OPEN ACCESS

ARTICLE

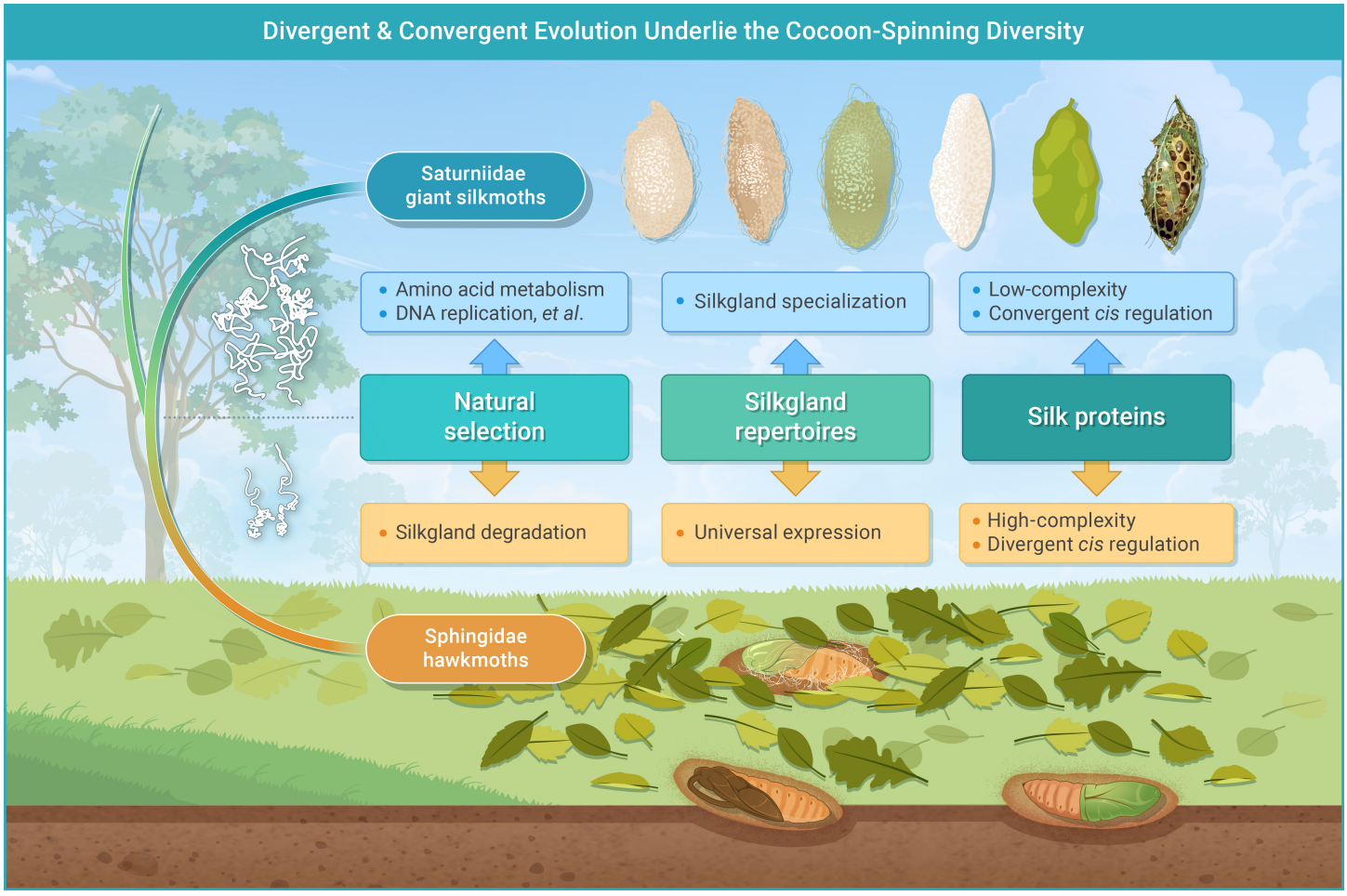
Divergent and convergent evolution underlie the cocoon-spinning diversity

Yong Cui,[1,9,10] Gang-Qi Fang,[2,10] Jian Zhong,[3,8,10] Xing-Rong Bai,[3,10] Han Xue,[1,10] Mei-Yan Yi,[1] Bao-Zhen Chen,[1] Ya-Nan Zhu,[4] Wen Wang,[4,5] Yan-Qun Liu,[6,*] Shuai Zhan,[2,*] and Hui Xiang[1,4,7,*]

*Correspondence: xianghui@m.scnu.edu.cn (H.X.); szhan@sibs.ac.cn (S.Z.); liuyanqun@syau.edu.cn (Y.-Q.L.)

Received: December 18, 2024; Accepted: June 17, 2025; Published Online: June 30, 2025; https://doi.org/10.59717/j.xinn-life.2025.100144

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- The Saturniid lineage diverged from Sphingids but convergently evolved silk-spinning cocoons with Bombycids.
- The two lineages differ in expression and functions of lineage-specific, positively selected, silk gland–specific genes.
- Bombycoidea retain diverse silk proteins; cocoon-silk lineages converge in fibroin regulation unlike non-cocoon ones.

ARTICLE

OPEN ACCESS

www.the-innovation.org/life

Divergent and convergent evolution underlie the cocoon-spinning diversity

Yong Cui,[1,9,10] Gang-Qi Fang,[2,10] Jian Zhong,[3,8,10] Xing-Rong Bai,[3,10] Han Xue,[1,10] Mei-Yan Yi,[1] Bao-Zhen Chen,[1] Ya-Nan Zhu,[4] Wen Wang,[4,5] Yan-Qun Liu,[6,*] Shuai Zhan,[2,*] and Hui Xiang[1,4,7,*]

[1]Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology, Guangzhou Key Laboratory of Insect Development Regulation and Application Research, Institute of Insect Science and Technology, School of Life Sciences, South China Normal University, Guangzhou 510631, China

[2]Key Laboratory of Plant Design, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[3]Sericultural and Apicultural Research Institute, Yunnan Academy of Agricultural Sciences, Mengzi 661101, China

[4]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Kunming 650201, China

[5]School of Ecology and Environment, Northwestern Polytechnical University, Xi'an 710072, China

[6]College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang 110866, China

[7]Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China

[8]Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Kunming 650204, China

[9]College of Animal Science and Technology, Jiangxi Agricultural University, Nanchang 330045, China

[10]These authors contributed equally to this work

*Correspondence: xianghui@m.scnu.edu.cn (H.X.); szhan@sibs.ac.cn (S.Z.); liuyanqun@syau.edu.cn (Y.-Q.L.)

Received: December 18, 2024; Accepted: June 17, 2025; Published Online: June 30, 2025; https://doi.org/10.59717/j.xinn-life.2025.100144

Citation: Cui Y., Fang G.-Q., Zhong J., et al. (2025). Divergent and convergent evolution underlie the cocoon-spinning diversity. *The Innovation Life* **3**:100144.

Silk production and use are widespread and highly diversified in insects, yet the evolution and molecular bases underlying this diversity remain enigmatic. The Bombycoidea insects consist of sublineages that spin large cocoons (silk lineages) and non-silk cocoons, providing an ideal model system to study how cocoon-spinning behaviors rapidly diversify between species. By analyzing 11 Bombycoidea genomes, this study explored the genomic and transcriptomic divergence underlying the differences in silk traits between sublineages. Specific genes presented in silk lineage Saturniids exhibit biased expression and genes under selection in Saturniids are related to silk production, which contrasts with those in the non-silk-cocoon lineage Sphingids. Moreover, the gene repertoires of silk glands in the silk lineage species were dominated by genes with specialized expression and enriched in amino acid metabolism. We further characterized a highly dynamic catalog of silk protein genes within Bombycoidea, particularly including an extreme diversity of fibroin sequences across species, as well as the convergent regulation of fibroin expression between the Bombycoidea sublineages that spin large cocoons. This study suggests that the silk lineage Saturniid may diverge from its sister non-silk-cocoon lineage Sphingid, but converge with the outgroup Bombycid in spinning silk cocoons and uncovers a comprehensive scenario in which diversifying and convergent selection act on different processes of silk use, jointly shaping the diversity of ecological adaptations in insects.

INTRODUCTION

Silk production and secretion are widespread in insects. Twenty-three lineages of silk-producing insects have been documented and multiple independent origins of silk production across these lineages have been inferred.[1] However, not all these insect lineages produce massive silk or use silk. Even among very closely-related species, whether and how silk is used may vary greatly and have evolved discontinuously. To date, the genomic basis underlying diversity of silk-spinning in insects remains largely unknown.

Of the many outcomes of silk use, silk cocoons attract broad attention due to their natural diversity and commercial application for manufacture. Spinning silk cocoons is considered a protective strategy against adverse environments and is employed by several lineages, particularly in Lepidoptera. More than 150,000 described lepidopteran species spin cocoons.[1-3] The Bombycoidea is an ecologically diverse superfamily and renowned for its characteristic cocoons. This superfamily includes the domestic silkworm, *Bombyx mori*, from the family Bombycidae, which has been commercially exploited for sericulture and serves as a study model of Lepidoptera. Different families of Bombycoidea display highly diverse forms of cocoon spinning. Notably, species of Saturniidae (also called giant silkmoths) produce a variety of characteristic cocoons, including the largest cocoon in nature. The cocoons of this family are largely different compared with those of Bombycidae and even vary among the species within the family in terms of weight, shape, and color (Table 1). By contrast, most species of the family Sphingidae (also called hawkmoths) spin debris pieces with limited silk to build an "earthy" cocoon (Table 1). Thus, Bombycoidea provides an ideal model system to study the rapid diversification and evolution of cocoon-spinning behaviors.

Silk production is influenced by a series of factors, including silk gland development, the function of silk proteins synthesized in the gland, and nutritional status. Much of the knowledge regarding silk components and silk production is developed in *B. mori*.[4,5] In *B. mori*, the silk fibroin consists of fibroin heavy chain (encoded by *fib-H*) and fibroin light chain (encoded by *fib-L*) that form heterodimers and a fibrohexamerin (encoded by *p25*) protein that prompts the formation of fibroin complexes.[6] Silk fibroins were coated with sericins (encoded by *sericin*).[6] However, this composition model of silk proteins is not conserved and has been reported to be partially lost in certain lepidopteran species.[7,8] In addition, sequences of silk proteins were different among different lepidopteran species.[9,10] Thus, obtaining complete, continuous genomic resources is critical for the precise investigation of silk protein evolution across species.

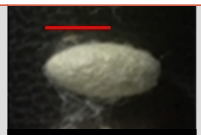
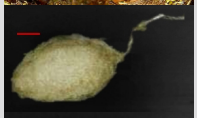
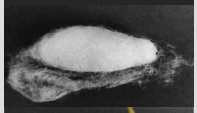
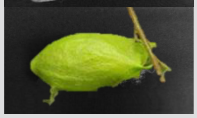
Here, we focus on Saturniidae, Bombycidae, and Sphingidae (SBS), three families that display great diversity in cocoon traits and harbor the majority of the described species in Bombycoidea. By generating six high-quality *de novo* reference genomes and integrating them with seven other publicly available resources, we performed large-scale comparative genomic analyses on the representative species to uncover the genomic and molecular bases underlying the diversification of silk traits between taxa. These findings will not only improve the understanding of the adaptive roles of silk use but also provide a valuable genomic resource for future studies on silk production in insects.

MATERIALS AND METHODS

Genome sequencing and *de novo* assembly

To carry out comparative genomic analysis, we firstly sequenced and assembled the genomes of five Bombycoidea species and one Sphingidae species (Table S1). The pupae of Saturniidae species, i.e., *Antheraea assama*, *Dictyoploca japonica*, *Rhodinia fugax*, and the 5th instar larvae of Saturniidae *Samia ricini* were provided by the Sericultural and Apicultural Research Institute, Yunnan Academy of Agricultural Sciences, Mengzi County, Yunnan Province, China. The 5th instar larvae of Saturniidae *A. yamamai* were provided by Nanling Oak silkworm Agricultural Development Co., Ltd, Huaihua, Hunan Province, China. The pupae and the 5th instar larvae of Sphingidae *Clanis bilineata* were collected from Cangshan Farm, Linyi, Shandong Province, China. Genomic DNA was isolated from an individual pupa or silk gland of the larvae, using the next standard protocol. Briefly, tissue was ground in liquid nitrogen, then homogenized and digested in SDS buffer with

Table 1. Cocoon features of the Bombycoidea species.

Family	Species	Image	Features
Bombycidae	*Bombyx mandarina*		Small light-yellow cocoon with thin silk fiber
Bombycidae	*B.mori*		Middle to big cocoon with thin to thick fiber, diverse color and shapes by domestication
Sphingidae	*Clanis bilineata*		"Earthy" cocoon
Sphingidae	*Manduca sexta*		"Earthy" cocoon
Sphingidae	*Hyles vespertilio*		"Earthy" cocoon
Saturniidae	*Antheraea pernyi*		Very big cocoon with thick fiber, varied in light yellow and green
Saturniidae	*A. yamamai*		Very big green cocoon with thick fiber
Saturniidae	*A.assama*		Very big darkorange cocoon with thick fiber
Saturniidae	*Dictyoploca japonica*		Very big brown "lattice" cocoon, uncharactered fiber
Saturniidae	*Samia ricini*		Middle white cocoon with thin and brittle fiber
Saturniidae	*Rhodinia fugax*		Very big emerald cocoon with thin and brittle fiber

proteinase K, followed by phenol and chloroform extraction, RNase treatment, and isopropanol precipitation.

