Analyzing the gut microbiome of *Samia cynthia ricini*

A thesis submitted

in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

By

Biju Bharali

Centre for the Environment
Indian Institute of Technology Guwahati
Kamrup 781039, Assam, India
2023

Analyzing the gut microbiome of *Samia cynthia ricini*

A thesis submitted

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

by

Biju Bharali
Roll Number: 176152007

Under the supervision of

Prof. Utpal Bora, *PhD*
Department of Biosciences & Bioengineering, IIT Guwahati
&
Prof. Karuna Kalita, *PhD*
Department of Mechanical Engineering, IIT Guwahati

Centre for the Environment

Indian Institute of Technology Guwahati

Kamrup 781039, Assam, India

2023

Dedication

To my mother

Dalimi Pachani Bharali


भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati

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 "**Analyzing the gut microbiome of *Samia cynthia ricini***" is the result of investigation carried out by me under the supervision of Prof. Utpal Bora and Prof. Karuna Kalita 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 acknowledgement has been made wherever the work of other investigators is referred.

Biju Bharali

Guwahati
2nd March 2023

Biju Bharali
Roll No. – 176152007
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 "**Analyzing the gut microbiome of *Samia cynthia ricini***" is the result of investigation carried out by **Biju Bharali (Roll No. – 176152007)** under our supervision at 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 our knowledge and belief.

6/3/2023

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

Prof. Karuna Kalita
Centre for the Environment
Indian Institute of Technology Guwahati,
Kamrup 781039, Assam, India

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 construct it into a beautiful one.

I am extremely grateful and indebted to my thesis supervisors Prof. Utpal Bora and Prof. Karuna Kalita for giving me the opportunity to pursue my dream of efficient research understanding significances of microbes in the environment and to explore a new aspect of omics and data science. I thank Prof. Bora for introducing me to the various paths of environmental research, and enriching my knowledge regarding the importance of balance and preserving the environment. He introduced me to the world of academic research and taught me adept academic and other life skills. I'm deeply grateful to Prof. Kalita who continuously encouraged me to explore multiple approach to the field and provided me with many invaluable advices. I also would like to thank both Prof. Bora and Prof. Kalita for allowing me to conduct my research independently and for teaching me scientific communication.

I would like thank my doctoral committee members Prof. Ranjan Tamuli, Prof. Mihir Kr. Purkait, Prof. Subhendu Sekhar Bag and former member of my doctoral committee Prof. Chandan Mahanta for their valuable suggestions, motivation and scientific guidance which always helped me to make my work better.

I would also like to convey my gratitude to the Centre for the Environment and the Institutional Biotech Hub at the Centre for the Environment for providing me with all the necessary facilities to pursue my research.

I sincerely acknowledge the financial support from Ministry of Human Resource Development (MHRD), Government of India for providing me fellowship as well as Department of Biotechnology (DBT), Government of India and Central Silk Board for funding our laboratory; Param-Ishan, IIT Guwahati’s high-performance computing cluster and Central Instrumentation Facility, IITG for providing me the facilities to pursue my research. I would further like to extend my gratitude towards Dr. Kartig Neog, Central Muga Eri Research and Training Institute (CMER&TI), Jorhat (Assam) for providing me with necessary samples to conduct my research.

A note of gratitude to my closest friends Hasnahana, Vimal, Deepanjali, Nitin, and Pallav for always being by my side, cheering and inspiring throughout. A big thanks to my seniors Arghya, Papori, Sunita, Kabiraj, Jon, for their guidance and companionship; to my labmates Adhiraj, Dharitri, Pulak for their all-round support; my project mentor Dr. Ashish Prabhu for the admirable mentorship and inspiration.

I express my sincere gratitude to my parents Atul Bharali & Dalimi Pachani Bharali, and my siblings Manas and Mousumi for their overwhelming love, support and patience for all these years. I would like to thank my aunt Lata and aunt Rupa and uncle Bhago for their support and care. I would also like to thank my grandparents Lt. Ratneswar Bharali, Lt. Asha Bharali, Lt. Rajendra Pachani and Smt. Minu Pachani for their constant blessings.

\- Biju

TABLE OF CONTENTS

SYNOPSIS

Organisms, currently recognised as metaorganism are defined by the eukaryotic organism coevolved with the microbiota that shapes its state of being. Every metaorganism comprises a complex network of organisms and their metabolites are synergistically interdependent to survive in the environment along with modifying it. Organisms and their microbiome co-exist and have parallel evolutionary histories represented by co-phylogenetic patterns and are reported to have undergone co-speciation.

Holometabolous insects are a major influence in the ecosystem They occupy multiple niches in the ecosystem forming completely different phenotypes in the larval and adult stages. With ~18000 described species over the world, lepidopterans are the second most diverse insect group. However, only <0.1% of lepidopterans have been reported of their microbial assistance. The composition and significance of gut microbiome varies vastly among lepidopterans with no universal pattern.

As the phenotype and physiology of insects are dependent on the microbial composition, the composition and abundance of the organism-associated microbiota, especially the gut are dependent on the insect habitat, immunity and other physiochemical conditions. Studies have reported variations in the lepidopteran gut microbiome with changes in environmental conditions. Most studies report the complete remodel of the gut microbiota according to the development changes and have provided different theories about the gut microbiome composition, transmission and coevolution of these organisms.

Studies based on functional analyses of the gut microbiome of lepidopterans reported 85-97% of active species. Early larval stages reported a higher presence of functions involved in carbohydrate metabolism, and cell motility while late instar microbes assisted with functions involved in amino acid, cofactors and vitamin metabolism. In moths, functions such as energy metabolism, and replication pathways were more abundant.

Studying the gut microbiota is used to many advantages. In addition to gaining insights into the biology, evolution and dynamics of the host-gut microbiome, insect gut microbiota analyses could be used in developing effective methods for their better growth and health. They could be used as biomarkers for various traits of the organism and could be used to determine and eliminate various unwanted traits like diseases, pests, weak immunity and reduced growth. Novel microbes and their products from insect guts have also been developed for the biodegradation of complex waste and for blocking the transmission of insect-borne diseases.

Samia cynthia ricini (Lepidoptera: Saturniidae), a domesticated silkworm, producing unique warm Eri-silk is widely cultivated by the rural population of Assam and Northeast India and largely contributes to the ecological and economic aspects of these regions. However, very little is known about the respective organisms resulting in their poor and obsolete rearing practice and cultivation techniques and low productivity of silk. The health conditions of these organisms are weakening leading to a decrease in their population which also has negative effects on the environment due to their role in the ecosystem.

In an attempt to address the research gap in these areas, the present work explores the analysis of *S. cynthia ricini* concerning its gut microbiota

The following objectives were formulated to gain a deeper understanding of

1. Comparative Analyses of *S. cynthia ricini* and its gut microbiome under different diet conditions and growth stages.

2. Shotgun metagenome analysis of *S. cynthia ricini* 5th instar larva.

3. Comparative analyses with other lepidopteran larva metagenomes.

4. Metabolic profiling of *S. cynthia ricini* based on different diets and analysis the of the associated microbiome.

Chapter 1 of the thesis reviews the literature on the organism used in our study, and the significance of studying the host-microbiome association. It composes of previous studies carried out to understand the relationship better and the research gaps present.

Chapter 2 describes the identification gut microbiota structure of *S. cynthia ricini* through amplicon sequencing of 16s rRNA gene sequence comparative analyses of its gut microbiome under different diet conditions and growth stages. Changes in the gut microbiome affected by the environmental conditions are explored along with their role in the changes in the phenotypic characteristics Diversity analyses including α and β-diversity were analyzed. Differential abundances were calculated through the taxonomic and functional composition. Significant changes and patterns through the various categories were observed.

Objectives of the study include-

- Understanding the gut microbiota of *S. cynthia ricini* through amplicon sequencing and establishing the core microbiota involved.
- To analyze the changes in the gut microbiota structure through various diet conditions and growth stages.

- To predict the functions of the microbes involved in the phenotype changes through the changes in diet and growth stages.
- To optimise the methodology of analysing the amplicon-based gut microbiome.

Chapter 3 focuses on shotgun-based analyses of the metagenome of *S. cynthia ricini* 5th instar larva. 16s rRNA gene sequence microbiome-based microbiome analyses provide only the taxonomic abundance of an environment. It is more appropriate for composition profiling. Shotgun-based analyses, however, provide the total genomic DNA from a sample. Using gene coverage, species-level genetic variation and taxonomic composition along with functional profiling could be analyzed. Comparative analyses of annotation and identification of the gut microbiota through the techniques used were summarised.

Objectives of the study include-

- Determining the optimum conditions and tools for *S. cynthia ricini* metagenome annotation and assembly.
- Taxonomic and functional annotation of the sequenced metagenome.
- Reconstruction of metabolic pathways based on functional annotations.
- Comparison of abundances for 16s rRNA gene sequence microbiome taxonomic abundances and functional predictions.

Chapter 4 describes comparative analyses of three different lepidopteran species' larva metagenomes. Insects representing the most diverse groups of organisms have been reported to have a varied composition of gut microbiome assisting them to adapt to extremely diverse eco-environments. They have a wide range of food sources.

Objectives of the study include-

- Observation of a common pattern of taxonomic or functional abundances of the lepidopteran species.
- Analysing the significant differences of the microbiota associated with their phenotype, metabolic activities and/or behaviour (pests, etc.).

Chapter 5 encompasses the differential profile of metabolites of the gut of *S. cynthia ricini* based on different diets and determines the effects of microbiota associated.

Understanding the dynamic relationships between the structure of gut microbial communities and metabolites in insects during plant host shifts can contribute to a better understanding of herbivore ecology and nutrient acquisition, immunomodulation and physiogenesis in response to profound lifestyle changes.

Integrative analyses of the metabolomic and microbiome data assist in a deeper understanding of the metaorganism i.e., the host organism, the environmental factors and the microbiota involved. 16s rRNA gene sequence analysis doesn't always provide the complete assessment of microbes involved or their contribution to the metabolic activity profile of the host as other factors such as expression and non-linear associations are involved. The metabolic phenotype of the host is better understood with subtle differences leading to a more in-depth and sensitive analysis. Hence, integrative analyses uncover the inherent associations between the microbiota and the host metabolic phenotypes and have been used to elucidate the differences in gut microbes and host metabolic phenotypes in the context of different food compositions. These studies are used to establish a more concise hypothesis based on the conditions and the profile.

With the development of high-throughput analysis techniques involving precise instruments including the development of liquid chromatography–electrospray ionization mass spectrometry (LC-ESI-MS or simply LC-MS), easy sample preparation process and wide metabolite coverage, untargeted or global metabolomics aims at quantifying as many metabolites as possible in biological samples.

Objectives of the study include-

- Establishing the differential metabolic profile of the insect gut based on different host plants.
- Metabolite annotation and description.
- Establishing microbiome-intermediated correlations between host plants and metabolites.

Overall, the study summarised the gut microbes of *S. cynthia ricini* using high-throughput analytical techniques. The study included exploration and optimisation of the techniques involved via comparative studies. The gut microbiota of the silkworm was analyzed through various environmental conditions to understand the changes in the structure and functions associated and infer the microbes and their interactions and functions responsible for the phenotypic changes associated. The statistical analyses associated with the compositional, sparse, high dimensional nature of microbiome data were studied and optimised for the methodologies involved in an attempt to reproducibility of the studies. The gut microbiota of the organism was studied with comparison to other closely lepidopteran species to understand the inter-species microbiota differences and related behaviours. The microbiome changes of the gut were further studied concerning the metabolic changes of the gut based on host diet and the correlation between the microbial changes and metabolic changes was explored, providing an integrative output and a greater resolution of the gut microbiome. In summary, the outcome of these studies will contribute to reducing the research gaps existing with the silkworm and its importance, and assist in the development of better growth and development of the organism. This, on a broader perspective, would benefit the preservation of the ecologically significant organism and assist with its enhanced rearing techniques, also benefiting the rural sericulture economy of north-eastern regions of India.

LIST OF TABLES

LIST OF FIGURES

Figure5.3: A graph showing the correction curve. Samples eluting after median retention time indicated by a positive deviation, samples eluting before median retention time indicating negative deviation.

Figure5.4. A diagrammatic representation of the methodology applied.

Figure5.5: NMDS plot of Unpaired parametric t-test (Welch t-test, unequal variances) analysing the diversity pattern of the metabolites based on different diets.

Figure5.6: Cloud plot representation of the differentially abundant gut metabolites based on diet-types. Green represented upregulated features in Eri-leaves based samples, red represents features upregulated in Kesseru-leaves based samples. Feature indicators of the molecules are defined by intensity (m/z) and retention time(m). Log-fold change is depicted by the value of the size of the circles. Colour shades depicts the p-value and interpret the significance, where lower p-values were depicted by brighter red/green shades. corrected retention time (TIC) of total chromatograms were overlaid in grey in the background.

Figure5.7: Differential abundance heatmap representation of the putative identified metabolites based on a METLIN

Figure5.8: Heatmap describing the microbiome-metabolite correlation.

LIST OF ABBREVIATIONS

SCL: Samia Cynthia Larva

QIIME2: Quantitative Insights Into Microbial Ecology

PICRUSt2: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States

KEGG: Kyoto Encyclopedia of Genes and Genomes

COG: Cluster of Orthologous Genes

SPAdes: St. Petersburg genome assembler

PERMANOVA: Permutational Multivariate Analysis of Variance Using Distance Matrices

PLS-DA: Partial Least Squares Discriminant Analysis

ANCOM-BC: Analysis of Compositions of Microbiomes with Bias Correction

ALDEx: ANOVA-Like Differential Expression

DeSeq: Differential Gene Expression Analysis based on the Negative Binomial Distribution

OTU: Operational Taxonomic Unit

NMDS: Non-metric Multidimensional Scaling

PCoA: Principal Co-ordinate Analysis

rRNA: ribosomal Ribonucleic Acid

ARG: Antimicrobial Resistance Genes

LC-MS: Liquid chromatography–mass spectrometry

kDa: kilo Dalton

RRHD: Rapid Resolution High Definition

METLIN: Metabolite and Chemical Entity Database

cAMP: cyclic Adenosine Monophosphate

Chapter 1

Introduction and Review of Literature

Chapter 1

Introduction and Review of Literature

INTRODUCTION

Insects, with high levels of diversity, tolerance and adaptation properties constitute around 66% of all animal species[1]. They are found almost everywhere, occupying various niches of the ecosystem, exploiting most types of organic matter and being of various ecological and economic values. From predatory insects, parasitic insects and pests to agriculturally beneficial insects providing pollination, beeswax, silk, etc., insects are crucial to the environment[1–4].

Silkworms are one of the beneficial lepidopteron insects popular for the production of silk fibre, which is considered as the “Queen of textiles”[5–7]. Apart from the versatile uses of silk as a fabric, including formal wear, sleep wear, rugs, its applications have recently increased to a great extent in the biomedical technology from medical stiches to prosthetic arteries and in scientific research extensively as a genetic resource to deal with a varied range of biological analyses[8–11].

However, besides *Bombyx mori*, very less research has been carried out in other silkworm species, in spite of their great potential values[11,12]. Northeast India is the endemic region to some of the unique and important silkworm species viz., *Antheraea assamensis* producing the golden Muga silk, *B. textor* producing the white pat silk and *Samia cynthia ricini* producing the warm Eri silk[13,14]. The ecological and economical aspects of these silkworms are huge, thus increasing their usage and demand in the commercial industries[13]. However, very little is known about the respective organisms resulting in their poor and obsolete rearing practice and cultivation techniques and low productivity of silk. The health conditions of these organisms are weakening, leading to decrease in their population which have negative effects on the environment due to their role in the ecosystem[15–17].

Micro-organisms are omnipresent with an abundance of 99% among living organisms[18–20]. Multicellular organisms have been found to exist as meta-organisms comprising of both the macroscopic host and its symbiotic commensal microbiota deciding the fate of their various metabolic activities and hence playing a huge role in various functions of the organisms including the host immune system, hormone regulation, carbohydrate metabolism and nutrient transport, etc[21–24] (**Figure1.1**). Associations of microbial communities and almost all plants and animals have been established with a mutualistic relationship where they provide for each other's requirements including nutrients, vitamins, ions and defence[25,26]. The microbiota present in the organisms may contribute to the function of their hosts either via synthesis of metabolic components required by the host for their proper functioning or via horizontal transfer of genes that encodes proteins required by the host organism[27,28].

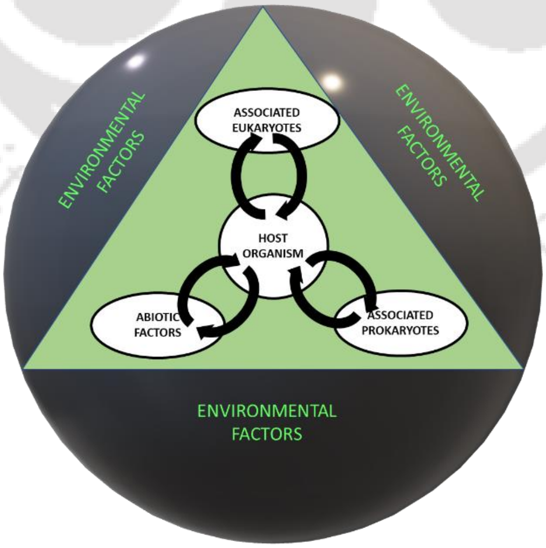


Figure1.1: Representation of a meta-organism comprising of the host organism, the microbiota (prokaryotic and eukaryotic), and the external factors associated and their inter-dependence.

The microbial community has been reported to contribute to many ecologically and economically important insects in their adaptation, heat tolerance, protection against pathogens or natural enemies, reproduction and vector competence[29–36]. Hence, a proper insight into the world of the microbiota co-existing with the macro-organisms and contributing to their major functions is crucial.

The analysis of the whole microbiota of a particular environment including those residing in multi-cellular organism as a part of their own system has been possible via the metagenomic analysis[37,38]. Metagenomics is the analysis of the total microbial genomic contents via direct collection of the samples present in a particular environment[39–41].

Following the observation of the ‘Great Plate Count Anomaly’, which states that only <1% of all known microorganisms are culturable; the co-dependence of micro-organisms to successfully grow and reproduce via mechanisms which includes acquisition and exchange of genetic resources and metabolites has become apparent[42,43]. Thus, metagenomic analysis with the recent advances in its associated techniques such as high-throughput sequencing for large-scale metagenomics using platforms like Illumina has led to the possibility of identification and analysis of unculturable micro-organisms. This results in a vast amount of information concerning the micro-organisms and their respective role in the environment they prevail[40,44,45].

Due to the role of the microbiota present in the organisms, extensive research to understand and analyze the micro-organisms present in these silkworm species will help us to understand the species and aid in their enhanced growth and development with high quality productivity of silk[46–48].

In this study, the metagenomic analysis of *S. cynthia ricini* (Eri silkworm) (n= 13) has been proposed. The Eri silkworm is the only other completely domesticated silkworm besides *B. mori*.[49] Unlike other common silkworms, Eri silkworms form open-ended cocoons enabling the moth to emerge easily[50,51]. Eri silk has excellent qualities combining the elegance of silk with the comfort of cotton and warmth of wool[51–53]. Besides the production of silk, ericulture is actively practiced in many households of the north-eastern regions of India due to their protein rich larvae and pupae, a delicacy for the tribal[49]. Understanding the microbiome of the Eri silkworm gut will aid in the proper growth and development of the organism and better quality of silk production.

REVIEW OF LITERATURE

The ecology of insects and lepidoptera

The diversity of insects comprises of ~4–8 million species around the world, i.e., ~3/4 of the total organisms; and ~0.1 million diverse species occurring in India[54–56]. Insect biomass with respect to all vertebrates has an overweight of 4:1 ratio (~200 million insects:1 human)[54]. Along with their diversity, the utilities and insect-derived resources are also quite huge and thus is very significant to mankind. However, due to climate change, anthropogenic activities and several other factors, the population of insects have declined by 40% and continuing at a rate of 2% yearly, around 8x more compared to other species[57–62].

Lepidoptera is the second most diverse group of insects following beetles. Many of lepidopteran species are widely known for their pest nature[63]. About 11,300 lepidopteran species are known to be found in India out of which less than 50% are well documented. The efforts of data collection, documentation and conservation of these species are more through personal approaches with a very small group associated. This negatively affects research in associated areas as well due to the lack of information. This along with growing human population collectively increases threat to the lepidopteran community many-fold[64].

The ecology and biology of silkworm

The first documented encounter with silk was around five thousand years ago, when a silkworm cocoon from a mulberry tree fell into the hot water while preparing a cup of tea by princess Xi Ling Shi and its silk thread started unravelling[65]. Silk plays a vital role in Indian community from requirement on a daily basis to being associated with its culture[66–68]. Sericulture is a huge source of income and employment in Indian economy with a high market value of both raw materials and processed goods. Cocoons are sold for up to 300 rupees a kilo, while raw silk up to 2,400 rupees[69,70]. It is a great agricultural prospect for farmers as it doesn't depend on the uncertainty of weather unlike for crops growth and eliminated the impact of natural calamities on their income source. The Ministry of Textiles, India implemented several initiatives for the promotion and development of Sericulture in the country[71].

The most widely known silkworm species is *B. mori*. It has a genome of size of 450Mbp, which is 1/6 of the human genome and 4x of the *Drosophila melanogaster* genome (International Silkworm Genome Consortium, 2008)[72]. Many variants with mutant strains and diverse morphological characters have been known[73,74]. It has a moderate body size (~75 mm larval length), and composed of organs including fat body, silk gland, haemolymph and the gut with a larval stage generation time for about 25–30 days[75]. In research, involvement of *B. mori* started with Toyama studying it to confirm Mendel's laws of heredity making its first confirmation in animal species[76].

Gradually they were used for numerous research studies comprising of genetics, medicine, economic and ecological-based areas. It was used as a model organism for the expression of human insulin receptor, having 40% amino-acid sequence similarity to human insulin; to evaluate drug efficacy for hyperuricemia, study methylation, CRISPR-Cas9 and epigenetic regulation[77–83].

Genome studies of the organism helps us understand the different variations of existing domesticated/semi-domesticated silkworms and study the genetic and morphological traits detecting mutations[84]. These studies would also help to further understand their development and contribute to their better growth and development and also increase the economic value of its products.

Domestication of silkworm has also been the studies for many years which is estimated to have started ~7500 years ago[84]. Continuous practices associated with domestication have led to *B. mori* losing many lepidopteran traits including larva colour, flight ability of adult moths, ability of foraging, and have acquired modified traits such reduced wingspan, decreased immunity and increased diet specificity[85,86]. This indicates a decreased suitability of *B. mori* as a model organism for various purposes and genetic research in regards to understanding lepidopteran insects.

Samia cynthia ricini

S. cynthia ricini (n= 13) is a multivoltine domesticated silkworm most frequently distributed in North-Eastern part of India, China and Japan[52]. It is commonly called as ‘Eri silkworm’ White or reddish warm Eri silk is produced by these silkworms. Some other rare eco-races of the species are also found across the Indo-Australian regions and the Palaearctic regions[87–89]. *S. cynthia ricini* is a polyphagous organism with the primary food being the leaves of castor plant (*Ricinus communis L.*). Other feeding plants include *Heteropanax fragrans* (Kesseru), *Ailenthus gradulosa*, *Evodia flaxinifolia*, and *Manihot utilissima*. The organism derives its name “Eri” from the Assamese word ‘era’ which translates to castor[90]. The Eri silkworm was studied to be domesticated from its closely related wild progenitor, the Himalayan *Samia canningi*. However, further DNA analysis has demonstrated that the species is a polyhybrid, created in captivity containing various other *Samia cynthia* races/subspecies genes[87,88]. Eri silkworms though were acclimatized in various regions of Europe and America but couldn’t take firm hold[91]. The generation time of domesticated *S. cynthia ricini* is up to 6 generations per year while the wild members complete around 3 generations per year[91].

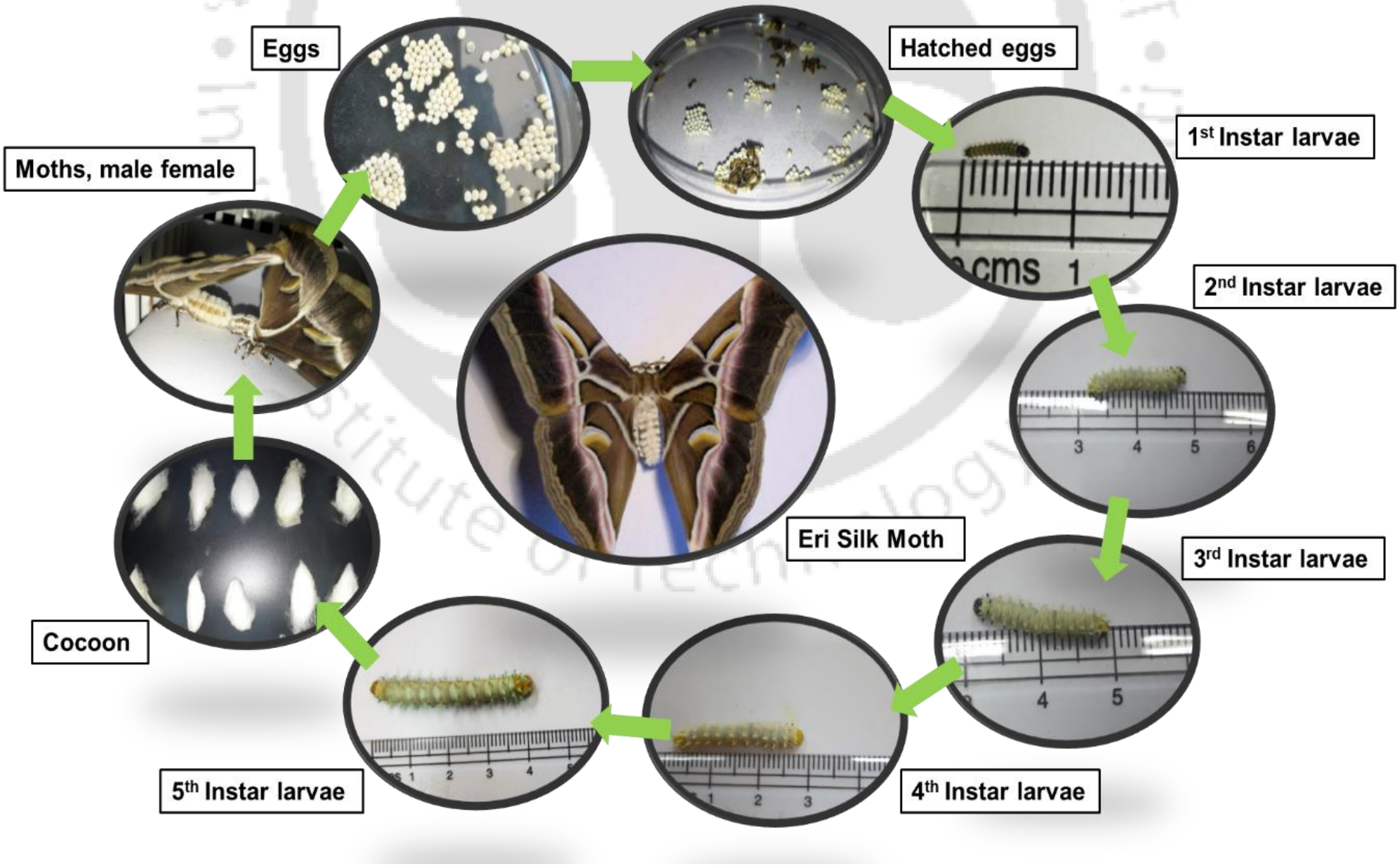
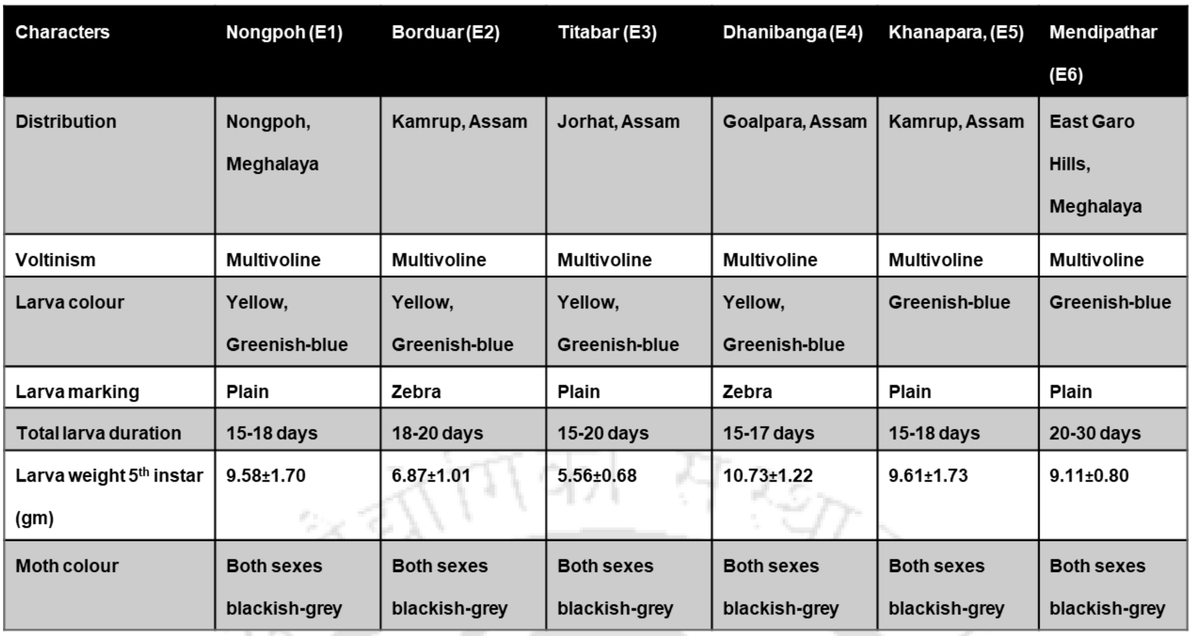


Figure1.2: Representation of the life cycle of *Samia cynthia ricini*

Belonging to the family of Saturniidae, *S. cynthia ricini* has the basic saturniid characters but based on its different generic origin, it has many distinguishing characteristics to that of other members of the species including *Bombyx* and *Antheraea*. *S. cynthia ricini* has 26 pairs of chromosomes[90]. Greenish yellow larva hatches from candid white, ovoid eggs which gradually changes to pure yellow by the third day. Colour segregations turning the body into yellow, green, cream, or blue are observed in the body from around late third instar onwards. A white powdery substance is seen to be covering the full matured fifth instar larvae. Various types of spots are observed through the later stages. E.g., unspotted, single-dotted, double-dotted, semi-zebra and zebra spotted. The pupa is similar to other Saturniidae species and weigh around an average of 2.6g. The cocoon is however distinguishable from other members of the species. They are soft, wooly, open-mouthed, elongated, peduncleless and unreelable. Colour polymorphism from brick red to creamy white are observed. Larger female moths are observed with a length of ~3cm while male moths being ~2.3cm. Wing span of female and male moths were ~15cm and ~13cm respectively[88,89,92].



Characters	Nongpoh (E1)	Borduar (E2)	Titabar (E3)	Dhanibanga (E4)	Khanapara, (E5)	Mendipathar (E6)
Distribution	Nongpoh, Meghalaya	Kamrup, Assam	Jorhat, Assam	Goalpara, Assam	Kamrup, Assam	East Garo Hills, Meghalaya
Voltinism	Multivoline	Multivoline	Multivoline	Multivoline	Multivoline	Multivoline
Larva colour	Yellow, Greenish-blue	Yellow, Greenish-blue	Yellow, Greenish-blue	Yellow, Greenish-blue	Greenish-blue	Greenish-blue
Larva marking	Plain	Zebra	Plain	Zebra	Plain	Plain
Total larva duration	15-18 days	18-20 days	15-20 days	15-17 days	15-18 days	20-30 days
Larva weight 5th instar (gm)	9.58±1.70	6.87±1.01	5.56±0.68	10.73±1.22	9.61±1.73	9.11±0.80
Moth colour	Both sexes blackish-grey	Both sexes blackish-grey	Both sexes blackish-grey	Both sexes blackish-grey	Both sexes blackish-grey	Both sexes blackish-grey

Table1.1: Summarised table representing phenotypic and genotypic variations of *Samia cynthia ricini* found in north-eastern regions of India (adapted from Vijayan et al., 2006) [92]

Unlike *B. mori*, *S. cynthia ricini* has been observed to have retained many domesticated associated traits that are lost in *B. mori*[89]. Having the advantage of being multivoline, it can be utilised for research and other beneficial purposes not limited by seasonal effects. Owing to its high genetic diversity, it can be explored significantly. They have also been reported to produce fertile hybrids with other wild Samia species including *Samia pryeri* and *Samia canningi* which are largely distributed through the regions of South-Eastern Asian countries[90]. Different endemic nature along with other traits such as host plants preferences are observed accordingly that have been enlightening through studies and are gradually being explored. *S. cynthia ricini* have been seen to be successfully reared uniformly and synchronously in large scale, efficient for economic purposes with better quality of the organism and its silk[91]. Functional analysis of its genes of interest have been carried out by using genome editing techniques such as Transcription activator-like effector nucleases (TALENs) and obtaining knockout gene cell-lines[93].

S. cynthia ricini assembled genome has a size of 485Mbp with 16,702 predicted protein coding genes. Its gene repertoire was similar to *B. mori* with some genes including fibroin and chorion seen to be more evolved[94].

All of these uniqueness and ease of its cohabitation, makes it a candidate species for sericulture and research purposes. Populations of *S. cynthia ricini* have been widely used for sericulture in different regions of North-east India and with their commercial exploration a wide range of morphological and genetic variations are observed such as different larval weight and body patterns, cocoon weight, silk content and fibre ratio, etc[88].

The gut microbiota and its effects on an organism

Being omnipresent, microbes are also present largely in different environments including most parts of almost all organisms and play important role in their genetic and metabolic characteristics[18,95]. The term "microbiota" is used to define the comprehensive population of microbes in a particular environment[96]. "Microbiota" and "microbiome" are often interchangeably used to define the present microbes in that particular environment. However, microbiota refers to the total population of microbes including archaea, bacteria, fungi, and viral species in the environment while microbiome is the entire combined meta-organism comprising of the microbial population, genome and the surrounding environment[97,98]. Gut is a vital organ in carrying out major metabolic functions related to the growth and development of an organism and widely studied for the purposes of understanding the microbiome of the living bodies[99]. In humans, bacterial cells have been observed to be 10x than the total human cell indicating the extent of their role in the human body[98]. The gut microbiota of many organisms is involved in many metabolic functions of the host organism such as digestion, growth, immunity, disease management coping with the surrounding environment[99,100]. They act as biomarkers of the species predicting the behaviour of different species under different circumstances such as a diseased state[101,102]. In many cases, bacterial metabolites were taken up by the host for the generation of other functional enzymes (co-factors) involved in host physiological activities[100]. These studies infer that the microbial population impacts and are also impacted by the host organisms on multiple levels and is a very dynamic process interdependent on each other that should be more deeply understood. Interactions of functions between different microbial species are essential to be analyzed as they are the downstream determining factors of an organism coping with the changes in the surrounding environment[96,99].

Numerous studies demonstrated the patterns and signatures of microbial communities in insect gut which also adapted according to changed environmental conditions e.g., changes in diet either by changes in the type of level of enzyme induction or by the change in microbial community[29,34,103]. Cellulose-rich diet in cockroaches was observed to have an increase in the population of protozoa in their hindgut[104]. Many sophisticated signals controlling behavioural aspects of insects such as guaiacol in locusts acting as an aggregation pheromone are also impacted by their gut microbial population[105]. Role of symbiotic bacterial community was observed in nymphal development of stinkbugs[106]. Metabolic activities of many microbial communities were observed to be controlled and coordinated by autoinducers for quorum sensing (QS)[107]. Metabolic activities such as production of pathogenicity factors, dyes, toxins, biofilms, bioluminescence were seen as a result of recognition of these signal molecules[26,36]. Sex-based difference in autoimmunity in mouse was found to be driven by hormone-dependent regulation where the gut microbiome played important role[108]. Similar studies concerning immunocompetence and susceptibility to pathogens in insects were also reported[22,109].

The gut microbiome of Lepidoptera

Lepidoptera larva is highly active in ingestion and digestion of food thus involving the various enzymatic reaction, solubility of the ingested food components, breaking down of the protein molecules[110,111]. Gut pH is very important in regulation of these enzymatic activities. Many members of the insect orders have highly alkaline gut regions. Most lepidopteran gut have a pH value ranging from 7.0-11.0 correlated with higher activities of trypsin and chymotrypsin[110,112]. These were also associated with the determination of their gut microflora[112].

Lepidopteran larval gut doesn't have any specialized structures such as the diverticula and the role of micro-organisms being involved in its nutrition and digestion has been explored[110]. It was observed that digestive enzyme producing bacterial population were higher in most lepidopteran gut which assisted with digestion of leaf constituents including xylan, starch, pectin and cellulose. Along with aid in digestion, they also improved the ability of the organisms to live on sub-optimal diets, provide vitamins, etc[48,112].

Both competitive and co-existing relationship of Lepidopteran gut bacteria were observed[113–117]. The total cultivable bacterial count of lepidopteran gut was observed to decrease through the increasing larval stages. However, the count of cellulolytic bacteria increased from the first to fifth instar larval gut[47,106,118,119]. These observations were statistically significant as well. The increase in cellulolytic bacteria was seen to be directly proportional to the ingesting and digestive capacity indicating the relationship between gut bacteria and digestion; and the decrease in the total bacterial count through the larval stages indicated a competitive relationship between the cellulolytic and other group of bacteria[106,110,112,120–122].

The gut microbiome of silkworms

Several studies analysing the gut bacteria of silkworm species, mostly *B. mori* have been carried out. Bacterial and fungal population have been reported to influence the growth and development of the silkworm along with its metamorphosis. Cellulolytic bacteria were observed to gradually increase through the larval growth stages of *B. mori*[47,119]. Numerous horizontal gene transfers have been reported which benefitted the organism with respect to its survival and prolificacy[81,109,119]. The silkworm gut microbiota has also reported to be impacted by forages. Co-evolution of the dietary changes and the changes in an organism is believed to have been an important mechanism in the formation of the host-gut microbe relationship[109,123]. pH of the digestive juices of the silkworm, *B. mori* which have pathogen resistance functions were reported to be impacted by the intestinal bacteria[47,124].

Metagenomics as a guide to the world of insect microbiome

Metagenomics is the direct genetic analysis of the total collection of all microbial genomes from a particular environment[125]. The concept of direct genetic analyses using sequencing tools started with the observation that the micro-organisms under a microscope collected from an environment wasn't corresponsive with that of the ones obtained after culturing in a plate[40]. Gradually, more evidences implied that microbes require special conditions to grow and were attempted to grow in conditions resembling their native growth environment. This was the start of microbial ecology which referred to the study of microbes and their environmental roles[126,127].

However, majority of them were still unculturable under laboratory conditions even with the proper resembling environment[126,128]. Q-fever causing *Coxiella burnetii* could be cultured after years of attempts and with hundreds of trial combinations of culture conditions[129]. Around 85-99% of bacteria were yet to be cultured and identified due to this huge disadvantage, making scientific knowledge about microbes and their potential uses drastically limited. This phenomenon was referred to as the “Microbial Dark Matter”[130,131]. Furthermore, micro-organisms were characterised using morphological and biochemical parameters which provided a rather limited resolution of the microbes and their applications. Many microbial products including antibiotics and their related genes were only characterised through sequencing technologies thus setting the foundation of metagenome analysis or metagenomics[128,130].

The method of extracting DNA directly from the environment started by Pace in 1985 which became a revolutionary step in the field of metagenomics[132]. The term metagenomics was coined by Jo Handelsman, Jon Clardy and Robert M. Goodman in 1998 and redefined and established by Kevin Chen and Lior Pachter in 2005[133,134].

Microbial community studies include studying the population as individual microbe groups responding to a particular environment and also analyses of the biochemical functions according to the change in the environment of the host organism[127,132]. Based on this, is categorised into structural and functional analyses. Structural metagenomics is focused on characterising the microbial communities of a particular environment and the patterns of changes accordingly. It includes taxonomic profiling, gene and metabolic pathways prediction. It is utilized for determining community structures, prediction of a sample character based on the microbiome structure, establishing the relationship between microbes and the environment[135–137]. Phylogenetic studies of Proteobacteria and Lux gene homologs inferred high levels of global similarity indicating similar lines of divergence and that the quorum sensing systems are very ancient arising early in the Proteobacteria lineage[138]. In humans, microbial genomic structure variants, that are tiny gene differences in the gut microbiota are studied and reported to have highly adaptive properties including antibiotic resistance and CRISPR-related activities[139]. Several other studies establishing the change in the structure of the microbial community with respect to the change in environment have been analyzed such as a diseased vs healthy organism, an area polluted with heavy metal, based on geographical and diet-based differences, etc[98,137,140].

Functional metagenomics, more focused on analysing the functions of the genes targets on studying gene expression, screening of functions, protein product characterization such as optimum temperature, pH, enzyme activity and its quantification, etc. It can be utilized to find novel enzymes with potential applications allowing high-resolution analysis. These are carried out by the developing sub-disciplines within the metagenomics that include metatranscriptomics i.e., study of the total microbial RNA of an environment, metaproteomics i.e., study of the total microbial protein of an environment and metabolomics i.e., studying the total metabolic compounds of an environment[137,141,142].

This aspect is useful for identification of the genes present in the microbial communities, analyze their functions, identification of novel genes and functional pertaining our importance for the better growth of the organism and commercial applications of economically important enzymes[136,137,141,142]. Several high activity enzymes such as lipase with a higher alkaline stability and specificity utilized in the food industry; proteases and esterases with higher pH and thermal stability, antimicrobial genes such as tubomycin, volacein, several novel multi-drug resistance genes; and also, biosynthetic pathways for important metabolites such as biotin, novel degradation pathways for pathways were identified through functional metagenomics[143,144].

Metagenomic studies provide us with the population and genetic studies but doesn't provide the actual presence and functions of the biochemical compounds, thus hindering the efforts to define a proper connection between the microbial strain, their network and the host phenotypes[145,146]. Use of the complementary sub-discipline such as metatranscriptomics, metaproteomics and metabolomics provide further levels of resolution for functional analysis of uncultivated microbial communities[147]. These studies together comprise meta'omics approach of microbial community analyses[148].

Meta'omic analyses are revolutionising the exploration of the diversity, functions, and inter-relationships among different organisms in the various ecological niches[149]. Gut metagenomics can provide an insight to the spontaneous, non-linear, and dynamic interaction among different gut residing microbes, thus, predicting the interaction network that can aid in developing novel genes and possible therapeutics[150,151].

Some of the problems during metagenome and other meta'omic analyses include sampling and cloning bias due to its complex nature. Many sampling locations are either extreme or unstable leading to further bias and has to be handled with great caution. Contamination through the steps of sampling and processing of genomic data (for metagenome/ metatranscriptome) and metabolites (metabolome) is also a major concern[152]. Development of second and third generation sequencing technologies has the advantage of minimal cloning bias as the library preparation didn't require bacterial hosts for cloning vectors[153]. This also helped in reducing DNA contamination. Downstream analyses and data mining is another challenging area in this field. Due to huge amount of data generated, it is imperative to have a clear direction of objective. Reproducibility of metagenomic analyses is another concern[154]. Handling of large comprised meta'omic datasets has further difficulties involving them being cumbersome while downloading and the attempts of successful association with appropriate metadata. Many missing links are observed through the processes[148]. Attempts have been made to minimise these challenges by adopting different methodologies such as standardisation of the data, adopting appropriate repositories and linking the metagenomic/microbiome data with their metabolomic data and thus reducing the gap for subsequent associated studies. The methods are prioritised for identifying confirmed novel links between the microbial genes/gene and the different metabolites[149,150].

Due to the huge amount of data and information generated through the meta-analyses of the gut microbial genome, statistical analyses are a major part of evaluating and navigating through the data and for a specified output[155].

The basic statistical approaches include normalization of the data[156]. The nature of microbiome/metagenome data is high-dimensional, structured, multivariate, and compositional thus having a large amount of sparse data. Compositional data are depicted as follows,

$x = (x_1, \ldots, x_k)$; $x_i > 0$, $i \in \{1, \ldots, k\}$ with a constraint or non-informative total sum, and defined as a matrix consisting of positive strict real numbers[157,158].

Most common techniques of data normalization include the log-ratio approach followed by rarefying of the data[156]. Most of meta'omics data are generated with exploratory aspects. Hence, appropriate measures for the process of a practical output are necessary. Diversity and variance analyses is an important aspect for analysing the data for further processing[159]. Common approaches of analysing the diversity includes richness analysis which attempts the representation of the total species in a sample; Evenness studies, which attempts the representation of the unequally represented microbial population in the communities where the dominant species are not significantly distinguished and many uncommon species are observed[160,161].

Following analysis of significant diversities, differential abundance analyses are performed which helps us quantify the differences observed, followed by the inference and the validity of the hypothesis. Due to its dimensionality, to demonstrate the global difference, multivariate approach is applied for differential abundance analyses, the most common being the Permutational Multivariate Analysis of Variance Using Distance Matrices (PERMANOVA) analysis which uses a multivariate Analysis of Variance Using Distance Matrices (ANOVA) distance-based approach[155,162]. Partial Least Squares Discriminant Analysis (PLS-DA) is another methodology preferred[161–163]. Following the establishment of significant global difference, univariate methodologies to attain the specific association with respect to the response is carried out. Common methodologies include classic approaches including Kruskal-Wallis, Welch's t-test or Wilcoxon rank-sum test or compositional data-based including ANCOM-BC and ALDEx2[155,157]. Integrative analyses increase the statistical influence of the studies and amplify the reproducibility and validity[164–166].

Uncovering the role of metagenome of the Eri silkworm and identification of the factors influencing its molecular mechanisms will help us understand what role its inhabitantmicrobiome plays in its basic biology and how it aids the silkworm in adjusting to hostile biotic as well as abiotic factors. These microbes might have significance in a better phenotype with enhanced growth and immunity and production of better silk in quality and quantity. This research will provide a foundation to develop enhanced culture techniques infused with health-promoting microbes. As Eri is already domesticated, a new type of probiotic feed may be developed for rearing these species indoor. Research in the field of the unexplored world of Eri silkworm will thereby sensitize the scientific community and provide us cues to educate the rural indigenous population on improved rearing methods. Ease of rearing and more yield will definitely increase the rate of human employment in the sericulture sector, thereby, strengthening the economic sector and ensuring gradual socio-economic growth of the rural populace of the nation. At the same time, rearing will also promote more tree plantation.

Using information from these studies and combination of the complementary processes helps reduce the complexity of microbial communities and the dynamic nature of their associations to different environments including their host environment, external environment and their interdependent interactions; construct microbial model communities and implement such models for subsequent studies appropriately.

Thus, based on our literature of review, with the aim to understand the gut microbes of the silkworm, *S. cynthia ricini* with respect to their role in its growth and development, and the different factors associated, to optimise the methodologies regarding its metagenome analyses and to explore the differences compared to other lepidopteran gut microbes the following objectives were formulated.

1. Comparative analyses of *Samia cynthia ricini* and its gut microbiome under different diet conditions and growth stages.

2. Shotgun metagenome analysis of *Samia cynthia ricini* 5th instar larva.
3. Comparative analyses of lepidoptera larva metagenome.
4. Metabolic profiling of Samia cynthia ricini based on different diets and analysis the of the associated microbiome

REFERENCES

1. Jankielsohn, A. & Jankielsohn, A. The Importance of Insects in Agricultural Ecosystems. *Adv in Entomology* **6**, 62–73 (2018).
2. Beneficial Insects - Predators, Parasitoids and Pollinators. (2016).
3. Lewis, O. T. & Gripenberg, S. Insect Seed Predators And Environmental Change. *J Appl Ecol* **45**, 1593–1599 (2008).
4. Tanda, A. S. Why Insect Pollinators Important in Crop Improvement? *Indian J. Entomol.* **84**, 223–236 (2022).
5. Sathish Kumar, T., Kumar, R. & Kumar, S. Review On Functional Characteristics Of Silk And Silk Blended Fabrics. Tierärztliche Praxis. doi:10.1016/B978-1-78242-311-9.00001-X, (2020).
6. Maurya, S. & Shyam, R. Silk: The Queen Of Fibres. *Rashtriya Krishi* **12**, 19–22 (2017).
7. Zhao, F., Nosch, M.-L., Unesco. & Zhongguo si chou bo wu guan. Textiles And Clothing Along The Silk Roads. 413 (2022).
8. Chand, S., Chand, S. & Raula, B. Usage of Silkworm Materials in Various Ground of Science and Research. *J. Nat. Fibers*. *https://doi.org/10.1080/15440478.2022.2139328* **20**, (2022).
9. Pugno, N. M., Motta, A. & Kaplan, D. Editorial: Frontiers in Silk Science and Technology. *Front Mater* **8**, 153 (2021).
10. Oduor, E. O. *et al.* Applications of Silk in Biomedical and Healthcare Textiles. *Textiles for Functional Applications* (2021) doi:10.5772/INTECHOPEN.96644.
11. Ebrahimi, D. *et al.* Silk–Its Mysteries, How It Is Made, and How It Is Used. *ACS Biomater Sci Eng* **1**, 864 (2015).
12. Tikader, A. K. V. and B. Saratchandra. Muga Silkworm, Antheraea Assamensis (Lepidoptera: Saturniidae)-An Overview Of Distribution, Biology And Breeding. *Eur. J. Entomol.* **110**, (2013).

13. Boro, P. & Borah, S. D. Biodiversity Of Sericigenous Insects In North-Eastern Region Of India-A Review. ~ *269* ~ *J. Entomol. Zool.* **8**, 269–275 (2020).

14. Silks of Northeast India Truly a Treasure and Pride of the Nation, Caleidoscope | Indian Culture, Heritage. https://www.caleidoscope.in/art-culture/silks-of-northeast-india-1.

15. Singh, T. & N. D. & K. R. Innovations in Silkworm Rearing and Importance Recent Advances. *J. Text. Assoc.* **82**, 87–90 (2021).

16. Chapter-12 Sericulture and Silk Industry.

17. Nuraeni, S. Gaps In The Thread: Disease, Production, And Opportunity In The Failing Silk Industry Of South Sulawesi. *Forest and Society* **1**, 110–120 (2017).

18. Gupta, A., Gupta, R. & Singh, R. L. Microbes and Environment. *Principles and Applications of Environmental Biotechnology for a Sustainable Future* 43 (2017) doi:10.1007/978-981-10-1866-4_3.

19. Finlay, B. J. & Clarke, K. J. Ubiquitous Dispersal Of Microbial Species. *Nature 1999 400:6747* **400**, 828–828 (1999).

20. Pascoal, F., Costa, R. & Magalhães, C. The Microbial Rare Biosphere: Current Concepts, Methods And Ecological Principles. *FEMS Microbiol Ecol* **97**, 227 (2021).

21. Bosch, T. C. G. & McFall-Ngai, M. J. Metaorganisms As The New Frontier. *Zoology (Jena)* **114**, 185 (2011).

22. Belkaid, Y. & Hand, T. W. Role of the Microbiota in Immunity and Inflammation. *Cell* **157**, 121 (2014).

23. Rowland, I. *et al.* Gut Microbiota Functions: Metabolism Of Nutrients And Other Food Components. *Eur J Nutr* **57**, 1 (2018).

24. Agamennone, V., Le, N. G., van Straalen, N. M., Brouwer, A. & Roelofs, D. Antimicrobial Activity And Carbohydrate Metabolism In The Bacterial Metagenome Of The Soil-Living Invertebrate *Folsomia candida*. *Sci Rep* **9**, (2019).

25. Christian, N., Whitaker, B. K. & Clay, K. Microbiomes: Unifying Animal And Plant Systems Through The Lens Of Community Ecology Theory. *Front Microbiol* **6**, 869 (2015).

26. Jandhyala, S. M. *et al.* Role Of The Normal Gut Microbiota. *World Journal of Gastroenterology: WJG* **21**, 8787 (2015).

27. Drew, G. C., Stevens, E. J. & King, K. C. Microbial Evolution And Transitions Along The Parasite–Mutualist Continuum. *Nat. Reviews Microbiol. 2021 19:10* **19**, 623–638 (2021).

28. Zilber-Rosenberg, I. & Rosenberg, E. Role Of Microorganisms In The Evolution Of Animals And Plants: The Hologenome Theory Of Evolution. *FEMS Microbiol Rev* **32**, 723–735 (2008).

29. Moran, N. A., Tran, P. & Gerardo, N. M. Symbiosis And Insect Diversification: An Ancient Symbiont Of Sap-Feeding Insects From The Bacterial Phylum Bacteroidetes. *Appl Environ Microbiol* **71**, 8802–8810 (2005).

