	0.121	beta-Alanyl-L-arginine	+	-	-	+	+	+	C9H19N5O3
131.067	0.879	cis-4-Hydroxy-L-proline	-	-	+	-	-	-	C4H9N3O2
160.078	0.118	D-Alanyl-D-alanine	-	-	+	+	+	-	C6H12N2O3
250.065	0.890	gamma-L-Glutamyl-L-cysteine	-	-	+	-	-	+	C8H14N2O5S
268.156	0.748	Histidylleucine	-	-	+	+	+	+	C12H20N4O3
323.140	0.635	L-4-Hydroxyphenylglycyl-L-arginine	-	-	-	+	-	-	C11H22N3O6P
174.127	0.729	L-Arginine	+	-	-	+	-	+	C6H14N4O2
147.072	0.665	L-Glutamate	+	+	+	-	+	+	C5H9NO4
115.074	0.854	L-Proline	-	-	-	-	-	+	C5H9NO2

168.967	0.693	L-Selenocysteine	-	-	-	-	+	+	C3H7NO2Se
105.059	0.568	L-Serine	-	-	-	+	-	-	C3H7NO3
119.064	0.821	L-Threonine	-	-	+	-	-	-	C4H9NO3
203.114	0.105	L-Tryptophanamide	-	+	+	+	+	+	C11H13N3O
181.068	0.701	L-Tyrosine	-	+	+	+	+	+	C9H11NO3
265.151	0.934	N(alpha)-Benzyloxycarbonyl-L-leucine	-	-	+	+	+	+	C14H19NO4
220.128	0.793	N'(omega)-Dihydroxy-N(omega)-methyl-L-arginine	-	-	-	-	-	+	C7H16N4O4
234.081	0.784	N-[(2S)-2-Amino-2-carboxyethyl]-L-glutamate	+	+	-	+	-	+	C8H12N2
335.149	0.204	N4-(Acetyl-beta-D-glucosaminyl)asparagine	-	+	+	+	+	+	C12H21N3O8
162.093	0.865	N6-Hydroxy-L-lysine	+	+	-	+	+	-	C7H16NO3
220.123	0.438	N-Acetyl-beta-D-glucosaminylamine	-	-	-	-	-	+	C8H16N2O5
189.070	0.834	N-Acetyl-L-glutamate	-	-	-	-	--	+	C7H11NO5
269.034	0.218	N-Acetyl-L-glutamate 5-phosphate	-	-	-	-	-	+	C7H12NO8P
209.057	0.599	N-Benzyloxycarbonylglycine	-	-	+	+	-	-	C10H11NO4
177.058	0.451	N-Formylmethionine	-	-	+	-	+	-	C6H11NO3S
148.064	0.821	O-Carbamoyl-L-serine	-	-	-	-	-	+	C6H12O4
261.045	0.313	Phosphotyrosine	-	+	-	+	-	-	C9H12NO6P
385.110	0.232	S-Inosyl-L-homocysteine	-	-	-	-	-	+	C14H19N5O6S

Note: presence: +, absence: -

Table 5. 4 The major sugar compounds identified in different samples

m/z	Retention time	Identified metabolite	M-L	S-L	M-G	S-G	M-SG	S-SG	Formula
318.199	0.882	(-)-Menthyl O-beta-D-glucoside	-	-	-	-	+	-	C16H30O6
179.071	0.857	D-Glucosamine	-	+	+	+	+	+	C6H13NO5
592.365	0.339	Nuatigenin 3beta-D-glucopyranoside	-	-	-	-	+	-	C33H52O9
410.260	0.085	Sterol 3-beta-D-glucoside	+	+	+	+	+	+	C23H38O6
220.128	0.793	1D-1-Guanidino-3-amino-1,3-dideoxy-scyllo-inositol	+	-	-	-	-	+	C7H16N4O4
220.123	0.438	N-Acetyl-beta-D-glucosaminylamine	-	-	-	-	-	+	C8H16N2O5
179.071	0.857	D-Glucosamine	-	+	+	+	+	+	C6H13NO5
182.077	0.149	Mannitol	-	-	-	+	+	+	C6H14O6
504.153	0.088	Raffinose	-	+	+	+	+	+	C18H32O16
335.145	0.410	N4-(Acetyl-beta-D-glucosaminyl)asparagine	-	-	+	-	-	-	C12H21N3O8

Note: presence: +, absence: -

5.3.3.3 Identification of pigment molecules

Biological pigment molecules are molecules found in living organisms that absorb and reflect light, giving them their characteristic colors. These pigments play important roles in biological processes such as photosynthesis, body colour formation, vision, and camouflage. In our study, more than twenty pigment molecules or their precursors were found to be present in the metabolome of the samples studied. The key pigment molecules identified in our study include phylloquinone, flavoxanthin, antheraxanthin, zeaxanthin, caloxanthin, myxol, pterin and others. The identified pigment molecules in the metabolomes of the gut leaf and silk gland samples is given in **Table 5. 5**.

5.3.3.4 Identification of other miscellaneous metabolites

Besides the amino acids, sugars and pigments, many other metabolites were identified which may play a crucial role in the biological processes of *Antheraea assamensis*. These metabolites are mainly antioxidants, alkaloids, antibacterial, antiviral or antifungal molecules, steroid hormones, secondary metabolites, and vitamins. The details of the identified metabolites belonging to the above-mentioned categories are given in Table **5.6**. The key alkaloids found in the leaf of both the host plants were, viz., Som and Mejankari were ancistrobrevine A, solacapine and tigloidine. Alkaloids identified from the larval gut metabolome include allocryptopine, lunarine, fabianine, acronycine, actinidine, and deoxytubulosine. Rapanone was of the major antioxidant molecule found in the majority of the samples. The larval gut samples were found to contain both rapanone and glutathione.

Table 5. 5 Major pigment molecules identified in the metabolomes of the studied samples

m/z	Retention time	Identified metabolites	M-L	S-L	M-G	S-G	M-SG	S-SG	Molecular Formula
430.287	0.784	4,4'-Diapolycopen-4-oate	+	+	-	-	-	-	C19H23N7O6
313.147	0.277	N-Feruloyltyramine	-		+	-	-	-	C22H29N7O5
237.103	0.504	Sepiapterin	+	+	-	+	+	-	C9H11N5O3
167.081	0.116	3-Methoxytyramine	-	-	-	+	-	-	C30H44O2
576.418	0.438	2-Hexaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinone	-	+	-	+	-	-	C28H37NO3
176.111	0.862	Canavanine;Serotonin	-	-	+	-	-	-	C16H17N
174.127	0.729	N-Methyltryptamine	-	-	-	+	-	-	C15H17N3
430.299	0.895	4,4'-Diapolycopen-4-oate	-	+	-	-	-	-	C13H21NO2
584.431	0.660	Flavoxanthin	-		+	-	+	+	C30H48
167.078	0.898	Pyridoxal	-	+	+	-	-	-	C31H48O2
138.059	0.610	para-Benzoquinone dioxime	-	+	-	-	-	-	C28H48O5
360.148	0.235	3-Methoxytyramine-betaxanthin	-	+	-	-	-	-	C21H21O11
568.422	0.729	Zeaxanthin	+	-	-	-	-	-	C21H21O12
584.423	0.216	Caloxanthin	-	-	+	-	+	+	C40H56O3
584.901	0.690	Myxol (xanthophyll)	-	-	+	-	+	+	C40H56O3
584.423	0.868	Capsanthin	-	-	+	-	+	+	C40H56O3
568.428	0.746	Lactucaxanthin	+	-	-	-	-	-	C40H56O2

