Genomics 116 (2024) 110841

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Genome sequencing and assembly of Indian golden silkmoth, *Antheraea assamensis* Helfer (Saturniidae, Lepidoptera)

Himanshu Dubey [a], A.R. Pradeep [a], Kartik Neog [b], Rajal Debnath [a,b], P.J. Aneesha [a], Suraj Kumar Shah [b], Indumathi Kamatchi [a], K.M. Ponnuvel [a], A. Ramesha [a], Kunjupillai Vijayan [c], Upendra Nongthomba [d], Utpal Bora [e], Sivaprasad Vankadara [a], K.M. VijayaKumari [b], Kallare P. Arunkumar [b,*]

[a] Seribiotech Research Laboratory, Central Silk Board, Kodathi, Bangalore, India
[b] Central Muga Eri Research and Training Institute, Central Silk Board, Jorhat, India
[c] Research Co-ordination Section, Central Silk Board, Bangalore, India
[d] Department of Developmental Biology and Genetics, Indian Institute of Science, Bangalore, India
[e] Department of Biosciences and Bioengineering, Indian Institute of Technology, Guwahati, India

ARTICLE INFO

Keywords:
Antheraea assamensis
Genome assembly
Lepidoptera
Saturniids

ABSTRACT

Muga silkworm (*Antheraea assamensis*), one of the economically important wild silkmoths, is unique among saturniid silkmoths. It is confined to the North-eastern part of India. Muga silk has the highest value among the other silks. Unlike other silkmoths, *A. assamensis* has a low chromosome number ($n = 15$), and ZZ/ZO sex chromosome system. Here, we report the first high-quality draft genome of *A. assamensis,* assembled by employing the Illumina and PacBio sequencing platforms. The assembled genome of *A. assamensis* is 501.18 Mb long, with 2697 scaffolds and an N50 of 683.23 Kb. The genome encompasses 18,385 protein-coding genes, 86.29% of which were functionally annotated. Phylogenetic analysis of *A. assamensis* revealed its divergence from other Antheraea species approximately 28.7 million years ago. Moreover, an investigation into detoxification-related gene families, CYP450, GST, and ABC-transporter, revealed a significant expansion in *A. assamensis* as compared to the *Bombyx mori*. This expansion is comparable to *Spodoptera litura*, suggesting adaptive responses linked to the polyphagous behavior observed in these insects. This study provides valuable insights into the molecular basis of evolutionary divergence and adaptations in muga silkmoth. The genome assembly reported in this study will significantly help in the functional genomics studies on *A. assamensis* and other *Antheraea* species along with comparative genomics analyses of Bombycoidea insects.

1. Introduction

The order Lepidoptera includes >160,000 species of which Bombycoid moths comprise silkmoths of economic importance. These silkmoths secrete diverse varieties of silk fibers. Silk production based on these moths, especially *Bombyx mori*, *Antheraea mylitta*, *Antheraea assamensis*, and *Samia ricini,* play a significant role in the rural economies of many developing nations. Native to the North-east part of India and named after the Assamese language word “Muga” for its amber-colored cocoon, the wild silkworm *A. assamensis* is known for its natural golden colour silk fiber with glossy fine texture and durability. The species predominantly thrives in the Brahmaputra valley and adjacent hills, exhibiting a polyphagous nature by primarily feeding on leaves of plants such as Som (*Persia bombycina*) and Soalu (*Litsaea monopetala*) abundant in the region [1]. This luminous golden muga silk has now secured Geographical Indication status, the recognition that confers an intellectual property right that it has its origin in the Assam region of India.

Despite being one of the economically important wild silkmoths, the genome of *A. assamensis* is least understood among the saturniid silkmoths [2]. Though the economic significance of *A. assamensis* is evident; it faces challenges, including decline in population and a depletion in genetic variability [3,4]. Unlike other silkmoths, *A. assamensis* has a low chromosome number ($n = 15$) [5], ZZ/ZO sex chromosome system [6] and fragmented populations with a narrow habitat range, probably experiencing genetic drift. Previous microsatellite marker-based studies have revealed low heterozygosity among muga populations collected

* Corresponding author.
E-mail address: arunkallare.csb@gov.in (K.P. Arunkumar).

https://doi.org/10.1016/j.ygeno.2024.110841
Received 28 July 2023; Received in revised form 19 March 2024; Accepted 3 April 2024
Available online 9 April 2024

from various locations in North-east India [4,7].

The muga raw silk production in India is limited to 261 metric tons (2022−23). A comprehensive study examining genotypic and phenotypic diversity across North-east India is imperative to enhance the muga silkworm's tolerance to pathogens, cocoon quality, fecundity and for addressing commercial traits like silk-fiber quality. Additionally, access to genomic resources is crucial for understanding insect behavior, pupal hibernation, sex determination, and to explore the evolutionary dynamics of saturniid silkmoths [4,7]. Comparative genome analysis of *A. assamensis* with well-studied insect like domesticated silkworm *B. mori* promises insights into the basic biological differences between domesticated and wild silkmoths. In this context, sequencing of the *A. assamensis* genome becomes crucial. In this study, we present a high-quality draft genome assembly of *A. assamensis*, derived from a male moth, aiming to contribute to the molecular understanding of this unique wild silkmoth. The computational analysis carried out has identified several novel genes, laying the foundation for further exploration of *A. assamensis* at the molecular level.

2. Materials and methods

2.1. Source material

Live cocoons of *A. assamensis* were collected from the food plant *Persea bombycina* in the Mokokchung forest of Nagaland, India and allowed to emerge at room temperature. After emergence, male moths were separated and used for genomic DNA isolation and whole genome sequencing. Mokokchung forest is hilly, located at an altitude of 1325 m above sea level and has an annual average temperature of 18 °C, 90% relative humidity and 2039 mm of average rainfall.

2.2. Library preparation and sequencing

The genomic DNA was isolated using DNeasy Blood and Tissue kit (Qiagen). The purity and integrity of the extracted genomic DNA was analyzed by Nanodrop (Thermo Scientific, USA) and agarose gel electrophoresis. High quality genomic DNA was used for whole genome sequencing on Illumina and PacBio sequencing platforms. For whole genome short-read sequencing on the Illumina platform, two libraries with fragment sizes of 250–550 bp and 300–700 bp were prepared with NEXTflex Rapid DNA sequencing kit (PerkinElmer, U.S.A.). The fragment size distribution of the short-read libraries was checked on Agilent TapeStation (Agilent, USA). To supplement the genome assembly process, two Illumina mate-pair and one SMRT bell PacBio libraries were constructed and sequenced on respective platforms. Illumina mate-pair libraries with 5–7 Kb and 7–10 Kb sizes were prepared with Nextera Mate Pair Library Preparation Kit (Illumina Inc., Austin, TX, USA). All the sequencing experiments were performed at Genotypic Technology Pvt. Ltd., Bangalore, India.

2.3. Qualitative analysis of raw reads

The quality of raw Illumina reads was checked using the FastQC tool [8] and TrimGalore-0.4.4 [9] was used for clipping of adapters and low-quality bases (Phred score < Q30) with a minimum read length of 50 bp. The mate-pair libraries were processed using NextClip [10] and Cutadapt [11] tools with a cut-off of Q30 and a minimum read length of 50 bp.

2.4. De-novo genome assembly and completeness assessment

The high quality processed paired-end, mate-pair and PacBio reads were used for the genome assembly of *A. assamensis*. Before proceeding with the genome assembly, processed reads were checked for contamination of microbial reads. The initial genome assembly was performed with different assemblers like Platanus [12], ABySS [13], SOAPdenovo [14], Minia [15] and MaSuRCA hybrid assembler [16]. Based on the N50 value and the number of contigs, genome assembly generated by the MaSuRCA hybrid assembler [16] was selected for the final assembly process. The final assembly was obtained by the GapCloser program [14] with paired-end and mate-pair libraries. To assess the completeness of assembled *A. assamensis* genome Universal Single-Copy Orthologs detection method implemented in the BUSCO (v5.2.2) pipeline [17] was utilized. Estimation of genome size and heterozygosity was carried out using Jellyfish v2.3.0 [18] and GenomeScope [19] programs with a k-mer size of 21 nucleotides.

2.5. Annotation of repetitive elements

To annotate repetitive elements in the *A. assamensis* genome, a *de-novo* repeat database was built using Extensive de-novo TE Annotator [20] and RepeatModeler2 programs [21]. A consensus custom *de-novo* repeat library of *A. assamensis* was then combined with the known repeats in RepBase repeat library edition 18 [22]. The combined repeat library was used for comprehensive repeat and transposable element masking in the *A. assamensis* using RepeatMasker program (v4.1.2) [23].

2.6. Gene prediction and functional annotation

To identify protein coding regions in *A. assamensis* genome, BRAKER2 pipeline [24] was employed. The pipeline integrates predictions from *ab-initio* gene prediction tools such as GeneMark-EP+ [25] and AUGUSTUS [26] with evidence based gene annotation. The evidence to identify the coding region in the *A. assamensis* genome was generated by two different approaches. Firstly, fourteen RNA-seq libraries available for *A. assamensis* at the NCBI SRA database under the accession PRJNA486234 (total 89 Giga bases), were aligned to the genome with the HISAT2 program for structural gene predictions. In the second approach, NCBI RefSeq invertebrate protein sequences were aligned with the Diamond (v 2.0.6) program [27] and used as external evidence to identify protein coding regions in the genome.

For functional annotation of the predicted genes in *A. assamensis,* OmicsBox Blast2GO methodology was utilized [28]. BLASTp [29] program with an e-value of $1e\text{-}^{5}$ was employed to align protein sequence derived from *A. assamensis* genome to the Swiss-Prot and NCBI-nr insect database. Functional domain identification through InterProScan [30], Gene ontology analysis and KEGG pathway analysis were also performed by OmicsBox Blast2GO methodology. The non-coding RNAs in the genome were identified using INFERNAL v1.0 [31] with the Rfam database [32] and tRNAs were identified using the tRNAscan-SE v1.3.1 program [33].

2.7. Phylogenomic analysis

A dataset of eleven insect genomes *A. assamensis* (this study), *A. pernyi* [34], *A. yamamai* [35], *A. mylitta* (GCA_014332785.1 AM_v1.0), *B. mori* [36], *D. melanogaster* [37], *D. plexippus* [38], *H. armigera* [39], *P. xuthus* [40], *S. litura* [41] and *S. ricini* [42] comprising of ten lepidopteran species and one dipteran species was selected for phylogenomic analysis. The dipteran insect *D. melanogaster* was used as an outgroup in this analysis. Single-copy orthologous genes conserved among the study group were identified using BUSCO [17]. The identified single-copy orthologous gene families were individually aligned with MUSCLE (v3.8.15) [43]. Poorly aligned regions from the multiple sequence alignment files were trimmed using trimAI (v1.4) [44] with the parameter "-automated 1". Trimmed alignments of all the single-copy genes families were concatenated to generate a final alignment supermatrix.

The aligned sequence supermatrix was used to reconstruct the maximum-likelihood (ML) phylogenetic tree using IQ-TREE (v2.1.4) [45] with the insect+F + R4 model, the best fit model according to the Bayesian information criterion identified by ModelFinder [46]. The

divergence times among the selected organisms were estimated, using the species tree generated in the previous step, employing the MCMC Tree program in PAML v4.9 [47] with four fold degenerate sites. Reference divergence times of the following species were retrieved from the Time Tree database [48] and were used as the calibration, namely *D. melanogaster-B. Mori* (223.8–344.7 mya), *A. yamamai–B. mori* (39.8–95.1 mya), *A. assamensis- S. ricini* (28.6–30.1 mya) and *S. litura-H. armigera* (50–60 mya).