30. Engelstädter, J. & Hurst, G. D. D. The Ecology And Evolution Of Microbes That Manipulate Host Reproduction. *Annu Rev Ecol Evol Syst* **40**, 127–149 (2009).

31. Brownlie, J. C. & Johnson, K. N. Symbiont-Mediated Protection In Insect Hosts. *Trends Microbiol* **17**, 348–354 (2009).

32. Ferrari, J. & Vavre, F. Bacterial Symbionts In Insects Or The Story Of Communities Affecting Communities. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 1389 (2011).

33. Akman Gündüz, E. & Douglas, A. E. Symbiotic Bacteria Enable Insect To Use A Nutritionally Inadequate Diet. *Proceedings of the Royal Society B: Biological Sciences* **276**, 987–991 (2012).

34. Engel, P. & Moran, N. A. The Gut Microbiota Of Insects – Diversity In Structure And Function. *FEMS Microbiol Rev* **37**, 699–735 (2013).

35. Aharon, Y. *et al.* Phylogenetic, Metabolic, And Taxonomic Diversities Shape Mediterranean Fruit Fly Microbiotas During Ontogeny. *Appl Environ Microbiol* **79**, 303–313 (2013).

36. Gupta, A. & Nair, S. Dynamics of Insect–Microbiome Interaction Influence Host and Microbial Symbiont. *Front Microbiol* **11**, (2020).

37. Blevins, S. M. & Bronze, M. S. Robert Koch And The ‘Golden Age’ Of Bacteriology. *International Journal of Infectious Diseases* **14**, (2010).

38. Burki, F. The Eukaryotic Tree Of Life From A Global Phylogenomic Perspective. *Cold Spring Harb Perspect Biol* **6**, (2014).

39. Sharpton, T. J. An Introduction To The Analysis Of Shotgun Metagenomic Data. *Front Plant Sci* **5**, (2014).

40. Escobar-Zepeda, A., de León, A. V. P. & Sanchez-Flores, A. The Road To Metagenomics: From Microbiology To DNA Sequencing Technologies And Bioinformatics. *Front Genet* **6**, 348 (2015).

41. Bashir, Y., Pradeep Singh, S. & Kumar Konwar, B. Metagenomics: An Application Based Perspective. (2014) doi:10.1155/2014/146030.

42. Connon, S. A. & Giovannoni, S. J. High-Throughput Methods for Culturing Microorganisms in Very-Low-Nutrient Media Yield Diverse New Marine Isolates. *Appl Environ Microbiol* **68**, 3878 (2002).

43. Harwani, D. The Great Plate Count Anomaly and the Unculturable Bacteria. *Int J Sci Res* **2**, 350–351 (2012).

44. Anyansi, C., Straub, T. J., Manson, A. L., Earl, A. M. & Abeel, T. Computational Methods for Strain-Level Microbial Detection in Colony and Metagenome Sequencing Data. *Front Microbiol* **11**, (2020).

45. Cibrián-Jaramillo, A. & Barona-Gómez, F. Increasing Metagenomic Resolution Of Microbiome Interactions Through Functional Phylogenomics And Bacterial Sub-Communities. *Front Genet* **7**, (2016).

46. Chen, B. *et al.* Gut Bacterial And Fungal Communities Of The Domesticated Silkworm (Bombyx mori) And Wild Mulberry-Feeding Relatives. *ISME J* **12**, 2252 (2018).

47. Dee Tan, I. Y. & Bautista, M. A. M. Bacterial Survey in the Guts of Domestic Silkworms, Bombyx mori L. *Insects* **13**, (2022).

48. Sun, B. F. *et al.* Multiple Ancient Horizontal Gene Transfers And Duplications In Lepidopteran Species. *Insect Mol Biol* **22**, 72–87 (2013).

49. Lakshmi NarayaNamma, V. Rearing of Eri Silk Worm and Its Economic Viability to Rainfed Castor Farmers. *Ind. Journ. of Ent. published Ref. No* (2105) doi:10.5958/IJE.2021.75.

50. Zhou, B. & Wang, H. Structure And Functions Of Cocoons Constructed By Eri Silkworm. *Polymers (Basel)* **12**, 1–18 (2020).

51. Mazzi, S., Zulker, E., Buchicchio, J., Anderson, B. & Hu, X. Comparative Thermal Analysis of Eri, Mori, Muga, and Tussar Silk Cocoons and Fibroin Fibers. *J Therm Anal Calorim* **116**, 1337–1343 (2014).

52. Tamta, M., Research, P. D. & Mahajan, S. The Novel Silk Fiber: Eri. *Intern. Journ. of Home Sci.* **101**, 101–104 (2021).

53. Melesse, G., Atalie, D. & Koyrita, A. Structural and Thermal Properties of Ethiopian Eri and Mulberry Silk Fibres. *Advances in Materials Science and Engineering* **2020**, (2020).

54. Lokeshwari, R. K. & Shantibala, T. A Review On The Fascinating World Of Insect Resources: Reason For Thoughts. *Psyche (London)* (2010) doi:10.1155/2010/207570.

55. Møller, A. P. *et al.* Abundance Of Insects And Aerial Insectivorous Birds In Relation To Pesticide And Fertilizer Use. *Avian Res* **12**, (2021).

56. Ghosh, A. K. Insect Biodiversity In India. *Orient Insects* **30**, 1–10 (1996).

57. Forister, M. L. *et al.* Fewer Butterflies Seen By Community Scientists Across The Warming And Drying Landscapes Of The American West. *Science (1979)* **371**, 1042–1045 (2021).

58. Cardoso, P. *et al.* Scientists' Warning To Humanity On Insect Extinctions. *Biol Conserv* **242**, (2020).

59. van Klink, R. *et al.* Meta-Analysis Reveals Declines In Terrestrial But Increases In Freshwater Insect Abundances. *Science (1979)* **368**, 417–420 (2020).

60. Dar, S. A. *et al.* Causes and Reasons of Insect Decline and the Way Forward. *Global Decline of Insects* (2021) doi:10.5772/INTECHOPEN.98786.

61. Sánchez-Bayo, F. & Wyckhuys, K. A. G. Worldwide Decline Of The Entomofauna: A Review Of Its Drivers. *Biol Conserv* **232**, 8–27 (2019).

62. van der Sluijs, J. P. Insect Decline, An Emerging Global Environmental Risk. *Curr Opin Environ Sustain* **46**, 39–42 (2020).

63. Insects - Insecta - The Animal Encyclopedia. https://www.thoughtco.com/insects-profile-130266.

64. Smetacek, P. Review of Indian Lepidoptera Collections and Their Significance In Conservation. *ENVIS Bulletin* **14**, 135–139 (2013).

65. History of Silk. https://www.thelightshadestudio.co.uk/Useful-Info/History-of-Silk.

66. The Silk Road: Connecting People and Cultures | Smithsonian Folklife Festival. https://festival.si.edu/2002/the-silk-road/the-silk-road-connecting-peoples-and-cultures/smithsonian.

67. The Role of Ethnic and Indigenous People of India and their Culture in the Conservation of Biodiversity. https://www.fao.org/3/xii/0186-a1.htm.

68. Eswarappa, K. Role of Women in Sericulture and Community Development A Study from a South Indian Village. *Sage Open* **3**, (2013).

69. Imports Of Raw Silk, Cocoons May Bring Down Prices | Business Standard News. https://www.business-standard.com/article/markets/imports-of-raw-silk-cocoons-may-bring-down-prices-110122800037_1.html.

70. Elumalai, D., Mohanraj, P., Ramamoorthy, R., Mohan, C. & Poovizhiraja, B. An Economics Analysis of Multiend and Automatic Reeling Units in Traditional and Non-Traditional Area of Tamil Nadu, India. *Int.J.Curr.Microbiol.App.Sci* **9**, 489–498 (2020).

71. Central Silk Board and Functioning of Central Silk Board Industry.

72. Mita, K. *et al.* The Genome Sequence Of Silkworm, Bombyx mori. *DNA Res* **11**, 27–35 (2004).

73. Furdui, E. M. *et al.* Genetic Characterization of Bombyx mori (Lepidoptera: Bombycidae) Breeding and Hybrid Lines With Different Geographic Origins. *J. Ins. Sci.* **14**, (2014).

74. Goldsmith, M. R., Shimada, T. & Abe, H. The Genetics And Genomics Of The Silkworm, Bombyx mori. *Annu Rev Entomol* **50**, 71–100 (2005).

75. Pandey, P., Tripathi, S. P. & Shrivastav, V. M. S. Effect Of Ecological Factors On Larval Duration Of Silkworm (Bombyx mori linn.). *JEOH* **6**, 123–125 (2006).

76. Yoshimaro Tan, B. A Study of Mendelian Factors in the Silkworm, Bombyx Mori. *J. Sci. Food Agric., Tohoku Imperial University, Sapporo, Japan* **5**, 91–113 (1913).

77. Chandrakanth, N., Ponnuvel, K. M., Moorthy, S. M., Sasibhushan, S. & Sivaprasad, V. Analysis Of Transcripts Of Heat Shock Protein Genes In Silkworm, *Bombyx mori* (Lepidoptera: Bombycidae). *Eur J Entomol* **112**, 676–687 (2015).

78. Mizoguchi, A. & Okamoto, N. Insulin-Like And IGF-Like Peptides In The Silkmoth *Bombyx mori*: Discovery, Structure, Secretion, And Function. *Front Physiol* **4**, (2013).

79. Meng, X., Zhu, F. & Chen, K. Silkworm: A Promising Model Organism in Life Science. *J Insect Sci* **17**, (2017).

80. Zhang, X. *et al.* Silkworms Can Be Used As An Animal Model To Screen And Evaluate Gouty Therapeutic Drugs. *J Insect Sci* **12**, (2012).

81. Li, B. *et al.* DNA Methylation Is Correlated with Gene Expression During Diapause Termination of Early Embryonic Development in the Silkworm (*Bombyx mori*). *Int J Mol Sci* **21**, (2020).

82. Ma, S. *et al.* CRISPR/Cas9 Mediated Multiplex Genome Editing And Heritable Mutagenesis of BmKu70 in Bombyx mori. *Sci. Rep. 2014 4:1* **4**, 1–6 (2014).

83. Wu, P. *et al.* DNA methylation in Silkworm Genome May Provide Insights Into Epigenetic Regulation Of Response to Bombyx mori Cypovirus Infection. *Sci. Rep. 2017 7:1* **7**, 1–13 (2017).

84. Yang, S. Y. *et al.* Demographic History And Gene Flow During Silkworm Domestication. *BMC Evol Biol* **14**, 185 (2014).

85. Lu, K. *et al.* Flight Muscle and Wing Mechanical Properties are Involved in Flightlessness of the Domestic Silkmoth, Bombyx mori. *Insects* **11**, (2020).

86. Xia, Q. *et al.* Complete Resequencing Of 40 Genomes Reveals Domestication Events And Genes In Silkworm (Bombyx). *Science (1979)* **326**, 433–436 (2009).

87. Botlagunta, M., Ghosh, S. & Kundu, S. Molecular Phylogeny Of Silk-Producing Insects Based On 16S Ribosomal RNA And Cytochrome Oxidase Subunit I Genes. *J Genet* **85**, 31–38 (2006).

88. Kumar, R., Singh, B. K., Ahmed, S. & Pathania, P. Diversity And Their Clarification Of Species Genus Samia (Lepidoptera : Saturniidae)in India and Their Prospects For Utilization. *Journal of Insect Science* **30**, (2017).

89. Pradeep, A. R. *et al.* Genetic Analysis Of Scattered Populations Of The Indian Eri Silkworm, *Samia cynthia ricini* Donovan: Differentiation of Subpopulations. *Genet Mol Biol* **34**, 502 (2011).

90. Samia cynthia. https://tpittaway.tripod.com/silk/s_cyn.htm.

91. Renuka, G. & Shamitha, G. Studies On The Economic Traits Of Eri Silkworm, *Samia Cynthia Ricini*, In Relation To Seasonal Variations. *Int J Adv Res (Indore)* **2**, 315–322 (2014).

92. Vijayan, K. *et al.* Genetic Diversity And Differentiation Among Populations Of The Indian Eri Silkworm, *Samia cynthia ricini*, Revealed By ISSR Markers. *Journal of Insect Science* **6**, 30 (2006).

93. Lee, J., Kiuchi, T., Kawamoto, M., Shimada, T. & Katsuma, S. Accumulation Of Uric Acid In The Epidermis Forms The White Integument Of *Samia ricini* Larvae. *PLoS One* **13**, e0205758 (2018).

94. Lee, J. *et al.* The Genome Sequence Of *Samia Ricini*, A New Model Species Of Lepidopteran Insect. (2020) doi:10.22541/au.158802211.15119012.

95. Bang, C. *et al.* Metaorganisms In Extreme Environments: Do Microbes Play A Role In Organismal Adaptation? *Zoology* **127**, 1–19 (2018).

96. Berg, G. *et al.* Role Of The Gut Microbiota In Nutrition And Health. *BMJ* **361**, 36–44 (2018).

97. Berg, G. *et al.* Microbiome Definition Re-Visited: Old Concepts And New Challenges. *Microbiome* **8**, 1–22 (2020).

98. Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. Defining the Human Microbiome. *Nutr Rev* **70**, S38 (2012).

99. de Vos, W. M., Tilg, H., van Hul, M. & Cani, P. D. Gut Microbiome And Health: Mechanistic Insights. *Gut* **71**, 1020–1032 (2022).

100. Rowland, I. *et al.* Gut Microbiota Functions: Metabolism Of Nutrients And Other Food Components. *Eur J Nutr* **57**, 1 (2018).

101. Veziant, J., Villéger, R., Barnich, N. & Bonnet, M. Gut Microbiota as Potential Biomarker and/or Therapeutic Target to Improve the Management of Cancer: Focus on Colibactin-Producing *Escherichia coli* in Colorectal Cancer. *Cancers (Basel)* **13**, (2021).

102. Conteville, L. C., Oliveira-Ferreira, J. & Vicente, A. C. P. Gut Microbiome Biomarkers And Functional Diversity Within An Amazonian Semi-Nomadic Hunter-Gatherer Group. *Front Microbiol* **10**, 1743 (2019).

103. Moran, N. A., Degnan, P. H., Santos, S. R., Dunbar, H. E. & Ochman, H. The Players In A Mutualistic Symbiosis: Insects, Bacteria, Viruses, And Virulence Genes. *Proc Natl Acad Sci U S A* **102**, 16919–16926 (2005).

104. Bertino-Grimaldi, D. *et al.* Bacterial Community Composition Shifts In The Gut Of *Periplaneta americana* Fed On Different Lignocellulosic Materials. *Springerplus* **2**, 1–11 (2013).

105. Nelson, K., Muge, E. & Wamalwa, B. Cellulolytic Bacillus Species Isolated From The Gut Of The Desert Locust Schistocerca gregaria. *Sci Afr* **11**, e00665 (2021).

106. Xue, H. *et al.* Gut Bacterial Diversity in Different Life Cycle Stages of Adelphocoris suturalis (Hemiptera: Miridae). *Front Microbiol* **12**, 670383 (2021).

107. Li, Y. H. & Tian, X. Quorum Sensing and Bacterial Social Interactions in Biofilms. *Sensors (Basel)* **12**, 2519 (2012).

108. Klein, S. L. & Flanagan, K. L. Sex Differences In Immune Responses. *Nat. Rev. Immun. 2016 16:10* **16**, 626–638 (2016).

109. Anand, A. A. P. *et al.* Isolation And Characterization Of Bacteria From The Gut Of Bombyx Mori That Degrade Cellulose, Xylan, Pectin And Starch And Their Impact On Digestion. *J. Ins. Sci.* **10**, (2010).

110. Kristensen, N. P. Lepidoptera, moths and butterflies: Volume 2: Morphology, physiology, and development. *Lepidoptera, Moths and Butterflies: Volume 2: Morphology, Physiology, and Development* 1–564 (2012) doi:10.1515/9783110893724/HTML.

111. Barbehenn, R. v. & Kristensen, N. P. Digestive And Excretory Systems. *Lepidoptera, Moths and Butterflies: Volume 2: Morphology, Physiology, and Development* 165–187 (2012) doi:10.1515/9783110893724.165/HTML.

112. Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M. & Fatouros, N. E. Bacterial Symbionts In Lepidoptera: Their Diversity, Transmission, And Impact On The Host. *Front Microbiol* **9**, 556 (2018).

113. Yang, F. Y. *et al.* Differential Profiles of Gut Microbiota and Metabolites Associated with Host Shift of *Plutella xylostella*. *Int J Mol Sci* **21**, 1–15 (2020).

114. Nagao-Kitamoto, H., Kitamoto, S., Kuffa, P. & Kamada, N. Pathogenic Role Of The Gut Microbiota In Gastrointestinal Diseases. *Intest Res* **14**, 127 (2016).

115. Bjornson, S. & Oi, D. Microsporidia Biological Control Agents and Pathogens of Microsporidia Biological Control Agents and Pathogens of Beneficial Insects Beneficial Insects. (2014).

116. Jaffar, S., Ahmad, S. & Lu, Y. Contribution Of Insect Gut Microbiota And Their Associated Enzymes In Insect Physiology And Biodegradation Of Pesticides. *Front Microbiol* **13**, 3495 (2022).

117. Ganeshprasad, D. N. & Jani, K. Gut Bacterial Inhabitants of Open Nested Honey Bee, Apis Florea. (2021) doi:10.21203/rs.3.rs-225332/v1.

118. Chen, B. *et al.* Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore Spodoptera littoralis. *Sci. Rep. 2016 6:1* **6**, 1–14 (2016).

119. Sun, X. *et al.* Relationship between Changes in Intestinal Microorganisms and Effect of High Temperature on the Growth and Development of Bombyx mori Larvae. *Int J Mol Sci* **23**, 10289 (2022).

120. Devi, S., Saini, H. S. & Kaur, S. Insecticidal and growth inhibitory activity of gut microbes isolated from adults of Spodoptera litura (Fab.). *BMC Microbiol* **22**, 1–14 (2022).

121. Yuning, L. *et al.* The Bacterial And Fungal Communities Of The Larval Midgut Of *Spodoptera Frugiperda* (Lepidoptera: Noctuidae) Varied By Feeding On Two Cruciferous Vegetables. *Sci. Rep. 2022 12:1* **12**, 1–12 (2022).

122. Mason, C. J. *et al.* Diet Influences Proliferation And Stability Of Gut Bacterial Populations In Herbivorous Lepidopteran Larvae. *PLoS One* **15**, e0229848 (2020).

123. Anand, A. A. P. *et al.* Isolation and Characterization of Bacteria from the Gut of Bombyx mori that Degrade Cellulose, Xylan, Pectin and Starch and Their Impact on Digestion. *J. Ins. Sci.* **10**, 107 (2010).

124. Saad, M. S. I., Elyamani, E. M. Y. & Helaly, W. M. M. Controlling Of Bacterial And Fungal Diseases That Contaminating Mulberry Silkworm, Bombyx mori By Using Some Plant Extracts. *Bulletin of the Nat. Res. Centre 2019 43:1* **43**, 1–9 (2019).

125. Bonilla-Rosso, G., Eguiarte, L. E., Romero, D., Travisano, M. & Souza, V. Understanding Microbial Community Diversity Metrics Derived From Metagenomes: Performance Evaluation Using Simulated Data Sets. *FEMS Microbiol Ecol* **82**, 37–49 (2012).

126. Prosser, J. I. *et al.* The Role Of Ecological Theory In Microbial Ecology. *Nature Reviews Microbiology 2007 5:5* **5**, 384–392 (2007).

127. Microbial Threats, I. of M. (US) F. on. Microbial Ecology and Ecosystems. (2009).

128. Gich, F., Janys, M. A., König, M. & Overmann, J. Enrichment Of Previously Uncultured Bacteria From Natural Complex Communities By Adhesion To Solid Surfaces. *Environ Microbiol* **14**, 2984–2997 (2012).

129. Eldin, C. *et al.* From Q Fever to *Coxiella burnetii* Infection: A Paradigm Change. *Clin Microbiol Rev* **30**, 115 (2017).

130. Simon, C. & Daniel, R. Metagenomic Analyses: Past and Future Trends. *Appl Environ Microbiol* **77**, 1153 (2011).

131. Dance, A. The Search For Microbial Dark Matter. *Nature* **582**, 301–303 (2020).

132. Christen, R. Analyzing natural microbial populations by rRNA sequences. *ASM News* **51**, 4–12 (1985).

133. Handelsman, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews* **68**, 669 (2004).

134. Chen, K. & Pachter, L. Bioinformatics For Whole-Genome Shotgun Sequencing Of Microbial Communities. *PLoS Comput Biol* **1**, 0106–0112 (2005).

135. Jiménez, D. J. *et al.* Structural and Functional Insights from the Metagenome of an Acidic Hot Spring Microbial Planktonic Community in the Colombian Andes. *PLoS One* **7**, e52069 (2012).

136. Ni, J. *et al.* Metagenomic Analysis Of Microbial Community Structure And Function In A Improved Biofilter With Odorous Gases. *Sci. Rep. 2022 12:1* **12**, 1–11 (2022).

137. Tripathi, L. K. & Nailwal, T. K. Metagenomics: Applications Of Functional And Structural Approaches And Meta-Omics. *Recent Adv. in Micro. Div.* 471–505 (2020) doi:10.1016/B978-0-12-821265-3.00020-7.

138. Nasuno, E. *et al.* Phylogenetically Novel LuxI/LuxR-Type Quorum Sensing Systems Isolated Using a Metagenomic Approach. *Appl Environ Microbiol* **78**, 8067 (2012).

139. Ramachandran, G. & Bikard, D. Editing The Microbiome The CRISPR Way. *Philos. Trans. Royal Soc. B: Biol. Sci.* **374**, (2019).

140. di Guglielmo, M. D., Franke, K., Cox, C. & Crowgey, E. L. Whole Genome Metagenomic Analysis Of The Gut Microbiome Of Differently Fed Infants Identifies Differences In Microbial Composition And Functional Genes, Including

An Absent CRISPR/Cas9 Gene In The Formula-Fed Cohort. *Hum Microb J* **12**, 100057 (2019).

141. Chistoserdova, L. Biotechnology and Genetic Engineering Reviews Functional Metagenomics: Recent Advances and Future Challenges. *Biotechnol Genet Eng Rev* **26**, 335–352.

142. Prayogo, F. A. *et al.* Metagenomic Applications In Exploration And Development Of Novel Enzymes From Nature: A Review. *Journal of Genetic Engineering & Biotechnology* **18**, (2020).

143. de Abreu, V. A. C., Perdigão, J. & Almeida, S. Metagenomic Approaches to Analyze Antimicrobial Resistance: An Overview. *Front Genet* **11**, (2020).

144. Berglund, F. *et al.* Identification And Reconstruction Of Novel Antibiotic Resistance Genes From Metagenomes. *Microbiome* **7**, 1–14 (2019).

145. Teeling, H. & Glöckner, F. O. Current Opportunities And Challenges In Microbial Metagenome Analysis—A Bioinformatic Perspective. *Brief Bioinform* **13**, 728 (2012).

146. Terrón-González, L., Genilloud, O. & Santero, E. Potential And Limitations Of Metagenomic Functional Analyses. in 1–43 (2014).

147. Hassa, J. *et al.* Metagenome, Metatranscriptome, And Metaproteome Approaches Unraveled Compositions And Functional Relationships Of Microbial Communities Residing In Biogas Plants. *Appl. Microbiol. Biotechnol.* vol. 102 5045–5063 Preprint at https://doi.org/10.1007/s00253-018-8976-7 (2018).

148. Fulci, V. Meta'omics: Challenges and Applications. *Int. J. Mol. Sci.* vol. 23 Preprint at https://doi.org/10.3390/ijms23126486 (2022).

149. Segata, N. *et al.* Computational Meta'omics For Microbial Community Studies. *Mol Syst Biol* **9**, 666 (2013).

150. Zhang, X., Li, L., Butcher, J., Stintzi, A. & Figeys, D. Advancing Functional And Translational Microbiome Research Using Meta-Omics Approaches. *Microbiome 2019 7:1* **7**, 1–12 (2019).

151. Djemiel, C. *et al.* Potential of Meta-Omics to Provide Modern Microbial Indicators for Monitoring Soil Quality and Securing Food Production. *Front Microbiol* **13**, 2349 (2022).

152. Thomas, T., Gilbert, J. & Meyer, F. Metagenomics - A Guide From Sampling To Data Analysis. *Microb Inform Exp* **2**, 3 (2012).

153. Head, S. R. *et al.* Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* **56**, 61 (2014).

154. Navgire, G. S. *et al.* Analysis And Interpretation Of Metagenomics Data: An Approach. *Biological Procedures Online 2022 24:1* **24**, 1–22 (2022).

155. Luz Calle, M. Statistical Analysis of Metagenomics Data. *Genomics Inform* **17**, (2019).

156. Pereira, M. B., Wallroth, M., Jonsson, V. & Kristiansson, E. Comparison Of Normalization Methods For The Analysis Of Metagenomic Gene Abundance Data. *BMC Genomics* **19**, 1–17 (2018).

157. Li, H. Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *https://doi.org/10.1146/annurev-statistics-010814-020351* **2**, 73–94 (2015).

158. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* **8**, 2224 (2017).

159. Odintsova, V., Tyakht, A. & Alexeev, D. Guidelines to Statistical Analysis of Microbial Composition Data Inferred from Metagenomic Sequencing. *Curr Issues Mol Biol* **24**, 17–36 (2017).

160. Roberts, F. S. Measurement of Biodiversity: Richness and Evenness. 203–224 (2019) doi:10.1007/978-3-030-22044-0_8.

161. Ruhl, I. A. *et al.* Microbial Functional Diversity Correlates with Species Diversity along a Temperature Gradient. *mSystems* **7**, (2022).

162. Bonilla-Rosso, G., Eguiarte, L. E., Romero, D., Travisano, M. & Souza, V. Understanding Microbial Community Diversity Metrics Derived From

Metagenomes: Performance Evaluation Using Simulated Data Sets. *FEMS Microbiol Ecol* **82**, 37–49 (2012).

163. Yang, C. *et al.* A Review Of Computational Tools For Generating Metagenome-Assembled Genomes From Metagenomic Sequencing Data. *Comput Struct Biotechnol J* **19**, 6301–6314 (2021).

164. Sun, Y. v. & Hu, Y. J. Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases. *Adv Genet* **93**, 147–190 (2016).

165. Krassowski, M., Das, V., Sahu, S. K. & Misra, B. B. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Front Genet* **11**, 1598 (2020).

166. Song, M. *et al.* A Review of Integrative Imputation for Multi-Omics Datasets. *Front Genet* **11**, 1215 (2020).

Chapter 2

Comparative Microbiome Analyses of *Samia cynthia ricini* Under Different Environmental Conditions

Chapter 2

Comparative microbiome analyses of *Samia cynthia ricini* under different diet conditions and growth stages

ABSTRACT

The domesticated silkworm, *Samia cynthia ricini*, is known for its production of unique warm Eri-silk, making it a vital contributor to the ecological and economic aspects of Assam and the north-eastern regions of India.

Gut microbiota plays an important role in the proper growth and development of organisms. Thus, in this study, we aimed to gain a comprehensive insight into the community gut microbiome of polyphagous *S. cynthia ricini* and explore potential changes based on two different diet types (Eri and Kesseru leaves) and the growth stages (Larva and Moth). Understanding of the symbiotic relationship between the Eri silkworm and its gut microbiota, emphasizing its ecological significance and potential applications in sericulture practices.

A total of 6341 features were generated. 11 Operational Taxonomic Units (OTUs) were consistently observed across all samples, belonging to phylum Proteobacteria, Firmicutus, Actinobacter, and Bacteriodetes. Notably, no significant differences in species diversity were observed among samples under different variables, such as larval instar and moth stages, diet-types, and larva-to-moth groups. However, moderate to high community diversity was noted among these groups, with the most considerable variation observed between larval and moth samples (p=0.001).

Differential abundance analysis revealed 20 significantly differentially abundant species, with [Thermi], Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Tenericutes, and WPS-2 standing out in relation to diet. Particularly, Bacteroidetes, Oscillospira, Prevotella and Lachnospira exhibited higher abundance in the adult stages, with Eri-fed female moths showing significantly elevated levels.

Predictive functional diversity between diet-types was also explored, revealing that pathways related to carbohydrate metabolism were more abundant in Kesseru-fed samples. On the other hand, larval microbiome-based functional predictions displayed prevalent pathways related to the degradation of aromatic compounds and detoxification. In contrast, moth samples showed increased abundance of various biosynthesis pathways, proteins associated with RNA transport, and immunogenic functions.

This study provides valuable insights into the gut microbiome of *S. cynthia ricini* and highlights the influence of diet and growth stages on microbial composition and functional potential.

INTRODUCTION

Organisms, currently recognised as metaorganism are defined by the eukaryotic organism coevolved with the microbiota that shapes its state of being. Every metaorganism comprises of a complex network of organisms and their metabolites synergistically interdependent on each other in order to survive in the environment along with modifying it[1]. Organisms and their microbiome co-exist and have parallel evolutionary histories represented by co-phylogenetic patterns and are reported to have undergone co-speciation[2–4].

Holometabolous insects are a major influence in the ecosystem They occupy multiple niches in the ecosystem forming completely different phenotypes in the larval and adult stages[5,6]. With ~18000 described species over the world, lepidopterans are the second most diverse insect group[7]. However, only <0.1% lepidopterans have been reported of their microbial assistance. The composition and significance of gut microbiome varies vastly among lepidopterans with no universal pattern[8].

As the phenotype and physiology of insects are dependent on the microbial composition, the composition and abundance of the organism-associated microbiota especially gut is dependent on the insect habitat, immunity and other physiochemical conditions[9–11]. Studies have reported variations in lepidopteran gut microbiome with the changes in different conditions[7,8,12–15]. Most studies report the complete remodel of the gut microbiota according to the development changes and have provided different theories about the gut microbiome composition, transmission and coevolution of these organisms[8,15].

The most dominant microbes present in most larval stages of lepidopteran insects are digestion aiding microbes. Strains of gut microbes such as *Enterobacter, E. casseliflavus, Burkholderia, Pseudomonas, Carnobacterium,* etc. from various insect gut were reported to degrade and detoxify plant compounds and toxins such as hydrogen cyanide (HCN), latex, resin acids, plants phenolics and other organic acids[12,16,17]. Gut microbiome has been known to play a profound role in mating and socializing behaviour of organisms. Microbial strains *Klebsiella, Pantoea, Enterococcus* were found comparatively in higher amounts in *Spodoptera* and other insect gut[8,15,17].

Studies based on functional analyses of the gut microbiome of lepidopterans, reported 85-97% of active species. Early larval stages reported to higher presence of functions involved in carbohydrate metabolism, cell motility while late instar microbes provided assistance with functions involved in amino acid, cofactors and vitamin metabolism. In moths, functions such as energy metabolism, replication pathways were more abundant[15,18,19].

Studying the gut microbiota is used to many advantages. In addition to gaining insights into the biology, evolution and dynamics of the host-gut microbiome, insect gut microbiota analyses could be used in developing effective methods of their better growth and health [20,21]. They could be used as biomarkers for various traits of the organism and could be used to determine and eliminate various unwanted traits like diseases, pests, weak immunity and reduced growth[22–26]. Novel microbes and their products from insect gut have also been developed for biodegradation of complex waste and blocking transmission of insect-borne diseases[27–31].

S. cynthia ricini (Lepidoptera: Saturniidae), a domesticated silkworm, producing unique warm Eri-silk is widely cultivated and by the rural population of Assam and Northeast India and largely contribute to the ecological and economic aspects of these regions[32–34]. However, very little is known about the respective organisms resulting in their poor and obsolete rearing practice and cultivation techniques and low productivity of silk. The health conditions of these organisms are weakening leading to decrease in their population which also has negative effects on the environment due to their role in the ecosystem[35,36].

In this study, the microbiome of the different growth stages of the polyphagous domesticated lepidopteran insect has been analyzed with respect to two different types of leaf diet.

The organism was reared in a controlled environment of room temperature 25˚C and 80% relative humidity. They were grouped into two different diet groups of *Ricinus communis Linn.* (Eri) and *Heteropanax fragrans* (Kesseru) as they are the most commonly fed food in the sericulture farms. Eri leaves are usually fed during the early larval stages for excess growth, cocoon quality and egg production while Kesseru leaves are fed during the later stages with the aim for a commercial crop production [37,38]. Larval and adult growth stages of both the group were analyzed in order to explore various traits including the core microbiome, temporal changes in the microbial composition with their growth stages and differences with respect to their diets.

Objectives of the study include-

- Understanding the gut microbiota of *S. cynthia ricini* through amplicon sequencing and establishing the core microbiota involved.
- To analyze the changes in the gut microbiota structure through various diet conditions and growth stages.
- To predict the functions of the microbes involved in the phenotype changes through the changes in diet and growth stages.

METHODOLOGY

Sample collection, rearing and growth characteristics observation

Eri silkworm (*S. cynthia ricini*) diseased-free layings (dfls) were acquired from Central Muga & Eri Research Institute, Jorhat, Assam, India (Lat: 26° 47′49.1″N Lon: 94°

19′35.0″E). They were reared in a domesticated semi-controlled environment with a room temperature of 25˚C and 80% relative humidity. Two types of leaf feeds were selected viz. *Ricinus communis Linn.* (Eri) and *Heteropanax fragrans* (Kesseru). Phenotypic differences including average length of each larval instar stages and egg viability were recorded for a comparative analysis of the growth rates of the two different diets. 14 samples were collected based on each larval and adult growth from the two different diet-types. A summarized table with sample name and description is given by **Table2.1**

Sample Name	Diet	Growth-Stage
SCLE1	Eri Leaves	Instar 1
SCLE2	Eri Leaves	Instar 2
SCLE3	Eri Leaves	Instar 3
SCLE4	Eri Leaves	Instar 4
SCLE5	Eri Leaves	Instar 5
SCLEF	Eri Leaves	Moth Female
SCLEM	Eri Leaves	Moth Male
SCLK1	Kesseru Leaves	Instar 1
SCLK2	Kesseru Leaves	Instar 2
SCLK3	Kesseru Leaves	Instar 3
SCLK4	Kesseru Leaves	Instar 4
SCLK5	Kesseru Leaves	Instar 5
SCLKF	Kesseru Leaves	Moth Female
SCLKM	Kesseru Leaves	Moth Male

Table2.1: Sample-metadata summarising sample names and variable categories.

DNA extraction and sequencing

Pooled samples based on equal weight were collected and stored from each instar of the larval stages and adult male and female moths fed on each diet-type for gut DNA extraction. Collected larval and moth samples were washed with 70% alcohol to minimize external microbial contamination[39,40]. DNA extraction was carried out using phenol/chloroform (PHEC) extraction methodology [41]. Cells were homogenized using (mention the specification and condition) in sterile phosphate-buffer solution (PBS) followed by SDS-based cell-membrane degradation. The extracts were then saturated with phenol (write condition) followed by separation of nucleic acids and proteins using chloroform: isoamyl-alcohol (24:1). Overnight precipitation of nucleic acids using isopropanol and 3 M sodium chloride for maximum DNA recovery was carried out. DNA samples were analyzed for their concentration and purity using 1% agarose gel electrophoresis and Nanodrop spectrophotometer[42,43]. Extracted DNA samples (A260/A280 ~1.8) were sent for 16s rRNA gene amplicon sequencing using Illumina NGS MiSeq (AgriGenome Technologies Pvt. Ltd.).

Sequence retrieval, quality-control and processing

Paired-end raw sequencing data (250bp) in FASTQ format were analyzed for quality and adapters using FastQC(v_0.11.9)[44]. Following removal of non-biological contaminants, the quality-controlled reads were imported to QIIME2-2022.2 with metadata mapping file including sample names, adapter barcodes and mapping categories of Life-stage (Larvae, Moth), Growth stages (Instar1-5, Moth_male, Moth_female), Diet (Eri_Leaves, Kesseru_Leaves)[45]. Sample metadata was modified and filtered accordingly based on different group analyses.

Reads were denoised for quality control and feature-table construction using DADA2 plugin thus mapping the feature identifiers (Amplicon Sequence Variants or ASVs) to the sequence reads they represent[46]. For quality filtering based on the FASTQC analyses, reads were truncated around 120bp. Obtained feature-table were subjected to denovo chimera check and removal followed by v-search open-reference OTU clustering using 99% identity reference sequences from Greengenes database to reduce FDRs[47,48]. The obtained clustered sequences were then filtered for mitochondrial and chloroplast sequences for further improved dataset.

Taxonomy classification and identification were done by training the 99% identity reference sequences from Greengenes database using Naive Bayes classifier[49]. Core microbiome was analyzed at 100% identity. Shared OTU numbers of different sample groups were visualised using Venn[50].

Compositional analyses

The obtained data were subjected to various exploratory and significance analyses with respect to various groups of metadata viz., growth stages (each larval instar stages and adult moths), life stages (compiled larval stages vs adults) and diet-types (Eri vs Kesseru leaves).

Taxonomic compositions were visualized using QIIME2 and R (v-4.2.0) phyloseq package[51,52]. The OTU outputs were aligned with multiple sequence alignment tool, mafft and a phylogeny with fasttree2 based on maximum likelihood was constructed with q-phylogeny[53,54]. Taxonomy-based and phylogeny-based α (observed otus, Shannon, Faith PD) and β-diversity analyses (Jaccard, weighed and unweighed UniFrac distances) were performed using diversity-plugin of QIIME2 and visualised through boxplots and PCoA NMDS plots respectively[55–58]. Kruskal-Wallis group significances were calculated for observed OTUs, Shannon and faith-PD phylogenetic-distance α diversity indices based on the diet and various groups of growth stages. PERMANOVA significances based on Pseudo-F analyses were calculated using unweighed UniFrac phylogenetic distance[59,60].

The filtered feature-table, taxonomy abundance table were exported to biom-formatted files for further analyses. Differential abundance (DA) against variables diet-types and life-stages were calculated using DeSeq2. The data was transformed following addition of pseudocount to avoid errors followed by Wald's test with alpha=0.05 for significance[52,61]. Multiple variable (diet and life-stage) based DA was analyzed by ANCOM-BC. The samples were normalized using additive log ratio (alr) transformation along with incorporation of sampling fraction into the model, estimated by the ratio of the library size to the microbial load reducing biasness in the analyses and holm-method was used to adjust p-values for multiple comparisons[62]. 20 samples with minimum p-value from the DA analyses outputs were selected for mapping using ComplexHeatmap (R 4.2.0)[63]. Phylogenetic tree representing evolutionary relationship of the microbiome present was visualised using MEGAN Community Edition (6.23.2)[64].

Differential microbial networks for larval and moth stages were generated using NetComi package (R 4.2.0). The clr transformed data was used for calculating differentiating correlation by Pearsons and significance was determined by Fisher's tests[65].

Functional prediction

The qiime2 filtered ASV feature-table and their representative sequences were exported to biom-formats for functional prediction using PICRUSt2 using EPA-ng Maximum-Likelihood based phylogenetic placement[66].

NMDS plotting was done based on unconstrained RDA analysis of centred-log ration (clr) transformed data using phyloseq to analyze the functional diversity among the group of samples[52]. ALDEx2 was used to calculate the functional DA between sample groups of diet-types and life-stages as preferred over ANCOM by other studies for functional prediction data[67]. Welch's t-test was carried out for significance and the DA was plotted through Bland- Altman log-ratio abundance (MA) plots and Effect (MW) plots[68,69].

Kegg-Orthology predicted outputs (KO) were used to find the related functions using KEGG-brite hierarchal map-file. Significant KO-derived functions and Metacyc-pathways were calculated with corrected BHE=0.1 (p=0.05), sorted and plotted using ggplot for the above sample groups[70–72].

RESULTS AND DISCUSSION

Silkworm rearing, growth observations

The reared silkworms were observed for their larval growth-rate using length measurements. Differences in phenotypes including growth curves and egg-viability were observed based on the two different diets. Larvae reared in Eri leaves were observed to have a higher growth curve and egg-viability rate compared to larvae feeding on Kesseru leaves.

Sequence processing, quality control and bioinformatic analyses

Paired-end 250bp 113215- 477230 reads were generated from the samples. A summary of the filtered output following demultiplexing and denoising of the samples is provided in Table2.2.

Sample-id	Input	Filtered	Percentage of input passed filter	Denoised	Merged	Percentage of input merged	Non-chimeric	Percentage of input non-chimeric
SCLE1	274989	245925	89.43	234527	212445	77.26	176241	64.09
SCLE2	280585	228509	81.44	221840	213177	75.98	193690	69.03
SCLE3	229700	170838	74.37	165618	160699	69.96	143542	62.49
SCLE4	149293	103102	69.06	101425	97566	65.35	78317	52.46
SCLE5	246061	207266	84.23	203619	193794	78.76	184845	75.12
SCLEF	113215	85623	75.63	81183	66811	59.01	42276	37.34
SCLEM	330290	126527	38.31	119896	89418	27.07	32703	9.9
SCLK1	167516	120398	71.87	117746	108255	64.62	68185	40.7
SCLK2	152280	111113	72.97	109436	104636	68.71	80647	52.96
SCLK3	159959	111203	69.52	109329	101584	63.51	60491	37.82
SCLK4	168904	96377	57.06	93810	86834	51.41	44061	26.09
SCLK5	162808	102294	62.83	99241	91310	56.08	53879	33.09
SCLKF	477230	200533	42.02	192900	152748	32.01	60634	12.71
SCLKM	361644	154512	42.72	147553	114555	31.68	46117	12.75

Table2.2: Summary of denoised filtered output.

The DADA2 feature-table output was collapsed to significant OTUs using 99% Greengenes database as it provided a more comprehensive output with respect to diverse and rare bacterial taxonomies with respect to the samples. Reduced number of features of archaeal diversity, comparatively, was however observed. A comparative output of the collapsed and filtered-feature (mitochondria and chloroplasts) is shown (**Table2.3**). The filtered feature-table provides a summary of SCLE5 having the highest feature count while SCLEM having the lowest.

Metric	Ufiltered_Table	Filtered_Table
Number of Samples	14	14
Number of features	6341	1139
Total Frequency	1265628	980296

Table2.3: Summary of final feature-table filtered output following OTU collapse and removal of mitochondria, chloroplast.

Microbiome composition

Taxa bar plots for taxonomic abundances, the highest dominant show significant differences (**Figure2.1**). A large number of unclassified bacteria has been observed in the larval stages. Furthermore, distinct patterns of taxonomy have been observed in samples from different environmental conditions including larval instar stages, larva to adult moths and different diets. Core microbiome at 100% was seen to be comprising of 11 OTUs mostly comprising of phylum *Proteobacteria* and *Firmicutus* with other dominant microbes belonging to phylum Bacteroidetes and Actinobacteria (**Table2.4**). Presence of *Plesiomonas shigelloides* have been observed in large amounts of Kesseru fed larva and adults. (**Figure2.2**).

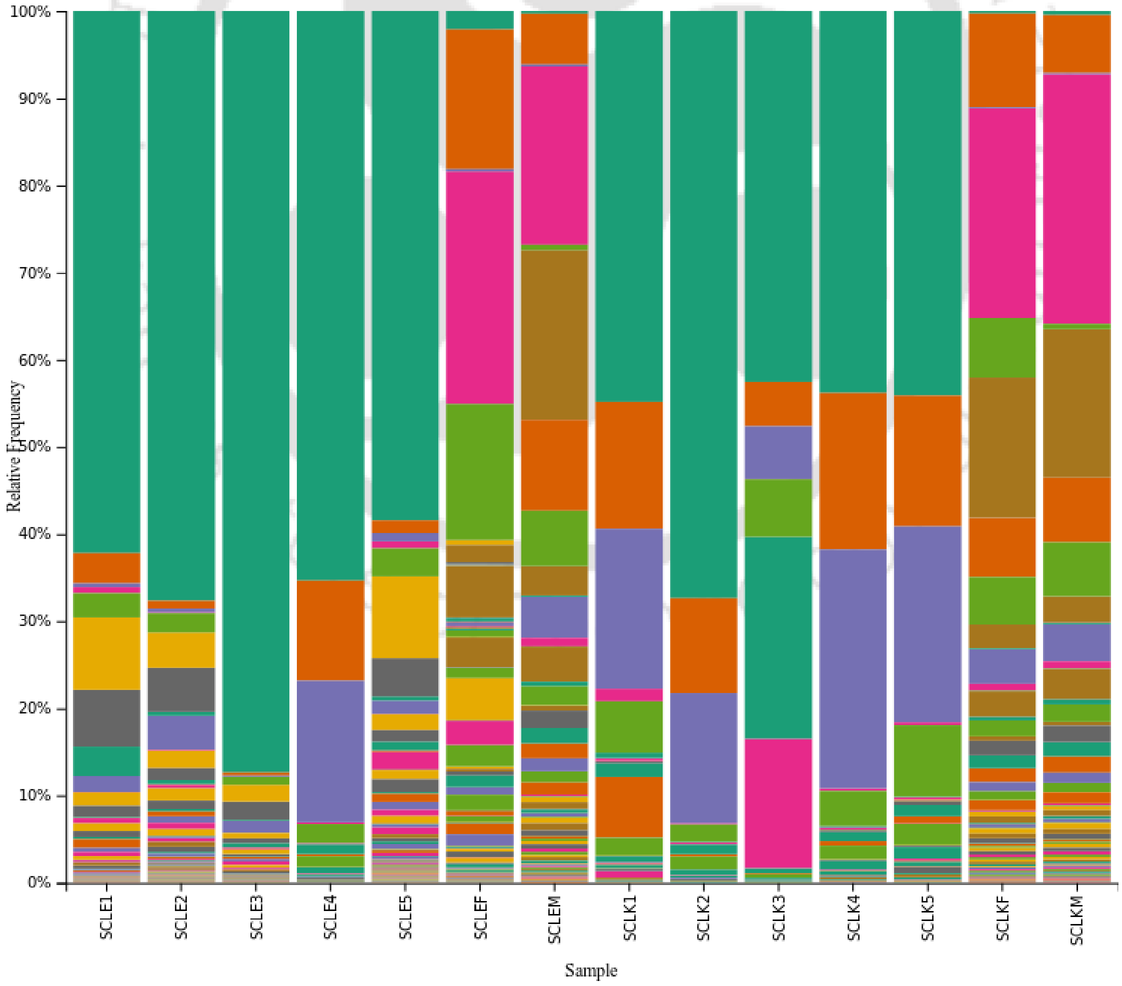


Figure2.1: Taxa-barplot representing microbial communities and the overall microbial diversity at the genus level (legend provided as supplementary svg file).

Feature ID	Taxonomy
656881	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__; s__
289174	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacteriales; f__Enterobacteriaceae; g__Plesiomonas; s__shigelloides
4438565	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudomonadaceae; g__Pseudomonas; s__
4406763	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__; s__
749805	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria
686593	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Rhizobiaceae; g__; s__
945326	k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae; g__; s__
73760	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__MKC10; f__; g__; s__
126133	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Moraxellaceae; g__Acinetobacter; s__
516809	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas; s__
212688	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__; s__

Table2.4: Core microbiome at 100% identity

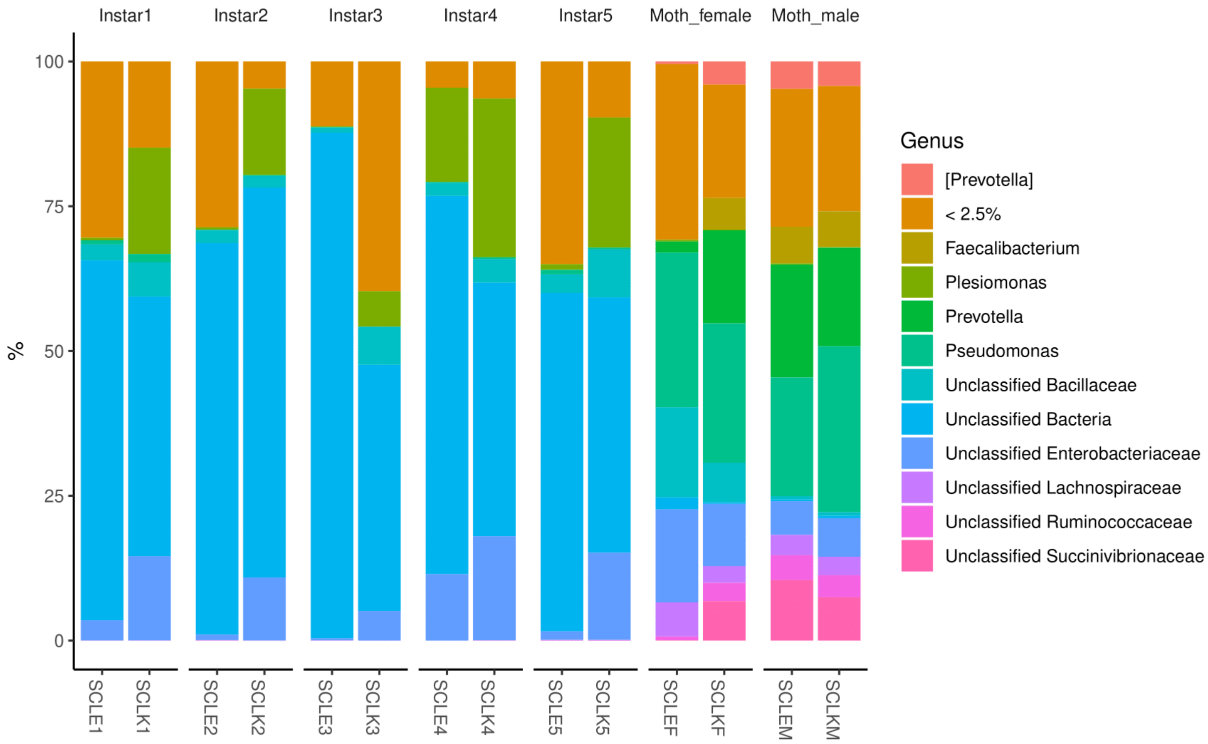


Figure2.2: Taxonomic abundance at the phylum level representing microbial community diversity among the samples. < 2.5% indicates the rare taxa in each group, with median relative abundance < 2.5%.

Variation in number of shared OTUs were observed with lowest observed in eri-fed larval stages while similar numbers were observed for the other groups viz. Kesseru-fed larval stages, adult moth from both the diet-types (**Figure2.3 A, B, C, D**).

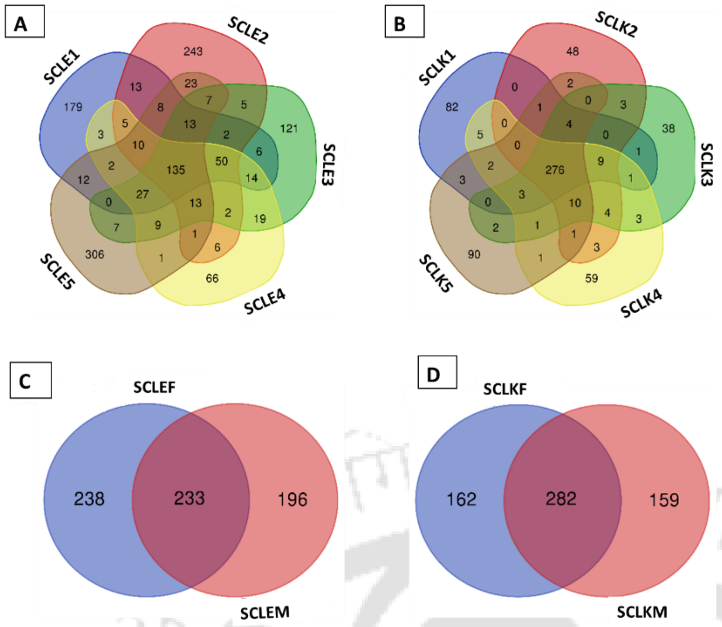
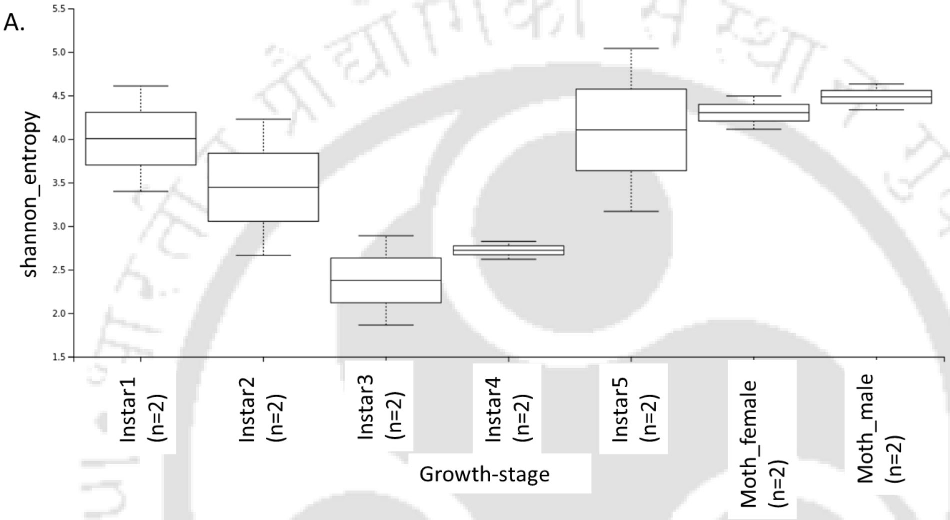


Figure2.3: Shared OTUs A. Eri-fed larval stages (SCLE1, SCLE2, SCLE3, SCLE4, SCLE5); B. Kesseru-fed larval stages (SCLK1, SCLK2, SCLK3, SCLK4, SCLK5); C. Adult moths- female and male from Eri-fed samples (SCLEF, SCLEM); Adult moths- female and male from Kesseru-fed samples (SCLKF, SCLKM).