568.428	0.895	Lutein	-	-	+	-	+	+	C40H56O2
584.423	0.610	Antheraxanthin	+	-	+	-	+	+	C40H56O3
195.091	0.904	6-(Hydroxymethyl)-7,8-dihydropterin	-	-	+	-	-	-	C7H9N5O2
163.038	0.396	Pterin	-	+		-	-	-	C6H5N5O
658.424	0.171	Fucoxanthin	-		+	-	-	-	C42H58O6
611.160	0.160	Cyanidin 3,5-O-diglucoside	-	+	-	-	-	-	C27H31O16
433.113	0.415	Pelargonidin 3-O-glucoside	+	-	-	+	-	-	C12137

Note: presence: +, absence: -

Table 5. 6 Some of the important identified metabolites belonging to different categories.

m/z	Retention time	Identified metabolites	Category	M-L	S-L	M-G	S-G	M-SG	S-SG	Molecular Formula
369.172	0.857	Allocryptopine	Alkaloid	-	-	-	+	-	-	C21H23NO5
435.257	0.857	Ancistrobrevine A		+	+	-	-	-	+	C27H33NO4
437.233	0.210	Lunarine		+	-	+	-	-	-	C26H31NO5
219.160	0.543	Fabianine		+	-	+	-	-	-	C14H21NO
432.378	0.865	Solacapine		+	+	-	-	-	-	C27H48N2O2
223.175	0.526	Tigloidine		+	+	-	-	-	+	C13H21NO2
174.134	0.299	N-Methyltryptamine		-	+	-	+	+	-	C27H33NO4
241.150	0.721	Valeroidine		+		-	-	-	+	C13H23NO3
189.070	0.834	6,7-Dimethoxyisoquinoline		-	-	-	-	-	+	C7H11NO5
321.127	0.479	Acronycine		-	-	+	-	-	-	C20H19NO3
147.110	0.346	Actinidine		-	-	+	+	+	-	C10H13N
283.171	0.243	Amabiline		-	-	-	-	+	+	C15H25NO4
459.276	0.895	Deoxytubulosine		-	-	+	+	+	-	C29H37N3O2
790.530	0.202	Oligomycin A	Antibiotc	-	-		+	-	-	C45H74O11
774.517	0.451	Oligomycin C		-	-	+	+	+	+	C45H74O10
776.492	0.254	Oligomycin D		-	-	+	+		-	C44H72O11
253.050	0.524	Sulfamethoxazole		+	-	-	-	+	+	C10H11N3O3S
338.229	0.934	Aphidicolin		-	+	+	+	-	-	C20H34O4

322.214	0.204	Rapanone	Antioxidant	-	+	+	+	+	+	C19H30O4
643.332	0.449	S-(PGA1)-glutathione		-	-	+	+	-	-	C30H49N3O10S
330.176	0.854	Carnosol		-	-	+		-	-	C20H26O4
253.127	0.091	Penciclovir	Antiviral	+	+	-	-	+		C10H15N5O3
281.168	0.615	Cycloheximide	Fungicide	-	-	+		+		C15H23NO4
137.075	0.596	Tyramine	Neurotransmitter	-	-	+	+	-	+	C8H11NO
148.070	0.937	trans-Cinnamate	Organic compound	-	+	-	-	-	+	C6H12O4
180.135	0.873	3-tert-Butyl-5-methylcatechol	Secondary metabolite	-	+	+	+	-		C11H16O2
139.043	0.621	4-Nitrophenol		-	-	-	-	-	+	C6H5NO3
139.044	0.371	4-Nitrophenol		-	-	-	-	-	+	C6H5NO3
421.259	0.901	Terpendole B		+	+	-	-	-	-	C27H35NO3
307.196	0.321	Dihydrocapsaicin			+	-	+	-	+	C21H25NO
618.417	0.241	Oleanolate 3-O-beta-D-glucoside		-	+	-		-	-	C36H57O8-
462.282	0.784	3-Dehydroecdysone	Steroid	+	-	+	-	-	-	C27H42O6
496.319	0.218	Polypodine B			-	+	-	-	-	C54H88O15
415.351	0.293	Tomatidine		+	+	-	-	-	-	C27H45NO2
464.360	0.130	Castasterone		+	+	-	-	-	-	C28H48O5
396.328	0.360	Ergosterol	Sterol	-	-	-	+	-	-	C28H44O

436.335	0.107	2-Phytyl-1,4-naphthoquinone	Vitamin	+	+	+	+	+	+	C05201
183.068	0.096	4-Pyridoxate		-	-	+	+	-	-	C8H9NO4
438.334	0.332	Demethylphylloquinol		+	-	-	-	+		C30H46O2
452.360	0.149	Phylloquinol		+	-	-	-	+	+	C31H48O2
286.217	0.565	Retinol		-	+	-	-			C20H30O
412.340	0.363	Vitamin D analogue		+	-	-	-	+	+	C28H44O2
249.050	0.177	Pyridoxine phosphate		+	-	+	-	-	-	C8H12NO6P

Note: presence: +, absence: -

5.4 Discussion

5.4.1 The gut metabolome of *Antheraea assamensis* larvae

The total number of identified metabolomes in the gut of larvae reared on Mejankari (M-G) and Som plants (S-G) were 363 and 359, respectively. These metabolomes can be classified into different categories a s described in **Figure 5. 6**. The major amino acid molecules found in the metabolomes of both gut samples were 4-hydroxy-L-threonine, L-tryptophanamide, L-tyrosine, N4-(acetyl-beta-D-glucosaminyl) asparagine, 3,4-dihydroxy-L-phenylalanine, histidylleucine, N(alpha)-benzyloxycarbonyl-L-leucine, D-alanyl-D-alanine and N-benzyloxycarbonylglycine (**Table 5. 3**). 4-hydroxy-L-threonine is an amino acid that has a hydroxyl (-OH) group attached to the fourth carbon atom of the L-threonine molecule. It is an essential amino acid that is used in the biosynthesis of proteins, and the addition of a hydroxyl group to its structure creates a new compound with unique properties. 4-hydroxy-L-threonine is an obligatory precursor molecule of vitamin B_6 metabolism (Tazoe et al., 2002; Wolf et al., 1995; Zhao, 1996). Tyrosine is an amino acid on which insects extensively rely for cuticle hardening and the melanization of pathogens as a crucial part of the insect immune system (Sterkel & Oliveira, 2017). 3,4-dihydroxy-L-phenylalanine, also known as L-DOPA, is an amino acid and a precursor to the neurotransmitter dopamine (Meiser et al., 2013). Across invertebrate and vertebrate species, dopamine plays conserved roles in the regulation of locomotion, pleasure, motivation, arousal, and memory. It is essential for the processes of melanization and sclerotization, which are required for immunological function and the development of an insect's exoskeleton (Verlinden, 2018).