2.8. Identification of duplicated genes

Duplicated genes are considered as one of the major sources of adaptive and non-adaptive evolution. To identify potentially duplicated gene families in the *A. assamensis* genome, we performed a self-BLAST search with an e-value cutoff of 1e-^{10} and identified the best hits within the genome. Subsequently, the blast results were processed through HSDFinder tool [49] (≥80% pairwise sequence identity and protein lengths variation ≤30 amino acids) to find highly similar duplicated genes. The ratio of synonymous / non-synonymous mutations in the duplicated gene families was calculated by Ka/Ks Calculator [50].

2.9. Comparative genomics analysis

To identify gene sequences shared and unique to the *A. assamensis* genome, we compared predicted genes with *A. pernyi* [34], *A. yamamai* [35], *A. mylitta* (GCA_014332785.1 AM_v1.0), *B. mori* [36], *D. melanogaster* [37], *D. plexippus* [38], *H. armigera* [39], *P. xuthus* [40], *S. litura* [41] and *S. ricini* [42] genomes using OrthoFinder2 program [51]. The parameters used were diamond_ultra_sens mode, e-value of 1^{-10}, query and subject coverage of 70%. Analysis of collinearity and synteny between *A. assamensis*, and *A. pernyi* genomes was carried out using the MCScanX toolkit [52] and visualization of syntenic blocks was performed by the SynVisio tool [53].

Following the gene family clustering and divergence estimation, the expansion and contraction of gene families were analyzed using CAFÉ v4.1 [54] with a probabilistic graphical model (PGM) to calculate the probability of transition in each gene family from parent to child nodes in the phylogeny.

2.10. Identification of detoxification gene families

2.10.1. Identification of the cytochrome P450 (CYP) gene family

To annotate the cytochrome *p*450 gene family in *A. assamensis,* BLASTp search was used against the NCBI-nr insecta database with an e-value of 10$^{-5.}$ The functional domain information for the protein sequences were obtained using InterProScan [30]. For comparative analysis, the annotated CYP450 gene families of *B. mori* and *S. litura* were obtained from previously published studies [36,41]. Amino acid sequences of CYP450 were aligned by MAFFT program version 7 [55] and model selection was conducted using an automatic selection of amino acid selection model using RaxMLv8 [56]. The maximum likelihood tree was inferred by using the LG + Gamma+I model in RaxML [56]. To evaluate the confidence of the tree topology, the bootstrap method was applied with 100 replications using the rapid bootstrap algorithm [57].

2.10.2. Identification of Glutathione S-transferase (GST) gene family

Annotation of GST members in *A. assamensis* genome was carried out using the previously reported GST sequences from *B. mori* and *S. litura* [36,41] as a query. Significant hits (BLASTp, e-value10^{-5}) were further processed through NCBI conserved domain database [58] to categorize the sequences in different GST classes. A comparative analysis of the GSTs identified in *A. assamensis* was performed with GSTs from *S. litura* and *B. mori* using MEGA XI [59] and the phylogenetic tree was constructed using the maximum-likelihood method with LG + Gamma amino acid substitution model with 500 bootstrap replicates.

2.10.3. Identification of ABC transporter gene family

ABC transporter genes in *A. assamensis* genome were identified using BLASTp search with *B. mori* ABC transporter [60] as query (e-value10^{-5}). Additionally, we also searched ABC transporter candidates in *A. assamensis* with ABC_scan [61]. The results from both the BLASTp and ABC_scan were combined to obtain the final set of potential ABC transporter candidates in *A. assamensis*. A similar search was also performed to annotate the ABC transporter genes in *S. litura* genome. The maximum-likelihood tree was generated using RaxMLv8 [56] with 500 bootstrap replicates by using LG + Gamma amino acid substitution model.

3. Results

3.1. Assembly of A. assamensis genome

We used a single male moth for the whole-genome sequence assembly of *A. assamensis* (Fig. 1). The main reason for the selection of male moth for sequencing was the homogametic nature of lepidopteran males which reduces the complexity of the genome assembly process. A combination of different library sequencing strategies (paired-end and mate-pair with different insert sizes) on the Illumina platform was followed to generate a total of 188.15 million paired reads with ~113 x coverage. Information on libraries and the data generated in this study is provided in Supplementary Table 1. Additionally, long-read sequencing using the PacBio RS II platform generated a total of 227,694 subreads with average length of 9 Kb and N50 value of 14.9Kb (Supplementary Table 2). We analyzed the quality of raw data using FastQC [8] and removed the low quality bases and adapter contamination using Trim Galore [9], NextClip [10] and Cutadapt tools [11]. After raw read quality control, a total of 156.28 million high quality paired-end reads with ~94 x coverage were available for the genome assembly.

Before initiating the genome assembly process, we performed k-mer size (21-mer) distribution analysis using Illumina pair-end reads to estimate the genome size and level of heterozygosity in *A. assamensis* genome with the Jellyfish program [18]. The k-mer frequency plot generated by GenomeScope showed a single peak, suggesting the presence of a low level of heterozygosity (0.39%) in *A. assamensis* genome (Supplementary Fig. S1 and Supplementary Table 3). Based on the 21-mer distribution, the genome size of *A. assamensis* was estimated to be 443.37 Mb.

The high quality *de-novo* draft genome of *A. assamensis* was generated by utilizing Illumina and PacBio reads with MaSuRCA hybrid assembler [16] and employing the SOAP Gapcloser [14] to close the gap within each scaffold. After gap closing the final assembly of

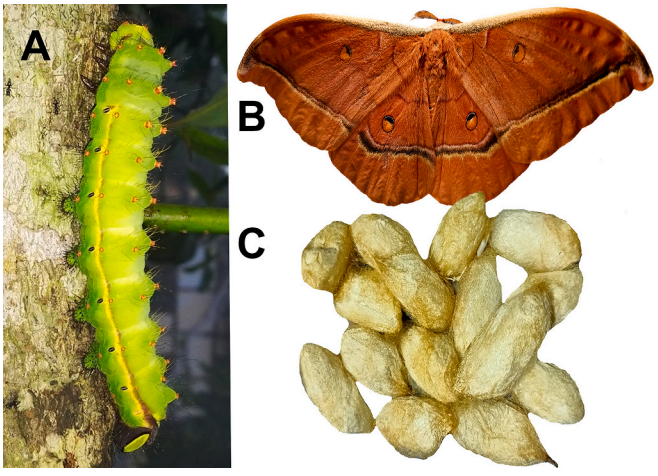

Fig. 1. Different stages of Muga silkworm. Fig. (A) shows the *A. assamensis* 5th instar larva, (B) Silkmoth, and (C) Cocoon.

A. assamensis consisted of 2,697 scaffolds (>1 kb), with an N50 length of 683.23 Kb. The genome has a GC ratio of 34.00% which is similar to other saturniid and *B. mori* genomes [34–36]. A comparative summary statistics of the assembled *A. assamensis* genome with other saturniid silkmoths is given in Table 1.

The quality of assembled genome was analyzed by assessing the presence of Universal Single-Copy Orthologs genes using the BUSCO pipeline with lepidoptera_odb10 which contains 5,286 sequences. Out of the total, 5,181 (98%) complete BUSCO genes were present in *A. assamensis* genome assembly, among these 179 sequences were complete duplicated and 34 BUSCO genes were fragmented (Supplementary Table 4). Approximately 1.4% of BUSCO genes (71 genes) were predicted to be missing in the current genome assembly. We also analyzed the length distribution of the 20 lengthiest genes present in the insect genome (Supplementary Table 5) and found that all the genes were accurately identified in our genome assembly. The results from BUSCO and the 20 lengthiest gene analyses showed that the current genome assembly is of sufficient quality for downstream analysis.

3.2. Repeat identification

Genomes of higher eukaryotes contain a considerable amount of repetitive elements. To characterize the repeat content in the *A. assamensis* genome we developed a custom repeat library using RepeatModeler2 [21] with RECON [14], RepeatScout [62], and TRF [63]. We also employed the Extensive de-novo TE Annotator tool to generate a filtered non-redundant TE library for *A. assamensis* genome. The resulting libraries of both the programs were further curated using the CENSOR [64] and finally combined with the RepBase RepeatMasker library to comprehensively annotate the repeat elements in *A. assamensis* genome with RepeatMasker program and RMBlast as a search engine. The percentage of different repetitive elements identified in *A. assamensis* genome is summarized in Supplementary Table 6. A total of 247.25 Mb amounting to 49.35% of the total genome was identified as repeats in *A. assamensis* genome which is higher than *B. mori* (46.84%) and *A. yamamai* (37.33%) but lower than *A. pernyi* (60.74%). The predominant repeat element was DNA elements, occupying a total of 112.14 Mb (22%) of the total genome. The second largest family of classified elements in *A. assamensis* genome was the Class I retrotransposons (15%). Compared to *A. yamamai* [35] and *B. mori* [36], our analysis showed the dominance of DNA transposons over the LINE elements which constitute only 7.93% of the total identified repeat elements in the *A. assamensis* genome. Approximately 11% of the identified repeat elements were not classified in this genome (Supplementary Table 6).

3.3. Gene prediction and annotation

We applied the BRAKER2 pipeline [24] to structurally annotate the gene models for *A. assamensis* genome. This pipeline provides the advantage of both *ab-initio* gene predictors (GeneMark-EP+ and AUGUSTUS) and evidence-based gene prediction. We used RNA-seq available for *A. assamensis* at the NCBI-SRA database and NCBI RefSeq invertebrate protein sequences to generate evidence for the coding region in the genome. Table 2 shows the comparative summary statistics for the gene prediction for the analyzed genome generated by the Genome Annotation Generator tool along with other Saturniidae moths. A total of 18,385 genes were identified in the *A. assamensis* genome. The average gene length in this genome is 9,750 bp, which is comparable to the mean gene length (9,745 bp) of the *A. pernyi* genome [34]. The number of exons and introns per gene were 6 and 5 respectively. Similar to genome assembly, we utilized BUSCO to assess the completeness of our gene prediction. Our analysis revealed the presence of over 96% of lepidoptera-specific single-copy genes in our gene annotation (Supplementary Table 4), emphasizing the accuracy and comprehensiveness of our gene prediction approach. Moreover, a BLAST analysis was conducted on *A. assamensis* protein-coding genes using the Swiss-Prot database (Release 2023_05). The results showed that 5,807 proteins from *A. assamensis* displayed over 70% coverage with the proteins in the Swiss-Prot database (Supplementary Table 7). The lower number of BLAST hits observed may be attributed to the presence of fewer than 10,000 reviewed protein sequences belonging to the class Insecta in the Swiss-Prot database. Furthermore, since <32% of *A. assamensis* genes showed hits with the Swiss-Prot database, we conducted a BLASTp search against the NCBI-nr database. This search was specifically limited to insect sequences, utilizing an e-value of 1e-5, to assign functions to the predicted genes in *A. assamensis*.

