DNA Research, 2025, **32**, dsaf029
https://doi.org/10.1093/dnares/dsaf029
Advance access publication 24 October 2025
Resource Article: Genomes Explored

OXFORD

Resource Article: Genomes Explored

Life inside a bag: multiomics insights into the bagworm species *Eumeta crameri*

Abhisek Chakraborty, Shruti Mahajan[†], Vishnu Prasoodanan P.K.[†], Akhilesh Shailendra Khamkar, and Vineet K. Sharma*

MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal - 462066, Madhya Pradesh, India

*Corresponding author. MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal - 462066, Madhya Pradesh, India (Email: vineetks@iiserb.ac.in).
†Equal contributions.

Abstract

Bagworms are commonly known for the well-organized case or bag surrounding them constructed using their silk and plant materials. To understand the genetic basis of these unique characteristics in bagworms, we performed multiomics analyses of a bagworm species, *Eumeta crameri*. The genome and transcriptome sequencing of *E. crameri* were used to construct the nuclear genome with a size of 668.2 Mb, N50 value of 6.6 Mb, and 13,554 coding genes, which was further assembled into 31 pseudochromosomes. The mitochondrial genome had a size of 15.6 Kb. We established the phylogenetic position of *E. crameri* with respect to 54 other insect species. The comparative analyses of *E. crameri* with other Lepidopterans revealed the adaptive evolution of genes related to primary metabolic pathways, defense, molting, and metamorphosis, and silk formation in the bagworm species. We also showed the ultrafine nature of the *E. crameri* silk fibres. Further, we performed the gut microbiome sequencing for *E. crameri* and constructed a gut microbial gene catalogue, which revealed the unique composition of the gut microbiome and its significance for host metabolism and defense. Together, the results provide multifaceted insights into the biological processes that support the well-organized holometabolous metamorphosis inside the bags of *E. crameri*.

Keywords: bagworm; genome; transcriptome; gut microbiome; adaptive evolution; metabolism; metamorphosis; silk.

1. Introduction

Eumeta crameri, commonly known as bagworm moth, belongs to the Psychidae family (order Lepidoptera) of Arthropods containing more than 1,300 species throughout the world.[1] More than 100 species of bagworms exist in the Indian subcontinent,[1] among which *E. crameri* is more commonly found.[2] The most interesting characteristics of bagworms are the production of silk fibres and the formation of well-organized portable cases or bags using their silk and other environmental materials.[3] Both male and female bagworm larvae can construct these small cases after hatching from the eggs, and only the head and thorax remain protruding out while dragging the case during movement.[2] The bags surrounding them serve in many ways, such as protecting against predatory attacks, maintaining microclimatic conditions for their development, and acting as camouflage.[4]

Eumeta crameri is commonly found in India and Bangladesh and resides on multiple host plant species, including *Acacia* sp. and *Litchi* sp.[2] It spins and decorates its cocoon throughout its larval life with the help of twigs, bark, and thorns from its host plant. *Eumeta crameri* shows at least four larval instars in its life cycle, and it renovates the case once between two consecutive larval instars once the case becomes smaller for its body size.[2] The pattern in which the length of the longest stick increases with each instance of case renovation is in agreement with Dyar's principle.[5] Also, the choice of case constructing material depends on its perception of the signal to optimize energy expenditure.[6] Regulation of metabolic energy utilization is a major process in the life cycle stages of holometabolous insects such as the bagworm species.[7,8] The carbohydrate and lipid metabolism rates are regulated depending on the life cycle stages, and insects also regulate the metabolic processes for physiological adaptations during a food shortage, ie, starvation stress.[8,9] *Eumeta crameri* has also been identified as a seasonal pest, showing damage potential in economically important plant species in India.[10]

Silk is a structural protein produced by the Arthropods, which aids in various functions such as foraging, reproduction, and locomotion.[11,12] The silk-producing ability has evolved convergently across the Arthropod classes.[13] Although there are only a few silk types observed in the Arthropods, the structural and mechanical properties differ.[14,15] Even within the Lepidoptera order, different mechanical properties of silk have evolved in Bombycidae (*Bombyx mori*) and Saturniidae, despite having a similar usage of silk.[16] In bagworms, a highly ordered hierarchical structure results in significant strength and toughness of the

Received 18 March 2025; revised 30 September 2025; accepted 6 October 2025

silk that is comparable to that of silkworm; however, the tensile strength is remarkably lower than that of spider silk.[15]

The gut microbiome has a role in determining the behaviour of insect hosts, and the host–microbiome interactions via symbiotic relationships can also influence the evolution of host species.[17] The gut bacteria can assist the host Lepidopteran species by aiding in nutrient acquisition, digestion, pathogen protection, and overcoming antiherbivore defense mechanisms of plants.[18,19] The microbiome composition of Lepidopteran insects has a wide range of intra-species and inter-species variation, which depends on various factors such as habitat, host plants, and insect developmental stage.[18] In silkworm and other insect gut microbiomes, the abundance of genes associated with carbohydrate breakdown and other metabolic processes and antimicrobial resistance were observed, which are involved in host digestion, development, and disease resistance activities.[18,19]

Despite the importance of bagworms in the production of silk fibres and the intriguing life inside the bag, the genome sequences of only four bagworm species were available so far.[15,20] Therefore, in this study, we sequenced the genome of bagworm species *E. crameri* using Illumina short reads and 10× Genomics synthetic long reads. We also performed sequencing of the transcriptome and gut microbiome of *E. crameri* using Illumina short read technology. The functional and evolutionary analyses revealed the adaptive evolution of genes involved in metabolic processes, metamorphosis, and silk-forming mechanisms, and indicated that the unique composition of the genome and gut microbiome could aid in the host metabolism and survival in the different life cycle stages.

2. Materials and methods

2.1. Sample preparation for genomic DNA and RNA extraction

A single and the same ($n = 1$) *E. crameri* larva was used for genomic DNA and RNA extraction and sequencing. The genomic DNA was extracted from about half of the larval body, similar to a previous study,[15] and the remaining half was used for RNA extraction. The genomic DNA extraction was performed using the Blood and Cell Culture DNA Mini kit (Qiagen, USA). The species was identified using amplification and Sanger sequencing of the marker gene cytochrome c oxidase subunit I (COI). The extracted genomic DNA was used to prepare the library using TruSeq Nano DNA library preparation kit and was sequenced on Illumina X Ten and NovaSeq 6000 platform (Illumina, CA, USA) for 150 bp paired-end sequencing. For 10× genomics synthetic long read sequencing, the library was prepared on a Chromium instrument using Chromium Genome Library & Gel Bead Kit v2 (10× Genomics, CA, USA). The 10× Genomics library was sequenced on NovaSeq 6000 (Illumina, CA, USA). The RNA was extracted using TRIzol reagent (Invitrogen, USA) and was used for library preparation with a TruSeq stranded mRNA library preparation kit following the manufacturer's protocol. The library was sequenced on the NovaSeq 6000 platform (Illumina, CA, USA) to generate paired-end reads.

2.2. Sample preparation for metagenomic DNA extraction

For faecal sample collection, the *E. crameri* larvae were cultured in sterile tip boxes. The boxes were lined with aluminium foil, which was cleaned with 70% ethanol to provide a sterile environment. The lids of the boxes were replaced by sterile mesh, and the larvae were observed for faecal sample collection. The faecal samples were collected immediately after defecation. The faecal samples were transferred to sterile Eppendorf tubes with sterile forceps and stored at −20 °C for metagenomic DNA extraction.

The metagenomic DNA from collected faecal samples of 19 *E. crameri* late instar larvae ($n = 19$), including the larva that was used for genomic DNA and RNA sequencing, was extracted using a DNeasy PowerSoil kit (Qiagen, USA) with a few modifications. The metagenomic DNA was used to amplify the V3 region of the 16S rRNA gene, and libraries were prepared following the 16S rRNA amplicon sequencing library preparation protocol (Illumina, CA, USA). Among the 19 larvae, the metagenomic DNA from six individuals ($n = 6$) was taken for whole metagenomic shotgun sequencing library preparation using a TruSeq Nano DNA library preparation kit (Illumina, CA, USA). The 16S rRNA amplicon and shotgun metagenome libraries were sequenced on the Illumina NovaSeq 6000 platform (Illumina, CA, USA) to generate 150 bp paired-end reads.

2.3. Genomic data preprocessing and genome assembly

Illumina short read sequencing data of *E. crameri* was quality-filtered using Trimmomatic v0.39[21] with the parameters —"TRAILING:20 LEADING:20 SLIDINGWINDOW:4:20 MINLEN:60". Barcode-filtering and quality-filtering of 10× Genomics reads were performed using proc10xG python scripts (https://github.com/ucdavis-bioinformatics/proc10xG) and Trimmomatic v0.39[21] with the same parameters, respectively. RNA-Seq data was also quality-filtered using Trimmomatic v0.39 with the same parameters used for genomic data processing.[21]

Before genome assembly, Jellyfish v2.2.10[22] was used with the quality-filtered Illumina genomic short reads to calculate k-mer count frequencies, which were further used to predict genome size and genomic heterozygosity using GenomeScope v2.0.[23] Using Illumina short read data (quality-filtered), the k-mer value was predicted using KmerGenie v1.7051.[24] Preprocessed Illumina short reads and quality-filtered 10× Genomics reads were used for genome assembly using SPAdes v3.15.3[25] with k-mer values of 87, 97, 107, 117, and 127 (best k-mer value as predicted by Kmergenie). The assembly obtained using the k-mer value of 107 was selected for downstream analyses for its best assembly statistics (assembly v0.1). Further, barcoded 10× Genomics reads were used for de novo assembly using Supernova v2.1.1 with "–maxreads = all" and other default parameters,[26] and Tigmint v1.2.2[27] was used in default mode for correction of mis-assembly regions using barcode-processed (using Longranger basic v2.2.2) 10× Genomics reads (assembly v0.2).

Both assembly v0.1 and assembly v0.2 were further scaffolded with pre-processed Illumina short reads using Platanus-allee v2.2.2 (default parameters)[28] and barcode-processed 10× Genomics reads using ARCS v1.2.2 (default parameters).[29] Preprocessed RNA-Seq reads were also used to scaffold the assemblies using AGOUTI v0.3 with default parameters.[30] The scaffolded assembly v0.1 and scaffolded assembly v0.2 were merged to construct a single, more contiguous genome assembly of *E. crameri* using Quickmerge v0.3 with default parameters.[31,32] Gap-closing

on this assembly was performed using GapCloser v1.12 (with "-l 155" option)[33] with preprocessed 10× Genomics reads and Illumina short reads. Further, single base errors, local misassemblies, or small indels were fixed using pre-processed short reads using Pilon v1.23 with default parameters.[34] After polishing, scaffolds with lengths of ≥5 Kb were retained and inspected for the presence of any contamination with BlobTools using BWA-MEM mapping of pre-processed Illumina short reads and BLASTN to the NCBI-nt database.[35,36] Scaffolds that were assigned to any phylum other than "Arthropoda" were removed to construct the final genome assembly of *E. crameri*.

The scaffold-level genome assembly of *E. crameri* was further assembled into pseudochromosomes via genome-wide synteny alignment with the chromosome-scale genome assembly of another bagworm species *Luffia ferchaultella* (GenBank accession—GCA_949709985.1),[20] using Satsuma2 "Chromosemble" (default parameters).[37,38]

BUSCO v5.6.0 was used to assess the genome assembly completeness with the Insecta_odb10 dataset of single-copy orthologous genes.[39] Further, quality-filtered Illumina short reads and pre-processed 10× Genomics reads were separately mapped on the genome assembly using BWA-MEM v0.7.17 (default parameters) to calculate the read mapping percentage.[36] Quality-filtered RNA-Seq reads were also mapped on the genome assembly using HISAT v2.2.1 with default parameters.[40]

The mitochondrial genome of *E. crameri* was assembled from the quality-filtered Illumina short reads using GetOrganelle v1.7.6[41] with animal_mt as a seed database. The circular mitochondrial genome was annotated using GeSeq in CHLOROBOX[42] with the reference mitochondrial genomes of Phychids (*E. variegata*, *Mahasena oolona*, and *Dahlica ochrostigma*) and four other Lepidopteran species (*Danaus*, *Bombyx*, *Heliconius*, and *Melitaea*) available in NCBI RefSeq database.