D-glucosamine, a sugar molecule found in the larval gut (**Table 5. 4**) plays an important role in the development and growth of insects (J. K. Chen et al., 2011; Merzendorfer & Zimoch, 2003b; Ohta et al., 2014). It is a precursor molecule for chitin, which is a structural polysaccharide that forms the exoskeleton of insects and other arthropods. It has also been found to be present inside the silk gland of the larvae. Chitin provides rigidity and protection to the insect

exoskeleton, allowing it to support the weight of the insect and protect it from predators and environmental stressors. The synthesis of chitin requires the enzymatic conversion of D-glucosamine to N-acetylglucosamine, which is then polymerized into chitin (Merzendorfer & Zimoch, 2003a). Sterol 3-beta-D-glucoside (S3G), a sugar molecule identified from the larval gut (**Table 5. 4**) is a type of sterol glucoside, which is a class of plant-derived compounds (Ferrer et al., 2017; Shimamura, 2020) that have been found to play a variety of roles in insects. S3G is functionally related to ergosterol and has also been shown to have antifungal and antibacterial properties (Neumann et al., 2010; Normile et al., 2020), which can help to protect insects from pathogenic microorganisms. This may be particularly important in insects that feed on plant tissues, which can harbor a variety of fungal and bacterial pathogens. Another important sugar molecule found in most of the studied samples including the larval guts is raffinose. It is a trisaccharide composed of galactose, glucose, and fructose (Sengupta et al., 2015). This oligosaccharide is mainly found in plants and is often used as a carbohydrate reserve. Insects that feed on these plants have evolved the ability to digest and utilize raffinose as a source of energy (Carneiro et al., 2004; Harvey, 1974; Peterbauer et al., 2001).

Besides the amino acid and sugar molecules, many other crucial metabolites have been identified from the gut of *A. assamnesis* larvae such as alkaloids; antibiotics, antiviral or antifungal compounds; steroids, pigment molecules and vitamins (**Table 5.6**). Actinidine, a natural cyclopentanoid monoterpene alkaloid derived from a five-carbon isoprene unit was found to be present in the gut of the larvae reared on both the host plants. It was reported earlier to be present in plants and insect species containing high amount of iridoid compounds (Shi et al., 2020). It was found to show defensive properties in several rove beetle species (Chow & Lin, n.d.; Francke & Schulz, 1999). Oligomycin C is a natural product and a member of the oligomycin family of macrolide antibiotics. It is produced by certain strains of *Streptomyces* bacteria and has been found to have antibacterial, antifungal, and antitumor activities (Chakraborty et al., 2020; Dame et al., 2016). Another compound, tyramine is a neuroactive

compound found in insects that works as neuromodulator, neurotransmitter and neurohormone (Finetti et al., 2021; Ryglewski et al., 2017). 2-Phytyl-1,4-naphthoquinone, commonly known as vitamin K is a phylloquinone that consists of 1,4-naphthoquinone bearing a phytyl group at position 2 (Ryall & Silverman, 2006). It was found to be contained by many edible insects. However, its functional role in the biological processes of insects is obscure.

5.4.2 The leaf metabolome of two host plants of *A. assamensis*

In the leaf metabolome of Som plant, a total of 365 metabolites were identified in our study. However, we could identify only 96 metabolites in the leaf of the Mejankari plants. The number of amino acid compounds identified in the leaves was smaller in comparison to the larval gut and silk gland metabolome (**Table 5. 3**). On the other hand, the leaf samples were found to contain a significant number of alkaloids, secondary metabolites, pigment molecules and vitamins (kindly refer **Table 5. 5** and **Table 5. 6**). Pelargonidin 3-O-glucoside, is a natural compound that belongs to a class of pigments called anthocyanins. It is commonly found in various plant sources, such as fruits, vegetables, and flowers, and is responsible for their red, purple, and blue colours (Su et al., 2020). This compound has been detected in the leaf of the Som plant in our study. 2-Phytyl-1,4-naphthoquinone or vitamin K is found to be present in all the studied samples including the host plant leaves. 4,4'-Diapolycopen-4-oate is a carotenoid molecule that belongs to the family of apocarotenoids (Weedon, 2013). It is derived from the carotenoid lycopene, which is a red pigment found in many fruits and vegetables, such as tomatoes, watermelon, and grapefruit. 4,4'-Diapolycopen-4-oate was studied for its potential health benefits, including antioxidant and anti-inflammatory properties (Simkin, 2021). It was also found to have a protective effect against oxidative stress. Studies have also suggested that 4,4'-Diapolycopen-4-oate may have a role in regulating immune function and preventing inflammation (Bohn, 2019). Another pigment molecule antheraxanthin was found to be present only on the leaf metabolome of the Mejankari plant (M-L). Antheraxanthin is a type of xanthophyll pigment that belongs to the carotenoid family. It is found in certain plants, algae,

and bacteria, where it plays a role in photosynthesis. Antheraxanthin is a yellow pigment that absorbs light in the blue-green range of the spectrum and reflects yellow light (Sun et al., 2020; Xie et al., 2013). On the other hand, pterin, a yellowish-red pigment that is synthesized by various organisms, including plants, bacteria, and humans (Basu & Burgmayer, 2011; Feirer & Fuqua, 2017; Noiriel et al., 2007) was found only in the leaf of Som plant but not in the leaf of Mejankari.

5.4.3 The silk gland metabolome of *Antheraea assamensis*

The silk produced by *Antheraea assamensis* is unique among other insect silks in terms of its colour, texture, strength and durability.(Chetia et al., 2017; Mohanty et al., 2006; Tikader et al., 2013). Interestingly, larvae reared on the Mejankari plant (*Litsea citrata*) produces cocoons with a shining white colour instead of the usual golden yellow cocoon (Devi et al., 2021). From the metabolomic analysis of the silk gland of 5th instar larvae of *A. assamensis*, we have identified 327 metabolites in the silk gland of larvae reared on the Mejankari plant and 348 in that of the larvae reared on the Som plant. Both the silk glands were found to be rich in amino acids, sugars, pigments, alkaloids and other metabolites (kindly refer to **Table 5. 3- Table 5. 7**). The major amino acid molecules identified in both the silk glands are (2R,3R)-3-methylglutamyl-5-semialdehyde-N6-lysine; 1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine; 3-(3,4-dihydroxypyridin-1-yl)-L-alanine; 4-hydroxy-L-threonine; histidylleucine; L-glutamate; L-selenocysteine and L-tyrosine (**Table 5. 3**). However, there were several amino acids molecules found to be present either in M-SG or S-SG. For example, 3,4-dihydroxy-L-phenylalanine; gamma-L-Glutamyl-L-cysteine; L-arginine; L-proline; N-acetyl-beta-D-glucosaminylamine; and O-carbamoyl-L-serine. were found only in the silk gland of larvae reared on the Som plant (S-SG). However, Only D-alanyl-D-alanine and N6-hydroxy-L-lysine were found to be present in the silk gland of larvae reared on the Mejankari plant (M-SG). Silk protein has a distinctive amino acid composition where glycine, alanine, serine, and tyrosine make up more than 85% of the total amino acid content (X. Chen et al.,

2022). A change in diet was reported to change the amino acid composition of *Bombyx mori* silk (Craig et al., 2000). The amino acid composition of Muga silk is well studied. However, one research carried by Hazarika and Saikia, found that the concentration of amino acid content varies with respect to the host plant (Craig et al., 2000). In our study, amino sugar such as D-glucosamine, and sugars like mannitol and raffinose were also identified in both the silk gland samples. Sugars are an important component of insect silk glands (Shimada, 1979). Bombyx mori silk gland was found to contain trehalose (Shimada et al., 1980). However, no trehalose was identified in any of the silk glands.

The key pigment molecules identified in both the silk glands are antheraxanthin, lutein, capsanthin, myxol, caloxanthin and flavoxanthin (**Table 5. 5**). However, except sepiapterin, a yellow-coloured pteridine pigment (Gao et al., 2013) observed only in M-SG, no other distinction was observed in the pigment composition between the silk glands of larvae reared on Som and Mejankari plant.

5.5 Conclusion

The key focus of this objective was to identify the metabolites present in the gut and silk gland of *Antheraea assamensis* 5th instar healthy larvae and its variation with respect to the leaf metabolome of host plants. Metabolomics is the study of small molecules, called metabolites, produced by cellular metabolism. It has become an increasingly popular field of research in recent years because it provides a comprehensive view of the metabolic status of an organism or a biological sample. However, the metabolomic analysis presents several challenges such as the complexity and diversity of the metabolome, a dynamic range of metabolite concentrations and variability in sample preparation.