Genomes of *A. assama*, *D. japonica* and *S. ricini* were sequenced using PacBio long-read sequencing platform Sequel II and Illumina sequencing platform X-Ten according to manufacturer's instructions. Library construction and sequencing were performed by the Berry Genomics Co. Ltd. (Beijing, China). Canu v1.7[11] was used to correct and trim reads using "genomeSize=450m corOutCoverage=500 corMinCoverage=2 minReadLength=2000". The longest 40-55X reads were used to assemble the genome using SMARTdenovo v1.0,[12] with "-c 1 -k 31" for *D. japonica* and *S. ricini* and "-c 1 -k 23" for *A. assama*. ARROW v2.2.2[13] was used to polish the assembly using long reads. Pilon v1.22[14] was used to further polish the assembly using Illumina reads. Purge_haplotigs v1.0.4[15] was used to remove the redundancy for two rounds. Genomes of *A. yamamai* and *C. bilineata* were sequenced on Oxford Nanopore Technology sequencing platform MinION and Illumina sequencing platform X-Ten according to manufacturer's instructions. Library construction and sequencing were performed by Nextomics Co. Ltd. (Wuhan, China). Canu v1.7[11] was used to correct and trim reads using "genomeSize=450m corOutCoverage=500 corMinCoverage=2 minReadLength=2000". The longest 45X corrected reads were used to assemble the genome using SMARTdenovo v1.0[12] using "-k 21 -c 1". Genome assembly was polished using Racon v1.4.11 [16] with long reads and using pilon v1.22[14] with Illumina reads. Finally, genome assembly was processed for redundancy removal using purge_haplotigs v1.0.4.[15]

The genome of *R. fugax* was sequenced on Illumina sequencing platform X-Ten. Library construction and sequencing were performed by the Berry Genomics Co. Ltd. (Beijing, China).Two short-insert paired-end libraries (500 bp and 800 bp) were constructed and three long-insert mate-pair libraries (3 kb, 8 kb, and 12 kb) were constructed. For each short insert library, 3 μg of DNA was fragmented, end-repaired, A-tailed, and ligated to Illumina paired-end adapters. The ligated fragments were size-selected at 500 bp and 800 bp and amplified by LM-PCR. For each long insert size mate-pair library, 20 μg of genomic DNA was sheared to the desired insert size. DNA fragments were end-repaired using biotinylated nucleotide analogues (Illumina) and circularized by intramolecular ligation. Circular DNA molecules were sheared with Adaptive Focused Acoustic (Covaris) to an average size of 500 bp. Biotinylated fragments were purified on magnetic beads (Invitrogen, USA), end-repaired, A-tailed and ligated to Illumina paired-end adapters, size-selected again and amplified by LM-PCR. Library construction and sequencing were entrusted to the Chinese National Human Genome Center (Shanghai, China).

Low-quality bases of raw reads were filtered out using Platanus_trim v1.0.7[17] and then subjected to genome assembly. Short reads were first assembled into contigs using DiscovarDeNovo.[18] Contigs were further assembled into scaffolds with the mate pair libraries using Scaffmatch v3.[19] Finally, intra-scaffold gaps were filled with short reads by GapCloser (available in SOAPdenovo).[20]

Assembled contigs of *A. yamamai*, *D. japonica* and *C. bilineata* were further anchored to chromosomes based on chromosome conformation contact (Hi-C) mapping. Released Hi-C sequencing data of *A. yamamai* (DRX465204) and *D. japonica* (SRX15765496) were downloaded from the NCBI SRA database. For *C. bilineata*, Hi-C sequencing libraries were constructed according to manufacturer's instructions and sequenced by Novagene Co. Ltd. (Beijing, China). Hi-C reads were processed by Juicer v1.5.7[21] using default parameters to generate the contact information, which was subsequently used to group the chromosomes using 3d-dna v180922[22] with'q 1 -sort-output -build-gapped-map' Assembled contigs of *S. ricini* were mounted to the published *S. ricini* chromosome[8] using RagTag v2.1.0.[23]

The completeness of the assembled genomes was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO v3.0.2) with 1658 homologous genes of insecta_odb9 (https://busco-archive.ezlab.org/v2/datasets/insecta_odb9.tar.gz).[24]

Detection and removal of contaminant sequences in genome assemblies

To identify and eliminate potential contamination in the assembled genome, sequences were aligned against the NCBI non-redundant nucleotide (nt) database using the BLASTN algorithm with stringent parameters: -max_target_seqs 10 -max_hsps 1 -evalue 1e-25. For each unique query position, the top-scoring hit was selected based on alignment quality. Taxonomic classification of these hits was then analyzed to detect sequences of foreign origin, including bacterial, fungal, or viral taxa. Contaminant sequences were systematically excised from the assembly based on their positional context: regions flanking contaminants located at sequence termini were entirely removed, while internal contaminants were resolved by splitting and trimming the affected segment. In addition, we also aligned the next-generation sequencing (NGS) data to the reference genomes and analyzed the sequencing depth. We investigated regions that showed notably lower depth than average, which indicated contamination, for potential misassemblies. These were corrected, where necessary, by splitting the sequences. For the Hi-C-enhanced assemblies, we also used Juicebox in the software Juicer v1.5.7[22] to visualize the Hi-C contact heatmaps, which helped in manually refining and validating the chromosomal structures.

RNA-seq

To annotate gene structures and to analyze spatial expression pattern based on transcriptomes of two representative species, *A. yamamai* for Saturniidae and *C. bilineata* for Sphingidae, we generated RNA-seq on the corresponding species and tissue samples. The live animals were brought into the lab for dissection. Different tissues from the mid-stage of the final instar larvae of the five species (*A. assama*, *S. ricini*, *D. japonica*, *R. fugax* and *C. bilineata*) were dissected respectively, quickly frozen in liquid nitrogen and stored at -80°C (see Table S4) for further use. The stored samples were packed in dry ice and transported to the Novogene Co. Ltd. for RNA isolation, purification, library construction and sequencing. Each candidate sample was ground in liquid nitrogen and total RNA was independently isolated and purified using the TRIzol extraction reagent (Thermo Fisher Scientific, USA). Construction of the cDNA library and paired-end RNA-seq (Illumina) were carried out in accordance with the manufacturer's protocol. Moreover, RNA-seq data of *A. yamamai*, *B. mori*, *Manduca sexta*, *Hyphantria cunea* and *Spodoptera litura* were also used for subsequent analyses. Statistics of transcriptome sequencing data are listed in Table S4. Raw data were filtered using seqtk v1.0[25] and then the resultant clean data were used for further analyses.

Identification of the sex chromosome

We conducted BLASTP for protein sequences of the genes located in the sex chromosome of *B.mori* against the annotated proteins in *A. yamamai*, *C. bilineata*, and *D. japonica* that have chromosome-level assemblies, with an e-value cutoff with 1e-5. Next, we identified best hits between *B.mori* and every other species. We detected the ortholog chromosome by genomic syntenic analysis by MCScanX.[26] The synteny plot was generated using online software synvisio (https://synvisio.github.io/).

Genome annotation

We annotated transposable elements and the predicted genes in the assembled genomes. For each of our assembled genomes, we identified repetitive sequences and transposable elements using RepeatMasker v4.1.1[27] against a *de novo* repeat library that was built by RepeatModeler v2.0.1,[27] as well as the arthropod set of Repbase v1.40.

Three independent approaches were generated for gene prediction, using the repeat-masked genome assembly. For homology-based annotation, gene sets of *B. mori*, *Danaus plexippus*, *Helicoverpa armigera*, *Papilio xuthus*, *Plutella xylostella*, and *S. litura* were used for homology search by TBLASTN with e<10^{-5}. Prediction of genic structure was generated by Genewise v2.4.1. *ab initio* prediction was generated using AUGUSTUS v3.[28] As to transcript-based annotation, all available clean RNA-seq reads for each species were mapped to the corresponding assembled genome using STAR[29] and prediction of transcripts were generated using StringTie v2.0.[30] Finally, EVidenceModeler v1.1.1[31] was used to merge all the independent predicted gene sets.

Annotation of gene function was generated by searching the homology with such databases as NCBI RefSeq and UniProt. In addition, GeneOntology annotation, KEGG KO annotation and protein domain annotation were generated using InterProScan (IPR) software with all the corresponding databases embedded.

Orthology analysis

For the gene set of each candidate species, the longest transcript for each gene and the predicted protein sequence was used for orthology identification. OrthoFinder v2.3.3[32] was used to sort orthologs across species using "-S blast -M msa".

Phylogenetic analysis, divergence time estimation as well as Gene family analyses

Data for comparative genomics analysis contained *A. assama*, *A. yamamai*, *S. ricini*, *D. japonica*, *R. fugax*, *C. bilineata* and other seven species from published papers including *B. mori*,[33] *A. pernyi*,[34] *Hyles vespertilio*,[35] *M. sexta*,[36] *B. mandarina*[37], *H. cunea*[38] and *S. litura*.[39] Single-copy gene families were retrieved from the identified orthologs and used for phylogenetic tree construction with RAxML software[40] based on the maximum likelihood method using GAMMA and the JTT model for 100 replicates of bootstrapping (parameters were set as -f a -x 12345 -p 12345 -# 100 -m PROTGAMMAJTT -s fasta -n tre -k -T 20). r8s v1.81[41] was applied to estimate the 95% confidence intervals of the differentiation times (parameters were set as MRCA asb Aper Slit; fixage taxon=asb age=90; MRCA asc Bmor Msex; constrain taxon=asc min_age=70 max_age=80), where the published timings for the divergence of different species were obtained with the TimeTree database.[42]

Lineage-specific genes were further identified from the above identified orthologs, with the highest stringency. Specifically, a gene was defined as lineage-specific if it was present in all species of one lineage but absent in any species of the other lineages. To verify the Saturniidae-specific genes and Sphingidae-specific genes that we further analyzed in this study, we retrieved all the newly chromosome-level genomes from the three SBS (Saturniidae, Bombycidae, and Sphingidae) species in CNCB (China National Center for Bioinformation) database, specifically, genome of *Trilocha varians* (GCA_030269945.2) from Bombycidae, genomes of *Actias luna* (GCA_039707435.1), *A. mylitta* (GCA_014332785.1) and *Saturnia pavonia* (GCA_947532125.1) from Saturniidae, and genomes of *Amorpha juglandis* (GCA_949126905.1), *Cephonodes hylas* (GCA_030295005.1), *Deilephila elpenor* (GCA_949752805.1), *Deilephila porcellus* (GCA_905220455.2), *Hemaris fuciformis* (GCA_907164795.1), *Hyles euphorbiae* (GCA_023078785.2), *Hyles lineata* (GCA_030180105.1), *Laothoe populi* (GCA_905220505.1), *Lapara coniferarum* (GCA_949316025.1), *Mimas tiliae* (GCA_905332985.1), *Sphinx pinastri* (GCA_947568825.1) and *Theretra japonica* (GCA_033459515.1) from Sphingidae.The protein sequences of lineage-

specific genes were aligned to each genome sequence of the other lineages using miniprot,[43] setting an e-value threshold of 10^{-5} and a sequence coverage cutoff of 30% to filter false positives. Totally, 83 out of the 110 originally identified Sphingidae-specific genes were successfully validated through testing with the newly included genomes of three Saturniidae and one Bombycidae species; and 75 out of the 118 Saturniidae-specific genes were successfully validated using genomes from twelve Sphingidae and one Bombycidae species.

Gene family expansion and contraction analyses were performed using CAFÉ 4.2.1.[44] For each branch and node, an "expanded and contracted gene family" with both an overall *P* value (family-wide *P* value) and an exact *P* value (Viterbi method) ≤0.01 was defined as a "significantly expanded and contracted gene family."

Ancestral state reconstruction of cocooning

To study the association of cocooning with the divergence of Saturniidae and Sphingidae, we used the R package phytools 2.0[45] to reconstruct the evolution of cocooning in the SBS group based on the phylogeny of the SBS we constructed, through a Mk trait evolution model.