2.4. Genome annotation

The scaffold-level genome assembly of *E. crameri* was used to identify the repetitive regions in the genome. RepeatModeler v2.0.1 (with "-LTRStruct" and other default parameters)[43] was used to construct a de novo repeat library, which was used for repeat-masking the genome using RepeatMasker v4.1.0 with default parameters (http://www.repeatmasker.org/).

Before coding gene prediction, pre-processed RNA-Seq data of *E. crameri* were assembled de novo using Trinity v2.14.0 with default parameters.[44] The transcriptome assembly and protein sequences of phylogenetically closer organisms (order: Lepidoptera) obtained from Ensembl[45] Metazoa release 54—*Danaus plexippus*, *Bombyx mori*, *Heliconius melpomene*, and *Melitaea cinxia* were used as empirical evidence in MAKER[46] v3.01.04 genome annotation pipeline (using AUGUSTUS v3.2.3[47] as ab initio gene finder) for coding gene prediction from *E. crameri* repeat-masked genome assembly. Coding genes with AED value <0.5 and length of ≥150 nucleotides were retained to construct the high-confidence gene set. Further, coding gene prediction was also performed for *E. variegata* genome assembly obtained from a previous study[15] using the same steps in MAKER pipeline, and a high-confidence coding gene set was also constructed for this bagworm species. The coding gene sets of both *Eumeta* species were mapped against NCBI-nr database and Swiss-Prot database[48] using BLASTP (e-value 10^{-5}), and mapped against Pfam[49] database using HMMER v3.3.2[50] (e-value 10^{-5}).

Noncoding RNAs were also identified in *E. crameri* genome assembly. tRNAscan-SE v2.0.11[51] with default parameters and Barrnap v0.9 with "–kingdom euk" option were used for tRNA and rRNA prediction (https://github.com/tseemann/barrnap), respectively. MirGeneDB v2.1 database[52] was used for the identification of miRNAs using BLASTN (query coverage 90%, sequence identity 90%).

MCScanX was used for performing intra-species synteny analysis and gene duplication analysis for *E. crameri* species, with five genes necessary to call a collinear block.[53,54] Inter-species synteny analysis was also performed between *E. crameri* and *E. variegata* using MCScanX with the same parameter.

2.5. Gene expression analysis

Quality-filtered RNA-Seq reads of *E. crameri* species were mapped to the coding gene set of *E. crameri* constructed in this study using Kallisto v0.46.0 with default parameters.[55] RNA-Seq data of *E. variegata* from previous study[15] was also quality-filtered using Trimmomatic v0.39[21] with the same parameters as used for *E. crameri*, and mapped to the coding gene set of *E. variegata* constructed in this study using Kallisto with default parameters.

2.6. PSMC analysis

Comparative demographic history for *E. crameri* and *E. variegata* species was constructed using pair-wise sequentially Markovian coalescent (PSMC).[56] The pre-processed Illumina reads were mapped to the assembled genome of *E. crameri* constructed in this study, and the quality-filtered Illumina reads of *E. variegata* from the previously available study[15] were mapped to its genome assembly using BWA-MEM.[36] The consensus diploid sequences were constructed using SAMtools (v1.9) "mpileup" and BCFtools v1.9 in a way that the minimum read depth and maximum read depth remain one-third and twice the average read depth, respectively, and consensus sites with quality score <20 were filtered.[57–59] The parameters for PSMC analysis were set to "-N25 -t15 -r5 -p4+25*2+4+6" with 100 bootstrap values, a mutation rate of 3×10^{-9}, and a generation time of 1 yr.[2,60]

2.7. Phylogenetic analysis

For the construction of a genome-wide species phylogenetic tree, protein sequences from all the insect orders available in Ensembl[45] Metazoa Release 54 (Diptera, Lepidoptera, Coleoptera, Hemiptera, Phthiraptera, and Isoptera) were obtained (one species per genus). Along with these species, *E. crameri* and *E. variegata* protein sequences obtained from the MAKER pipeline were also used for the phylogenetic tree construction.

The protein sequences of all 55 species were used for the construction of orthogroups using OrthoFinder v2.5.4 with default parameters,[61] and fuzzy one-to-one orthogroups were identified using KinFin v1.0 with default parameters[62] (Table S1). Among these orthogroups, only those containing proteins from all 55 selected species were considered, and only the longest protein for each species was retained. The filtered one-to-one orthogroups were individually aligned using MAFFT v7.490 with default parameters,[63] and the multiple

sequence alignments were processed for removing the empty sites and concatenation using BeforePhylo v0.9.0 (https://github.com/qiyunzhu/BeforePhylo). The resultant alignment was used to construct the phylogenetic tree using RAxML v8.2.12[64] with 100 bootstrap values and the "PROTGAMMAAUTO" substitution model.

2.8. Analyses of evolutionary signatures

2.8.1. Genes with higher rate of evolution

The protein sequences of *E. crameri* and *E. variegata* obtained in this study, and proteins of four other Lepidopteran species (*Danaus plexippus*, *Bombyx mori*, *Heliconius melpomene*, and *Melitaea cinxia*) obtained from Ensembl Metazoa Release 54 were used for the construction of orthogroups using OrthoFinder v2.5.4 with default parameters.[61] The orthogroups were processed to contain only the longest sequence for each of the six species, and the filtered orthogroups were individually aligned using MAFFT v7.490 with default parameters.[63]

The orthogroups alignments were used for construction of individual gene phylogenetic trees with RAxML v8.2.12[64] using "PROTGAMMAAUTO" substitution model and 100 bootstrap values. "adephylo" package available in R was used to identify the *E. crameri* and *E. variegata*-specific genes with higher rate of evolution compared to the orthologous genes in other selected Lepidopteran species.[65]

2.8.2. Genes showing unique amino acid substitution with functional impact

The orthogroups alignments were also used to identify the genes in *Eumeta* species with unique amino acid substitutions compared to other species in the orthogroups. In this analysis, the amino acid positions that were identical in other species but different in *E. crameri* and *E. variegata* were identified as the uniquely substituted amino acids in *E. crameri* and *E. variegata*, respectively. Gaps and ten amino acids around a gap in the alignments were ignored. The impact of these unique substitutions on the protein function was predicted using SIFT (Sorting Intolerant From Tolerant) with UniProt as a reference database.[66] The unique substitutions that the protein function is "intolerant" to can subsequently alter the phenotype of a species, therefore, carrying evolutionary significance.[67]

2.8.3. Genes with positive selection

The protein sequence alignments of the orthogroups were converted into the corresponding codon alignments using PAL2NAL v14 with "-nogap" and "-nomismatch" options.[68] Using the resulting codon alignments, the branch-site model utilized by the "CodeML" program in PAML v4.9a was used to identify the positively selected genes in the two *Eumeta* species.[69] In this analysis, the null model ("fix_omega = 1") assumed $\omega \leq 1$ for all codons in all branches, whereas the alternative model ("fix_omega = 0") assumed $\omega > 1$ for the foreground branch. To compare the two models, a likelihood-ratio test (LRT) was performed, and *P*-values were estimated using the chi-square test (df = 1). The *Eumeta* genes qualifying against the null model with FDR-corrected *P*-values <0.01 were identified as positively selected genes. Further, Bayes Empirical Bayes (BEB) analysis was performed, and codon sites with >95% probability for the foreground lineage were extracted as positively selected codon sites for the *Eumeta* species. If the positively selected genes contained no positively selected codon site, they were excluded as false positives.

Eumeta genes showing more than one of the three evolutionary signatures—positive selection, higher rate of evolution, and unique substitution with functional impact were termed MSA (multiple signatures of adaptive evolution) genes.[70,71]

2.9. Gene family evolution analysis

Gene family expansion/contraction analysis was performed using CAFÉ v5 across the 55 species that were used for species phylogenetic tree construction.[72] The longest isoforms for the proteins of all 55 species were used for the all-versus-All BLASTP search (with "-seg yes" and "-outfmt 7" options), followed by clustering and gene family filtering (to remove the gene families containing ≥100 gene copies for ≥1 species) as suggested for the CAFÉ analysis.[73] In this analysis, the ultrametric species tree was constructed using the divergence time between *E. variegata* and *Agrilus planipennis* obtained from TimeTree v5.[74] A multilambda model was used in CAFÉ v5, where species from the Lepidoptera order were assigned a separate lambda value compared to the other insect species. Using the estimated lambda values, the large gene families that were filtered out in the earlier step were also analysed. The expanded gene families with ≥5 expanded genes were considered as the "Highly expanded" gene families.[75]

2.10. Shotgun metagenomic data analysis

The shotgun metagenomic reads for six *E. crameri* individuals were quality-filtered using Trimmomatic v0.39[21] with the parameters: "TRAILING:20 LEADING:20 SLIDINGWINDOW:4:20 MINLEN:60". The quality-filtered reads were further used to remove host genomic contamination using the whole genome assembly constructed in this study and to remove human contamination with BMTagger v3.101. The resultant reads were used for taxonomic classification using Kraken v2 with default parameters,[76] and the reads that could be classified as prokaryotic reads (paired and unpaired) were utilized for sample-specific assembly as well as co-assembly using metaSPAdes v3.15.3[77] with default parameters.

The resultant contigs with lengths ≥300 bases were retained and used for the taxonomic assignment of contigs using CAT with default parameters.[78] Contigs that were identified as Eukaryotic contigs using CAT were removed. The filtered contigs were used for gene prediction from the sample-specific assemblies and the co-assembly using Prodigal[79] v2.6.3 with "-p meta" option. Genes predicted from these assemblies were concatenated, and CD-HIT v4.8.1[80] was used with a sequence identity threshold of 1.00 to construct a non-redundant gene catalogue for *E. crameri* gut microbiome. This gene catalogue was further used for CAT taxonomic assignment, and the genes assigned as Eukaryotic were removed to construct the filtered gut microbial gene catalogue.

Gene abundance in each sample was calculated after mapping the filtered reads onto the gene catalogue using SOAP v2.21 with "-m 100", "-x 2000", and other default parameters.[81] Two types of alignments were used for sequence-based profiling[82]—(i) Entire paired-end read mapped onto a gene and (ii) one end of a paired-end read mapped and the other end remained unmapped. The mapped read was counted as

one copy in both the cases. The read count was also normalized based on the gene length as $b_{i=}x_{i/}L_i$. Thus, the calculation of the relative abundance of gene i was as follows: $a_i = b_i/\sum_j b_j = x_i/L_i/\sum_j x_j/L_j$, where a_i is the relative abundance of gene i in sample S, x_i is the number of mapped reads for gene i in sample S, L_i is the length of gene i, j is the number of genes, and b_i is copy number of gene i in sequenced data from sample S. The relative gene abundance was further normalized by read depth (by dividing with the number of mapped reads for each corresponding sample). The relative gene abundance table was used to construct a gene proportion table.

2.11. Amplicon data analysis

The 16S rRNA amplicon paired-end reads for 19 individuals were quality-filtered using Trimmomatic v0.39[21] with the same parameters used for shotgun metagenomic data filtering. Quality-filtered paired-end reads were analysed using QIIME 2 pipeline.[83,84] The paired-end reads were denoised and filtered for the presence of any chimeric sequences using DADA2[85] with the parameters: "–p-trim-left-r 0", "–p-trim-left-f 0", "–p-trunc-len-r 200", and "–p-trunc-len-f 240", to construct the feature table that was used for downstream analyses. q2-feature-classifier plugin was trained on the V3 region using SILVA v138 reference database[86] and was used for the taxonomic classification of ASVs (amplicon sequence variants). The relative phylum abundance was calculated after the taxonomic assignment of the ASVs, based on the detected ASV counts in each sample.