Diversity analyses

Rarefaction at a sampling depth of 1000 was performed based on the observed features and rarefaction curve for generating diversity outputs at an even depth.

Species richness in each sample and groups of interest were analyzed through α-diversity indices. Observed OTU features, Shannon's and Faith-PD indices-based richness plots revealed the adult stages to be higher in species richness compared to the larval samples (p=0.08) (**Figures2.4C and 2.5C**). Lower values were observed in the middle larval stages (Instar3 and Instar4). Shannon's index-based richness was lowest in the 3rd instar while 4th instar was the lowest based on phylogenetic distance-based Faith PD analysis **(Figures2.4A and 2.5A)**. However, the differences through the larval stages were not significant. Species richness difference based on observed OTUs in samples fed on Kesseru leaves were significantly lower than those fed on Eri (p=0.05). However, Shannon's entropy difference exhibited no significance (p=0.2) **(Figure2.4B)**. Evenness, calculated using Pielou's index which considers both richness and diversity values found most groups to be moderately even sharing most features through one or more samples **(Figures2.6A, B, C)**. Lowest but unsignificant amount of evenness was observed between adult and larval stages (p=0.12).



Diet

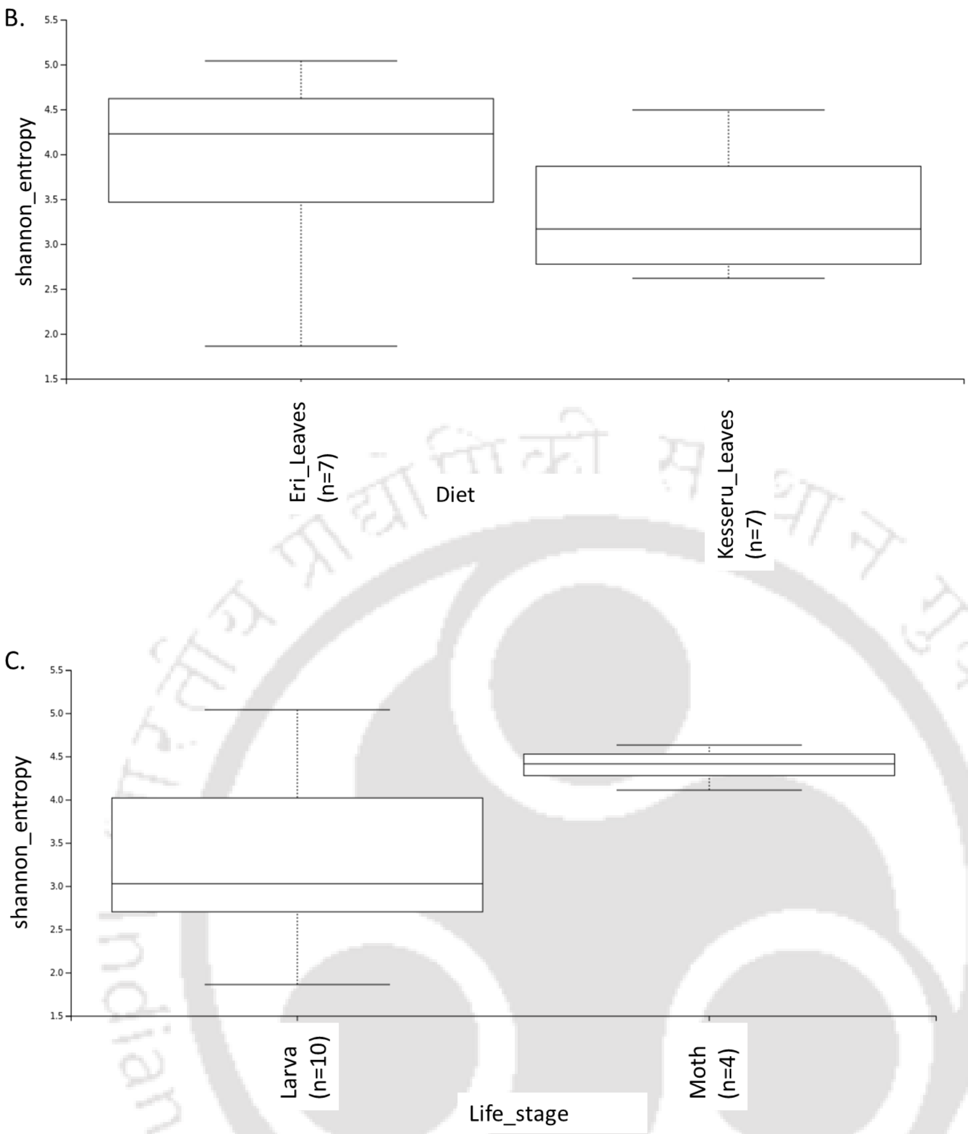
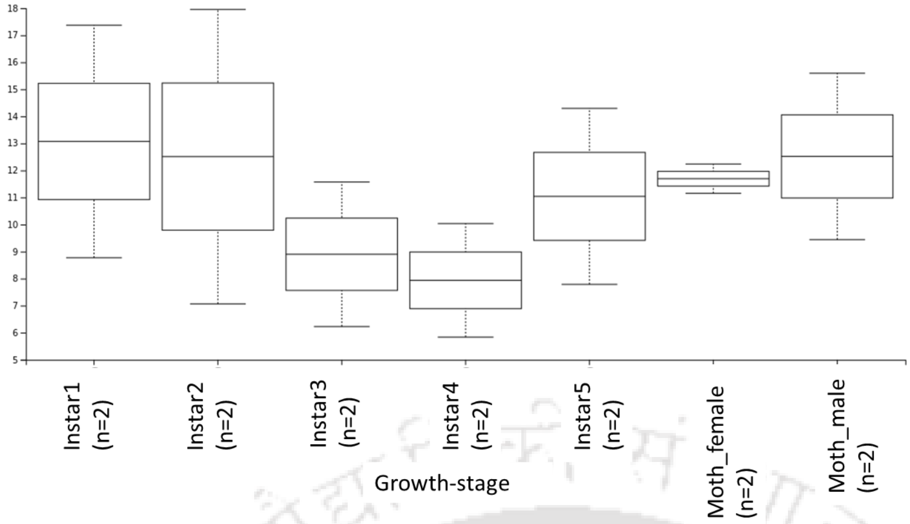
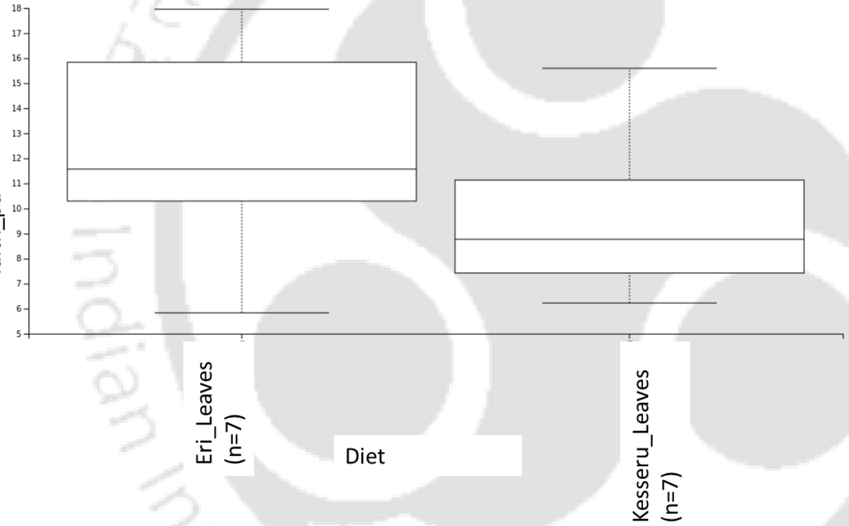


Figure2.4: Shannon's entropy depicting species richness among sample groups. A. Growth stages (Instar1, Instar2, Instar3, Instar4, Instar5, Moth female, Moth male) (p= 0.2); B. Diet (Eri-fed samples, Kesseru-fed samples) (p=0.2); C. Life-stages (Larvae, Moth) (p=0.08). [OTU-based]

A.


faith_pd
18
17
16
15
14
13
12
11
10
9
8
7
6
5
Instar1 (n=2)
Instar2 (n=2)
Instar3 (n=2)
Instar4 (n=2)
Instar5 (n=2)
Moth_female (n=2)
Moth_male (n=2)
Growth-stage

B.


faith_pd
18
17
16
15
14
13
12
11
10
9
8
7
6
5
Eri_Leaves (n=7)
Diet
Kesseru_Leaves (n=7)

C.

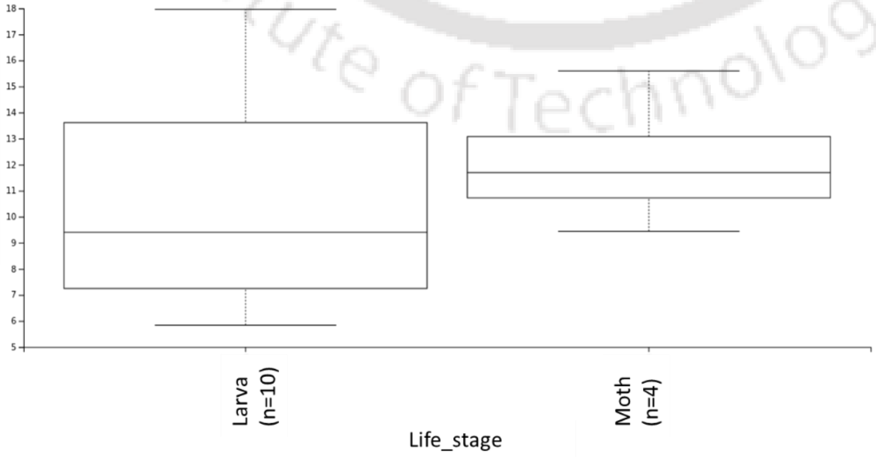
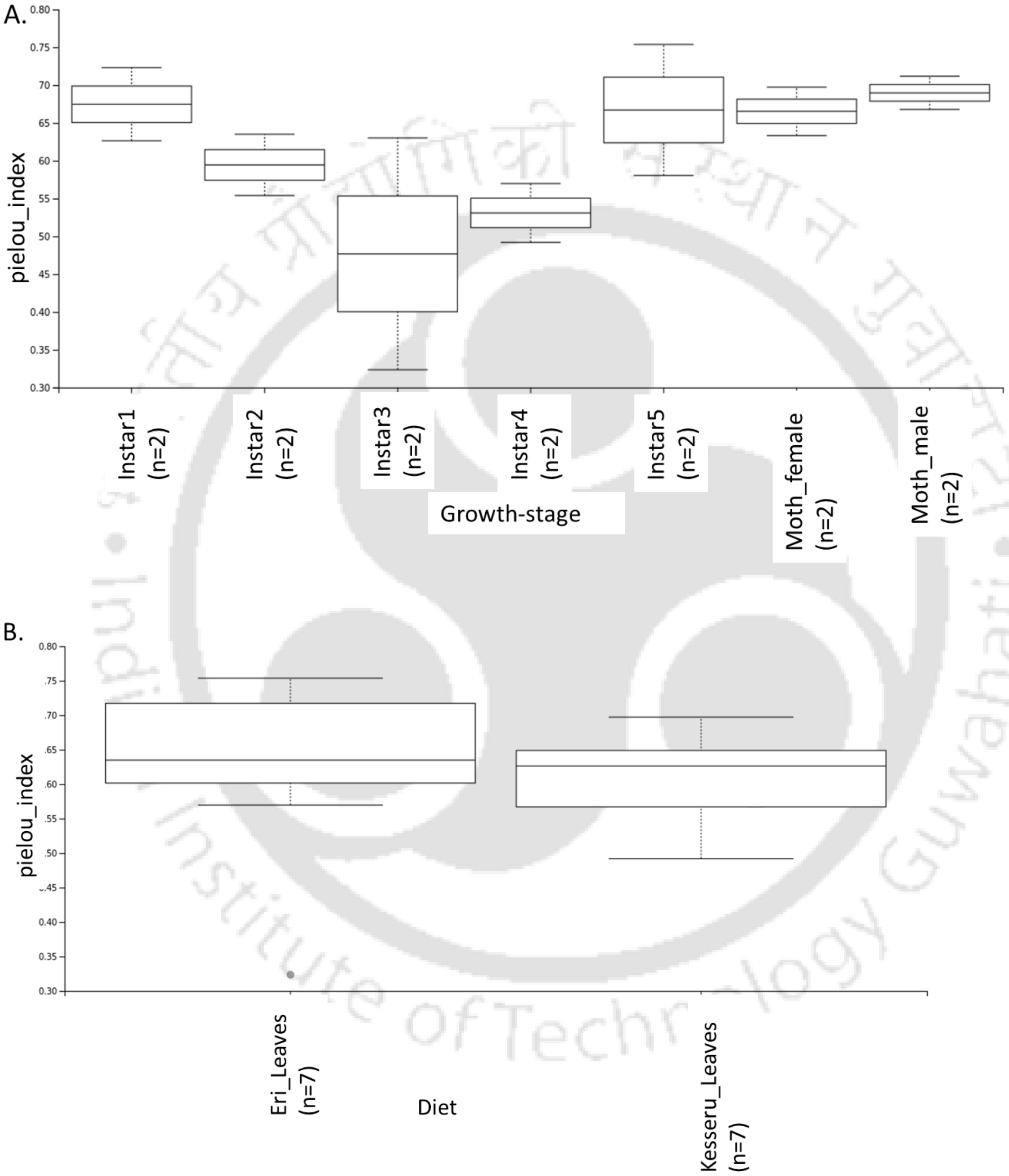

faith_pd
18
17
16
15
14
13
12
11
10
9
8
7
6
5
Larva (n=10)
Moth (n=4)
Life_stage

Figure2.5: Boxplots depicting species richness based on Faith's phylogenetic diversity (PD) among sample groups A. Growth stages (Instar1, Instar2, Instar3, Instar4, Instar5, Moth female, Moth male) (p=0.8); B. Diet (Eri-fed samples, Kesseru-fed samples) (p=0.2); C. Life-stages (Larvae, Moth) (p=0.39). [Phylogenetic-distance based].



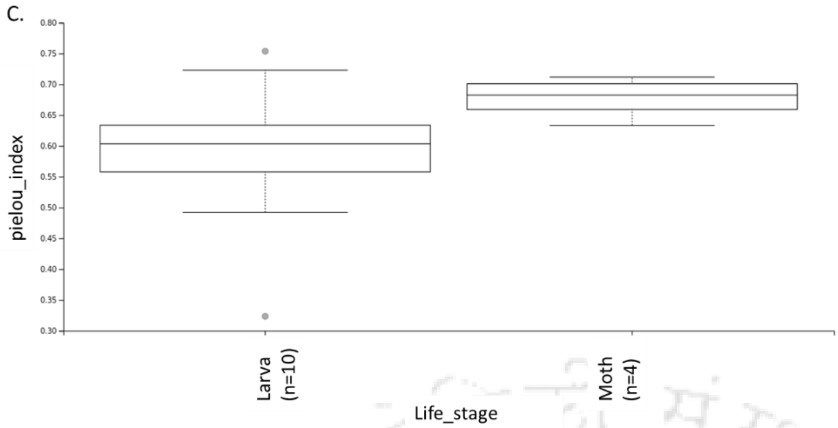


Figure2.6: Boxplots depicting species evenness based on Pielou's index among sample groups A. Growth stages (Instar1, Instar2, Instar3, Instar4, Instar5, Moth female, Moth male) (p=0.3); B. Diet (Eri-fed samples, Kesseru-fed samples) (p=0.27); C. Life-stages (Larvae, Moth) (p=0.11).

Diversity analyses through the community using phylogeny (weighted and un-weighted unifrac distance) and non-phylogeny-based (Jaccard's index) methods revealed the highest diversity to be between larvae and moth sample groups observed using NMDS PCoA (**Figures2.7A, B, C**). Species diversity between larval and moth sample groups was highly significant, calculated using PERMANOVA significance based on Weighted-Unifrac (p=0.01). Other groups including diet (p=0.13) and growth stages (p=0.16) were moderately significant. High diversity between larval and moth stages were reported in other studies of lepidoptera microbiome analyses.

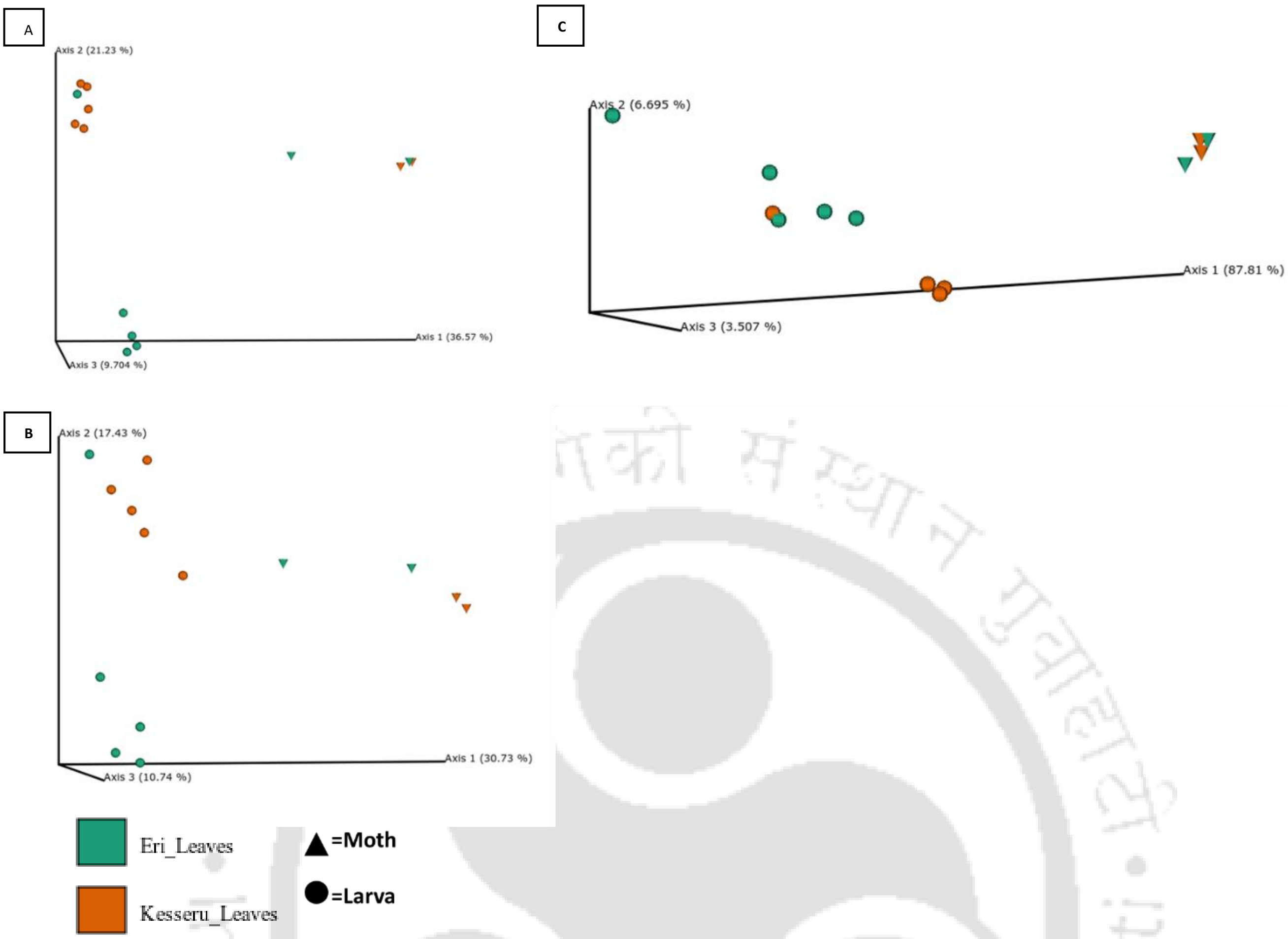


Figure2.7: PCoA plots depicting β-diversity among samples A. Jaccard's index; B. Weighed UniFrac Distance (phylogenetic-distance based quantitative measuring branch length); C. Unweighed UniFrac Distance (phylogenetic-distance based qualitative).

β-diversity significance between various group samples were carried out using PERMANOVA summarized in Table. Pseudo-F test statistic was used with 999 permutations **(Table2.5)**.

Group	p-value	Test statistic
Growth stages	0.167	1.271346
Diet	0.13	1.44053
Life-stages (Larva vs Moth)	0.001	4.94633

Table2.5: Summary of β-diversity significance by PERMANOVA between different sample groups. Sample size=14. Number of groups for growth stages=7, diet=2, life-stages=2.

Phylogenetic trees showing sample group diversity and taxonomic abundances were visualised using MEGAN6 **(Figure2.8)**.

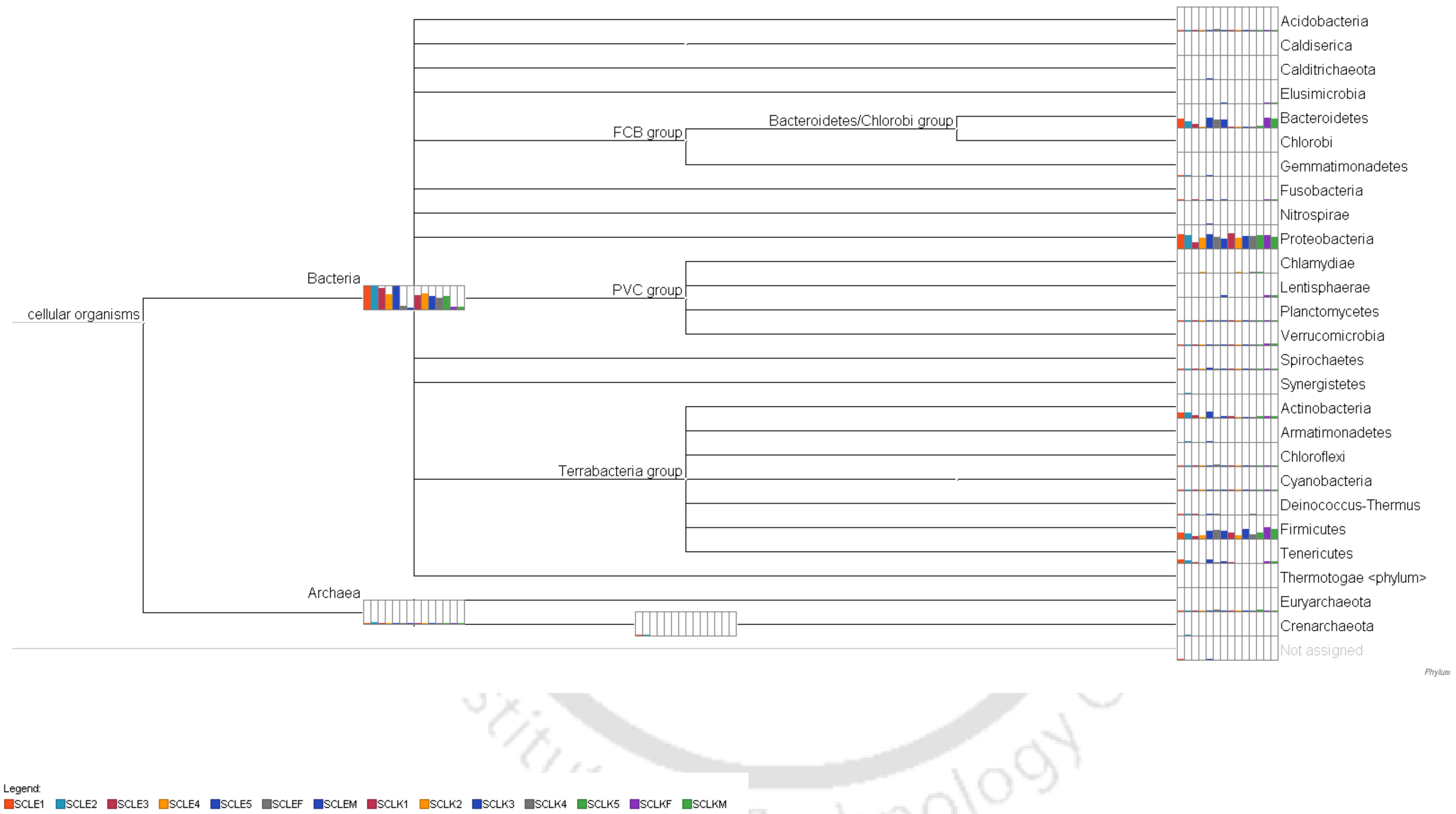


Figure2.8: Cladogram representing evolutionary relationship between the taxa present in the samples.

Differential abundances

Differential abundance analyses were done considering single variables (diet and life-stage) using DeSeq2 (**Figures2.9 and 2.10**). ANCOM-BC was used to analyze based on multiple variables considered together **(Figure2.11)**.

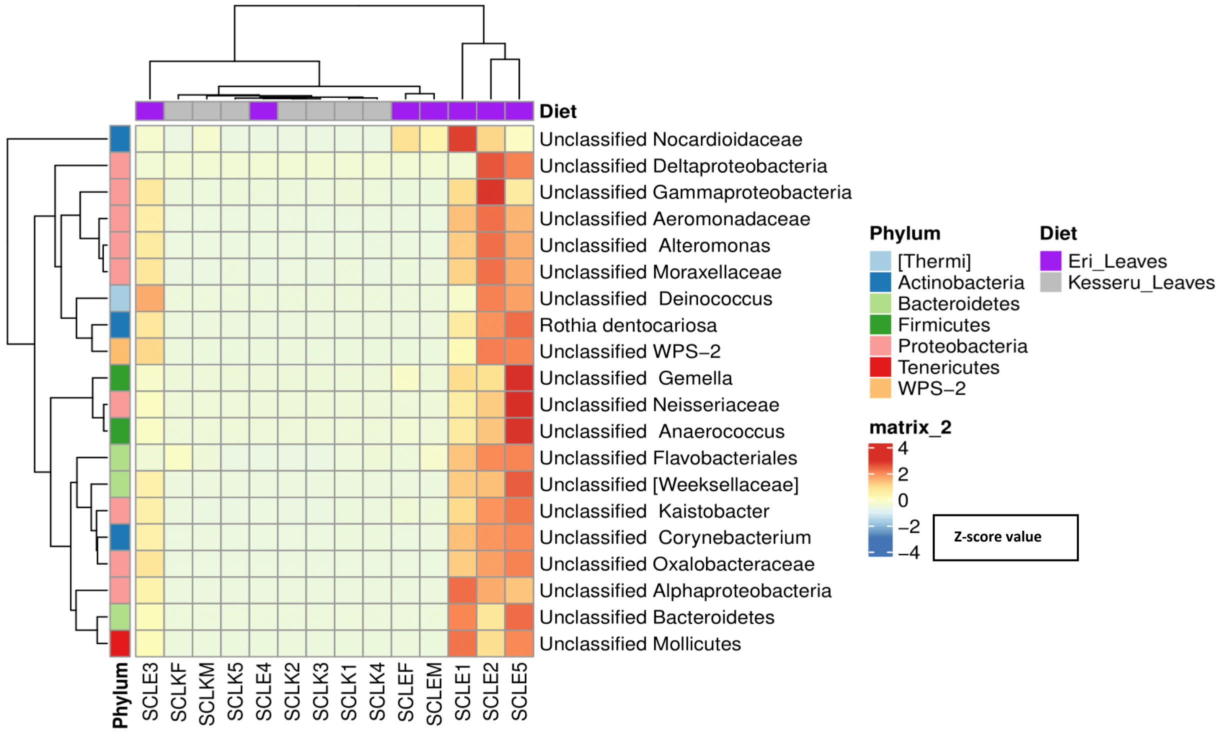


Figure2.9: Microbiome differential abundances based on diet as a variable (Eri-leaves and Kesseru-Leaves). 20 most significant species mapped ($p< 0.05$).

Phylum such as [Thermi], Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Tenericutes and WPS-2 were the most differentially abundant with respect to diet. Eri-fed leaves was seen to have a higher abundance of most species such as Kaistobacter, Mollicutes, Dienococcus, Alteromonas, Aeromonadaceae, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, etc. Eri-fed leaves were also observed to have a more diverse presence of microbiome abundance.

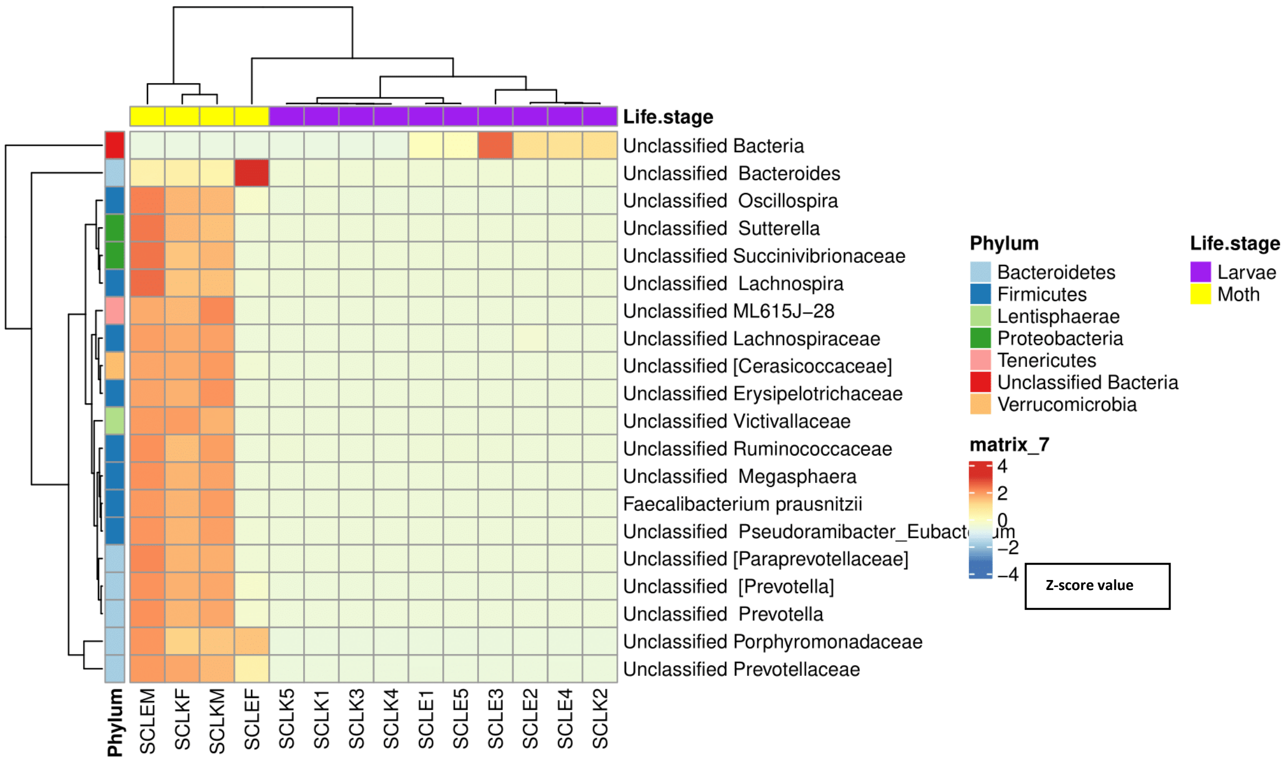


Figure2.10: Microbiome differential abundance based on life-stage as a variable (Larvae and Moth). 20 most significant species mapped ($p < 0.05$).

Highly distinct pattern for life-stage based differential abundance was observed. Larval stages showed a higher abundance of unclassified bacteria especially the Eri-fed larval samples. Bacteroidetes were seen to be more abundant in the adult stages with Eri-fed female moth having very high abundance. Other species such as Oscillospira, Sutterala, Succinivibrionaceae, *Faecebacterium prausnitzii*, several species of unclassified Prevotella and Lachnospira, etc. were highly abundant in adult moths with respect to larvae.

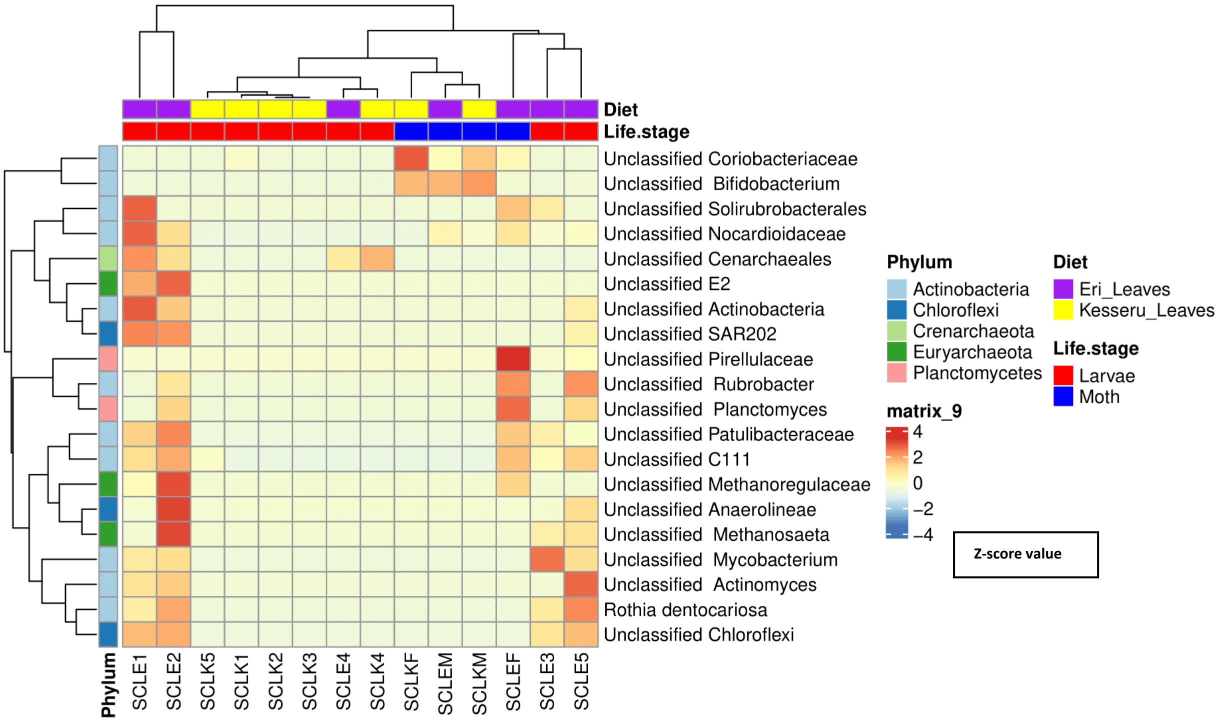


Figure2.11: Microbiome differential abundance based on variables diet and life-stages. 20 most significant species mapped ($p< 0.05$).

Differential analysis taking both variables i.e., diet-types and life-stages into account was analyzed using ANCOM-BC. The most significant phylum differentially abundant with respect to life-stages were Actinobacteria, Chloroflexi, Crenarchaeota, Euryarchaeota, Plancomycetes which consisted of species including *Coriobacteriaceae, Bifidobacterium, Nocardioiceae, Pirellulaceae, Chloroflexi, Methanosaeta, Actinomyces, Mycobacterium*, etc. The adult moth of both the diets had comparatively similar microbiome pattern irrespective of the diet. However, female moth of Eri-fed leaves higher abundances of Pirellulaceae, Rubrobacter, Planctomyces, Patulibacteriaceae, Methanoregulaceae and unclassified C11. Bacterial samples belonging to Coriobacterium and Nocardioiceae were slightly differentially expressed in the adult moths with respect to diet.

A comparative network based on previous observations was constructed using where two nodes are connected if they are differentially associated between the two groups **(Figure2.12)**. NetComi (R) was used applying Pearson's correlations for estimating associations between OTUs and Fisher's z-test differentially correlated OTUs. The data was clr transformed and managed for low FDRs.

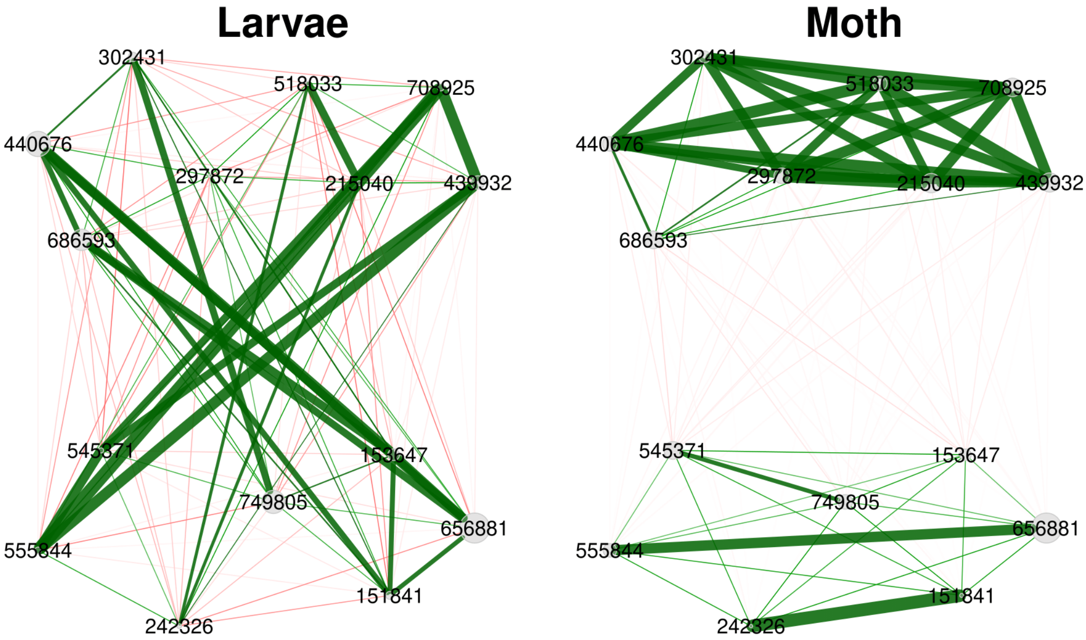


Figure2.12: Differential association networks between sample groups larva and adult moths. Green indicated positive associations while red indicated negative associations. Significant OTUs differentially associated: 302431 (Clostridiales); 518033 (Firmicutes); 708925 (Gallibacterium); 440676 (Bacillaceae); 297872(Lachnospiraceae); 215040 (Dialister); 439932 (Streptococcus); 686593 (Rhizobiaceae); 545371 (Lactobacillus); 749805 (Betaproteobacteria); 153647 (Methanosphaera); 555844 (Bacillaceae); 656881 (Enterobacteriaceae); 242326 (Bacteroidetes); 151841 (Bacteroidales).

A significant differential network was observed among the species. In larval stages positive correlations between species such as Gallibacterium, Bacillaceae, Streptococcus were observed. Higher correlations among species such as

Clostridiales, Firmicutes, Gallibacterium, Lachnospiraceae, Dialister, Streptococcus were observed in moths. Some interactions such as between Bacillaceae and Enterobacteriaceae observed in adult moths were absent in larval samples where interactions between species such as Rhizobiaceae and Enterobacteriaceae were rather present.

Silkworm larvae reared in two different types of diets i.e., Eri and Kesseru leaves were observed to have a difference in their larval growth rate and egg viability rates. Similar observations have been reported by several other studies before [37,73,74].

Although a shared microbiome among all the samples was observed, it was low. Compared to other groups of samples, larval samples fed on Eri leaves showed higher diversity and shared drastically a smaller number of microbes. The number of OTUs including unclassified bacteria were also higher than other sample groups implying its greater potential and more information to be explored.

The bacteria associated with adult moths consisted of those more involved in immunological functions (Erysipelotrichaceae, Lachnospira, Ruminiococcaceae), homeostatis (Ruminiococcaceae), chemoorganotrophy (Succinivibrionaceae, members of Lachnospira), etc[75–78]. Some species found in adult gut were similar to the ones found in mammal vaginal environment (Megasphaera, Prevotella) apart from the gut and were associated with immunological functions[79–81]. Species such as members of Bulkholderia were also reported in high numbers in other adult lepidoptera guts[15].

Rothio spp. present significantly in Eri-fed samples were reported to contribute to degrading and detoxifying gluten implying similar contributions[82]. Nocardioidaceae decomposed carbohydrates and utilizes excessive C and N sources[83]. Mollicutes was also reported to have unique N-associated metabolic capabilities and Kaistobacter with disease suppression[84,85].

Strains of Bifidobacterium and Lactobacilli have been associated with nutrition and categorised as probiobiotics[86]. They have also been artificially incorporated in artificial silkworm feed successfully observing healthier growth[87–89]. In our study, apart from a minimal presence in the 5th instar of Eri-fed larval samples, Bifidobacterium has not been found in any other larval samples. Moth samples however had a significant presence of the species, mostly from Kesseru-fed samples. The presence of the microbe has also been known to be associated with immunity, higher resilience to social stress and secretion of GABA[90]. It has been reported to play a role in socializing in several animal studies[90,91]. Other microbes associated with similar functions were Oscillospira which has also been found in moth samples irrespective of the diet differences[92]. Oscillospira sp. has been commonly found in the gut of many animals but have never been able to be cultured in-vitro[93].

A large number of species was categorised under unclassified bacteria. A very common and economically important strain found in most lepidopterans, *Serratia* has not been identified in our study. It could be absent altogether in the species of *Samia* of our study or could be included in the unidentified strains.

Most lepidoptera studies reported changes in gut microbiota through the larval development. Although a change was observed through the larval growth changes, the pattern was random and insignificant. However, the microbiome abundance and change pattern of the instar stages were similar for both the diet-types. Early larval stages larger feature count decreasing through increasing instars, being lowest at the 4th instar and then drastically increasing to the highest during the 5th instar in both the diet-types. Species richness was also less in the 3rd and 4th instar stages. Adult moths had lower feature counts to that of larval stages but had higher species richness. When compared based on diet-types, Eri-fed larva had much higher number of features compared to Kesseru-fed larva but moths derived from Kesseru-fed had higher number of features. These patterns were not observed in other lepidoptera-microbiome studies. Community difference was Based on diet differences, a broader ordination was observed for organisms fed on Eri leaves compared to those fed on Kesseru.

The compositional difference has been significant in some of the aspects such as species richness but moderately significant in community diversity which could indicate other genetic and environmental factors responsible. It could also be due to methodological factors while sampling or sequencing or other inconsistencies while processing. Our study doesn’t distinguish between living populations (residents) and dormant or dead population (transient) which could also be a factor in the inconsistencies found.

Predictive functional analyses

NMDS PCA diversity plots generated from KO-derived function and Metacyc pathway dissimilarities along the categories of diet and life-stages are summarised through figures (**Figures2.13A and B)**.

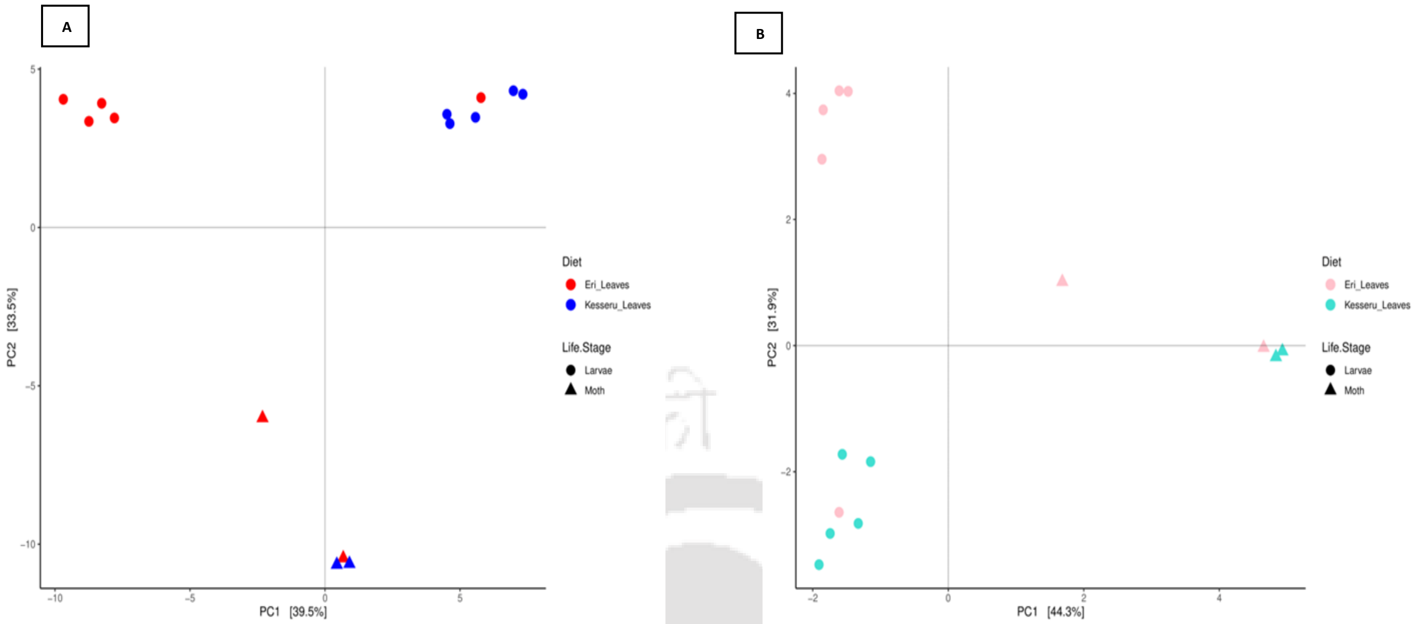


Figure2.13: NMDS plot of functional prediction representing A. KO-derived L3-functions B. Metacyc pathways

Bland- Altman log-ratio abundance plots and Dispersion plots representing differential abundance and dispersion of functions based on diet and life-stages are depicted in **Figures2.14A, B, C, D**.

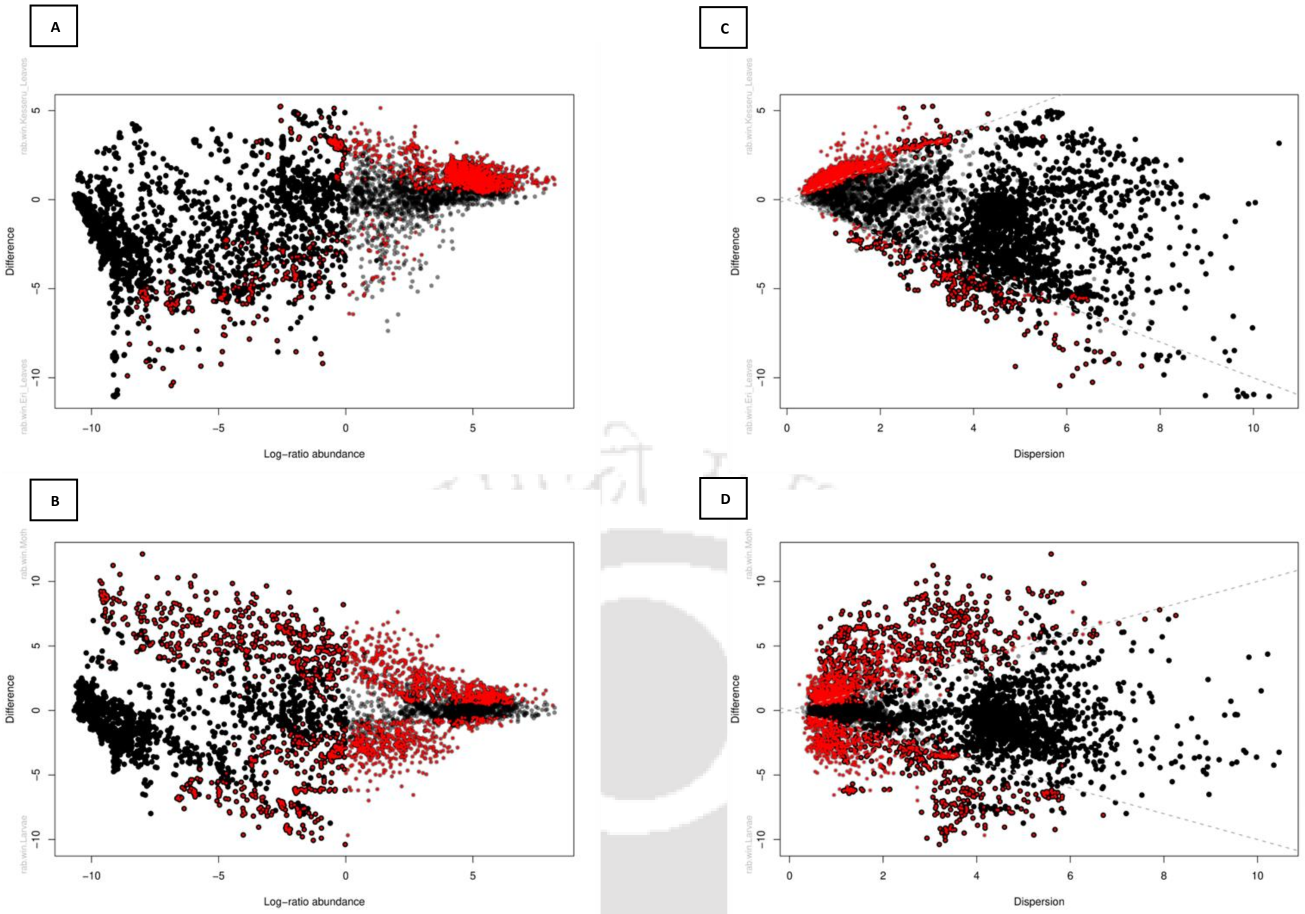


Figure2.14: Bland- Altman log-ratio abundance (MA) plots depicting Kegg-Orthology differential abundance between sample groups A. Diet-types (Eri-fed vs Kesseru-fed) B. Life-stages (Larva vs Adult); Effect (MW) plots depicting dispersion of differential abundance between sample groups C. Diet-types (Eri-fed vs Kesseru-fed) D. Life-stages (Larva vs Adult)

Bar plot showing significant functions (Kegg-derived) and pathways (Metacyc-derived) varying with respect to larval and adult life-stages based on ALDEx2 outputs are represented **(Figures2.15A and B, 16A and B)** ($p<0.05$).