The identification of unknown metabolites is a major challenge in metabolomic analysis. Despite the advancements in metabolomic techniques, many metabolites remain unidentified or poorly characterized. In our experiment, we tried to qualitatively elucidate the larval gut, silk gland and host plant leaf metabolomes of *Anthearaea assamensis* larvae reared on two

different host plants. The results of this experiment identified hundreds of metabolites in the studied samples through LC-HRMS analysis which would expand our knowledge of the metabolite composition of the studied samples. Many of the metabolites may be the microbial origin. Multi-omic approaches are necessary to draw a broader and better picture of the host-microbiome-metabolome interaction. Future studies for the quantification of the metabolites and their structural and functional characterization through liquid chromatography-mass spectrometry, NMR and other allied techniques, will further extend our knowledge in this regard.

References

Alvares, A. P. (1978). Research review. Interactions between environmental chemicals and drug biotransformation in man. *Clinical Pharmacokinetics*, *3*(6), 462–477. https://doi.org/10.2165/00003088-197803060-00004

Basu, P., & Burgmayer, S. J. N. (2011). *Pterin chemistry and its relationship to the molybdenum cofactor*. https://doi.org/10.1016/j.ccr.2011.02.010

Bedia, C. (2022). Metabolomics in environmental toxicology: Applications and challenges. *Trends in Environmental Analytical Chemistry*, *34*, e00161. https://doi.org/10.1016/J.TEAC.2022.E00161

Bohn, T. (2019). Carotenoids and markers of oxidative stress in human observational studies and intervention trials: Implications for chronic diseases. In *Antioxidants* (Vol. 8, Issue 6). MDPI. https://doi.org/10.3390/antiox8060179

Cao, Y. Y., Peng, L. L., Jiang, L., Thakur, K., Hu, F., Tang, S. M., & Wei, Z. J. (2020). Evaluation of the Metabolic Effects of Hydrogen Sulfide on the Development of Bombyx mori (Lepidoptera: Bombycidae), Using Liquid Chromatography-Mass Spectrometry-Based Metabolomics. *Journal of Insect Science*, *20*(2). https://doi.org/10.1093/JISESA/IEAA008

Carneiro, C. N. B., Isejima, E. M., Samuels, R. I., & Silva, C. P. (2004). Sucrose hydrolases from the midgut of the sugarcane stalk borer Diatraea saccharalis. *Journal of Insect Physiology*, *50*(11), 1093–1101. https://doi.org/10.1016/J.JINSPHYS.2004.09.011

Chakraborty, M., Mahmud, N. U., Muzahid, A. N. M., Fajle Rabby, S. M., & Islam, T. (2020). Oligomycins inhibit Magnaporthe oryzae Triticum and suppress wheat blast disease. *PLoS ONE*, *15*(8). https://doi.org/10.1371/JOURNAL.PONE.0233665

Chen, J. K., Shen, C. R., Yeh, C. H., Fang, B. S., Huang, T. L., & Liu, C. L. (2011). N-Acetyl Glucosamine Obtained from Chitin by Chitin Degrading Factors in Chitinbacter

tainanesis. *International Journal of Molecular Sciences*, *12*(2), 1187. https://doi.org/10.3390/IJMS12021187

Chen, Q., Liu, X., Zhao, P., Sun, Y., Zhao, X., Xiong, Y., Xu, G., & Xia, Q. (2015). GC/MS-based metabolomic studies reveal key roles of glycine in regulating silk synthesis in silkworm, Bombyx mori. *Insect Biochemistry and Molecular Biology*, *57*, 41–50. https://doi.org/10.1016/J.IBMB.2014.12.007

Chen, X., Ye, A., Wu, X., Qu, Z., Xu, S., Sima, Y., Wang, Y., He, R., Jin, F., Zhan, P., Cao, J., & Zhou, W. (2022). Combined analysis of silk synthesis and hemolymph amino acid metabolism reveal key roles for glycine in increasing silkworm silk yields. *International Journal of Biological Macromolecules*, *209*, 1760–1770. https://doi.org/10.1016/J.IJBIOMAC.2022.04.143

Chetia, H., Kabiraj, D., Singh, D., Mosahari, P. V., Das, S., Sharma, P., Neog, K., Sharma, S., Jayaprakash, P., & Bora, U. (2017). De novo transcriptome of the muga silkworm, Antheraea assamensis (Helfer). *Gene*, *611*, 54–65. https://doi.org/10.1016/j.gene.2017.02.021

Chow, Y. S., & Lin, Y. M. (n.d.). *ACTINIDINE, A DEFENSIVE SECRETION OF STICK INSECT, MEGACRANIA ALPHEUS WESTWOOD (ORTHOPTERA: PHASMATIDAE)*. Retrieved March 2, 2023, from http://meridian.allenpress.com/jes/article-pdf/21/2/97/1552100/0749-8004-21_2_97.pdf

Clish, C. B. (2015a). *Metabolomics: an emerging but powerful tool for precision medicine*. https://doi.org/10.1101/mcs.a000588

Clish, C. B. (2015b). Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harbor Molecular Case Studies*, *1*(1), a000588. https://doi.org/10.1101/MCS.A000588

Craig, C. L., Riekel, C., Herberstein, M. E., Weber, R. S., Kaplan, D., & Pierce, N. E. (2000). Evidence for Diet Effects on the Composition of Silk Proteins Produced by Spiders.

Molecular Biology and Evolution, *17*(12), 1904–1913. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A026292

Dame, Z. T., Islam, M. T., Helmke, E., von Tiedemann, A., & Laatsch, H. (2016). Oligomycins and pamamycin homologs impair motility and induce lysis of zoospores of the grapevine downy mildew pathogen, Plasmopara viticola. *FEMS Microbiology Letters*, *363*(16), 167. https://doi.org/10.1093/FEMSLE/FNW167

Devi, B., Chutia, M., & Bhattacharyya, N. (2021). Food plant diversity, distribution, and nutritional aspects of the endemic golden silk producing silkworm, Antheraea assamensis – a review. *Entomologia Experimentalis et Applicata*, *169*(3), 237–248. https://doi.org/10.1111/EEA.13021

Dong, H. L., Zhang, S. X., Tao, H., Chen, Z. H., Li, X., Qiu, J. F., Cui, W. Z., Sima, Y. H., Cui, W. Z., & Xu, S. Q. (2017). Metabolomics differences between silkworms (Bombyx mori) reared on fresh mulberry (Morus) leaves or artificial diets. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-11592-4

Feirer, N., & Fuqua, C. (2017). Pterin function in bacteria. *Pteridines*, *28*(1), 23–36. https://doi.org/10.1515/PTERID-2016-0012/ASSET/GRAPHIC/J_PTERID-2016-0012_FIG_002.JPG

Ferrer, A., Altabella, T., Arró, M., & Boronat, A. (2017). Emerging roles for conjugated sterols in plants. *Progress in Lipid Research*, *67*, 27–37. https://doi.org/10.1016/J.PLIPRES.2017.06.002

Finetti, L., Roeder, T., Calò, G., & Bernacchia, G. (2021). The Insect Type 1 Tyramine Receptors: From Structure to Behavior. *Insects*, *12*(4), 315. https://doi.org/10.3390/INSECTS12040315

Francke, W., & Schulz, S. (1999). Pheromones. *Comprehensive Natural Products Chemistry*, 197–261. https://doi.org/10.1016/B978-0-08-091283-7.00052-7