2.12. Functional annotation

2.12.1. Coding gene set

Eumeta crameri and *E. variegata* coding genes were annotated using KAAS v2 genome annotation server and eggNOG-mapper v2 with default settings.[87,88] Using the outputs, KEGG pathway enrichment analysis was performed for the *E. crameri* coding gene set with the coding genes of all six Lepidopteran species used in this study as a background using Fisher's exact test and then adjusted using the Benjamini–Hochberg procedure (adjusted P-value <0.01). *Eumeta crameri* and *E. variegata* genes showing evolutionary signatures were also analysed for KEGG pathway enrichment using Fisher's exact test (adjusted P-value <0.01) with the complete coding gene set of *E. crameri* and *E. variegata* as background, respectively. Further, the CAZy (Carbohydrate-Active enZYmes) families assigned to the *E. crameri* coding genes using eggNOG-mapper were used to calculate the abundance of CAZy families in the genome.

2.12.2. Gut microbial gene catalogue

The *E. crameri* gut microbial gene catalogue was also annotated using KAAS v2 genome annotation server and eggNOG-mapper v2 with default settings.[87,88] The gene proportion table constructed in the earlier step was used to calculate KO (Kegg Orthology) and EC abundance from the eggNOG-mapper results. The CAZy families assigned to the gut microbial genes using eggNOG-mapper were used to calculate the abundance of CAZy families in the gut microbiome.

2.13. Silk sample preparation for SEM imaging

Eumeta crameri bags were treated in chloroform for the removal of any wax. Treated bags were degummed in water having 0.5% (w/w) sodium carbonate and 10% ethylenediamine solution at 85 °C for 1 h with a solution-to-sample ratio of 20:1. The degummed bags were washed thoroughly in warm water and dried in ambient conditions. Samples were observed under High-Resolution Field Emission Scanning Electron Microscope (HR FESEM) at Central Instrumentation Facility, IISER Bhopal.

3. Results

3.1. Whole genome and transcriptome sequencing

A total of 115.5 Gb of Illumina short read data and 65 Gb of 10× Genomics synthetic long read data were generated for *E. crameri* species in this study corresponding to 180.1× coverage of Illumina short reads and 101.3× coverage of 10× Genomics reads based on the estimated genome size of 641.8 Mb. Further, 8.7 Gb (86.3 million reads) of RNA-Seq data was also generated for *E. crameri* species (Table S2).

3.2. Genome assembly

The genome of *E. crameri* was estimated to contain 0.77% heterozygosity (Fig. S1). The assembled genome size of *E. crameri* was 668.2 Mb, close to the estimated genome size. The genome assembly contained 1,777 scaffolds (the longest scaffold of 21 Mb) with an N50 value of 6.6 Mb and 33.53% GC content (Table S3, Fig. S2). 95.01% of the quality-filtered Illumina short reads, 95.04% barcode-filtered 10× Genomics reads, and 92.69% quality-filtered RNA-Seq reads were mapped on the final genome assembly. Further, 92.4% complete BUSCOs (91.3% complete and single-copy BUSCOs) were found in *E. crameri* genome assembly using the Insecta_odb10 database. The scaffold-level genome was used to construct a pseudochromosome-level genome assembly of *E. crameri* consisting of 31 pseudochromosomes with an N50 value of 26.3 Mb (Table S3).

The constructed circular mitochondrial genome of *E. crameri* had a size of 15,550 bases (Fig. S3). The mitochondrial genome was AT-rich (A = 41.63%, T = 39.23%, G = 6.97%, C = 12.17%). Thirteen protein-coding genes, 20 tRNAs, and two rRNA genes were identified in the mitochondrial genome assembly of *E. crameri*, similar to *E. variegata* mitochondrial genome.[89] All seven enzymes (*ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5*, and *ND6*) of complex I, the largest complex of the mitochondrial electron transport chain (ETC), were present in the mitochondrial genome assembly constructed in this study.

3.3. Genome annotation

The de novo repeat library for *E. crameri* species contained 2,135 sequences, and repeat-masking of genome using this repeat library showed the presence of 39.9% repetitive regions (with 31.42% interspersed repeats) (Table S4). Quality-filtered RNA-Seq reads were de novo assembled to generate 67,202 transcripts (N50 value = 2,315 bases) (Table S5), which were used as empirical evidence for coding gene prediction of *E. crameri* species. 14,659 coding genes were identified, of which 13,900 (94.82%) were retained after AED value-based filtering. Furthermore, 13,554 coding genes were present in the final high-confidence coding gene set after length-based filtering. In this study, we also annotated the publicly available genome assembly of the other *Eumeta* species (*E. variegata*) because of the presence of several gene prediction artefacts as mentioned in the previous study.[15] The repetitive regions consisted of 42.7% (with 35.12% of interspersed repeats) of the *E. variegata* genome, and 12,332 coding genes were identified in the final high-confidence coding gene set.

97.46% and 98.67% of the coding genes could be mapped using three publicly available databases—Swiss-Prot, NCBI-nr, and Pfam for *E. crameri* and *E. variegata* species, respectively (Table S6). Further, 89.84% and 92.17% of the coding genes could be assigned as "Insecta"-specific genes using eggNOG-mapper for *E. crameri* and *E. variegata* species, respectively. Among the various KEGG pathways assigned to the *E. crameri* gene set, genes involved in primary metabolic processes were noteworthy (Table S7, Figs. S4 and S5), similar to the gene set of *E. variegata*.[15] Among the coding genes of the two *Eumeta* species, 9,282 *E. crameri* genes (68.48%) had high expression (TPM value >1), and 8,949 *E. variegata* genes (72.57%) had high expression (TPM value >1). Additionally, 487 tRNAs, 115 rRNAs, and 31 miRNAs were identified in *E. crameri* genome assembly.

3.4. Collinearity and orthologous gene clustering

Collinearity analysis identified the presence of 8.66% of *E. crameri* genes in the intra-species collinear blocks. Additionally, among the coding genes of *E. crameri* genome, 40.53% were dispersed duplicated, 4.71% were proximally duplicated, 10.36% were tandem duplicated, and 35.75% were singleton genes. Further, 5,041 *E. crameri* genes (37.19%) and 4,787 *E. variegata* genes (38.82%) were present in the 527 inter-species collinear blocks identified between *E. crameri* and *E. variegata*.

Among the six Lepidopteran species, *B. mori* had the highest number of species-specific gene clusters (396 clusters), followed by *D. plexippus* (147 clusters) and *E. crameri* (111 clusters) (Fig. 1), and *E. crameri* and *E. variegata* species had 8,281 shared gene clusters between them (Fig. S6).

3.5. Comparative demographic history of the two *Eumeta* species

Effective population size (N_e) was estimated for *E. crameri* species using PSMC, which showed the presence of two population bottleneck events between 0.1 and 0.2 million years ago (mya), and ~1 mya, similar to other Lepidopteran insects[91] (Fig. 2a). The decreases in effective population size appear to be the result of repeated glaciation events in the Pleistocene epoch or the "ice age", similar to other Lepidopterans such as *Pieris rapae*.[92] However, in the case of *E. variegata* species, only one population bottleneck event was observed ~0.2 mya. Further, PSMC analysis suggested that the effective population size is smaller in *E. variegata* species despite a recent increase (Fig. 2b). The difference in the effective population size and the occurrence of population bottleneck events in these *Eumeta* species could be a true representation given the similar observations in *Kallima* species[91] or could be an artefact due to the lower genomic coverage by Illumina short reads data in the case of *E. variegata* species.[15]

3.6. Phylogenetic position and evolution of gene families

One hundred and thirty-eight fuzzy one-to-one orthogroups with 147,397 alignment positions were used for phylogenetic analysis across 55 insect species including *E. crameri* and *E. variegata*, spanning seven phylogenetic orders. The two *Eumeta* species were placed in the same clade with *Bombyx mori* as the closest Lepidopteran species. Species from the Diptera order were found to be closer to the Lepidopteran species (Fig. 3). The relative positions of the insect orders observed in this phylogenetic tree were also in accordance with the previous studies.[93]

A total of 24,060 gene families across 55 insect species were used for gene family evolution analysis after clade-based and size-based filtering. 882 and 767 gene families underwent expansion and contraction in *E. crameri* species, respectively. In *E. variegata* species, 766 gene families were expanded, and 1,152 gene families were contracted. Among the expanded gene families of the two *Eumeta* species, 18 and 19 gene families were highly expanded in *E. crameri* and *E. variegata*,

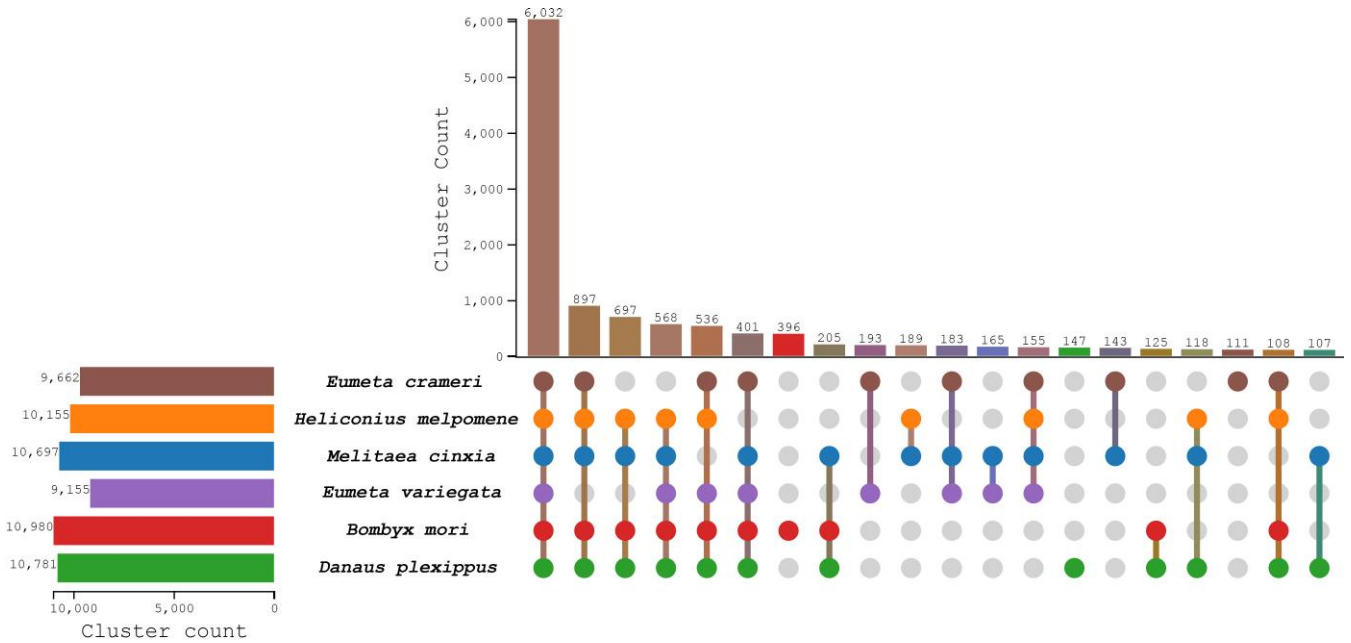

Fig. 1. UpSet plot showing the shared orthologous clusters among *E. crameri*, *E. variegata*, and four other Lepidopteran species.[90]

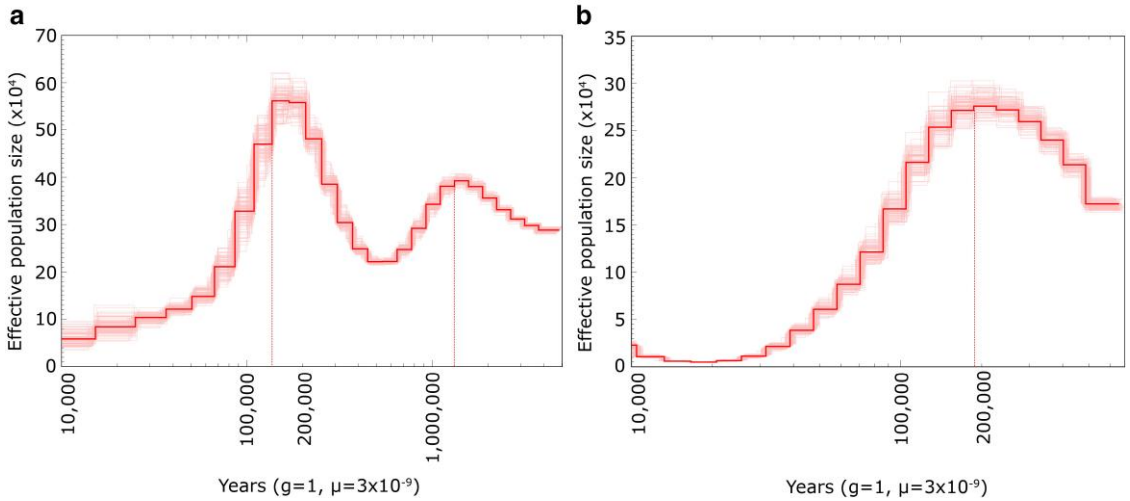

Fig. 2. Demographic histories of bagworm species. The light pink lines represent the bootstrap values used in PSMC analysis. a) *E. crameri*, b) *E. variegata*.