Additionally, we also performed protein domain searches using InterproScan [30] and GO term annotation using OmicsBox. The functional annotation showed around 86% of the genes predicted in *A. assamensis* have a significant match with the NCBI-nr insect sequences. Moreover, 71.32% and 63.47% of the total sequences were annotated with domain and GO term information, respectively (Supplementary Table 8, Fig. 2A). The GO term distribution under the molecular function category showed a large proportion of GO terms were related to the binding activity such as organic cyclic compound binding, heterocyclic compound binding activity, ion binding and hydrolase activity (Fig. 2B). Under the cellular component and biological process categories, terms associated with membrane, intracellular anatomical structure and organic substance metabolic process, primary metabolic processes were abundant. The distribution of 6 major categories of enzymes, transferase, hydrolases, lyases, isomerases, ligases, and translocases are provided in Supplementary Fig. S2.

Table 1
Genome assembly statistics of *A. assamensis* and its comparison with other Saturniidae silkworms.

	A. assamensis	*A. yamamai*	*A. pernyi*	*S. ricini*
No. of scaffolds	2697	7723	423	155
L50	196	265	20	7
N50 (kb)	683.23	734.7	13,766.17	21,366.39
GC content (%)	34.00	34.07	37.03	34.26
Longest scaffold length (Mb)	4.95	3.17	2.99	33.97
Mean scaffold length (kb)	185.83	85.64	1719.41	2906.32
Median scaffold length (kb)	40.77	1.71	20	28.0
Shortest scaffold length (bp)	1000	123	1000	568
Total (Mb)	501.18	661.38	727.31	450.48

Table 2
Comparative gene prediction analysis of the *A. assamensis* genome with selected saturniidae moths.

	A. assamensis	*A. pernyi*	*A. yamamai*	*S. rcini*
Total genome length (Mb)	501.18	727.31	661.38	450.48
Number of genes	18,385	21,431	15,481	18,078
Number of exons	114,188	94,995	87,344	111,452
Number of introns	95,803	75,370	71,863	93,374
Total gene length (Mb)	179.26 (35.8%)	208.85 (28.7%)	170.53 (25.8%)	186.28 (41.4%)
Total CDS length (Mb)	24.2 (4.8%)	25.54 (3.5%)	20.33 (3.1%)	23.30 (5.2%)
Longest gene (kb)	171.1	98.83	179.80	209.36
Longest exon (kb)	13.79	14.96	24.65	14.58
Mean gene length (bp)	9,750	9,745	11,016	10,304
Mean exon length (bp)	212	269	233	211
Mean intron length (bp)	1616	2070	2092	1739
Mean CDS length (bp)	1316	1301	1313	1289
Mean no. of exons per mRNA	6	5	6	6
Mean no. of introns per mRNA	5	4	5	5

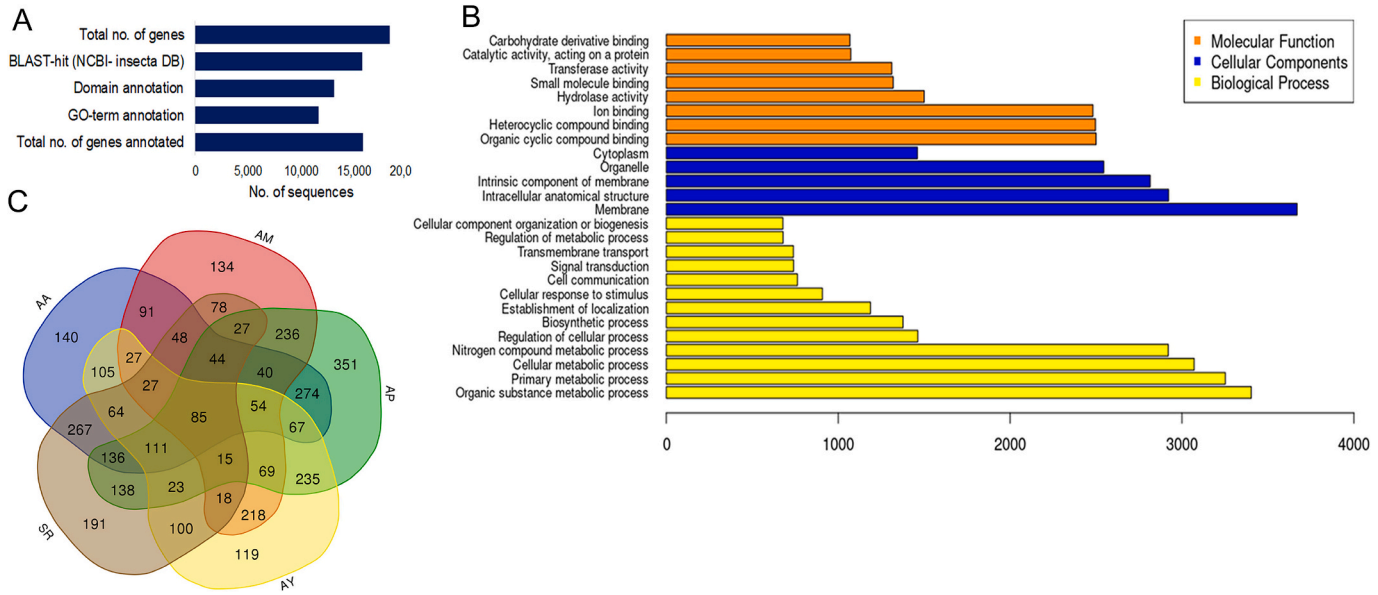

Fig. 2. Functional annotation of genes identified in *A. assamensis* genome. (A) Genes with significant Blast hits, domains and GO term annotations, (B) representation of top GO terms identified in *A. assamensis* genome, and (C) shows the gene clusters uniquely present in analyzed Saturniidae species derived from orthology analysis of 11 insect species (AA- *Antheraea assamensis*; AM- *Antheraea mylitta*; AP- *Antheraea pernyi*; AY- *Antheraea yamamai*; SR- *Samia ricini*).

Noting the important role of non-coding RNAs in the regulation of gene expression and cellular homeostasis, we identified different ncRNAs such as rRNAs, tRNAs snRNAs, miRNAs, and Signal recognition particle RNAs (Supplementary Table 9). A total of 713 copies of tRNAs were detected in *A. assamensis* genome. We compared the predicted numbers of tRNA genes in *A. assamensis* with four other Saturniidae silkmoths: *A. pernyi* [34], *A. mylitta*, *A. yamamai* [35], and *S. ricini* [42]. The analysis revealed 705, 725, and 1165 tRNAs in the genomes of *A. mylitta*, *A. yamamai*, and *S. ricini* respectively. In a prior study, Duan et al. (2020) reported 766 tRNA genes in *A. pernyi*. These findings suggest that the anticipated number of tRNA genes among *Antheraea* species typically ranges between 700 and 800 copies (Supplementary Table 10).

3.4. Phylogenetic analysis and divergence time estimation

To analyze the taxonomic position of *A. assamensis* within the Saturniidae family, we constructed a phylogenetic tree based on the 877 conserved one-to-one orthologous genes identified from BUSCO [17] analysis among the eleven selected genomes. Multiple sequence alignment of the 877 genes from the selected species was conducted using MUSCLE (v3.8.15) [43]. After trimming poorly aligned regions with trimAl (v1.4) [44], the supermatrix generated for the trimmed alignments contained 463,306 amino-acid sites.

The maximum-likelihood species tree generated using the insect+F + R4 substitution model showed all the selected organisms from the Saturniidae family clustered together, confirming the phylogenetic arrangement of *A. assamensis* in Saturniidae with other *Antheraea* species. The analyzed Bombycoidae insect species were classified as a sister clade to Noctuidae (Fig. 3), which further strengthens the grouping of Lepidoptera. Our analysis showed the divergence of Bombycidae lineage from the Saturniidae approximately ~64.61 mya, which is close to the median value of divergence time (~67 mya) reported in the TimeTree database [48]. The analysis revealed the divergence of *A. assamensis* from other *Antheraea* species approximately ~28.7 mya, (Fig. 3). A previous study based on the fossil evidence reported that the last common ancestor of *A. pernyi* and *A. yamamai* diverged from the Bombycidae lineage ~81.5 million years ago, and the divergence time between sister species *A. pernyi* and *A. yamamai* was estimated to be ~30.3 million years ago [34,35]. The placement of *A. assamensis* in comparison to other *Antheraea* species in the inferred species tree

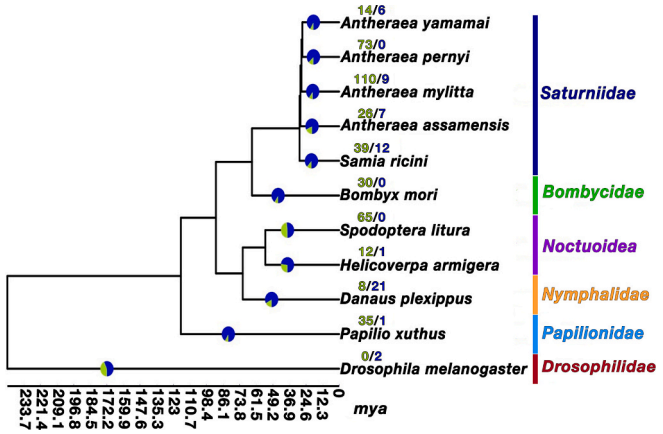

Fig. 3. Phylogenomic analysis of *A. assamensis*. The tree shows the estimated divergence time of *A. assamensis* with other Antheraea species. The species tree was calibrated using the estimated divergence times between *D. melanogaster*-*B. mori* (223.8–344.7 mya), *A. yamamai*–*B. mori* (39.8–95.1 mya), *A. assamensis*-*S. ricini* (28.6–30.1 mya) and *S. litura* - *H. armigera* (50–60 mya). The pie chart on the branches of each node shows the number of the expanded (green) and contracted (blue) gene families among the selected species. The numbers of significantly ($p < 0.05$) expanded and contracted gene families are given above the node names. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

suggests that the *A. assamensis* could be the basal species of the genus *Antheraea*.

3.5. Comparative genome analysis

We used the OrthoFinder program [51] to identify orthologous genes among the eleven genomes *i.e.*, *A. assamensis* (this study), *A. pernyi* [34], *A. yamamai* [35], *A. mylitta* (GCA_014332785.1 AM_v1.0), *B. mori* [36], *D. melanogaster* [37], *D. plexippus* [38], *H. armigera* [39], *P. xuthus* [40], *S. litura* [41] and *S. ricini* [42]. Representative protein sequences of these genomes were aligned against each other using DIAMOND BLAST [27]. A total of 185,092 genes were present in these eleven genomes out of

which 137,993 (74.52%) genes clustered in 18,683 orthogroups (Supplementary Table 11). Of all the gene clusters only 8.7% (1,626) clusters were common in all the analyzed genomes. From the total of 18,683 orthogroups we extracted 3,532 gene clusters uniquely shared by the analyzed Saturniid species (Fig. 2C). *In-silico* functional GO enrichment analysis of these gene clusters revealed the majority of genes belonging to the metabolic process such as cellular metabolic process, macromolecule metabolic process, nitrogen compound metabolic process, organic cyclic compound metabolic process, heterocyclic metabolic process, nucleobase-containing compound metabolic process and aromatic compound metabolic process (Supplementary Table 12a).

Similarly, we analyzed the genes present only in *A. assamensis*. The analysis revealed a total of 2,421 genes are uniquely present in this genome. For the majority of these genes, BLAST hits were not found in the NCBI-nr database and only 335 genes were functionally annotated. However, the expression of 935 genes was found in transcriptome data (Supplementary Table 12b and 12c), suggesting that these are true genes but their function remains to be elucidated. The 3,532 gene clusters unique to the studied sericigenous moths of the Saturniidae family, also contain genes exclusive to particular species. These genes may confer adaptive advantages, helping these species thrive in their peculiar habitat.