Gao, J., Wang, J., Wang, W., Liu, C., & Meng, Y. (2013). Isolation, Purification, and

Identification of an Important Pigment, Sepiapterin, from Integument of the lemon Mutant of the Silkworm, Bombyx mori. *Journal of Insect Science*, *13*, 118. https://doi.org/10.1673/031.013.11801

Goswami, A., & Devi, D. (2021). Structural insight on the liquid silk from the middle silk gland of non-mulberry silkworm Antheraea assamensis. *Https://Doi.Org/10.1080/07391102.2021.2017347*. https://doi.org/10.1080/07391102.2021.2017347

Harvey, G. T. (1974). NUTRITIONAL STUDIES OF EASTERN SPRUCE BUDWORM (LEPIDOPTERA: TORTRICIDAE): I. SOLUBLE SUGARS. *The Canadian Entomologist*, *106*(4), 353–365. https://doi.org/10.4039/ENT106353-4

Ibarra-Estrada, E., Palma-Tenango, R. M. S.-H. and M., Ibarra-Estrada, E., & Palma-Tenango, R. M. S.-H. and M. (2016). Metabolomics as a Tool in Agriculture. *Metabolomics - Fundamentals and Applications*. https://doi.org/10.5772/66485

Kim, S. J., Kim, S. H., Kim, J. H., Hwang, S., & Yoo, H. J. (2016). Understanding Metabolomics in Biomedical Research. *Endocrinology and Metabolism*, *31*(1), 7. https://doi.org/10.3803/ENM.2016.31.1.7

Lankadurai, B. P., Nagato, E. G., & Simpson, M. J. (2013). Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environmental Reviews*, *21*(3), 180–205. https://doi.org/10.1139/ER-2013-0011/ASSET/IMAGES/ER-2013-0011TAB3.GIF

Lelli, V., Belardo, A., Timperio, A. M., Lelli, V., Belardo, A., & Timperio, A. M. (2021). From Targeted Quantification to Untargeted Metabolomics. *Metabolomics - Methodology and Applications in Medical Sciences and Life Sciences*. https://doi.org/10.5772/INTECHOPEN.96852

Meiser, J., Weindl, D., & Hiller, K. (2013). Complexity of dopamine metabolism. *Cell Communication and Signaling*, *11*(1), 1–18. https://doi.org/10.1186/1478-811X-11-

34/FIGURES/5

Merzendorfer, H., & Zimoch, L. (2003a). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *The Journal of Experimental Biology*, *206*(Pt 24), 4393–4412. https://doi.org/10.1242/JEB.00709

Merzendorfer, H., & Zimoch, L. (2003b). Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology*, *206*(24), 4393–4412. https://doi.org/10.1242/JEB.00709

Moghadam, N. N., Thorshauge, P. M., Kristensen, T. N., de Jonge, N., Bahrndorff, S., Kjeldal, H., & Nielsen, J. L. (2018). Strong responses of Drosophila melanogaster microbiota to developmental temperature. *Fly*, *12*(1), 1. https://doi.org/10.1080/19336934.2017.1394558

Mohanty, N., Das, H. K., Mohanty, P., & Mohanty, E. (2006). Modification of Muga Silk by Methyl-Methacrylate. II. *Http://Dx.Doi.Org/10.1080/10601329508019151*, *32*(sup7), 1103–1111. https://doi.org/10.1080/10601329508019151

Muñoz-Benavent, M., Pérez-Cobas, A. E., García-Ferris, C., Moya, A., & Latorre, A. (2021). Insects' potential: Understanding the functional role of their gut microbiome. *Journal of Pharmaceutical and Biomedical Analysis*, *194*, 113787. https://doi.org/10.1016/J.JPBA.2020.113787

Neumann, A., Baginski, M., & Czub, J. (2010). How do sterols determine the antifungal activity of amphotericin B? Free energy of binding between the drug and its membrane targets. *Journal of the American Chemical Society*, *132*(51), 18266–18272. https://doi.org/10.1021/JA1074344/SUPPL_FILE/JA1074344_SI_001.PDF

Noiriel, A., Naponelli, V., Gregory, J. F., & Hanson, A. D. (2007). Pterin and Folate Salvage. Plants and Escherichia coli Lack Capacity to Reduce Oxidized Pterins. *Plant Physiology*, *143*(3), 1101–1109. https://doi.org/10.1104/PP.106.093633

Normile, T. G., McEvoy, K., & del Poeta, M. (2020). Steryl Glycosides in Fungal Pathogenesis:

An Understudied Immunomodulatory Adjuvant. *Journal of Fungi*, *6*(1). https://doi.org/10.3390/JOF6010025

Ohta, T., Ido, A., Kusano, K., Miura, C., & Miura, T. (2014). A Novel Polysaccharide in Insects Activates the Innate Immune System in Mouse Macrophage RAW264 Cells. *PLoS ONE*, *9*(12). https://doi.org/10.1371/JOURNAL.PONE.0114823

Oliveira, N. C., Phelan, L., Labate, C. A., & Cônsoli, F. L. (2022). Non-targeted metabolomics reveals differences in the gut metabolic profile of the fall armyworm strains when feeding different food sources. *Journal of Insect Physiology*, *139*, 104400. https://doi.org/10.1016/J.JINSPHYS.2022.104400

Perez de Souza, L., Alseekh, S., Naake, T., & Fernie, A. (2019). Mass Spectrometry-Based Untargeted Plant Metabolomics. *Current Protocols in Plant Biology*, *4*(4), e20100. https://doi.org/10.1002/CPPB.20100

Peterbauer, T., Lahuta, L. B., Blöchl, A., Mucha, J., Jones, D. A., Hedley, C. L., Gòrecki, R. J., & Richter, A. (2001). Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. *Plant Physiology*, *127*(4), 1764–1772. https://doi.org/10.1104/PP.010534

Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., & Picotti, P. (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell*, *172*(1–2), 358-372.e23. https://doi.org/10.1016/J.CELL.2017.12.006

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*(1), 1–11. https://doi.org/10.1186/1471-2105-11-395/TABLES/3

Raza, M. F., Wang, Y., Cai, Z., Bai, S., Yao, Z., Awan, U. A., Zhang, Z., Zheng, W., & Zhang, H. (2020). Gut microbiota promotes host resistance to low-temperature stress by

stimulating its arginine and proline metabolism pathway in adult Bactrocera dorsalis. *PLOS Pathogens*, *16*(4), e1008441. https://doi.org/10.1371/JOURNAL.PPAT.1008441

Rinschen, M. M., Ivanisevic, J., Giera, M., & Siuzdak, G. (2019). Identification of bioactive metabolites using activity metabolomics. *Nature Reviews Molecular Cell Biology 2019 20:6*, *20*(6), 353–367. https://doi.org/10.1038/s41580-019-0108-4

Ríos Peces, S., Díaz Navarro, C., Márquez López, C., Caba, O., Jiménez-Luna, C., Melguizo, C., Carlos Prados, J., Genilloud, O., Vicente Pérez, F., & Pérez del Palacio, J. (2017). *Untargeted LC-HRMS-Based Metabolomics for Searching New Biomarkers of Pancreatic Ductal Adenocarcinoma: A Pilot Study*. https://doi.org/10.1177/1087057116671490

Roberts, L. D., Souza, A. L., Gerszten, R. E., & Clish, C. B. (2012). Targeted Metabolomics. *Current Protocols in Molecular Biology*, *CHAPTER*(SUPPL.98), Unit30.2. https://doi.org/10.1002/0471142727.MB3002S98