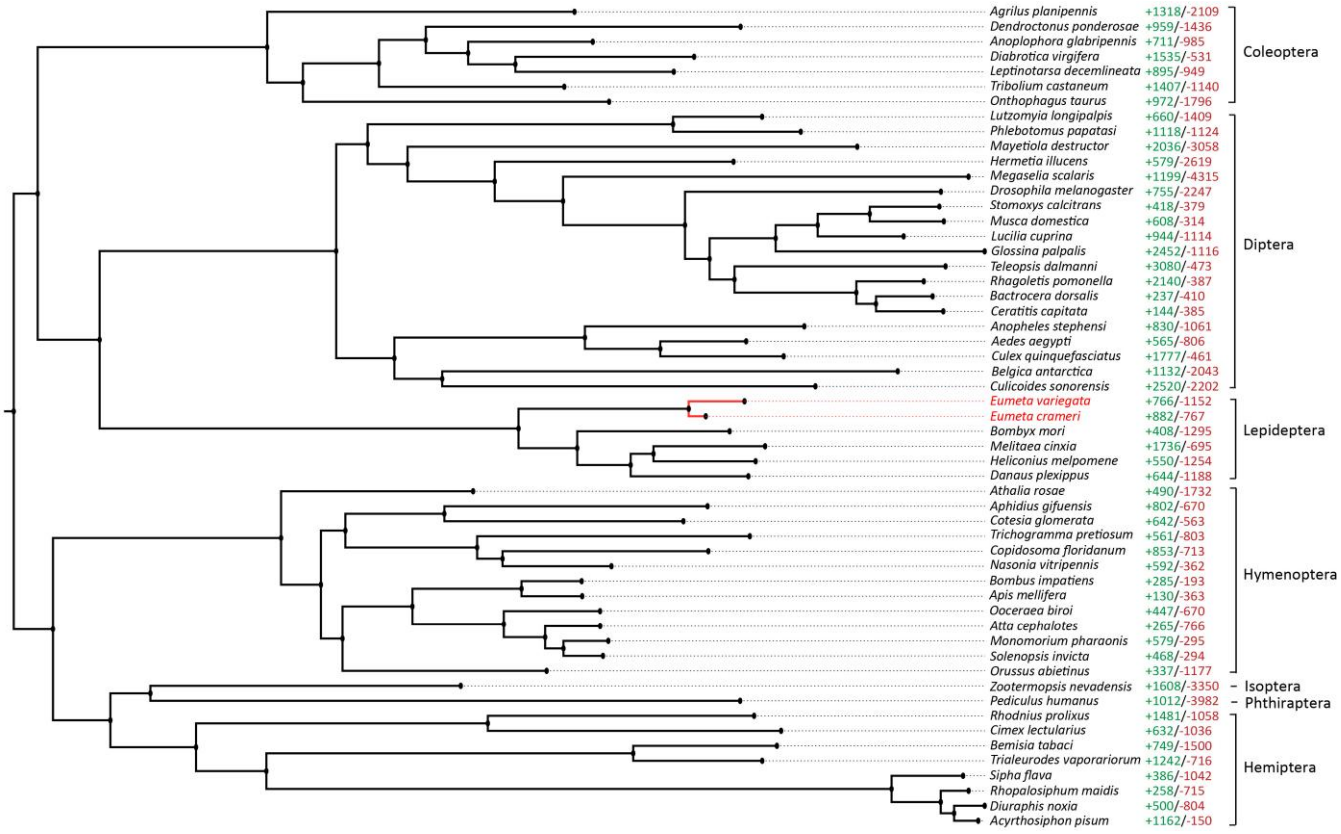

Fig. 3. Phylogenetic position of *E. crameri* and *E. variegata* with respect to 53 other species spanning seven insect orders. The numbers in green and red indicate the number of expanded and contracted gene families, respectively, in each species.

respectively, in terms of gene numbers (≥5 expanded genes) (Table S8). Further, among the large gene families (containing ≥100 gene copies for ≥1 species), one gene family was highly expanded in each of the *Eumeta* species. Overall, among the highly expanded gene families, *CYP9* in *E. crameri* (expanded by 10 genes), and *CYP6* in *E. variegata* (expanded by 13 genes) are notable examples, which are well-known cytochrome P450 families involved in resistance to insecticides and are often highly expanded in Lepidopteran pests.[94]

3.7. Genes with evolutionary signatures

To analyse the genes with evolutionary signatures, 6,032 orthogroups were identified across the six Lepidopteran species including *E. crameri* and *E. variegata* (Fig. 1). In *E. crameri* species, 97 genes had a higher rate of evolution, 351 genes were positively selected ($P < 0.01$), and 613 genes had unique amino acid substitution with functional impact. Further, 88 genes showing Multiple Signatures of Adaptive evolution (MSA) were identified, among which four had all three

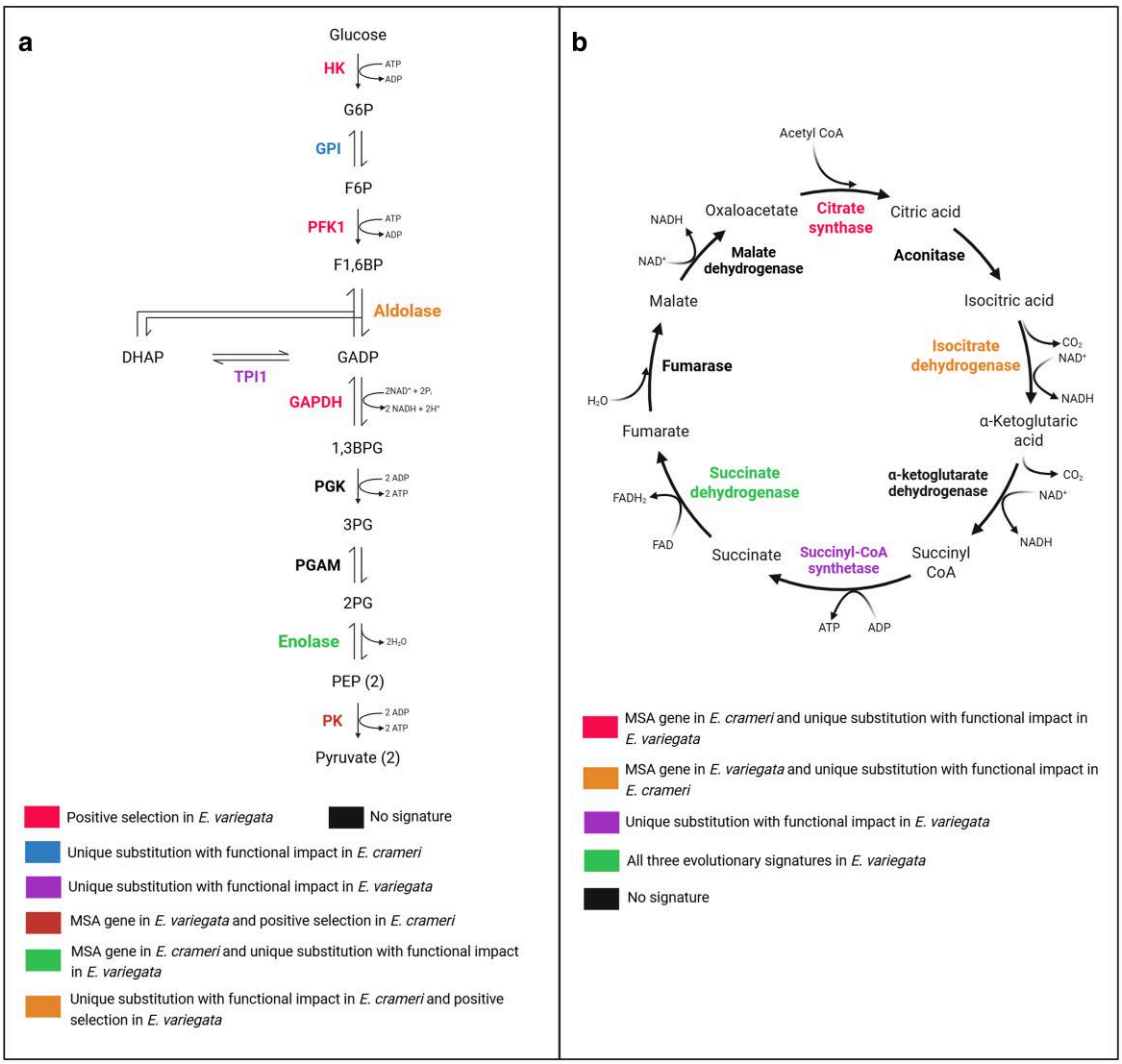

Fig. 4. Evolutionary signatures in glycolysis and citric acid cycle of the two *Eumeta* species. The term "MSA gene" represents the genes showing ≥2 evolutionary signatures. a) Glycolysis pathway. HK, hexokinase; GPI, glucose-6-phosphate isomerase; PFK1, phosphofructokinase-1; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGAM, phosphoglycerate mutase; PK, pyruvate kinase. b) Citric acid cycle. Created in https://BioRender.com.

evolutionary signatures—positive selection, higher rate of evolution, and unique amino acid substitution with functional impact. In *E. variegata* species, 229 genes had a higher rate of evolution, 707 genes were positively selected ($P < 0.01$), and 836 genes contained unique amino acid substitution with functional impact. Additionally, 237 genes were identified as MSA genes, and 13 genes had all three evolutionary signatures in *E. variegata*.

Among the 88 MSA genes in *E. crameri* species, 59 genes (67%) were supported by gene expression data (TPM > 1), of which 37 genes could be assigned KO numbers. Among the 237 MSA genes in *E. variegata* species, 153 genes (64.6%) were supported by gene expression data (TPM > 1), of which 107 genes could be assigned KO numbers. The genes with evolutionary signatures in *E. crameri* and *E. variegata* were distributed in various KEGG pathways and COG categories (Tables S9–13).

3.7.1. Adaptive evolution of metabolism-related genes in *Eumeta* species

Metabolic activities are essential in the life cycle of holometabolous insects, including bagworms, since the rate of metabolism regulates the insect metamorphosis through different life cycle stages and aids in environmental adaptability.[7,9] In both the *Eumeta* species, genes involved in primary metabolic pathways, such as glycolysis and citric acid cycle, showed evolutionary signatures (Figs. 4a, b). Among the other key metabolic genes showing evolutionary signatures in *E. crameri*, *PEPCK* (gluconeogenesis), phosphoglucomutase (glycogenesis), glucose 1-dehydrogenase (pentose phosphate pathway), *ACAA2*, stearoyl-CoA desaturase, glucuronosyltransferase (fatty acid metabolism), *ASAH2*, *DEGS* (sphingolipid metabolism), *ABAT* (amino acid metabolism) were notable. Similarly, in *E. variegata*, alpha-trehalase, alpha-amylase (starch and sucrose metabolism), *ACAA2*, glucuronosyltransferase, *HSD17B12* (fatty acid metabolism), *ASAH2* (sphingolipid metabolism), and glutathione S-transferase were notable among the other primary metabolism-related genes with evolutionary signatures. These results indicate that bagworms have plausibly evolved the metabolic genes compared to the other Lepidopterans to cope with the energy requirements during food shortages and the nonfeeding life cycle stages during their unique and complex life cycle inside the bags.[9]

3.7.2. Adaptive evolution of genes involved in molting and metamorphosis

Molting and metamorphosis of bagworm moths occur inside the bags in a unique and well-organized manner, involving case renovation events in the larval stages.[5] In support of this, genes related to insect molting and metamorphosis showed evolutionary signatures, such as one chitinase-domain containing gene was MSA gene (TPM > 1), and chitin synthase, two other chitin-binding domain genes, and two insect cuticle genes were positively selected in *E. crameri*. Further, four chitin-binding domain proteins, juvenile hormone binding protein (*JHBP*), and juvenile hormone epoxide hydrolase (*JHEH*) showed unique substitution with functional impact in *E. crameri*. *JHEH* also showed gene family expansion in *E. crameri*. It is important to mention that chitinases aid in chitin metabolism in insects to replace their old cuticles during molting or ecdysis,[95] *JHEH* functions in degradation pathways of the metamorphosis-preventing juvenile hormones in insects,[96] and *JHBP* plays a role in transporting and protection of juvenile hormone.[97] Further, Ecdysteroids are the principal molting hormones in insects, and ecdysteroid kinase (regulating ecdysteroid production) showed unique substitution with functional impact, and a cytochrome P450 gene family (catalyses 20-hydroxyecdysone biosynthesis) was highly expanded in *E. crameri*.[98]

Similarly, in *E. variegata*, one chitin metabolism gene showed a higher rate of evolution, and chitin synthase, two chitin-binding domain proteins, and one insect cuticle gene were positively selected. Further, three chitin-binding domain genes, two chitin metabolism genes, *JHEH* and *JHBP*, showed unique substitution with functional impact. Ecdysteroid kinase showed unique substitution with functional impact in *E. variegata*, similar to *E. crameri*.