The comparative CAFE analysis among the selected species showed a significant expansion of 26 gene families and 7 were significantly contracted ($p < 0.05$) in *A. assamensis* (Fig. 3).

3.6. Genome-wide identification of detoxification-related gene families

3.6.1. Identification and phylogeny of CYP450 genes in A. assamensis

The CYP450 gene family is a group of enzymes involved in the detoxification of xenobiotics and the metabolism of a wide range of compounds in insects and other organisms. Studies have shown that the CYP450 gene family has undergone multiple rounds of gene duplication and divergence in insects, leading to the evolution of multiple subfamilies with distinct functions. For example, some insect CYP450s are involved in the detoxification of insecticides, while others are involved in the metabolism of pheromones and other signaling compounds, some insects have evolved specialized CYP450s that are involved in the detoxification of plant secondary metabolites [41,65]. This adaptation has likely contributed to the success of these insects as herbivores and has shaped the evolution of plant-insect interactions [66]. The CYP450 monooxygenases play a vital role in insect survival and adaptation to their environment. The increasing availability of insect transcriptomes and genomes is contributing to our understanding of their diversity and functional significance [67]. In our study, we identified 130 candidate CYP450 genes in *A. assamensis* and compared them to *S. litura* and *B. mori* (Table 3). The phylogenetic analysis of 344 CYP450 proteins from three insect species (*A. assamensis, S. litura* and *B. mori*) showed that insect CYPs have evolved into several groups over time. The results showed that CYP450 genes in *A. assamensis* were grouped into four clans: CYP2, CYP3, CYP4, and mitochondrial clans (Fig. 4). The CYP3 and CYP4 clades had the majority of these genes, which is consistent with other insect species [68]. The CYP3 clan is involved in xenobiotic metabolism and insecticide resistance and gene expression of some of the members can be induced by various agents such as phenobarbital, pesticides, and natural products [69]. Therefore, the expansion of genes in the CYP3 clan can be directly correlated with the survival of insect under challenging environmental conditions. Since the genes in this cluster are also involved in the metabolism of natural xenobiotic compounds present in different host plants, the expansion of these genes may provide an opportunity to broaden the host range. Gene expansions were also common within the CYP4 clan particularly in the CYP450-4C1-like clusters (AaP450_83–90), (SlP450_61–81) and CYP450-4 L6 cluster (AaP450_105–107, AaP450_115–117) in *A. assamensis* and *S. litura* when compared to *B. mori* respectively (Fig. 4). Previous studies have suggested that genes from the CYP4L family have a role in odorant degradation in other species [41]. The results of the study are consistent with previous findings that expansions in CYP3 and CYP4 clans are involved in xenobiotic metabolism and enhance the insect's ability to detoxify various toxic substances allowing the insect to feed on a wide range of plants and become tolerant to insecticides.

Table 3
Analysis of detoxification related gene families in *A. assamensis* and their comparison with *S. litura* and *B. mori*.

Detoxification related gene families	Clan	*A. assamensis*	*S. litura*	*B. mori*
Cytochrome P450 (CYP450)	Clan 2	8	11	7
	Clan 3	55	53	28
	Clan 4	53	56	35
	Mitochondrial	14	11	13
	total	**130**	**131**	**83**
Glutathione-S-transferase (GST)	ε	15	19	8
	δ	8	3	4
	ω	3	3	4
	σ	5	6	2
	θ	1	1	1
	ζ	2	2	2
	Microsomal	5	3	0
	Uncharacterized	5	3	2
	Metaxin	1	2	0
	Total	**45**	**42**[a]	**23**
ABC transporter	ABC-a	10	11	8
	ABC-b	12	13	9
	ABC-c	13	17	10
	ABC-d	2	2	2
	ABC-e	1	1	1
	ABC-f	3	3	3
	ABC-g	18	22	15
	ABC-h	3	7	3
	Total	**62**	**76**	**51**

[a] Cheng et al. 2017, reported total 47 GST sequences in *S. litura* but only 42 unique GST sequence were retrieved from the NCBI database.

3.6.2. Comparative phylogeny of Glutathione-S transferase of A. assamensis

The Glutathione S-transferase gene family plays an important role in detoxification and cellular defense against oxidative stress in insects and other organisms. The evolution of the GST gene family in insects has been shaped by a combination of gene duplication, divergence, and loss events, leading to the diversification of GST functions in different insect groups [70]. In *A. assamensis*, a total of 45 genes encoding GSTs proteins were identified which includes 40 cytosolic GSTs and 5 microsomal GSTs (Table 3). A comparison of GST sequences among the *A. assamensis, S. litura* and *B. mori* revealed that *A. assamensis* has an approximately similar number of GSTs as *S. litura* [41] and a higher number than *B. mori* [71]. Based on the phylogenetic analysis, the GSTs are classified into delta, epsilon, omega, sigma, theta and zeta, metaxin, and microsomal. Among them, the largest and most abundant class of GST in *A. assamensis* is the epsilon class followed by delta (Fig. 5) The phylogenetic tree suggested that the epsilon GSTs are further distributed in 5 different sub-clades, indicating internal variation among these members. Many studies have recognized the specific functions of epsilon and delta GSTs regarding insecticide resistance, and they are widely recognized as important contributors to the development of insecticide resistance in insects [72,73]. The expansion of the epsilon and delta GST class in polyphagous insects may provide an adaptive advantage against the plant secondary metabolites present in a wide variety of host plants [41].

3.7. Identification and Phylogeny of ABC transporters

The ATP-binding cassette transporter gene family is a large and

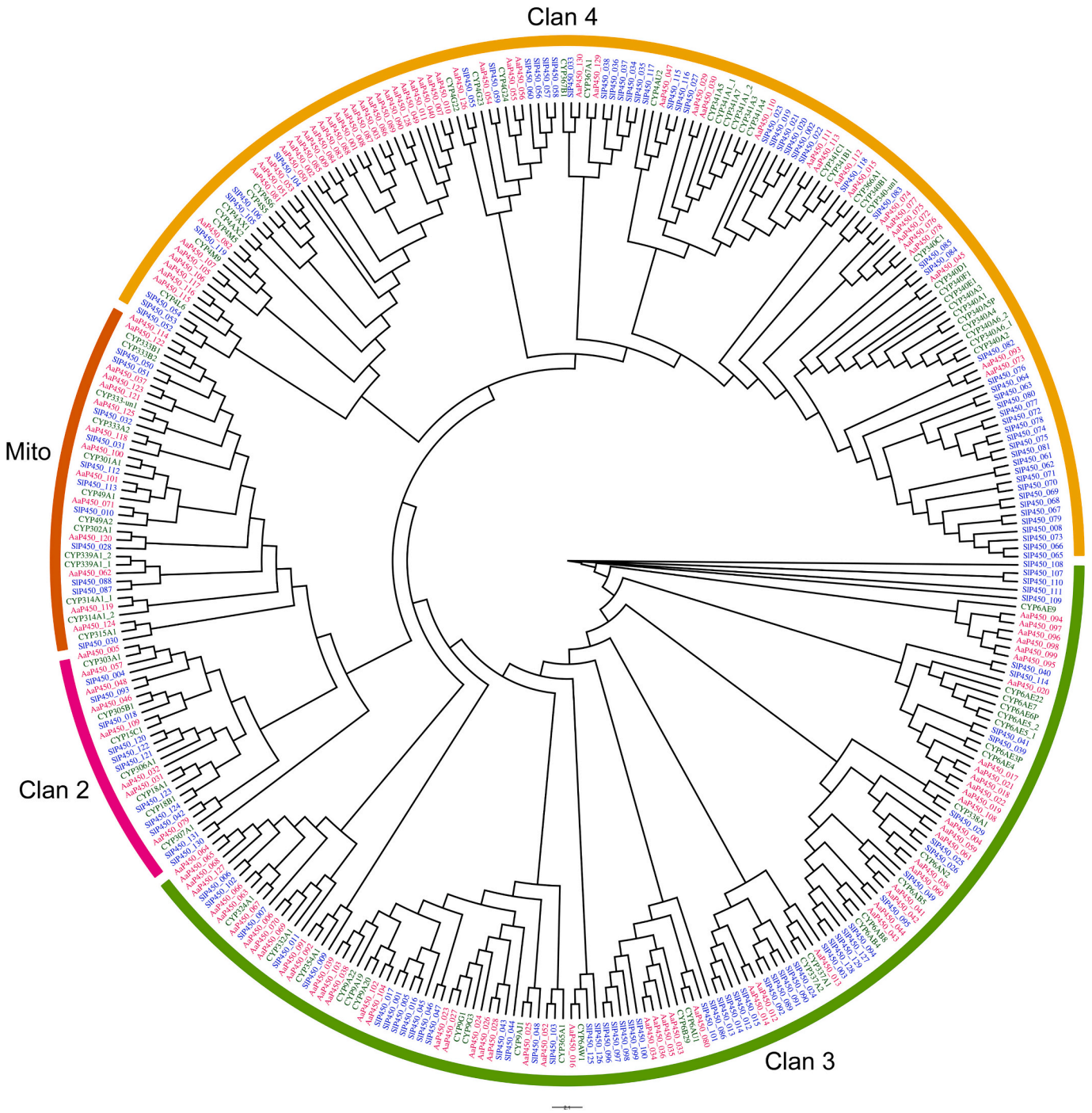

Fig. 4. Phylogenetic analysis of CYP450 gene family among in *A. assamensis*, *S. litura* and *B. mori* genomes. Node names in blue, pink and green colors represent *S. litura*, *A. assamensis* and *B. mori* sequences respectively. Based on sequence similarity CYP450 sequences have been clustered in 4 major groups, Clan 2, Clan 3, Clan 4 and Mitochondrial CYPs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

diverse group of genes that encode for transmembrane proteins involved in the transport of a wide variety of molecules across cell membranes. These transporters utilize ATP as an energy source and play an important role in the defense against xenobiotics, including insecticides, multiple drugs, and allelochemicals [41,74] In *A. assamensis*, the ABC transporter genes were annotated and compared with *S. litura*, and *B. mori*. Our analysis showed a total of 62 ABC transporter genes in the *A. assamensis* genome, 76 in *S. litura*, and 51 in *B. mori* (Table 3), further these genes were categorized into 8 subfamilies (a to h) (Fig. 6). The subfamily ABC-g contains half transporters, the genes belonging to this subfamily either contain NBD-TMD or TMD-NBD domain organization. Our analysis showed, *S. litura* contains 22 half transporters while *A. assamensis* and *B. mori* have 18 and 15 copies of half transporters respectively. The subfamilies ABC-b and ABC-c, that confer multidrug resistant, showed a higher number in *S. litura* and *A. assamensis*. The majority of the full transporter (NBD-TMD-NBD-TMD) belonging to subfamilies ABC-a, ABC-b and ABC-c were identified in *S. litura* [35], *A. assamensis*, and *B. mori* [15] (Fig. 6). The results suggest that the *S. litura* and *A. assamensis* have expansion of subfamilies (ABC-b and ABC-c) of ABC transporter genes that may contribute towards enhanced resistance to xenobiotics [60,75]. The presence of full transporters in all three species suggests that they are capable of active transport of