Quantification of gene expression and identification of tissue-specific genes

The independent RNA-sequenced clean reads were mapped to the annotated gene set and determined expression profiles as fragments per kilobase of transcript per million mapped reads (FPKM) by RSEM v1.3.040.[46]

To infer the potential functions of the silk gland, we sought to analyze the spatial expression pattern based on the transcriptomes of two representative species, *A. yamamai* for Saturniidae and *C. bilineata* for Sphingidae (Table S4). RNA-seq data from the anterior silk gland, posterior silk gland, fatbody, midgut, integument, malpighian tubule, brain, hemocyte, testis and ovary of *A. yamamai* and those from the silk gland, fatbody, midgut, integument, Malpighian tubule, brain, antenna and wing of *C. bilineata* were used to calculate the tissue-specific expression. For each species, genes showing tissue-specific expression were identified by a Tau-value based approach.[47] Firstly, we calculated the Tau value for all genes. If a gene had a Tau value larger than 0.8 and the expression level for the silk gland was larger than 1 and more than twice that in the tissue with the second highest expression, the gene was then identified as being silk-gland-specific highly expressed genes. Chi-square test was used to test the statistical significance of over-representation concerned with the percentage of the foreground genes against that of the background genes. For instance, we took *A. yamamai* as a model to test the overrepresentation of Saturniidae-specific genes in a hemolymph-specific expression mode. Chi-square test was used to test whether the percentage of those hemolymph-specific expressed ones in all the Saturniidae-specific gene orthologs in *A. yamamai* was significantly higher than the percentage of all hemolymph-specific expressed genes among all the expressed genes in *A. yamamai*. Similar tests were also used in other cases.

Evolutionary analyses

We used the single-copy orthologous genes to identify the rapidly evolving genes. A variety of maximum likelihood tests were used for adaptive evolution using PAML 4.8.[48,49] We applied several different models in the program CODEML.[49] Branch models allow the *w* ratio to vary across lineages in a phylogeny,[50] and branch-site models[51-53] allow the *w* ratio to vary among sites along branches on the tree, such that positive selection can be detected in specific codon sites along particular lineages of interest. For the branch models, we compared the 1-ratio model, which estimates a single *w* ratio for all SBS lineages, with the 2-ratio model, which estimates a *w* ratio for the branch connecting Saturniidae / Sphingidae (foreground branch) and a second *w* ratio for all other branches (background branches). A likelihood ratio test was used to test the significance of the better model. These tests were performed between the null model and the alternative model to determine whether the rates of evolution were significantly different between the foreground branch and the background branches. For the branch-site model, LRT was conducted to compare a model that allowed sites to be under positive selection on the foreground branch with the null model in which sites could evolve either neutrally and under purifying selection. BEB was used to calculate the posterior probability for each site. BEB output was used to determine which sites are under positive selection in cases in which the LRT was significant.

Functional enrichments

KEGG enrichments were generated in two species, i.e., *A. yamamai* and *C. bilineata*, respectively, using the OmicShare online platform (http://www.omicshare.com/tools), in order to explore the functional implication of the interested gene sets. Specifically, as to the enrichment analysis on the 113 genes under selection in Saturniidae, the orthologous genes in *A. yamamai* were selected as targets and used its whole annotated genes as background. Similarly, as to analysis on the 369 genes under selection in Sphingidae, the orthologous genes in *C. bilineata* were selected as targets and used its whole annotated genes as background. As to enrichment analyses on the 451 and 226 genes (termed as SGH genes) with extremely high expression in the silk glands of *A. yamamai* and *C. bilineata,* respectively, the SGH genes were set as targets and the whole annotated gene set of the corresponding species was set as background. Hypergeometric and FDR multiple tests were utilized to identify significantly enriched pathways.

Local genomic synteny analysis

We generated local genomic synteny analysis on regions covering the *fib-L* and *p25* respectively, to illustrate evolutionary loss or gain of the genes. Specifically, we conducted TBLASTN for FibL and P25 protein sequences from *B.mori* against other 9 species' genomes with an e-value cut-off of $1x10^{-5}$, and combined local alignments with the SOLAR. Next, we identified the best TBLASTN hits between *B.mori* and every other species on the basis of the following parameters: alignment score, alignment rate and identity. The synteny plot was generated using an in-house-generated script.

Codon usage and tRNA prediction

To investigate the translational features of the fibroin genes, we used two approaches, i.e., the codon usage frequency of the fibroin genes using CodonW1.4.2[54] (https://sourceforge.net/projects/codonw/), and tRNA prediction using tRNAscan-SE 2.0.9.[55]

Semi-quantitative PCR validation of expression of silk genes in *C. bilineata*

The silk gland, fat body, midgut, epidermis and Malpighian tube were dissected from the 6th instar larvae of *C. bilineata* and were ground in liquid nitrogen, respectively. RNA was extracted from each tissue using TRIzol (Invitrogen, USA) and purified with chloroform. Two μg purified RNA was treated with 2 units of DNase I to remove trace amounts of genomic DNA. Reverse transcription was performed using the Reverse Transcriptase M-MLV Kit according to the manufacturer's protocols (TaKaRa, China). The primers used for semi-quantitative PCR of *fib-H*, *fib-L* and *p25* respectively, are listed in Table S14. We have confirmed that the RT-PCR primers span an exon boundary.

Bmfib-H promoter cloning via overlap extension PCR

The upstream regulatory region of *Bmfib-H*, spanning from -2211 bp to +24 bp, was amplified based on the *B. mori* genome sequence obtained from SilkDB 3.0 and subsequently cloned into the pMD-18 T vector (TaKaRa, China). Truncated fragments of the *Bmfib-H* regulatory region, starting at position +24 bp and extending to -202 bp, were generated through PCR amplification using the *Bmfib-H*-Promoter (-2211/+24 bp) plasmid as a template. These fragments were then ligated into the luciferase reporter plasmid, pGL3-basic vector (Promega, USA). Mutant *Cbfib-H*-Promoter (-181/+24 bp) constructs, with mutations at sites -181 to -166 nt, were created using the overlap extension PCR method, employing the *Bmfib-H*-Promoter (-202/+24 bp) plasmid as the template. The presence of mutations was confirmed via DNA sequencing, and the resulting construct was inserted into the pGL3-basic vector for subsequent activity assays.

Luciferase assay for promoter activity

A luciferase assay was used to compare the activity of *fib-H* promoter in

B.mori and *C. bilineata*. BmE cells were inoculated into culture media in 24-well culture plates (Corning, USA) and cultured for 12h. Cell transfection and co-transfection were conducted when the cell density achieved 80%. To normalize the firefly luciferase activity, the Renilla luciferase vector pRLSV40 (Promega, USA) was co-transfected with each of the pGL3-derived vectors. For transfection, a mix of 50 μL containing 800 ng of *Bmfib-H*-Promoter (−202/+24bp) or *Cbfib-H*-Promoter (−181/+24bp) -pGL3 reporter plasmid DNA, 50 ng internal control plasmid and 1.5 μL Fugene HD transfection reagent (Promega, USA) in the Opti-MEM reduced serum medium (Invitrogen, USA) was added to BmE cells in Grace medium. The cells were cultured for 48 h at 28 °C before the promoter activity assay. The transfected cells were washed twice with PBS and lysed in 150 μL Passive Lysis Buffer (Promega, USA). Samples were centrifuged at 1,000 g for 5 min at room temperature. The supernatant was used to analyze the luciferase activity using the Dual-Luciferase Assay System according to the instructions of the Dual-Luciferase® Reporter Assay System Kit (Promega, USA). The luciferase activity was normalized to the Renilla luciferase activity. All assays were repeated at least three times. The luciferase activity was represented as mean ± the standard error (SE). Statistical significance of the luciferase activity was analyzed using Student's t-test.

Electrophoretic mobility shift assay (EMSA) on binding activity of *cis*-element of *fib-H*

EMSA was conducted to compare the binding activity of the cis-element of *fib-H* in the three different species with nucleoproteins of *B.mori* silk gland, using the Light Shift Chemiluminescent EMSA Kit (Thermo Fisher Scientific, USA). Oligonucleotides from the -202 to -184nt sequence upstream *fib-H* start codon of the *C. bilineata* and *B.mori* were labeled with biotin at the 5' end and heated at 95 °C for 10 min in 50 mM Tris-acetate buffer at pH 4.1 and slowly cooled to room temperature. The biotin-labeled oligonucleotides were synthesized by Invitrogen.

The oligonucleotide probes used are shown in Table S14. Reactions were conducted in a 20 μL mix (containing 1 × binding buffer, 2.5% glycerol, 0.05% NP-40, 50 mM KCl, 5 mM $MgCl_2$, 4 mM EDTA, 1 μg poly dI-dC, 2 μg nuclear proteins and 20 fmol of a biotinylated end-labeled probe) at room temperature for 20 min. For the competition assay, cold probes (unbiotinylated) were added to the binding reaction. The samples were then separated in 6% polyacrylamide gels on ice at 100 V for 1.5 h. After electrophoresis, the gels were blotted onto a Nylon positively charged membrane (Amersham Biosciences, USA). The membranes were then developed using the Light Shift Chemiluminescent EMSA kit according to the manufacturer's protocol.

RESULTS

High-quality genomic resources of representative Bombycoidea species

We selected 11 representative species with diversified cocoon features from Saturniidae, Sphingidae, and Bombycidae for comparison (Table 1). Six *de novo* reference genomes were generated to fill the paucity of high-quality genomic resources for these taxa (Table S1). Long-read sequencing technology was employed to generate highly continuous reference genomes for five of them, including *Antheraea yamamai*, *A. assama*, *Clanis bilineata*, *Dictyoploca japonica*, and *Samia ricini* (Table S1). Hi-C sequencing data of *C. bilineata*, *A. yamamai*, *D. japonica* and the published chromosome assembly of *S. ricini* were used to further improve the contiguity of the four corresponding genomes, which enabled the generation of four chromosome-level genomes from the two respective focal families in the present study (Table S2 & S3). Genome assemblies, generated using Hi-C data, underwent rigorous contamination checks and were manually curated. For *A. yamamai* and *C. bilineata*, no contamination was detected. However, *D. japonica* exhibited minimal contamination, accounting for only 0.03% of the genome, which was primarily from protists and fungi and restricted to non-chromosomal sequences. The microbial contamination was checked and removed or corrected then (see Materials and Methods). These assembled genomes ranged in size from 453 Mb to 618 Mb, with the contig N50 sizes all reaching the megabase level (1.8 - 9.6 Mb) (Table S2). There are 31 chromosomes in *A. yamamai*, 31 chromosomes in *D. japonica*, 15 chromosomes in *S. ricini*, and 29 chromosomes *in C. bilineata*. There are 425 scaffolds in *A. assama*. Given that the genome of *R. fugax* was sequenced by a short-read Illumina sequencing platform, the number of scaffolds is 21,734, which is much more than the other species (Table S2). Sex chromosomes were also identified in the chromosome-level genomes using the sex chromosome of *B.mori* as reference (Figure S1 & Table S2).

Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis revealed that all these assemblies exhibited high levels of completeness in terms of gene content (94.6 - 97.9%) (Table S2). We have also generated a draft genome assembly, based on high-coverage Illumina sequencing data for *R. fugax* (Table S1). Although this 533 Mb assembled genome is relatively fragmented, with an N50 size of 158 Kb, the 96.4% complete BUSCO level suggests that this species' genome is as complete as other highly continuous genome assemblies in terms of gene content (Table S2). The proportion of annotated repetitive sequences in the genomes of six species ranges from 43.0% to 55.1% (Table S2). Based on the repeat-masked genome (Table S2), we integrated transcriptome data (Table S4), gene prediction algorithms, and homologs from related lepidopteran species to generate a consensus gene set for each of the six Bombycoidea species (Methods), resulting in 16,828 - 19,676 protein-coding genes across genomes (Table S2). We note that, during this study being conducted, the genomes of *A. yamamai* and *S. ricini* have been reported by other research groups.[8,56] Based on the quality comparisons, all our assemblies are of better genome completeness and contiguity than those previously published.

For comparative genomic analyses, we also included publicly available genomic data for five additional Bombycoidea species (*B. mori*, *B. mandarina*, *Manduca sexta*, *A. pernyi*, and *Hyles vespertilio*) (Table S2), forming a dataset consisting of 11 annotated whole genomes from the three main families of Bombycoidea. Despite the considerable variation in genome size (398.6 - 727.3 Mb), the necessary quality and reasonable features enable these genomes to be comparable for further cross-species analyses (Figure 1 & Table S2).

Genome evolution within Bombycoidea

To study the evolution of silk traits in the context of lineage divergence, we first investigated the phylogenetic relationships within Bombycoidea. To accomplish this, we characterized and grouped orthologs across all involved bombycoid species and two outgroup species (*Hyphantria cunea* and *Spodoptera litura*; Table S2). Overall, all bombycoid species were found to encode a large proportion of universal genes, ranging from 44.56% to 64.33% (Figure 1). Based on the 2,656 single-copy universal genes, we inferred a maximum-likelihood (ML) phylogenomic tree with each node being solidly supported. As expected, this ML tree clearly recovered the three main sublineages (termed SBS for Saturniidae, Bombycidae, and Sphingidae) within Bombycoidea (Figure 1A). Notably, the two bombycid species were placed as the evolutionary cousins,[57] while saturniid and sphingid species were clustered as sister clades sharing a common ancestor parallel to the bombycid sublineage (Figure 1A). The results of previous studies regarding the phylogeny across SBS are inconsistent.[58,59] The whole genome-based, well-supported phylogeny generated in the current study supports the previous transcriptome-based results and helps clarify the phylogeny across SBS.

We further inferred the divergence time across these taxa based on genomic alignment and calibrated using the inferred resource for timelines,[42] tracing the common ancestor of SBS back to approximately 60 million years ago (Mya) (Figure 1A). Further divergence between Saturniidae and Sphingidae occurred approximately 55 Mya, which was possibly the earliest stage when sphingid species ceased to spin silk cocoons. In addition, the massive speciation within Saturniidae was estimated to begin at approximately 21 Mya (Figure 1A).