Roques, S., Deborde, C., Guimas, L., Marchand, Y., Richard, N., Jacob, D., Skiba-cassy, S., Moing, A., & Fauconneau, B. (2020). Integrative Metabolomics for Assessing the Effect of Insect (Hermetia illucens) Protein Extract on Rainbow Trout Metabolism. *Metabolites*, *10*(3). https://doi.org/10.3390/METABO10030083

Ryall, R. P., & Silverman, R. B. (2006). Synthesis of 2-Substituted-3-Phytyl-1,4-naphthoquinone Epoxides. *Http://Dx.Doi.Org/10.1080/00397919008052785*, *20*(3), 431–438. https://doi.org/10.1080/00397919008052785

Ryglewski, S., Duch, C., & Altenhein, B. (2017). Tyramine actions on Drosophila flight behavior are affected by a glial dehydrogenase/reductase. *Frontiers in Systems Neuroscience*, *11*, 68. https://doi.org/10.3389/FNSYS.2017.00068/BIBTEX

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted metabolomics strategies – Challenges and Emerging Directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897.

https://doi.org/10.1007/S13361-016-1469-Y

Schmid, R., Heuckeroth, S., Korf, A., Smirnov, A., Myers, O., Dyrlund, T. S., ... & Pluskal, T. (2023). Integrative analysis of multimodal mass spectrometry data in MZmine 3. Nature biotechnology, 41(4), 447-449.

Sehnal, F., & Akai, H. (1990). Insect silk glands: their types, development and function, and effects of environmental factors and morphogenetic hormones on them. *International Journal of Insect Morphology and Embryology*, *19*(2), 79–132. https://doi.org/10.1016/0020-7322(90)90022-H

Sehnal, F., & Sutherland, T. (2008). Silks produced by insect labial glands. *Prion*, *2*(4), 145. https://doi.org/10.4161/PRI.2.4.7489

Sengupta, S., Mukherjee, S., Basak, P., & Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Frontiers in Plant Science*, *6*(AUG), 656. https://doi.org/10.3389/FPLS.2015.00656/BIBTEX

Shimada, S. (1979). Glucose Enzymatically appearing from silk gland homoseuutes of the silkworm, Bombyx mori. *Agricultural and Biological Chemistry*, *43*(2), 203–208. https://doi.org/10.1080/00021369.1979.10863435

Shimada, S., Kido, M., Kamada, A., & Asano, S. (1980). Trehalose in the silk glands of the silkworm, Bombyx mori. *Insect Biochemistry*, *10*(2), 175–177. https://doi.org/10.1016/0020-1790(80)90070-0

Shimamura, M. (2020). Structure, metabolism and biological functions of steryl glycosides in mammals. *Biochemical Journal*, *477*(21), 4243. https://doi.org/10.1042/BCJ20200532

Shi, Q., He, Y., Chen, J., & Lu, L. (2020). Thermally Induced Actinidine Production in Biological Samples. *Journal of Agricultural and Food Chemistry*, *68*(44), 12252–12258. https://doi.org/10.1021/ACS.JAFC.0C02540/ASSET/IMAGES/LARGE/JF0C02540_0006.JPEG

Simkin, A. J. (2021). Carotenoids and Apocarotenoids in Planta: Their Role in Plant

Development, Contribution to the Flavour and Aroma of Fruits and Flowers, and Their Nutraceutical Benefits. *Plants 2021, Vol. 10, Page 2321*, *10*(11), 2321. https://doi.org/10.3390/PLANTS10112321

Snart, C. J. P., Hardy, I. C. W., & Barrett, D. A. (2015). Entometabolomics: applications of modern analytical techniques to insect studies. *Entomologia Experimentalis et Applicata*, *155*(1), 1–17. https://doi.org/10.1111/EEA.12281

Späth, J., Brodin, T., McCallum, E., Cerveny, D., Fick, J., & Nording, M. L. (2022). Metabolomics reveals changes in metabolite profiles due to growth and metamorphosis during the ontogeny of the northern damselfly. *Journal of Insect Physiology*, *136*, 104341. https://doi.org/10.1016/J.JINSPHYS.2021.104341

Sterkel, M., & Oliveira, P. L. (2017). Developmental roles of tyrosine metabolism enzymes in the blood-sucking insect Rhodnius prolixus. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854). https://doi.org/10.1098/RSPB.2016.2607

Su, H., Xie, L., Xu, Y., Ke, H., Bao, T., Li, Y., & Chen, W. (2020). Pelargonidin-3- O-glucoside Derived from Wild Raspberry Exerts Antihyperglycemic Effect by Inducing Autophagy and Modulating Gut Microbiota. *Journal of Agricultural and Food Chemistry*, *68*(46), 13025–13037. https://doi.org/10.1021/ACS.JAFC.9B03338/ASSET/IMAGES/LARGE/JF-2019-03338U_0008.JPEG

Sun, K. M., Gao, C., Zhang, J., Tang, X., Wang, Z., Zhang, X., & Li, Y. (2020). Rapid formation of antheraxanthin and zeaxanthin in seconds in microalgae and its relation to non-photochemical quenching. *Photosynthesis Research*, *144*(3), 317–326. https://doi.org/10.1007/S11120-020-00739-6/FIGURES/4

Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y., & Merritt, D. J. (2009). Insect Silk: One Name, Many Materials. *Https://Doi.Org/10.1146/Annurev-Ento-112408-085401*, *55*, 171–188. https://doi.org/10.1146/ANNUREV-ENTO-112408-085401

Tazoe, M., Ichikawa, K., & Hoshino, T. (2002). Biosynthesis of Vitamin B6 in Rhizobium: In Vitro Synthesis of Pyridoxine from 1-Deoxy-D-xylulose and 4-Hydroxy-L-threonine. *Bioscience, Biotechnology, and Biochemistry*, *66*(4), 934–936. https://doi.org/10.1271/BBB.66.934

Tikader, A., Vijayan, K., & Saratchandra, B. (2013). Muga silkworm, antheraea assamensis (Lepidoptera: Saturniidae) - An overview of distribution, biology and breeding. *European Journal of Entomology*, *110*(2), 293–300. https://doi.org/10.14411/eje.2013.096

Verlinden, H. (2018). Dopamine signalling in locusts and other insects. *Insect Biochemistry and Molecular Biology*, *97*, 40–52. https://doi.org/10.1016/J.IBMB.2018.04.005

Vinayavekhin, N., & Saghatelian, A. (2010). Untargeted metabolomics. In *Current Protocols in Molecular Biology* (Issue SUPPL. 90). https://doi.org/10.1002/0471142727.mb3001s90

Wang, X., Li, Y., Liu, Q., Tan, X., Xie, X., Xia, Q., & Zhao, P. (2019). GC/MS-based metabolomics analysis reveals active fatty acids biosynthesis in the Filippi's gland of the silkworm, Bombyx mori, during silk spinning. *Insect Biochemistry and Molecular Biology*, *105*, 1–9. https://doi.org/10.1016/J.IBMB.2018.12.009

Wolf, E., Hill, R. E., Sayer, B. G., & Spenser, I. D. (1995). 4-Hydroxy-L-threonine, a committed precursor of pyridoxol (vitamin B6). *Journal of the Chemical Society, Chemical Communications*, *9*(13), 1339–1340. https://doi.org/10.1039/C39950001339

Wu, X., Chen, X., Ye, A., Cao, J., He, R., Pan, M., Jin, F., Ma, H., & Zhou, W. (2022). Multi-tissue metabolomic profiling reveals potential mechanisms of cocoon yield in silkworms (Bombyx mori) fed formula feed versus mulberry leaves. *Frontiers in Molecular Biosciences*, *9*, 852. https://doi.org/10.3389/FMOLB.2022.977047

Xie, X., Gao, S., Gu, W., Pan, G., & Wang, G. (2013). Desiccation Induces Accumulations of Antheraxanthin and Zeaxanthin in Intertidal Macro-Alga Ulva pertusa (Chlorophyta).