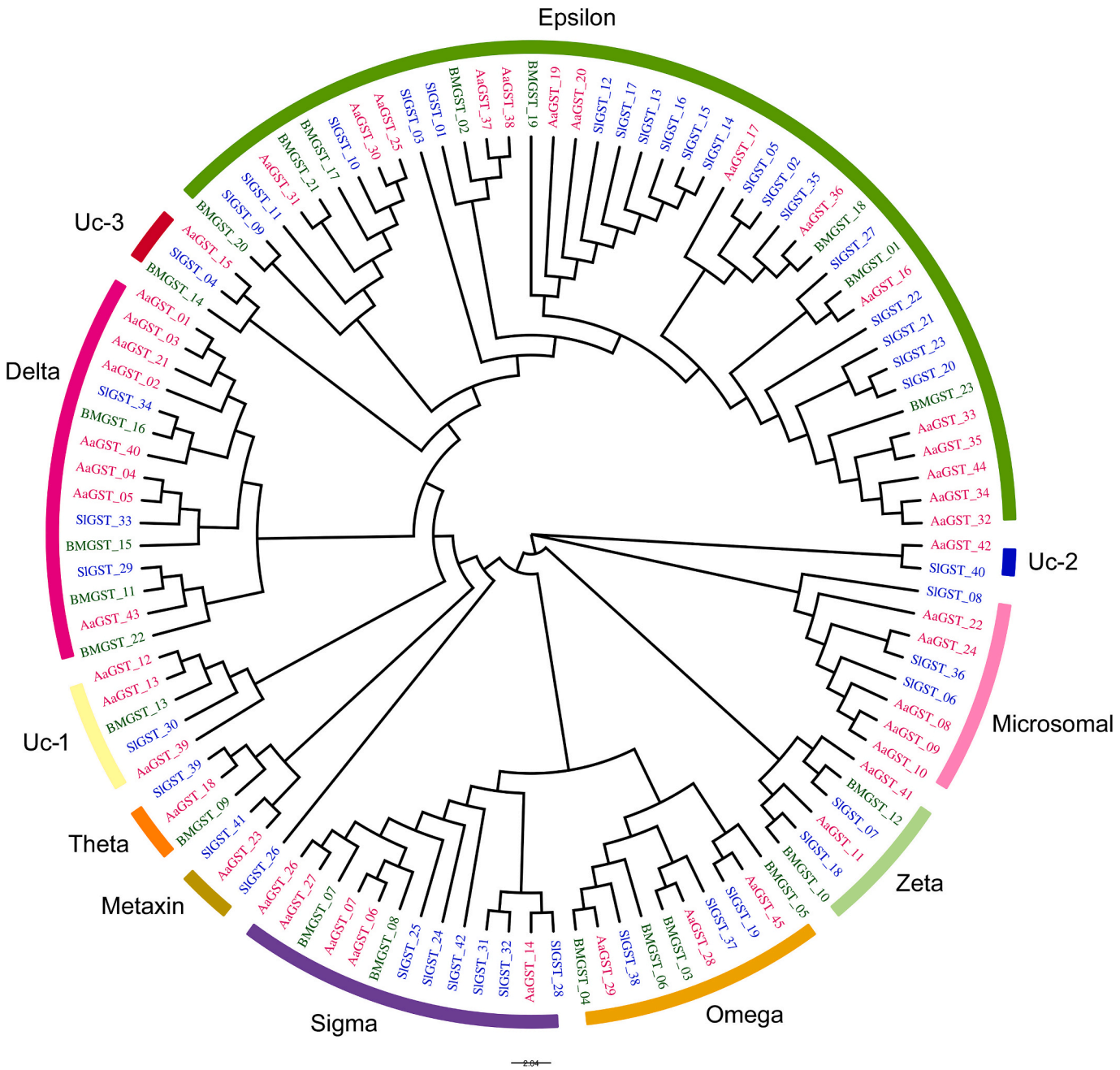

Fig. 5. Phylogenetic analysis of GST gene family among in *A. assamensis*, *S. litura* and *B. mori* genomes. Node names in blue, pink and green colors represent *S. litura*, *A. assamensis* and *B. mori* sequences respectively. Based on the similarity sequences have been clustered in different groups, Epsilon, Delta, Sigma, Omega, Zeta, Theta, Metaxin and microsomal classes. Three clusters Uc-1, Uc-2 and Uc-3 represents the functionally classified GST sequences. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

molecules across cell membranes. However, the ABC-e and ABC-f subfamilies, which contain two ATP binding domains but lack TMDs, did not show any expansion. These proteins are not involved in molecule transport but are active in other functions that are essential for cell viability [76].

3.8. Collinearity analysis between A. assamensis and A. pernyi genomes

We used MCScanX [52] to search syntenic blocks between *A. assamensis* and *A. pernyi* genomes based on the protein sequence homology identified through a reciprocal best-hit approach with an e-value cutoff of $1e^{-10}$. We selected the *A. pernyi* genome for synteny analysis because it belongs to the same genus (*Antheraea*) as *A. assamensis* and has a chromosomal level assembly [34]. Syntenic blocks represent the order of homologous genes derived from a common ancestor shared between the genomes. In our analysis, we found a total of 592 collinear blocks between 49 chromosomes of *A. pernyi* and 463 scaffolds of *A. assamensis* genomes. A total of 7551 (41%) genes of *A. assamensis* were present in syntenic blocks with 7536 (35.16%) genes *A. pernyi*. Fig. 7 shows the syntenic blocks between *A. assamensis* scaffolds and chromosome 1 of the *A. pernyi* genome. The functional annotation of these genes showed that they belong to diverse cellular processes (Supplementary Table 13a). Further, GO enrichment analysis showed the over-represented GO terms such as organic cyclic compound binding, protein binding, nucleoside phosphate binding, and signaling receptor regulator activity (Supplementary Table 13b).

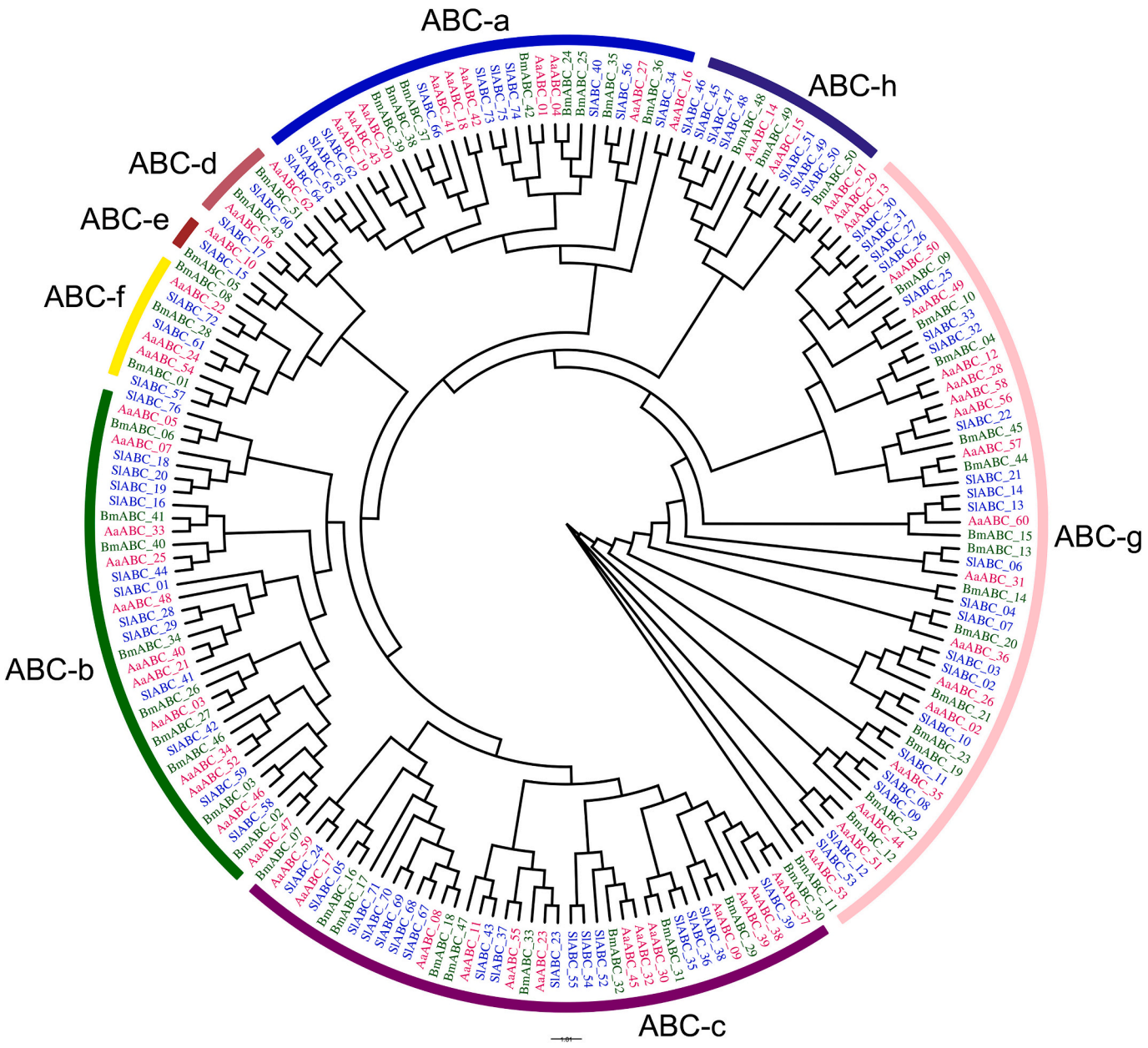

Fig. 6. Phylogenetic analysis of ABC transporter gene family among in *A. assamensis*, *S. litura* and *B. mori* genomes. Node names in blue, pink and green colors represent *S. litura*, *A. assamensis* and *B. mori* sequences respectively. Based on the similarity sequences have been clustered in ABC-a to h families. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

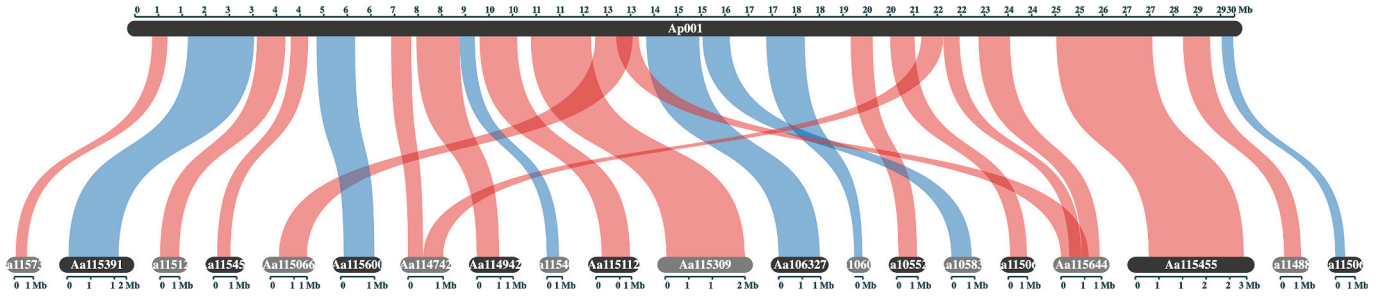

Fig. 7. Collinearity analysis between *A. assamensis* and *A. pernyi* genomes. Scaffolds of *A. assamensis* genome shows synteny based on collinear homologous gene anchor with chromosome 1 of *A. pernyi* genome.