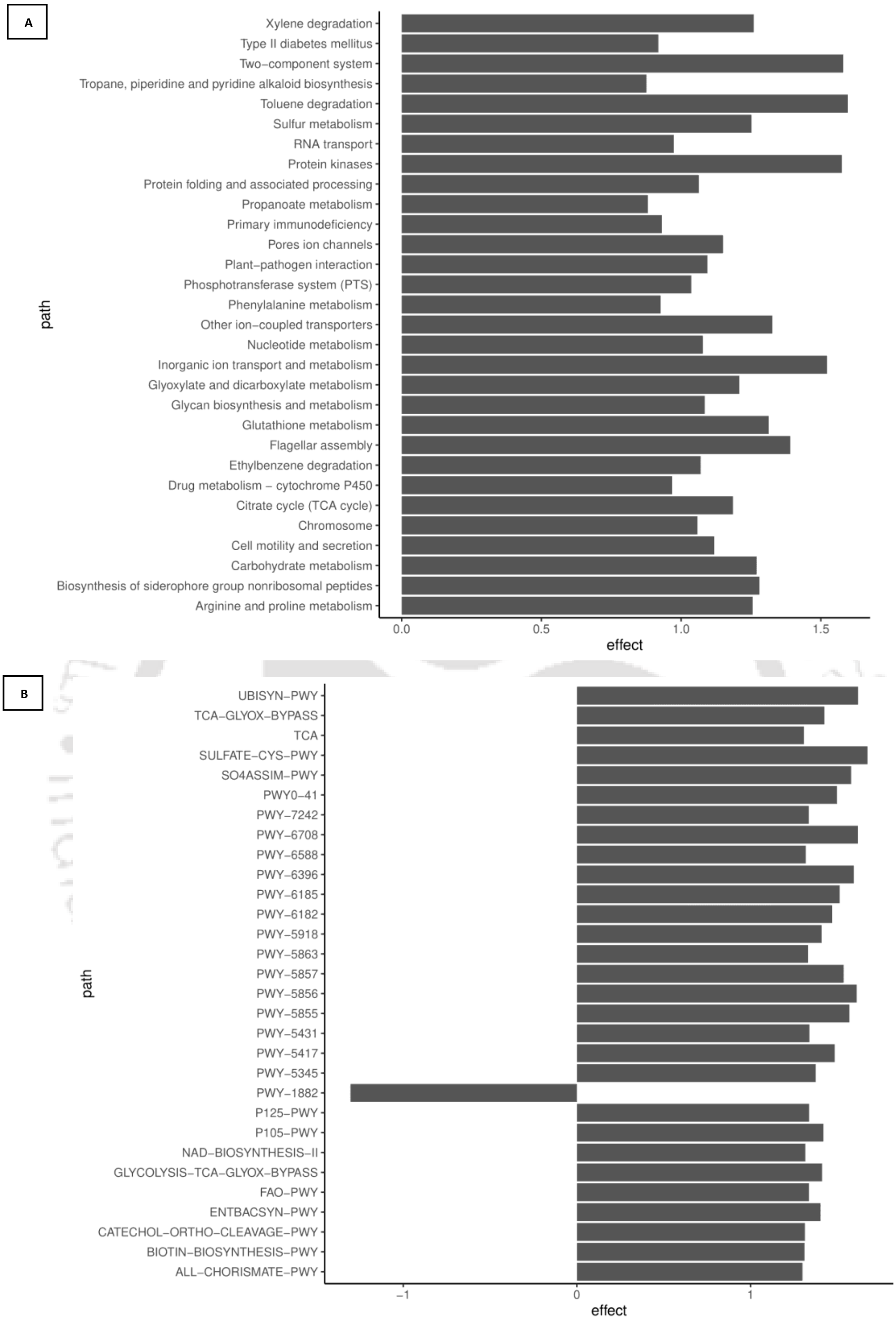
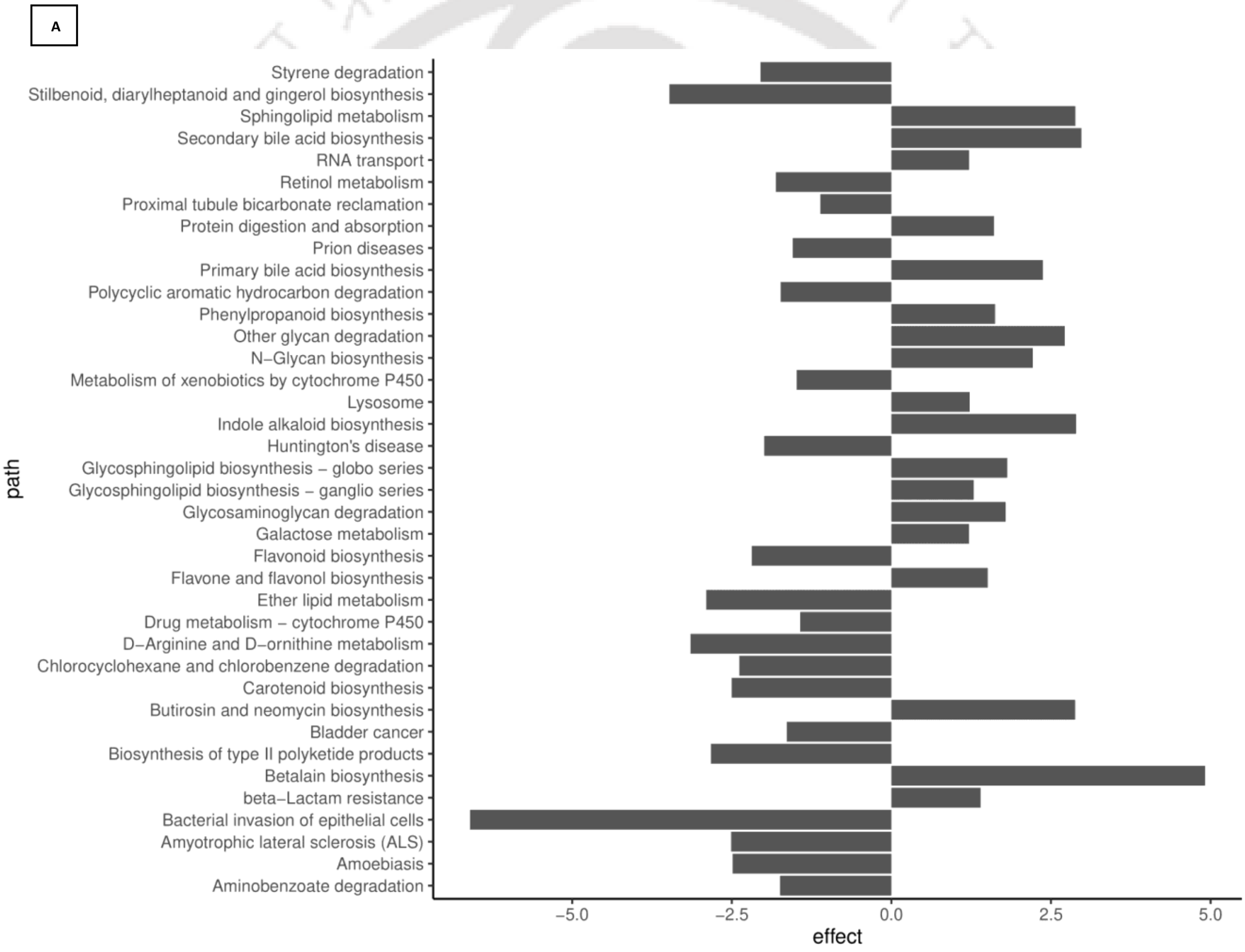


Figure2.15: Most significant A. functions (Kegg-derived) and B. pathways (Metacyc-derived) based on their differential abundance between different diet-types [Effect scores > 0: enriched in Eri_Leaves; < 0: enriched in Kesseru_Leaves]

Significant differences of functions enriched in Eri-fed leaves were not observed compared to samples fed with Kesseru-leaves. Metacyc pathway, PWY-1882 (Superpathway of C1 compounds oxidation to CO_2) enriched in eri-leaves fed samples. Functions including xylene degradation, toluene degradation, plant-pathogen interaction, carbohydrate metabolism, etc. although found in Eri-fed samples were comparatively found to be enriched in Kesseru-fed samples.



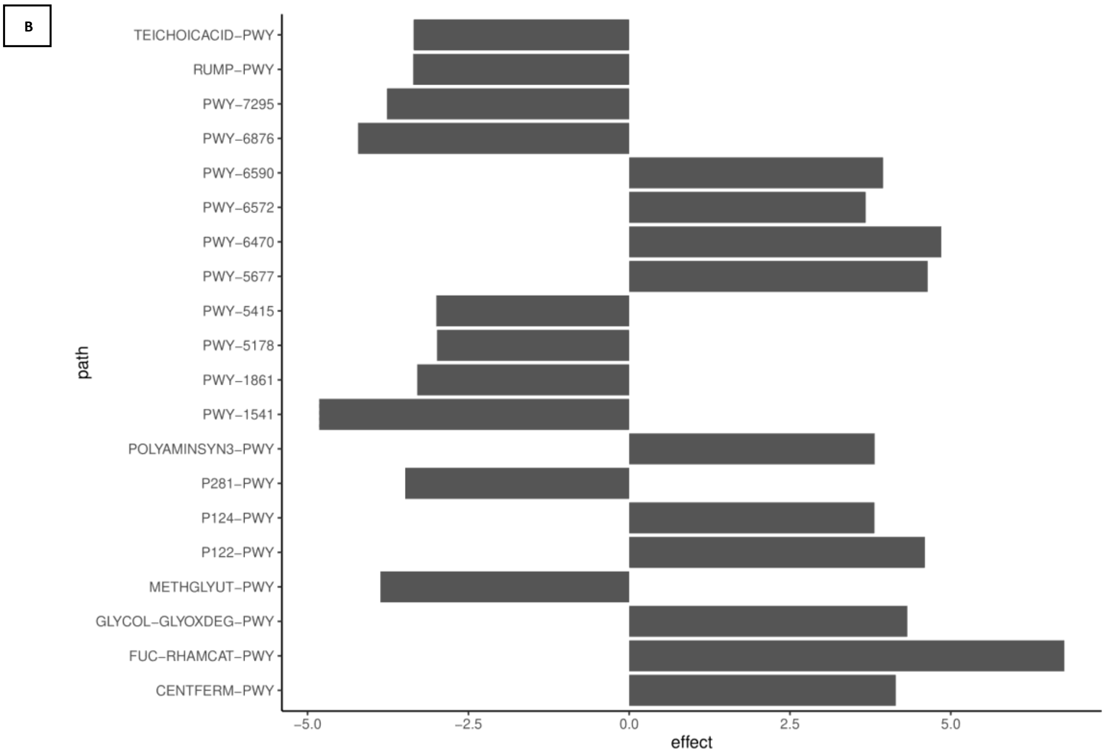


Figure2.16: Most significant A. functions (Kegg-derived) and B. pathways (Metacyc-derived) based on their differential abundance between different life-stages. [Effect scores > 0: enriched in larvae; < 0: enriched in moth]

KO functions such as styrene degradation, hydrocarbon degradation, metabolism of xenobiotics, chlorohexane- chlorobenzene degradation, bacterial invasion of epithelial cells, aminobenzoate degradation were observed to be enriched in larval stages. Metacyc pathways that included teichoic-acid pathway, formaldehyde oxidation/detoxification (PWY), arabinose degradation, catechol degradation, toluene degradation, formaldehyde assimilation, taurine degradation, 3 phenylpropanoate degradation, methylglyoxal degradation was observed to be enriched in the larval stages. On the other hand, moth samples showed increase in function such as sphingolipid biosynthesis, RNA transport, bile biosynthesis, protein absorption, Phenylpropanoid biosynthesis, Lysosome, Glycosphingolipid biosynthesis, Galactose metabolism, Flavone and flavanol biosynthesis, antibacterial and antifungal compounds (butirosin, neomycin and polyketide products) biosynthesis, β-lactam resistance, betalain biosynthesis, etc. Metacyc pathways showed enrichment of pathways such as sulphate degradation, acidogenesis, several fermentation pathways, Bifidobacterium shunt, Polyamine biosynthesis, etc.

KO-derived functions that were significantly different in the diet types were mostly to be the ones enriched in organisms fed in Kesseru leaves. In Eri leaves fed samples, the diversity of microbiome present and predicted functional pathways through the growth stages were larger. No significant pathways or functions except PWY-1882 was found to be enriched in Eri-fed samples, which could be due to the high diversity of microbes in Eri-fed samples. Functions including xylene degradation, toluene degradation, carbohydrate metabolism were comparatively found to be higher in Kesseru-leaves fed samples.

Functional prediction between larval and moth stages have given distinct and significant differences in various functions and pathways. Several KO-derived functions contributing to lipid biosynthesis and metabolism were found to be enriched in adult moth stages. Several biosynthesis pathways were observed in moth stages. On the contrary several enzymes contributing to degradation pathways including plant lignin, aromatic compounds and toxins were found in the larval stages. Pathways contributing to maintaining the pH gradient and electron transport were also found to be comparatively abundant in the larval stages.

CONCLUSION

Understanding an important species representing a large group of insects is of great significance. A comprehensive insight into the microbial composition of *S. cynthia ricini* has been reported. Detailed patterns based on the various factors influencing it have been studied. A core microbiome of 11 species comprising of beneficial microbes belonging to phyla Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria.

Low to moderate microbial diversity through each larval stages (p=0.167), diet types (p=0.13) were observed while a very high diversity (p=0.001) through larva to moth transitions were observed suggesting complete restructuring. Compared to other groups of samples, larval samples fed on Eri leaves showed higher diversity. The bacteria associated with adult moths consisted of those involved in immunological functions and also play a role in socializing while the bacteria associated with larval stages are involved in digestive and detoxifying functions. Further, differences in larval gut microbiome with respect to diet while similar microbiome pattern of moths irrespective of the diet were observed.

These findings could be used to various advantages for the commercially important silkworm species including for better growth and development of the species and enhanced sericulture practices. Diseased species could be understood and treated better as well. Novel microbes with potential could be further explored.

Currently, there are minimum studies of other lepidopteran studies have been reported. Hence, the current study would contribute to the research gaps in insect biology and evolution. Similar studies with respect to other insects would contribute to reducing the research gaps further.

REFERENCES

1. Gilbert, S. F., Sapp, J. & Tauber, A. I. A Symbiotic View Of Life: We Have Never Been Individuals. *Q Rev Biol* **87**, 325–341 (2012).
2. Groussin, M., Mazel, F. & Alm, E. J. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host & Microbe* **28**, 12–22 (2020).
3. Gupta, A. & Nair, S. Dynamics of Insect–Microbiome Interaction Influence Host and Microbial Symbiont. *Front. in Microbiol.* **11**, 1357 (2020).
4. Kikuchi, Y. *et al.* Host-Symbiont Co-Speciation And Reductive Genome Evolution In Gut Symbiotic Bacteria Of Acanthosomatid Stinkbugs. *BMC Biology* **7**, 1–22 (2009).
5. Wilbur, H. M. Complex Life Cycles. *https://doi.org/10.1146/annurev.es.11.110180.000435* **11**, 67–93 (2003).
6. Rolff, J., Johnston, P. R. & Reynolds, S. Complete metamorphosis of insects. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, (2019).
7. Zhang, X., Zhang, F. & Lu, X. Diversity and Functional Roles of the Gut Microbiota in Lepidopteran Insects. *Microorganisms* **10**, 1234 (2022).
8. Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M. & Fatouros, N. E. Bacterial Symbionts In Lepidoptera: Their Diversity, Transmission, And Impact On The Host. *Front. in Microbiol.* **9**, (2018).
9. Engel, P. & Moran, N. A. The Gut Microbiota Of Insects – Diversity In Structure And Function. *FEMS Microbiology Reviews* **37**, 699–735 (2013).
10. Hasan, N. & Yang, H. Factors Affecting The Composition Of The Gut Microbiota, And Its Modulation. *PeerJ* **7**, (2019).
11. Douglas, A. E. Multiorganismal Insects: Diversity and Function of Resident Microorganisms. *Annu Rev Entomol* **60**, 17 (2015).

12. Mason, C. J. *et al.* Diet Influences Proliferation And Stability Of Gut Bacterial Populations In Herbivorous Lepidopteran Larvae. *PLOS ONE* **15**, e0229848 (2020).

13. Gomes, S. I. F. *et al.* Microbiomes Of A Specialist Caterpillar Are Consistent Across Different Habitats But Also Resemble The Local Soil Microbial Communities. *Animal Microbiome 2020 2:1* **2**, 1–12 (2020).

14. Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P. & Fierer, N. Caterpillars Lack A Resident Gut Microbiome. *Proc Natl Acad Sci U S A* **114**, 9641–9646 (2017).

15. Chen, B. *et al.* Biodiversity and Activity of the Gut Microbiota across the Life History of the Insect Herbivore *Spodoptera littoralis*. *Sci. Rep. 2016 6:1* **6**, 1–14 (2016).

16. Jing, T. Z., Qi, F. H. & Wang, Z. Y. Most Dominant Roles Of Insect Gut Bacteria: Digestion, Detoxification, Or Essential Nutrient Provision? *Microbiome 2020 8:1* **8**, 1–20 (2020).

17. Ma, Q. *et al.* Gut Bacterial Communities of *Lymantria xylina* and Their Associations with Host Development and Diet. *Microorganisms* **9**, (2021).

18. Pandiarajan, J. & Krishnan, M. Comparative Bacterial Survey In The Gut Of Lepidopteran Insects With Different Bionetwork. *Microbiology 2018 87:1* **87**, 103–115 (2018).

19. Brinkmann, N., Martens, R. & Tebbe, C. C. Origin And Diversity Of Metabolically Active Gut Bacteria From Laboratory-Bred Larvae Of *Manduca sexta* (Sphingidae, Lepidoptera, Insecta). *Appl. and Environ. Microbiol.* **74**, 7189–7196 (2008).

20. Engel, P. & Moran, N. A. Functional And Evolutionary Insights Into The Simple Yet Specific Gut Microbiota Of The Honey Bee From Metagenomic Analysis. *Gut Microbes* **4**, 60 (2013).

21. Henry, L. P., Bruijning, M., Forsberg, S. K. G. & Ayroles, J. F. The Microbiome Extends Host Evolutionary Potential. *Nat. Comm. 2021 12:1* **12**, 1–13 (2021).

22. Rowland, I. *et al.* Gut Microbiota Functions: Metabolism Of Nutrients And Other Food Components. *Eur. J. of Nutri.* **57**, 1 (2018).

23. Lei, W. te, Huang, K. Y., Jhong, J. H., Chen, C. H. & Weng, S. L. Metagenomic Analysis Of The Gut Microbiome Composition Associated With Vitamin D Supplementation In Taiwanese Infants. *Sci. Rep. 2021 11:1* **11**, 1–15 (2021).

24. Ren, Z. *et al.* Gut Microbiome Analysis As A Tool Towards Targeted Non-Invasive Biomarkers For Early Hepatocellular Carcinoma. *Gut* **68**, 1014–1023 (2019).

25. Veziant, J., Villéger, R., Barnich, N. & Bonnet, M. Gut Microbiota as Potential Biomarker and/or Therapeutic Target to Improve the Management of Cancer: Focus on Colibactin-Producing Escherichia coli in Colorectal Cancer. *Cancers (Basel)* **13**, (2021).

26. Allaband, C. *et al.* Microbiome 101: Studying, Analyzing, and Interpreting Gut Microbiome Data for Clinicians. *Clin Gastroenterol Hepatol* **17**, 218 (2019).

27. Hammer, T. J. & Bowers, M. D. Gut Microbes May Facilitate Insect Herbivory Of Chemically Defended Plants. *Oecologia 2015 179:1* **179**, 1–14 (2015).

28. Show, B. K. *et al.* Insect Gut Bacteria: A Promising Tool For Enhanced Biogas Production. *Rev. in Environ. Sci. and Biotechnol.* **21**, 1–25 (2022).

29. Francoeur, C. B. *et al.* Bacteria Contribute To Plant Secondary Compound Degradation In A Generalist Herbivore System. *mBio* **11**, 1–18 (2020).

30. Kohl, K. D. & Denise Dearing, M. The Woodrat Gut Microbiota As An Experimental System For Understanding Microbial Metabolism Of Dietary Toxins. *Front. in Microbiol.* **7**, 1165 (2016).

31. Krishnan, M. *et al.* Insect gut microbiome - An Unexploited Reserve For Biotechnological Application. *Asian Pacific J. Trop. Biomed.* **4**, S16 (2014).

32. Shifa, K., E. G. & Sori, W. Rearing Performance of Eri-Silkworm (*Samia cynthia ricini* Boisduval) (Lepidoptera: Saturniidae) Fed with Different Castor (*Ricinus communis L.*) Genotypes. *J. Entomol.* **11**, 25–33 (2013).

33. Renuka G & Shamitha G. Studies on The Economic Traits of Eri Silkworm, *Samia cynthia ricini*, In Relation to Seasonal Variations. *Int J. of Adv. Res.* (2014).

34. Sharma, P. & Chandra Kalita, J. A Comparative Study on Six Strains of Eri Silk Worm (Samia Ricini; Donovan) Based on Morphological Traits. **2**, 506–511 (2013).

35. Ahmed, S. A., Singh, N. I. & Sarkar, C. R. Role of Forest Biodiversity In Conservation of Non-Mulberry (Vanya) Silk In India. _*Mun. Ent. Zool* **10**, (2015).

36. Velayudhan, K., Balachandran, N., RadhaKrishnan, S., Singh, B. K. & and Jayaprakash, P. Biodiversity in Eri Silkworm *Samia Ricini (Donovan*) Genetic Resources and its Conservation. *J. Aq. Biol. Fish.* **2**, 817–824 (2014).

37. Host Plants of Eri Silkworm | Handloom Textiles & Sericulture | Government of Assam, India. https://hts.assam.gov.in/portlet-innerpage/host-plants-of-eri-silkworm.

38. Improved Varieties of Silkworm Food Plants | Churchandpur, Manipur. https://silks.csb.gov.in/churchandpur/improved-varieties-of-silkworm-food-plants/.

39. Zhang, N. *et al.* Contribution Of Sample Processing To Gut Microbiome Analysis In The Model Lepidoptera, Silkworm *Bombyx mori*. *CSBJ* **19**, 4658–4668 (2021).

40. Keen, J. N., Austin, M. K., Huang, L. S., Messing, S. & Wyatt, J. D. Efficacy of Soaking in 70% Isopropyl Alcohol on Aerobic Bacterial Decontamination of Surgical Instruments and Gloves for Serial Mouse Laparotomies. *J Am Assoc Lab Anim Sci* **49**, 832 (2010).

41. Gerasimidis, K. *et al.* The Effect Of DNA Extraction Methodology On Gut Microbiota Research Applications. *BMC Research Notes* **9**, (2016).

42. García-Alegría, A. M. *et al.* Quantification of DNA through the nanodrop spectrophotometer: Methodological validation using standard reference material and sprague dawley rat and human DNA. *Int. J. Anal. Chem.* **2020**, (2020).

43. Olson, N. D. & Morrow, J. B. DNA Extract Characterization Process For Microbial Detection Methods Development And Validation. *BMC Research Notes* **5**, 668 (2012).

44. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

45. Bolyen, E. *et al.* Reproducible, Interactive, Scalable And Extensible Microbiome Data Science Using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).

46. Callahan, B. J. *et al.* DADA2: High-Resolution Sample Inference From Illumina Amplicon Data. *Nat. Methods 2016 13:7* **13**, 581–583 (2016).

47. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A Versatile Open-Source Tool For Metagenomics. *PeerJ* **2016**, e2584 (2016).

48. McDonald, D. *et al.* An Improved Greengenes Taxonomy With Explicit Ranks For Ecological And Evolutionary Analyses Of Bacteria And Archaea. *The ISME Journal 2012 6:3* **6**, 610–618 (2011).

49. Bokulich, N. A. *et al.* Optimizing Taxonomic Classification Of Marker-Gene Amplicon Sequences With QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 1–17 (2018).

50. Draw Venn Diagram. https://bioinformatics.psb.ugent.be/webtools/Venn/.

51. RStudio | Open Source & Professional Software For Data Science Teams - RStudio. https://www.rstudio.com/.

52. McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* **8**, e61217 (2013).

53. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE* **5**, e9490 (2010).

54. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A Novel Method For Rapid Multiple Sequence Alignment Based On Fast Fourier Transform. *Nuc. Acids Res.* **30**, 3059–3066 (2002).

55. Faith, D. P. Conservation Evaluation And Phylogenetic Diversity. *Biol. Conserv.* **61**, 1–10 (1992).

56. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative And Qualitative Beta Diversity Measures Lead To Different Insights Into Factors That Structure Microbial Communities. *Appl Environ Microbiol* **73**, 1576–1585 (2007).

57. Konopiński, M. K. Shannon Diversity Index: A Call To Replace The Original Shannon's Formula With Unbiased Estimator In The Population Genetics Studies. *PeerJ* **2020**, (2020).

58. Lozupone, C. & Knight, R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* **71**, 8228 (2005).

59. Davenport, J. M. New Approximations To The Exact Distribution Of The Kruskal-Wallis Test Statistic. *Comms. in Stat. - Theory and Methods* **5**, 1335–1348 (1976).

60. Kelly, B. J. *et al.* Power And Sample-Size Estimation For Microbiome Studies Using Pairwise Distances and PERMANOVA. *Bioinformatics* **31**, 2461–2468 (2015).

61. Love, M. I., Huber, W. & Anders, S. Moderated Estimation Of Fold Change And Dispersion for RNA-seq Data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

62. Lin, H. & Peddada, S. das. Analysis of Compositions Of Microbiomes With Bias Correction. *Nat. Comms. 2020 11:1* **11**, 1–11 (2020).

63. Gu, Z., Eils, R. & Schlesner, M. Complex Heatmaps Reveal Patterns And Correlations In Multidimensional Genomic Data. *Bioinformatics* **32**, 2847–2849 (2016).

64. Huson, D. H. *et al.* MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology* **12**, 1004957 (2016).

65. Peschel, S., Müller, C. L., von Mutius, E., Boulesteix, A. L. & Depner, M. NetCoMi: Network Construction And Comparison For Microbiome Data in R. *Brief. in Bioinfo.* **22**, 1–18 (2021).

66. Douglas, G. M. *et al.* PICRUSt2 For Prediction Of Metagenome Functions. *Nat. Biotech. 2020 38:6* **38**, 685–688 (2020).

67. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS ONE* **8**, (2013).

68. WELCH, B. L. The Significance of The Difference Between Two Means When the Population Variances are Unequal. *Biometrika* **29**, 350–362 (1938).

69. Altman, D. G. & Bland, J. M. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician* **32**, 307 (1983).

70. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: integrating viruses and cellular organisms. *Nuc. Acids Res.* **49**, D545–D551 (2021).

71. Caspi, R. *et al.* The MetaCyc Database Of Metabolic Pathways And Enzymes - A 2019 Update. *Nucleic Acids Research* **48**, D445–D453 (2020).

72. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. Royal Stat. Soc.: Series B (Methodological)* **57**, 289–300 (1995).

73. (PDF) Effect of Differential Feeding on Eri Silk Worm (*Samia Ricini Donovan*) and its Pupal Protein Concentration. https://www.researchgate.net/publication/319178276_Effect_of_Differential_Feeding_on_Eri_Silk_Worm_Samia_Ricini_Donovan_and_its_Pupal_Protein_Concentration.

74. Baruah, M. & Bari, P. Studies on Larval Weight and Shell Ratio of Eri Silkworm (Philosamia ricini) on Castor, Kesseru and Treated Kesseru by Foliar Spray. (2012).

75. Kaakoush, N. O. Insights into the Role of Erysipelotrichaceae in the Human Host. *Front. Cell. Inf. Microbiol.* **5**, 84 (2015).

76. Canani, R. B. *et al.* Specific Signatures Of The Gut Microbiota And Increased Levels Of Butyrate In Children Treated With Fermented Cow's Milk Containing Heat-Killed *Lactobacillus paracasei* CBA L74. *App. Env. Microbiol.* **83**, (2017).

77. Jiang, W. *et al.* Dysbiosis Gut Microbiota Associated With Inflammation And Impaired Mucosal Immune Function In Intestine Of Humans With Non-Alcoholic Fatty Liver Disease. *Sci. Rep. 2015 5:1* **5**, 1–7 (2015).

78. Ng, S. H., Stat, M., Bunce, M. & Simmons, L. W. The Influence Of Diet And Environment On The Gut Microbial Community Of Field Crickets. *Ecol. and Evol.* **8**, 4704 (2018).

79. van Teijlingen, N. H. *et al.* Vaginal Dysbiosis Associated-Bacteria *Megasphaera elsdenii* And *Prevotella timonensis* Induce Immune Activation Via Dendritic Cells. *J Reprod Immunol* **138**, (2020).

80. Iljazovic, A. *et al.* Perturbation Of The Gut Microbiome By *Prevotella spp*. Enhances Host Susceptibility To Mucosal Inflammation. *Muc. Immun. 2020 14:1* **14**, 113–124 (2020).

81. Tainchum, K. *et al.* Bacterial Microbiome in Wild-Caught Anopheles Mosquitoes in Western Thailand. *Front. in Micro.* **11**, 965 (2020).

82. Zamakhchari, M. *et al.* Identification of *Rothia* Bacteria as Gluten-Degrading Natural Colonizers of the Upper Gastro-Intestinal Tract. *PLoS ONE* **6**, (2011).

83. Zhang, M. *et al.* Variations Of Earthworm Gut Bacterial Community Composition And Metabolic Functions In Coastal Upland Soil Along A 700-Year Reclamation Chronosequence. *Sci. Total Env.* **804**, 149994 (2022).

84. Sapountzis, P., Zhukova, M., Shik, J. Z., Schiott, M. & Boomsma, J. J. Reconstructing The Functions Of Endosymbiotic Mollicutes In Fungus-Growing Ants. *Elife* **7**, (2018).

85. Liu, X. *et al.* Using Community Analysis To Explore Bacterial Indicators For Disease Suppression Of Tobacco Bacterial Wilt. *Sci. Rep. 2016 6:1* **6**, 1–11 (2016).

86. Fijan, S. Microorganisms with Claimed Probiotic Properties: An Overview of Recent Literature. *Int. J. Environ. Res.* **11**, 4745 (2014).

87. Zuko, Y. & Maeda, K. The Effect Of Intestinal Bifidobacterium On The Output Of *Bombyx mori*. *Peer J Preprints* (2018).

88. Soliman, S. Micro-Organisms Supplementation to Mulberry Silkworm, Bombyx mori L.. . *Egypt. Acad. J. Biol. Sci.* (2017).

89. Nishida, S., Ono, Y. & Sekimizu, K. Lactic Acid Bacteria Activating Innate Immunity Improve Survival In Bacterial Infection Model Of Silkworm. *Drug Disc. & Therap.* **10**, 49–56 (2016).

90. Yang, C. *et al.* Bifidobacterium In The Gut Microbiota Confer Resilience To Chronic Social Defeat Stress In Mice. *Sci. Rep.* **7**, (2017).

91. Socała, K. *et al.* The Role Of Microbiota-Gut-Brain Axis In Neuropsychiatric And Neurological Disorders. *Pharm. Res.* **172**, 105840 (2021).

92. Maltz, R. M. *et al.* Social Stress Affects Colonic Inflammation, The Gut Microbiome, And Short Chain Fatty Acid Levels And Receptors. *J Pediatr Gastroenterol Nutr* **68**, 533 (2019).

93. Dance, A. The Search For Microbial Dark Matter. *Nature* **582**, 301–303 (2020).

Chapter 3

Shotgun metagenome analysis of *Samia cynthia ricini* 5th instar larva

Chapter 3

Shotgun metagenome analysis of *Samia cynthia ricini* 5th instar larva

ABSTRACT

In this study, we employed shotgun sequencing-based metagenome analysis to explore the taxonomic and functional composition of the microbial community associated with the 5th instar larva of *Samia cynthia ricini*. By comparing the results with traditional amplicon microbiome analyses, we aimed to highlight the advantages of shotgun sequencing in providing a more comprehensive understanding of the microbial genes present.

For taxonomic composition analysis, we identified the dominant microbial groups in the larval sample, with Proteobacteria, Firmicutes, Cyanobacteria, and Bacteroidetes being the most prevalent. Interestingly, we observed differences in abundance between the taxonomic profiles derived from 16s rRNA gene sequence reads obtained through amplicon-based sequencing, 16s rRNA gene sequence reads acquired from shotgun sequencing, and total reads derived from shotgun-based metagenomics, excluding 16s rRNA gene sequences. Particularly, the taxonomic profile based on total shotgun reads revealed a higher diversity of microbial species across various domains, emphasizing the benefits of shotgun metagenomics in capturing a more comprehensive microbial community.

In the functional profile analysis, we annotated the microbial genes and reconstructed biological pathways using MinPath to gain insights into the microbial functions present in the larval environment. Our results highlighted the prevalence of pathways involved

in amino acid metabolism, lipid metabolism, and the metabolism of vitamins and cofactors. Additionally, a high abundance of secondary metabolism genes was observed, suggesting the potential for diverse biosynthetic activities within the microbial community.

Furthermore, we identified various genes associated with xenobiotics metabolism, drug metabolism, and cellular mechanisms, including defense mechanisms, 2-component regulatory systems, cytoskeleton-peroxisome interactions, post-translational modifications, chaperones, and signal transduction mechanisms. These findings shed light on the potential functional roles of the microbial community in the larval environment, providing valuable insights into the interactions between the host and its associated microbiome.

INTRODUCTION

The emergence and gradual development of applications of next-generation sequencing technologies is very rapid and along with it more in-depth methodologies for metagenomics and related studies derived[1–3].

Shotgun metagenomics is a powerful tool for microbiome analysis in insect guts. In contrast to 16s rRNA gene sequencing-based microbiome identification, it allows a more comprehensive analysis of microbial communities[4]. Detailed view of microbial populations, their taxonomic composition and abundances and the functional potential along with the gene abundances are obtained helping us understand the interaction between the microbes and their role in the surrounding environment's function. E.g., it can be used to identify potential antibiotic-resistance genes within a microbial community, allowing for a more accurate prediction of the susceptibility of an insect gut to certain medications[5–7].

16S rRNA gene sequencing is a targeted approach that amplifies and sequences a conserved region of the 16S rRNA gene, which is present in bacterial and archaeal genomes[8,9]. In contrast, shotgun metagenome sequencing is a non-targeted approach that sequences all the genes present in a sample, providing a much more comprehensive view of the gut microbiome[5,10].

Bias generation due to the use of single phylogenetic marker such as the 16S ribosomal gene or its variable regions is observed using amplicon sequence-based microbiome studies[11]. Many studies reported other drawbacks of amplicon sequencing including 16s rRNA gene sequence including species level low resolution because of variation in copy number and the fact that they compose of only <0.1% of the total genome and the higher chances of horizontal transfer of the gene. Low and insignificant amplification of marker genes were observed in many studies[12–15].

Compared to 16S rRNA gene sequencing, shotgun metagenomics has several advantages. It provides a much more detailed view of microbial communities in an environment, allowing for a better understanding of the functional roles that different microbes play. It can also be used to identify genetic markers associated with antibiotic-resistance, which can be used to guide treatment strategies[7,16,17]. Further, microbial community identification via 16s rRNA reads could also be extracted from shotgun sequencing based reads[18,19].

Despite its advantages, shotgun metagenomics is not without its drawbacks. It is much more expensive and time-consuming than 16S rRNA gene sequencing, and the data produced is much more complex, making it more difficult to interpret. Additionally, shotgun metagenomics requires more sophisticated laboratory equipment, which may not be available in many laboratories[10,20].

Many studies compared the performance of the two approaches in terms of their ability to detect microbial diversity, taxonomic resolution, and functional capabilities[17,18,20]. The authors found that 16S rRNA gene sequencing has a higher detection sensitivity than shotgun metagenome sequencing, with the latter having greater resolution and providing more detailed information about microbial populations[20]. 16S rRNA gene sequencing was found to have superior performance for detecting low-abundance microbial species[14,21]. 16S rRNA gene sequencing and shotgun metagenome sequencing are complementary approaches for studying the gut microbiome. The two approaches can be used in combination to provide a more comprehensive assessment of the gut microbiome[22,23].

The first comprehensive study of the silkworm gut microbiome was conducted by Chen et al., who used high-throughput sequencing to analyze the composition of microbial communities from adult and larval stages of *B. mori*[24]. This study revealed that the gut microbiota of silkworm larvae was composed of Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes phyla, and that the microbial communities of adult and larval stages differed significantly in terms of their species composition[24]. Subsequent studies have further characterized the silkworm gut microbiome, finding a core microbiome of 11 genera and a wide diversity of bacteria, including several novel species[24]. In addition to investigating the composition of the silkworm gut microbiome, a number of studies have sought to characterize its functional capabilities by investigating the metatranscriptome and metaproteome of the silkworm. Studies have identified several important functional roles of the silkworm gut microbiome, including the production of enzymes for digestion, protection from pathogens and toxins, synthesis of vitamins and other essential molecules, and regulation of host development[25,26]. These functional capabilities of the silkworm gut microbiome have potential applications to silk production, and may be exploited for the improvement of silkworm rearing and silk production[27]. Overall, recent advances in silkworm metagenomics have shed light on the composition, dynamics and functional roles of the silkworm gut microbiome[28–30]. Further research is needed to explore the potential applications of gut microbial communities to silk production and to identify novel microbial species that could be exploited for the improvement of rearing conditions.

Significances of the study include-

- Determining the optimum conditions and tools for *S. cynthia ricini* metagenome annotation and assembly.
- Taxonomic and functional annotation of the sequenced metagenome.
- Reconstruction of metabolic pathways based on functional annotations.
- Comparison of abundances for 16s rRNA gene sequence microbiome taxonomic abundances and functional predictions.

METHODOLOGY

Sample collection and DNA extraction

Eri silkworm (*S. cynthia ricini*) diseased-free layings (dfls) were acquired from Central Muga & Eri Research Institute, Jorhat, Assam, India (Lat: 26° 47′49.1″N Lon: 94° 19′35.0″E). They were reared in a domesticated semi-controlled environment with a room temperature of 25˚C and 80% relative humidity. Gut tissue samples pooled from three 5th instar larva reared on *Ricinus communis Linn.* (Eri) following surface sterilization of the sample in 70% alcohol. The samples were stored in phosphate buffer solution at −80 °C with (storing for DNAase) for further processing.

DNA extraction was carried out using phenol/chloroform (PHEC) extraction methodology[31]. Cells were homogenized using using a Vibrio Sonics^{TM} for 5m in sterile phosphate-buffer solution (PBS) followed by SDS-based cell-membrane degradation. The extracts were then saturated with phenol followed by separation of nucleic acids and proteins using chloroform: isoamyl-alcohol (24:1). Overnight precipitation of nucleic acids using isopropanol and 3 M sodium chloride for maximum DNA recovery was carried out. DNA samples were analyzed for their concentration and purity using 1% agarose gel electrophoresis and Nanodrop spectrophotometer. Extracted DNA samples (A260/A280 ~1.8) were sent for shogun sequencing using Illumina NGS MiSeq (Eurofins Technologies Pvt. Ltd.).

Sequence retrieval, quality-control and processing

Paired-end raw sequencing data (250bp) in FASTQ format were analyzed for quality and adapters using FastQC(v_0.11.9)[32]. The reads were then subjected to quality control using Trimmomatic (v 0.32). Adapters, low quality bases towards 3′ end and duplicated reads were trimmed from the raw reads with phred quality score of ≤ 30 using Trimmomatic (v 0.32)[33].

Processing, assembly and annotation

16s rRNA gene sequence reads were sorted and filtered out from the quality-controlled reads. The filtered reads were subjected to assemblers including SPADES v12.0, metaSPAdes, MEGAHIT and metaVELVET. SPADES output was chosen for further analyses based on the assembly quality[34–36]. The assemble contigs were analyzed using MetaQUAST for characteristics including N50, number of contigs, etc[37] (**Table3.1**). The extracted 16s rRNA gene sequence reads from sortmerna were classified using RDP classifier trained on NCBI database[38]. The hierarchical structure of the classification was viewed using MEGAN and KRONA charts[39,40].

Criteria	SPADES	MetaSPAdes	MEGAHIT	metaVELVET
N50	1153	903	928	716
Number of contigs	189501	217655	215422	24053
Number of contigs >1000bp	75489	43802	54526	22526

Table3.1: Comparative summary of assembly statistics of assemblers with SCL_5E shotgun sequences as analyzed by QUAST.

Annotation of the identified protein sequences were carried out through taxonomic, structural and functional categories.

Taxonomic analyses of the assembled CDS were carried out by homology-based assignments using MEGAN+DIAMOND methodology[41]. DIAMOND formatted NCBI nr database was indexed followed by alignment of the assembled reads against it. MEGAN was then used for taxonomic binning using the naïve LCA algorithm. Contaminants were filtered out. The classification was represented in the form of a phylogenetic tree.

Functional annotation was performed using PROKKA (v1.14.6)[42]. PROKKA uses unsupervised machine learning gene prediction algorithm, Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm) for finding gene and translation. The subsequent processes are automated via PROKKA including location of RNA regions and open reading frames (ORFs), translation of ORFs to proteins, homology searching of proteins via BLAST v10.0 and HMMER v3.1) against public databases (NCBI-CDD, PFAM, TIGRFAM) and PROKKA custom databases for annotation.

Proteins/enzymes annotated are involved in various reactions for various cellular and metabolic activities depicted by different pathways. The pathways are predicted by considering the presence of the set of enzymes to be the indicator and thus mapping using a naïve/direct approach. However, this methodology leads to false positive, as many pathways require only few of the present proteins/enzymes. MinPath (Minimal set of Pathways) uses a parsimony approach for biological pathway reconstructions yielding a more conserved and appropriate functional diversity, eliminating excessive predicted pathways[43].

The annotated PROKKA outputs were filtered based on KEGG Ortholog (KO) IDs for a compatible input using in-house shell scripts and mapped through MinPath based on KEGG and MEtaCyC databases.

Antibiotic resistance genes major classes were annotated and identified using ARDB and the relative abundances of the subtypes were summarised.

Mapping reads and quantifying genes

Annotation output provides with number of genes and their annotations of a particular sample from the assembled reads. The information on the quantitative details is then achieved by mapping the raw reads to the assembled reads using Bowtie2 (v 2.5.1)[44]. The output alignment BAM file is then processed with Picard for duplicate reads removal[45].

Following that, coverage of each annotated gene is obtained. This is obtained by creating a custom Gene Transfer File (GTF) with the information of the gene regions and structure using in-house shell script thus defining the regions. Then, using htseq2-count (v 2.0.2) the number of reads mapped to each gene were counted sorted by alignment coordinate[46].

Normalizing to Transcripts Per Million (TPM)

Taking into account that longer reads are mapped more than shorter reads due to their length and hence would result in erroneous abundance values, normalization of the read counts is carried out. TPM (Transcripts Per Million) method of normalization was adopted[47]. The average read length of our sample and length of all genes present are required for TPM calculation, as follows,

$$TPM = \frac{rg \; x \; rl}{flg \; x \; T} . 10^6$$

where, rg: reads mapped to gene g

rl: read length

flg: feature length

T: sum of rgxrl/flg for all genes

The average read length is ~150bp calculated from the raw reads. The gene lengths were extracted from the read count file using in-house shell script. TPM was calculated using a python script.

The obtained filtered KO and COG IDs from PROKKA annotated files and the normalised gene coverage value files were used to summarize and define the annotation. KRONA tools were used for the summarisation and plotting of interactive plots. The data was transformed to KRONA chart compatible hierarchal system using in-house python script. The summarized pathways based on KO, MetaCyc and COG databases annotation files were interactively visualized as KRONA charts and explored accordingly.

Comparative taxonomy and functional analyses

Comparative descriptions between the output of taxonomic analyses done using the 16s sequenced data through QIIME2, 16s rRNA gene sequence extracted reads from shotgun sequenced samples using SortmeRNA and from protein sequences identified using MEGAN+DIAMOND were normalised (copies per million i.e., CPM), summarised and visualised using MEGAN and iVenn[48,49]. CPM is defined by the total read number mapped against each gene normalised by the length of the gene of interest. It is calculated as follows:

$$CPM = \frac{Number\ of\ mapped\ reads\ to\ a\ gene\ of\ interest * 10^6}{Total\ read\ counts}$$

Comparative summaries of the predicted KEGG pathways of the16s rRNA gene sequence derived data and shotgun annotated data were also visualized through kronacharts, iVenn, MS-excel and R. Overview of the metagenome-based identified pathways were visualised using iPATH3[50]. An overview of the methodology is represented beloa (**Figure3.1**).

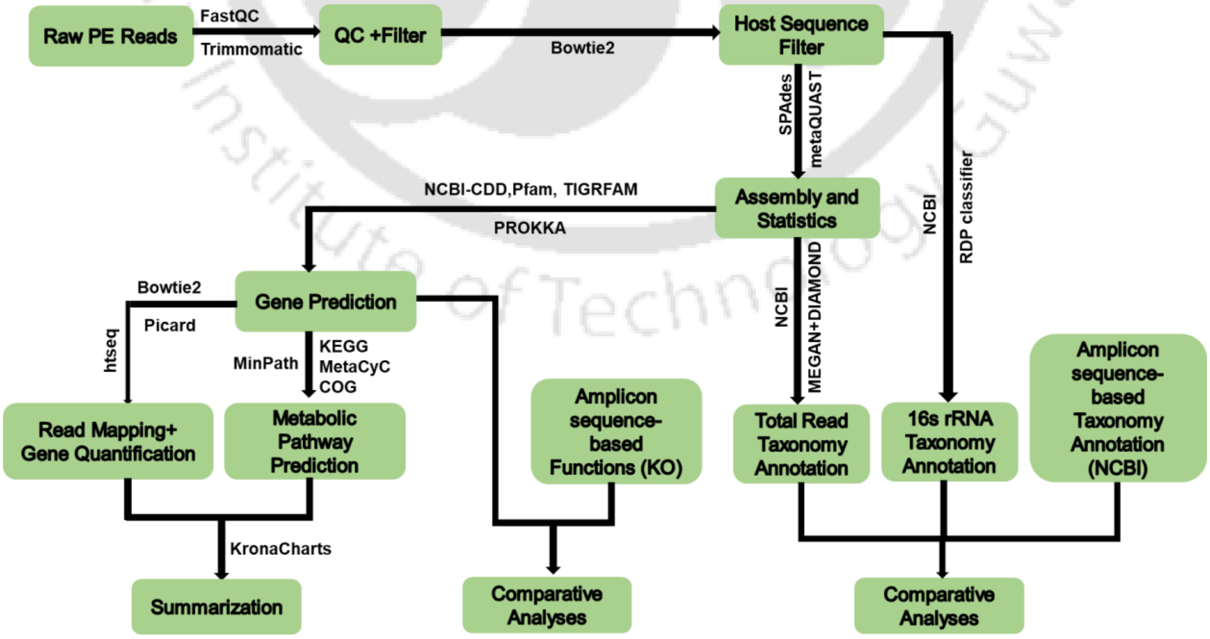


Figure3.1: Schematics of the methodology summarised.

RESULTS AND DISCUSSION

Metagenomic assembly of *S. cynthia ricini*

High-throughput sequencing of the 5th instar *S. cynthia ricini* larva sample, SCL_5E generated 10,517,215 paired-end raw reads. Post-quality control and host sequence contamination removal, 4,233,587 paired-end raw reads were retained. The raw reads were deposited in the NCBI Short Read Archive (SRA) database.

Assembly quality was found to be optimum using SPADES and hence the output was of the analysis from SPADES was chosen for taxonomic and functional annotation. The assembly statistics of the sample using SPADES analyzed via MetaQUAST is summarised below (**Table3.2**)

Description	***S. cynthia ricini***
Total length	202665010
Contigs (>= 10000bp)	7
Contigs (>= 1000bp)	75489
Contigs	189501
Largest contig	19053
4N50	1153
GC (%)	35.91

Table3.2: Annotation statistics of SCL_5E using SPADES.

A total of 21,500 CDS (NCBI and UniProt) reads were predicted through Prodigal.

Annotation statistics is summarised in **Table3.3**.

	Number of reads annotated
CDS	**21500**
KEGG	**24452**
Pfam	**4127**
COG	**1102**
ARDB	**504**

Table3.3: Annotation summary of SCL_5E.

Taxonomy composition

Community analysis based on 16s rRNA gene sequence reads resulted in identification of 79% bacterial population and 21% archaeal population. Proteobacteria was the most abundant phylum comprising of 41% of the bacterial population followed by Firmicutes, Cyanobacteria and Bacteroidetes comprising of 16%, 12% and 11% of the bacterial population respectively (**Figure3.1**). This observation has been found to be similar with several taxonomy composition studies of silkworm *Bombyx mori* and some other insects including *Plutella xylostella, Spodoptera litura, Culex pipens*, etc[27,51–53]. Members of Proteobacteria are reported to play an important role in colonization of healthy gut anaerobes by maintaining a reduced redox potential[54].

Euryarchaeota and Thermoprotei are the most abundantly found occupying 41% each of archaea population. Members of these groups are associated with assisting in bowel movements and polyamine production, respectively[55]. Strains of methanogenic Euryarchaeota such as Halobacteria and Methanobacter were highly abundant. Several strains of Woesearchaeota sp. were observed (**Figure3.1**). They were reported to be associated with anaerobic carbon cycling in the gut[56].

Taxonomic analysis of total reads excluding 16s rRNA gene sequence analyzed with Megan+ DIAMOND revealed a high density but low diversity of microbial population (**Figure3.2a and b**). Highest abundance of bacterial population (89%) is observed followed by 7.9% of fungal strains abundance, 2.1% of viral and 0.17% of archaeal presence were identified (**Figure3.2b**). 49% of total identified sequences were identified to belong to phylum Proteobacteria. 24% was identified as Firmicutes, followed by 11.8% Actinobacteria, 5.2 % of Ascomycota (Fungi), and 2.1% of Bacteroidetes. Cyanobacteria was seen to consist of only 0.34% of the total reads identified. At species level identification, *Pseudomonas aeruginosa* was most abundant (12%) (**Figure3.3**). Various strains of *Bacillus* and *Burkholderia* were highly abundant. These microbes play a part in pathogen resistance, prevent secondary bacterial infection, detoxification and host immunity[57,58].