PLOS ONE, *8*(9), e72929. https://doi.org/10.1371/JOURNAL.PONE.0072929

Yang, L., Zhao, Z., Luo, D., Liang, M., & Zhang, Q. (2022). Global Metabolomics of Fireflies (Coleoptera: Lampyridae) Explore Metabolic Adaptation to Fresh Water in Insects. *Insects*, *13*(9), 823. https://doi.org/10.3390/INSECTS13090823/S1

Yang, Q., Zhang, A. H., Miao, J. H., Sun, H., Han, Y., Yan, G. L., Wu, F. F., & Wang, X. J. (2019). Metabolomics biotechnology, applications, and future trends: a systematic review. *RSC Advances*, *9*(64), 37245–37257. https://doi.org/10.1039/C9RA06697G

Zhang, A., Sun, H., Yan, G., Wang, P., & Wang, X. (2015). Metabolomics for Biomarker Discovery: Moving to the Clinic. *BioMed Research International*, *2015*. https://doi.org/10.1155/2015/354671

Zheng, X., Yu, J., Cairns, T. C., Zhang, L., Zhang, Z., Zhang, Q., ... & Ma, Y. (2019). Comprehensive improvement of sample preparation methodologies facilitates dynamic metabolomics of Aspergillus niger. Biotechnology Journal, 14(3), 1800315.

Zhao, G. (1996). 4-Phospho-hydroxy-l-threonine is an obligatory intermediate in pyridoxal 5′-phosphate coenzyme biosynthesis in Escherichia coli K-12. *FEMS Microbiology Letters*, *135*(2–3), 275–280. https://doi.org/10.1016/0378-1097(95)00466-1

Weedon, B. C. L. (Ed.). (2013). Carotenoids–4: Main Lectures Presented at the Fourth International Symposium on Carotenoids, *Berne*, Switzerland, 25-29 August 1975. Elsevier.

Chapter 6

Summary and Prospects

CHAPTER 6

Summary and Prospects

Antheraea assamensis Helfer is an endemic Lepidopteran insect that produces a unique type of silk known as Muga silk. Its distribution is confined to the northeastern part of India and some of its adjoining areas. This silk is highly valued for its natural golden colour, durability, tensile strength and shine and is considered one of the most expensive natural fibers in the world. Moreover, the Muga silk's fibroin component is a promising material for tissue engineering. The gut microbiota refers to the diverse community of microorganisms that live in the digestive tract of animals, including both vertebrates and invertebrates. These microbes play several important roles in host biology including digestion and absorption of nutrients, development of the host organism's immune system, protection against pathogens and regulation of host metabolism. In the present study, we have identified and characterized the gut microbial community of *Antheraea assamensis*, its variation with change in the host plant and the impact of host leaf-associated microbiota on larval gut microbial composition through marker gene analysis. A Comprehensive investigation to understand the functional potential of the gut microbiome was carried out through shotgun metagenomic sequencing. The gene expression profiling and identification of the active gut microbiota were conducted through metatranscriptomic analysis of the gut of 5^{th} instar healthy larvae of *A. assamensis*. Additionally, an untargeted metabolic analysis was A chapter-wise summary of the thesis work is given below.

6.1 Structural analysis of the gut metagenome of *A. assamensis* and its comparison with host leaf-associated microbiota

In this chapter, we reported the gut microbial and archaeal community of *A. assamensis*, the impact of host leaf-associated microbiota on gut microbial composition and how the gut microbial community changes with the change in the host plant. The predictive functional profiling of the gut microbiota and host leaf-associated microbiota was carried out through

PICRUSt2 analysis. This study is the first-ever culture-independent analysis of the gut microbial community of *A. assamensis*. The results of this analysis showed that the 5th instar larva of *A. assamensis* contains a highly diverse and unique gut microbial community in comparison to other lepidopteran species. The host leaf-associated microbiota occupied a major portion of the total larval gut microbiota. However, the microbial diversity of the larval gut was observed to be higher than the host leaf-associated microbiota. Besides a core set of microbiota, intraspecific variation of the gut microbial community was also observed among larvae reared separately on three different host plants, namely, *Machilus bombycina*, *Litsea polyantha* and *Litsea citrata*. Predictive functional analysis of the gut microbiome revealed several key interactions between the host and its gut microbial community including host leaf digestion, metabolite detoxification, chitinase production and fat body metabolism.

6.2 Functional analysis of *Antheraea assamensis* gut metagenome.

Here, we described the functional potential of the gut microbial community of *A. assamensis* larvae through shotgun metagenomic sequencing. For this, we have examined the larvae raised on *Machilus bombycina*, the most frequently used host plant for silk rearing to produce commercial Muga silk. The results of this study have demonstrated the functional roles of the gut microbial population on host biology. A total of 72063 coding sequences were annotated from the assembled shotgun sequence reads from which 270 enzymes were identified with their EC numbers. Further, 1003 orthologous genes belonging to 24 COG categories were identified using the coding sequences. The identified enzymes were utilized to find 405 and 100 metabolic pathways through MinPath against the MetaCyc and KEGG databases, respectively. The assembled shotgun sequences revealed 97% Bacteria, 2.4% Archaea, and 0.06% Viruses upon taxonomic classification. The shotgun metagenomic analysis revealed thousands of microbial species with better resolution than the marker gene analysis. Functional identification of the metagenome revealed several key functions of the larval gut microbiota

which may help in the vital physiological processes of *A. assamensis*.

6.3 Chapter 4: Gene expression profiling of *A. assamensis* gut metagenome

In this chapter, we have reported the metabolically active gut microbial community of *A. assamensis* larvae and their gene expression profiling through metatranscriptomic analysis. For the metatranscriptome analysis, total RNA from the gut of 5^{th} instar larvae of A. assamensis was isolated. polyA-mRNA enriched libraries were prepared with the help of TruSeq RNA Library Preparation Kit and sequenced. Metatranscriptome annotation of the assembled transcripts resulted in the identification of ~7000 bacterial and 2708 archaeal active protein-coding genes. Taxonomic annotation of the assembled transcripts identified 1091 bacterial and 490 archaeal taxa at the genus level. In this analysis, several key metabolically active functions with potential benefits towards the host organism such as food digestion, detoxification of plant secondary metabolite, resistance to the oxidative stress inside the gut, and vitamin synthesis have been reported. This would help us to differentiate between the beneficial and harmful active gut microbes of a healthy larva, their adaptation and their mechanism of action inside the gut environment. In addition, interactome analysis considering biomolecules from the host gut and residential microbiota at different biological conditions will give a broader detail about the protein-protein interactions, signaling pathways, disease occurrence, differential gene expression etc. in the future.

6.4 Untargeted metabolomic analysis of *Antheraea assamensis*

Metabolites are the end products of cellular processes, and their levels can provide insight into the metabolic state of an organism. Untargeted metabolomics of insect gut refers to the comprehensive analysis of the metabolites present in the gut of insects without prior knowledge of their identity. In this chapter, we described the global metabolome of the host plant leaf, gut, and silk gland of *A. assamensis* larvae in their fifth instar, as well as how changes in the host plant affect the metabolite composition of the gut and silk gland. For this, we took into

consideration the larvae raised on Som and Mejankari plants. The results of this experiment identified hundreds of metabolites from the larval gut, silk gland and host plant leaf. The identified metabolites can be clustered into different categories like amino acids, sugars, fatty acids, alkaloids, secondary metabolites, pigments vitamins and others. Change in host plant was found to change the metabolite composition of the larval gut and the silk gland. The finding of this study will help to extend our knowledge of the metabolomes of the studied samples, which have not been reported to date.