3.8. Silk genes of *E. crameri*

Silk production is an attractive characteristic in bagworm species required for case construction. Fibroin heavy chain (H-Fib) is the major silk gene among the genes encoding structural silk proteins.[99] The bagworm H-Fib gene has a very repetitive domain, similar to other species. In the repetitive domain of the *E. crameri* H-Fib gene (Fig. S7), poly-alanine [A_n] motif and the alternating glycine-alanine [$(GA)_n$] motif were found, similar to *E. variegata*, combining the characteristics of both Bombycidae and Saturniidae H-Fib genes.[15] This unique arrangement of repetitive motifs is responsible for the high tensile strength of the bagworm silk.[15,100] The presence of the GGYG motif was also observed in this repetitive domain of the *E. crameri* H-Fib gene. Further, the fibrohexamerin (P25) gene encoding another protein that forms the core of the silk fibre was also found in the *E. crameri* genome, similar to *E. variegata*[15] (Fig. S8).

Apart from these well-known structural silk genes, a previous study combined silk-gland transcriptomic and silk proteomic analysis to verify the presence of a few other silk gene candidates in Lepidopteran species,[101] which showed evolutionary signatures in *Eumeta* species. In *E. crameri*, sulphatase and peroxidase genes showed unique substitutions with functional impact and positive selection, and antichymotrypsin 2-like gene showed unique substitution. Additionally, mucin 2 and trypsin-like protease gene families were highly expanded in *E. crameri*. Zonadhesin, antichymotrypsin 2-like, and serine protease gene families also showed gene family expansion in this species. Similarly, *E. variegata* also showed evolutionary signatures in silk component genes—antichymotrypsin 2-like gene was positively selected, sulphatase and trypsin-like genes showed higher rate of evolution and unique substitution with functional impact, peroxidase and venom protease genes showed unique substitution with functional impact. Further, sulphatase, trypsin-like, and mucin 2 gene families showed gene expansion, and the antichymotrypsin 2-like family was contracted in *E. variegata*.

3.9. Amplicon and shotgun metagenome analyses revealed the taxonomic composition of the gut microbiome

Taxonomic classification of filtered shotgun metagenomic reads (25,961,105 reads) using Kraken identified nine phyla with >1% relative abundance in all six individuals. The Pseudomonadota phylum showed the highest relative abundance (mean 30.1%) in all individuals, followed by Bacillota (mean 26.27%) and Bacteroidota (mean 13.89%) (Fig. 5). The abundance of these bacterial phyla in the gut microbiota was also the case for the silkworm and other Lepidopteran species, where they aid in host metabolism and defense responses.[102,103] From the Kraken analysis, *Clostridium*, belonging to the phylum Bacillota, was the most abundant genus in all individuals (in the range of 6.13%—6.84%). *Clostridium* was also found to be abundant in the gut microbiota of other Lepidopteran insects such as *Spodoptera littoralis* and *Helicoverpa armigera*,[104] where it showed high metabolic activity in the insect gut aiding in the host's digestion and metabolism and enhancing host immunity.[105]

A total of 146,313,980 (mean 7,700,736 ± 6,987,425) paired-end reads were generated from 16S rRNA amplicon sequencing of 19 *E. crameri* larvae. In QIIME 2 analysis, 2,635 Amplicon Sequence Variants (ASVs) were detected across all 19 samples after denoising, which were used for analysing the taxonomic composition and phylum abundance. Taxonomic classification of the ASV sequences after excluding the unassigned sequences and sequences assigned to "Chloroplast" helped to identify 21 phyla in *E. crameri* individuals (present in at least one individual), with Bacillota and Pseudomonadota as the most abundant phyla.

3.10. Metagenomic analysis revealed the gut microbial gene catalogue

A total of 25,961,105 (mean 4,326,851 ± 794,647) shotgun metagenome sequencing reads (both paired-end and unpaired) for six *E. crameri* larvae were used for assembly. Metagenome assembly and subsequent filtering steps resulted in an average of 81,528 contigs per sample and 113,336 contigs for the co-assembly (≥300 bases), which were used for gene prediction. A total of 166,427 non-redundant genes were identified from all individuals, of which 72,390 genes were retained after length-based filtering (≥150 nucleotides) and further contamination removal to represent the first gut microbial gene catalogue of *E. crameri* species.

Functional annotation of the gene catalogue showed the abundance of various metabolic pathways-related genes (Fig. S9) and the presence of 488 CAZy families. 3,405 KOs and 1,445 ECs were identified in this gene catalogue, among which K05577 (*ndhF*—NAD(P)H-quinone oxidoreductase subunit 5) was the most abundant KO (in the range of 1.63%—2.15%) in all the samples, and EC:1.6.5.3 (NADH:

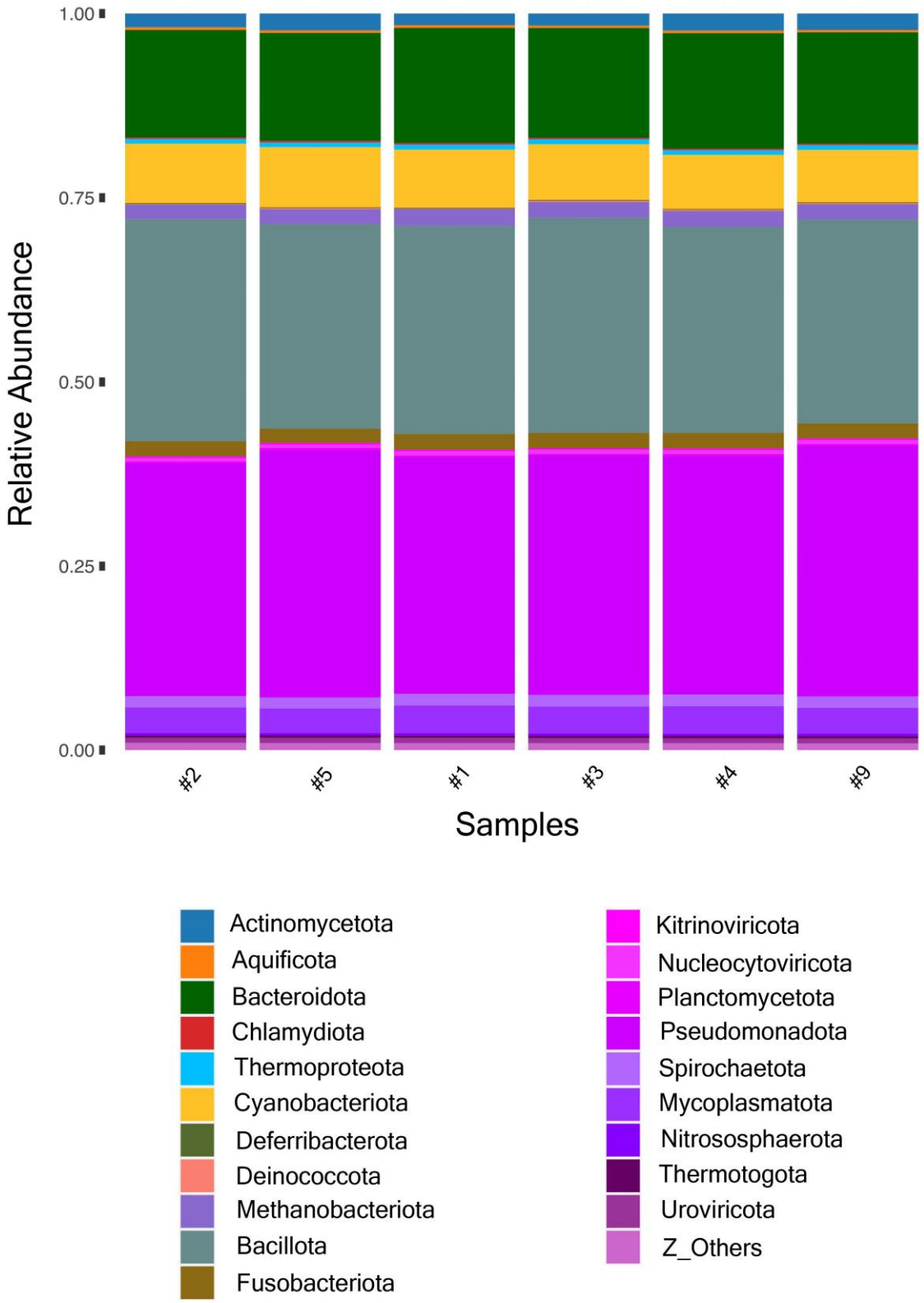

Fig. 5. Relative abundance of the microbial phyla obtained from the Kraken analysis of shotgun metagenomic reads.

ubiquinone reductase) was the most abundant EC (in the range of 3.35%—4.3%) in all the samples. *ndhF* is a component of electron transport chain (ETC) complex I and thus functions in oxidative phosphorylation.[106] Further, among the highly abundant top 20 KOs and ECs, 17 KOs and 13 ECs were common in all the samples (Table S14). Among these 17 KOs, 11 were involved in key metabolic pathways such as oxidative phosphorylation.

3.11. CAZy analysis showed the unique composition of the gut microbiome and genome

GH13 (Glycoside Hydrolase Family 13) was the most abundant CAZy family (26.02%) in the gut microbial gene catalogue, and CE10 (Carbohydrate Esterase Family 10) was the most abundant CAZy family (21.19%) in the *E. crameri* coding gene set. Overall, Glycoside Hydrolase (GH) gene families were the most abundant (57.38%) CAZy families in the gut microbiome of *E. crameri*. In contrast, glycosyltransferases (GTs) were the most abundant (52.12%) in the *E. crameri* genome.

We identified 27 CAZy families in the *E. crameri* gut microbiome that were not present in the host genome. Similarly, the *E. crameri* genome contained 35 unique CAZy families (Table S15). Among the unique CAZy families, GT51 (12.09%) and GH3 (6.76%) were the most abundant in the gut microbiome, and GT31 (6.36%) and GT27 (4.24%) were the most abundant in the genome.

3.12. SEM imaging of silk fibres

Silk is the key structural component of the case constructed by the bagworm species. Degummed bags of *E. crameri* showed the presence of long and fine silk fibres. Silk threads had a clean and smooth surface consisting of a pair of silk fibres. Each one of this pair had an average fibre diameter of 3.09 ± 0.43 μm (mean ± sd) at a voltage of 20 kV (Fig. 6), which is similar to those of other bagworm species (*E. minuscula* = ~2 μm,[107] *Theriodopteryx ephemeraeformis* = 2.9 ± 1 μm[108]) and thinner than the silkworm species (10–16 μm[109]).