Gene family evolution within Bombycoidea

Reconstruction of the ancestral state of cocooning characters within SBS inferred that non-cocooning in sphingid species may be at a relatively derived position (Figure S2). Therefore, we proposed that sphingid species have lost the ability to produce massive silk since the lineage divergence, rather than the two silk lineages, the Bombycidae and Saturniidae, evolved in parallel with superb capability of producing silk. We thus mainly performed comparisons between Saturniidae and Sphingidae, leaving Bombycidae as an outgroup, and began analyses by investigating the gene family evolution. Newly gained

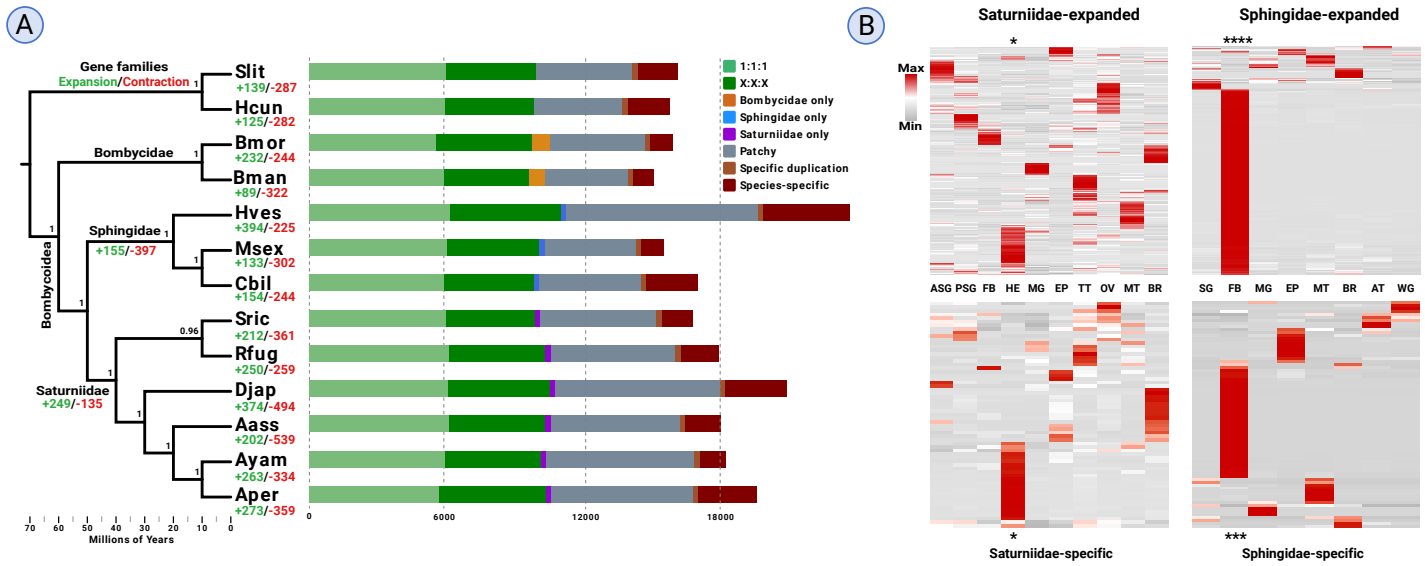


Figure 1. Genome evolution of Bombycoidea species (A) The maximum-likelihood phylogeny based on single-copy ortholog proteins in Bombycoidea species and two outgroup species (*Spodoptera litura* and *Hyphantria cunea*). The numbers labeled in the nodes of the tree show the bootstrap support value. The divergent time scale is shown at the bottom of the tree. Numbers of expanded and contracted gene families were shown in green and red, respectively. The gene set of each species was subdivided to represent different types of orthology cluster as indicated. 1:1:1, universal single-copy gene families across all examined species, but absence and/or duplication in at most one genome is tolerated; X:X:X, other universal genes; Bombycoidea only, gene families unique to Bombycoidea; Sphingidae only, gene families unique to Sphingidae; Patchy, all remaining orthologs; Specific duplication, duplicated genes specific to the species; Species-specific, species-specific genes. *Silt, S. litura; Hcun, H. cunea; Bman, Bombyx mandarina; Bmor, B. mori; Cbil, Clanis bilineata; Msex, Manduca sexta; Hves, Hyles vespertilio; Aper, Antheraea pernyi; Ayam, A. yamamai; Aass, A. assama; Djap, Dictyoploca japonica; Sric, Samia ricini; Rfug, Rhodinia fugax*. The abbreviations mean the same in the following captions. (B) Heatmap of tissue expression (FPKM) of Saturnidae-expanded genes in the representative species, *A. yamamai* (left) and Sphingidae-expanded genes in the representative species, *C. bilineata* (right). ASG, anterior silkgland; PSG, posterior silkgland; FB, fat body; HE, haemolymph; MG, midgut; EP, epidermis; TT, testis, OV, ovary; MT, Malpighian tubule; BR, brain; SG, silk gland; AT, antenna; WG, wing. The abbreviations means the same in the following captions.

genes, including lineage-specific expanded gene families, may confer unique traits in focal lineages.[60] We characterized 75 and 83 lineage-specific gene families, as well as 266 and 194 lineage-expanded gene families in Saturniidae and Sphingidae, respectively (Figure 1B & Datasets A-D). As expected, most of these newly evolved gene families were unavailable with known functions (Datasets A-D). To infer their potential functions, we sought to analyze the spatial expression profiles based on 20 tissue transcriptomes of two species, *A. yamamai* and *C. bilineata*, representing Saturniidae and Sphingidae, respectively (see Methods; Table S4). As a result, both lineage-specific present and expanded gene families of Saturniidae showed diverse expression patterns across *A. yamamai* tissues; of them, hemolymph was significantly overrepresented with genes under expression specialization (p = 0.02 for lineage-specific present and p = 0.03 for expanded gene families, chi-square test; Figure 1B; Datasets A & C). In contrast, newly evolved gene families in Sphingidae exhibited dramatic specialization in the fat body of *C. bilineata* (p = 1.4e-4 for lineage-specific present and p < 2.2e-16 for expanded gene families, chi-square test; Figure 1B, Datasets B & D).

For orthologous genes, we compared the molecular evolution signatures at the site level to identify those subject to lineage-specific selection. A branch likelihood ratio analysis was performed to characterize genes subject to natural selection in the focal lineage (see Methods; Datasets E-F). Functional annotation of the 113 genes under selection in Saturniidae revealed significant enrichment (hypergeometric test, adjusted p < 0.05) in 22 biological pathways (Table S5). Interestingly, many of these pathways have been reported to have functional relationships with the development of silk glands in *B. mori*, including ribosome, protein processing in endoplasmic reticulum, and DNA replication pathways (Table S5), which may involve in the processes of massive DNA replication and enormous protein accumulation that accompanies the dramatic growth of silk glands.[4] In addition, we found 11 genes of the Ras signaling pathway that were subject to selection in Saturniidae (Table S5 & Figure S3). Previous findings suggest that *Ras* activity may regulate the fibroin production and silk yield in *B. mori*.[61,62] The genes under selection in Saturniidae indicate a constraint on selectively advantageous silk production between the two discontinuous lineages that both spin great cocoons. By contrast, genes under selection in Sphingidae were only significantly enriched in one pathway, the regulation of actin cytoskeleton (Table S5), which has not been reported with known functional relationship with silk production.

Gene repertoires in the silk glands of Saturniidae and Sphingidae

Lepidopterans always produce silks in the larval silk glands, regardless of whether the species spins a silk cocoon for pupation or not.[1,63] We next compared the transcriptomes of the silk glands between Saturniidae and Sphingidae, still using *A. yamamai* and *C. bilineata* as the respective representatives. As expected, *A. yamamai* silk glands are morphologically remarkably larger than those of *C. bilineata* (Figure 2A). We characterized 451 and 226 genes with extremely high expression in the silk glands (termed SGH genes) of *A. yamamai* and *C. bilineata*, respectively (Figure 2B; Datasets G and H). Notably, only eight SGH genes were simultaneously highly expressed in the two species, indicating diverse gene repertoires between the species (Figure 2B). By cross-analyzing the spatial profiles (Table S4), we found that 35% of *A. yamamai* SGH orthologs are expressed in a specific tissue of *C. bilineata*, either the silk glands (5%) or any other tissues (56%), and 39% are universally expressed across different tissues of *C. bilineata* (Figure 2C). By contrast, the great majority (89%) of *C. bilineata* SGH genes exhibit universal expression in *A. yamamai* (Figure 2C). *A. yamamai* SGH genes were mainly enriched in metabolism-related pathways, particularly the amino acid metabolism (Figure 2D). Specifically, threonine is a branched-chain amino acid that can be converted into silk fibroin.[64] It has been suggested that genes in branched-chain amino acids are related to heavier silk glands and higher silk yield in the silkworm.[65] By contrast, *C. bilineata* SGH genes were significantly enriched in pathways related to degradation, such as autophagy, apoptosis, and lysosome (Figure 2D). These pathways may be associated with the morphological reduction in the size of silk gland in *C. bilineata* (Figure 2A), given their similar effects described in *B. mori*.[66-68]

Highly dynamic catalog of silk protein genes within Bombycoidea

We next investigated the evolution of silk protein genes across Bombycoidea species. Based on manual curation and local synteny surveys, we generated a full list of silk protein genes for bombycoid species and found that, except *fib-H*, all silk protein genes have undergone either loss or duplication across Bombycoidea species (Figure 3A & Tables S6-S9). The genomic locus encoding *fib-L* is completely absent in all investigated saturniids, although the neighboring genes are well-maintained in a conserved

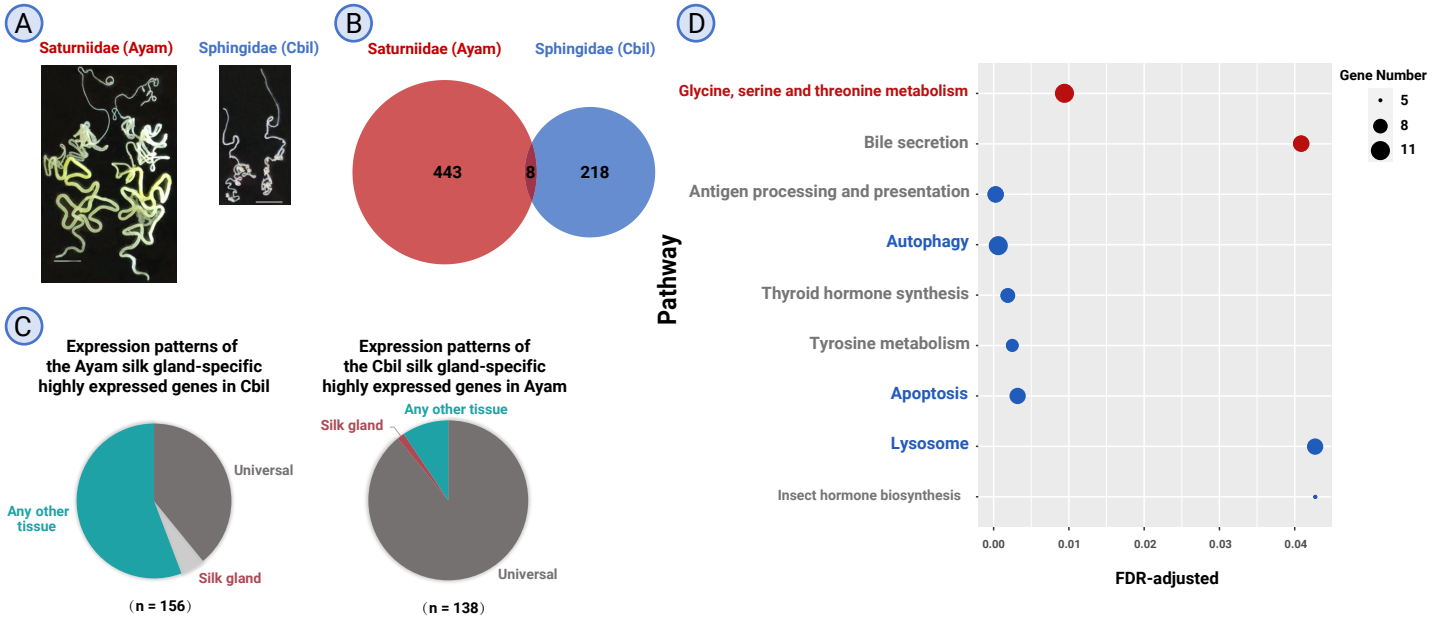


Figure 2. Gene repertoires of silkg lands between Saturniidae and Sphingidae (A) Silk glands of the representative species of Saturniidae, *A. yamamai* (Ayam;left) and those of Sphingidae, *C. bilineata* (Cbil;right). (B) Venn diagram of the silk gland-specific highly expressed genes in Ayam and Cbil. (C) Proportion of genes with different expression patterns of Ayam silk gland-specific highly expressed genes in Cbil (Left) and *vice versa* (right). D. Significantly enriched pathways of silk gland-specific highly expressed genes in Ayam (red dots) and Cbil (blue dots).

syntenic block across the three families (Figure 3B & Table S7). Thus, the loss of *fib-L* in saturniids is likely to attribute to the common ancestor of this lineage.