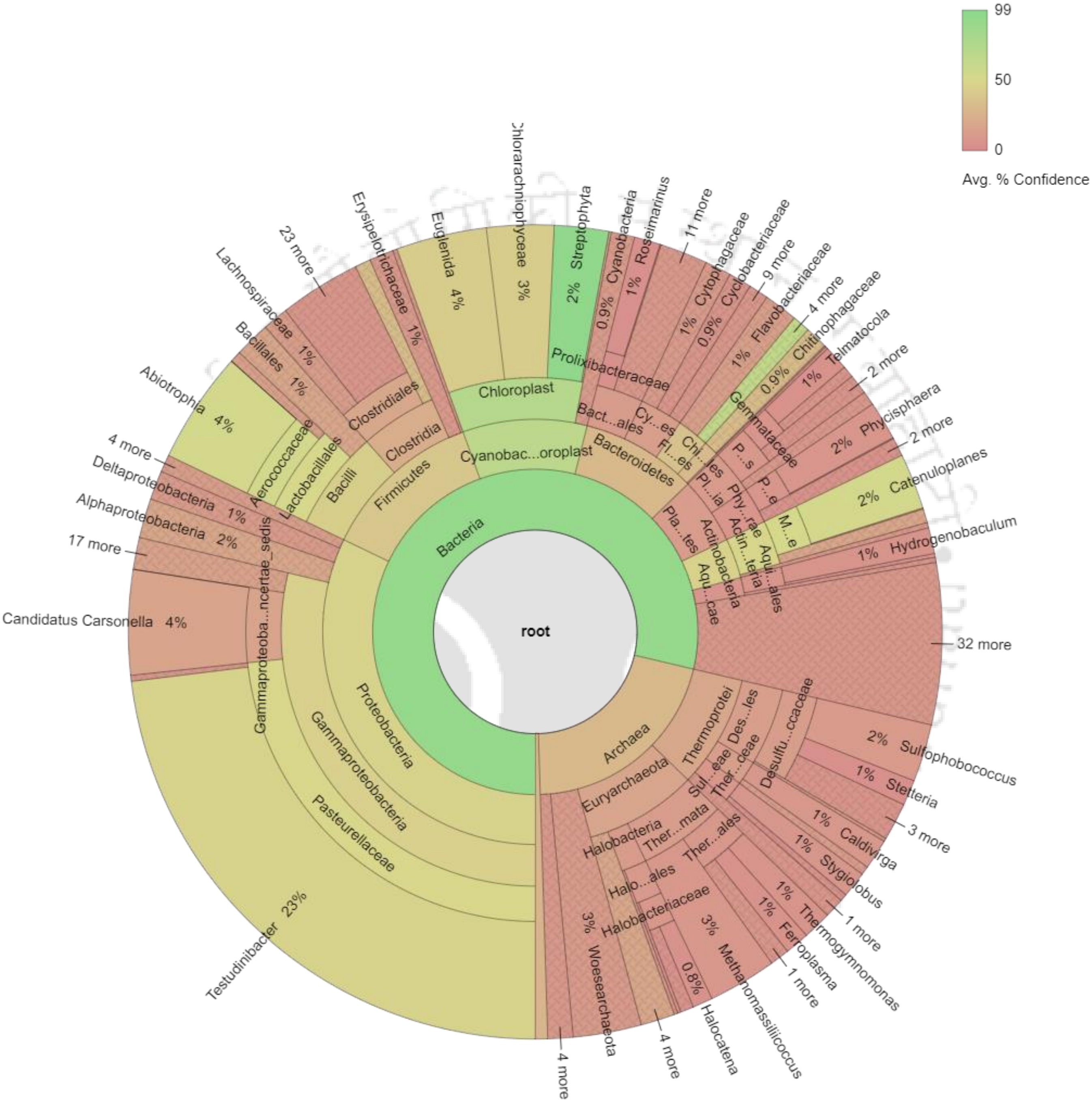
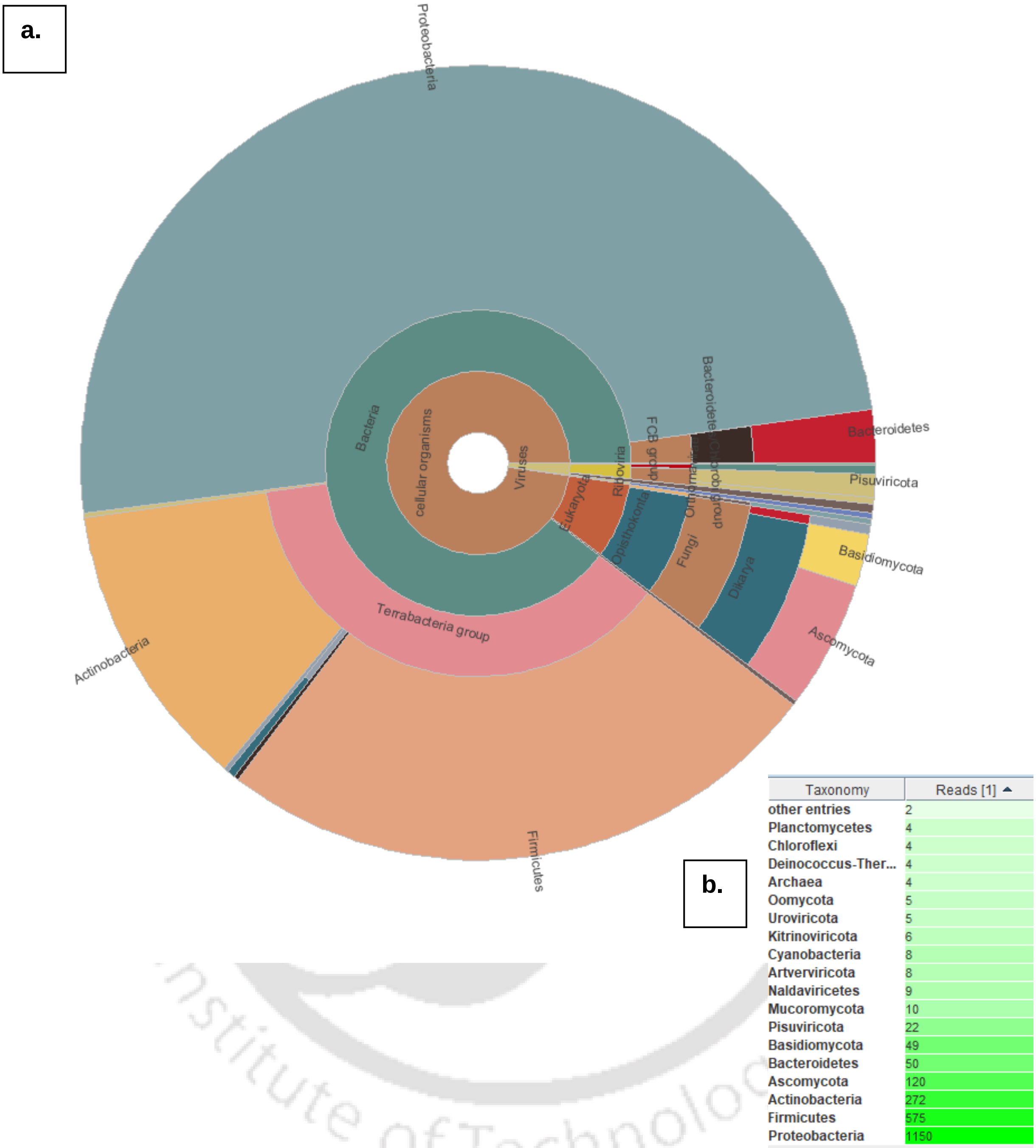


Figure3.1: Radial chart representation of taxonomic abundances through 16s rRNA gene sequence community analysis of shotgun sequenced data (SortmeRNA-based)



Taxonomy	Reads [1] ▲
other entries	2
Planctomycetes	4
Chloroflexi	4
Deinococcus-Ther...	4
Archaea	4
Oomycota	5
Uroviricota	5
Kitrinoviricota	6
Cyanobacteria	8
Artverviricota	8
Naldaviricetes	9
Mucoromycota	10
Pisuviricota	22
Basidiomycota	49
Bacteroidetes	50
Ascomycota	120
Actinobacteria	272
Firmicutes	575
Proteobacteria	1150

Figure3.2: a. Radial representation of the taxonomic composition of total reads at phylum level

b. Heatmap based on read counts and percentage abundances of phylum observed.

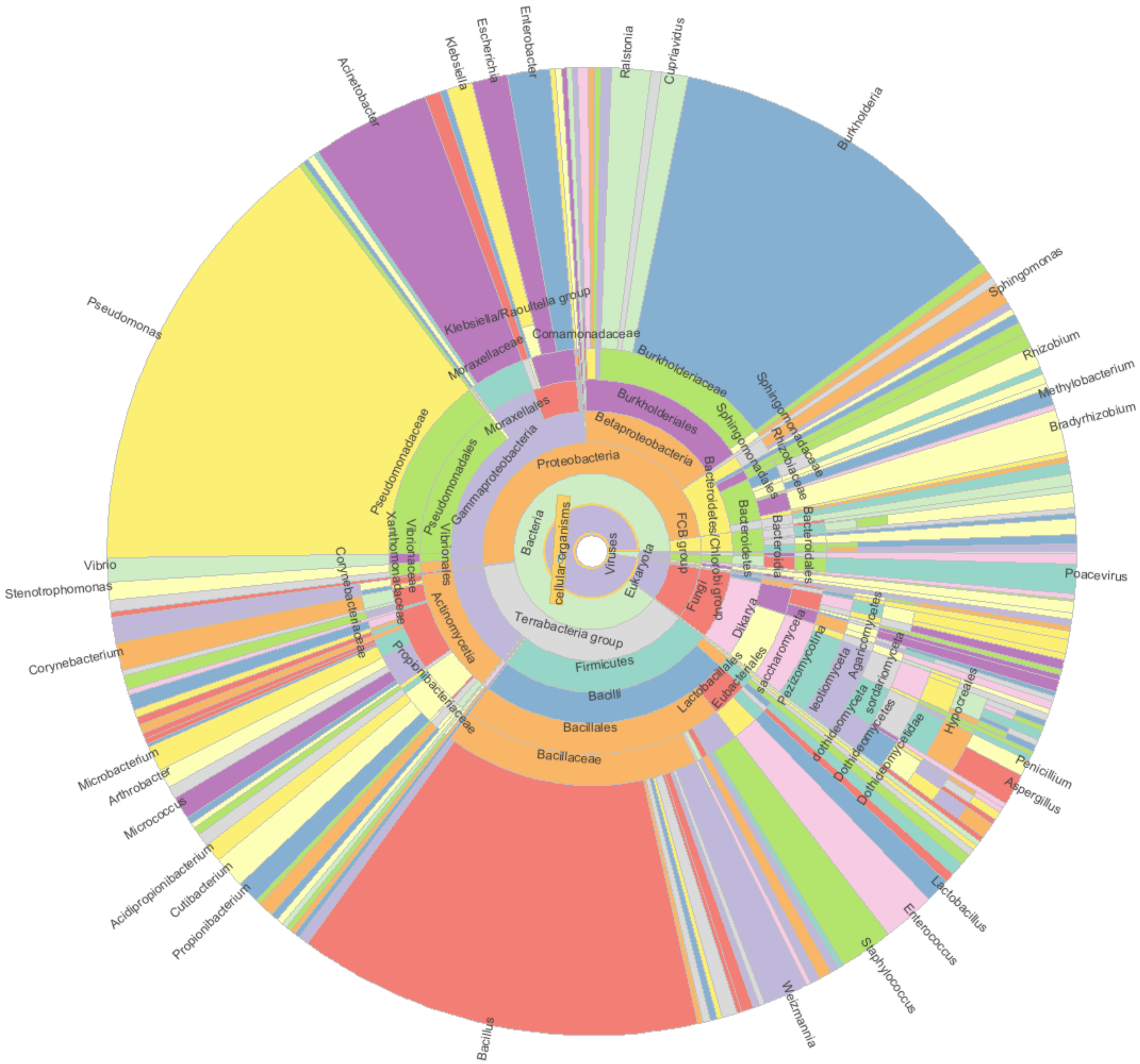


Figure3.3: Radial Chart representation taxonomic abundance identified through total reads at genus level.

Comparative taxonomy analyses

16s rRNA gene sequenced (V3-V4 region) reads were analyzed for the taxonomic composition using QIIME2 against NCBI. Comparative taxonomic analyses using SortmeRNA identified, MEGAN+DIAMOND identified and QIIME2 identified reads of the *S. cynthia ricini* 5th instar larva samples resulted in highest number of features identified through the 16s rRNA gene sequence reads extracted from the shotgun sequence sample at the genus level, confirming a higher resolution through shotgun analyzed samples (**Table3.4**).

Almost all the QIIME2 identified taxonomies were shared by the ones which were MEGAN+DIAMOND-based and SortmeRNA-based identified (**Figure3.4**). Number of features were low for the QIIME2-based 16s rRNA gene sequence analyzed samples. Almost all reads were classified under bacteria. Further, 58.41% of the reads remained unidentified at the domain level as unclassified bacteria. Most reads remained unclassified after the phylum level (**Figure3.5**). Total reads-based identification resulted in lowest number of identified features but higher abundances of the identified features. On the other hand, the highest diversity of features identified were through the SortmeRNA-based samples.

MEGAN+DIAMOND methodology, while providing a deeper insight into the identified taxonomies at the genus and species level identified higher reads of most species identified (**Figure3.6a, b, c, d**). It also identified higher number at the genus and species level many of which many were unique and not found through other methods. Following, CPM normalised counts many bacterial species such as members of *Serratia, Vibrio, Acidipropionibacterium, Bortdetella, Hyphomicrobium,* *Bradyrhizobium, Halomonas*, etc. were identified only in MEGAN+DIAMOND sample. 33 fungal species included members of Aspergillus, Saccharomycetaceae, Helotiales, Penicillium, were identified.

In many studies, fungal populations have been found to play a major role in abundance and function in lepidoptera gut including many silkworm species[24,59–61]. In this study, members of Ascomycota and Basidiomycota were the most dominant similar to other studies. Members of Mucormycota and Oomycota were also observed. Aspergillus and Penicillium in many insect gut microbiome studies have been known to degrade xenobiotics, produce antifungal substances against pathogenic fungi and have phytotoxic properties[62,63]. Many fungal species are also associated with high digestion capability by pest insects[64].

Yeast strains including members of Saccharomyces and *Candida*, observed in our study has been known to provide nutritional advantages to the host. They have been studied to be involved in amino acid and fatty acid metabolism and also metamorphosis[65]. Drosophila melanogaster gut microbiome study showed that they would prefer and develop on plants containing fungi or yeast even when mycotoxin harming the insects are produced by them[62,66,67]. A few pathogenic strains of *Fusarium* have been observed in this study[68].

Virus from members of *B. mori* nucleopolyhedrovirus, Martellivirus, Sugarcane streak mosaic virus, Caulimovirus, Poacevirus were identified only in MEGAN+DIAMOND. Low counts of synthetic construct virus were also observed. Strains such as Serratia sp., are known to present in most silkworm and other lepidopteran species. They are associated with modulating virulence factors of related pathogenic strains[27,69–72]. They are also commercially utilized as a source of proteolytic enzymes Serralysin for research purposes as promoting hemolymph bleeding and Serratiopaptidase for their anti-inflammatory properties[73,74]. These enzymes are known to increase the pathogenicity of the strain[75]. Studies on *B. mori* metagenome reported that in spite of having a lower taxonomic abundance was found to be dominant in the RNA-based studies (metatranscriptomics), indicating their higher metabolic activity and role in the gut. Among the viral strains, many plant-viral strains have been identified. Studies have reported of plant viral transmission through insects. The transmission potential is highly associated with the vector capability of other microbial population in the gut. These transmissions are dependent on the virus-protein interaction manipulating the vector and the insect host. These viruses have also been reported to alter the insect gut microbiome or not be associated with subsequent transmission. Some of the viral strains found in low abundance like *B. mori* nucleopolyhedrovirus, belonging to Baculovirus are highly pathogenic in silkworm *B. mori*. Previous studies showed that the strain was pathogenic in Eri silkworms due to the high expression of SUH gene, linked with facilitating infection[76]. Presence of these viral strains along with a few pathogenic bacterial strains could indicate a low amount of infection in the studies samples.

These observations based on MEGAN+DIAMOND analysis resulted in deeper insights into the taxonomic presence of the 5th larval stage of *S. cynthia ricini* and also helped us with understanding their functional role better. Community analysis using 16s rRNA gene sequence reads provided with a higher diversity of bacterial population better for analysing alterations based on environment. Total read based taxonomy analysis was found to be advantageous compared to the other methodologies to a great extent. The low read and diversity of amplicon sequenced samples provides us a difference in the analysis depth.

Cladogram showed the phylogenetic relationship of the taxa identified through the different techniques along with the insights in the microbial species identified (**Figure3.7**). Fungal and viral groups and members from the PVC and Terrabacteria group were mostly identified using total reads while members from the FCB group including Bacteroidetes were found to be relatively low. Archaeal species were also relatively low in MEGAN+DIAMOND-based identification. SortmeRNA-based sample showed presence of almost all the groups but in relatively low abundance.

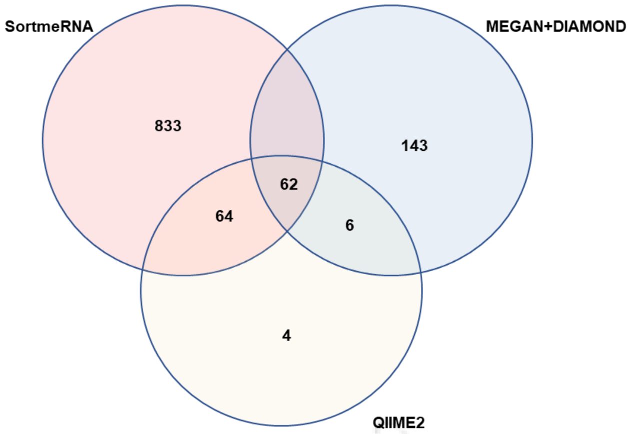


Figure3.4: Venn representation of shared and unique features identified using SortMeRNA (16s rRNA gene sequence reads from shotgun sequence-based data), MEGAN+DIAMOND (total reads of shotgun sequence-based data) and QIIME2 (16s rRNA gene sequence analysis-based) at genus level.

Domain	SortmeRNA	MEGAN+DIAMOND	QIIME2
Archaea	97	4	1
Bacteria	2172	2067	2301
Eukaryota	0	184	0
Viruses	0	50	0

Table3.4: Comparative summary of the number of features identified through SortmeRNA, MEGAN+DIAMOND, QIIME2

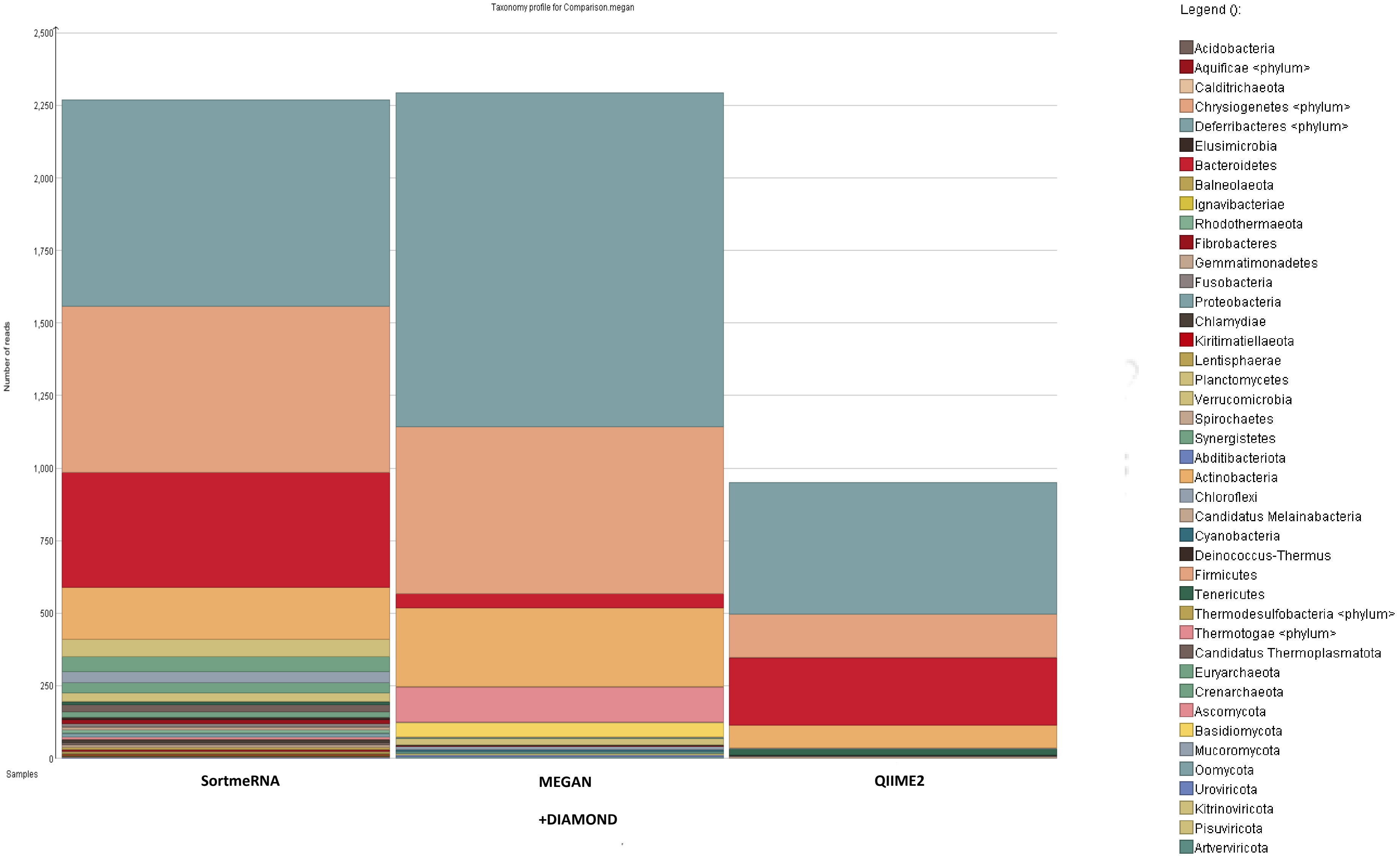


Figure3.5: Relative taxonomic abundance (normalised) plot at phylum level summarising the differences observed via the approaches applied

Figure3.6a: Heatmap representation of the differential abundance of taxonomic composition at the species level

Figure3.6b: Heatmap representation of the differential abundance of taxonomic composition at the species level

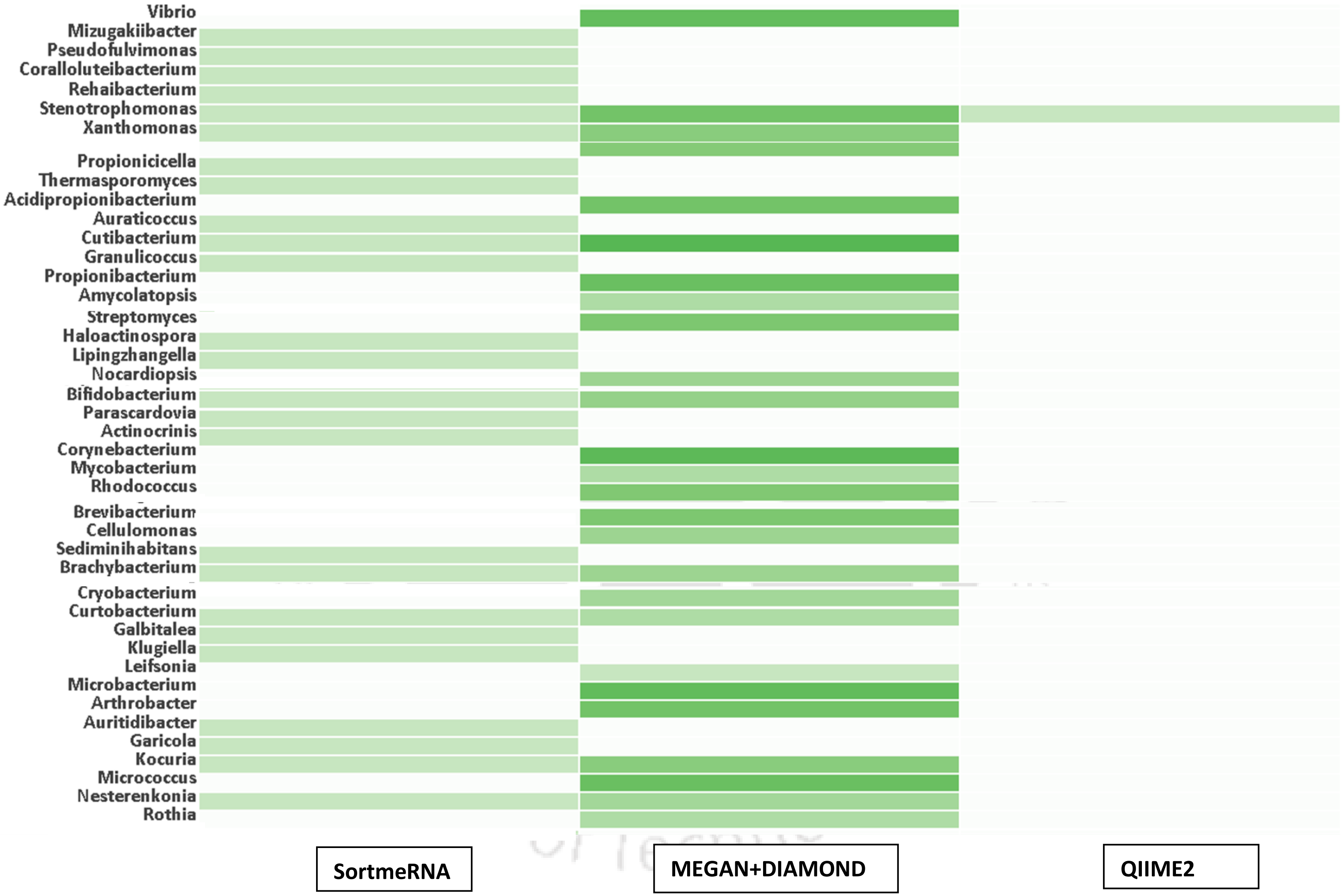


Figure3.6c: Heatmap representation of the differential abundance of taxonomic composition at the species level

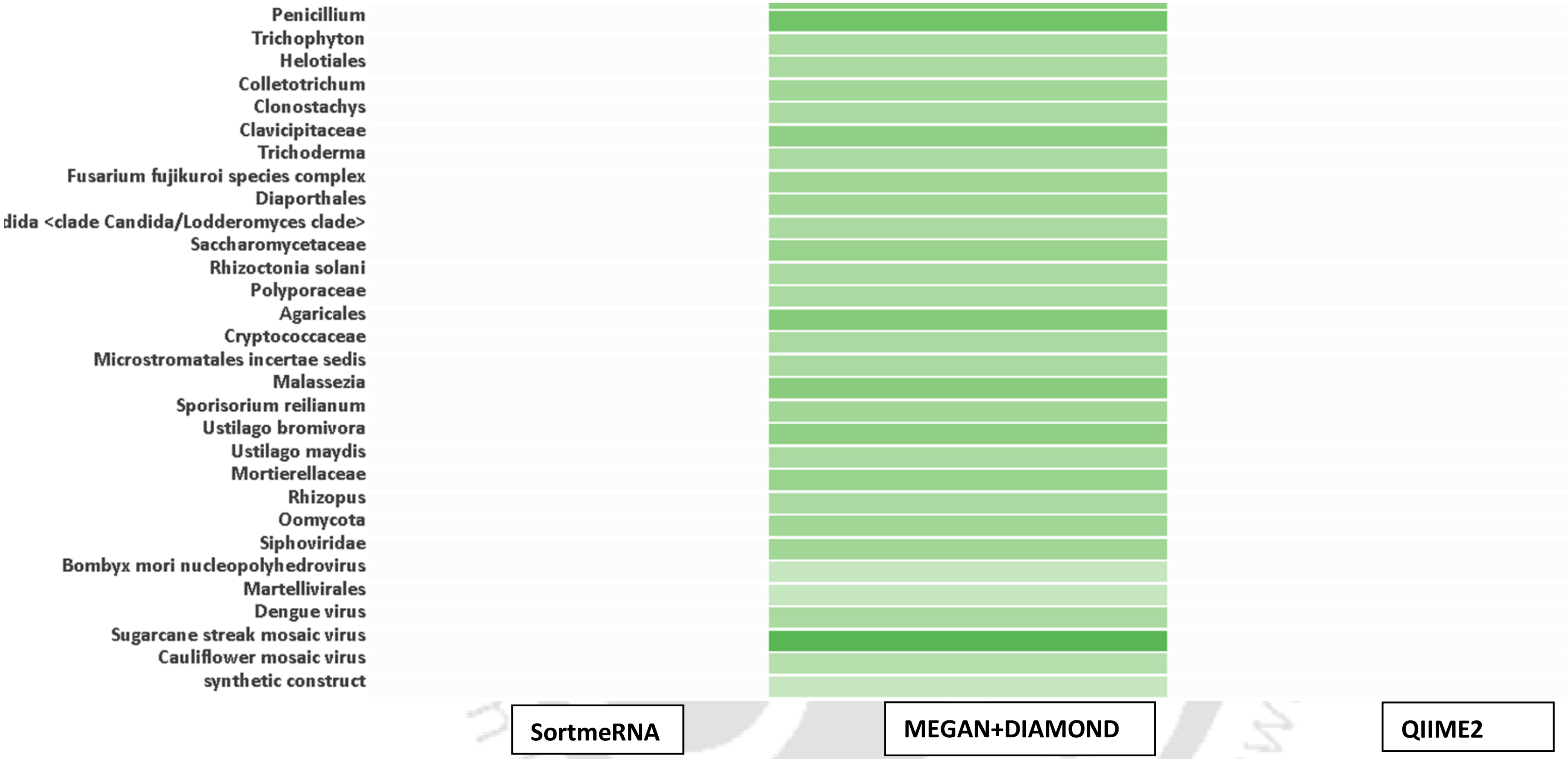


Figure3.6d: Heatmap representation of the differential abundance of taxonomic composition at the species level

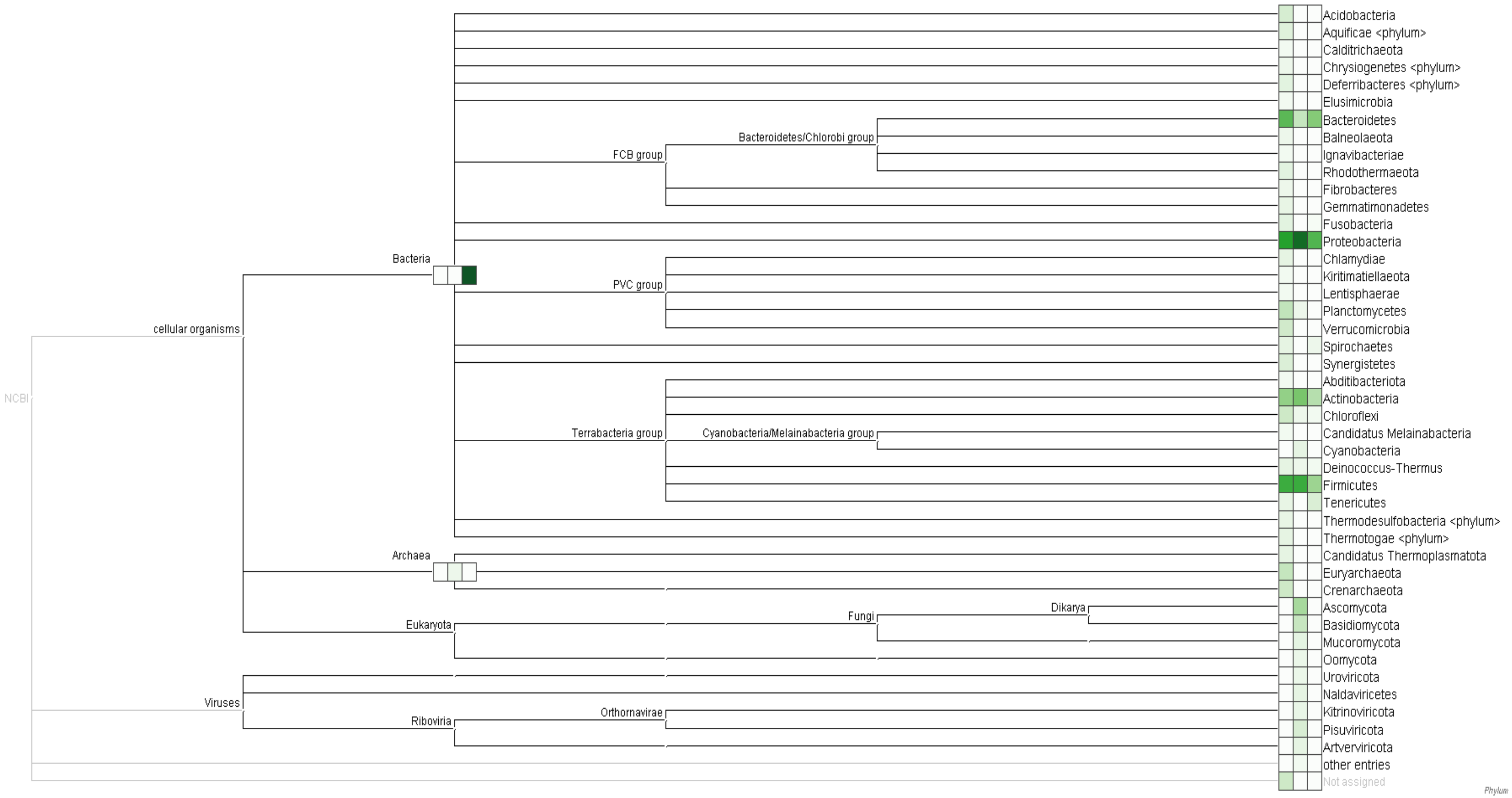


Figure3.7: Cladogram of the 3 groups of samples (SortmeRNA, MEGAN+DIAMOND, QIIME2)

Functional analyses

Functional annotation and pathway inferences based on pathway reconstruction using KEGG pathways showed the 83% of the genes present to be involved in metabolism, 11% involved in genetic information processing, 5% in cellular processes and 1% in two-component system (**Figure3.8**). Genes involved in carbohydrate metabolism were observed to be very high comprising of 21% of total identified genes and 25% of genes involved in metabolism. Other major metabolism involved genes comprised of lipid metabolism and amino acid metabolism,15% and 12% of the total genes respectively. 8% of the genes are involved in metabolism of vitamins and cofactors and 4% in metabolism of xenobiotics and degradation. Peroxisome related pathways were seen to comprise of 50% of the cellular processes followed by cell growth and death (33%) and bacterial chemotaxis (17%).

MetaCyc database-based reconstruction of pathways detailing only the metabolic pathways, resulted in 54% of the genes to be involved in various biosynthesis pathways, 33% in degradation pathways, 4% in energy metabolism, 2% in fermentation-based pathways (**Figure3.9**). 38% of biosynthesis pathways comprised of biosynthesis of secondary metabolites. This indicated that the gut microbial genes present were highly responsible for metabolic pathways mediating ecological interactions responsible for the interactive network of the meta-organism. Flavonoid biosynthesis (38%) and antibiotics biosynthesis (13%) were some of the pathways present in higher amounts. Genes involved in cofactor and vitamins biosynthesis were highly abundant (15% of biosynthesis pathways) and was more than carbohydrate biosynthesis (10%), lipid biosynthesis (10%) and polyamine biosynthesis (7%). Aromatic compound degradation was one of the pathways found highly abundant

(16%), followed by secondary metabolites degradation (12%), amino acid degradation (11%), sugars and polysaccharide degradation (9%), fatty acid degradation (8%). Plant hormone degradation pathways involved genes of significant amount were also seen to be present. 25% of energy metabolism pathway genes were contributed to TCA variants, 17% to glycolysis variants, 5% to methanogenesis. Genes contributing to chitin degradation to ethanol (4%), long fatty acid ester synthesis for microdiesel production (4%) were also significantly present. COG database-based pathway reconstruction showed 44% comprising of pathways involved in metabolism, 38% in cellular processing and signalling, 10% in information storage and processing (**Figure3.10**). Highest abundance of metabolism comprised of amino acid transport and metabolism (23% of metabolism), followed by lipid metabolism and secondary metabolites biosynthesis, transport and catabolism (17% each of metabolism). Among cellular processes and signalling, highest abundance of cytoskeleton pathway genes (36% of cellular processes and signalling) was observed. Genes involved in post translational mechanism and chaperones (16% of cellular processes) were also highly abundant. Other genes involved in cell cycle control, cell division, defence mechanism, envelope biogenesis, cell motility and signal transduction mechanisms were also observed.

Based on these observations it is apparent that microbial population have a dynamic role in the different pathways of the host majority being metabolic pathways post-translation. Microbial genes annotated were involved with amino acid and lipid metabolism indicating their importance in the primary metabolic pathways of the host[64]. Secondary metabolism was also highly regulated by microbial genes including the degradation of plant derived aromatic compounds and toxins[54]. Numerous other functions including vitamins and co-factor biosynthesis aiding the insect development were present in high amounts. Lipids, fatty acids and their intermediates are responsible for metabolic pathways that promote signalling, energy metabolism, aid in digestive abilities, formation of structural components and produce hormones[27]. They can also act as biomarkers[77,78]. Based on the vast range of functions annotated, the role of the present microbes in specific required area could be analyzed.

Comparative analysis

Normalised relative abundance of KEGG-derived pathways (L3-level) derived pathways of the annotated genes compared to the predicted pathways derived from 16s rRNA amplicon sequenced sample showed that all the predicted functions were encoded by genes present in the shotgun annotated samples (**Figure3.11a**). However, the relative functional abundance of most of the shotgun-based samples were higher compared to the amplicon sequence predicted based samples (**Figure3.11b**). KEGG-derived pathways including fatty acid biosynthesis, various amino acid metabolism including tryptophan, histidine, valine, leucine and tryptophan metabolism, drug metabolism, foliate metabolism, various carbohydrate including galactose and pentose phosphate pathway and various vitamins and cofactor metabolism pathways including thiamine metabolism and pantothenate and CoA metabolism are highly abundant in shotgun based annotated samples compared to QIIME-PICRUSt2 predicted samples. Some pathways such as purine and pyrimidine metabolism, two-component metabolism, and some unknown functions were observed to be highly abundant in the PICRUSt2 predicted samples indicating the actual functions to be either less abundant or be involved in different functions than predicted. Comparing the amplicon-based predictive function annotation and the shotgun-based gene annotated functions, it is seen that the gene-base annotation resulted in a more precise analyses with more depth.

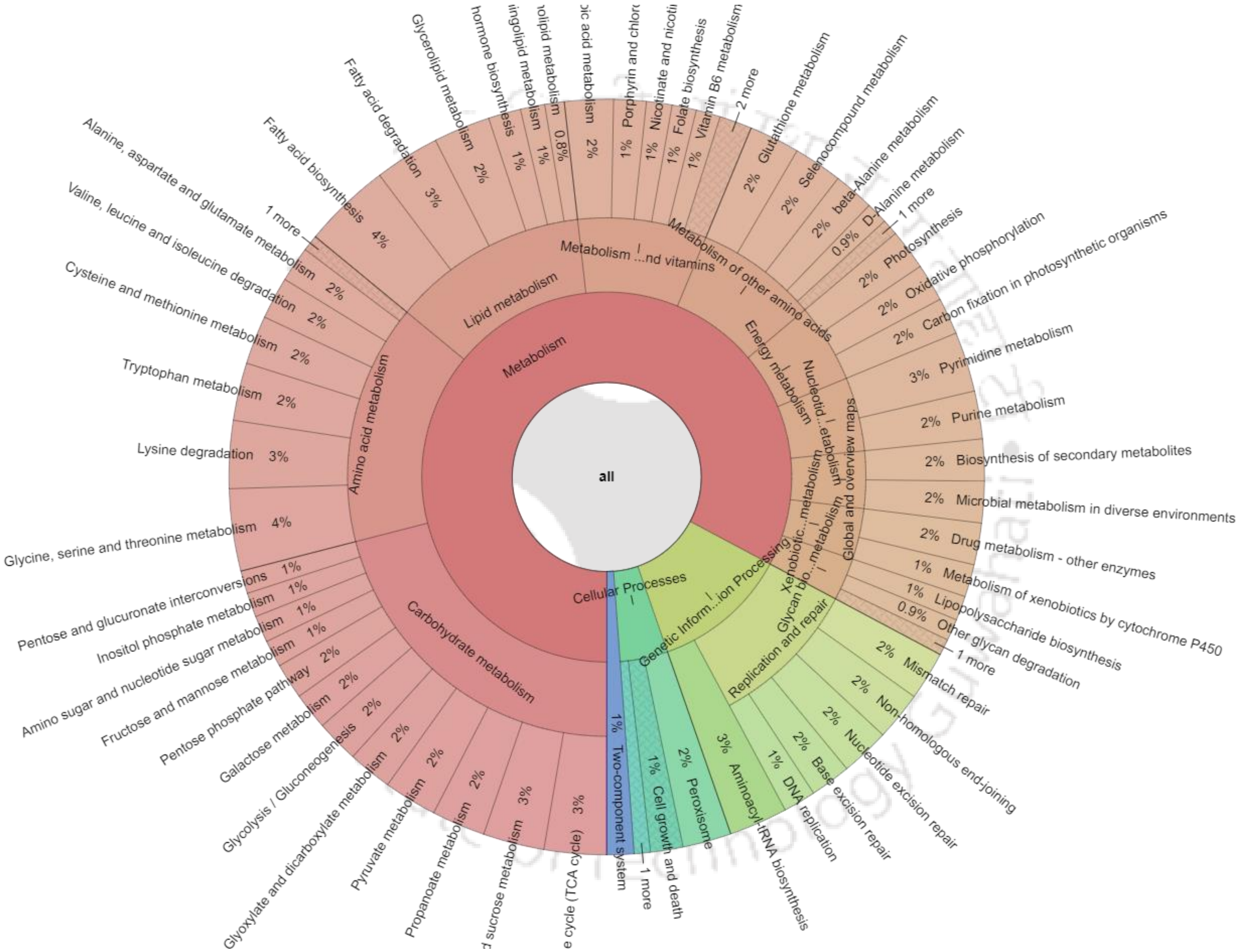


Figure3.8: Radial chart representation of abundance of pathways based on KEGG pathway database

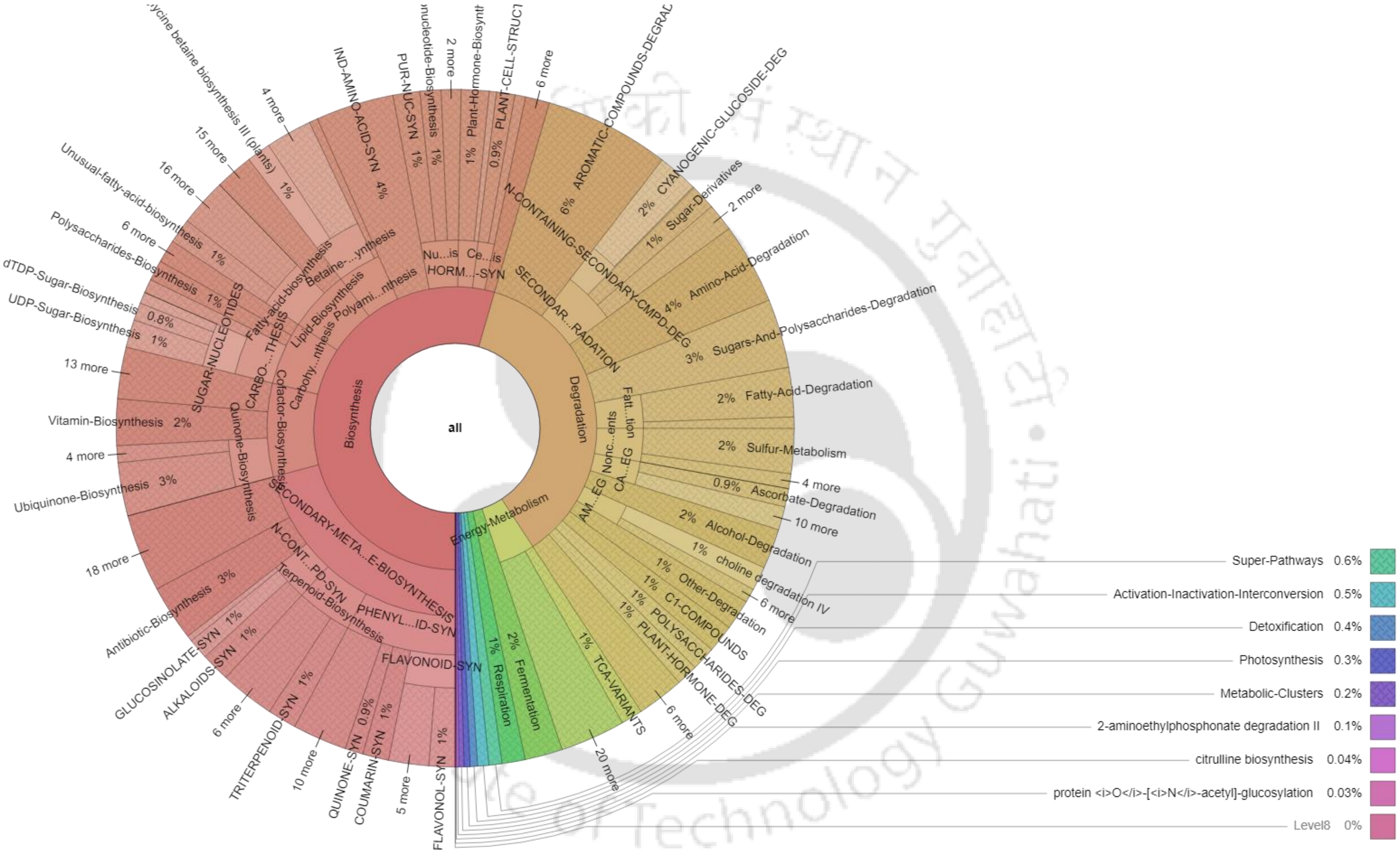


Figure3.9: Radial chart representation of abundance of metabolic pathways based on Metacyc pathway database

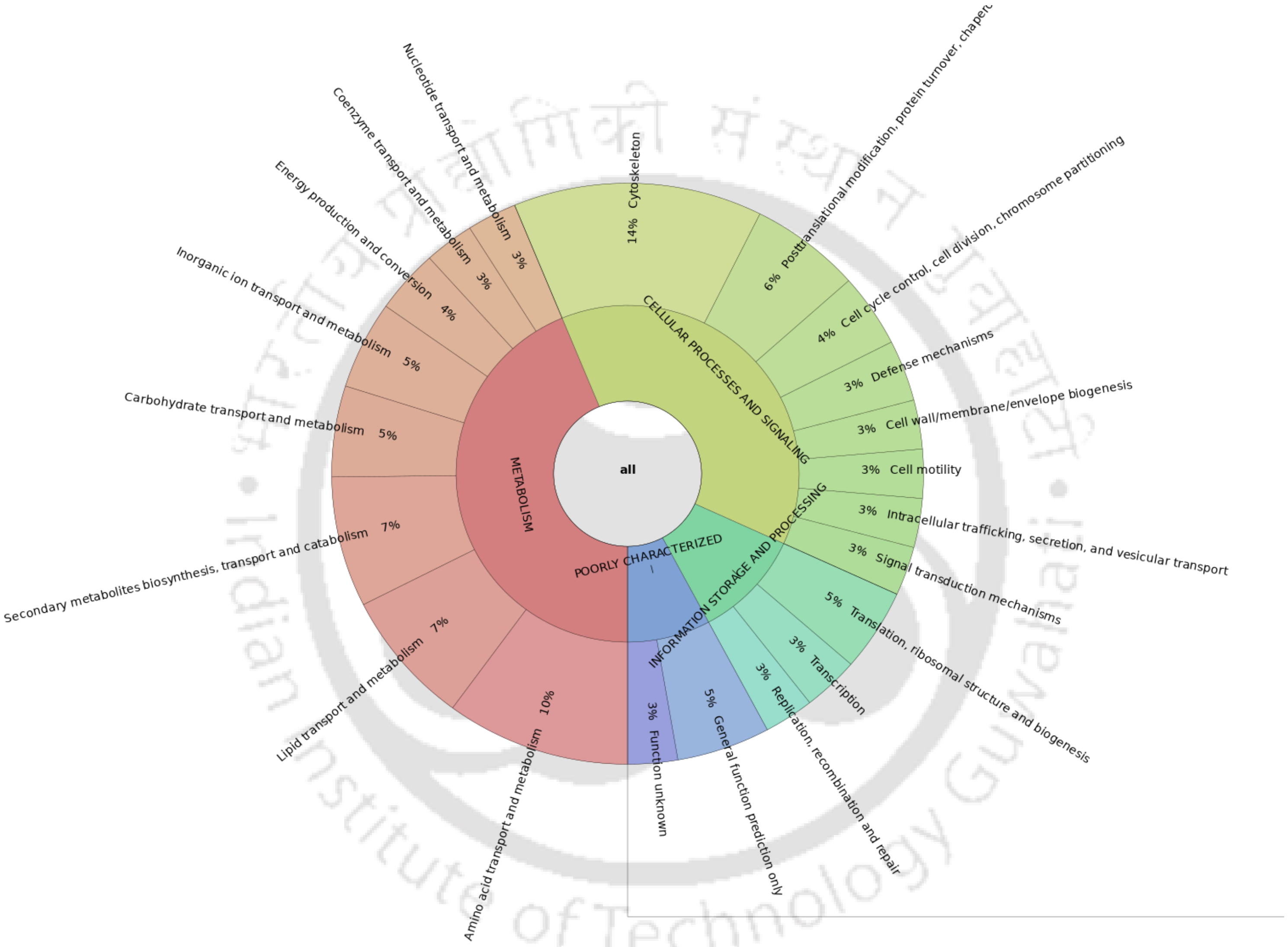


Figure3.10: Radial chart representation of abundance of pathways based on COG database

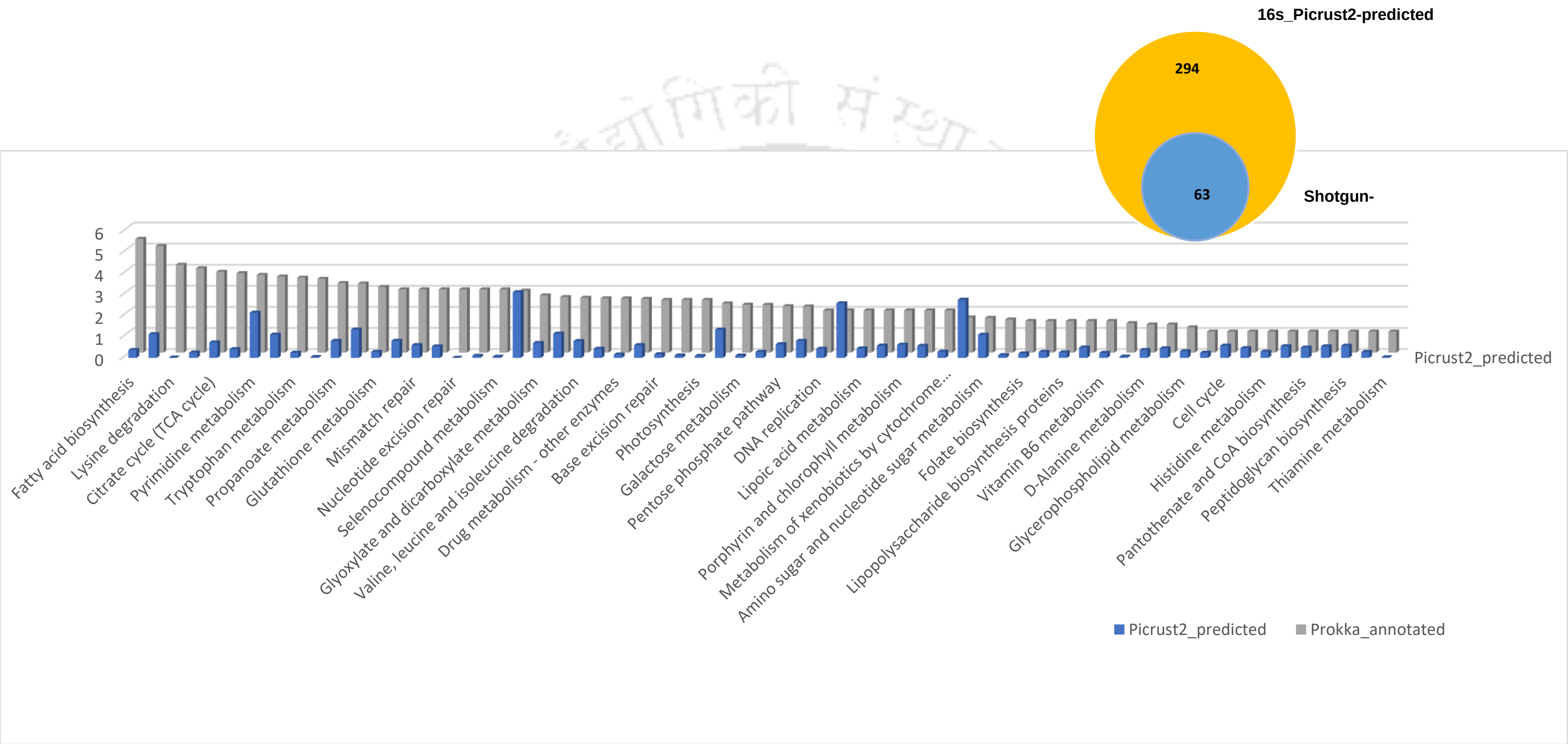


Figure3.11: a. Venn diagram representing the shared functions between PICRUSt2 predicted pathways and shotgun-based annotated pathways.

b. Comparative bar plots representing the abundance differences between PICRUSt2 predicted pathways and shotgun-based annotated pathways (KEGG-L3)

Antimicrobial Resistant Genes

Antimicrobial resistant genes were identified using RGI software against the assembled reads resulted in most abundance of B-lactamase class of antimicrobial resistance genes (**Table3.5**). They were mostly observed in Clostridium, Proteobacteria, Firmicutus, Actinobacteria, Caulobacterales, Saccharomycetes and Basidiomycota. B-lactamases provided resistance multi-drug resistance against compounds containing lactam rings. They have been observed in many studies to have a relatively higher speed and have been found to be relatively more abundant in several insect species. Other observed genes included from classes lipopeptides and tetracyclins from Proteobacteria, Actinobacteria and Firmicutus; Pyrimidine from Proteobacteria and Actinobacteria and Glycopeptides from Pseudomonas and Proteobacteria. iPATH view of the reconstructed pathways was observed showing the contribution of microbial genes in energy, carbohydrates, amino acids, nucleotide, cofactors and vitamins, lipids metabolism and biosynthesis and degradation of secondary metabolites, xenobiotics biodegradation and metabolism (**Figure3.12**). Genes of some pathways including glycan biosynthesis and metabolism, metabolism of terpenoid and polyketide were seen to be completely absent from the annotated metagenome. The overview of the genes involved in antibiotic resistance gene pathways are shown in **Figure3.13**. inferring a majority of the metabolic pathways involved compared to the pathways of the total genes present.

Most of the genes associated with carbohydrate metabolism, other amino acid metabolism, lipid metabolism belonged from the ARG classes inferring the amount of role of metabolism of lipid, amino acid and carbohydrate in drug resistance.

ARDB class	Organism	Subtypes (Genes)	Abundance
B-lactamase	Clostidium sp., Proteobacteria, Basidiomycota, Firmicutus, Actinobacteria, Caulobacterales, Saccharomycetes	*gloB, rnjA, ybbE, comEC, yqjP, vicX, rnz, pbpX, ampC, bla, yhfl, rnjB, YSH, pksB_1, glx2, ytnP, nylB*	0.45
Lipopeptides	Proteobacteria, Actinobacteria, Firmicutus	*slmA* *tnp3514a, puuR_1, PPA1638, rpoB, rimK, recR, amaB*	0.11
Tetracyclins	Actinobacteria, Bacteroidetes, Firmicutus	*tetR, tetW*	0.08
Pyrimidine	Proteobacteria, Actinobacteria	*dfrA,dfrK*	0.011
Glycopeptides	Pseudomonas, Proteobacteria	*rpoB, phoP*	0.003

Table3.5: Summarised representation of the normalised relative abundance of the classes of Antimicrobial Resistance Genes (ARGs) annotated.