4. Discussion

Bagworm moths are commonly known for their case-forming behaviour and production of silk fibres. Although the bagworm family consists of more than 1,000 species[4] and is known for their ultrafine silk fibres with high tensile strength,[15,108] the genome sequences of only four bagworm species have been reported so far.[15,20] In this study, we performed Illumina short read and 10× Genomics synthetic long read sequencing to construct the scaffold-level genome assembly of *E. crameri* species with an N50 value of 6.6 Mb, which is 20 times greater than the previously assembled bagworm (*E. variegata*) genome with an N50 value of 324 Kb.[15] The presence of >90% complete BUSCOs and >90% read mapping statistics also attested to the genome assembly quality. Further, we performed the mitochondrial genome assembly and annotation of *E. crameri*, which will be helpful for future evolutionary studies.

The percentage of repetitive elements present in the genome assembly of *E. crameri* and *E. variegata* was similar to other Lepidopteran insects such as *B. mori*.[110] The identification of the coding genes in *E. crameri* genome with a combination of ab initio and homology-based approaches in the MAKER pipeline with stringent parameters helped in construction of a high-confidence gene set. The number of coding genes identified for *E. crameri* species in this study was similar to other Lepidopteran insects (such as *B. mori*, *D. plexippus*, and *H. melpomene*).

However, since the coding gene set of *E. variegata* had several prediction artefacts as mentioned in the previous study,[15] we performed the coding gene prediction for *E. variegata* species using the previously available genome assembly[15] and identified a similar number of genes as *E. crameri*. Using the coding gene sets of both the species constructed in this study, a comprehensive genome-wide phylogeny across 55 species spanning seven insect orders was constructed, which will be helpful for future evolutionary studies. Further, the usage of species from Lepidoptera order helped in the precise identification of the genes that exhibited adaptive evolution in both the *Eumeta* species since the genetic distance of the selected species may affect such analyses.[70] The adaptively evolved genes of the two bagworm species (*E. crameri* and *E. variegata*) were enriched in pathways related to metabolism and defense ($p\_adj < 0.01$, Figs. S10 and S11), which was also supported by the gene expression data ($TPM > 1$).

Bagworm moths are holometabolous insects that have four stages in the life cycle—egg, larva, pupa, and adult. However, bagworms uniquely complete their life cycle inside the bags. The metabolic activities are essential for insect metamorphosis since protein, carbohydrate, and lipid levels vary in different stages of the insect life cycle.[8] For example, during the nonfeeding pupal stage, lipid and carbohydrate interconversion is essential for energy gain.[7] Further, during the life cycle, insects (including bagworms) often face food shortages, ie, starvation stress, and in this time period, insects regulate the physiological processes such as carbohydrate and lipid metabolism to obtain energy.[9] Bagworm moths undergo diapause in the egg stage inside the bag,[4,111] and in the diapause stage, fats are the primary food reserve.[112] Further, the interactions between nutrient storage and metabolism influence the diapause time period, and the energy expenditure in this stage also determines post-diapause fitness.[112] In this study, one of the key findings was the enrichment of *E. crameri* coding genes ($p\_adj < 0.01$) associated with various metabolic pathways such as carbohydrate, fatty acid, and lipid metabolism (Fig. S4, Table S7). Further, we found that the genes involved in primary metabolic processes such as carbohydrate and lipid metabolism were adaptively evolved in both the *Eumeta* species (Figs. 4a, b, S10, and 11, Tables S9–S12). These findings could explain the functional and evolutionary importance of metabolic activities in the unique and complex life cycle of the bagworm species compared to other Lepidopterans.

Molting and metamorphosis of bagworm moths are well-organized processes inside the bags, with case renovation occurring between two larval instar stages.[2] Ecdysteroids and the juvenile hormone (JH) are the primary hormones that control insect molting and metamorphosis—ecdysteroids induce molting of the cuticle and drive the metamorphic development, whereas JH controls the entry into metamorphosis and regulates the growth of larva.[113] In this study, evolutionary signatures were observed in chitinases, chitin-binding domain-containing genes, genes regulating the degradation and transport of JH, and genes regulating ecdysteroid levels. Further, the most abundant CAZy family found in *E. crameri* genome, ie, CE10, is involved in juvenile hormone metabolism.[114] Taken together, these results hint towards the genomic basis of the well-organized life cycle of *Eumeta* species.

Bagworm moths, including the *Eumeta* species, were also described as pest species that damage and affect the growth of the host plant at different stages.[10] Such Lepidopteran insect pests possess a higher number of detoxification genes such as glutathione S-transferase (*GST*), cytochrome P450, ABC transporters, and carboxylesterase (*COE*), and show resistance to insecticides.[110] This study also found a cytochrome P450 family (highly expanded), a *COE* family (highly expanded), and *GST* family to be expanded in *E. crameri*, and a cytochrome P450 family and an ABC transporter gene

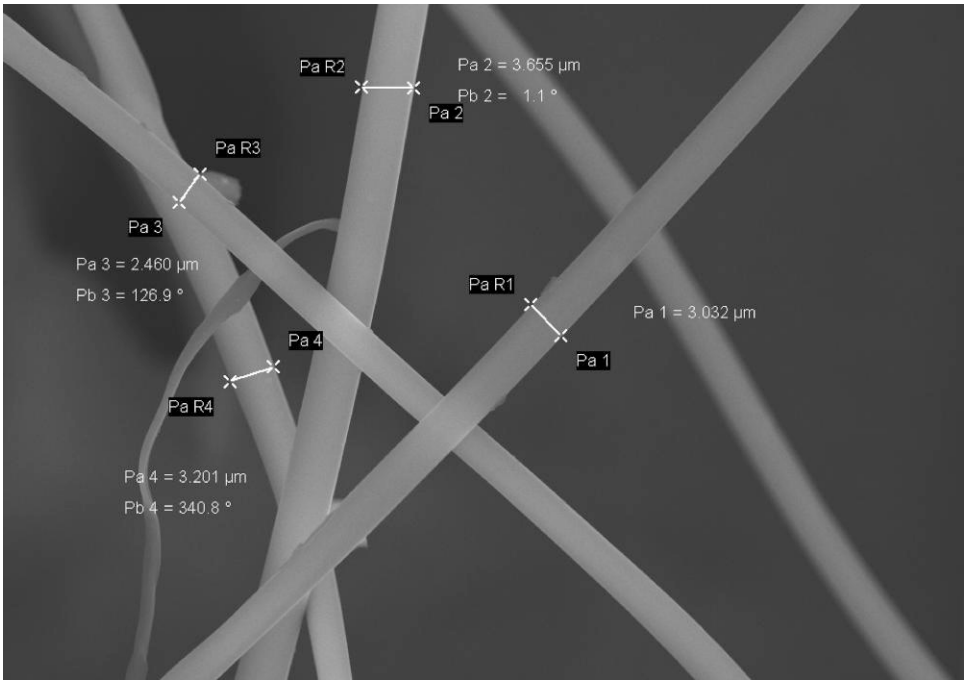

Fig. 6. Scanning electron microscope (SEM) image of *E. crameri* silk fibres (magnification = 3 K X, EHT = 20 kV, WD = 7 mm).

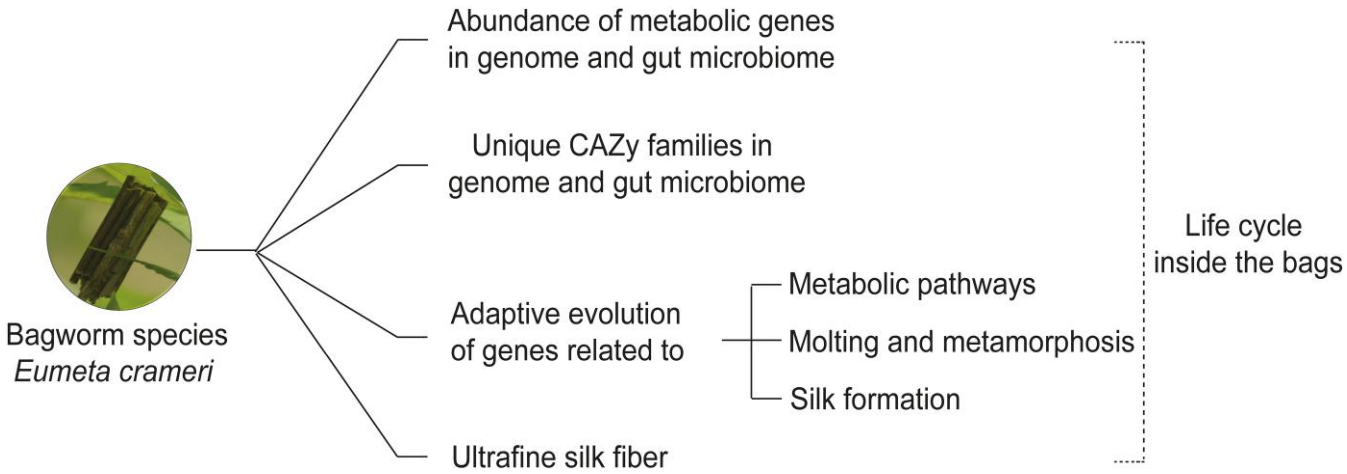

Fig. 7. Factors assisting the life cycle of bagworm species.

subfamily to be highly expanded in *E. variegata* (Table S8). Further, the highly expanded DDE superfamily in both the *Eumeta* species also aids in their adaptation to environmental stresses.[115] Expansion of these gene families results in the detoxification of a broader range of insecticides and plant toxins and environmental adaptations, which plays a critical role in the polyphagous behaviour of bagworm species.[4,110]

Silk is a natural and versatile biomaterial mainly observed in Arthropods, and bagworm moths uniquely use this silk to construct their surrounding bag in larval stages and to hang pupae using the silk attachments. In this study, we found that the candidate genes required for silk formation in Lepidopteran insects showed evolutionary signatures and gene family evolution in *Eumeta* species that can plausibly explain their silk-producing and case-forming behaviour required for survival. Previous studies on *E. variegata* silk had revealed that toughness, high tensile strength, and other mechanical properties are similar to silkworm moths and other saturniids, which have evolved to protect and support the larva, and proposed that bagworm silk can be considered for sustainable structural materials.[15,100] In this study, we observed the ultrafine nature of *E. crameri* silk (SEM image) extracted from its bag, similar to other bagworm species.[100,108] Measurements of different physical parameters are required to strengthen the structural properties of *E. crameri* silk. However, noting the similarity in the adaptive evolution of silk candidate genes and the similar ultrafine nature of the silk in both the *Eumeta* species, it is tempting to speculate that the mechanical properties of *E. crameri* silk could have evolved in a similar manner.

One of the objectives of this study was the investigation of the gut microbiota composition of *E. crameri* larvae. Pseudomonadota was the most abundant phylum from the

Kraken analysis, which was also the case in many other Lepidopterans such as *Metisa plana* (another bagworm species), *B. mori*, *Acronicta major*, *Diaphania pyloalis*, and *Spodoptera frugiperda*.[102,116,117] The other two abundant phyla obtained from our analysis—Bacillota and Bacteroidota were also found among the abundant phyla in the gut microbiota of these Lepidopterans.[102,116,117] The comparatively lower abundance of the phyla Actinomycetota, Methanobacteriota, and Spirochaetota in the *E. crameri* gut microbiota is also similar to *S. frugiperda*.[117]

Similar to other Lepidopterans, *E. crameri* gut contains a large microbial community, which provides metabolic, defense, and other benefits to the host species.[18,103] The gut microbiome of Lepidopteran insects aids in host immunity by creating a chemical barrier against the invading bacterial species,[103] which was also supported by the fact that ABC transporter genes (function in insecticide resistance) were abundant in the *E. crameri* gut microbial gene catalogue (Fig. S9). In this study, we found an abundance of gut microbial genes related to various metabolic pathways, which correlates with the results obtained from the genomic analysis and is further supported by the abundance of distinct CAZy families in the genome and gut microbiome. Metabolic genes were also a significant fraction of the silkworm gut microbiome, as observed in a previous study.[118]

We found the distinct composition of the *E. crameri* gut microbiome and genome from the CAZy analysis, which were predominated by the GHs and the GTs, respectively. GHs, such as the most abundant GH13 in *E. crameri* gut microbiome, benefit the host by degrading the insect-indigestible plant polysaccharides,[119] similar to the termite gut microbiome.[120] In contrast, GTs are primarily involved in insect survival and developmental processes by aiding in tissue differentiation, metamorphosis, and detoxification.[121] In this study, we also found the unique CAZy families in the *E. crameri* gut microbiome that were not present in the genome and vice versa. The plant cellular component-degrading (eg, cell wall) gene families such as GH3, GH77, GH5, GH51, GH43, and others (Table S15) were noteworthy among the unique CAZy families present in the gut microbiome.[122] In the genome, the two unique and most abundant CAZy families, GT31, and GT27, are involved in insect developmental processes.[121] These findings indicate that the *E. crameri* gut microbiome supports the host species by providing additional genes for the metabolism of plant components, which is essential for the survival of the bagworm species.