3.9. Duplicated gene families in A. assamensis genome

We identified high-similarity homologous genes within the *A. assamensis* genome using a self-BLAST search approach using *A. assamensis* protein sequences. The obtained results were parsed through the HSDFinder tool [49] to get gene sequences with a minimum of 80% sequence identity and 30 amino acid length variation. With this, we identified 687 paralogous gene families in the *A. assamensis* genome. A total of 1,726 paralogous genes were present in these families which is equivalent to 9.38 % of the total identified genes in this genome. The top gene families based on the number of sequences present in the family were neuro-filament medium polypeptide-like (32 copies), chorion class A protein Ld19-like (24 copies), histone gene family (H2A, H2B, H3, and H4; 15 copies of each gene), tubulin beta chain-like (13 copies), histone H1B-like (11 copies). The details of duplicated gene families are given in Supplementary Table 14a. Out of 687 paralogous gene families, 307 gene families were tandemly duplicated. The majority of tandemly duplicated gene families were of small size. Since we found a significant fraction of the total genes (9.38%) present as duplicated genes, to investigate whether these genes are under selection constraint or not, we calculated pairwise Ka/Ks values. The Ka/Ks ratio for four gene pairs was found to be >1 (p-value <0.05) which shows the sign of diversifying selection on these genes. The functional analysis of these genes showed their association with insect immunity (Moricin-like and Histidine-rich glycoprotein-like) and circadian rhythm-related genes (Circadian clock-controlled protein-like) (Supplementary Table 14b). A large fraction of the paralogous gene families (572 gene pairs) were shown to have a Ka/Ks ratio of <1 suggesting these genes are under purifying selection (Supplementary Table 14c). The normal functioning of these genes may be essential for the silkworm, hence the purifying selection acts as a guard to remove the deleterious mutations from the genome.

3.10. Silk genes of A. assamensis

A recent report describes the gene coding for silk fibroin-H chain from muga silkworm along with structurally important motifs responsible for its much sought-after properties. The fibroin-H of *A. assamensis* has longer, numerous, and relatively uniform repeat motifs with lower serine content that assume tighter β-crystals and denser packing, which are speculated to be responsible for its acclaimed properties of higher tensile strength and higher refractive index responsible for golden luster [77].

From the functionally annotated data set, we searched the genes related to silk proteins such as silk fibroin, sericin, and p25 genes. We identified a single homolog of a fibroin-heavy chain and two copies of p25 genes (Table 4). An earlier study on Saturniidae silkworm reported the presence of only a single fibroin-heavy gene in these insects [42].

Sericin is a group of serine-rich cocoon proteins produced by silkworms. It is known as glue protein, which binds fibroin fibers. Three genes coding for sericin proteins have been identified and characterized in *B. mori* [78–80]. The most abundant silk sericin of *B. mori* is encoded by the single gene Bombyx sericin 1 (Serl) which gives rise to mature mRNAs through differential splicing of the primary transcript [78,81]. This maturation is a tissue- and developmentally- regulated process and the corresponding sericin proteins can be visualized as distinct layers piling up around fibroin in the middle silk gland. The structure of this gene is characterized by the presence of a large central alternative exon that encodes an internally repetitive sequence.

Not much is known about the sericin genes of saturniid silkmoths except for two studies on sericin protein extraction from cocoons [82,83]. Sequences of Bombyx sericin 1, sericin 2, and sericin 3 were obtained from the NCBI database and were aligned with the genomes of saturniid silkmoths ie., *A. assamensis* (this study), *A. pernyi* [34], *A. yamamai* [35], *A. mylitta* (GCA_014332785.1 AM_v1.0), and *S. ricini* [42] and Bombycidae silkmoth *i.e.*, *B. mandarina*. A significant number of hits were obtained against sericin 1 and sericin 2 in all the genomes analyzed, but no hits were obtained against sericin 3 in saturniid silkmoth genomes. Hits against sericin 3 were obtained only in the *B. mandarina* genome.

Table 4
Silk and pigment binding related genes in *A. assamensis* genome.

S. no.	Gene ID	Description	Protein length	*E*-value
1	WT.00 g049290.m01-v1.0.a1	Fibroin heavy chain	2423	1.02E-34
2	WT.00 g102360.m01-v1.0.a1	Sericin	259	2.56E-70
3	WT.00 g081480.m01-v1.0.a1	Sericin1	1351	0
4	WT.00 g099650.m01-v1.0.a1	Silk protein P25	239	1.25E-82
5	WT.00 g135720.m01-v1.0.a1	Fibroin P25	969	3.26E-96
6	WT.00 g146160.m01-v1.0.a1	Carotenoid-binding protein	296	0

4. Discussion

Muga silk produced by *A. assamensis* is the sole source of natural golden brown colored silk, and therefore, this species is invaluable to the silk industry. Muga silk production has an important role in employment generation and improving rural income. Thus, the development of *A. assamensis* sericulture helps the farmers of North-eastern states economically. However, due to outdoor rearing, *A. assamensis* crops are easily affected by disease outbreaks and seasonal fluctuations in weather conditions, resulting in high annual economic losses. Traditional breeding methods did not yield satisfactory results for obtaining *A. assamensis* breeds resilient to biotic or abiotic stresses. Moreover, the lack of genomic information has significantly hampered the genetic research and molecular breeding work on *A. assamensis*. Considering the economic importance and lack of genomic resources, here, we report a high-quality draft genome of *A. assamensis*. By employing two different sequencing platforms and implementing three diverse library strategies, we successfully assembled the genome of *A. assamensis*. The assembled genome, spanning 501.8 Mb in 2697 scaffolds with an N50 value of 683.23 Kb and a genome BUSCO score of 98% reflects the positive outcomes of adopting long-read sequencing along with mate-pair Illumina libraries. This approach enhances resolution in complex genomic regions and enables the assembly of more contiguous and accurate genomes. However, it's worth noting that our current assembly, based on approximately 4× long-read data, presents an opportunity for further refinement through the addition of more long-read data to achieve an even more complete genome assembly. Approximately 49% of the *A. assamensis* genome was identified to be composed of repetitive elements. In contrast to the *A. pernyi* and *B. mori* [34,36], we found over representation of DNA elements (22%) than the LINE elements (7.93%) in the *A. assamensis* genome. The genome is predicted to contain 18,385 protein-coding genes with a mean gene length of 9750 bp (Table 2). Through BLAST homology and domain search, functional annotation of 86.29% of the predicted genes was achieved. We identified a single homolog of a fibroin-heavy chain (2423 aa) (Table 4) in the *A. assamensis*. Unlike the *B. mori* silk fiber, the saturniid moths contain only a single fibroin chain (H-fibroin) and the gene for fibroin light chain is believed to be absent in these moths [34,35,42]. The absence of fibroin light chain in Saturniidae moths could be a common feature of sericigenous insects from this family. The liquid silk structural analysis from the posterior and middle silk gland of *A. assamensis* showed that α helical structure is predominant in the posterior silk gland while β-sheet, random coil, and β-turn structural components are predominant conformations in the liquid silk form middle silk gland [84,85]. The genome annotation also revealed that the *A. assamensis* genome has two copies of p25 (Table 4). Further, our analysis identified a single homolog of the

carotenoid-binding protein responsible for controlling the cocoon colour trait in *B. mori* [86]. Although, the homolog of carotenoid-binding protein is present in the *A. assamensis* genome, whether this gene is responsible for the golden colour of the muga silk, requires in-depth molecular analysis.

The comparative genome analysis of *A. assamensis* with ten selected genomes revealed that 78.33% (14,402) of the *A. assamensis* genes were clustered with gene sequences from the selected organisms. In the *A. assamensis* genome, a total of 140 species-specific clusters (Fig. 2C) containing 379 genes were identified. Comparative genome analysis revealed a total of 2421 genes unique to *A. assamensis*, including those found in species-specific clusters and singletons. Out of the 2421 genes, BLAST hits for only 335 genes were found in NCBI-nr insecta sequences. A large proportion of *A. assamensis* specific genes are uncharacterized and do not have homologs in other organisms. Out of the total 2421 genes, expression support for 935 genes was found in RNA-seq data.

We performed synteny analysis between *A. assamensis* and *A. pernyi* genomes based on the protein homology. Although both the organisms belong to the same genus they show a large variation in chromosome numbers (*A. assamensis* $n = 15$ and *A. pernyi* $n = 49$). Our analysis showed that 41% of *A. assamensis* genes were present in 592 syntenic blocks with the *A. pernyi* genome. The phylogenetic analysis revealed that *A. assamensis* diverged before the last common ancestor of *A. mylitta, A. pernyi,* and *A. yamamai* (Fig. 3).

Duplicated genes provide source genetic material for the acquisition and evolution of new functions in the genome through mutation, drift, and selection. We analyzed the presence of duplicated gene families in the *A. assamensis* genome. A total of 687 duplicated gene families were identified in the *A. assamensis* genome (Supplementary Table 14a). The analysis revealed the highest number of neuro-filament medium polypeptide-like (32 copies) followed by chorion class A protein Ld19-like (24 copies), histone gene family (H2A, H2B, H3, and H4; 15 copies of each gene) gene families. Further, to analyze the role of evolutionary forces acting on these paralogous genes, we calculated the pairwise Ka/Ks ratio. Our analysis showed a total of 576 gene pairs where the Ka/Ks ratio was significantly (P-value <0.05) less than one (Ka/Ks < 1) (Supplementary Table 14c), suggesting that these genes are under negative selection (purifying selection) and changes in nucleotide sequences of these genes may be deleterious hence these changes are removed from the next generation. Out of the total identified paralogous gene families only four pairs of genes showed signs of diversifying selection (Ka/Ks >1; P-value <0.05) these genes include antimicrobial peptide gene "Moricin-like", multifunctional, and immunity-related gene "Histidine-rich glycoprotein-like" and circadian rhythms related genes "Circadian clock-controlled protein-like" (Supplementary Table 14b). The functional divergence of immunity-related genes is known in many organisms. Divergence in these genes may provide the host with new functionality for the recognition and removal of the new pathogen or its sub-variant.

The evolution of the GST, ABC transporter, and CYP450 gene families in insects is associated with their feeding behavior. These gene families play important roles in the detoxification of toxic compounds, including plant secondary metabolites and insecticides, which insects encounter through their diet. Studies have shown that the evolution of GST, ABC transporter, and CYP450 gene families in insects has been shaped by a combination of gene duplication, divergence, and loss events, leading to the diversification of their functions and the adaptation of insects to their diets and toxic environments [41,65,66,69]. For example, some GSTs in insects have evolved to metabolize specific classes of compounds, such as glucosinolates in cruciferous plants, while others are involved in the detoxification of insecticides [72,73]. Similarly, different ABC transporters have evolved to transport different classes of compounds, such as phospholipids and lipophilic compounds, while CYP450s are involved in the metabolism of a wide range of compounds, including plant secondary metabolites and insecticides [41,65,74]. Overall, the GST, ABC transporter, and CYP450 gene families play crucial roles in the feeding behavior of insects, as they help insects to adapt to their diets and to cope with the toxic compounds present in them. In *A. assamensis* a total of 130 CYP450, 45 GST, and 62 ABC transporter gene families were identified (Table 3). Comparative analysis showed the expansion of these gene families in polyphagous insects (*S. litura* and *A. assamensis*) in comparison to monophagous (*B. mori*). The expansion of these gene families in *S. litura* and *A. assamensis* may provide an adaptive advantage to feed on different host plants.