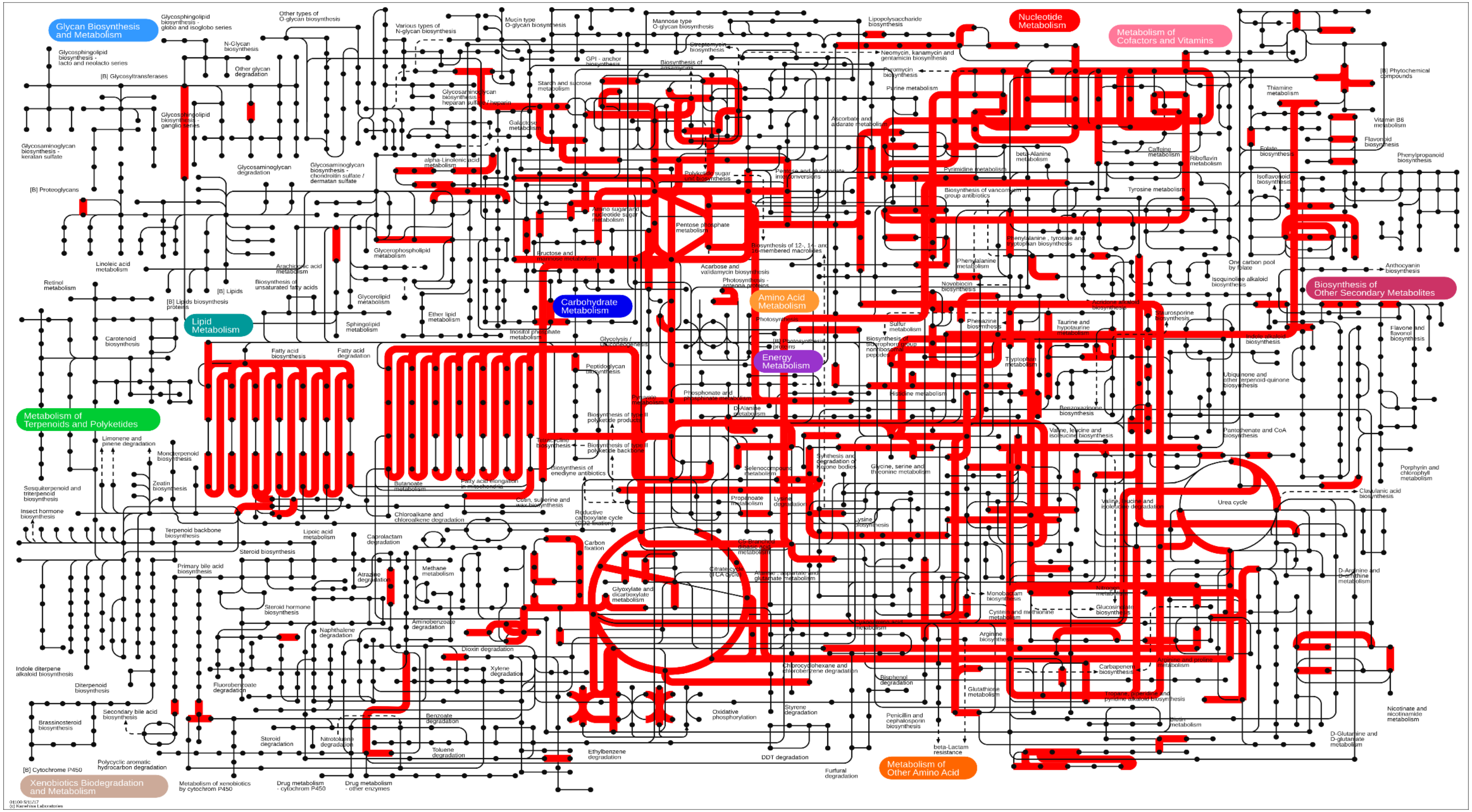


Figure3.12: iPATH visualisation representing a comprehensive summary of the metagenome identified genes involved in various pathways.

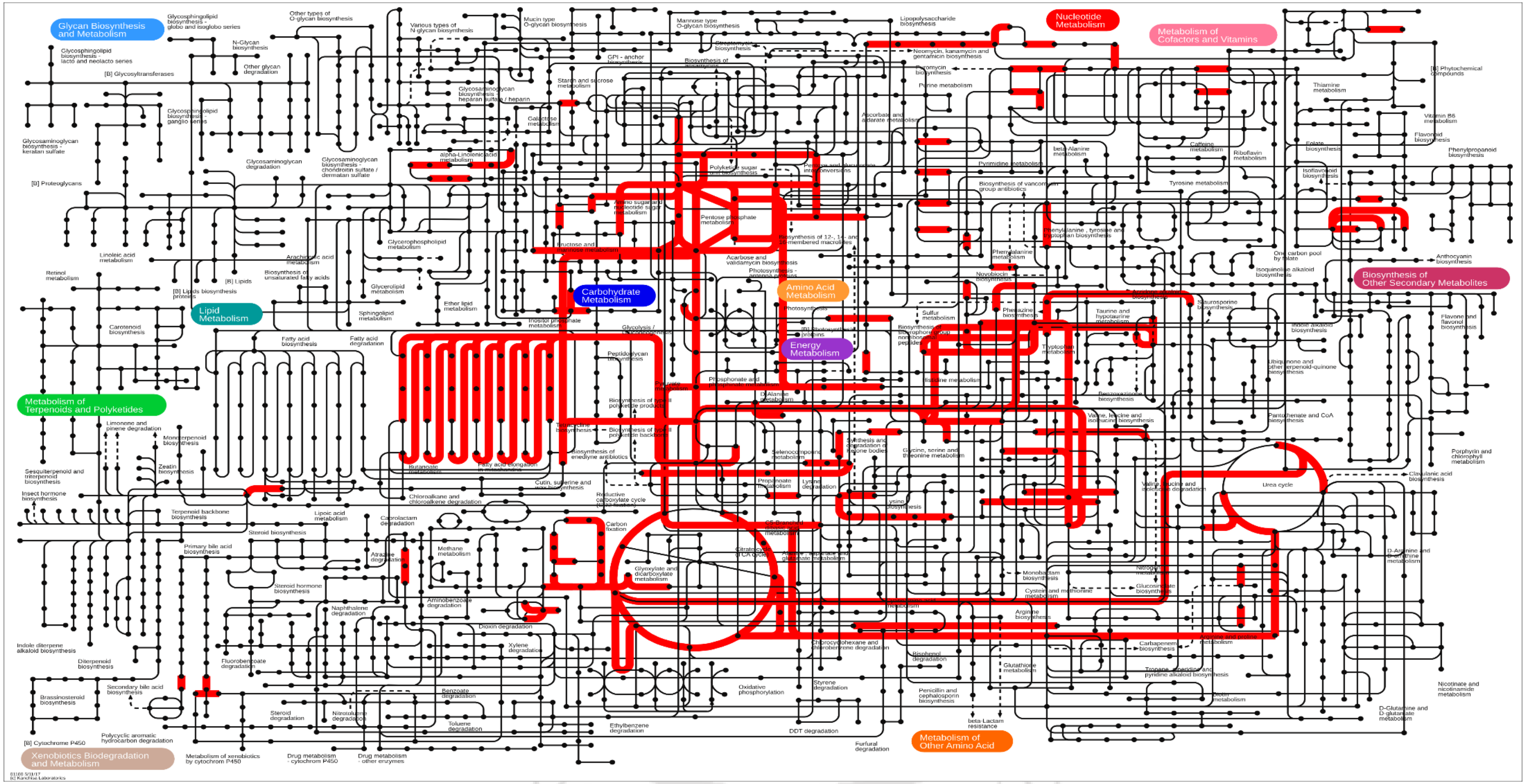


Figure3.13: iPATH representation of the present antimicrobial resistance genes involved pathways.

CONCLUSION

In the present study, the metagenome of *S. cynthia ricini* was explored with shotgun sequencing of the silkworm larva. The results demonstrated the higher depth of analyses of the metagenome of the sample. Detailed view of the taxonomic and functional abundances of the gut microbes were outlined. Optimization of the methodologies for the analyses regarding the organism were explored and summarised. Comparative details with respect to amplicon sequence based microbial population of the sample was also studied resulting in 89% abundance of bacterial population, 7.9% of fungal strains abundance, 2.1% of viral and 0.17% of archaeal presence revealing a high percentage of fungal strains in the gut microbiota of the silkworm. Important strains including *Serratia, Vibrio, Acidipropionibacterium, Bortdetella, Hyphomicrobium, Bradyrhizobium, Halomonas* could only be identified using shotgun analyses. It also revealed that shotgun analyses helped in identification of a high density but low diversity of bacterial population. Most of the genes associated with carbohydrate metabolism, amino acid metabolism, lipid metabolism belonged from the ARG classes.

This study hence, helps us with the basic insights on the 5th instar larva metagenome carried out via shotgun analyses. It also focuses on advantages and disadvantages of shotgun analyses over amplicon sequence-based analyses and usage of standardised and optimised techniques for our study according to our requirements.

Further integrative analyses could be explored for understanding the interactive metabolic activities of the microbiome comparative analyses observing the role of microbes associated with various diseases in the silkworm would lead to the development of better strains with higher resistance against those diseases.

REFERENCES

1. Zhang, L. *et al.* Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front Microbiol* **12**, 3847 (2021).
2. de Vries, J. J. C. *et al.* Recommendations for the introduction of metagenomic next-generation sequencing in clinical virology, part II: bioinformatic analysis and reporting. *Journal of Clinical Virology* **138**, 104812 (2021).
3. Fatma¸, F. *et al.* Metagenomics Next Generation Sequencing (mNGS): An Exciting Tool for Early and Accurate Diagnostic of Fungal Pathogens in Plants. *Journal of Fungi 2022, Vol. 8, Page 1195* **8**, 1195 (2022).
4. Joseph, T. A. & Pe'er, I. An Introduction to Whole-Metagenome Shotgun Sequencing Studies. *Methods Mol Biol* **2243**, 107–122 (2021).
5. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol. 2017 35:9* **35**, 833–844 (2017).
6. Ji, B. & Nielsen, J. New insight into the gut microbiome through metagenomics. *Adv Genomics Genet* **5**, 77–91 (2015).
7. Sanabria, A. M., Janice, J., Hjerde, E., Simonsen, G. S. & Hanssen, A. M. Shotgun-metagenomics based prediction of antibiotic resistance and virulence determinants in Staphylococcus aureus from periprosthetic tissue on blood culture bottles. *Sci. Rep. 11:1* **11**, 1–13 (2021).
8. Jovel, J. *et al.* Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol* **7**, 459 (2016).
9. Xiang, K. L. *et al.* Organization, Phylogenetic Marker Exploitation, and Gene Evolution in the Plastome of Thalictrum (Ranunculaceae). *Front Plant Sci* **13**, 1334 (2022).
10. Bharti, R. & Grimm, D. G. Current challenges and best-practice protocols for microbiome analysis. *Brief Bioinform* **22**, 178–193 (2021).

11. Pearson, T. *et al.* Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci U S A* **101**, 13536–13541 (2004).

12. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Comm. 10:1* **10**, 1–11 (2019).

13. Louca, S., Doebeli, M. & Parfrey, L. W. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* **6**, 1–12 (2018).

14. Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D. & Konstantinidis, K. T. Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS One* **9**, e93827 (2014).

15. Kameoka, S. *et al.* Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1–V2 and V3–V4 primer sets. *BMC Genomics* **22**, 1–10 (2021).

16. Xia, X. *et al.* Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Front Microbiol* **8**, 663 (2017).

17. Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun* **469**, 967–977 (2016).

18. Peterson, D. *et al.* Comparative Analysis of 16S rRNA Gene and Metagenome Sequencing in Pediatric Gut Microbiomes. *Front Microbiol* **12**, 1651 (2021).

19. Zeng, F., Wang, Z., Wang, Y., Zhou, J. & Chen, T. Large-scale 16S gene assembly using metagenomics shotgun sequences. *Bioinformatics* **33**, 1447–1456 (2017).

20. Durazzi, F. *et al.* Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci. Rep. 11:1* **11**, 1–10 (2021).

21. Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L. & Leddy, M. B. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One* **15**, (2020).

22. Festa, S. *et al.* Integrating Shotgun Metagenomics, 16s Rrna Gene Metabarcoding and Culture Approaches: A Better Outlook for Functional Profiling of a Pah-Contaminated Soil. *SSRN Electronic Journal* (2022) doi:10.2139/SSRN.4302698.

23. Purushothaman, S., Meola, M. & Egli, A. Combination of Whole Genome Sequencing and Metagenomics for Microbiological Diagnostics. *International Journal of Molecular Sciences 2022, Vol. 23, Page 9834* **23**, 9834 (2022).

24. Chen, B. *et al.* Gut bacterial and fungal communities of the domesticated silkworm (Bombyx mori) and wild mulberry-feeding relatives. *The ISME Journal 2018 12:9* **12**, 2252–2262 (2018).

25. Dong, H. L. *et al.* Differences in gut microbiota between silkworms (Bombyx mori) reared on fresh mulberry (Morus alba var. multicaulis) leaves or an artificial diet. *RSC Adv* **8**, 26188–26200 (2018).

26. Chen, B. *et al.* Gut bacteria of the silkworm Bombyx mori facilitate host resistance against the toxic effects of organophosphate insecticides. *Environ Int* **143**, 105886 (2020).

27. Chen, B. *et al.* Comparative shotgun metagenomic data of the silkworm Bombyx mori gut microbiome. *Scientific Data 2018 5:1* **5**, 1–10 (2018).

28. Javaid, A. *et al.* Isolation and characterization of bacteria associated with silkworm gut under antibiotic-treated larval feeding. *Brazilian Journal of Biology* **84**, (2021).

29. Chen, B. *et al.* Gut bacteria of the silkworm Bombyx mori facilitate host resistance against the toxic effects of organophosphate insecticides. *Environ Int* **143**, 105886 (2020).

30. Bay, V., Gür, S. & Bayraktar, O. Plant-derived tormentic acid alters the gut microbiota of the silkworm (Bombyx mori). *Sci Rep* **12**, 13005 (2022).

31. Sambrook, J. & Russell, D. W. Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harb Protoc* **2006**, pdb.prot4455 (2006).

32. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

33. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114 (2014).

34. Liang, K. ching & Sakakibara, Y. MetaVelvet-DL: a MetaVelvet deep learning extension for de novo metagenome assembly. *BMC Bioinformatics* **22**, 1–21 (2021).

35. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res* **27**, 824–834 (2017).

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

37. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).

38. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* **42**, D633 (2014).

39. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res* **17**, 377 (2007).

40. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **12**, 1–10 (2011).

41. Bağcı, C., Patz, S. & Huson, D. H. DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences. *Curr Protoc* **1**, e59 (2021).

42. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

43. Ye, Y. & Doak, T. G. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLoS Comput Biol* **5**, e1000465 (2009).

44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4* **9**, 357–359 (2012).

45. Zhao, Q. A Study on Optimizing MarkDuplicate in Genome Sequencing Pipeline. doi:10.1145/3309129.3309134.

46. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166 (2015).

47. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**, 281–285 (2012).

48. Heberle, H., Meirelles, V. G., da Silva, F. R., Telles, G. P. & Minghim, R. InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, (2015).

49. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology 2019 37:8* **37**, 852–857 (2019).

50. Darzi, Y., Letunic, I., Bork, P. & Yamada, T. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res* **46**, W510–W513 (2018).

51. Xia, X. *et al.* Metagenomic Sequencing of Diamondback Moth Gut Microbiome Unveils Key Holobiont Adaptations for Herbivory. *Front Microbiol* **8**, 663 (2017).

52. Xia, X., Lan, B., Tao, X., Lin, J. & You, M. Characterization of Spodoptera litura Gut Bacteria and Their Role in Feeding and Growth of the Host. *Front Microbiol* **11**, 1492 (2020).

53. Wang, Y. T. *et al.* Metagenome Sequencing Reveals the Midgut Microbiota Makeup of Culex pipiens quinquefasciatus and Its Possible Relationship With Insecticide Resistance. *Front Microbiol* **12**, 228 (2021).

54. Jing, T. Z., Qi, F. H. & Wang, Z. Y. Most dominant roles of insect gut bacteria: digestion, detoxification, or essential nutrient provision? *Microbiome 2020 8:1* **8**, 1–20 (2020).

55. Horz, H. P. & Conrads, G. The discussion goes on: What is the role of Euryarchaeota in humans? *Archaea* **2010**, (2010).

56. Huang, W. C. *et al.* Comparative genomic analysis reveals metabolic flexibility of Woesearchaeota. *Nat Commun* **12**, (2021).

57. Xu, Y., Buss, E. A. & Boucias, D. G. Culturing and characterization of gut symbiont Burkholderia spp. from the Southern chinch bug, Blissus insularis (Hemiptera: Blissidae). *Appl Environ Microbiol* **82**, 3319–3330 (2016).

58. Ilinskaya, O. N., Ulyanova, V. v., Yarullina, D. R. & Gataullin, I. G. Secretome of Intestinal Bacilli: A Natural Guard against Pathologies. *Front Microbiol* **8**, (2017).

59. Joseph, I., Chellaiah, D. E. & Singh, A. J. A. R. Studies on the influence of Beauveria bassiana on survival and gut flora of groundnut caterpillar, Spodoptera litura Fab. *553 Journal of Biopesticides* **3**, 553–555 (2010).

60. Indiragandhi, P. *et al.* Cultivable bacteria associated with larval gut of prothiofos-resistant, prothiofos-susceptible and field-caught populations of diamondback moth, Plutella xylostella and their potential for, antagonism towards entomopathogenic fungi and host insect nutrition. *J Appl Microbiol* **103**, 2664–2675 (2007).

61. Yuning, L. *et al.* The bacterial and fungal communities of the larval midgut of Spodoptera frugiperda (Lepidoptera: Noctuidae) varied by feeding on two cruciferous vegetables. *Scientific Reports 2022 12:1* **12**, 1–12 (2022).

62. Gouba, N., Raoult, D. & Drancourt, M. Plant and Fungal Diversity in Gut Microbiota as Revealed by Molecular and Culture Investigations. *PLoS One* **8**, e59474 (2013).

63. Sharma, M. K. *et al.* Effect of Dietary Supplementation of Probiotic Aspergillus niger on Performance and Cecal Microbiota in Hy-Line W-36 Laying Hens. *Animals* **12**, (2022).

64. Zhang, X. ; *et al.* Diversity and Functional Roles of the Gut Microbiota in Lepidopteran Insects. *Microorganisms 2022, Vol. 10, Page 1234* **10**, 1234 (2022).

65. Zhang, L., Zhan, H., Xu, W., Yan, S. & Ng, S. C. The role of gut mycobiome in health and diseases. *Therap Adv Gastroenterol* **14**, (2021).

66. Ramírez-Camejo, L. A., Maldonado-Morales, G. & Bayman, P. Differential Microbial Diversity in Drosophila melanogaster: Are Fruit Flies Potential Vectors of Opportunistic Pathogens? *Int J Microbiol* **2017**, (2017).

67. Majumder, R., Sutcliffe, B., Taylor, P. W. & Chapman, T. A. Fruit host-dependent fungal communities in the microbiome of wild Queensland fruit fly larvae. *Scientific Reports 2020 10:1* **10**, 1–12 (2020).

68. Guerre, P. Mycotoxin and Gut Microbiota Interactions. *Toxins 2020, Vol. 12, Page 769* **12**, 769 (2020).

69. Sikorowski, P. P., Lawrence, A. M. & Inglis, G. D. Effects of Serratia marcescens on Rearing of the Tobacco Budworm (Lepidoptera: Noctuidae). *American Entomologist* **47**, 51–60 (2001).

70. Johnston, P. R. & Rolff, J. Host and Symbiont Jointly Control Gut Microbiota during Complete Metamorphosis. *PLoS Pathog* **11**, e1005246 (2015).

71. Wang, X. *et al.* Variability of Gut Microbiota Across the Life Cycle of Grapholita molesta (Lepidoptera: Tortricidae). *Front Microbiol* **11**, 1366 (2020).

72. Xia, X. *et al.* Gut microbiota mediate insecticide resistance in the diamondback moth, Plutella xylostella (L.). *Front Microbiol* **9**, 25 (2018).

73. Garcia, C. J. *et al.* Serralysin family metalloproteases protects Serratia marcescens from predation by the predatory bacteria Micavibrio aeruginosavorus. *Scientific Reports 2018 8:1* **8**, 1–11 (2018).

74. Jadhav, S. B., Shah, N., Rathi, A., Rathi, V. & Rathi, A. Serratiopeptidase: Insights into the therapeutic applications. *Biotechnology Reports* **28**, e00544 (2020).

75. Mason, C. J., Peiffer, M., Felton, G. W. & Hoover, K. Host-Specific larval lepidopteran mortality to pathogenic Serratia mediated by poor diet. *J Invertebr Pathol* **194**, 107818 (2022).

76. Liang, Y. *et al.* Midgut membrane protein BmSUH facilitates Bombyx mori nucleopolyhedrovirus oral infection. *PLoS Pathog* **18**, e1010938 (2022).

77. Joynson, R., Pritchard, L., Osemwekha, E. & Ferry, N. Metagenomic analysis of the gut microbiome of the common black slug arion ater in search of novel lignocellulose degrading enzymes. *Front Microbiol* **8**, 2181 (2017).

78. Kandel Gambarte, P. C. & Wolansky, M. J. The gut microbiota as a biomarker for realistic exposures to pesticides: A critical consideration. *Neurotoxicol Teratol* **91**, 107074 (2022).

Chapter 4

Comparative analyses of lepidoptera larva metagenome

CHAPTER 4

Comparative analyses of lepidoptera larva metagenome

ABSTRACT

Insects, representing one of the most diverse groups of organisms, harbor a complex gut microbiome that plays a crucial role in aiding their adaptation to extreme and diverse environmental conditions, driven by their wide range of food sources.

This study presents a comparative analysis of the metagenomes of three different lepidoptera larva species: *Samia cynthia ricini* (Family: Saturniidae), *Bombyx mori* (Family: Bombycidae), and *Plutella xylostella* (Family: Plutellidae). The primary aim was to investigate the structural and functional diversity of the microbiota characterizing each organism and identify the role of their distinct environments in shaping microbial interactions.

Our analysis revealed intriguing differences among the three lepidoptera species. *S. cynthia ricini* exhibited the highest diversity of microbial species, highlighting its unique gut microbiome composition. On the other hand, *B. mori* displayed a higher abundance of gut microbiota compared to the other two species, indicating a potentially more enriched microbial community.

Distinct microbial compositions were observed in *P. xylostella*, setting it apart from both *S. cynthia ricini* and *B. mori*. Notably, the presence of specific microbes such as

Klebsiella, Pantoea, and various microsporidia distinguished *P. xylostella*'s gut microbiota.

Functional analysis of the metagenomes revealed contrasting patterns in the abundance of microbial genes associated with different pathways. *S. cynthia ricini* demonstrated a higher abundance of genes related to digestion, nutrition assistance, and immunity, likely contributing to its adaptability to diverse environments.

In contrast, *P. xylostella* exhibited a higher abundance of genes involved in cell envelope, cellular processes, signalling, and membrane transport pathways, hinting at potential adaptations that contribute to its survival and interaction with its unique ecological niche.

Overall, our study provides valuable insights into the diverse structural and functional aspects of gut microbiota in these lepidoptera larvae. By uncovering the distinct microbial compositions and functional attributes, this research advances our understanding of the intricate interplay between insects and their gut microbiota in response to their respective environments.

INTRODUCTION

Based on their diversity and composition, the gut microbes have different level of effects on important metabolic functions of the host such as degradation of many metabolites, manipulated reproduction, nitrogen recycle, production of many pigments and other aspects favouring the health and fitness of the insect[1–5]. Many gut microbes have convergently co-evolved with their insect host, gradually reduced their genome size, undergoing co-speciation and developing complementary functions with each other[6–9].

Lepidoptera is a highly diverse, distinctively recognised and widely studied family represented by many ecologically and economically important moths and butterflies[10]. In spite the biological studies, the importance of its host-associated microbes in the organisms were only recently appreciated more focused on heritable endosymbionts due to their sex determination system. Unlike other arthropods, female lepidopterans are heterogametic (ZW) and male (ZZ). Heritable endosymbionts such as Wolbachia were reported to be associated with this manipulation through cytoplasmic maternal inheritance[11]. These endosymbionts were found to be highly prevalent in lepidopterans with higher chances of vertical transmission. These differences along with different populations of other microbial community than other arthropods were assumed to be playing major role in the alternative sex determination system[12]. This were further evaluated for implications of other host genetic diversity and wider co-inherited microbial network responsible[7].

Presently, out of ~11300 species of lepidoptera found in India, bacterial associates of <0.1% species have been studied implying a huge amount of information and resources to be explored[10,13].

Plutella xylostella (Diamondback moth) (Family: Plutellidae) (**Table4.1**) larvae are phytophagous lepidoptera, with their dominant host plant being glucosine containing plants of the Brassicaceae family. It is widely considered as a pest regarding its destructive nature towards its food sources[14].

B. mori (Domestic silkworm) (Family: Bombycidae) (**Table4.1**) are polyphagous lepidopterans as well, feeding primarily on mulberry leaves. They are one of the economically important being the primary producer of silk. Some phenotypic traits of domestic silk moths are different from most members in the genus *Bombyx*, e.g., lost flying abilities and colour pigments[15,16].

Samia cynthia ricini (Eri silkworm) (Family: Saturniidae) (**Table4.1**), the organism of interest in this study is another polyphagous lepidoptera mainly feeding on castor leaves. They are comparatively, larger than the other two organisms with a high feeding and growth rate. Despite being domesticated, they have retained their flying abilities with a large wingspan. They have a high degree of breeding rate and are mostly resilient to environmental conditions[17,18].

With the aim of understanding the phenotypic and metabolic differences and the surrounding environmental conditions, a comparative study of the gut microbiota of the respective microbiota was performed. The role of these microbes in the host metabolism and adaptation were analyzed using shotgun metagenomics.

In our previous chapter, we have established the design and optimization of a shotgun metagenomics analysis workflow for maximum species identification concerning our study. In this chapter, we have utilized the metagenome data of *S. cynthia ricini* analyzed in the previous chapter and adopted the workflow established previously for a comparative analysis with other curated lepidopteran metagenomic datasets from public repositories.

S.ricini	*B. mori*	*P.xylostella*
Kingdom: Animalia	**Kingdom:** Animalia	**Kingdom:** Animalia
Phylum: Arthropoda	**Phylum:** Arthropoda	**Phylum:** Arthropoda
Subphylum: Hexapoda	**Subphylum:** Hexapoda	**Subphylum:** Hexapoda
Class: Insecta	**Class:** Insecta	**Class:** Insecta
Infraclass: Neoptera	**Infraclass:** Neoptera	**Infraclass:** Neoptera
Subclass: Pterygota	**Subclass:** Pterygota	**Subclass:** Pterygota
Order: Lepidoptera	**Order:** Lepidoptera	**Order:** Lepidoptera
Superfamily: Bombycoidea	**Superfamily:** Bombycoidea	**Superfamily:** Yponomeutoidea
Family: Saturniidae	**Family:** Bombycidae	**Family:** Plutellidae
Subfamily: Saturniinae Boisduval	**Subfamily:** Bombycinae	**Subfamily:** Plutellinae
Genus: Samia	**Genus:** Bombyx	**Genus:** Plutella
Subject: Samia cynthia (Linnaeus)	**Subject:** Bombyx Linnaeus (Linnaeus)	**Subject:** Plutella xylostella (Linnaeus)

Table4.1: Comparative taxonomic hierarchy of the organisms for the study: *S. cynthia ricini, B. mori* and *P. xylostella*

Significances of the study include-

- Observation of a common pattern of taxonomic or functional abundances of the lepidopteran species.
- Analysing the significant differences of the microbiota associated with their phenotype, metabolic activities and/or behaviour (pests, etc.).

METHODOLOGY

Sample collection and DNA extraction

DNA extraction was carried out using phenol/chloroform (PHEC) extraction methodology[19]. Cells were homogenized using (mention the specification and condition) in sterile phosphate-buffer solution (PBS) followed by SDS-based cell-membrane degradation. The extracts were then saturated with phenol (write condition) followed by separation of nucleic acids and proteins using chloroform: isoamyl-alcohol (24:1). Overnight precipitation of nucleic acids using isopropanol and 3 M sodium chloride for maximum DNA recovery was carried out. DNA samples were analyzed for their concentration and purity using 1% agarose gel electrophoresis and Nanodrop spectrophotometer[20,21]. Extracted DNA samples (A260/A280 ~1.8) were sent for 16s rRNA amplicon sequencing using Illumina NGS MiSeq (AgriGenome Technologies Pvt. Ltd.).

Sequence retrieval, quality-control and processing

Plutella xylostella metagenome sequences (Illumina HiSeq 2000) were retrieved from NCBI-SRA (SRR1292478). *B. mori* P50 metagenome (Illumina HiSeq 2500) were retrieved from NCBI-SRA database (SRR6062243). Paired-end (PE) FASTQ files were extracted from the SRA-accessions using combination of prefetch + fasterq-dump (SRA-toolkit). The PE reads were assessed for quality using fastqc (v 0.11.9).

Paired-end raw sequencing data (250bp) in FASTQ format were analyzed for quality and adapters using FastQC(v_0.11.9)[22]. Following removal of non-biological contaminants, the quality-controlled reads.

The methods were adapted from *S. cynthia ricini* metagenome sequence processing. The filtered sequences were then assembled using SPADES v12.0 and annotated using prokka (NCBInr, tigrfam, pfam, etc.)[23].

Functional annotation was performed using PROKKA [24]. PROKKA uses unsupervised machine learning gene prediction algorithm, Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm) for finding gene and translation. The subsequent processes are automated via PROKKA including location of RNA regions and open reading frames (ORFs), translation of ORFs to proteins, homology searching of proteins via BLAST and HMMER) against public databases (NCBI-CDD, PFAM, TIGRFAM) and PROKKA custom databases for annotation.

Proteins/enzymes annotated are involved in various reactions for various metabolic activities depicted by different pathways. The pathways are predicted by considering the presence of the set of enzymes to be the indicator and thus mapping using a naïve/direct approach. However, this methodology leads to false positive, as many pathways require only few of the present proteins/enzymes. MinPath (Minimal set of Pathways) uses a parsimony approach for biological pathway reconstructions yielding a more conserved and appropriate functional diversity, eliminating excessive predicted pathways[24].

The annotated PROKKA outputs were filtered based on KEGG Ortholog (KO) IDs for a compatible input using in-house shell scripts and mapped through MinPath MEtaCyC databases.

Based on the higher diversity and abundance of the species across the domain observed through SCL_5E shotgun metagenome analysis, obtained assembly were analyzed for their taxonomic composition using MEGAN+DIAMOND based on CDS[25].

Normalization was carried out based on CPM and the differential abundance based on z-score were represented through heatmap using MEGAN and the differences were explored[26].

The identified taxonomies were combined and used to generate a common tree using NCBI tree generator at the genus level. the comparative composition, diversity[27]. The annotation files with the absolute counts of the respective abundances were used to explore phylogenetic relationship and relative abundances were explored via dot plots using iTOLv3[28]. Species present in *S. cynthia ricini* were taken as a cut-off for a better view of the tree at the genus level.

Mapping reads and quantifying genes

Annotation output provides with number of genes and their annotations of a particular sample from the assembled reads. The information on the quantitative details is then achieved by mapping the raw reads to the assembled reads using Bowtie2 (v 2.5.1)[29]. The output alignment BAM file is then processed with Picard for duplicate reads removal[30].

Following that, coverage of each annotated gene is obtained. This is obtained by creating a custom Gene Transfer File (GTF) with the information of the gene regions and structure using in-house shell script thus defining the regions. Then, using htseq2-count (v 2.0.2) the number of reads mapped to each gene were counted sorted by alignment coordinate[31].

The average read length is ~150bp calculated from the raw reads. The gene lengths were extracted from the read count file using in-house shell script. TPM was calculated using a python script.

The obtained filtered KO and COG IDs from PROKKA annotated files and the normalised gene coverage value files were used to summarize and define the annotation. KRONA tools were used for the summarisation and plotting of interactive plots. The data was transformed to KRONA chart compatible hierarchal system using in-house python script[32]. The metabolic pathways based on KO, MetaCyc, COG and SEED-based databases were summarised. Comparative functional analyses were then observed based on different features. An overview of the methodology is represented below (**Figureure4.1**)

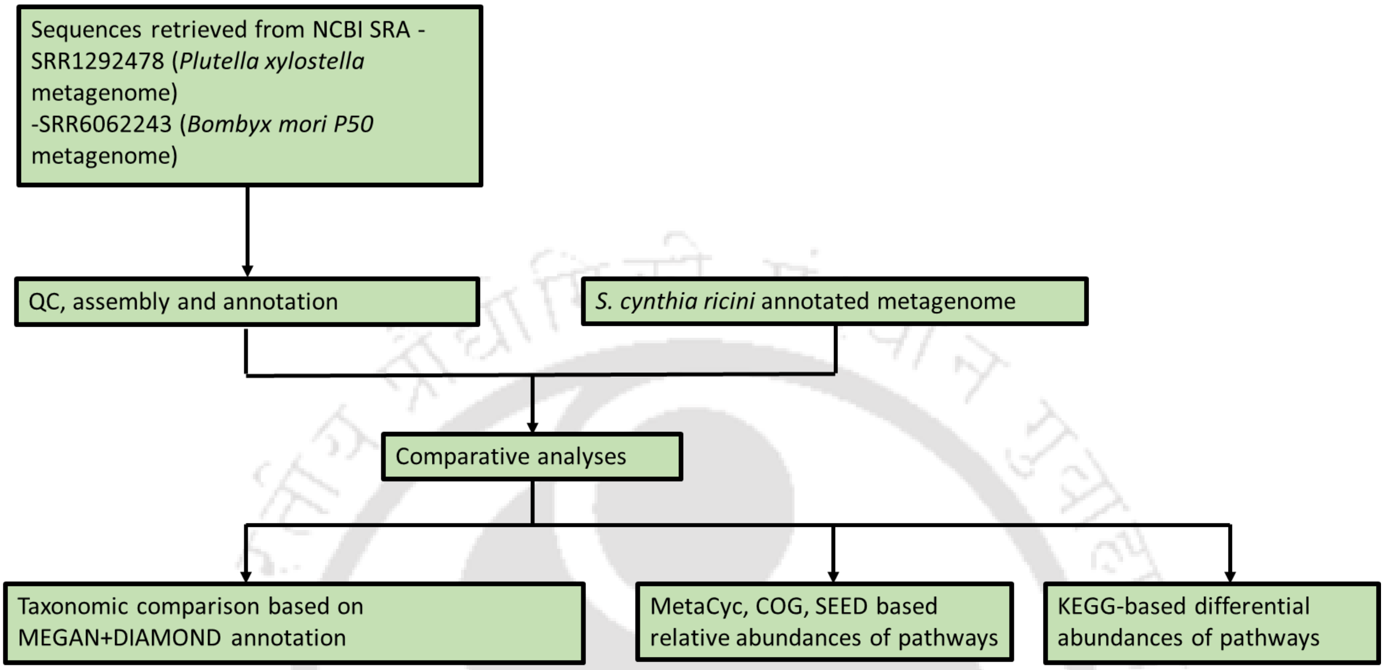


Figure4.1: Schematic representation of the methodology applied for comparative shotgun analyses of lepidopteran metagenome sample from *S. cynthia ricini*, *B. mori P50* and *P. xylostella*.

RESULTS AND DISCUSSION

The assembly statistics calculated using Quast were summarised as below (**Table4.2**)

	S. cynthia ricini	***B. mori***	***P. xylostella***
Total length	202665010	51423570	37579489
Contigs (>= 10000bp)	7	63	58
Contigs (>= 1000bp)	75489	5035	6123
Contigs	189501	27852	20959
Largest contig	19053	18177	44740
N50	1153	1214	1133

GC (%)	35.91	50.2	40.04

Table4.2: Summarised representation of comparative assembly statistics

The assembly statistics showed the number of contigs of more than 10000bp generated for *S. cynthia ricini* was 7. However, the number for *B. mori* was 63 and *P. xylostella* was 58, indicating a huge number of fragmented sequences for *S. cynthia ricini* data. N50 was 1153, 1214 and 1133 for *S. cynthia ricini*, *B. mori* (P50) and *P. xylostella* respectively.

Number of contigs generated were highest for *B. mori* (P50) followed by *P. xylostella*. The GC content was very low for *S. cynthia ricini* compared to the other samples.

Comparative taxonomic analyses

Exploring the differential abundance heatmap, significant differences in the taxonomy presence and abundances through the species metagenome samples were observed (**Figure4.3a,b,c,d**). *S. cynthia ricini* comprised of higher abundances of *Prevotella copri*, various species from Burkholderiales including Achromobacter xylosoxidans, different strains of Burkholderia, and strains of Ralstonia, Aeromonas caviae, Acinetobacter indicus, Pseudomonas aeruginosa, Pseudomonas stutzeri, Vibrio choleri, strains of Corynebacterium, Mycobacterium, Microbacterium saperdae, strains of Kocuria, Micrococcus, strains of Acidipropionibacterium, strains of Bacillus, Priestia megaterium, Weizmannia coagulans, Staphylococcus epidermidis compared to the other two species. Higher fungal species abundance such as strains of Aureobasidium, Aureobasidium, Sporisorium reilianum, strains of Ustilago, and viral strains including *B. mori* nucleopolyhedrovirus, Dengue virus, Sugarcane streak mosaic virus, Cauliflower mosaic virus were also observed. *B. mori* metagenome had higher abundance of numerous species of Bradyrhizobium, members of Alphaproteobacteria, Acidovorax avenae, strains of Comamonas, strains of Delftia, Variovorax paradoxus, numerous strains of Acinetobacter, Pseudomonas, Stenotrophomonas, Rhodococcus, Glutamicibacter, Lysinibacillus, Mammaliicoccus, and Enterococcus. *P. xylostella* comprised of populations of strains of Citrobacter, Enterobacter, Klebsiella, Lelliottia, fungal species *Astraeus odoratus* and various microsporidian species such as *Nosema, Enterocytozoon, Enterospora, Vittaforma corneae, Spraguea lophii, Encephalitozoon, Ordospora colligate, Hamiltosporidium magnivora, Trachipleistophora hominis, Vavraia culicis, Thelohania contejeani, Anncaliia algerae, Tubulinosema ratisbonensis, Rhizopus arrhizus and Thelohanellus kitauei*. Microbial populations of *S. cynthia ricini* were observed to be more diverse comprising of abundant species across the domains compared to the other samples. It was observed to contain populations of fungal species and viral species in abundance and a wide variety of bacterial populations. *B. mori* metagenome was enriched with similar groups of bacterial population. Many bacterial species including Pseudomonas, Bacillus, etc. were observed to more abundant in *S. cynthia ricini* and *B. mori*. *P. xylostella* consisted of groups of Enterobacter, E. coli, Klebsiella, Pantoea, Yersinia in higher abundance. Population of Serratia in *P. xylostella* was very high compared to *B. mori* and S.ricini (lowest). Microsporidia population was abundantly found in *P. xylostella*.

Prevotella sp., an anaerobic group of bacteria, mostly reported to be uncultivable, are a major player in non-cellulolytic carbohydrate degradation, responsible for aiding the host in digestion of cell wall polysaccharides such as xylan[33]. Previous reports have demonstrated the higher abundance of Proteobacteria in pest insect gut such as *Schistocerca gregaria* and associated them to be involved in the defensive mechanisms of the host providing resistance against plant pathogens and toxins[34]. Microsporidia have been associated with insect pathogenicity and weaken their immune system. Species including Enterococcus were isolated from *P. xylostella* and observed to have a positive effect on resistance against chlorpyriphos[35]. However, Serratia from the same sample were seen to have a reduced effect on the resistance against the pesticide[35]. Previous studies reported the dominant presence of Gammaproteobacteria in *P. xylostella* gut[14]. In this study, it is observed that Gammaproteobacteria population were significantly higher in P. xylostella compared to the other two lepidopterans.

Phylogenetic representation visualised using common tree and absolute counts showed higher counts of most bacterial population in *B. mori*. *S. cynthia ricini* had lower counts of microbes but a more diverse population. *P. xylostella* had a sparse diversity of microbiome population compared to *S. cynthia ricini* and *B. mori* with a lower count of most species. Few digestion-aiding bacterial population including several species of Enterobacter, Klebsiella, etc were more abundant in *P. xylostella* **(Figure4.4)**.

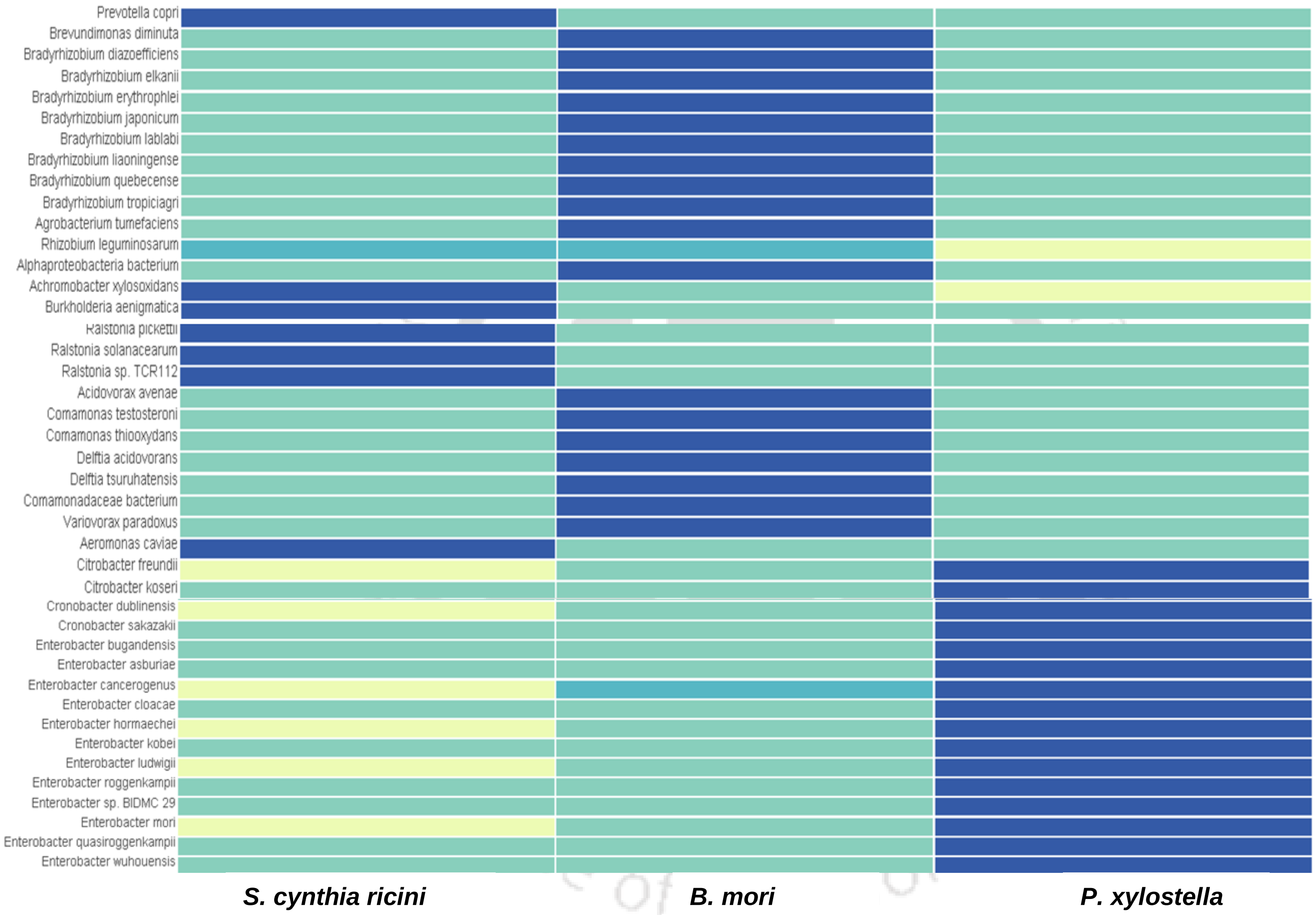


Figure4.3a: Heatmap representation of the differential abundances of identified taxonomies through the samples.

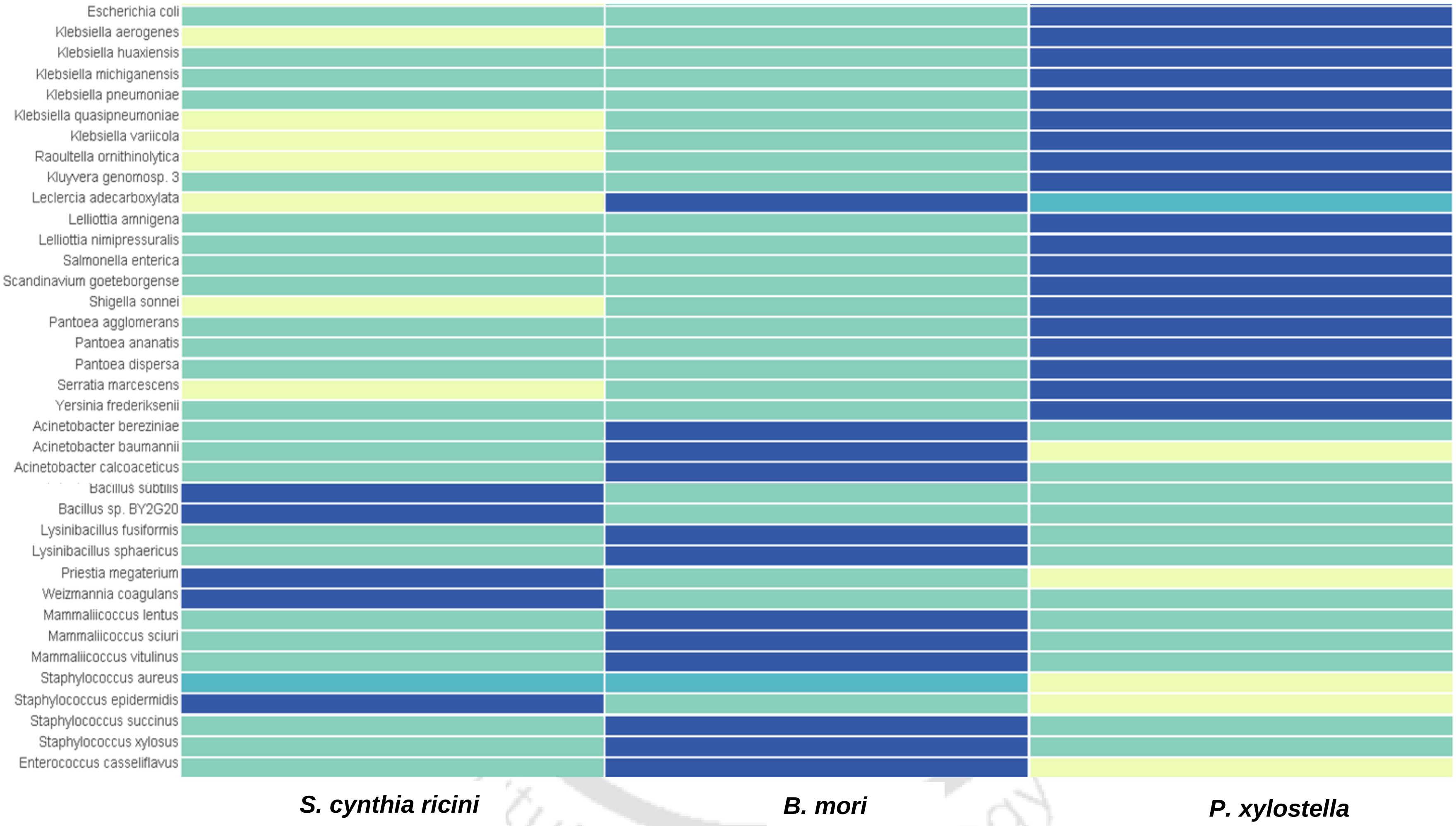


Figure4.3b: Heatmap representation of the differential abundances of identified taxonomies through the samples.

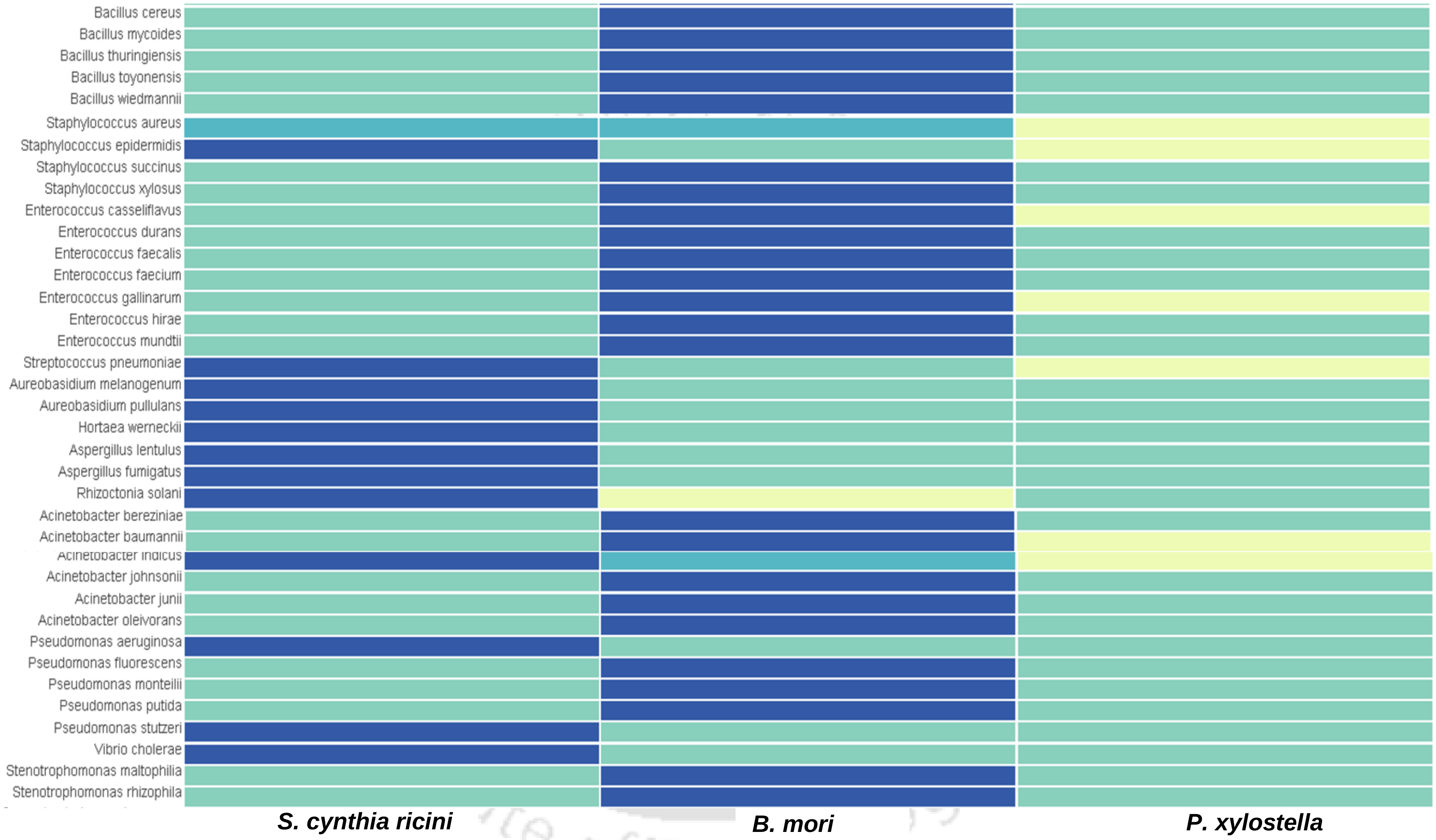


Figure4.3c: Heatmap representation of the differential abundances of identified taxonomies through the samples.