Taken together, although the adaptive evolution observed in the *E. crameri* genes is still putative and requires further investigation using population polymorphism data, it is tempting to speculate that bagworm species have evolved the metabolic pathways, molting and metamorphosis processes, and silk-forming mechanisms, which have assisted in their holometabolous metamorphosis inside the bag constructed using silk and plant materials. Further, the gut microbiome of *E. crameri* species, with its unique genes, also aids in defense and metabolism of plant material, and thus adds additional support for their survival inside the bags (Fig. 7). Although the precise role of gut microbiome on silk production is not clearly understood to date, studies have shown that the gut microbes can affect the commercial traits of silkworm species such as cocoon quality, silk length and size, and suggested that the application of probiotics could improve silk productivity.[123,124] Therefore, the availability of the genome sequence of bagworm species *E. crameri* and the functional and evolutionary insights obtained from the genomic, transcriptomic, and gut microbiome analysis in this study will serve as valuable resources for future studies on the Psychidae family and other Lepidopteran insects.

Acknowledgments

A.C. and S.M. thank the Council of Scientific and Industrial Research (CSIR) for the fellowship. V.P.P.K. and A.S.K. thank DST-INSPIRE for the fellowship. We thank the SEM imaging facility at IISER Bhopal.

Author contributions

V.K.S. conceived and coordinated the project. A.S.K. and A.C. collected the samples. S.M. performed genomic DNA-RNA extraction, species identification assay, and microbial DNA extraction and prepared the genomic and metagenomic samples for sequencing. A.C. and V.K.S. designed the computational framework of genomic analysis. A.C., V.P.P.K., and V.K.S. designed the computational framework of metagenomic analysis. A.C. performed all the genomic analyses presented in the study and functional annotation of gene sets. A.C. and V.P.P.K. performed all shotgun metagenome and 16S rRNA amplicon analyses presented in the study and constructed the figures. A.S.K. and A.C. performed the sample preparation for the SEM imaging of silk material. A.C., V.K.S., and V.P.P.K. interpreted the results. A.C., V.K.S., and S.M. wrote the manuscript. All the authors have read and approved the final version of the manuscript.

Supplementary material

Supplementary data are available at DNARES online.

Funding

The authors thank the intramural research funds provided by IISER Bhopal and acknowledge the grant by Science and Engineering Research Board (Grant number - CRG/2022/001833).

Conflict of interest

None declared.

Data availability

The whole genome, transcriptome, and gut microbiome sequencing reads of *E. crameri* species have been deposited in the NCBI SRA database under the BioProject accession PRJNA1051493. The genome and transcriptome assemblies, coding gene set, gut microbial gene catalogue, and gene identifiers for the genes showing adaptive evolution are available at https://metabiosys.iiserb.ac.in/Bagworm. The genome and transcriptome assemblies are also available at the Indian Biological Data Centre with INDA accession INRP000468.

Ethics declarations

This study was approved by the Institutional Ethics Committee (IEC), IISER Bhopal, India.

References

1. Shah SK, Biswas O, Mazumder A, Banerjee A. Indian Bagworms (Insecta: Lepidoptera: Psychidae). Zoological Survey of India; 2015.
2. Agrawal A, Pati AK. Larval case renovation—a unique behaviour in bagworm moth, *Eumeta crameri* Westwood. Curr Sci. 2003:85: 1674–1679. https://www.jstor.org/stable/24109970
3. Sugiura S. Bagworm bags as portable armour against invertebrate predators. PeerJ. 2016:2016:e1686. https://doi.org/10.7717/peerj.1686
4. Rhainds M, Davis DR, Price PW. Bionomics of bagworms (Lepidoptera: Psychidae). Annu Rev Entomol. 2009:54:209–226. https://doi.org/10.1146/annurev.ento.54.110807.090448
5. Agrawal A, Pati AK. Increment in the size of an inanimate object in the case of bagworm moth, *Eumeta crameri* Westwood (Lepidoptera: Psychidae) obeys Dyar's law. Curr Sci. 2002:83: 71–73. https://www.jstor.org/stable/24106080
6. Pati AK, Agrawal A. Hierarchical perception of stimuli during case construction in the bagworm moth *Eumeta crameri* (Lepidoptera: Psychidae). J Insect Behav. 2000:13:667–677. https://doi.org/10.1023/A:1007839926460
7. Agrell I. The aerobic and anaerobic utilization of metabolic energy during insect metamorphosis. Acta Physiol Scand. 1953:28: 306–335. https://doi.org/10.1111/j.1748-1716.1953.tb00984.x
8. Gilbert L, Schneiderman H. Some biochemical aspects of insect metamorphosis. American Zoologist. 1961:1:11–51. https://doi.org/10.1093/icb/1.1.11
9. Zhang DW, Xiao ZJ, Zeng BP, Li K, Tang YL. Insect behavior and physiological adaptation mechanisms under starvation stress. Front Physiol. 2019:10:163. https://doi.org/10.3389/fphys.2019.00163
10. Kumar V, Anal AKD, Nath V. Prevalence of some threatening pests and disease of Litchi (*Litchi chinensis* Sonn). in Bihar state of India. J Appl Hortic. 2014:16:235–240. https://doi.org/10.37855/jah.2014.v16i03.39
11. Arakawa K et al. 1000 spider silkomes: linking sequences to silk physical properties. Sci Adv. 2022:8:6043. https://doi.org/10.1126/sciadv.abo6043
12. McKim SA, Turner TL. The evolution of silk production in Crustacea. J Crustac Biol. 2024:44:ruae056. https://doi.org/10.1093/jcbiol/ruae056
13. Sutherland TD, Young JH, Weisman S, Hayashi CY, Merritt DJ. Insect silk: one name, many materials. Annu Rev Entomol. 2010:55:171–188. https://doi.org/10.1146/annurev-ento-112408-085401
14. Craig CL. Evolution of arthropod silks. Annu Rev Entomol. 1997:42:231–267. https://doi.org/10.1146/annurev.ento.42.1.231
15. Kono N et al. The bagworm genome reveals a unique fibroin gene that provides high tensile strength. Commun Biol. 2019:2:148. https://doi.org/10.1038/s42003-019-0412-8
16. Malay AD et al. Relationships between physical properties and sequence in silkworm silks. Sci Rep. 2016:6:1–11. https://doi.org/10.1038/srep27573
17. Lewis Z, Lizé A. Insect behaviour and the microbiome. Curr Opin Insect Sci. 2015:9:86–90. https://doi.org/10.1016/j.cois.2015.03.003
18. Voirol LRP, Frago E, Kaltenpoth M, Hilker M, Fatouros NE. Bacterial symbionts in lepidoptera: their diversity, transmission, and impact on the host. Front Microbiol. 2018:9:556. https://doi.org/10.3389/fmicb.2018.00556
19. Engel P, Martinson VG, Moran NA. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 2012:109:11002–11007. https://doi.org/10.1073/pnas.1202970109
20. Whiteford S. The genome sequence of the virgin bagworm, *Luffia ferchaultella* (Stephens, 1850). Wellcome Open Res. 2025:10:108. https://doi.org/10.12688/wellcomeopenres.23768.1
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014:30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
22. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011:30: 2114–2120. https://doi.org/10.1093/bioinformatics/btr011
23. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432. https://doi.org/10.1038/s41467-020-14998-3
24. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014:30:31–37. https://doi.org/10.1093/bioinformatics/btt310
25. Bankevich A et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012:19: 455–477. https://doi.org/10.1089/cmb.2012.0021
26. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017:27: 757–767. https://doi.org/10.1101/gr.214874.116
27. Jackman SD et al. Tigmint: correcting assembly errors using linked reads from large molecules. BMC Bioinformatics. 2018:19:393. https://doi.org/10.1186/s12859-018-2425-6
28. Kajitani R et al. Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. Nat Commun. 2019:10:1702. https://doi.org/10.1038/s41467-019-09575-2
29. Yeo S, Coombe L, Warren RL, Chu J, Birol I. ARCS: scaffolding genome drafts with linked reads. Bioinformatics. 2018:34: 725–731 https://doi.org/10.1093/bioinformatics/btx675
30. Zhang SV, Zhuo L, Hahn MW. AGOUTI: improving genome assembly and annotation using transcriptome data. Gigascience. 2016:5:131. https://doi.org/10.1186/s13742-016-0136-3
31. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. 2016:44: e147. https://doi.org/10.1093/nar/gkw654
32. Bisht MS, Mahajan S, Chakraborty A, Sharma VK. A high-quality genome assembly of *Annona squamosa* (custard apple) provides functional insights into an emerging fruit crop. DNA Res. 2025:32:dsaf007. https://doi.org/10.1093/dnares/dsaf007
33. Luo R et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 2012:27:18. https://doi.org/10.1186/2047-217X-1-18
34. Walker BJ et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014:9:e112963. https://doi.org/10.1371/journal.pone.0112963
35. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. G3 Genes|Genomes|Genetics. 2020:10:1361–1374. https://doi.org/10.1534/g3.119.400908
36. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. [preprint]. arXiv 2013; arXiv:1303.3997. doi: https://doi.org/10.48550/arXiv.1303.3997
37. Grabherr MG et al. Genome-wide synteny through highly sensitive sequence alignment: satsuma. Bioinformatics. 2010:26: 1145–1151. https://doi.org/10.1093/bioinformatics/btq102
38. Chakraborty A, et al. Genome sequencing and de novo and reference-based genome assemblies of *Bos indicus* breeds. Genes & Genomics. 2023:45:1399–1408. https://doi.org/10.1007/s13258-023-01401-w
39. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351
40. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019:378:907–915. https://doi.org/10.1038/s41587-019-0201-4
41. Jin JJ et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020:21: 1–31. https://doi.org/10.1186/s13059-020-02154-5

42. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019:47:W59–W64. https://doi.org/10.1093/nar/gkz238
43. Flynn JM et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117
44. Haas BJ et al. De novo transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013:8:1494–1512. https://doi.org/10.1038/nprot.2013.084
45. Yates AD et al. Ensembl genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Res. 2022:50:D996–D1003. https://doi.org/10.1093/nar/gkab1007
46. Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics. 2014:48:4–11. https://doi.org/10.1002/0471250953.bi0411s48
47. Stanke M et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006:34:W435–W439. https://doi.org/10.1093/nar/gkl200
48. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000:28:45–48. https://doi.org/10.1093/nar/28.1.45
49. Bateman A. The Pfam protein families database. Nucleic Acids Res. 2004:32:138D–1141. https://doi.org/10.1093/nar/gkh121
50. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011:32:138D–1141. https://doi.org/10.1093/nar/gkr367
51. Chan PP, Lin BY, Mak AJ, Lowe TM. TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021:49:9077–9096. https://doi.org/10.1093/nar/gkab688
52. Fromm B et al. MirGeneDB 2.1: toward a complete sampling of all major animal phyla. Nucleic Acids Res. 2022:50:D204–D210. https://doi.org/10.1093/nar/gkab1101
53. Wang Y et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012:40:e49–e49. https://doi.org/10.1093/nar/gkr1293
54. Bisht MS, Singh M, Chakraborty A, Sharma VK. Genome of the most noxious weed water hyacinth (*Eichhornia crassipes*) provides insights into plant invasiveness and its translational potential. iScience. 2024:27:110698. https://doi.org/10.1016/j.isci.2024.110698
55. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-Seq quantification. Nat Biotechnol. 2016:345:525–527. https://doi.org/10.1038/nbt.3519
56. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011:475:493–496. https://doi.org/10.1038/nature10231
57. Li H et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009:25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
58. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011:27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509
59. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing and comparative analysis of *Ficus benghalensis* and *Ficus religiosa* species reveal evolutionary mechanisms of longevity. iScience. 2022:25:105100. https://doi.org/10.1016/j.isci.2022.105100
60. Keightley PD et al. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. Mol Biol Evol. 2015:32:239–243. https://doi.org/10.1093/molbev/msu302
61. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y
62. Laetsch DR, Blaxter ML. KinFin: software for taxon-aware analysis of clustered protein sequences. G3 Genes, Genomes, Genet. 2017:7:3349–3357. https://doi.org/10.1534/g3.117.300233
63. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30:772–780. https://doi.org/10.1093/molbev/mst010
64. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014:30:772–780. https://doi.org/10.1093/bioinformatics/btu033
65. Jombart T, Dray S. Adephylo: exploratory analyses for the phylogenetic comparative method. Bioinformatics. 2010: 26:1907–1909. https://doi.org/10.1093/bioinformatics/btq292
66. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003:31:3812–3814. https://doi.org/10.1093/nar/gkg509
67. Agaba M et al. Giraffe genome sequence reveals clues to its unique morphology and physiology. Nat Commun. 2016:7:1–8. https://doi.org/10.1038/ncomms11519
68. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006:34:W609–W612. https://doi.org/10.1093/nar/gkl315
69. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:24:1586–1591. https://doi.org/10.1093/molbev/msm088
70. Chakraborty A, Mondal S, Mahajan S, Sharma VK. High-quality genome assemblies provide clues on the evolutionary advantage of blue peafowl over green peafowl. Heliyon. 2023:9:e18571. https://doi.org/10.1016/j.heliyon.2023.e18571
71. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties. Front Plant Sci. 2023:14:1260414. https://doi.org/10.3389/fpls.2023.1260414
72. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020:36:5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022
73. Chakraborty A, Mahajan S, Jaiswal SK, Sharma VK. Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. Commun Biol. 2021:4:1–12. https://doi.org/10.1038/s42003-021-02720-y
74. Kumar S et al. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol. 2022:39:1–6. https://doi.org/10.1093/molbev/msac174
75. Mahajan S, Chakraborty A, Bisht MS, Sil T, Sharma VK. Genome sequencing and functional analysis of a multipurpose medicinal herb *Tinospora cordifolia* (Giloy). Sci Rep. 2024:14:1–17. https://doi.org/10.1038/s41598-024-53176-z
76. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019:20:1–13. https://doi.org/10.1186/s13059-019-1891-0
77. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017:27:824–834. https://doi.org/10.1101/gr.213959.116
78. Von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol. 2019:20:1–14. https://doi.org/10.1186/s13059-019-1817-x
79. Hyatt D et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010:11:1–11. https://doi.org/10.1186/1471-2105-11-119
80. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006:22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158
81. Li R et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009:25:1966–1967. https://doi.org/10.1093/bioinformatics/btp336

82. Dhakan DB et al. The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. Gigascience. 2019:8:1–20. https://doi.org/10.1093/gigascience/giz004
83. Bolyen E et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019:378:852–857. https://doi.org/10.1038/s41587-019-0209-9
84. Prasoodanan PKV et al. Metagenomic exploration of Andaman region of the Indian Ocean. Sci Rep. 2024:14:1–16. https://doi.org/10.1038/s41598-024-53190-1
85. Callahan BJ et al. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016:137:581–583. https://doi.org/10.1038/nmeth.3869
86. Quast C et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013:41:D590–D596. https://doi.org/10.1093/nar/gks1219
87. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007:35:W182–W185. https://doi.org/10.1093/nar/gkm321
88. Huerta-Cepas J et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017:34:2115–2122. https://doi.org/10.1093/molbev/msx148
89. Arakawa K, Kono N, Ohtoshi R, Nakamura H, Tomita M. The complete mitochondrial genome of *Eumeta variegata* (Lepidoptera: Psychidae). Mitochondrial DNA: Part B Resour. 2018:3:812–813. https://doi.org/10.1080/23802359.2018.1495119
90. Sun J et al. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res. 2023:51:W397–W403. https://doi.org/10.1093/nar/gkad313
91. Wang S et al. The evolution and diversification of oakleaf butterflies. Cell. 2022:185:3138–3152.e20. https://doi.org/10.1016/j.cell.2022.06.042
92. Zheng L et al. Population genomics provides insights into the genetic diversity and adaptation of the *Pieris rapae* in China. PLoS One. 2023:18:e0294521. https://doi.org/10.1371/journal.pone.0294521
93. Ishiwata K, Sasaki G, Ogawa J, Miyata T, Su ZH. Phylogenetic relationships among insect orders based on three nuclear protein-coding gene sequences. Mol Phylogenet Evol. 2011:58:169–180. https://doi.org/10.1016/j.ympev.2010.11.001
94. Yao PH, Mobarak SH, Yang MF, Hu CX. Differential detoxification enzyme profiles in C-corn strain and R-rice strain of *Spodoptera frugiperda* by comparative genomic analysis: insights into host adaptation. BMC Genomics. 2025:26:1–18. https://doi.org/10.1186/s12864-024-11185-2
95. Merzendorfer H, Zimoch L. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. J Exp Biol. 2003:206:4393–4412. https://doi.org/10.1242/jeb.00709
96. Vasquez DDN et al. Simultaneous silencing of juvenile hormone metabolism genes through RNAi interrupts metamorphosis in the cotton boll weevil. Front Mol Biosci. 2023:10:1073721. https://doi.org/10.3389/fmolb.2023.1073721
97. Niewiadomska-Cimicka A et al. Juvenile hormone binding protein core promoter is TATA-driven with a suppressory element. Biochim Biophys Acta—Gene Regul Mech. 2011:1809:226–235. https://doi.org/10.1016/j.bbagrm.2011.02.001
98. Sonobe H et al. Purification, kinetic characterization, and molecular cloning of a novel enzyme, ecdysteroid 22-kinase. J Biol Chem. 2006:281:29513–29524. https://doi.org/10.1074/jbc.M604035200
99. Heckenhauer J et al. Characterization of the primary structure of the major silk gene, h-fibroin, across caddisfly (Trichoptera) suborders. iScience. 2023:26:107253. https://doi.org/10.1016/j.isci.2023.107253
100. Yoshioka T, Tsubota T, Tashiro K, Jouraku A, Kameda T. A study of the extraordinarily strong and tough silk produced by bagworms. Nat Commun. 2019:10:1469. https://doi.org/10.1038/s41467-019-09350-3
101. Volenikova A, et al. Genome sequence and silkomics of the spindle ermine moth, *Yponomeuta cagnagella*, representing the early diverging lineage of the ditrysian Lepidoptera. Commun Biol. 2022:5:1–13. https://doi.org/10.1038/s42003-022-04240-9
102. Chen B et al. Gut bacterial and fungal communities of the domesticated silkworm (*Bombyx mori*) and wild mulberry-feeding relatives. ISME J. 2018:129:2252–2262. https://doi.org/10.1038/s41396-018-0174-1
103. Zhang X et al. Diversity and functional roles of the gut microbiota in lepidopteran insects. Microorganisms. 2022:10:1234. https://doi.org/10.3390/microorganisms10061234
104. Tang X et al. Complexity and variability of gut commensal microbiota in polyphagous lepidopteran larvae. PLoS One. 2012:7:e36978. https://doi.org/10.1371/journal.pone.0036978
105. Shao Y, Arias-Cordero E, Guo H, Bartram S, Boland W. In vivo pyro-SIP assessing active gut microbiota of the cotton leafworm, *Spodoptera littoralis*. PLoS One. 2014:9:e85948. https://doi.org/10.1371/journal.pone.0085948
106. Battchikova N, Aro EM. Cyanobacterial NDH-1 complexes: multiplicity in function and subunit composition. Physiol Plant. 2007:131:22–32. https://doi.org/10.1111/j.1399-3054.2007.00929.x
107. Wolff JO, et al. Strength of silk attachment to *Ilex chinensis* leaves in the tea bagworm *Eumeta minuscula* (Lepidoptera. Psychidae). J R Soc Interface. 2017:14:20170007. https://doi.org/10.1098/rsif.2017.0007
108. Reddy N, Yang Y. Structure and properties of ultrafine silk fibers produced by *Theriodopteryx ephemeraeformis*. J Mater Sci. 2010:45:6617–6622. https://doi.org/10.1007/s10853-010-4752-5
109. Zhao HP, Feng XQ, Shi HJ. Variability in mechanical properties of *Bombyx mori* silk. Mater Sci Eng C. 2007:27:675–683. https://doi.org/10.1016/j.msec.2006.06.031
110. Kawamoto M et al. High-quality genome assembly of the silkworm, *Bombyx mori*. Insect Biochem Mol Biol. 2019:107:53–62. https://doi.org/10.1016/j.ibmb.2019.02.002
111. Rivers DB, Antonelli AL, Yoder JA. Bags of the bagworm *Thyridopteryx ephemeraeformis* (Lepidoptera: Psychidae) protect diapausing eggs from water loss and chilling injury. Ann Entomol Soc Am. 2002:95:481–486. https://doi.org/10.1603/0013-8746(2002)095[0481:BOTBTE]2.0.CO;2
112. Hahn DA, Denlinger DL. Meeting the energetic demands of insect diapause: nutrient storage and utilization. J Insect Physiol. 2007:53:760–773. https://doi.org/10.1016/j.jinsphys.2007.03.018
113. Truman JW. The evolution of insect metamorphosis. Curr Biol. 2019:29:R1252–R1268. https://doi.org/10.1016/j.cub.2019.10.009
114. Zhang D et al. Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. Insect Mol Biol. 2012:21:235–245. https://doi.org/10.1111/j.1365-2583.2011.01130.x
115. Jiang F, Liang L, Wang J, Zhu S. Chromosome-level genome assembly of *Bactrocera dorsalis* reveals its adaptation and invasion mechanisms. Commun Biol. 2022:5:1–11. https://doi.org/10.1038/s42003-021-02966-6
116. Ting A, Abidin CMRZ, Hamid NH, Azzam G, Salim H. Uncovering the microbiota of bagworm *Metisa plana* (Lepidoptera: Psychidae) in oil palm plantations in Malaysia. Trop Life Sci Res. 2023:34:185–218. https://doi.org/10.21315/tlsr2023.34.1.11
117. Ugwu JA, Liu M, Sun H, Asiegbu FO. Microbiome of the larvae of *Spodoptera frugiperda* (J. E. Smith) (Lepidoptera: Noctuidae) from maize plants. J Appl Entomol. 2020:144:764–776. https://doi.org/10.1111/jen.12821
118. Chen B, et al. Comparative shotgun metagenomic data of the silkworm *Bombyx mori* gut microbiome. Sci Data. 2018:5:1–10. https://doi.org/10.1038/sdata.2018.285
119. Liu N et al. Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. ISME J. 2019:13:104–117. https://doi.org/10.1038/s41396-018-0255-1

120. Marynowska M et al. A holobiont approach towards polysaccharide degradation by the highly compartmentalised gut system of the soil-feeding higher termite *Labiotermes labralis*. BMC Genomics. 2023:24:1–20. https://doi.org/10.1186/s12864-023-09224-5
121. Nagare M, Ayachit M, Agnihotri A, Schwab W, Joshi R. Glycosyltransferases: the multifaceted enzymatic regulator in insects. Insect Mol Biol. 2021:30:123–137. https://doi.org/10.1111/imb.12686
122. Badhan A et al. Mechanistic insights into the digestion of complex dietary fibre by the rumen microbiota using combinatorial high-resolution glycomics and transcriptomic analyses. Comput Struct Biotechnol J. 2022:20:148–164. https://doi.org/10.1016/j.csbj.2021.12.009
123. Sun Z, et al. Effects of transient high temperature treatment on the intestinal flora of the silkworm *Bombyx mori*. Sci Rep. 2017:7:3349. https://doi.org/10.1038/s41598-017-03565-4
124. Anishȧ SP, Yasmin N, Devi SV. Effects of probiotic supplemented diet on growth performance of silkworm *Bombyx mori* and improved characteristics of cocoon and silk. J Pharm Res Int. 2022:34:73–78. https://doi.org/10.9734/jpri/2022/v34i30A36271