Prospects

- Identification and functional characterization of the fungal community and unicellular eukaryotes inhabiting the gut of *A. assamnesis* will elucidate their taxonomic identification and interaction with the host plant.
- Microbial community analysis during different developmental stages of *A. assamensis* can provide valuable information for understanding the microbial community dynamics and the interactions between the host and its gut microbiota during molting where the insect undergoes significant physiological and morphological changes.
- Analysis of temporal, spatial and disease-related variation in the composition and function of the gut microbiota will expand our knowledge of the gut microbiome of *A. assamensis* to a substantial extent.
- We have found that methanogenic archaea are an integral component of the gut microbial community of *A. assamensis*. Further research is necessary to identify if *A. assamensis* emits methane at any of its developmental stages and its contribution towards global methane emission.

- Comparative metatranscriptomic analysis of healthy and diseased larvae can provide a comprehensive understanding of the functional changes occurring in microbial communities associated with diseased and healthy hosts. These findings can help to develop probiotics and to identify new targets for treating major bacterial, fungal and viral diseases of *A. assamensis*.
- Quantification of the global metabolome of different body parts and physiological conditions and their relation with the host microbial community will further extend our knowledge of the holobiont

List of publication

Research Paper

1. Structural and functional analyses of the gut microbial community of Antheraea assamensis Helfer and its comparison with host leaf-associated microbiota (Under review)
2. De novo metetranscriptomic exploration of the gut microbiome of *A. assamensis* Helfer (Under preparation)

Book Chapter:

1. Ojha, Sunita, **Dharitri Saikia,** and Utpal Bora. "Nanopharmaceuticals: Synthesis, Characterization, and Challenges." *Nanopharmaceuticals: Principles and Applications Vol. 3*. Springer, Cham, 2020. 81-138.
2. Chetia, Hasnahana, Debajyoti Kabiraj, Biju Bharali, Sunita Ojha, Manash Pratim Barkataki, **Dharitri Saikia**, Tinka Singh, Ponnala Vimal Mosahari, Pragya Sharma, and Utpal Bora. "Exploring the benefits of endophytic fungi via omics." In *Advances in Endophytic Fungal Research*, pp. 51-81. Springer, Cham, 2019.

List of conference/Workshop/Symposia

Conferences:

1. Poster presented at SeriTech 2017 at IIT Guwahati
2. Submitted abstract and poster at conference on "Evolutionary Emergence of Life Cycles", Max Planck Intstitute for Evolutionary Biology, Germany, October 2018

Curriculum vitae

DHARITRI SAIKIA

Centre for the Environment
Indian Institute of Technology Guwahati,
Kamrup 781039, Assam, India
Phone: +91-8876663258
dharitrisaikia@iitg.ac.in

I seek a challenging career, which provides an opportunity to showcase my talent & experience toward meeting organizational goals and thereby climb the career ladder.

Education

PhD	IIT Guwahati, Centre for The Environment	Presented Synopsis
MSc	Gauhati University, Botany (8.26)	Passed: 2016
BSc	Darrang College, Gauhati University, Botany (8.7)	Passed: 2014
HSSLC	Darrang College, Gauhati University, (70%)	Passed: 2011
HSLC	Chariduar Higher Secondary School (85%)	Passed: 2009

Honors and Awards

Qualified GATE 2017

Permanent address

C/O: Jadav Saikia
Village: Chariduar, Near Nevil ME School
PO: Chariduar; District: Sonitpur, Assam-784103

Research Experience

After qualifying my masters in 2016, I have worked as a project fellow at Tezpur Central University and IIT Guwahati where I got the opportunity to learn about the foundations of scientific research. I qualified GATE in 2017 in life sciences and joined as PhD student at Centre for the Environemnt, IIT Guwahati where I have accumulated a strong background in

several research fields that will help to ensure my future success. Over the last six years, I have gained proficiency in high-throughput sequence analysis, Metagenomic and Metatranscriptomic analysis and visualization, DNA isolation from environmental samples, PCR amplification, cloning, microbiological techniques, cell culture and Nanoparticle synthesis.

Work Experience

1. Project Title: "Sequencing Genomes of Some Bacteria That Invades/Resides in Tomato Plant."

Institute: Department of Molecular Biology and Biotechnology (MBBT), Tezpur University, Tezpur-784028

Project Investigator: Dr. S.K Ray, Professor

Position: Project Assistant

Duration: 01/08/2016 To 09/10/2016

2. Project Title: "Exploration and Characterization of Seri-Bio resources of North East India for potential textile and non-textile applications"

Department of Biosciences and Bioengineering,

Indian Institute of Technology Guwahati, Assam-781039

Project Investigator: Dr. Utpal Bora, Professor

Position: Junior Rescarch Fellow

Duration: December 2016 to May 2017

3. Worked as a Teaching Assistant in NPTEL online certificate course entitled "Genome Editing and Engineering" from July to October 2022.

Publications

Book Chapter

1. Ojha, Sunita, **Dharitri Saikia**, and Utpal Bora. "Nanopharmaceuticals: Synthesis, Characterization, and Challenges." Nanopharmaceuticals: Principles and Applications Vol. 3. Springer, Cham, 2020. 81-138.

2. Chetia, Hasnahana, Debajyoti Kabiraj, Biju Bharali, Sunita Ojha, Manash Pratim Barkataki, **Dharitri Saikia,** Tinka Singh, Ponnala Vimal Mosahari, Pragya Sharma, and Utpal

Bora. "Exploring the benefits of endophytic fungi via omics." In Advances in Endophytic Fungal Research, pp. 51-81. Springer, Cham, 2019.

Conferences/Symposium

1. Submitted abstract and poster at conference on “Evolutionary Emergence of Life Cycles”, Max Planck Intstitute for Evolutionary Biology, Germany, October 2018
2. Poster presented at SeriTech 2017 at IIT Guwahati
3. Kalita, J. J., Kabiraj, D., Bharali, B., Saikia, D., Bora, U. Applications of Aptamer conjugated nanomaterials in biosensing and therapeutics. International Conference on Advances in Nanotechnology (ICAN-2017), January, 2017

Professional Training

Training Area: R-Programming
Institute: Coursera from Johns Hopkins University
Duration: 4 Week online credit course

Symposium/Conference/Seminar/Workshop Co-Organizer/Member

6. Kalita, J. J., Kabiraj, D., Bharali, B., Saikia, D., Bora, U. Applications of Aptamer conjugated nanomaterials in biosensing and therapeutics. International Conference on Advances in Nanotechnology (ICAN-2017), January, 2017.
7. Submitted abstract and poster at conference on “Evolutionary Emergence of Life Cycles”, Max Planck Intstitute for Evolutionary Biology, Germany, October 2018.
8. Submitted abstract on “Traditional Practices of Assam: A conservation Perspective” in Biodiverse-2018, IIT Guwahati.
9. Poster presented at SeriTech 2017 at IIT Guwahati.
10. Best oral presentation award (technical session 2, venue 4) on the topic “Understanding the Gut microbial Community of Lepidoptera through Meta-omics” in “International Conference on Challenges and Prospects of Bioresource Conservation in Eastern Himalaya with Special Reference to UN-Sustainable Development Goal” organized by Department of Zoology, Gauhati University. 3-4 May 2023.