The evolution of *p25* is relatively complicated due to the duplication events. Here, potential *p25* homologs were characterized in all focal species and phylogenetic analysis was subsequently conducted to differentiate the orthologs (*p25*) from the retained homologs (*p25h*) (Figure S4 & Tables S8-S9). The deep split suggests ancient functional diversification between *p25* and *p25h*. Although *p25* is present in all investigated sphingids, it was found to be widely absent in non-*Antheraea* saturniid species (Figure 3A & Table S8). The genomic synteny surveys also supported this pattern and revealed frequent genomic rearrangements around the original locus encoding *p25* (Figure 3C & Table S8). Identification of *p25* in *Antheraea* corrects the previously reported loss in *A. peryni* (Dataset I);[34] However, the characterized copies of *Antheraea p25* are all phylogenetically distant from those of sphingids, bombycids, and even species outside Bombycoidea (Figure S4). Furthermore, *p25* exhibited extremely low expression in the silk glands of all *Antheraea* species (Figure S4), with the expression level (fragments per kilobase of transcript per million mapped reads, FPKM) ranging from 0 to 2.9. Based on this finding, we speculate that *p25* has been under pseudogenization in Antheraea that causes loss-of-function, as in other saturniids. On the other hand, we found that *p25h* has undergone frequent duplications in most species, with the greatest expansion in saturniids and bombycids (Table S9). Interestingly, the phylogenetic analysis and genomic arrangement revealed that the duplication respectively occurred in the common ancestor of each sublineage of Bombycoidea and was maintained along with further speciation within each sublineage (Figure S4).

Like *p25h*, we found a similar evolutionary pattern of genes encoding *sericin* within Bombycoidea. That is, there were frequent duplications of *sericin* across species and more copies in the two silk lineages, saturniids and bombycids, than in sphingids (Figure 3A & Table S10). These duplications included an independent cluster of *sericin* homologs that exclusively consisted of saturniid homologs and was closely related to bombycid *sericin 1* (Figure S5A). Interestingly, all members of this clade were highly expressed in the silk glands and exhibited relatively strong signatures of selection constraints (low dN/dS values, see Methods) (Figure S5A,B). Given that *B. mori sericin 1* encodes the main sericin components of cocoon silk proteins,[69] we suppose that the *sericin 1* orthologs play a common role in synthesizing silk proteins for cocoon spinning in Bombycoidea.

Extreme diversity of fibroin sequences within Bombycoidea

fib-H is the only silk protein gene that has not undergone loss or duplication across Bombycoidea; that is, all investigated species maintain a strictly single copy of *fib-H* in the genome (termed *fib-H* as fibroin as below). The complete gene structure of fibroin has been characterized in very few species due to its complex properties.[70,71] Here we were able to characterize the complete full-length fibroin sequences in multiple saturniid and sphingid species based on the highly-continuous genome assembly (Table S6 & Dataset I). As expected, the fibroin sequences exhibited signs of rapid evolution across species. With the exception of the 5′ and 3′ non-repetitive ends, most coding regions in all characterized fibroins were composed of repeated motifs that varied greatly in length, from 7,740 bp (*A. pernyi*) to 10,911 bp (*H. vespertilio*) (Figure 4A).

In all characterized bombycoid fibroins, we found that these iterated simple amino acid repeats are further organized into higher-order ensemble repeats (Figure 4B & Table S11). A previous study has reported 12 ensemble repeats in *B. mori* fibroin (Figure 4A), each consisting of alternate arrays of two subdomains but sharing variable lengths and organization.[9] In saturniid and sphingid fibroins, we also characterized ensemble repetitive domains that occurred in immediate arrays (n = 12-24, with the number varying among species; Figure 4B). Each of these ensemble domains consists of three or four alternate types of subdomains (Figure 4B & Table S11). We note that, unlike the submains of *B. mori* fibroin being filled with Gly-X units, all characterized subdomains in saturniid and sphingid fibroins are separated by a polyalanine block (Table S11). However, like *B. mori*, the ensemble repetitive domains are of highly variable lengths, orders, and compositions across fibroins of different saturniid species (Figure 4B). Moreover, these domains exhibit evident variations among very closely related species, such as *A. yamamai* and *A. pernyi* (Figure S6). By contrast, the organization of ensemble domains in sphingid fibroins is much more stringent, each being composed of almost all four alternate subdomains in the same order (Figure 4B). Moreover, the subdomains of sphingid fibroins are of a relatively higher level of sequence complexity compared with those of saturniid (Table S11). Overall, we observed an extreme diversity of bombycoid fibroin sequences, along with a divergent evolutionary scenario between the sublineages spinning silk cocoons, in which fibroins underwent a dynamic evolution with low sequence complexity, and the sublineage with "earthy" cocoons, in which fibroins evolved with great complexity and stringent organization.

As previously described in *B. mori* and even spiders,[1,9,72] the protein-coding

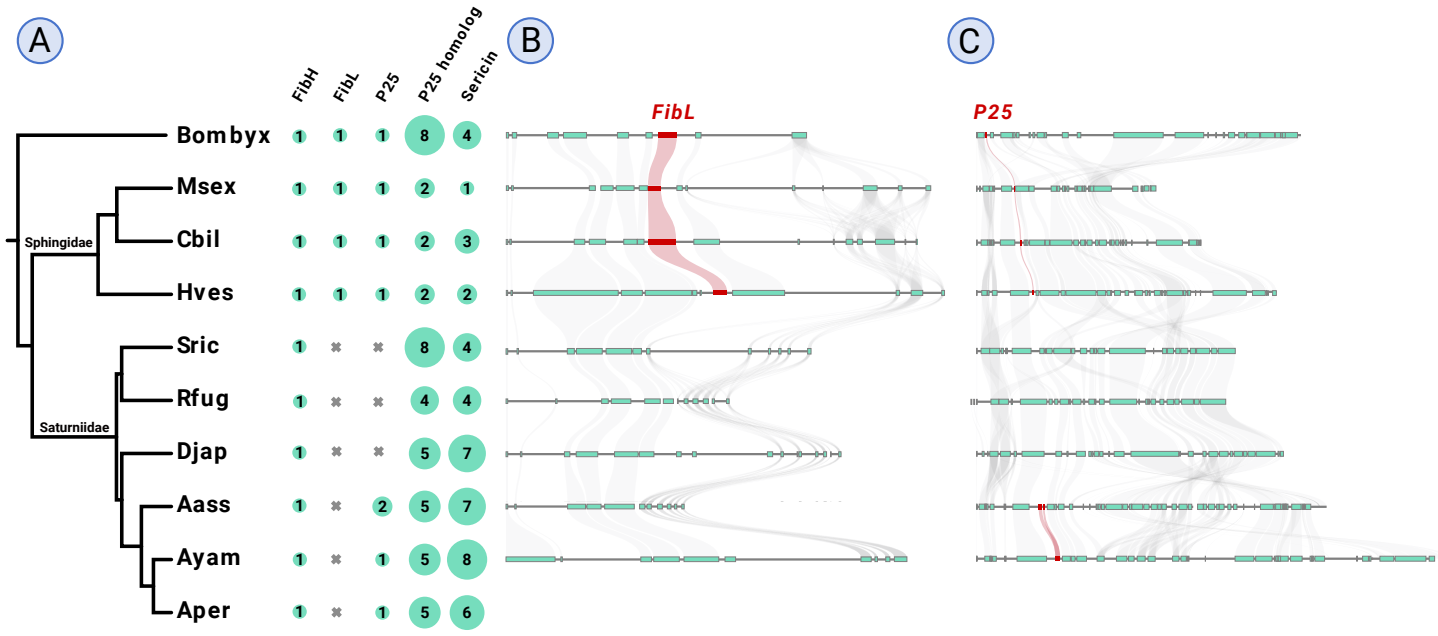


Figure 3. Catalog of silk protein genes and homologs (A) Number of gene copies of the silk protein genes and *p25* homologs in the Bombycoidea species. '×', gene loss. (B) Synteny of *Fib-L*, indicating that it is lost in Saturniidae species. (C) Synteny of *p25*, indicating that it is lost in some Saturniidae species while remaining in *A. assama* (Aass) and *A. yamamai* (Ayam).

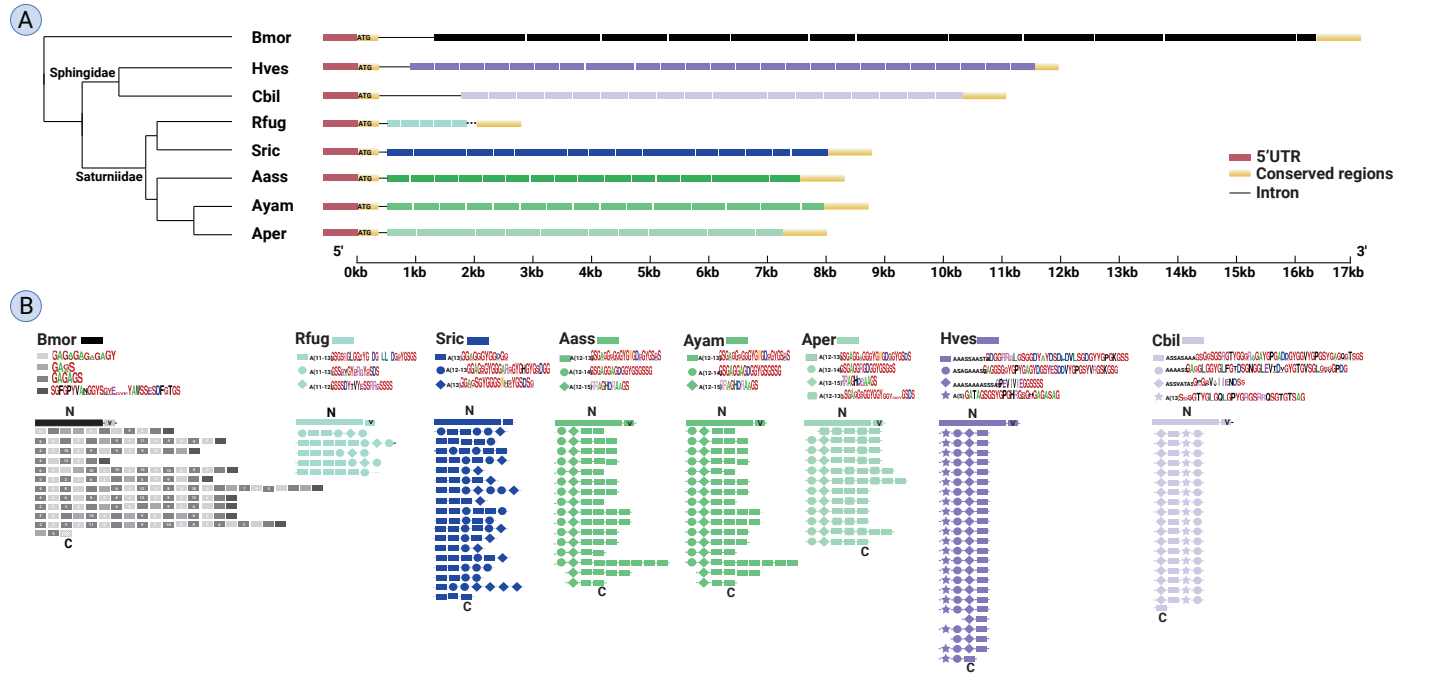


Figure 4. Dynamic evolution of fibroin sequences (A) Genomic structure of *fib-H* in the Bombycoidea species. Repetitive blocks in each species indicate the Repetitive sequences of *fib-H* . (B) Arrangement of the FibH. Sequence logo of the four domains of the repetitive sequences of FibH were shown in the upper panels and the arrangement of them are shown in lower panels.

sequences of saturniid and sphingid fibroins are remarkably enriched in three amino acid residues, i.e., alanine (Ala), glycine (Gly), and serine (Ser), which range from 59.4% to 88.2% (Figure 5 & Table S12). However, several different features shape the great diversity of fibroin sequences across the three families. First, the sum proportions of these amino acids in saturniid (64.8-83.8%) and bombycid (88.2%) fibroins are greater than those in sphingid ones (59.4-62.7%) (Figure 5 & Table S12), indicating a more skewed amino acid composition in the fibroins of the cocoon-spinning sublineages. Second, the most dominant residue of saturniid fibroins is alanine (31.3-45.4%), unlike *B.mori* fibroin, which is composed of 45.9% Gly (Figure 5 & Table S12). The evolved predominant residue largely correlates with the diverse repeat units of fibroins among different sublineages (Dataset J). In *B. mori* fibroin, the bulk of the repeat region is dominated by iterations of a Gly-X dipeptide where X represents Ala in 65%, Ser in 23%, and a small subset of other amino acids.[9] Instead, saturniid and sphingid fibroins are dominated by interspersed polyalanine blocks (Figure 4B, Figure 5, Table S11 & Dataset J).