The genomic resources generated in this study will immensely benefit the future genomic and molecular intervention for the improvement of muga silkworm. The analysis of paralogous gene families showed that some gene families are under evolutionary selection constraints. The unique gene families present in this genome may be involved in providing adaption abilities to this silkworm in its peculiar habitat requirements. The draft genome provides a foundation for in-depth analysis of various genes and genetic loci linked to silk production, immune response, insect behavior, and sex determination.

With complete genome sequences now available for five saturniid silkmoths, *A. assamensis*, *A. mylitta*, *A. pernyi*, *A. yamamai,* and *S. ricini*, future research should focus on comparative genomics, particularly insect behavior. These wild silkmoths exhibit distinct behaviors for example, *A. assamensis* demonstrates positive geotropism and negative phototropism during late larval stages, *S. ricini* displays crowding behavior which is an essential factor for indoor rearing of silkworms, mature larvae of *A. mylitta* spin cocoon on trees and they produce strong peduncle to hang the cocoon. Such mutually exclusive behavioral patterns are fascinating to study and the availability of genome sequence information facilitates such studies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2024.110841.

Disclosures

None.

CRediT authorship contribution statement

Himanshu Dubey: Investigation, Formal analysis, Data curation, Conceptualization, Methodology, Resources, Validation, Writing – original draft, Writing – review & editing. **A.R. Pradeep:** Writing – original draft, Supervision, Resources, Methodology, Data curation, Conceptualization. **Kartik Neog:** Supervision, Project administration, Methodology, Data curation, Conceptualization. **Rajal Debnath:** Data curation, Formal analysis, Methodology, Resources. **P.J. Aneesha:** Formal analysis, Data curation. **Suraj Kumar Shah:** Formal analysis, Data curation. **Indumathi Kamatchi:** Formal analysis, Data curation. **K. M. Ponnuvel:** Writing – original draft, Supervision, Resources, Data curation. **A. Ramesha:** Writing – original draft, Resources, Methodology, Formal analysis, Data curation. **Kunjupillai Vijayan:** Supervision, Data curation. **Upendra Nongthomba:** Writing – original draft, Supervision, Methodology, Data curation. **Utpal Bora:** Supervision, Data curation. **Sivaprasad Vankadara:** Data curation, Project administration, Supervision, Writing – original draft. **K.M. VijayaKumari:** Writing – original draft, Supervision. **Kallare P. Arunkumar:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that the research was conducted without any commercial or financial relationship that could pose conflict of interest.

Data availability

Primary genome sequence reads generated in this study are available

in the NCBI repository under the BioProject: PRJNA478112.

Acknowledgments

The authors acknowledge financial support from the Central Silk Board, the Ministry of Textiles, Government of India through research grants AIT05016MI and AIT-5872.

References

[1] Tikader A, Vijayan K, Saratchandra B. Muga silkworm, Antheraea assamensis (Lepidoptera: Saturniidae)-an overview of distribution, biology and breeding. *Eur. J. Entomol.*. 1102013;.
[2] K. Arunkumar, A. Tomar, T. Daimon, T. Shimada, J. Nagaraju, WildSilkbase: an EST database of wild silkmoths, BMC Genomics (2008), https://doi.org/10.1186/1471-2164-9-338.
[3] M.S. Jolly, S.K. Sen, T.N. Sonwalker, G.K. Prasad, FAO Agriculture Services Bulletin: Non-Mulberry Silks, FAO, Rome, 1981.
[4] K.P. Arunkumar, L. Kifayathullah, J. Nagaraju, Microsatellite markers for the Indian golden silkmoth, Antheraea assama (Saturniidae: Lepidoptera), Mol. Ecol. Resour. (2009), https://doi.org/10.1111/j.1755-0998.2008.02414.x.
[5] Deodikar GB, Bhuyan B, Kshirsagar K, Chowdhury S. Cytogenetic Studies in Indian Silkworms. 2. Chromosome Number in Muga Silk-Worm Antheraea Assamensis Westwood. *Curr. Sci.* JSTOR; 31:2471962;.
[6] M.L. Gupta, R.C. Narang, Karyotype and Meiotic Mechanism in Muga Silkmoths, Antheraea compta Roth. and *A. assamensis* (Helf.) (Lepidoptera: Saturniidae), Genetica (1981), https://doi.org/10.1007/BF00057539.
[7] K.P. Arunkumar, A.K. Sahu, A.R. Mohanty, A.K. Awasthi, A.R. Pradeep, S.R. Urs, et al., Genetic Diversity and Population Structure of Indian Golden Silkmoth (Antheraea Assama), PLOS ONE. Public Library of Science, 2012, https://doi.org/10.1371/journal.pone.0043716.
[8] B. Bioinformatics, FastQC: A Quality Control Tool for High Throughput Sequence Data, Babraham Institute, Cambridge, UK, 2011.
[9] Krueger F. Trim Galore: a wrapper tool around Cutadapt and FastQC.
[10] R.M. Leggett, B.J. Clavijo, L. Clissold, M.D. Clark, M. Caccamo, NextClip: an analysis and read preparation tool for Nextera long mate pair libraries, Bioinformatics. (2014), https://doi.org/10.1093/bioinformatics/btt702.
[11] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.journal. (2011), https://doi.org/10.14806/ej.17.1.200.
[12] R. Kajitani, K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, et al., Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads, Genome Res. (2014), https://doi.org/10.1101/gr.170720.113.
[13] J.T. Simpson, K. Wong, S.D. Jackman, J.E. Schein, S.J.M. Jones, I. Birol, ABySS: a parallel assembler for short read sequence data, Genome Res. (2009), https://doi.org/10.1101/gr.089532.108.
[14] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, Gigascience. (2012), https://doi.org/10.1186/2047-217X-1-18.
[15] R. Chikhi, G. Rizk, Space-efficient and exact de Bruijn graph representation based on a bloom filter, Algorithms Mol. Biol. (2013), https://doi.org/10.1186/1748-7188-8-22.
[16] A.V. Zimin, G. Marçais, D. Puiu, M. Roberts, S.L. Salzberg, J.A. Yorke, The MaSuRCA genome assembler, Bioinformatics. (2013), https://doi.org/10.1093/bioinformatics/btt476.
[17] Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. Oxford University Press; 31:3210–22015;.
[18] Marçais G, Kingsford C. JELLYFISH–fast, parallel k-mer counting for DNA. *Bioinformatics*. 27:764–702011;.
[19] G.W. Vurture, F.J. Sedlazeck, M. Nattestad, C.J. Underwood, H. Fang, J. Gurtowski, et al., GenomeScope: fast reference-free genome profiling from short reads, in: Bioinformatics, Oxford Academic, 2017, https://doi.org/10.1093/BIOINFORMATICS/BTX153.
[20] S. Ou, W. Su, Y. Liao, K. Chougule, J.R.A. Agda, A.J. Hellinga, et al., Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline, in: Genome Biology, BioMed Central Ltd., 2019, https://doi.org/10.1186/S13059-019-1905-Y/FIGURES/6.
[21] J.M. Flynn, R. Hubley, C. Goubert, J. Rosen, A.G. Clark, C. Feschotte, et al., RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families, National Academy of Sciences, Proceedings of the National Academy of Sciences of the United States of America, 2020, https://doi.org/10.1073/PNAS.1921046117.
[22] W. Bao, K.K. Kojima, O. Kohany, Repbase Update, a Database of Repetitive Elements in Eukaryotic Genomes, Mobile DNA. BioMed Central Ltd., 2015, https://doi.org/10.1186/s13100-015-0041-9.
[23] M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences, Curr. Protoc. Bioinformatics (2009), https://doi.org/10.1002/0471250953.bi0410s25.
[24] T. Brůna, K.J. Hoff, A. Lomsadze, M. Stanke, M. Borodovsky, BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database, in: NAR Genomics and Bioinformatics, Oxford Academic, 2021, https://doi.org/10.1093/NARGAB/LQAA108.
[25] T. Brůna, A. Lomsadze, M. Borodovsky, GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR genomics and bioinformatics*, NAR Genom Bioinform (2020), https://doi.org/10.1093/NARGAB/LQAA026.
[26] M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: ab initio prediction of alternative transcripts, in: Nucleic Acids Research, Oxford Academic, 2006, https://doi.org/10.1093/NAR/GKL200.
[27] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods (2015), https://doi.org/10.1038/nmeth.3176.
[28] S. Götz, J.M. García-Gómez, J. Terol, T.D. Williams, S.H. Nagaraj, M.J. Nueda, et al., High-throughput functional annotation and data mining with the Blast2GO suite, in: Nucleic Acids Research, Oxford Academic, 2008, https://doi.org/10.1093/NAR/GKN176.
[29] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, et al., BLAST+: architecture and applications, in: BMC Bioinformatics, BioMed Central, 2009, https://doi.org/10.1186/1471-2105-10-421/FIGURES/4.
[30] P. Jones, D. Binns, H.Y. Chang, M. Fraser, W. Li, C. McAnulla, et al., InterProScan 5: genome-scale protein function classification, in: Bioinformatics, Oxford Academic, 2014, https://doi.org/10.1093/BIOINFORMATICS/BTU031.
[31] Nawrocki EP. Annotating functional RNAs in genomes using Infernal. In: Gorodkin J, Ruzzo WL, editors. *RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods*. Totowa, NJ: Humana Press;.
[32] I. Kalvari, E.P. Nawrocki, J. Argasinska, N. Quinones-Olvera, R.D. Finn, A. Bateman, et al., Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics*, John Wiley & Sons, Ltd, 2018, https://doi.org/10.1002/CPBI.51.
[33] P.P. Chan, T.M. Lowe, tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. *Methods in Molecular Biology*, Humana, New York, NY, 2019, https://doi.org/10.1007/978-1-4939-9173-0_1.
[34] J. Duan, Y. Li, J. Du, E. Duan, Y. Lei, S. Liang, et al., A chromosome-scale genome assembly of Antheraea pernyi (Saturniidae, Lepidoptera), in: Molecular Ecology Resources, John Wiley & Sons, Ltd, 2020, https://doi.org/10.1111/1755-0998.13199.
[35] S.R. Kim, W. Kwak, H. Kim, K. Caetano-Anolles, K.Y. Kim, S.B. Kim, et al., Genome sequence of the Japanese oak silk moth, Antheraea yamamai: the first draft genome in the family Saturniidae, in: GigaScience, Oxford Academic, 2018, https://doi.org/10.1093/GIGASCIENCE/GIX113.
[36] M. Kawamoto, A. Jouraku, A. Toyoda, K. Yokoi, Y. Minakuchi, S. Katsuma, et al., High-quality genome assembly of the silkworm, Bombyx mori, Insect Biochem. Mol. Biol. Pergamon (2019), https://doi.org/10.1016/J.IBMB.2019.02.002.
[37] K. Berlin, S. Koren, C.S. Chin, J.P. Drake, J.M. Landolin, A.M. Phillippy, Assembling large genomes with single-molecule sequencing and locality-sensitive hashing, Nat. Biotechnol. 33 (6) (2015), https://doi.org/10.1038/nbt.3238. Nature Publishing Group.
[38] S. Zhan, C. Merlin, J.L. Boore, S.M. Reppert, The monarch butterfly genome yields insights into long-distance migration, in: Cell, Cell Press, 2011, https://doi.org/10.1016/J.CELL.2011.09.052.
[39] Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, et al. Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive Helicoverpa pest species. *BMC biology*. Springer; 15:1–302017;.
[40] H. Nishikawa, T. Iijima, R. Kajitani, J. Yamaguchi, T. Ando, Y. Suzuki, et al., A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly, in: Nature Genetics 47:4, Nature Publishing Group, 2015, https://doi.org/10.1038/ng.3241.
[41] T. Cheng, J. Wu, Y. Wu, R.V. Chilukuri, L. Huang, K. Yamamoto, et al., Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest, Nat. Ecol. & Evol. 1 (11) (2017), https://doi.org/10.1038/s41559-017-0314-4. Nature Publishing Group.
[42] J. Lee, T. Nishiyama, S. Shigenobu, K. Yamaguchi, Y. Suzuki, T. Shimada, et al., The genome sequence of Samia ricini, a new model species of lepidopteran insect, in: Molecular Ecology Resources, John Wiley & Sons, Ltd, 2021, https://doi.org/10.1111/1755-0998.13259.
[43] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, in: Nucleic Acids Research, Oxford Academic, 2004, https://doi.org/10.1093/NAR/GKH340.
[44] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, in: Bioinformatics, Oxford Academic, 2009, https://doi.org/10.1093/BIOINFORMATICS/BTP348.
[45] B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. Von Haeseler, et al., IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, in: Molecular Biology and Evolution, Oxford Academic, 2020, https://doi.org/10.1093/MOLBEV/MSAA015.
[46] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. Von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, in: Nature Methods, Nature Publishing Group, 2017, https://doi.org/10.1038/nmeth.4285.
[47] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. (2007), https://doi.org/10.1093/molbev/msm088.
[48] S. Kumar, G. Stecher, M. Suleski, S.B. Hedges, TimeTree: a resource for timelines, Timetrees, and divergence times, Mol. Biol. Evol. (2017), https://doi.org/10.1093/molbev/msx116.
[49] X. Zhang, Y. Hu, D.R. Smith, HSDFinder: a BLAST-based strategy for identifying highly similar duplicated genes in eukaryotic genomes, Front Bioinform (2021), https://doi.org/10.3389/fbinf.2021.803176.
[50] D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, in: Genomics,