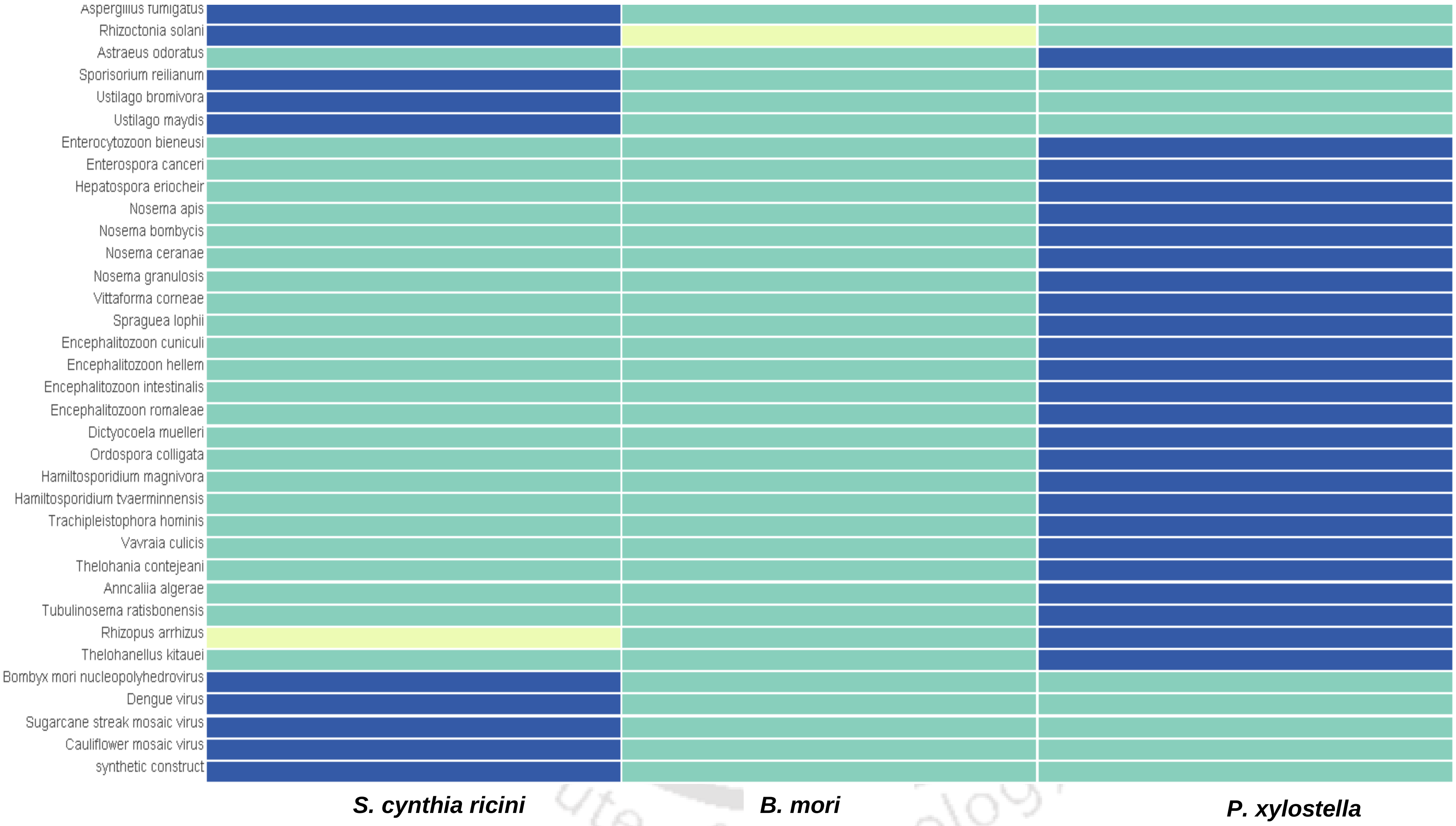


Figure4.3d: Heatmap representation of the differential abundances of identified taxonomies through the samples.

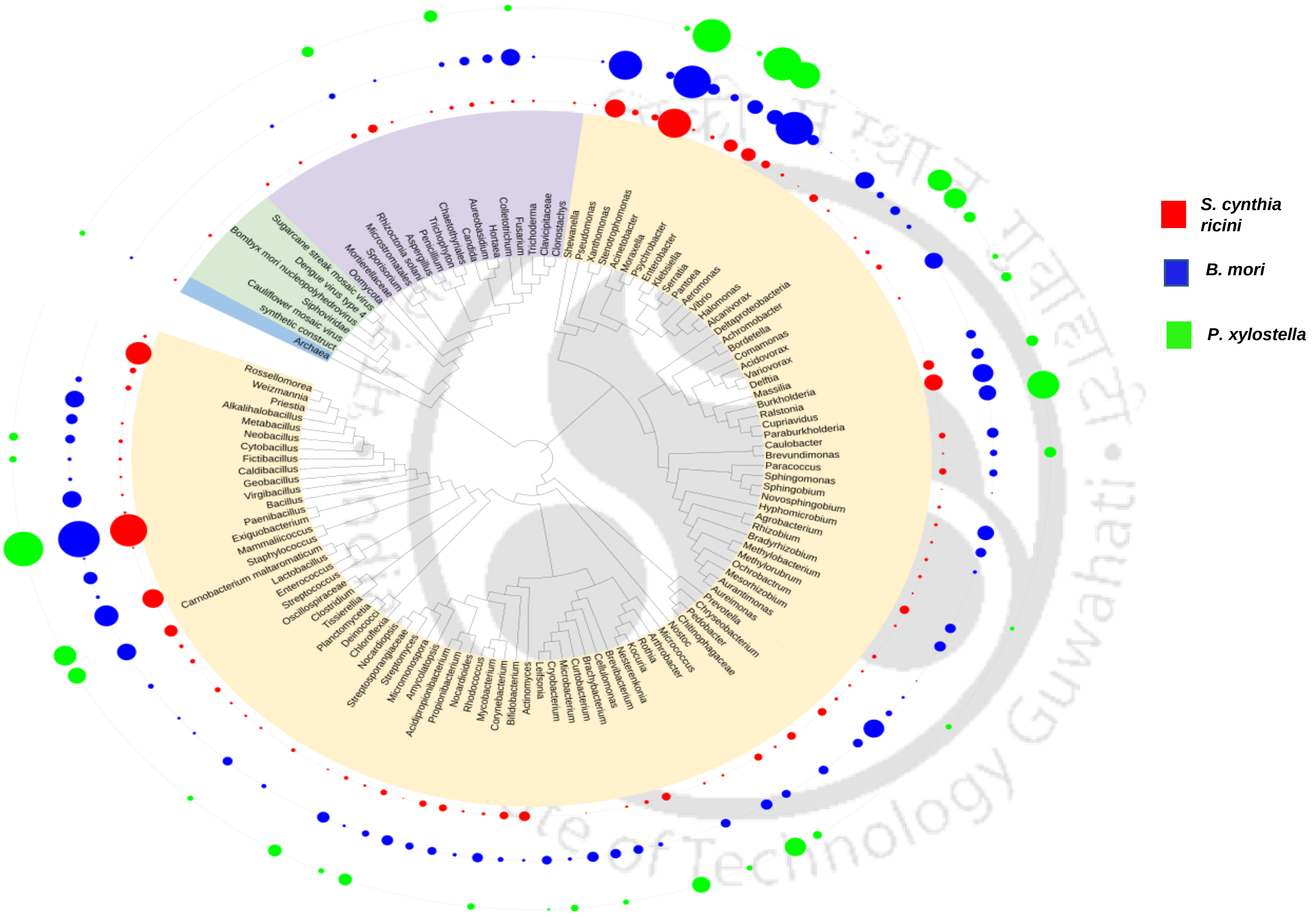


Figure4.4: Phylogenetic tree representation of comparative taxonomic composition and abundances of the metagenomes of *S. cynthia ricini*, *B. mori* and *P. xylostella*.

Comparative functional analyses

Relative functional abundance based on reconstructed biological pathways showed a difference in the abundance of genes involved in metabolic pathways based on MetaCyc database. 60% and 59% of *B. mori* and *P. xylostella* genes accounted for functions involved in biosynthesis pathways respectively (**Figure4.5a,b,c**). However, *S. cynthia ricini* consisted of 54% biosynthesis pathways related genes. On the other hand, degradation pathways related genes comprised of 28% of *B. mori*, 24% of *P. xylostella* in identified metagenome genes. Energy-metabolism related genes comprised of 4% in *S. cynthia ricini* and *B. mori* each while 2% in *P. xylostella*. 2% comprised of fermentation contributing genes in all the samples. Secondary metabolite biosynthesis related genes were lesser in *S. cynthia ricini* (38% of biosynthesis) compared to those in *B. mori* (53% of biosynthesis) and *P. xylostella* (47% of biosynthesis) while secondary metabolite degradation related genes were higher in *P. xylostella* (14% of degradation) than *S. cynthia ricini* and *B. mori* (11% of degradation each). Co-factor and vitamin biosynthesis related genes were higher in *S. cynthia ricini* (15% of biosynthesis) and *P. xylostella* (11% of biosynthesis) compared to (8% of biosynthesis).

Higher degradation-based pathways in *S. cynthia ricini* could infer to its higher capability of leaf ingestion and digestion along with its body-size and fat composition compared to the other species. Its higher abundance of secondary pathways related genes infers the higher amount of plant compound degradation involved. Higher abundance of co-factor and vitamins related genes inferred the involvement of the microbiota of *S. cynthia ricini* to be more active in the nutrition of the host along with the maintenance of its metabolic homeostasis.

Observing relative abundance of COG-based pathways, metabolic pathways involved genes were higher (57%) in the microbiome of *S. cynthia ricini* compared to *B. mori* (51%) and *P. xylostella* (43%) (**Figure4.6**). Cellular processing and signalling mechanisms related genes in *P. xylostella* were the highest (27%) in comparison to *B. mori* (24%) and *S. cynthia ricini* (20%). Information storage and processing mechanisms related genes were lower in *S. cynthia ricini* (21%) compared to the other species (24%, each). SEED-based pathway comparison resulted in higher abundance of energy-related metabolic pathways to be highly abundant in *S. cynthia ricini* genes (16%) (**Figure4.6**). While in *B. mori* and *P. xylostella*, their relative abundances were very low (9% each). Protein processing gene involved pathways were also higher in *S. cynthia ricini* (12%) than *B. mori* (9%) and *P. xylostella* (8%). Genes of pathways related to cell envelope, membrane transport and cell signalling were observed to be very high in *P. xylostella* (4%, 13% and 1.78% respectively). They comprised of 1%, 7% and 0.7% respectively in *S. cynthia ricini* and 2%, 10% and 1% respectively in *B. mori*. Other metabolism pathway-based genes were relatively higher in *S. cynthia ricini* and *B. mori* (40% each) in comparison to *P. xylostella* (34%). Abundance of other pathways were relatively similar.

Based on the COG and SEED-based pathway abundances, metabolism related genes of *S. cynthia ricini* were higher than that of the others and most of them were directed to energy and protein processing related pathways indicating the involvement of the microbiota with the metabolic activities of the host directly or indirectly either by producing host related metabolites or assisting them with the production of required metabolites. It is also inferred that they directly provide energy to the host assisting in a higher capability of degradation pathways. Cellular processing and signalling mechanism pathways including cell envelope related pathways in *P. xylostella* is very high compared to *S. cynthia ricini* and *B. mori*. *P. xylostella*, being a pest organism are highly destructive to mostly cruciferous plants. The higher abundance of these microbial pathway genes in the organism is indicative of their assistance of processing of environmental signal processing of the pest insect and also induced stress tolerance. Further, membrane transport pathways-based genes were higher in *P. xylostella*. Observing the taxonomic abundance, various microsporidia populations were abundantly present in the organism. Several microsporidia species such as Nosema, Sprague, Serratia were reported to have reduced genome with minimal energy requirements and rely on the host for nutrient uptakes via various membrane transport systems and being pathogenic to the host[36,37]. Microsporidia have been reported to act as natural pest population regulating agents, maintaining the fecundity of the pest host[38]. Dominance of these microbes were reported in previous intraspecies-based studies as well[39,40]. Higher abundance of the pathway genes with respect to other organism imply their significant role in the pest adaptation.

Differential abundance was observed using KEGG-based pathways (**Figure4.8**). Most genes of the metabolic pathways including fatty acid biosynthesis, lipoic acid metabolism, folate acid biosynthesis, steroid hormone biosynthesis, pentose phosphate pathway, thiamine, glycerophospholipid metabolism were abundant in *B. mori* compared to the other samples. Comparatively, *P. xylostella* had the least count of genes of the different pathways. However, 2-component pathway related genes were highly abundant in *P. xylostella*. Genes of pathways including metabolism of xenobiotics, porphyrin and chlorophyll metabolism, fatty acid degradation, starch and sucrose metabolism, amino acid such as glycine, serine, threonine, and tryptophan metabolism, purine and pyrimidine metabolism, biosynthesis of secondary metabolites, base excision repair were found comparatively abundant in *S. cynthia* *ricini. B. mori* was seen to have higher counts of reads for most pathways related genes, however, the comparative counts for various digestion aiding and nutrition acquisition pathways were higher in *S. cynthia ricini* indicating their higher growth and survival rate and their digestive abilities. The higher presence of amino acid and xenobiotics metabolism genes were also indicative of their immunity and resistance to pathogens compared to *B. mori*.

Based on these analyses, gut microbiota of *S. cynthia ricini* and *B. mori* were found to be more similar. However, more similarities in the taxonomic and functional composition of *P. xylostella* and *S. cynthia ricini* involved in defence mechanisms and degrading pathways accounting for their higher growth rate and immunity.

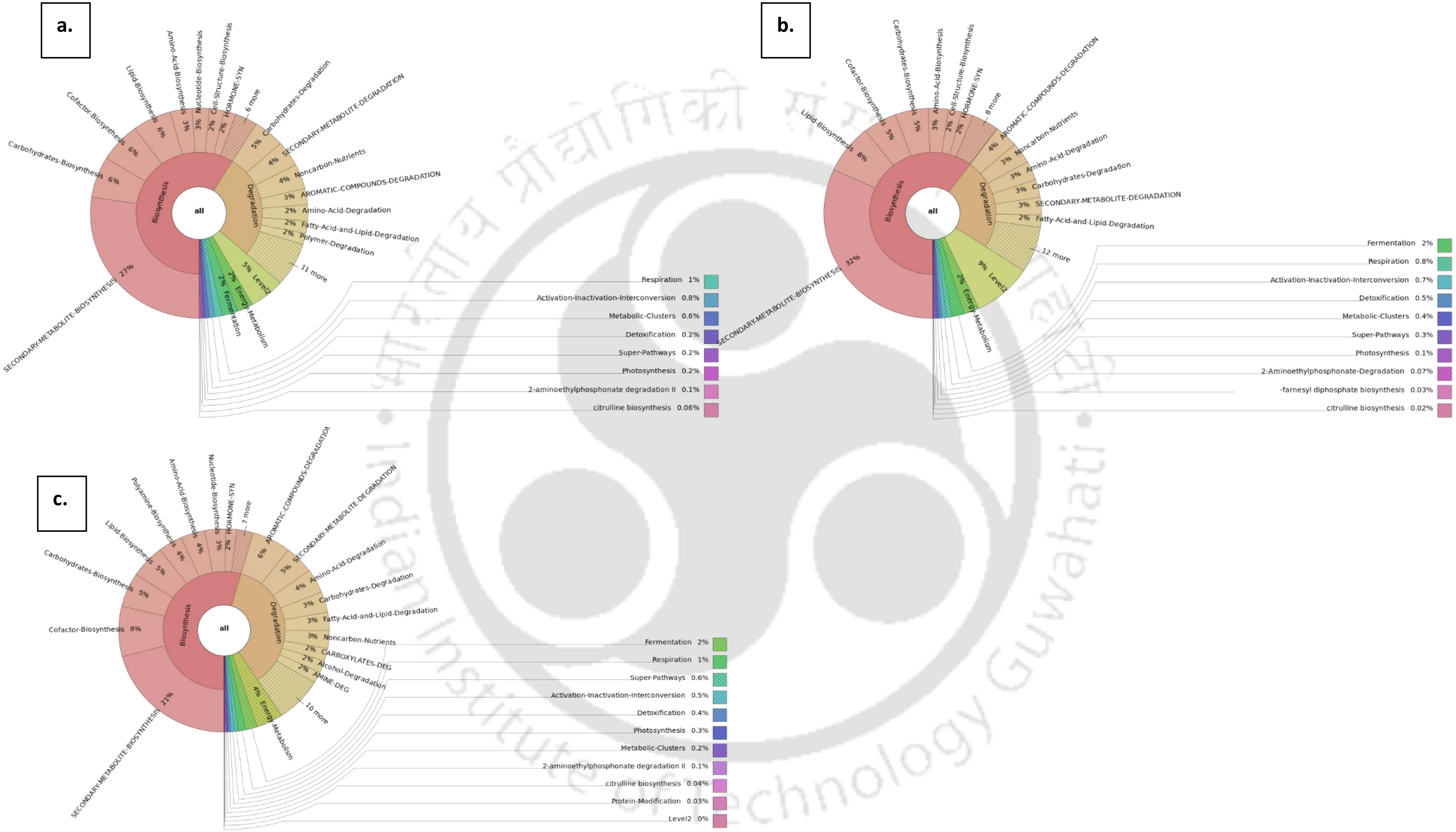


Figure4.5a,b,c: Radial chart representation of MetaCyc Level 2-based relative abundance of various metabolic pathways: a. *P. xylostella* b. *B.mori* c. *S. cynthia ricini*

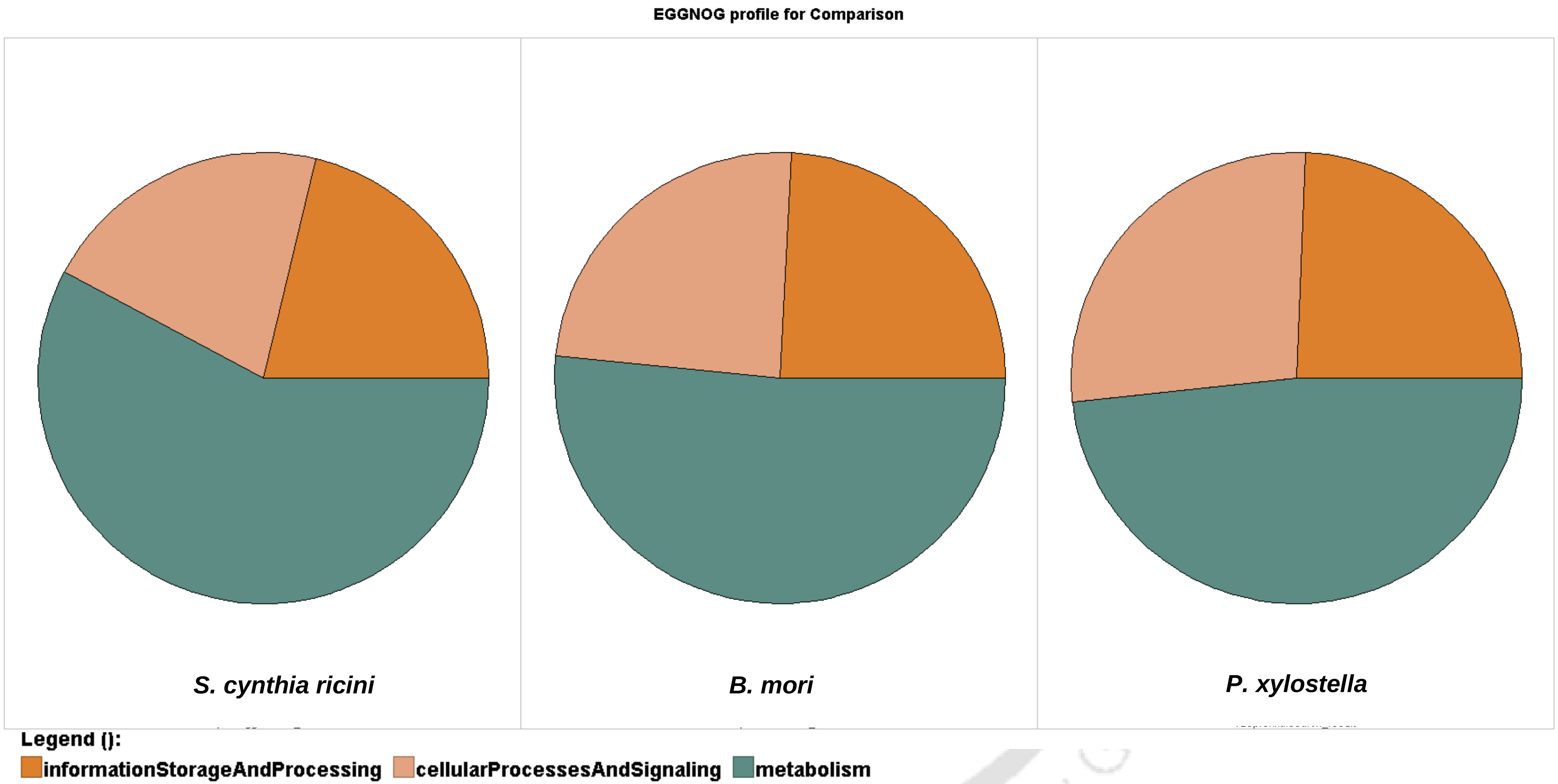


Figure4.6: COG- based comparison of the relative abundance of genes involved in various pathways *of the samples S. cynthia ricini, P. xylostella* and *B. mori*

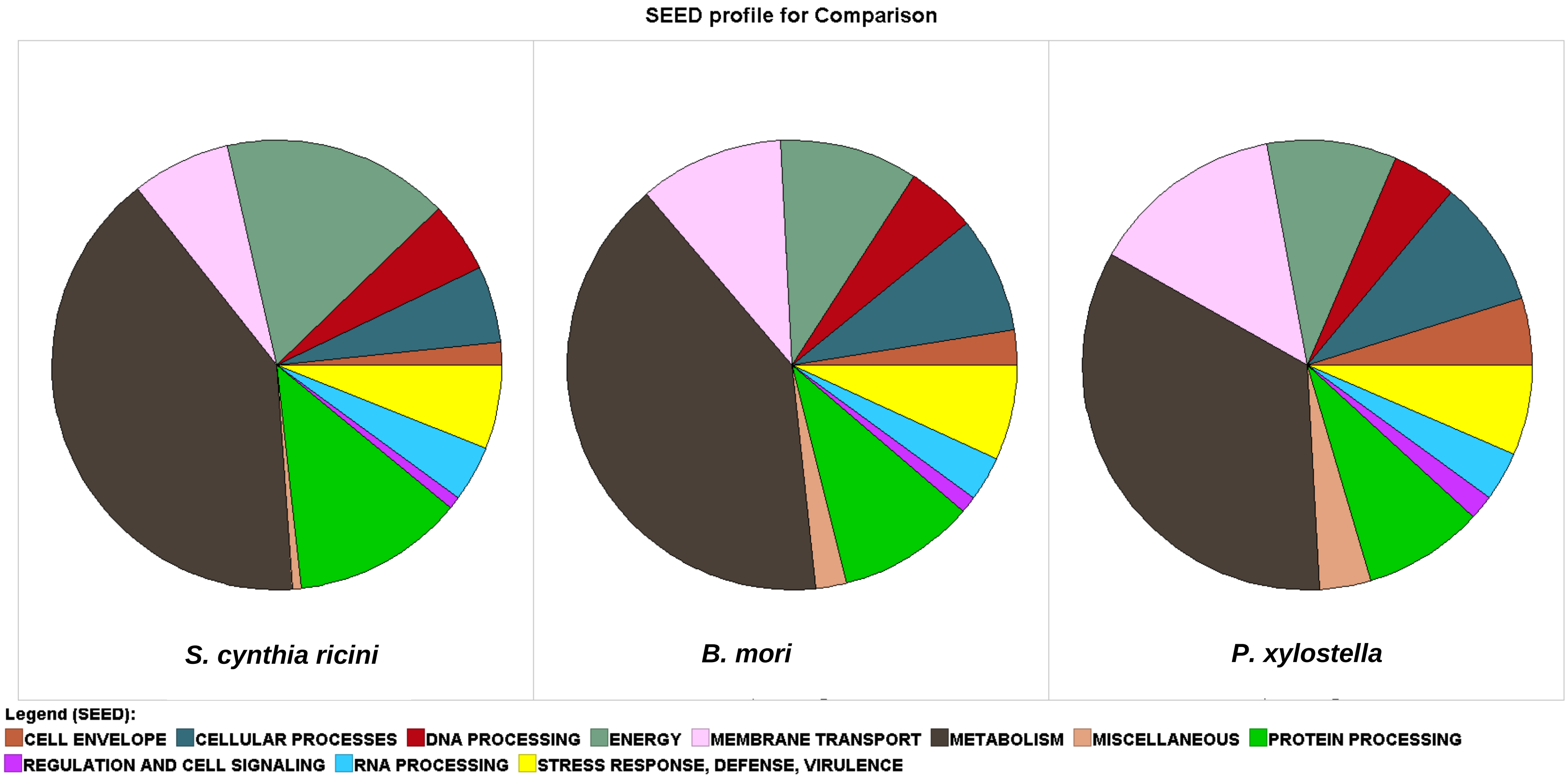


Figure4.7: SEED-based comparison of the relative abundance of genes involved in various pathways of the samples *S. cynthia ricini, P. xylostella* and *B. mori*

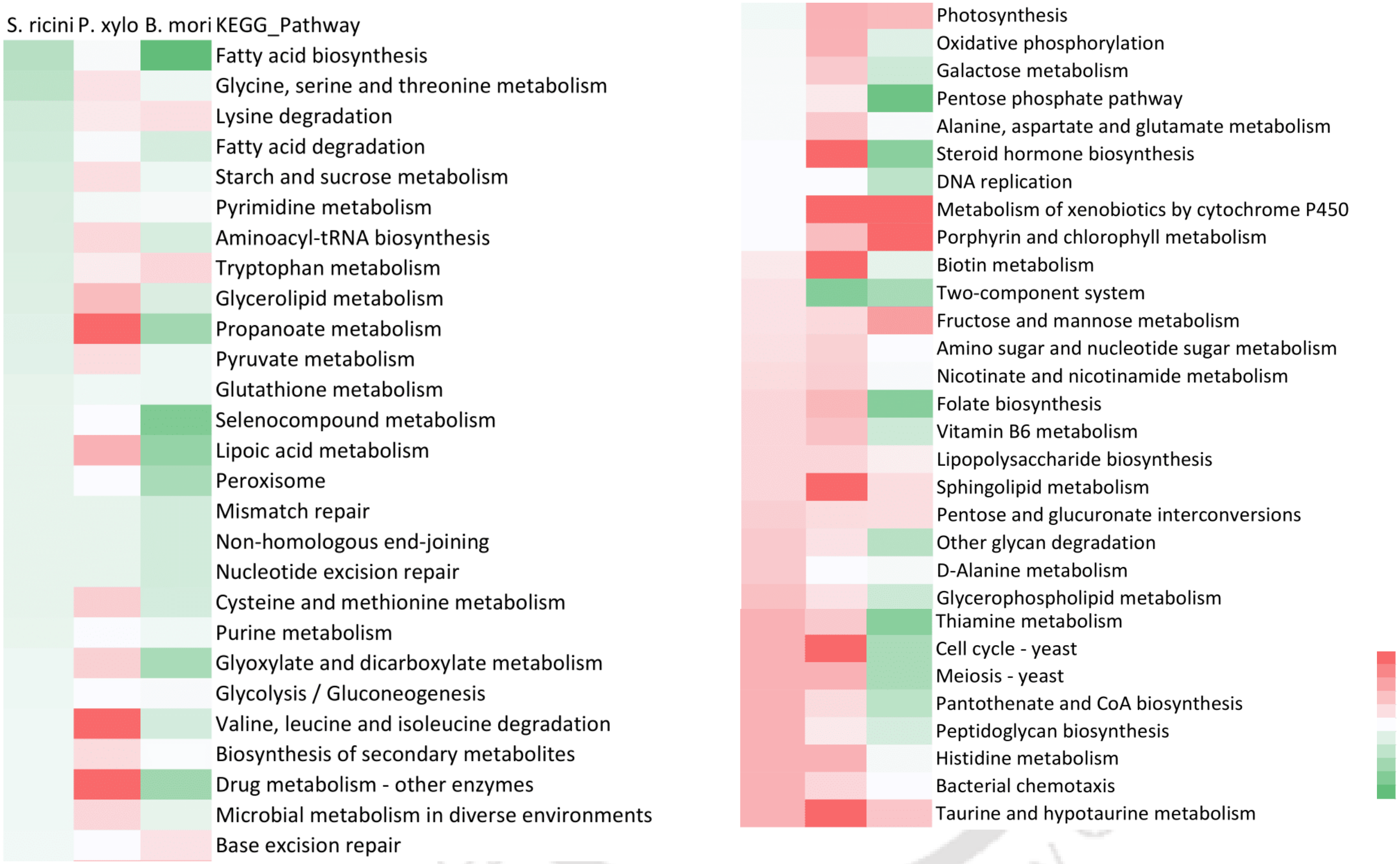


Figure4.8: Heatmap representation of differential abundance of most significant KEGG-based pathways of the samples *S. cynthia ricini, P. xylostella* and *B. mori*.

CONCLUSION

Lepidoptera being one of the largest groups if insects, having a wide range of diversity and ecological significance, this study attempts to understand the microbial diversity associated with it. Differences in the taxonomic and functional abundance among the species were summarised with the aim to understand their role in their ecological and phenotypic variances. *S. cynthia ricini* metagenome showed a higher abundance of degradation-based pathways, secondary pathways and co-factor and vitamin related genes inferring the involvement of the microbiota of *S. cynthia ricini* to be more active in the digestive capabilities and nutrition of the host along with the maintenance of its metabolic homeostasis. Microbial populations of *S. cynthia ricini* were observed to be more diverse comprising of abundant species across the domains compared to the other samples. Microbiome of *B. mori* was seen to be more similar to *S. cynthia ricini* than *P. xylostella. P. xylostella* microbiota was seen to consist of a higher abundance of species of Proteobacteria, microsporidia which are associated with defensive mechanisms of the host providing resistance against plant pathogens and toxins revealing the involvement of its microbiota in its pest nature.

This study, attempts to reduce the research gap in the familiarity of lepidoptera microbiome allowing for a better measure for their growth, development and further research. The pest related microbiota was studied which could assist in their better regulation.

Further analyses could be done with respect to the plant hosts, the lepidoptera insects and pests, the microbes and the surrounding environment in an integrative manner using system biology techniques. Understanding the microbial signatures in more in-depth integrative and validation studies based on the hypothesis generated could also help predict disease patterns.

REFERENCES

1. Johnston, P. R. & Rolff, J. Host and Symbiont Jointly Control Gut Microbiota during Complete Metamorphosis. *PLoS Pathog* **11**, e1005246 (2015).

2. Zhao, M., Lin, X. & Guo, X. The Role of Insect Symbiotic Bacteria in Metabolizing Phytochemicals and Agrochemicals. *Insects* **13**, (2022).

3. Yuning, L. *et al.* The Bacterial and Fungal Communities of The Larval Midgut of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Varied by Feeding on Two Cruciferous Vegetables. *Sci. Rep. 2022 12:1* **12**, 1–12 (2022).

4. Zhang, X. *et al.* Diversity and Functional Roles of the Gut Microbiota in Lepidopteran Insects. *Microorganisms 2022, Vol. 10, Page 1234* **10**, 1234 (2022).

5. Ottman, N., Smidt, H., de Vos, W. M. & Belzer, C. The Function of Our Microbiota: Who Is Out There and What Do They Do? *Front Cell Infect Microbiol* **2**, 104 (2012).

6. Groussin, M., Mazel, F. & Alm, E. J. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host Microbe* **28**, 12–22 (2020).

7. Gupta, A. & Nair, S. Dynamics of Insect–Microbiome Interaction Influence Host and Microbial Symbiont. *Front Microbiol* **11**, 1357 (2020).

8. Kikuchi, Y. *et al.* Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biol* **7**, 1–22 (2009).

9. Co-evolution and Co-speciation of Host-Gut Bacteria Systems. *Cell Host Microbe.* doi:10.1016/J.CHOM.2020.06.013.

10. Smetacek, P. Review of Indian Lepidoptera Collections and Their Significance In Conservation. *ENVIS Bulletin* **14**, 135–139 (2013).

11. Kageyama, D., Narita, S. & Watanabe, M. Insect Sex Determination Manipulated by Their Endosymbionts: Incidences, Mechanisms and Implications. *Insects* **3**, 161 (2012).

12. Kageyama, D. *et al.* Feminizing Wolbachia Endosymbiont Disrupts Maternal Sex Chromosome Inheritance in a Butterfly Species. *Evol Lett* **1**, 232 (2017).

13. Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M. & Fatouros, N. E. Bacterial Symbionts in Lepidoptera: Their Diversity, Transmission, and Impact on The Host. *Front Microbiol* **9**, (2018).

14. Huang, B., Shi, Z., Hou, Y. & Lapointe, S. Host Selection Behavior and the Fecundity of *Plutella xylostella* (Lepidoptera: Plutellidae) on Multiple Host Plants. *Jour. of Ins. Sci.* **14**, (2014).

15. Goldsmith, M. R., Shimada, T. & Abe, H. The Genetics and Genomics of The Silkworm, *Bombyx Mori*. *Annu Rev Entomol* **50**, 71–100 (2005).

16. Furdui, E. M. *et al.* Genetic Characterization of Bombyx mori (Lepidoptera: Bombycidae) Breeding and Hybrid Lines with Different Geographic Origins. *Journal of Insect Science* **14**, (2014).

17. *Samia cynthia*. https://tpittaway.tripod.com/silk/s_cyn.htm.

18. Renuka, G. & Shamitha, G. Studies on the Economic Traits of Eri silkworm, *Samia cynthia ricini*, in Relation to Seasonal Variations. *Int J Adv Res (Indore)* **2**, 315–322 (2014).

19. Sambrook, J. & Russell, D. W. Purification of Nucleic Acids by Extraction with Phenol:Chloroform. *Cold Spring Harb Protoc* **2006**, pdb.prot4455 (2006).

20. García-Alegría, A. M. *et al.* Quantification of DNA through the nanodrop spectrophotometer: Methodological validation using standard reference material and sprague dawley rat and human DNA. *Int J Anal Chem* **2020**, (2020).

21. Olson, N. D. & Morrow, J. B. DNA Extract Characterization Process for Microbial Detection Methods Development and Validation. *BMC Res Notes* **5**, 668 (2012).

22. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

23. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A New Versatile Metagenomic Assembler. *Genome Res* **27**, 824–834 (2017).

24. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

25. Bağcı, C., Patz, S. & Huson, D. H. DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences. *Curr Protoc* **1**, e59 (2021).

26. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN Analysis of Metagenomic Data. *Genome Res* **17**, 377 (2007).

27. Generate a Common Tree for a set of taxa. https://www.ncbi.nlm.nih.gov/guide/howto/gen-com-tree/.

28. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res* **49**, W293–W296 (2021).

29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods 2012 9:4* **9**, 357–359 (2012).

30. Zhao, Q. A Study on Optimizing MarkDuplicate in Genome Sequencing Pipeline. *ICBRA '18.* doi:10.1145/3309129.3309134.

31. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python Framework to Work with High-throughput Sequencing Data. *Bioinformatics* **31**, 166 (2015).

32. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive Metagenomic Visualization in a Web browser. *BMC Bioinformatics* **12**, 1–10 (2011).

33. Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P. & Forano, E. Microbial Degradation of Complex Carbohydrates in the Gut. *Gut Microbes* **3**, 289 (2012).

34. Dillon, R. & Charnley, K. Mutualism Between The Desert Locust Schistocerca Gregaria and Its Gut Microbiota. *Res Microbiol* **153**, 503–509 (2002).

35. Xia, X. *et al.* Gut Microbiota Mediate Insecticide Resistance in The Diamondback Moth, *Plutella Xylostella* (L.). *Front Microbiol* **9**, 25 (2018).

36. Peyretaillade, E. *et al.* Extreme Reduction and Compaction of Microsporidian Genomes. *Res Microbiol* **162**, 598–606 (2011).

37. Jespersen, N., Monrroy, L. & Barandun, J. Impact of Genome Reduction in Microsporidia. *Exp Suppl* **114**, 1–42 (2022).

38. Bjornson, S. & Oi, D. Microsporidia Biological Control Agents and Pathogens of Microsporidia Biological Control Agents and Pathogens of Beneficial Insects Beneficial Insects. (2014).

39. Indiragandhi, P. *et al.* Cultivable Bacteria Associated with Larval Gut of Prothiofos-Resistant, Prothiofos-Susceptible and Field-Caught Populations of Diamondback Moth, *Plutella Xylostella* and Their Potential for, Antagonism Towards Entomopathogenic Fungi and Host Insect Nutrition. *J Appl Microbiol* **103**, 2664–2675 (2007).

40. Xia, X. *et al.* Gut Microbiota Mediate Insecticide Resistance in The Diamondback Moth, Plutella Xylostella (L.). *Front Microbiol* **9**, 25 (2018).

Chapter 5

Metabolic profiling of *Samia cynthia ricini* based on different diets and analysis the of the associated microbiome.

Chapter 5

Metabolic profiling of *Samia cynthia ricini* gut based on different diets and analysis the of the associated microbiome

ABSTRACT

The dynamic relationship between gut microbiota and host metabolites plays a crucial role in understanding the ecology, physiogenesis, nutrient acquisition, and immunomodulation of organisms, particularly in response to changes in their surrounding environment, including their plant-based diets. Amplicon-based microbiome analysis often falls short in providing a complete assessment of microbial contributions to the host's metabolic activity, considering factors such as expression and non-linear associations. In contrast, metabolic phenotyping offers a more sensitive and in-depth approach to unraveling subtle differences in host-microbiota interactions.

In this study, we aimed to analyze the differential profile of gut metabolites in *Samia cynthia ricini* larvae, based on two diet sources, Eri and Kesseru leaves, while investigating the effects of associated gut microbiota. LC-MS/MS-based spectral profile analysis generated a total of 3553 features. The metabolite profiles exhibited a distinct separation based on diet category, consistent with the diversity clusters observed through amplicon microbiome analysis. Utilizing OPLS-DA analyses, we identified 56 differentially abundant metabolites, comprising plant, insect, and

microbial origins. Annotation with METLIN and PubChem databases characterized 33 differentially abundant metabolites.

In Eri-based gut samples, 17 metabolites were upregulated, while 16 metabolites were upregulated in Kesseru-based gut samples. Notably, a significant portion of the differentially abundant metabolites originated from fungal sources, suggesting a substantial role of the fungal microbiome in the larval gut.

Correlation analysis of amplicon sequenced microbiome and differentially abundant metabolites revealed a complex interaction between various microbes and metabolites, underscoring the intricate interplay between gut microbiota and host metabolism.

Overall, our findings provide valuable insights into the diet-dependent metabolic profiles in *S. cynthia ricini* larvae and shed light on the multifaceted interactions between gut microbiota and host metabolites. This study enhances our understanding of the ecological and physiological implications of gut microbial communities in the context of the host's dietary preferences.

INTRODUCTION

The intermediates and final products of metabolism define the phenotypic characters of an organism; responsible for growth, development and functioning of an organism[1]. Metabolites, primary and secondary involved with all chemical reactions in an organism, are also responsible for its interaction, adaptation and defence to various surrounding environmental conditions[2]. Based on different environmental factors, metabolites and metabolic activities differ among the populations of the organism[3].

Metabolome is the comprehensive set of molecules (<1.5 kDa) in a particular biological sample[4]. Analysis of the metabolome helps understand the role of different metabolites responsible for the proper functioning of the organism in a particular environment[5]. It gives direct information of the biochemical activity of the cells thus best representing the phenotype[6–8].

Lepidoptera larva have a very species-specific digestive tract maintained to extract nutrients from the plants tissues along with minimizing harmful effects of various secondary metabolites from plants[9]. Some of these specific traits include highly reduced chemical state, low oxygen tension and extremely alkaline environment[10,11]. These conditions along with the plant-based inhibitory compounds and the counter-reactive gut produced metabolites makes it an extreme environment for the microbiome present[9,12]. The insect also has to protect itself against the foliage-derived pathogenic microbes[9]. Studies of the lepidopteran larval gut microbiome has been seen to be very unstable and varied providing different theories throughout. Some studies on lepidoptera gut have reported a transient microbiome while other studies have reported a residential gut microbiome with minimal population[13,14]. Gypsy moth larva were reported to have 15 phylotypes as a core residential population[15]. Similar studies were seen in *Hepialus gonggaensis*[16]. Based on these microbiome, and different host plants, feeding habits, pathogens, etc. A variety of metabolic compounds and enzymes were isolated from saturniids, Rothschildia lebeau, Automeris zugana with potential functions relating to its digestive capability enhancement such as cellulase, chitinase, amylase, xylanase, etc[17].

Insects and plants have some co-occurring metabolites, mostly secondary metabolites assisting them through various functions despite the phylogenetic distance between them. Most of these compounds were based on quinolones, glycosides, terpenes structures. The co-occurrences could be due to evolutionary adaptation by insects to minimize the toxicity of plant toxins or by plants to interfere with the insect functions[18]. Plant produced, Methylthiophenols attract wasp pollinators by mimicking female wasp pheromones thus interfering with the insect metabolism[19]. They could also have developed independently having parallel functions. (E)-β-Farnesene is a stress induced hormone by aphids (alarm pheromone) and several plant species (damage-induced)[20,21]. Linamarin, produced by both plants and insects acts as anti-insect defence in both the groups[22,23]. The compounds could have different biosynthetic origin or might be involved with different pathways or intermediates or the final product[18].

Gut microbes can affect the host gut metabolite either from bacteria-derived metabolites reacting with host dietary substrates or by modifying the host molecules[24]. Rootworm larva were reported to take up benzoxazine glucoside from maize roots and activate it to 6-methoxy-2-benzoxazolinone (MBOA) when exposed to entomopathogenic bacteria from nematode *Heterorhabditis bacteriophora* thus reporting the copying of two-component defence system[25].

Metabolome is the direct indicator of the state of an organism determining the homeostatic condition for that organism with respect to the environment[26]. Understanding the metabolic profile and their variation corresponding to the environment helps establish the signature metabolites that are related to the changes in the metabolic route activities and is a method of related pathway analysis[27].

The microbiome associated metabolic profile is also dependant on the surrounding environmental factors such as diet, stress, xenobiotics activities, etc. which helps interpret the complex interaction of the host, microbiome and the environment[28]. It provides information on the different host and microbial metabolite functions and their dependence on each other and also their dependence to the environment and the alterations along with it[29]. Gut microbes and their metabolites have important role in maintaining host metabolism and metabolites both directly and indirectly and changes in any of the factor could cause dysbiosis[24].

Thus, metabolomics increases our knowledge of the microbiome more deeply and their direct role in the host metabolism transforming nutrients, toxins and other abiotic factors affecting the host. Microbes and microbial communities play a major role in critical biogeochemical cycles and can assist in developing predictive biomarkers for various environmental stressors[30,31]. Microbiomes are also considered as biological reactors which can transform the surrounding available resources to products that might be either beneficial or harmful to the particular environment, e.g., their role in bioremediation[32]. Changes in the gut microbiome also results in change in other factors of the host environment[33,34]. Metabolome can also describe the interaction and communication between the microbial population themselves such as quorum sensing and analyze the signalling processes involved[35–37].

Analysing the metabolic profile and its change under different conditions could provide insights into their role in contributing to the difference in the phenotype based on the particular environment, e.g., the metabolites responsible for a desired phenotype, diseased state, effects of a toxin or a drug, etc. and help attain the optimum personalised measures for the efficient growth of the organism with better potential. It helps predict functions of unknown genes based on phenotypic observations using gene manipulation techniques such as deletion or insertion[31,35,38,39].

The microbiome-metabolome based information are complementary to the other omics-based data and helps us understand the internal and external interactions of the microbial communities along with the description of the biological systems they are associated with[34]. This intensive analysis of the mechanisms of communication mechanisms within and surrounding the microbial communities could revolutionise the aspects of disease control, environmental conservation, agricultural development, optimising the strategies[32,33].

Unlike metagenomics and other omics-based data, metabolomics doesn’t rely on sequencing techniques. The mechanism involved in metabolomics include complementary chromatographic techniques i.e., liquid chromatography (LC), gas chromatography (GC) along with detection techniques i.e., nuclear magnetic resonance (NMR) and mass spectrometry (MS)[40,41]. Identification and quantification of the metabolites (small molecules) are analyzed using these methods[42]. The produced spectral datasets consist of the different peak patterns based on the sample which could be stored in respective databases, allowing in reducing the gap in information and allowing further automated analyses and references thus following a high-throughput analysis methodology. Data analysis being a critical step in the interpretation of the information, numerous standardized techniques and technologies, software, data repositories are being constantly developed and upgraded[40,43]. This facilitates thorough analyses and data integration.

Metabolomics data, unlike metagenomics data are not standardized with respect to both format and presence of open-access repositories. The data obtained then is required to be re-processed jointly which required higher expertise and more varied usage of bioinformatic processes. Keeping the data organized and correctly mapped is a tedious and highly challenging task[40].

In this study, the metabolic changes with respect to two different diets of *S. cynthia ricini* are explored. The differential profile is determined along with annotation and characterisation of the metabolites involved is described. Further, the correlation of the amplicon sequence identified gut microbial profile based on the two diet-types with the metabolic profile was established. The role of the correlated microbes in the differentially abundant metabolites were explored in an attempt to understand the complex interaction of the microbiome.

Purposes of the present study include-

- Establishing the differential metabolic profile of the insect gut based on different host plants.
- Metabolite annotation and description.
- Establishing microbiome-metabolite correlations to understand the associations between the differentially abundant gut microbes and metabolites.

METHODOLOGY

Sample collection and storage

Eri silkworm (*S. cynthia ricini*) diseased-free layings (dfls) were acquired from Central Muga & Eri Research Institute, Jorhat, Assam, India (Lat: 26° 47′49.1″N Lon: 94° 19′35.0″E). They were reared in a domesticated semi-controlled environment with a room temperature of 25˚C and 80% relative humidity. Two types of leaf feeds were selected viz. *Ricinus communis Linn.* (Eri) and *Heteropanax fragrans* (Kesseru).

4 samples were collected from the 1st and 5th larval growth stages from the two different diet-types (**Table5.1**)

Sample Name	**Diet**	**Growth-Stage**
SCLE1	Eri Leaves	Instar 1
SCLE5	Eri Leaves	Instar 5
SCLK1	Kesseru Leaves	Instar 1
SCLK5	Kesseru Leaves	Instar 5

Table5.1: Metadata of the samples and the respective categories

Pooled samples based on equal weight were collected and stored from each instar of the larval stages and adult male and female moths fed on each diet-type for gut DNA extraction. Collected larval and moth samples were washed with 70% alcohol to minimize external microbial contamination. Samples were freeze dried and stored. The freeze-dried samples were solubilised using 300 µL mix of methanol/water (1:1) followed by homogenization through a Sonics Vibra-Cell™ (25 Hz, 7 min).

Metabolite extraction and profiling

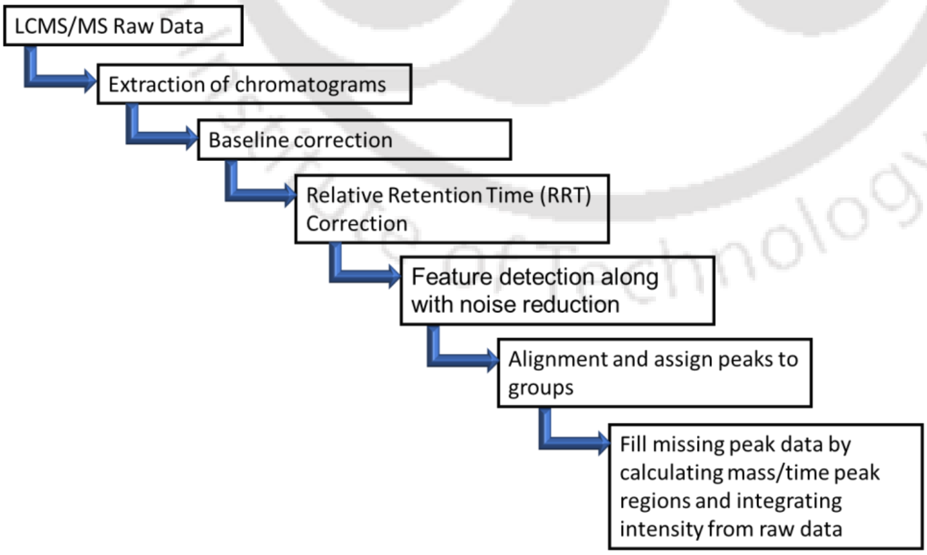


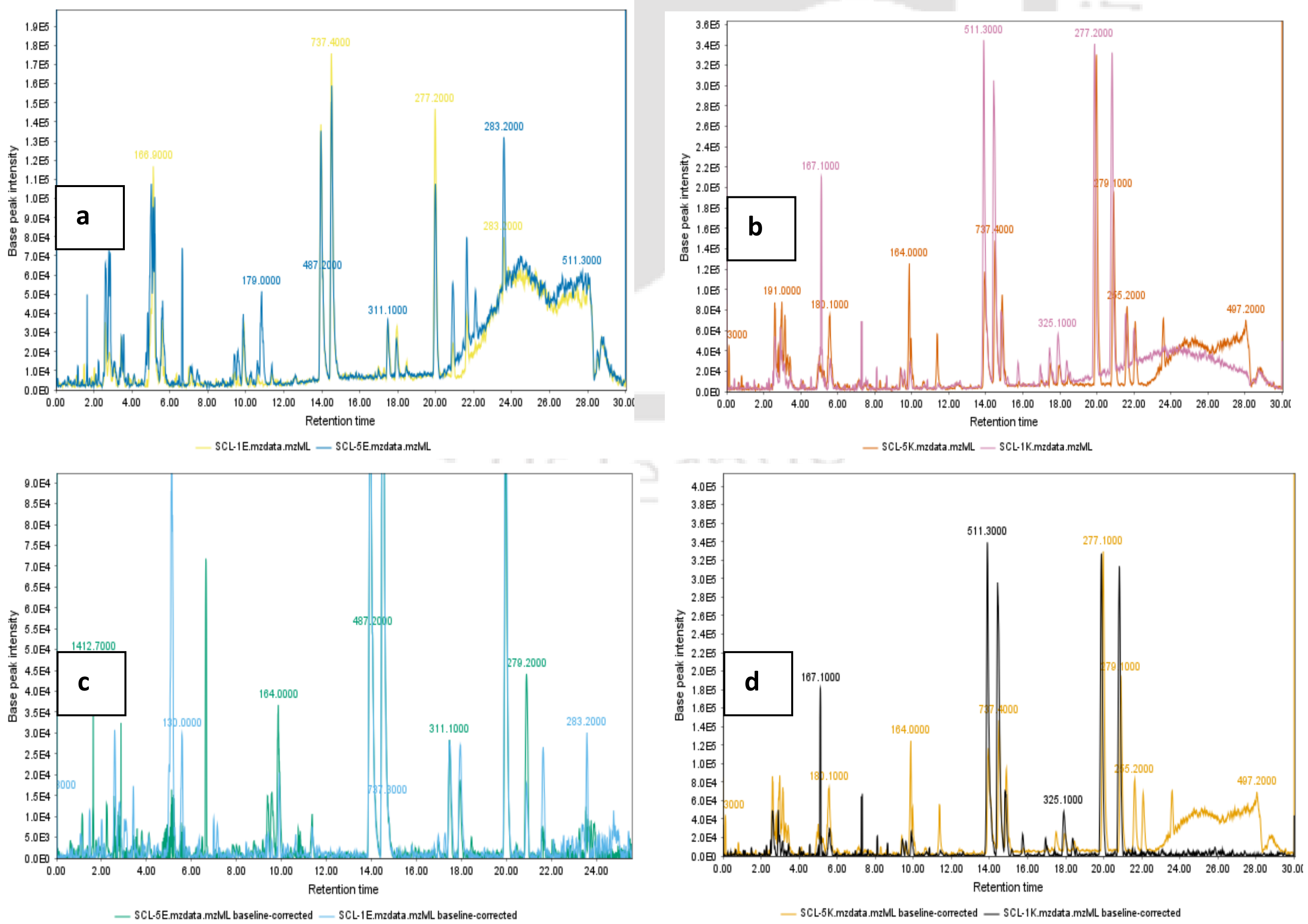
Figure5.1: Processing and filtration of LC-MS/MS data.

Metabolite extraction was carried out using adopted protocol from Yang *et. al*. The homogenized samples were transferred to 2ml Eppendorf tubes followed by their centrifugation at 12,000rpm for 10m at 4°C. Supernatants were then collected and the solutions were evaporated in a vacuum desiccator (Polycab) for 60m at 45 °C. The remaining solutes were then solubilised in a 200µl solution of 1:1 methanol: water each. Filtration was carried out using 0.22-µmol Millipore Filters and the solutions were transferred into chromatography vials.