Synchronization between the fibroin codon usage evolution and tRNA duplication

In addition to the diversity of amino acid compositions, the codon usage encoding the dominant residues varied greatly across the three families (Figure 4B & Table S12). For example, the most biased codon for Ala in *B.*

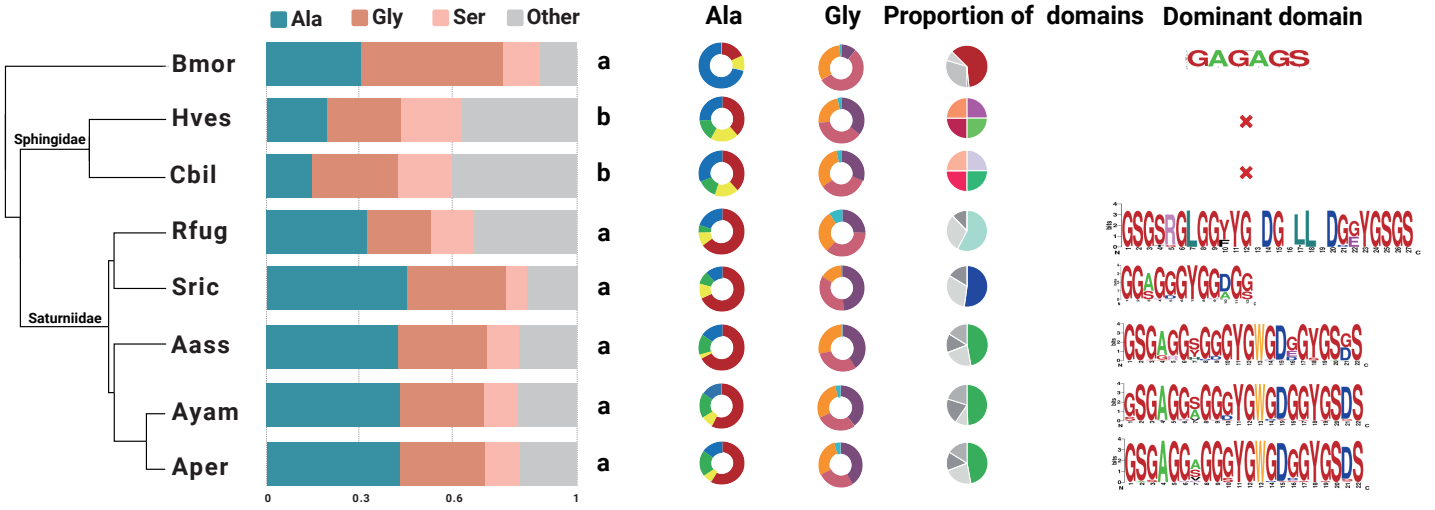


Figure 5. Analyses of the dominant amino acids of FibH Proportion of the dominant and other amino acids of FibH was shown in the left penal, proportion of the three dominant codons of Ala as well as Gly, were show in the middle two panels, and proportion of the four domains as well as the sequence logo of the most dominant domain of repetitive sequences of FibH, were shown in the two right panels. '×', there is no the most dominant domain.

mori fibroin is GCT (71%), in contrast to that in saturniids, which is GCA (58.2-74.3%) (Figure 5 & Table S12). Similarly, one of the preferential codons for Gly in saturniid fibroins, GGC, (15.6-31.9%) is sparsely used by *B. mori* (4.1%) (Figure 5 & Table S12). To support the massive synthesis of silk proteins during a limited period, the biased codon usage for fibroin amino acids is suggested to be correlated with the effective accumulation of tRNAs with cognate anticodons.[73,74] To test whether this adaptation was synchronized with the evolution of fibroin codon usages, we performed a genome-wide search of tRNAs in all investigated Bombycoidea species (Table S13). Interestingly, of all characterized $tRNA^{Ala}$ in saturniid genomes, the proportion of isotypes with the anticodon that recognizes the major codon of saturniid fibroins (GCA) (28.3 - 45.0%) was substantially higher than those in bombycid and sphingid ones (15.5 - 26.3%) (Table S13). As compensation, the ratio of isotypes recognizing GCU, the predominant codon in bombycids, decreased from 64.6% in *B. mori* to ~40-45% in saturniid species (Table S13). However, this co-evolution pattern was absent for $tRNA^{Gly}$, of which the isotype recognizing GGC/GGU dominated in all investigated species (all >50%) (Table S13). Given the rapid turnover of tRNA in the genome (Table S13), we hypothesize a major role of the alanine composition evolution in shaping the great diversity of fibroin sequences in Bombycoidea. Indeed, the alanine compositions are of the greatest variation range across Bombycoidea species (14.5 - 45.4%) (Figure 5 & Table S12).

Convergent regulation of fibroin expression between the silk lineages

Another prerequisite for spinning large cocoons is the efficient expression of fibroins under the spatial and temporal control. Considering the nutrition and energy required to produce fibroins, the expression of fibroins is strictly tissue-specific in most cases; in *B. mori* and some other lepidopterans, fibroins are specifically expressed in the posterior silk glands.[6,38,75] By investigating the spatial expression patterns across multiple tissues by both transcriptome sequencing and PCR (Table S4 & S14), we found, as expected, a remarkable expression specialization of fibroin in the posterior silk glands of *A. yamamai* and *B. mori* (Figure 6A). However, almost no expression of fibroin was detected in the silk glands of *C. bilineata* (Figures 6A & S7). The level of fibroin expression in the fat body of *C. bilineata* was much lower than those in the silk glands of *A. yamamai* and *B. mori* (Figure 6A). The tissue shift and reduction in the expression of fibroin may underlie the loss of silk-cocoon-producing behavior in Sphingidae.

The fibroin regulation mechanism in *B. mori* has been extensively studied, with several regulatory elements being characterized in the 5'-flanking sequence of fibroin.[76,77] All characterized bombycoid fibroins consist of a short first exon that is relatively conserved across species, an intron, and a long second exon encoding the major repetitive contents (Figure 4A). The flanking region upstream of the first exon is also relatively conserved (Figure S8). However, a segment replacement (−202 to −184 nt) located in the potential binding site of a main fibroin transcription factor, FMBP-3, attracted our attention (Figure S8). This replacement consistently and specifically occurred in the three sphingid species (Figure 6B), implying a potential role in the reduced and shifted expression of fibroin in these species. We first employed an electrophoretic mobility shift assay to test whether this replacement affected the binding activity (see Methods). The experiments demonstrated that the motifs (−202 to −184 nt) of *B. mori* and *A. yamamai* could both bind to the nuclear proteins extracted from the silk glands of *B. mori*, while there was a negative binding effect of the corresponding *C. bilineata* copy (Figure 6C & Figure S9). We further performed a dual-luciferase reporter assay to determine the influence of this replacement on the promoter activity of fibroin (see Methods). To accomplish this, the *B. mori* upstream sequence (−202 to +24 nt) containing this locus was cloned into the vector and used to evaluate the promoter activity in *B. mori* cells. In comparison to the original copy, the mutant clone with a corresponding replacement of the *C. bilineata* specific sequence (−181 to −166 nt) showed significantly decreased luciferase expression (Student's t-test, $p = 0.003$) (Figure 6D), which supports the possibility that the focal replacement in *C. bilineata* may cause reduced promoter activity in the expression of fibroin. Altogether, these results demonstrate that a specific replacement in the promoter region may lead to the shift and reduction in the expression of *fibroin* in sphingid species that spin earthy cocoons.

Previous studies have also indicated the potential role of the intron in regulating the expression of fibroin in *B. mori*.[77,78] Moreover, it is hypothesized that natural selection likely favors short introns in highly expressed genes, which may reduce the cost of transcription and facilitate more accurate splicing.[79,80] Interestingly, we found that most saturniid species maintain a short intron (120-150 bp) between the two exons of *fibroin* (Figure 4A), with the only exception being *D. japonica*, a special species that constructs hollow cocoons (Table 1). More strikingly, the intron lengths of saturniid fibroins are even much shorter than those of the bombycid one (971 bp) (Figure 4A). It was noteworthy that the expression of fibroin was much higher in the silk glands of *A. yamamai* (8-fold) than in those of *B. mori* (Figure 6A). This adaptation may explain the more efficient silk production and greater cocoon size in some wild silk moths of Saturniidae (Table 1).

DISCUSSIONS AND CONCLUSIONS

Divergent andconvergent adaptation in terms of silk-cocoon spinning

In this study, we elucidated the genomic and transcriptomic basis underlying the evolutionary divergence of large-silk-cocoon producing Saturniids from non-silk-cocoon producing Sphingids, and convergence with their cousin, the silk-cocoon producing Bombycids. This revealed how they adapted to spin extraordinarily large cocoons in natural environments.

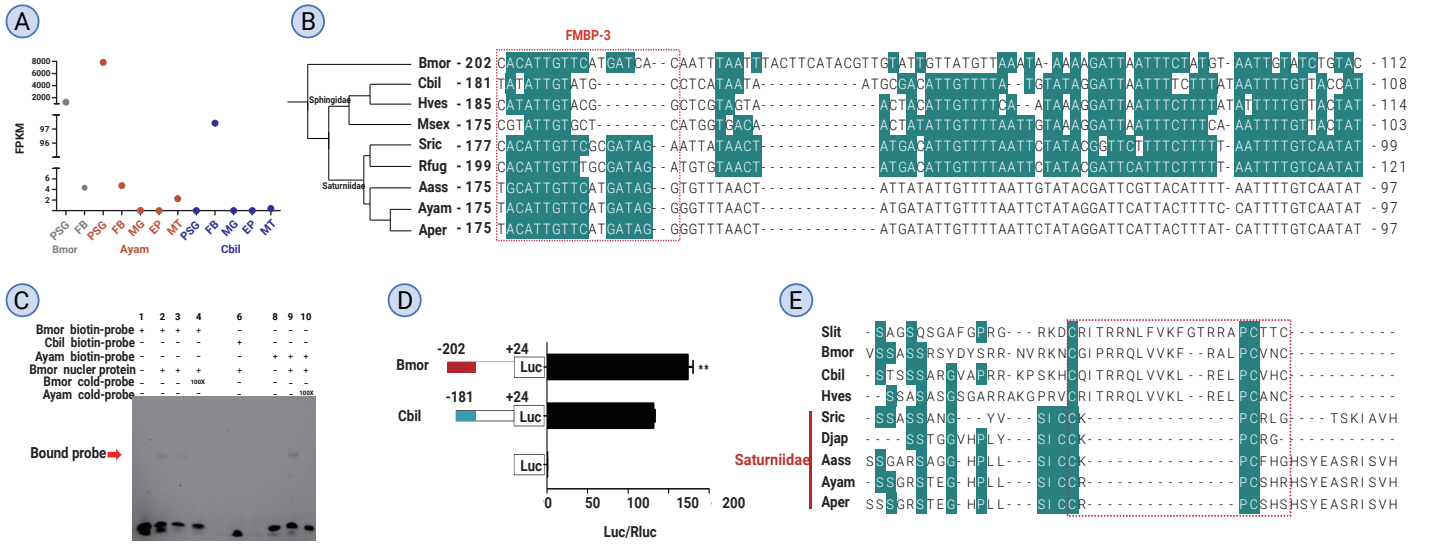


Figure 6. Convergent evolution of fibroin expression and binding (A) Expression level of *fib-H* in different tissues of *B. mori* (Bmor), *A. yamamai* (Ayam) and *C. bilineata* (Cbil), indicating *fib-H* is higly expressed in the silkgland of Bmor and Ayam but not expressed in the silkgland of Cbil. (B) Sequencing alignments of the upstream region of *fib-H* in the Bombycoidea species. The red dashed box indicates the putative cis-element, FMBP-3. (C) Electrophoretic mobility shift assay (EMSA) on binding activity the candidate cis-elements in the upstream region of *fib-H* in Bmor, Ayam for the nuclear proteins extracted from silk gland of Bmor. Probes were designed from of the upstream region between −202 to −184 nt of *Bmor fib-H* and *Ayam fib-H* as well as the upstream region between −181 to −166 nt of *Cbil fib-H*, respectively. (D) Comparison of *fib-H* promoter activities between Bmor and Cbil. Reporter plasmids containing the upstream region of *Bmor fib-H, and the* chimeric *Bmor fib-H* promoter with −202 to −184 nt of *Bmor fib-H* replaced by −181 to −166 nt of *Cbil fib-H*,were assayed. (E) Alignment of the C-terminal region of *fib-H* among Bombycoidea species and the outgroup species.

Saturniidae-specific presented and expanded genes/families displayed expression bias with significant preferences toward the hemolymph (Figure 1), a major organ for resource transportation and utilization (Blow and Douglas 2019), implying an active motivation in transporting resources for silk production. This pattern contrasts with Sphingidae-specific genes, which showed highly biased expression in the fat body, a typical organ for resource storage in pupae and adults. Secondly, genes subject to natural selection and silk gland specific expressed genes in Saturniidae, were both diverged in function to those in Sphingidae, and showed convergent evolution with *B. mori*. They are related to silk gland development or processes of amino acid metabolism. Generally, these features suggest active amino acid metabolism in the silk lineage, which is essential to meet the high demand of amino acids by the silk glands to synthesize massive silk proteins.[37,65]

Specifically, to achieve the massive, specialized expression of fibroin proteins in the silk glands, saturniid species maintain a short intron that may favor high expression as well as a relatively conserved 5′ coding and flanking region that harbors ancestral elements regulating the expression specialization in the silk glands (Figure 6). Finally, we observed synchronized evolution between the specific codon usage of the dominant amino acid residues in the fibroin sequence and the increased copies of the corresponding tRNA isotypes in the genome (Tables S12 and S13). This adaptation further favors fibroin synthesis via the accumulation of specific tRNAs for transferring required amino acids.