Proteomics & Bioinformatics, Elsevier, 2010, https://doi.org/10.1016/S1672-0229(10)60008-3.
[51] D.M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics, in: Genome Biology, BioMed Central Ltd., 2019, https://doi.org/10.1186/S13059-019-1832-Y/FIGURES/5.
[52] Y. Wang, H. Tang, J.D. Debarry, X. Tan, J. Li, X. Wang, et al., MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, Nucleic Acids Res. (2012), https://doi.org/10.1093/nar/gkr1293.
[53] V. Bandi, C. Gutwin, J.N. Siri, E. Neufeld, A. Sharpe, I. Parkin, Visualization tools for genomic conservation, Methods Mol. Biol. (2022), https://doi.org/10.1007/978-1-0716-2067-0_16.
[54] T. De Bie, N. Cristianini, J.P. Demuth, M.W. Hahn, CAFE: a computational tool for the study of gene family evolution, Bioinformatics. (2006), https://doi.org/10.1093/bioinformatics/btl097.
[55] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. (2013), https://doi.org/10.1093/molbev/mst010.
[56] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. (2014), https://doi.org/10.1093/bioinformatics/btu033.
[57] A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RAxML web servers, Syst. Biol. (2008), https://doi.org/10.1080/10635150802429642.
[58] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, et al., CDD: NCBI's conserved domain database, Nucleic Acids Res. (2015), https://doi.org/10.1093/nar/gku1221.
[59] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. (2016), https://doi.org/10.1093/molbev/msw054.
[60] S. Liu, S. Zhou, L. Tian, E. Guo, Y. Luan, J. Zhang, et al., Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*, BMC Genomics (2011), https://doi.org/10.1186/1471-2164-12-491.
[61] S. Denecke, I. Rankić, O. Driva, M. Kalsi, N.B.H. Luong, B. Buer, et al., Comparative and functional genomics of the ABC transporter superfamily across arthropods, BMC Genomics (2021), https://doi.org/10.1186/s12864-021-07861-2.
[62] A.L. Price, N.C. Jones, P.A. Pevzner, De novo identification of repeat families in large genomes, in: Bioinformatics, Oxford Academic, 2005, https://doi.org/10.1093/BIOINFORMATICS/BTI1018.
[63] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, in: Nucleic Acids Research, Oxford Academic, 1999, https://doi.org/10.1093/NAR/27.2.573.
[64] O. Kohany, A.J. Gentles, L. Hankus, J. Jurka, Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor, in: BMC Bioinformatics, BioMed Central, 2006, https://doi.org/10.1186/1471-2105-7-474/FIGURES/2.
[65] B. Feng, K. Zheng, C. Li, Q. Guo, Y. Du, A cytochrome P450 gene plays a role in the recognition of sex pheromones in the tobacco cutworm, *Spodoptera litura*, Insect Mol. Biol. (2017), https://doi.org/10.1111/imb.12307.
[66] J.-C. Simon, E. d'Alençon, E. Guy, E. Jacquin-Joly, J. Jaquiéry, P. Nouhaud, et al., Genomics of adaptation to host-plants in herbivorous insects, Brief Funct Genom. (2015), https://doi.org/10.1093/bfgp/elv015.
[67] W. Zhang, W. Chen, Z. Li, L. Ma, J. Yu, H. Wang, et al., Identification and characterization of three new cytochrome P450 genes and the use of RNA interference to evaluate their roles in antioxidant defense in Apis cerana cerana Fabricius, Front. Physiol. (2018), https://doi.org/10.3389/fphys.2018.01608.
[68] D.-D. Wei, E.-H. Chen, T.-B. Ding, S.-C. Chen, W. Dou, J.-J. Wang, De novo assembly, gene annotation, and marker discovery in stored-product pest Liposcelis entomophila (Enderlein) using transcriptome sequences, PLoS One (2013), https://doi.org/10.1371/journal.pone.0080046.
[69] L. Yu, W. Tang, W. He, X. Ma, L. Vasseur, S.W. Baxter, et al., Characterization and expression of the cytochrome P450 gene family in diamondback moth, Plutella xylostella (L.), Sci. Rep. (2015), https://doi.org/10.1038/srep08952.
[70] M. Schwartz, V. Boichot, S. Fraichard, M. Muradova, P. Senet, A. Nicolai, et al., Role of insect and mammal glutathione transferases in Chemoperception, Biomolecules. (2023), https://doi.org/10.3390/biom13020322.
[71] Q. Yu, C. Lu, B. Li, S. Fang, W. Zuo, F. Dai, et al., Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*, Insect Biochem. Mol. Biol. (2008), https://doi.org/10.1016/j.ibmb.2008.08.002.
[72] N. Durand, M.-A. Pottier, D. Siaussat, F. Bozzolan, M. Maïbèche, T. Chertemps, Glutathione-S-transferases in the olfactory organ of the noctuid moth Spodoptera littoralis, diversity and conservation of chemosensory clades, Front. Physiol. (2018), https://doi.org/10.3389/fphys.2018.01283.
[73] H. Shi, L. Pei, S. Gu, S. Zhu, Y. Wang, Y. Zhang, et al., Glutathione S-transferase (GST) genes in the red flour beetle, Tribolium castaneum, and comparative analysis with five additional insects, Genomics. (2012), https://doi.org/10.1016/j.ygeno.2012.07.010.
[74] C. Wu, S. Chakrabarty, M. Jin, K. Liu, Y. Xiao, Insect ATP-binding cassette (ABC) transporters: roles in xenobiotic detoxification and Bt insecticidal activity, Int. J. Mol. Sci. (2019), https://doi.org/10.3390/ijms20112829.
[75] R. Labbé, S. Caveney, C. Donly, Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera, Insect Mol. Biol. (2011), https://doi.org/10.1111/j.1365-2583.2010.01064.x.
[76] A.S. Strauss, D. Wang, M. Stock, R.R. Gretscher, M. Groth, W. Boland, et al., Tissue-specific transcript profiling for ABC transporters in the sequestering larvae of the phytophagous leaf beetle Chrysomela populi, PLoS One (2014), https://doi.org/10.1371/journal.pone.0098637.
[77] K. Adarsh Gupta, K. Mita, K.P. Arunkumar, J. Nagaraju, Molecular architecture of silk fibroin of Indian golden silkmoth, Antheraea assama, Sci. Rep. 5 (1) (2015), https://doi.org/10.1038/srep12706. Nature Publishing Group.
[78] A. Garel, G. Deleage, J.C. Prudhomme, Structure and organization of the Bombyx mori sericin 1 gene and of the sericins 1 deduced from the sequence of the Ser 1B cDNA, Insect Biochem. Mol. Biol. Pergamon (1997), https://doi.org/10.1016/S0965-1748(97)00022-2.
[79] H. Okamoto, E. Ishikawa, Y. Suzuki, Structural analysis of sericin genes. Homologies with fibroin gene in the 5′ flanking nucleotide sequences, J. Biol. Chem. Elsevier (1982), https://doi.org/10.1016/S0021-9258(18)33412-4.
[80] Y. Takasu, H. Yamada, T. Tamura, H. Sezutsu, K. Mita, K. Tsubouchi, Identification and characterization of a novel sericin gene expressed in the anterior middle silk gland of the silkworm Bombyx mori, in: Insect Biochemistry and Molecular Biology, Pergamon, 2007, https://doi.org/10.1016/J.IBMB.2007.07.009.
[81] P. Couble, J.J. Michaille, A. Garel, M.L. Couble, J.C. Prudhomme, Developmental switches of sericin mRNA splicing in individual cells of Bombyx mori silkgland, in: Developmental Biology, Academic Press, 1987, https://doi.org/10.1016/0012-1606(87)90496-9.
[82] R. Ahmad, A. Kamra, S.E. Hasnain, Fibroin silk proteins from the nonmulberry silkworm Philosamia ricini are biochemically and immunochemically distinct from those of the mulberry silkworm Bombyx mori, DNA Cell Biol. (2004), https://doi.org/10.1089/104454904322964742.
[83] R. Dash, S. Mukherjee, S.C. Kundu, Isolation, purification and characterization of silk protein sericin from cocoon peduncles of tropical tasar silkworm, Antheraea mylitta, Int. J. Biol. Macromol. Elsevier (2006), https://doi.org/10.1016/J.IJBIOMAC.2006.03.001.
[84] A. Goswami, N. Goswami, A. Bhattacharya, P. Borah, D. Devi, Composition and in silico structural analysis of fibroin from liquid silk of non-mulberry silkworm Antheraea assamensis, Int. J. Biol. Macromol. (2020), https://doi.org/10.1016/j.ijbiomac.2020.08.232.
[85] A. Goswami, D. Devi, Structural insight on the liquid silk from the middle silk gland of non-mulberry silkworm Antheraea assamensis, J. Biomol. Struct. Dyn. (2023), https://doi.org/10.1080/07391102.2021.2017347.
[86] T. Sakudoh, H. Sezutsu, T. Nakashima, I. Kobayashi, H. Fujimoto, K. Uchino, et al., Carotenoid silk coloration is controlled by a carotenoid-binding protein, a product of the yellow blood gene, Proc. Natl. Acad. Sci. U. S. A. (2007), https://doi.org/10.1073/pnas.0702860104.