Liquid Chromatograph Mass Spectrometry (LC–MS/MS) was carried out using Agilent 6410 system coupled to a Triple Quad MS-MS mass spectrometer (Agilent Technologies). A C-18 RRHD column (2.1 × 100 mm, 1.7 µm particle size; Agilent Technologies) was used for separation of the samples. Gradient elution was carried out with 0.1% formic acid aqueous solution as Solvent A and 0.1% formic acid-acetonitrile solution solvent B. The following gradient program was carried out: 99.9% A for 0–2m; 75% A for 2–6m; 20% A for 6–10m; 10% A for 10–12m; 0.1% A for 12–21m; 0.1% A for 21–23m; 99.9% A for 23–24m; and 99.9% A for 24–26m. Flow rate was maintained at 0.3ml/m and volume of injection of samples was 5µl. Temperature of the sample flow was maintained at 4°C, and column temperature was at 40 °C. The detection of compounds was carried out in negative (−) Electrospray Ionisation (ESI) mode of mass spectrometry. Range of scan was taken from 100–1500m/z. The electrospray ionization conditions were: 900l/h desolvation gas flow, 450˚C desolvation temperature, 120˚C source temperature, 25l/h cone gas flow, 6eV low collision energy, 50eV high collision energy and 2.5kV (-) capillary coltage.

The obtained spectral scan results were interactively visualised using MzMine (v) and processed for peak detection, alignment, and further filtration based on the chromatograms and methodology applied. Control sample was prepared by mixing all samples in equal amounts and evenly spaced among injections and used for baseline correction of raw samples preparing quality control (QC) samples. The obtained centroid data was then processed by XCMS for feature detection using centWave method for noise filtration with maximal tolerated m/z deviation in consecutive scans (Δ m/z) of 15 ppm, a minimum peak width of 10s and maximum peak width of 120s. A signal-to-noise ratio >20:1 was filtered. A summarised methodology of processing is represented in **Figure5.1**.

Relative retention time (RRT) correction was carried out using obiwarp method (Step size, profStep = 1) followed by peak integration/chromatogram alignment with parameters including width of overlapping m/z slices to use for creating peak density chromatograms i.e. mzwid =0.015, minimum sample fraction required in sample groups for its validity i.e. minfrac =0.5, and bandwidth (standard deviation or half width



at half maximum) of gaussian smoothing kernel applied to the peak density chromatogram for peaks grouping through the samples i.e. bw =5 (**Figure5.3**).

Figure5.2: Chromatograms visualised as sample groups a. Raw chromatogram of Eri-fed sample SCL_E1, SCL_E2, b. Raw chromatogram of Kesseru-fed samples SCL_K1, SCL_K5, c. Baseline corrected Eri-fed samples d. Baseline corrected Kesseru-fed samples.

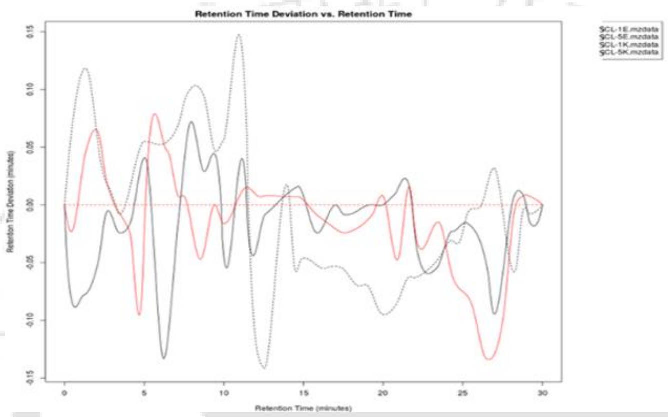


Figure5.3: A graph showing the correction curve. Samples eluting after median retention time indicated by a positive deviation, samples eluting before median retention time indicating negative deviation.

Statistical analyses were carried out using three conditions for filtration including analysis of variance (ANOVA) via Welch t-test (unequal variances) with a p-value ≤ 0.05, a fold change (FC) ≥ 2, and a minimum coefficient of variation (CV) ≤ 30 reducing “false discovery rate”. The obtained diversity was visualised via NMDS plotting.

The differences in the metabolites between the different sample groups based on ion changes was analyzed using OPLS-DA via a paired approach and visualised as a cloud plot. The differential abundance heatmap plot was visualised based on the values obtained. Differentiated filtered ions were annotated for their respective compounds using the standard Metabolite and Chemical Entity Database (METLIN) (http://metlin.scripps.edu).

Microbiome-metabolome interaction

Microbiome obtained from 16s rRNA gene sequence-based analyses previously was used for microbiome-metabolite correlation with the aim of deeper insights into the contribution of the microbiome present on the diet-based differential metabolite profile using Pearson correlation. Pairwise combination of each microbial taxonomic abundances and metabolite intensities were analyzed using Pearson correlation. Benjamini–Hochberg method was used to adjust p-values with a significance level <0.01 to adjust the false discovery rate. The differential abundances of microbes and metabolite intensities were visualised using the ComplexHeatmap (R 4.2.0).

A summarised representation of the methodology has been described below (**Figure5.4**)

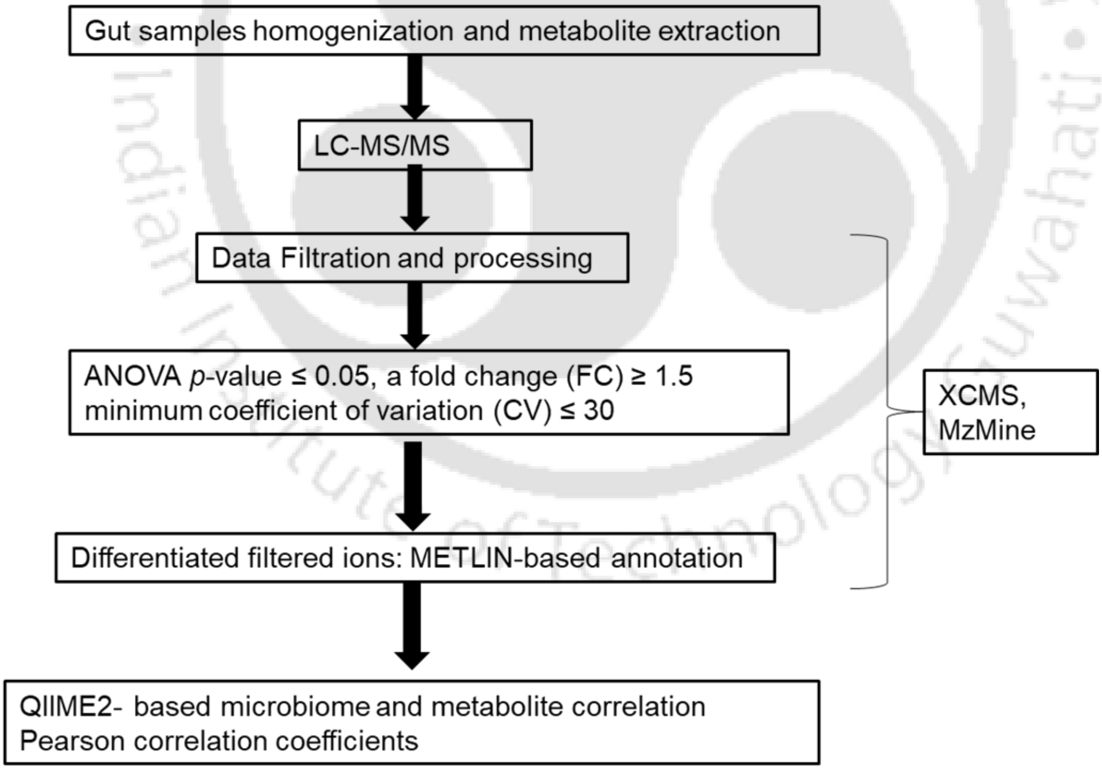


Figure5.4. A diagrammatic representation of the methodology applied.

RESULTS AND DISCUSSION

Diversity analysis

Following processing of the chromatograms of the samples, a total of 3553 (+) features were detected. Welch's ANOVA t-test showed the sample diversity to be divided into two distinct groups, thus demonstrating metabolic profiles which is diet-dependent, visualised through the NMDS-plot (**Figure5.5**). Further, Eri samples-based features were more diverse compared to Kesseru fed sample features. This observation was similar to the 16s rRNA gene sequence analysis-based bacterial microbiome diversity as summarised in Chapter 2 diet-based diversity plots, indicating the interaction of host-diet, the gut microbiome and the metabolome and the involvement of the gut microbiome in the alteration of the metabolites.

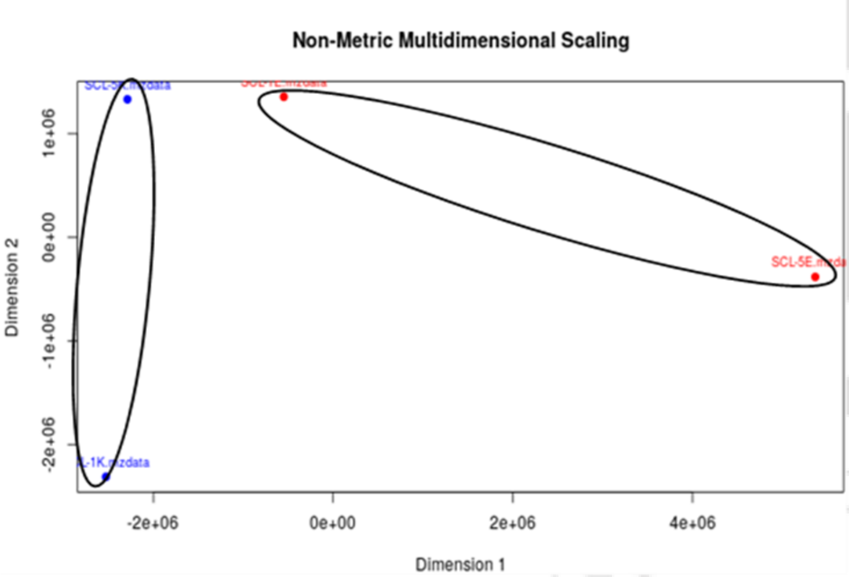


Figure5.5. NMDS plot of Unpaired parametric t-test (Welch t-test, unequal variances) analysing the diversity pattern of the metabolites based on different diets.

Differential abundance and annotation

A differential profile with 56 significant metabolites were observed based on OPLS-DA analysis visualised through a differential cloud plot with a p-value <0.05 (**Figure5.6**).

Using METLIN database with exact m/z value for mass determination and putative annotation, 33 metabolites were annotated and manually described associated with their involved characteristics, pathways or functions using PubChem, KEGG and MetaCyc databases summarised in Table5.1. 17 of the annotate metabolites were observed to be significantly upregulated in Eri-based gut metabolites and 16 of them were upregulated in Kesseru-based gut metabolites of the samples. Most differentially abundant compounds comprised of fatty acids and derivatives, vitamin and co-factor their intermediates and derivatives, antibiotics and antibacterial compounds, many organic compounds (lipids, phenols, steroids, ester) and organic acids. Some other compounds including amino acids, sugar derivatives and sodium salt differences were also observed. This indicated the involvement of the upregulated/downregulated metabolites in the insect metabolic activities and their association with the phenotype differences observed with respect to the different diet-types.

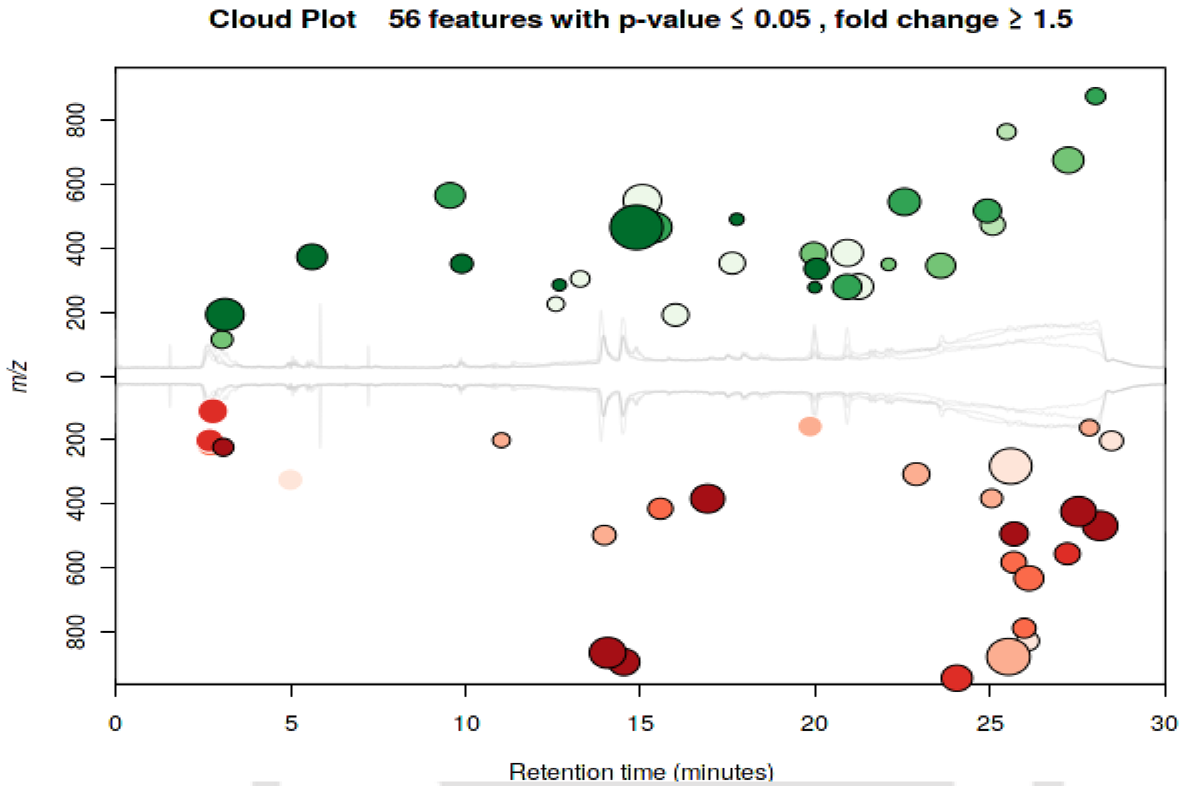


Figure5.6: Cloud plot representation of the differentially abundant gut metabolites based on diet-types. Green represented upregulated features in Eri-leaves based samples, red represents features upregulated in Kesseru-leaves based samples. Feature indicators of the molecules are defined by intensity (m/z) and retention time(m). Log-fold change is depicted by the value of the size of the circles. Colour shades depicts the p-value and interpret the significance, where lower p-values were depicted by brighter red/green shades. corrected retention time (TIC) of total chromatograms were overlaid in grey in the background.

Based on the OPLS-DA analysis heatmap, various fatty acid, intermediates and derivatives including methyl myristic acid, 1-Hexadecanoyl-2-hexanoyl-sn-glycero-3-phosphocholine, N-(5-hydroxy-pentyl) arachidonoyl amine, 2-Cyanoethyl, Nostocyclopeptide A2, numerous organic acid such as emodic acid, muconic acid, mycophenolic acid, etyrapone dichloroasterrate, various vitamins and co-factor derivatives including Benfotiamine, Precorrin 6Y, cholecalciferol, Hydroxypentanoyl-CoA, antibacterial compounds including Clioquinol, Cetrimonium, Biapenem, Actinorhodin, Urdamycin G, etc. nucleotides and amino acids including 8-(4-

Chlorophenylthio)-cAMP sodium, Granisetron metabolite 4 glucuronide, N2-(2-Carboxymethyl-2-hydroxysuccinoyl)arginine, etc., lipids and organic compounds including 4,4'-Sulfonyldiphenol, (E)-3-(Methoxycarbonyl)pent-2-enedioate, N-acylsphingosines, Honaucin A, Stearamide, etc., sugar derivative D-Mycinose were found to be differentially abundant through the two different diet groups (**Figure5.7**) (**Table5.2**). Most antibacterial compounds including Cetrimonium, Biapenem, Urdamycin G, etc. were found to be more enriched in Kesseru-fed samples. Organic acids emodic acid, Xipamide O-glucuronide, are highly enriched in Kesseru-fed samples. Vitamins derivatives such as Benfotiamine, 4,4'-Sulfonyldiphenol, Precorrin 6Y were also enriched in Kesseru-fed samples while co-factor Hydroxypentanoyl-CoA and vitamin D3 derivative cholecalciferol were enriched in Eri-fed samples. Many organic compounds, amino acid, lipids and sugar derivatives were found to more in Eri-fed samples. A large number of differentially abundant metabolites were of fungal source e.g., Emodic acid, Erythroskyrin from Penicillium, Urdamycin G, Mycinose-D from Streptomyces, Saccharomyces, Aspergillus.

Biapenem is an antibiotic sourced from Streptomyces has been found to be potent against a wide range of gram-negative anaerobic bacteria[44]. Stremocyes derived Actinorhodin is a polyketide antibiotic mostly reported to be induced through stressed conditions[45]. Granisetron metabolite 4 glucuronides are amino acids with anti-vomit properties[46]. Studies have shown gut bacterial enzymes (GUS) metabolizing glucuronides in the gut and playing important role in the health of the host[47]. Rifapentine are organic compounds with antibacterial properties[48]. This indicates the involvement of diet-based gut microbes playing a role in xenobiotics metabolism.

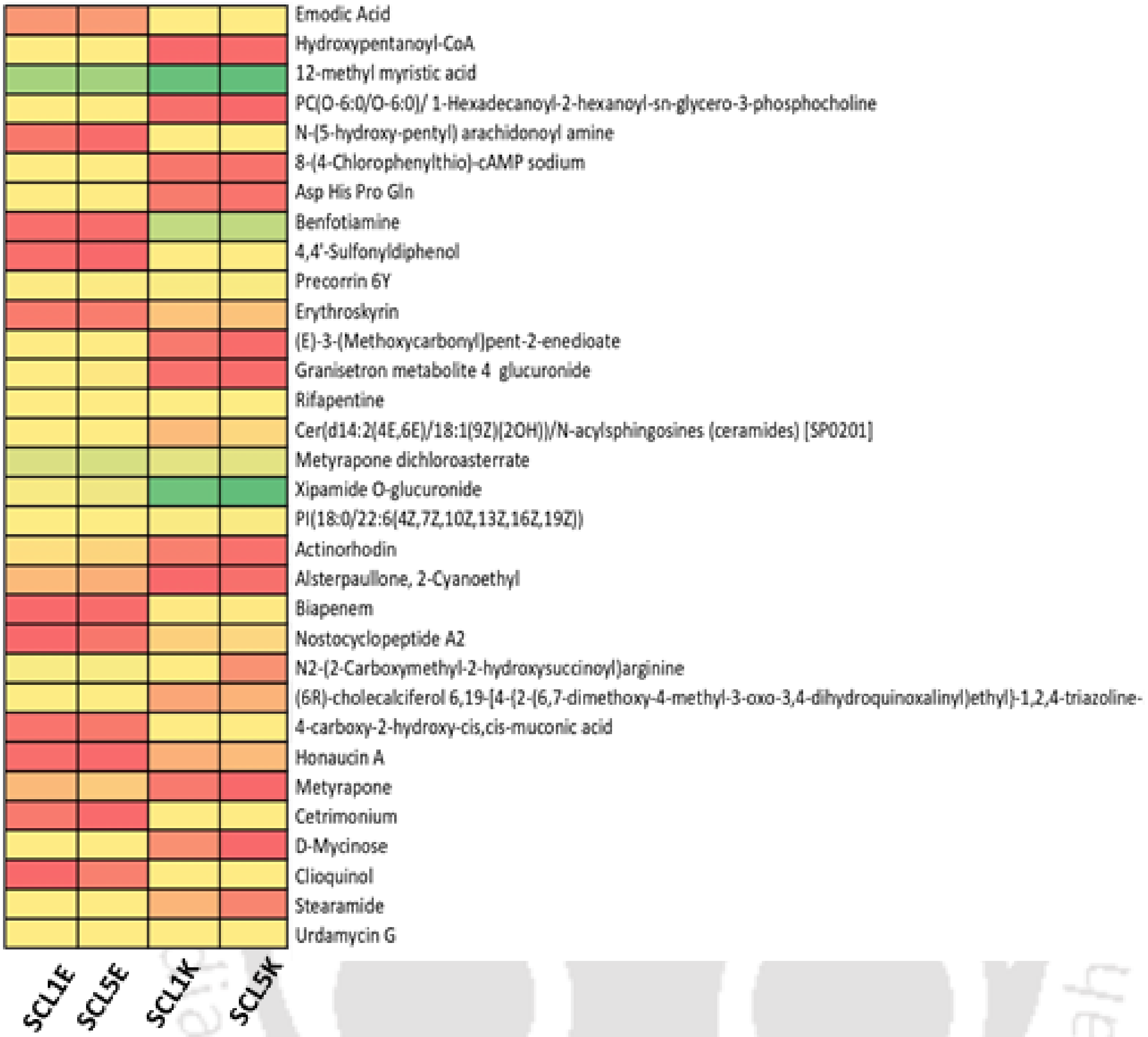


Figure5.7. Differential abundance heatmap representation of the putative identified metabolites based on a METLIN

Compound	Description
Emodic Acid	Organic acid
Hydroxypentanoyl-CoA	Derivative of Acyl Coenzyme A
12-methyl myristic acid	Methylated Fatty Acid, Antifouling agent
PC(O-6:0/O-6:0)/ 1-Hexadecanoyl-2-hexanoyl-sn-glycero-3-phosphocholine	Derivative of fatty acid phosphocholine
N-(5-hydroxy-pentyl) arachidonoyl amine	Unsaturated fatty acid
8-(4-Chlorophenylthio)-cAMP sodium	Thionucleotides
Asp His Pro Gln	Peptide
Benfotiamine	Thiamine derivative
4,4'-Sulfonyldiphenol	Phenol
Precorrin 6Y	Vitamin B12 intermediate
Erythroskyrin	Sodium salt
(E)-3-(Methoxycarbonyl) pent-2-enedioate	Organic compound
Granisetron metabolite 4 glucuronide	Amino acid with anti-vomit properties
Rifapentine	Organic compounds with antibacterial properties
Cer(d14:2(4E,6E)/18:1(9Z)(2OH))/N-acylsphingosines (ceramides) [SP0201]	Sphingolipids
Metyrapone dichloroasterrate	Methylated ester
Xipamide O-glucuronide	Mycophenolic acid glucuronide
PI(18:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	Phosphatidylinositol
Actinorhodin	Benzoisochromanequinone with antibacterial properties
Alsterpaullone, 2-Cyanoethyl	Derivative of saturated fatty acid
Biapenem	Antibiotic
Nostocyclopeptide A2	Fatty acid
N2-(2-Carboxymethyl-2-hydroxysuccinoyl) arginine	Bacteria induced metabolism
(6R)-cholecalciferol	Vitamin D3 derivative
4-carboxy-2-hydroxy-cis, cis-muconic acid	Organic acid
Honaucin A	Anti-inflammatory compound
Metyrapone	Steroid inhibitor
Cetrimonium	Antibacterial ammonium cation
D-Mycinose	Sugar derivative
Clioquinol	Antibacterial
Stearamide	Phototransduction
Urdamycin G	Glycoside

Table5.2: PubChem database-based description of the compounds annotated

Microbe-metabolite correlation

Positive correlation in the above analysis indicated the change in abundance of the respective microbes in the gut along with the abundance of the metabolites based on the diet. A negative correlation indicated no change to opposite abundance of the microbes with respect to the change in the metabolite abundance (**Figure5.8**).

From the above representation, *Rothia sp.* is observed to have a positive correlation with methylmyristic acid, a class of methylated fatty acid. *Rothia sp*. is largely associated with the degradation of oligosaccharides and protein molecules to simple structures[49]. Methylmyristic acid, a yeast derived metabolite, more abundant in Kesseru-based insect gut. It is reported to raise the low-density lipoprotein (LDL) upon consumption[50,51]. This correlation could indicate *Rothia sp*. being directly involved the metabolism of the fatty acid or the increased LDL in the body.

Methanspaera, an archae is seen to be positively associated with metabolites upregulated in Eri-based gut and are associated with methanogenesis in other insect guts. Eri-based samples were mostly observed to have higher abundance of metabolites involved in metabolism of sugar, fatty acids, amino acids and their derivatives. Most of these were fungi-derived metabolites. The stimulation of methanogenic microbes by anaerobic fungi and the methanogenic archaea increasing the sugar and fibre degrading ability has been reported in several studies. Thus, Methanospaera could be either directly involved in the digestion and energy associated metabolism of the host or it could be involved in a symbiotic relationship of the gut fungi.

Nostocylopeptide, fatty acid is positively correlated with proteobacteria including *Enterobacteriaceae, Pleisomonas, Pseudomonas*, etc. Nostocyclopeptide are found to be more abundant in Kesseru-based gut. Nostocyclopeptide A2 is a cyclopeptide naturally found in Cyanobacteria, was reported as an antitoxin against microcystins[52]. Proteobacteria members, although present in large amount in Eri-based samples were more abundant in Kesseru-based samples comparatively, thus indicating their role in aiding Nostocyclopeptide involved metabolism related to toxin degradation.

Deinococcus has a negative correlation with thionucleotides. Thionucleotide, 8-(4-Chlorophenylthio)-cAMP sodium was found to be upregulated in Eri-based gut while higher abundance of Deinococcus were found in Kesseru-based gut. Deinococcus sp. are associated with reducing oxidative stress in several insect gut environment[53]. 8-(4-Chlorophenylthio)-cAMP sodium is a protein kinase agonist[54]. 8-(4-Chlorophenylthio)-cAMP sodium could thus be responsible the decrease of the bacteria in the gut or be associated with more stressed environment in Kesseru-diet gut.

Biapenem, antibiotic positively correlated with microbes such as Kaitobacter, Acinetobacter, Bacteroidetes indicating a role of these organisms in the antibiotic metabolism. Most positive correlated interaction of microbes were seen to be with compounds related to fatty acid and organic acids indicating their possible role in energy based and detoxification metabolism. Stearamide, a naturally occurring organic compound found in plants and fungus such as *Beauveria bassiana* (Family: Cordycipitaceae), associated with signal transduction has been studied to be related with gut bacterial population including SJA4, Paenibacillaceae[55].

Mycinose-D, a deoxyallose derivative sourced from fungi such as Streptomycesis were associated with increased ability of the antibacterial agents such as Tylosin to enter bacterial cells[56]. It was found to be positively related with many bacteria including Pleisomonas shigelloides, Enterobacteriaceae, Ruminococcus, etc. Metyrapone, a pyrimidine derivative regulates corticosteroids which is associated with stress management and adaptation[57]. Bacterial population were not seen to be highly associated with its abundance. Honaucin A, a cyanobacterium derived compound was reported to have anti-inflammatory properties and involved in bacterial quorum sensing associated with bacterial populations such as SJA-4, Comamonadaceae[58].

Muconic acid is a commercially beneficial dicarboxylic acid derivative used industrially for the production of nylon[59]. It is associated with degradation of xenobiotics. Positive correlation of bacteria including Bacteroidetes and Pseudomonas were observed. These bacterial populations were reported to have similar roles from reported studies as well.

Hence, here we have established the microbes playing a role in the differential metabolite profile based on the diet differences in *S. cynthia ricini*. The effect of these microbes could be direct by altering the metabolic functions or indirect through playing a role in the increase/decrease of other microbes that are directly responsible for the differences in the metabolite abundances.

Figure5.8. Heatmap describing the microbiome-metabolite correlation.

CONCLUSION

In this study, we have studied the metabolic profile of *S. cynthia ricini* gut with respect to two diet-types and established the related bacterial species involved with the differential abundance of those metabolites.

Eri samples-based features were more diverse compared to Kesseru fed sample features which were similar to the feature diversity observed in amplicon studies carried out in Chapter 2, thus indicating major interaction of host-diet, the gut microbiome and the metabolome. A differential profile with 56 significant metabolites were observed, out of which most were originated from fungal sources indicating a larger role of fungal microbiome in the digestive and related other functions of the silkworm. Further, the correlation of these metabolites and the diet-based differentially abundant bacterial microbiome analyzed in Chapter 2 showed a diverse complex network as well.

Hence, the current study gives us a more sensitive outlook to the complex network of the gut microbiome and metabolites change with respect to the change in diet. These observations could help manipulate the microbiota accordingly for a desired change in metabolites for desired phenotypes or functions and hence a desired yield.

These observations interpreting their role in metabolism of the host organism helps us understand the microbiome of it and the complex network. This could be further analyzed for the quantitative analysis of the differential abundant metabolites with deeper analyses and could be used as a basis for hypothesis generation and also further integrative analyses with a larger acquisition of data. These findings could be used to various advantages for the commercially important and other related organism.

REFERENCES

1. Fani, R. The Origin and Evolution of Metabolic Pathways: Why and How did Primordial Cells Construct Metabolic Routes? *Evol.: Educ. Outreach* **5**, 367–381 (2012).
2. Coutinho, I. D. *et al.* Identification Of Primary And Secondary Metabolites And Transcriptome Profile Of Soybean Tissues During Different Stages Of Hypoxia. *Data Brief* **21**, 1089 (2018).
3. Junkins, E. N., McWhirter, J. B., McCall, L.-I. & Stevenson, B. S. Environmental Structure Impacts Microbial Composition And Secondary Metabolism. *ISME Comms. 2022 2:1* **2**, 1–10 (2022).
4. Nalbantoglu, S. & Nalbantoglu, S. Metabolomics: Basic Principles and Strategies. *Mol. Med.* (2019) doi:10.5772/INTECHOPEN.88563.
5. Clish, C. B. Metabolomics: An Emerging But Powerful Tool For Precision Medicine. *Cold Spring Harb Mol Case Stud* **1**, a000588 (2015).
6. Zhang, A., Sun, H., Xu, H., Qiu, S. & Wang, X. Cell Metabolomics. *OMICS* **17**, 495 (2013).
7. Wishart, D. S. Metabolomics For Investigating Physiological And Pathophysiological Processes. *Physiol Rev* **99**, 1819–1875 (2019).
8. Peng, B., Li, H. & Peng, X. X. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* **6**, 628 (2015).
9. Wu, K. *et al.* Gut immunity in Lepidopteran insects. *Dev Comp Immunol* **64**, 65–74 (2016).
10. Zhang, N. *et al.* Contribution Of Sample Processing To Gut Microbiome Analysis In The Model Lepidoptera, Silkworm *Bombyx mori*. *Comput Struct Biotechnol J* **19**, 4658–4668 (2021).
11. War, A. R. *et al.* Mechanisms Of Plant Defense Against Insect Herbivores. *Plant Signal Behav* **7**, 1306 (2012).

12. Voirol, L. R. P., Frago, E., Kaltenpoth, M., Hilker, M. & Fatouros, N. E. Bacterial Symbionts In Lepidoptera: Their Diversity, Transmission, And Impact On The Host. *Front Microbiol* **9**, (2018).

13. Šigut, M. *et al.* Fungi Are More Transient Than Bacteria In Caterpillar Gut Microbiomes. *Scientific Reports 2022 12:1* **12**, 1–12 (2022).

14. Hammer, T. J., Janzen, D. H., Hallwachs, W., Jaffe, S. P. & Fierer, N. Caterpillars Lack A Resident Gut Microbiome. *Proc Natl Acad Sci U S A* **114**, 9641–9646 (2017).

15. Broderick, N. A., Raffa, K. F., Goodman, R. M. & Handelsman, J. Census of the Bacterial Community of the Gypsy Moth Larval Midgut by Using Culturing and Culture-Independent Methods. *Appl Environ Microbiol* **70**, 293 (2004).

16. Yu, H. *et al.* Analysis Of The Intestinal Microflora In *Hepialus gonggaensis* Larvae Using 16S Rrna Sequences. *Curr Microbiol* **56**, 391–396 (2008).

17. Pinto-Tomás, A. *et al.* Enzymatic Activities Of Bacteria Isolated From The Digestive Tract Of Caterpillars And The Pupal Content of *Automeris zugana* and *Rothschildia lebeau* (Lepidoptera: Saturniidae). *Rev Biol Trop* **55**, 401–415 (2007).

18. Beran, F., Köllner, T. G., Gershenzon, J. & Tholl, D. Chemical Convergence Between Plants And Insects: Biosynthetic Origins And Functions Of Common Secondary Metabolites. *New Phytol* **223**, 52–67 (2019).

19. Wei, J. *et al.* Plants Attract Parasitic Wasps To Defend Themselves Against Insect Pests By Releasing Hexenol. *PLoS One* **2**, (2007).

20. Yang, L. *et al.* Caterpillar-Induced Rice Volatile (E)-β-Farnesene Impairs the Development and Survival of Chilo suppressalis Larvae by Disrupting Insect Hormone Balance. *Front Physiol* **13**, 960 (2022).

21. Joachim, C. & Weisser, W. W. Does the Aphid Alarm Pheromone (E)-β-farnesene Act as a Kairomone under Field Conditions? *J Chem Ecol* **41**, 267–275 (2015).

22. Seigler, D. S. Plants of the Northeastern United States That Produce Cyanogenic Compounds. *Econ Bot* **30**, 395–407 (1976).

23. Davis, R. H. & Nahrstedt, A. Biosynthesis Of Cyanogenic Glucosides In Butterflies And Moths: Effective Incorporation Of 2-Methylpropanenitrile And 2-Methylbutanenitrile Into Linamarin And Lotaustralin By Zygaena And Heliconius Species (Lepidoptera). *Insect Biochem* **17**, 689–693 (1987).
24. Rooks, M. G. & Garrett, W. S. Gut Microbiota, Metabolites And Host Immunity. *Nat Rev Immunol* **16**, 341 (2016).
25. Tolerance Of Western Corn Rootworm Larvae (Coleoptera: Chrysomelidae) To 6-Methoxy-2-Benzoxazolinone, A Corn Semiochemical For Larval Host Location. https://agris.fao.org/agris-search/search.do?recordID=US9501360.
26. Balcerczyk, A., Damblon, C., Elena-Herrmann, B., Panthu, B. & Rautureau, G. J. P. Metabolomic Approaches to Study Chemical Exposure-Related Metabolism Alterations in Mammalian Cell Cultures. *Int J Mol Sci* **21**, 1–22 (2020).
27. Lankadurai, B. P., Nagato, E. G. & Simpson, M. J. Environmental metabolomics: An emerging approach to study organism responses to environmental stressors. *Environ. Rev.* **21**, 180–205 (2013).
28. Rowland, I. *et al.* Gut Microbiota Functions: Metabolism Of Nutrients And Other Food Components. *Eur J Nutr* **57**, 1 (2018).
29. Tu, P. *et al.* Gut Microbiome Toxicity: Connecting the Environment and Gut Microbiome-Associated Diseases. *Toxics 2020, Vol. 8, Page 19* **8**, 19 (2020).
30. Vernocchi, P., del Chierico, F. & Putignani, L. Gut Microbiota Profiling: Metabolomics Based Approach To Unravel Compounds Affecting Human Health. *Front Microbiol* **7**, 1144 (2016).
31. Tang, Z. Z. *et al.* Multi-Omic Analysis Of The Microbiome And Metabolome In Healthy Subjects Reveals Microbiome-Dependent Relationships Between Diet And Metabolites. *Front Genet* **10**, 454 (2019).
32. Chandran, H., Meena, M. & Sharma, K. Microbial Biodiversity and Bioremediation Assessment Through Omics Approaches. *Front. in Environ. Chem.* **1**, 9 (2020).

33. Kang, W. N., Jin, L., Ma, H. Y. & Li, G. Q. Integrated Microbiome–Metabolome Analysis Reveals Stage-Dependent Alterations in Bacterial Degradation of Aromatics in *Leptinotarsa decemlineata*. *Front Physiol* **12**, 739800 (2021).

34. Chen, M. X., Wang, S. Y., Kuo, C. H. & Tsai, I. L. Metabolome Analysis For Investigating Host-Gut Microbiota Interactions. *JFMA* **118**, S10–S22 (2019).

35. Lee-Sarwar, K. A., Lasky-Su, J., Kelly, R. S., Litonjua, A. A. & Weiss, S. T. Metabolome–Microbiome Crosstalk and Human Disease. *Metabolites* **10**, (2020).

36. Bauermeister, A., Mannochio-Russo, H., Costa-Lotufo, L. v., Jarmusch, A. K. & Dorrestein, P. C. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol* **20**, 143 (2022).

37. Tang, J. Microbial Metabolomics. *Curr Genomics* **12**, 391 (2011).

38. Swanson, H. I. Drug Metabolism by the Host and Gut Microbiota: A Partnership or Rivalry? *Drug Metab. Dispos.* **43**, 1499 (2015).

39. Li, H., He, J. & Jia, W. The Influence Of Gut Microbiota On Drug Metabolism And Toxicity. *Expert Opin Drug Metab Toxicol* **12**, 31 (2016).

40. Alonso, A., Marsal, S. & Julià, A. Analytical Methods In Untargeted Metabolomics: State Of The Art In 2015. *Front Bioeng Biotechnol* **3**, 23 (2015).

41. Segers, K., Declerck, S., Mangelings, D., Heyden, Y. vander & Eeckhaut, A. van. Analytical Techniques For Metabolomic Studies: A Review. *Bioanalysis*. *https://doi.org/10.4155/bio-2019-0014* **11**, 2297–2318 (2019).

42. Zhou, B., Xiao, J. F., Tuli, L. & Ressom, H. W. LC-MS-based Metabolomics. *Mol Biosyst* **8**, 470–481 (2012).

43. Li, X. *et al.* LC-MS-Based Metabolomic Approach Revealed the Significantly Different Metabolic Profiles of Five Commercial Truffle Species. *Front Microbiol* **10**, 2227 (2019).

44. Kemung, H. M. *et al.* Streptomycesas a Prominent Resource Of Future anti-MRSA Drugs. *Front Microbiol* **9**, 2221 (2018).

45. Robertsen, H. L. & Musiol-Kroll, E. M. Actinomycete-Derived Polyketides as a Source of Antibiotics and Lead Structures for the Development of New Antimicrobial Drugs. *Antibiotics* **8**, (2019).

46. Nakamura, H. *et al.* CYP1A1 Is a Major Enzyme Responsible for the Metabolism of Granisetron in Human Liver Microsomes. *Curr Drug Metab* **6**, 469–480 (2005).

47. Pellock, S. J. & Redinbo, M. R. Glucuronides in the Gut: Sugar-driven Symbioses between Microbe and Host. *J Biol Chem* **292**, 8569 (2017).

48. Pyta, K. *et al.* Specific Interactions between Rifamycin Antibiotics and Water Influencing Ability to Overcome Natural Cell Barriers and the Range of Antibacterial Potency. *ACS Infect Dis* **5**, 1754–1763 (2019).

49. Wei, G., Darwish, G., Oppenheim, F. G., Schuppan, D. & Helmerhorst, E. J. Commensal Bacterium *Rothia aeria* Degrades and Detoxifies Gluten via a Highly Effective Subtilisin Enzyme. *Nutrients 2020, Vol. 12, Page 3724* **12**, 3724 (2020).

50. Designed Research; Y, K. S. S. & Performed Research; Y, K. S. S. Myristate Can Be Used As A Carbon And Energy Source For The Asymbiotic Growth Of Arbuscular Mycorrhizal Fungi. **117**, 25779–25788 (2006).

51. Galdiero, F. *et al.* Beneficial Effects Of Myristic, Stearic Or Oleic Acid As Part Of Liposomes On Experimental Infection And Antitumor Effect In A Murine Model. *Life Sci* **55**, 499–509 (1994).

52. Jokela, J. *et al.* A Novel Cyanobacterial Nostocyclopeptide Is A Potent Antitoxin Against Microcystins. *Chembiochem* **11**, 1594–1599 (2010).

53. Slade, D. & Radman, M. Oxidative Stress Resistance in Deinococcus radiodurans. *Microbiol Mol Biol Rev* **75**, 133 (2011).

54. Emery, A. C., Eiden, M. v. & Eiden, L. E. Separate Cyclic AMP Sensors for Neuritogenesis, Growth Arrest, and Survival of Neuroendocrine Cells. *J Biol Chem* **289**, 10126 (2014).

55. Sun, B., He, X. & Zhang, W. Findings on the Relationship Between Intestinal Microbiome and Vasculitis. *Front Cell Infect Microbiol* **12**, 812 (2022).

56. Bate, N. & Cundliffe, E. The Mycinose-Biosynthetic Genes Of Streptomyces Fradiae, Producer Of Tylosin. *J Ind Microbiol Biotechnol* **23**, 118–122 (1999).

57. Sigalas, P. D., Garg, H., Watson, S., McAllister-Williams, R. H. & Ferrier, I. N. Metyrapone in Treatment-Resistant Depression. *Ther Adv Psychopharmacol* **2**, 139 (2012).

58. Choi, H. *et al.* Honaucins A–C, Potent Inhibitors of Eukaryotic Inflammation and Bacterial Quorum Sensing: Synthetic Derivatives and Structure-Activity Relationships. *Chem Biol* **19**, 589 (2012).

59. Abdolmohammadi, S., Hernández, N., Tessonnier, J. P. & Cochran, E. W. Bioadvantaged Nylon From Renewable Muconic Acid: Synthesis, Characterization, And Properties. *ACS Symposium Series* **1310**, 355–367 (2018).

Chapter 6

Summary and Future Prospects

Chapter 6

Summary and Future Prospects

Summary

Samia cynthia ricini (Lepidoptera: Saturniidae) is a domesticated silkworm, producing unique warm Eri-silk. It has a wide prospect with respect to its ecological and research significance. In Assam and Northeast India, the silkworm is cultivated by the rural population and largely contribute to the ecological and economic aspects of these regions. However, their poor and obsolete rearing practice and cultivation techniques have led to a weakening of their health conditions, decrease in its population and low productivity of silk. Very little research has been carried out regarding the health and utilization of the silkworm.

Gut microbes play an important role in the phenotype of organism including the silkworm. The gut microbiota also changes in structure and function based on the surrounding environmental conditions affecting the growth and development of the organism. The structure and functions changes with conditions such as various growth stages, diet, diseases, etc. The gut microbiota contributes to a large amount of metabolic and cellular processes of the host organism directly or indirectly. The gut microbiota, the host and the surrounding environment are independent on each other. Understanding the gut microbiota of Eri silkworm and the microbiome, it could be manipulated for a better growth and silk yield of the silkworm.

In an attempt to address the research gap in these areas, present work explores the analysis of *S. cynthia ricini* with respect to its gut microbiota. The gut structure and functions were explored using amplicon and shotgun metagenome techniques. The changes of the phenotype based on growth stages and diet and the gut microbiota change along with it were reported. In addition, various comparative studies were reported. In addition, the silkworm metagenome was compared to other lepidopteran insects observing the role of the respective microbiota in their environmental behaviour. A microbiome-metabolite correlation was also reported with an attempt to understand the microbiome-metabolite changes with the different host plants and the role of microbes associated.

A chapter-wise summary is provided.

Comparative microbiome analyses of *Samia cynthia ricini* under different diet conditions and growth stages

In this study, a comprehensive insight into the microbial composition of *S. cynthia ricini* has been reported. The structure of the gut microbiota based on the various factors influencing it including growth stages, diet and life-stage i.e., the larval vs moth stages have been studied. The microbes with significant changes were analyzed inferring their involvement. Further predictive functions of the gut microbiota present were also studied with respect to the changes in the abovementioned environmental conditions to understand the possible functions associated with the changes.

These findings could be used to various advantages for the commercially important and other related organism. Novel microbes with potential could be further explored. Further, these studies would contribute to the research gaps in insect biology and evolution.

The basis of microbiome composition of the organism and their changes with respect to different conditions can help also help with deeper analyses combined with shotgun sequencing to understand the microbial genome composition and important details including Metagenome-assembled genomes (MAGs), Horizontal Gene Transfers (HGTs) and Antibiotic Resistance Genes (ARGs). The study could be explored for further insights consisting of multi-omics approaches.

Shotgun metagenome analysis of *Samia cynthia ricini* 5th instar larva

In the present study, shotgun sequence of the silkworm larva gut was analyzed to construct the metagenome of *S. cynthia ricini* was explored with. Higher resolutions of taxonomic and functional abundances were reported. Optimization of the methodologies for the analyses regarding the organism were explored and summarised. Contig assembly and microbial gene annotations were carried out using various databases followed by reconstructed predictive functions and the functional abundance were reported. Comparative details with respect to amplicon sequence based microbial population of the sample was also studied. Comparative taxonomic and functional abundance summaries indicating the advantages of applying different approaches based on our analyses of interest were reported. The major classes of Antimicrobial Resistance Genes (ARGs) annotated, is reported here which is an important aspect to understand the mechanism associated with antibiotic resistance of pathogenic strains and horizontal gene transfers (HGTs).

Interactive metabolic activities of the microbiome comparative analyses observing the role of microbes associated with various diseases in the silkworm could be further studied via integrative analyses could be explored for understanding that would lead to the development of better strains with higher resistance against those diseases.

Metagenome Assembled Genomes (MAGs) exploration is another aspect of further studies, thus identifying novel important microbial strains and the associated genes. Established resistomes based on the Antimicrobial Resistant Genes (ARGs) is an important aspect to understand the mechanism associated with antibiotic resistance of pathogenic strains and horizontal gene transfers (HGTs).

Comparative analyses of lepidoptera larva metagenome

Lepidoptera being one of the largest groups if insects, having a wide range of diversity and ecological significance, this study attempts to understand the microbial diversity associated with it. Differences in the taxonomic and functional abundance among the species were summarised with the aim to understand their role in their ecological and phenotypic variances. Patterns of microbial signatures were observed. This study, attempts to reduce the research gap in the familiarity of lepidoptera microbiome allowing for a better measure for their growth, development and further research. The pest related microbiota was studied which could assist in their better regulation.

Further analyses could be done with respect to the plant hosts, the lepidoptera insects and pests, the microbes and the surrounding environment in an integrative manner using system biology techniques. Understanding the microbial signatures could also help predict disease patterns.

Metabolic profiling of *Samia cynthia ricini* based on different diets and analysis the of the associated microbiome

In this study, we have studied the metabolic profile of *S. cynthia ricini* with respect to two diet-types and established the related bacterial species involved with the differential abundance of those metabolites. Detailed patterns based on the various environmental factors influencing it have been studied. These findings could be used to various advantages for the commercially important and other related organism.

These observations interpreting their role in metabolism of the host organism helps us understand the microbiome of it and the complex network. This could be further analyzed for the quantitative analysis of the differential abundant metabolites with deeper analyses and could be used as a basis for hypothesis generation and also further integrative analyses with a larger acquisition of data.

Future Prospects of the study

Future prospects of the studies include

- In-depth studies of other environmental conditions affecting the growth and health of the organism. Other diet plants of the host are to be explored.
- Studies on various diseases of the organism to observed the structural and functional diversity associated hence creating a predictive model for diseased species would be carried out. This could help in prevention and better management of vital diseases of the silkworm.
- Integrative analyses could be carried out with higher resolution of metabolic profile and including other meta’omics-based techniques such as

metatranscriptomics and metaproteomics to generate a more in-depth knowledgebase of the microbiome of the silkworm.

- Formulation of a microbiota-based diet with the required combinatorial structure of beneficial microbes could be undergone which could provide for an enhanced growth of the silkworm with better immunity.
- Metabologenomics, a new aspect of study in the sub-field of metagenomics to be explored combining the shotgun metagenome and gene cluster assembly techniques in correlation with the metabolic profiles generating a deeper and specific insight. Novel microbes with potential could be further explored. Further, these studies would contribute to the research gaps in insect biology and evolution.
- With the development of more powerful tools and machine learning, a deeper analysis could be carried out. Gradual development of prediction software for retention time has been developed for a more successful mass detection of the small molecules from complex mixtures including food, microbial culture, plasma and hemolymph. This increases the identification rate during the LC-MS/MS process and a improved interpretation and quantification of the data.
- Production of better yield and quality of silk dependent on the gut microbiota could be explored for the economic benefit of the rural population of Northeast India.

CURRICULUM VITAE

BIJU BHARALI

PERSONAL INFORMATION:

Date of birth: 25th July, 1992
Nationality: Indian
Present Address: Subansiri Hostel, Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India
Email: bijub@iitg.ac.in, j4biju@gmail.com
Phone no: +918753928570

EDUCATIONAL QUALIFICATIONS:

NAME OF EXAMINATION	***YEAR OF PASSING***	***SCHOOL/ COLLEGE***	***BOARD/ UNIVERSITY***	***PERCENTAGE***
Bachelor of Science (BOTANY)	2013	COTTON COLLEGE, GUWAHATI	GAUHATI UNIVERSITY	71.2%
Master of Science (BIOTECHNOLOGY)	2015	ST. ANTHONY’S COLLEGE, SHILLONG	NORTH EASTERN HILL UNIVERSITY	64.9%

RESEARCH EXPERIENCE:

Junior Research Fellow, IIT Guwahati **August 2016-July 2017**

- Department of Biotechnology, Ministry of Science and Technology, Government of India funded project.
- Area of Research: “Aptamer probes for detection of celiac disease epitopes of gluten in commercial varieties of Indian wheat, rye and barley.”

Junior Research Fellow, IIT Guwahati, **January 2016-July 2016**

- Department of Biotechnology, Ministry of Science and Technology, Government of India funded project.

- Area of Research: “Effect of molecular chaperones on the heterologous expression of human interferon gamma in *Pichia pastoris* expression system.”

Junior Research Fellow, IIT Guwahati **January 2016- July 2016**

- Area of Research: “Development of high biomass and lipid yielding medium for newly isolated yeast”.
- Skills: Isolation, purification and identification of novel microbial strains, Statistical models of nutrient optimization for microbial growth, Spectrophotometry.

Master’s dissertation, NEHU Shillong **January 2014-March 2015**

- Area of Research: “Isolation of Bacteriocin producers from milk and milk products-Comparative studies of Bacteriocin production and their partial purification and characterization and classification based on their molecular weight.”

Research Intern, IIT Guwahati, **December 2014- January 2015**

- Area of Research: “Production of cellulose from fungal strains supporting degradation of cellulosic waste.”

PUBLICATIONS

- Exploring the Benefits of Endophytic Fungi via Omics; In: Singh, B. (eds) Advances in Endophytic Fungal Research. Fungal Biology. Springer, Cham. https://doi.org/10.1007/978-3-030-03589-1_4

- Engineered nanomaterials in plants: Sensors, carriers, and bio-imaging; Comprehensive Analytical Chemistry, Elsevier; doi.org/10.1016/bs.coac.2019.09.009

- Development of high biomass and lipid yielding medium for newly isolated *Rhodotorula mucilaginosa*; Fuel; JFUE-D-18-00410

- Gene and process level modulation to overcome the bottlenecks of recombinant proteins expression in *Pichia pastoris*. March 2018; Current Pharmaceutical Biotechnology 19(15); DOI: 10.2174/1389201019666180329112827

- Engineering folding mechanism through Hsp70 and Hsp40 chaperones for enhancing the production of Recombinant human interferon gamma (rhIFN-γ) in *Pichia pastoris* Cell factory. February 2018; Chemical Engineering Science 181; DOI: 10.1016/j.ces.2018.02.003

CONFERENCES AND SEMINARS

- Overexpression of cytoplasmic chaperons to modulate the expression level of human interferon gamma (hIFN-γ) in Pichia pastoris, Bioprocessing India, CIAB Mohali, Punjab, India, 2016

- Overexpression of cytoplasmic chaperons to modulate the expression level of human interferon gamma (hIFN-γ) in Pichia pastoris, Reflux 2017, IIT Guwahati, India, 2017

- ‘Banking on insects: a roadmap to conservation and sustainability’; International Symposium on Biodiversity and Biobanking 2018 at Indian Institute of Technology Guwahati, Guwahati, India. (ABSTRACT PUBLICATION).

- ‘Diversity of insect mitochondrial genomes’; International Symposium on Biodiversity and Biobanking 2018 at Indian Institute of Technology Guwahati, Guwahati, India. (ABSTRACT PUBLICATION).

WORKSHOPS

- One Week Online Workshop on Environmental Genomics and Genome Editing organized by Knowledge Incubation for TEQIP, Centre for Educational Technology, IIT Guwahati, 2021.

- Winter School Program, Gifu University. “Isolation, Characterization and Identification of Bifidobacterium strains from Silkworm *Bombyx mori* gut using NGS-library preparation and sequencing.”, 2019.

- BIOCONVERSE at Manas National Park organized by DoS, BTC and CVSc, AAU, Khanapara, Guwahati, 2018

- Workshop on Advances in Environmental Protection and Sustainability (AEPS) organized by IIT Guwahati 2018.

CERTIFICATES:

Programming for Everybody (Python) September 2020
University of Michigan. Credential ID: DBFSGTCDU4J2.

Gut Check: Exploring Your Microbiome September 2020
University of Colorado Boulder. Credential ID: VGKCX2HUHRSZ

Introduction to Genomic Technologies September 2020
John Hopkins University. Credential ID: 6ML7A6MRUFC4

Bioinformatic Methods I September 2020
University of Toronto, Credential ID: U3CJ55PXBDAK

R Programming October 2020
John Hopkins University Credential ID: BRNQXWTUWN2L