Taken together, these adaptive features in Saturniidae highlight the importance of an effective supply of amino acids and fibroin molecules in massive silk production. Importantly, most of these adaptations display a different pattern in the sister lineage, Sphingidae, which spins earthy cocoons, but display convergence with Bombycidae, a more distant lineage that spins silk cocoons as well (Figure 1A). Adaptation to massive silk production underlies a parallel path favoring silk production between the two silk lineages.

The detailed compositions of fibroins, in terms of amino acid residues and codon usage, also varied greatly between the two silk lineages (Figures 4-5). Bombycidae Fibroins are dominated by iterations of Gly-X dipeptides, while saturniid fibroins are enriched in polyalanine blocks (Figure 4 & Dataset I), which is similar to spider dragline spidroins.[6,9] Both polyalanine and Gly-Ala regions can form β-sheets that are critical for the high tensile strength of silk fibers, but polyalanine sheets have a higher binding energy than the poly-Gly-Ala sheets.[82] The different secondary structures may contribute to the higher toughness of saturniid silk fibers compared with *B. mori* silk (Table 1).[7,82]

Although it is argued that the drawing process rather than the sequence is more important to silk fiber formation and properties,[83] it should be noted that the organization of fibroin sequences is not completely free; the assembly of high-tensile-strength silk fibers requires numerous iterations of simple amino acid motifs, either polyalanine or poly-Gly-Ala, to constitute β-sheets. Due to the properties of low-complexity and high iterations of these β-sheet motifs, the fibrion gene is prone to errors in DNA replication and transcription, leading to variation and diversity.[84-87] In line with this, it can be observed that the cocoons of very closely related species, such as the three *Antheraea* species, vary greatly in terms of cocoon shape and the organization of repetitive units (Table 1 & Figure 4B).

Diversification of silk protein contents

Silk proteins are largely diversified between the two silk lineages, in terms of the architectures and sequences of the fibroin core, reflecting the broad convergence in usage of silk proteins as raw material for silk-spinning in nature.[1] The core fibroins in *B. mori* are reported to be composed of fibroin heavy chain, fibroin light chain, and fibrohexamerin (P25) and the two types of fibroins form a heterodimer via disulfide bonds.[75] However, genes encoding the light chain and P25 have been both lost in Saturniidae (Figure 3). It has suggested that two heavy fibroins form a homodimer via disulfide bonds in *A. yamamai* as an alternate assembly form, and that the binding is possibly determined by the three cysteine (Cys) residues in the C-terminal of heavy fibroin.[81] The full-length sequences of saturniid fibroins obtained in this study helped to further test this hypothesis. Multiple alignment of the C-terminal sequences across the three families clearly demonstrated two distinct arrangement patterns of the three Cys residues between saturniids and non-saturniids (Figure 5E), supporting the potential difference in fibroin assembly. We also noted that saturniid genomes maintained more copies of genes encoding other distantly related fibrohexamerins and sericins which coat fibroins in silk proteins (Figure 3). Whether these duplication events have led to other aspects of differences in silk fibers merits further investigation.

Silk use for ecological adaptation is subject to diversifying selection

The effects of using silk for insect ecological adaptation in terms of selective advantages and disadvantage remain largely unknown. Within Bombycoidea, the sister sublineages make different choices for pupation. That is, saturniid species produce massive silk to spin characteristic cocoons, while sphingid species use limited silk to spin phytodetritus or earth particles for

earthy cocoons to envelop pupae (Table 1 & Figure 1). Silks are externally secreted fibrous, proteinaceous materials, and massively silk use as an ecological adaptative trait generally comes at the expense of nutrient consumption. Nutrition in the form of amino acids and proteins and the utilization of nitrogen substances are particularly important for the silk-cocoon-producing species that synthesize massive silk proteins.[88,89]

According to the analyses in the present study, genes involved in amino acid metabolism dominate the gene repertoires of saturniid (*A. yamamai*) silk glands, while apoptosis-related genes dominate those of sphingid (*C. bilineata*) silk glands (Figure 2). The distinct gene repertoires in the silk glands and associated expression profiles reveal diverse functional roadmaps leading to the diverse cocoon traits between species. That is, the silk glands of Saturniidae may recruit a set of specialized genes to enable high-efficiency amino acid metabolism to optimize the nitrogen supply for silk production, while Sphingidae co-opts a small number of apoptosis-related genes in the silk glands which may promote self-degradation. Amino acid metabolism, specifically those involving glutamine, plays a crucial role in nutrient supply in organisms.[89,90] Interestingly, Compared to those in sphingids, genes in glutamine-involved amino acid metabolism displayed significantly elevated ratios of nonsynonymous-to-synonymous substitutions (dN/dS) in saturniids (Figure S10). The elevation of dN/dS is due to either the relaxation of selection constraints or the existence of balancing selection, both of which indicate the potential for evolutionary change.[91] Previous studies that contrasted the domestic silkworm with its ancestor *B. mandarina* also found genes involving in amino acid metabolism were enriched with variations, while divergent alleles were found to be limited to the upstream regions, suggesting that artificial selection may act on the regulation of gene expression, to promote silk production.[5,37] Domestication of the silkworm did not alter the feeding habits and a specialized diet of mulberry leaves is irreplaceable for silk yielding in *B. mori* and its wild ancestor *B. mandarina.* Factually, dominant hosts of Bombycidae species are limited to Moraceae, Theaceae, and Aceraceae species.[92] While in the wild silk lineages such as Saturniidae species or the non-cocoon-silk lineages such as Sphingidae, host plants were significantly broadened.[92,93] In this study, the Saturniidae and Sphingidae species favor different host plants from that of the Bombycidae species. Compared to these non-silk-cocooning Sphingids, Saturniids may face more significant challenges in optimizing amino acid metabolism for massive silk production. As expected in this study, the Saturniids showed elevated dN/dS ratios in the amino acid metabolism. Insects may encounter changing or even adverse host conditions; as a consequence, shifts and adaptations to alternative hosts may either promote the optimization of nutrition metabolism to maintain silk production or favor alternative strategies that use less silk or even avoid using silk entirely. Thus, we propose that using silk for ecological adaptation is subject to diversifying selection for wild organisms.

REFERENCES

1. Sutherland T.D., Young J.H., Weisman S., et al. (2010). Insect silk: One name, many materials. *Annu. Rev. Entomol.* **55**:171–188. DOI:10.1146/annurev-ento-112408-085401
2. Mitter C., Davis D.R. and Cummings, M.P. (2017). Phylogeny and evolution of lepidoptera. *Annu. Rev. Entomol.* **62**:265–283. DOI:10.1146/annurev-ento-031616-035125
3. Yu T.T., Chang Z., Dong Z.W., et al. (2022). A glimpse into the biodiversity of insects in Yunnan: An updated and annotated checklist of butterflies (Lepidoptera, Papilionoidea). *Zool. Res.* **43**:1009–1010. DOI:10.24272/j.issn.2095-8137.2022.313
4. Xia Q., Li S. and Feng, Q. (2014). Advances in silkworm studies accelerated by the genome sequencing of Bombyx mori. *Annu. Rev. Entomol.* **59**:513–536. DOI:10.1146/annurev-ento-011613-161940
5. Yi M.Y., Yang X., Wang M., et al. (2025). Consistency, distinction, and potential metabolic crosstalk of nitrogen mobilization-related genes in silk production and silk gland biology. *Zool. Res.* **46**:446–458. DOI:10.24272/j.issn.2095-8137.2024.391
6. Yukuhiro K., Sezutsu H., Tsubota T., et al. (2016). Insect silks and cocoons: Structural and molecular aspects. E. Cohen and B. Moussian (eds). *Extracellular Composite Matrices in Arthropods* (Springer International Publishing), pp. 515-555. DOI:10.1007/978-3-319-40740-1_14.
7. Dong Y., Dai F., Ren Y., et al. (2015). Comparative transcriptome analyses on silk glands of six silkmoths imply the genetic basis of silk structure and coloration. *BMC Genomics* **16**:203. DOI:10.1186/s12864-015-1420-9
8. Lee J., Nishiyama T., Shigenobu S., et al. (2021). The genome sequence of Samia ricini, a new model species of lepidopteran insect. *Mol. Ecol. Resour.* **2**:327–339. DOI:10.1111/1755-0998.13259
9. Zhou C.Z., Confalonieri F., Medina N., et al. (2000). Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic Acids Res.* **28**:2413–2419. DOI:10.1093/nar/28.12.2413
10. Sezutsu H. and Yukuhiro K. (2015). The complete nucleotide sequence of the Eri-silkworm (samia cynthia ricini) fibroin gene. *J. Insect Biotechnol.* **83**:59–70. DOI:10.11416/jibs.83.3_059
11. Koren S., Walenz B.P., Berlin K., et al. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**:722–736. DOI:10.1101/gr.215087.116
12. Liu H., Wu S., Li A., et al. (2021). SMARTdenovo:A de novo assembler using long noisy reads. *GigaByte* **2021**:gigabyte15. DOI:10.46471/gigabyte.15
13. Chin C.S., Alexander D.H., Marks P., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**:563–569. DOI:10.1038/nmeth.2474
14. Walker B.J., Abeel T., Shea T., et al. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963. DOI:10.1371/journal.pone.0112963
15. Roach M.J., Schmidt S.A. and Borneman A.R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**:460. DOI:10.1186/s12859-018-2485-7
16. Vaser R., Sovic I., Nagarajan N., et al. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**(5):737–746. DOI:10.1101/gr.214270.116
17. Kajitani R., Toshimoto K., Noguchi H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**:1384–1395. DOI:10.1101/gr.170720.113
18. Love R.R., Weisenfeld N.I., Jaffe D.B., et al. (2016). Evaluation of DISCOVAR de novo using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics* **17**:187. DOI:10.1186/s12864-016-2531-7
19. Mandric I. and Zelikovsky A. (2015). ScaffMatch: Scaffolding algorithm based on maximum weight matching. *Bioinformatics* **31**:2632–2638. DOI:10.1093/bioinformatics/btv211
20. Luo R., Liu B., Xie Y., et al. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**(1):18. DOI:10.1186/2047-217X-1-18
21. Durand N.C., Shamim M.S., Machol I., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**:95–98. DOI:10.1016/j.cels.2016.07.002
22. Dudchenko O., Batra S.S., Omer A.D., et al. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**:92–95. DOI:10.1126/science.aal3327
23. Alonge M., Lebeigle L., Kirsche M., et al. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**:258. DOI:10.1186/s13059-022-02823-7
24. Waterhouse R.M., Seppey M., Simao F.A., et al. (2019). Using BUSCO to assess insect genomic resources. *Methods Mol. Biol.* **1858**:59–74. DOI:10.1007/978-1-4939-8775-7_6
25. Li H. (2012). seqtk Toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk.
26. Wang Y., Tang H., Debarry J.D., et al. (2012). MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49. DOI:10.1093/nar/gkr1293
27. Tarailo-Graovac M. and Chen N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **25**:4.10.11–14.10.14. DOI:10.1002/0471250953.bi0410s25
28. Hoff K.J. and Stanke M. (2013). WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res.* **41**:W123-128. DOI:10.1093/nar/gkt418.
29. Dobin A., Davis C.A., Schlesinger F., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI:10.1093/bioinformatics/bts635
30. Pertea M., Pertea G.M., Antonescu C.M., et al. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**:290–295. DOI:10.1038/nbt.3122
31. Haas B.J., Salzberg S.L., Zhu W., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**:R7. DOI:10.1186/gb-2008-9-1-r7
32. Emms D.M. and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238. DOI:10.1186/s13059-019-1832-y
33. Lu F., Wei Z., Luo Y., et al. (2020). SilkDB 3.0: Visualizing and exploring multiple levels of data for silkworm. *Nucleic Acids Res.* **48**:D749-D755. DOI: 10.1093/nar/gkz919.
34. Duan J., Li Y., Du J., et al. (2020). A chromosome-scale genome assembly of *Antheraea pernyi* (Saturniidae, Lepidoptera). *Mol. Ecol. Resour.* **20**:1372–1383. DOI:10.1111/1755-0998.13199

35. Pippel M., Jebb D., Patzold F., et al. (2020). A highly contiguous genome assembly of the bat hawkmoth Hyles vespertilio (Lepidoptera: Sphingidae). *Gigascience* **9**. DOI:10.1093/gigascience/giaa001.
36. Kanost M.R., Arrese E.L., Cao X., et al. (2016). Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, Manduca sexta. *Insect Biochem. Mol. Biol.* **76**:118–147. DOI:10.1016/j.ibmb.2016.07.005
37. Xiang H., Liu X., Li M., et al. (2018). The evolutionary road from wild moth to domestic silkworm. *Nat. Ecol. Evol.* **2**:1268–1279. DOI:10.1038/s41559-018-0593-4
38. Wu N.N., Zhang S.F., Li X.W., et al. (2019). Fall webworm genomes yield insights into rapid adaptation of invasive species. *Nat. Ecol. Evol.* **3**:105–115. DOI:10.1038/s41559-018-0746-5
39. Cheng T., Wu J., Wu Y., et al. (2017). Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. *Nat. Ecol. Evol.* **1**:1747–1756. DOI:10.1038/s41559-017-0314-4
40. Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313. DOI:10.1093/bioinformatics/btu033
41. Sanderson M.J. (2003). R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**:301–302. DOI:10.1093/bioinformatics/19.2.301
42. Kumar S., Stecher G., Suleski M., et al. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**:1812–1819. DOI:10.1093/molbev/msx116
43. Li H. (2023). Protein-to-genome alignment with miniprot. *Bioinformatics* **39**:btad014. DOI:10.1093/bioinformatics/btad014
44. De Bie T., Cristianini N., Demuth J.P., et al. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271. DOI:10.1093/bioinformatics/btl097
45. Revell L.J. (2024). Phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**:e16505. DOI:10.7717/peerj.16505.
46. Li B. and Dewey, C.N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**:323. DOI:10.1186/1471-2105-12-323
47. Yanai I., Benjamin H., Shmoish M., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**:650–659. DOI:10.1093/bioinformatics/bti042
48. Yang Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556. DOI:10.1093/bioinformatics/13.5.555
49. Yang Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591. DOI:10.1093/molbev/msm088
50. Yang Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573. DOI:10.1093/oxfordjournals.molbev.a025957
51. Yang Z. and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917. DOI:10.1093/oxfordjournals.molbev.a004148
52. Yang, Z. Wong W.S. and Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**:1107–1118. DOI:10.1093/molbev/msi097
53. Zhang J., Nielsen R. and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**:2472–2479. DOI:10.1093/molbev/msi237
54. Sun X., Yang Q. and Xia, X. (2013). An improved implementation of effective number of codons (nc). *Mol. Biol. Evol.* **30**:191–196. DOI:10.1093/molbev/mss201
55. Chan P.P., Lin B.Y., Mak A.J., et al. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**:9077-9096. DOI:10.1093/nar/gkab688.
56. Kim S.R., Kwak W., Kim H., et al. (2018). Genome sequence of the Japanese oak silk moth, Antheraea yamamai: the first draft genome in the family Saturniidae. *Gigascience* **7**:1–11. DOI:10.1093/gigascience/gix113
57. Walker J.M., van der Heijden E.S.M., Maulana A., et al. (2024). Common misconceptions of speciation. *Evol. J. Linn. Soc.* **3**:kzae029. DOI:10.1093/evolinnean/kzae029
58. Zwick A., Regier J.C., Mitter C., et al. (2011). Increased gene sampling yields robust support for higher-level clades within Bombycoidea (Lepidoptera). *Syst. Entomol.* **36**:31–43. DOI:10.1111/j.1365-3113.2010.00543.x
59. Hamilton C.A., St Laurent R.A., Dexter K., et al. (2019). Phylogenomics resolves major relationships and reveals significant diversification rate shifts in the evolution of silk moths and relatives. *BMC Evol. Biol.* **19**:182. DOI:10.1186/s12862-019-1505-1
60. Yoshioka Y., Suzuki G., Zayasu Y., et al. (2022). Comparative genomics highlight the importance of lineage-specific gene families in evolutionary divergence of the coral genus, Montipora. *BMC Ecol. Evol.* **22**:71. DOI:10.1186/s12862-022-02023-8
61. Ma L., Xu H., Zhu J., et al. (2011). Ras1(CA) overexpression in the posterior silk gland improves silk yield. *Cell Res.* **21**:934–943. DOI:10.1038/cr.2011.36
62. Ma L., Ma Q., Li X., et al. (2014). Transcriptomic analysis of differentially expressed genes in the Ras1(CA)-overexpressed and wildtype posterior silk glands. *BMC Genomics* **15**:182. DOI:10.1186/1471-2164-15-182
63. Craig C.L. (1997). Evolution of arthropod silks. *Annu. Rev. Entomol.* **42**:231–267. DOI:10.1146/annurev.ento.42.1.231
64. Bricteux-Gregoire S., Dewandre A. and Florkin M. (1960). Contributions to silkworm biochemistry. *XVI. Conversion of threonine into the glycine, serine and alanine of silk fibroin. Arch. Int. Physiol. Biochim.* **68**:281–284. DOI:10.3109/13813456009083549
65. Sun L., Sun B., Chen L., et al. (2024). Identification of genes associated with the silk gland size using multi-omics in silkworm (Bombyx mori). *Insect Mol. Biol.* **33**:1–16. DOI:10.1111/imb.12870
66. Hu W., Chen Y., Lin Y., et al. (2019). Developmental and transcriptomic features characterize defects of silk gland growth and silk production in silkworm naked pupa mutant. *Insect Biochem. Mol. Biol.* **111**:103175. DOI:10.1016/j.ibmb.2019.05.010
67. Cui Y., Zhu Y.A., Lin Y.J., et al. (2018). New insight into the mechanism underlying the silk gland biological process by knocking out fibroin heavy chain in the silkworm. *Bmc Genomics* **19**. DOI:ARTN 21510.1186/s12864-018-4602-4.
68. Hu W., Wang X., Ma S., et al. (2021). CRISPR-mediated endogenous activation of fibroin heavy chain gene triggers cellular stress responses in *Bombyx mori* embryonic cells. *Insects* **12**:552. DOI:10.3390/insects12060552
69. Dong Z., Zhao P., Wang C., et al. (2013). Comparative proteomics reveal diverse functions and dynamic changes of *Bombyx mori* silk proteins spun from different development stages. *J. Proteome. Res.* **12**:5213–5222. DOI:10.1021/pr4005772
70. Hwang J.S., Lee J.S., Goo T.W., et al. (2001). Cloning of the fibroin gene from the oak silkworm, Antheraea yamamai and its complete sequence. *Biotechnol. Lett.* **23**:1321–1326. DOI:10.1023/A:1010542011150
71. Sezutsu H. and Yukuhiro K. (2000). Dynamic rearrangement within the *Antheraea pernyi* silk fibroin gene is associated with four types of repetitive units. *J. Mol. Evol.* **51**:329–338. DOI:10.1007/s002390010095
72. Zhou S.Y., Dong Q.L., Zhu K.S., et al. (2021). Long-read transcriptomic analysis of orb-weaving spider Araneus ventricosus indicates transcriptional diversity of spidroins. *Int. J. Biol. Macromol.* **168**:395–402. DOI:10.1016/j.ijbiomac.2020.11.182
73. Garel J. (1976). Quantitative adaptation of isoacceptor tRNAs to mRNA codons of alanine, glycine and serine. *Nature* **260**:805–806. DOI:10.1038/260805a0
74. Garel J.P., Mandel P., Chavancy G., et al. (1970). Functional adaptation of tRNAs to fibroin biosynthesis in the silkgland of Bombyx mori L. *FEBS Lett.* **7**:327–329. DOI:10.1016/0014-5793(70)80196-x
75. Inoue S., Tanaka K., Arisaka F., et al. (2000). Silk fibroin of Bombyx mori is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6 : 6 : 1 molar ratio. *J. Biol. Chem.* **275**:40517–40528. DOI:10.1074/jbc.M006897200
76. Shimizu K., Ogawa S., Hino R., et al. (2007). Structure and function of 5'-flanking regions of Bombyx mori fibroin heavy chain gene: identification of a novel transcription enhancing element with a homeodomain protein-binding motif. *Insect Biochem. Mol. Biol.* **37**:713–725. DOI:10.1016/j.ibmb.2007.03.016
77. Takiya S., Kokubo H. and Suzuki Y. (1997). Transcriptional regulatory elements in the upstream and intron of the fibroin gene bind three specific factors POU-M1, Bm Fkh and FMBP-1. *Biochem J* **321**:645–653. DOI:10.1042/bj3210645
78. Zhao A., Zhao T., Zhang Y., et al. (2010). New and highly efficient expression systems for expressing selectively foreign protein in the silk glands of transgenic silkworm. *Transgenic Res.* **19**:29–44. DOI:10.1007/s11248-009-9295-7
79. Castillo-Davis C.I., Mekhedov S.L., Hartl D.L., et al. (2002). Selection for short introns in highly expressed genes. *Nat. Genet.* **31**:415–418. DOI:10.1038/ng940
80. Pickrell J.K., Pai A.A., Gilad Y., et al. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *Plos Genet.* **6**:e1001236. DOI:10.1371/journal.pgen.1001236
81. Tanaka K. and Mizuno S. (2001). Homologues of fibroin L-chain and P25 of Bombyx mori are present in Dendrolimus spectabilis and Papilio xuthus but not detectable in *Antheraea yamamai*. *Insect Biochem. Mol. Biol.* **31**:665–677. DOI:10.1016/s0965-1748(00)00173-9
82. Hakimi O., Knight D.P., Vollrath F., et al. (2007). Spider and mulberry silkworm silks as compatible biomaterials. *Compos. Part B-Eng.* **38**:324–337. DOI:10.1016/j.compositesb.2006.06.012
83. Shao Z., and Vollrath F. (2002). Surprising strength of silkworm silk. *Nature* **418**:741. DOI:10.1038/418741a
84. Li X., Zhao L., Jiang H., et al. (2009). Short homologous sequences are strongly associated with the generation of chimeric RNAs in eukaryotes. *J. Mol. Evol.* **68**:56–65. DOI:10.1007/s00239-008-9187-0
85. Tautz D. and Renz M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* **12**:4127–4138. DOI:10.1093/nar/12.10.4127
86. Lower S.E., Dion-Cote A.M., Clark A.G., et al. (2019). Special Issue: Repetitive DNA sequences. *Genes (Basel)* **10**:896. DOI:10.3390/genes10110896
87. Lovett S.T. (2004). Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* **52**:1243–1253. DOI:10.1111/j.1365-2958.2004.04076.x
88. Osanai M., Okudaira M., Naito J., et al. (2000). Biosynthesis of L-alanine, a major amino acid of fibroin in *Samia cynthia ricini*. *Insect Biochem Mol. Biol.* **30**:225–232. DOI:10.1016/s0965-1748(99)00120-4

89. Shinbo H., Konno K. and Hirayama C. (1997). The pathway of ammonia assimilation in the silkworm, *Bombyx mori*. *J. Insect Physiol.* **43**:959–964. DOI:10.1016/s0022-1910(97)00045-0
90. Hu D., Wang W., Zhao X., et al. (2023). Expression pattern of glutaminase informs the dynamics of glutamine metabolism. *The Innovation Life* **3**:100128. DOI:10.59717/j.xinn-life.2024.100128
91. Hunt B.G., Ometto L., Wurm Y., et al. (2011). Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc. Natl. Acad. Sci. USA* **108**:15936–15941. DOI:10.1073/pnas.1104825108
92. Zhu H. and Wang, L. (1996). Lepidoptera: Bombycidae, saturniidae and Thyrididae. *Fauna Sinica Insecta*, A.S. (Science Press), pp:299.
93. Zhu H. and Wang L. (1997). Lepidoptera Sphingidae. *Fauna Sinica Insecta*, A.S. (Science Press), pp:436.

FUNDING AND ACKNOWLEDGMENTS

We thank Mr. Xiu-Liang Yang for his kind gift of the A. yamamai sample, Ms. Lizhi Wang and Ms. Xiang-Min Wei for their help in sampling, Drs. Yan-dong Ren and Wei Liu for their help in data analysis, Dr. Kang-Kang Niu for his help in the experiments. We thank Profs. Sheng Li and Qi-Li Feng for their constructive comments. Yong Cui specially thanks his parents, Mr. Xiao-Fang Cui and Mrs. Shun-Lian Zhou for helping to collect the D. japonica larvae and cocoons in the burning sun.This work was supported by the National Science Foundation of China (grants 32270458, 32021001, 32225008, 32070411) the Guangdong Science Foundation (grant 2023A1515010657), the Laboratory of Lingnan Modern Agriculture Project (NZ2021019). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

H.X., S.Z. and W.W. conceived the project. S.Z. and H.X. designed and supervised the studies. Z.J., X.B C.Y. and Y.L,Y.Z provided the samples. Y.C., S.Z., G.F. H,X. and B.C. performed the analyses. Y.C., M.Y and B.C performed the experiments. Y.C., S.Z. and H.X. wrote and revised the manuscript with input from all authors. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

ETHICAL STATEMENT AND PATIENT CONSENT

Not applicable.

DATA AND CODE AVAILABILITY

All genome assemblies involved in this study have been deposited into the NCBI database under the Bioproject PRJNA922366. Transcriptome data have been deposited in SRA under accessions SAMN36998913–SAMN36998937.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/j.xinn-life.2025.100144

LEAD CONTACT WEBSITE

https://mp.weixin.qq.com/s/oPKn_-4Vw53ANsw-UuDaFg